DIGITALCOMMONS @WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 3 | Issue 2 Article 3

11-1-2004

Assessing Treatment Effects in Randomized Longitudinal Two-Group Designs with Missing Observations

James Algina
University of Florida, algina@ufl.edu

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

Journal of Modern Applied Statistical Methods
November, 2004, Vol. 3, No 2, 271-287

1538 – 9472/04/$95.00

Assessing Treatment Effects in Randomized Longitudinal Two-Group Designs with Missing Observations

James Algina
University of
Florida

H. J. Keselman
University of Manitoba

SAS's PROC MIXED can be problematic when analyzing data from randomized longitudinal two-group designs when observations are missing over time. Overall (1996, 1999) and colleagues found a number of procedures that are effective in controlling the number of false positives (Type I errors) and are yet sensitive (powerful) to detect treatment effects. Two favorable methods incorporate time in study and baseline scores to model the missing data mechanism; one method was a single-stage PROC MIXED ANCOVA solution and the other was a two-stage endpoint analysis using the change scores as dependent scores. Because the two-stage approach can lack sensitivity to detect effects for certain missing data mechanisms, in this article we examined variations of the single-stage approach under conditions not considered by Overall et al., in order to assess the generality of the procedure's positive characteristics. The results indicate when and when not it is beneficial to include a baseline score as a covariate in the model. As well, we provide clarification regarding the merits of adopting an endpoint analysis as compared to the single-stage PROC MIXED procedure.

Keywords: Randomized designs, repeated measurements, missing data, PROC MIXED

Introduction

Overall and his colleagues (Ahn, Tonidandel, & Overall, 2000; Overall, Ahn, Shivakumar, & Kalburgi 1999, Overall, Ghasser, & Fiore, 1996) have provided very valuable information to biopharmaceutical researchers regarding the analysis of data from randomized longitudinal two-group designs. In particular, they compared various formulations of SASs (SAS, 1995) PROC MIXED program for analyzing effects in repeated measures designs when data are missing over time, finding that many formulations did not provide effective Type I error control, while others lacked power to detect treatment effects. In their studies, they found that a number of analysis of covariance (ANCOVA) analyses, using baseline scores and time in study as covariates, provided effective Type I error control and were, among the procedures compared, relatively powerful to detect treatment effects. In particular, they found that a single-stage PROC MIXED (1999, p. 208) and several two-stage analyses (1999, pp. 205-209) provided good results. Among the two-stage analyses, the endpoint analysis had the largest estimated power.

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.

Algina and Keselman (2003) however, compared the single-stage PROC MIXED analysis and endpoint analysis, as well as others presented in the literature, and found that though Overall et al.'s (1999) two-stage endpoint procedure had power similar to that of the other procedures when data were missing completely at random, it was lacking in power to detect treatment effects when data were not missing completely at random (See discussion below).

For example, the two-stage power value in one condition was .26, while the other procedures investigated had values clustered around .60. On the other hand, Overall et al.'s single-stage procedure controlled rates of Type I error and was the most powerful (or second next most powerful in one case) procedure among those procedures that were never liberal. Moreover, with regard to bias and sampling variability its values were not very different from bias and sampling variability for the other procedures that did not exhibit liberal rates of Type I error. Thus, in the investigation reported herein, we only examined modifications of the single-stage PROC MIXED procedure enumerated by Overall et al. (1999) as well as another method to be described.

The variations of the Overall et al. (1999) PROC MIXED procedure that we investigated are based on their acknowledgement that there was some concern regarding "the propriety of ... including the baseline scores as *both* linear covariate and as one of the repeated measurements to which a linear regression model was fitted" (p. 267). Given the very positive operating characteristics of their approach to the analysis of longitudinal data with missing observations we thought it important to further investigate their method of analysis by comparing PROC MIXED models that do and do not include a baseline score as both a covariate and repeated measurement in the analysis. In addition, we vary other conditions such as drop out mechanism, number of repeated measurements, and pattern of parameters in order to assess the operating characteristics of their procedure over conditions not yet examined in order to assess the generality of their findings.

Missing Data Mechanisms

To set the stage for our investigation we first discuss conditions under which data may be missing in randomized longitudinal two-group designs.

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. 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.

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

$$\mathbf{y}_{ik} = \mathbf{X}\beta_{ik} + \varepsilon_{ik}$$

where $\mathbf{y}'_{ik} = (Y_{i1k},\cdots,Y_{iJk})$, β_{ik} is an unobservable r-dimensional random vector, ε_{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 $t_1,\ldots,t_J$ indexes time of measurement. We assume $\varepsilon_{ik} \sim N(0,\sigma^2\mathbf{I}_J)$.

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 $\beta'_{ik} = (\beta_{i0k}\ \beta_{i1k})$. The between-subjects model for β_{ik} is

$$\begin{bmatrix}\beta_{i0k}\\ \beta_{i1k}\end{bmatrix}=\begin{bmatrix}1 & z & 0 & 0\\ 0 & 0 & 1 & z\end{bmatrix}\begin{bmatrix}\gamma_{00}\\ \gamma_{01}\\ \gamma_{10}\\ \gamma_{11}\end{bmatrix}+\begin{bmatrix}u_o\\ u_1\end{bmatrix} \qquad (1)$$

where $z=0$ for the first treatment and 1 for the second treatment. More compactly

$$\beta_{ik}=\mathbf{W}\gamma+\mathbf{u}.$$

We assume that $\mathbf{u}\sim N(\mathbf{0},\mathbf{D})$.

In randomized longitudinal two-group designs, 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. The design considered in this paper is a special case of the longitudinal design considered by Little (1995). 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.

In order to clarify missing data mechanisms, we employ a random coefficients selection model perspective to the analysis of missing data in longitudinal data. 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 let $\mathbf{y}_{ik}$ 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. According to Little (1995), in this approach the joint distribution of $\mathbf{y}_{ik}$, β_{ik}, and R_{ik} is factored as

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

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

$$(\mathbf{y}_{ik}\mid\mathbf{X},\mathbf{W},\beta_{ik})\sim N(\mathbf{W}\gamma+\mathbf{Xu},\sigma^2\mathbf{I}_J)$$

and

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

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

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

(See, e.g., 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 $\mathbf{y}_{ik}$, β_{ik}, $\mathbf{X}$ or $\mathbf{W}$. The missing data mechanism is called covariate dependent (CD) if the probability of a particular data point being missing does not depend on either $\mathbf{y}_{ik}$, β_{ik}:

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

The missing data mechanism is called missing at random (MAR) if

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

that is, the probability of a particular data point being missing does not depend on either $\mathbf{y}_{miss,ik}$ or β_{ik}. Following Verbeke and Molenberghs (2000, p. 213), a missing data mechanism that does not meet any of these criteria can be referred to as missing not at random (MNAR). Consistent estimates for γ can be obtained from the likelihood for $\mathbf{y}_{obs,ik}$ and R_{ik}. However if the data are MCAR, CD, 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, CD or MAR, but the missing data mechanism is non-ignorable if the mechanism is MNAR.

Hedeker and Gibbons (1997) noted that, frequently, missing data are related to performance or other characteristics of participants. (See 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 a study. It should be noted that to legitimately ignore the missing data mechanism for estimation purposes, not only must the data be missing at random, but also, the parameters of the missing data mechanism must be independent of the parameters of the data model (Schafer, 1997). This independence or distinctness of parameters is quite realistic in many contexts (e.g,, Schafer, 1997, p. 11-15). When the missing data mechanism is ignorable, numerical results can easily be obtained with commercially available software, e.g., the SAS (1995) PROC MIXED program (see Littell et al., 1996).

Overall et al.'s (1999) Approach

Overall and his colleagues (See Overall et al., 1999) investigated an ANCOVA approach using the baseline score on Y $\left(Y_{i1k}\right)$ and the number of available measurements for participant i as covariates. Their model is

$$Y_{ijk} = \beta_{0ik} + \beta_{1ik} t_j + \varepsilon_{ijk}$$
$$\beta_{0ik} = \lambda_{00} + \lambda_{01} J_{ik} + \lambda_{02} z + \lambda_{03} Y_{i1k} + u_{i0}$$
$$\beta_{1ik} = \lambda_{10} + \lambda_{12} z + u_{i1}.$$

PROC MIXED code (See Overall et al., 1999, p. 208) for the model is

```
proc mixed method=ml;
class id group;
model score=nrm scr1 group time time*group/solution;
random intercept time/type=un subject=id;
```

The variable nrm is the number of measurements (time in study) available for a participant. The variable scr1 is the baseline score. As Overall et al. (1999, p. 193) note "The covariates entered the PROC MIXED model statement in numeric form by being excluded from the class statement."

In this article, we compare 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 as a function of the covariates included in the model.

Methodology

Four methods of examining the group by time interaction effect in a randomized longitudinal two-group design were examined. Specifically, the methods (with their acronyms) were:

(1) PROC MIXED analysis that presumes the data are missing at random (PMMAR),
(2) Overall et al.'s (1999) PROC MIXED analysis that uses scr1 as a covariate (SCR1),
(3) Overall et al.'s (1999) PROC MIXED analysis that uses nrm as a covariate (NRM),
(4) Overall et al.'s (1999) PROC MIXED analysis that uses scr1 and nrm as covariates (SCR1&NRM).

It should be noted that PMMAR is Overall et al.'s procedure without any covariates.

We investigated three factors in our study: number of equally spaced levels of the repeated measures variable (5 and 9), missing data mechanism (MCAR, MAR and MNAR), and covariance structure for the repeated measures. (The variations on the covariance structure are presented when we describe the model we used to simulate the data.) Overall and his colleagues (See Ahn, Tonidandel and Overall, 2000; Overall et al., 1999; Overall et al., 1996) examined the group by time interaction effect in a 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. For examining generality of results, 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:

$$Y_{ijk} = \beta_{0i} + \beta_{1i} t_j + \varepsilon_{ijk} \qquad (2).$$

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 β_{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 [$\gamma_{11} = 0$, where γ_{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 ($\gamma_{11} = 4.5$) when there were nine occasions and 12.5 in treatment 2 and 4.5 in treatment 1 ($\gamma_{11} = 8$) 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 (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. The variance for the residuals, conditional on time, was 240. In half of the conditions 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. In the other half of the conditions we changed the covariance to 12.42 from -12.42 and retained all other features of the design. Changing the covariance for the slope and intercept changes the covariance structure for the repeated measures. It should be noted that Overall et al. assumed that there was no between-subjects variation in the subject-specific slopes. The correlation in their repeated measurements was due to the random intercept and correlation in the residuals in equation (2). Thus, we investigated the performance of Overall et al.'s procedure, as well as the alternatives, for different correlation structures than Overall et al. employed.

Overall et al. (1999) investigated three variations on the missing data mechanism, which they called completely random, treatment dependent, and treatment and baseline dependent. In each 30% of the simulated participants dropped out of the study. In the completely random condition dropping out was not related to scores on the repeated measures, time, or the treatment indicator. Thus the completely random condition meets the requirements for a MCAR mechanism. In the treatment dependent condition, dropping out was not related to scores on the repeated measures or time but was related to the treatment indicator: two-thirds of the dropouts came from the treatment group. The treatment dependent condition meets the requirements for a CD mechanism.

In the treatment and baseline dependent condition, missing data were related to the random effects for the intercept with dropouts from the treatment group coming from those that had a subject-specific intercept above the mean and dropouts from control group coming from those that had a subject-specific intercept below the mean. Thus, the treatment and baseline dependent condition employed a MNAR missing data mechanism.

In our study, 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. To select missing observations we used the following model

$$Z_{ijk} = \theta_{1j} + \theta_2\beta_{0i} + \theta_3\beta_{1i} + \theta_4 Y_{i(j-1)k} + \theta_5 Y_{ijk}.$$

An observation was set as missing if $U_{ijk} < \phi\left(Z_{ijk}\right)$ where U_{ijk} is a uniformly distributed random variable and ϕ is the standard normal distribution. The missing data mechanism is MCAR if $\theta_2 = \theta_3 = \theta_4 = \theta_5 = 0$, MAR if $\theta_2 = \theta_3 = \theta_5 = 0$ and MNAR if θ_2, θ_3, or θ_5 is not equal to zero. In one MNAR mechanism only θ_2 and θ_3 were not equal to

zero (MNAR-SI). In the other MNAR mechanism, only θ_5 was not equal to zero (MNAR-Y). The values of θ_{1j} were selected to give cumulative missing data rates between 30% and 40% at the ninth occasion. In all conditions missing data conformed to a monotone drop out pattern. That is, if a simulated respondent had missing data on occasion *j*, the respondent had missing data on all subsequent occasions. Thus we investigated the performance of Overall et al.'s procedure, as well as variations, for different missing data mechanisms than Overall et al. employed. In particular our MAR and MNAR conditions were different than those employed by Overall et al. (1999).

Figure 1 shows estimated proportions of participants remaining in the study at each occasion in the non-null condition with a negative correlation between the slope and intercept and 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 drop out 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 group 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 2500 times. When there were five time points, the drop out rates for the jth time point (j = 1, …, 5) were equal to the drop out rates for the jth time point in the design with nine time points. All hypothesis tests were conducted with a nominal alpha of .05.

Results

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. Power differences were negligible in the MCAR conditions and in the MAR conditions when the correlation between the slope and intercept was positive and very small in the MAR conditions when the correlation between the slope and intercept was negative. Larger differences emerged in the MNAR conditions and clearly indicated lower power for the PMMAR procedure than for the other procedures. Among the remaining procedures NRM is the most powerful, though the advantage is fairly small, ranging from about .004 to .061.

Table 3 contains means and standard deviations (empirical standard errors) of the estimates for the MCAR and MAR conditions when $\gamma_{11}=0$. Table 4 contains the same information for the MNAR conditions. When $\gamma_{11}=0$ none of the procedures had an average estimate that was significantly different from zero and, across all conditions, empirical standard errors were fairly similar.

Table 5 contains means and standard deviations of the estimates for the MCAR and MAR conditions when $\gamma_{11}\neq 0$. As expected from theory PMMAR produced unbiased estimators under the MCAR and MAR missing data mechanisms. The other procedures produced unbiased estimators in the MCAR conditions and in the MAR conditions when the correlation between the slope and intercept was positive. When the correlation between the slope and intercept was negative with MAR data, all procedures, except PMMAR (No/No) produced slightly biased estimators. The estimator for the NMR (No/Yes) procedure was less biased, although the difference was small. (Biased estimators are delineated in the tables in bold face type.)

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

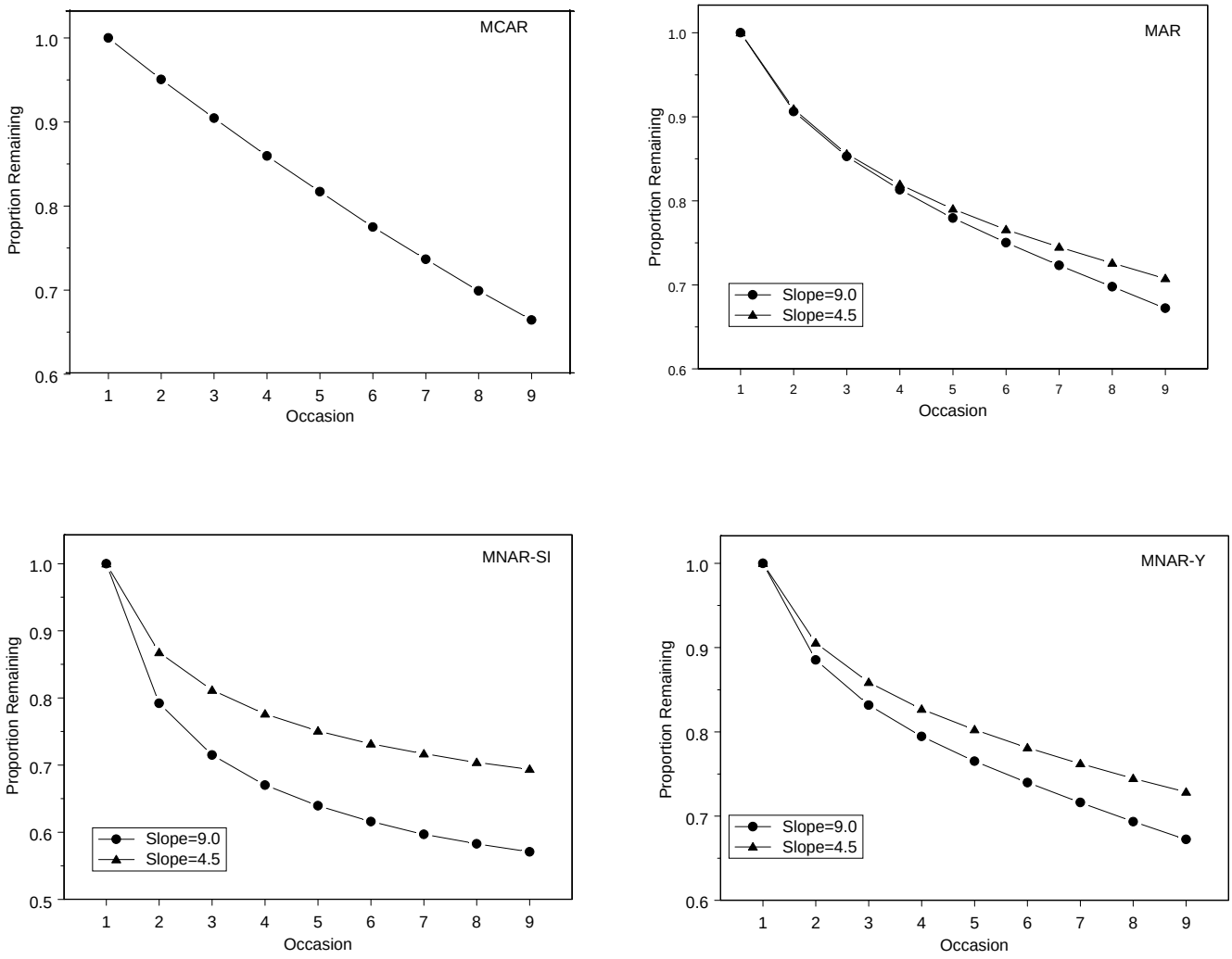

Table 1. Type I Error and Power Rates for MCAR and MAR Conditions.

Missing Data		Covariates		5-levels		9-levels	
Mechanism	Correlation	SCR1	NRM	Type I Error	Power	Type I Error	Power
MCAR	Positive	No	No	.046	.623	.056	.613
		Yes	No	.046	.624	.056	.607
		No	Yes	.046	.624	.056	.613
		Yes	Yes	.046	.625	.055	.608
	Negative	No	No	.052	.628	.054	.614
		Yes	No	.055	.623	.055	.601
		No	Yes	.052	.630	.054	.614
		Yes	Yes	.054	.623	.055	.601
MAR	Positive	No	No	.048	.592	.048	.604
		Yes	No	.052	.592	.049	.602
		No	Yes	.050	.594	.048	.604
		Yes	Yes	.053	.592	.047	.604
	Negative	No	No	.055	.616	.049	.607
		Yes	No	.052	.625	.045	.615
		No	Yes	.056	.632	.044	.631
		Yes	Yes	.052	.622	.044	.616

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

Missing Data		Covariates		5-levels		9-levels	
Mechanism	Correlation	SCR1	NRM	Type I Error	Power	Type I Error	Power
MNAR-SI	Positive	No	No	.056	.236	.061	.356
		Yes	No	.056	.363	.061	.414
		No	Yes	.055	.383	.060	.423
		Yes	Yes	.055	.363	.061	.414
	Negative	No	No	.052	.237	.056	.303
		Yes	No	.052	.418	.047	.453
		No	Yes	.049	.474	.049	.525
		Yes	Yes	.053	.420	.047	.464
MNAR-Y	Positive	No	No	.055	.507	.051	.553
		Yes	No	.055	.552	.051	.562
		No	Yes	.053	.556	.051	.571
		Yes	Yes	.054	.550	.051	.560
	Negative	No	No	.053	.514	.053	.519
		Yes	No	.050	.581	.052	.600
		No	Yes	.046	.604	.055	.622
		Yes	Yes	.050	.577	.052	.597

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

Missing Data		Covariates		5-levels		9-levels	
Mechanism	Correlation	SCR1	NRM	MEAN	SD	MEAN	SD
MCAR	Positive	No	No	0.068	3.544	0.044	2.056
		Yes	No	0.069	3.557	0.045	2.073
		No	Yes	0.067	3.544	0.044	2.056
		Yes	Yes	0.069	3.557	0.045	2.074
	Negative	No	No	0.047	3.518	0.023	2.018
		Yes	No	0.051	3.548	0.027	2.050
		No	Yes	0.048	3.518	0.023	2.019
		Yes	Yes	0.051	3.548	0.027	2.049
MAR	Positive	No	No	0.003	3.543	-0.016	2.066
		Yes	No	-0.017	3.579	-0.009	2.067
		No	Yes	-0.006	3.541	-0.016	2.037
		Yes	Yes	-0.015	3.568	-0.012	2.047
	Negative	No	No	-0.047	3.622	-.004	1.973
		Yes	No	-0.088	3.610	-.019	2.006
		No	Yes	-0.078	3.558	-.021	1.954
		Yes	Yes	-0.084	3.600	-.017	1.994

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

Missing Data		Covariates		5-levels		9-levels	
Mechanism	Correlation	SCR1	NRM	MEAN	SD	MEAN	SD
MNAR-SI	Positive	No	No	0.024	3.809	0.009	2.049
		Yes	No	0.022	3.810	-0.007	2.053
		No	Yes	0.024	3.765	0.004	2.039
		Yes	Yes	0.020	3.807	-0.006	2.049
	Negative	No	No	0.058	3.667	-0.010	2.015
		Yes	No	0.071	3.706	0.002	2.023
		No	Yes	0.054	3.610	-0.004	1.946
		Yes	Yes	0.071	3.702	0.004	2.018
MNAR-Y	Positive	No	No	-0.040	3.608	-0.054	1.929
		Yes	No	-0.035	3.601	-0.054	1.937
		No	Yes	-0.039	3.582	-0.053	1.928
		Yes	Yes	-0.036	3.599	-0.056	1.935
	Negative	No	No	-0.087	3.564	-0.022	2.037
		Yes	No	-0.097	3.506	-0.026	1.990
		No	Yes	-0.094	3.435	-0.020	1.954
		Yes	Yes	-0.098	3.503	-0.026	1.985

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

Missing Data		Covariates		5-levels		9-levels	
Mechanism	Correlation	SCR1	NRM	MEAN	SD	MEAN	SD
MCAR	Positive	No	No	8.080	3.523	4.573	1.996
		Yes	No	8.073	3.527	4.569	2.012
		No	Yes	8.081	3.523	4.573	1.996
		Yes	Yes	8.073	3.527	4.570	2.012
	Negative	No	No	8.080	3.513	4.500	2.073
		Yes	No	8.064	3.544	4.506	2.106
		No	Yes	8.083	3.514	4.501	2.073
		Yes	Yes	8.064	3.544	4.505	2.105
MAR	Positive	No	No	7.989	3.595	4.546	2.058
		Yes	No	8.037	3.630	4.530	2.060
		No	Yes	7.992	3.595	4.499	2.039
		Yes	Yes	8.018	3.620	4.501	2.040
	Negative	No	No	8.109	3.657	4.493	2.053
		Yes	No	**8.279**	3.670	**4.597**	2.054
		No	Yes	**8.199**	3.625	4.550	2.001
		Yes	Yes	**8.261**	3.661	**4.574**	2.043

Table 6 contains means and standard deviations of these estimates for the MNAR conditions when $\gamma_{11} \neq 0$. In these conditions the PMMAR (No/No) estimator was clearly more biased than were the other estimators. In most conditions the NRM estimator was less biased than were the SCR1 (Yes/No) and SCR1&NRM (Yes/Yes) estimator, though in many conditions the differences among the three procedures were negligible.

Discussion

We compared the performance of four data analysis procedures, which varied in terms of the covariates employed: no covariates, SCR1, NRM, and SCR1 and NRM. As expected from theory, when the missing data mechanism was MCAR or MAR there was no advantage to including SCR1, NRM, or both in the model. However, including SCR1 and/or NRM did not have a negative impact on the results. For the MNAR missing mechanisms including SCR1 and/or NRM improved power and reduced bias relative to the analysis without covariates. However, including SCR1 in addition to NRM did not enhance power or reduce bias relative to including only NRM as a covariate. And in some conditions including only NRM did enhance power and or reduce bias relative to the analyses that included SCR1 in addition to or in place of NRM.

Additional Results

Given our results and the fact that Overall et al. (1999) used both NRM and SCR1 as covariates and that Ahn et al. (2000) used only SCR1 as a covariate in their PROC MIXED analyses that included a random statement, but not a repeated statement, the question arises as to when is it necessary to employ SCR1 or SCR1 and NRM as covariates. To explore this question we simulated data using the treatment and baseline dependent missing data mechanism employed by Overall and his colleagues. In Overall et al. and Ahn et al. there was no between-subject random variation in the subject-specific slopes; accordingly, we included conditions like those studied by Overall and his colleagues as well as conditions in which there was between-subject random variation in the subject-specific slopes. For the latter analyses

$$\mathbf{D} = \begin{bmatrix} 15.21 & 0.00 \\ 0.00 & 82.81 \end{bmatrix}$$

and accordingly the treatment and baseline dependent missing data mechanism does not result in indirect selection on the slope, as would occur if the covariance between the subject-specific slopes and intercepts were non-zero. We refer to the treatment and baseline dependent conditions without slope variation as MNAR-I.NSV since the missing data mechanism is MNAR; that is, the missing data indicator is dependent on the intercept, and, there is no slope variations. The other conditions are referred to as MNAR-I.SV. Type I error rates and power results are presented in Table 7. The results indicate that when the probability of missing data depends on the subject-specific intercept, but not the slope, it is essential to control for SCR1 (baseline score) and the addition of NRM (number of repeated measurements) does not enhance control of the Type I error rate or power.

Results in Overall et al. (1999) suggest that a two-stage endpoint analysis is more powerful than the PROC MIXED analysis that includes SCR1 and NRM. Results in Ahn et al. (2000) suggest that the endpoint analysis is more powerful than the PROC MIXED analysis that includes SCR1 only. As noted in our introduction, Algina and Keselman (2003) did not find the endpoint analysis to be more powerful than the other procedures in the study. However, as noted above, Algina and Keselman simulated data with random variation in the subject-specific slopes, but Overall and his colleagues did not. To determine whether random variation in the subject-specific slopes accounts for the results with regard to the endpoint analysis, we estimated Type I error rates and power under the two MNAR-I missing data mechanisms. In all endpoint analyses both SCR1 and NRM were included as covariates. Results, shown in Table 7, indicate that the endpoint analysis controlled the Type I error rate

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

Missing Data		Covariates		5-levels		9-levels	
Mechanism	Correlation	SCR1	NRM	MEAN	SD	MEAN	SD
MNAR-SI	Positive	No	No	**5.005**	4.000	**3.279**	2.090
		Yes	No	**6.469**	3.988	**3.587**	2.083
		No	Yes	**6.644**	3.933	**3.610**	2.066
		Yes	Yes	**6.480**	3.984	**3.575**	2.078
	Negative	No	No	**4.779**	4.079	**3.041**	2.164
		Yes	No	**6.972**	4.122	**3.909**	2.190
		No	Yes	**7.392**	4.001	**4.077**	2.086
		Yes	Yes	**7.033**	4.116	**3.939**	2.180
MNAR-Y	Positive	No	No	**7.218**	3.679	**4.084**	1.972
		Yes	No	**7.532**	3.659	**4.143**	1.970
		No	Yes	**7.563**	3.644	**4.133**	1.964
		Yes	Yes	**7.520**	3.657	**4.123**	1.969
	Negative	No	No	**7.303**	3.661	**4.049**	2.060
		Yes	No	**7.839**	3.605	**4.398**	2.030
		No	Yes	7.950	3.546	4.434	1.985
		Yes	Yes	**7.846**	3.602	**4.394**	2.028

Table 7. Type I Error and Power Rates for MNAR-I Conditions.

Missing Data Mechanism	Data Analysis	Covariates SCR1	Covariates NRM	5-levels Type I Error	5-levels Power	9-levels Type I Error	9-levels Power
MNAR-I.NSV	Proc Mixed	No	No	.072	.521	.092	.714
		Yes	No	.053	.698	.058	.909
		No	Yes	.070	.525	.091	.717
		Yes	Yes	.053	.698	.059	.909
	Endpoint	Yes	Yes	.056	.522	.050	.557
MNAR-I.SV	Proc Mixed	No	No	.068	.477	.080	.508
		Yes	No	.054	.628	.056	.666
		No	Yes	.067	.480	.079	.510
		Yes	Yes	.055	.628	.056	.665
	Endpoint	Yes	Yes	.064	.461	.057	.383

Table 8. Type I Error and Power Rates for CD Conditions.

Missing Data Mechanism	Data Analysis	Covariates SCR1	Covariates NRM	5-levels Type I Error	5-levels Power	9-levels Type I Error	9-levels Power
CD.NSV	Proc Mixed	No	No	.053	.676	.046	.882
		Yes	No	.054	.688	.049	.879
		No	Yes	.054	.680	.046	.883
		Yes	Yes	.054	.686	.049	.878
	Endpoint	Yes	Yes	.050	.602	.051	.616
CD.SV	Proc Mixed	No	No	.052	.635	.058	.605
		Yes	No	.052	.626	.060	.598
		No	Yes	.052	.634	.057	.603
		Yes	Yes	.052	.626	.060	.598
	Endpoint	Yes	Yes	.051	.568	.054	.494

in all conditions, but was not more powerful than the PROC MIXED analysis that controlled the Type I error rate.

Algina and Keselman (2003) also did not include a treatment dependent drop-out condition like that included in Overall et al. (1999) and Ahn et al. (2000) and this may account for differences in terms of the endpoint analysis. Table 8 contains Type I error rates and power for a treatment dependent drop out condition like that included in Overall et al. and Ahn et al. We refer to this condition as CD. In CD conditions with between-subject random variation in the subject-specific slopes (CD.SV) reported in Table 8

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

However, we also conducted simulations with $D_{12} = 12.42$ and $D_{12} = 0.00$ and the general pattern of results was the same as those reported in Table 8: the endpoint analysis controls the Type I error rate, but was not more powerful than the PROC MIXED analyses.

Conclusion

In summary, we believe our results provide some clarification to the findings reported by Overall et al (1999) and Ahn et al. (2000), clarification, we believe, that adds to the importance of their contributions to the literature regarding the analysis of missing data in randomized longitudinal two-group designs.

First, with the regard to the controversy of including baseline scores as both independent and dependent variables in the analysis, our results show that it is not always necessary to include the baseline score as a covariate. Except when the distribution of the missing data depended exclusively on the subject-specific intercept, neither Type I error control nor power to detect effects was enhanced by including baseline as a covariate in addition to specifying the number of repeated measurements as a covariate in the model. However, it is also true that including the baseline as an additional covariate did not detract from control of the Type I error rate and detracted noticeably from power only when the probability of missing data depended on the subject-specific slopes and intercepts.

When the probability of a missing value depended on the subject-specific intercept, our results, along with those reported by Overall et al. (1999) indicate that the baseline score should be specified as a covariate in the model. In this case, however, no additional gains in terms of Type I error control or power to detect effects, is acquired by including as a second covariate number of repeated measurements.

Lastly, the findings from our study indicate that an endpoint analysis need not be more powerful than the single-stage PROC MIXED analysis presented by Overall et al. (1999). That is, like Overall et al., we also found the endpoint analysis to be effective in controlling the rate of Type I error and to have power similar to that for the single-stage procedures when data are not missing at random. However, our findings indicate that the single-stage strategy was never less powerful than the two-stage endpoint analysis and was in some cases substantially more powerful. In general, our recommendation for analysis would be to include both covariates in Overall et al.'s single-stage procedure.

Reference

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*, 265-286.

Algina, J. Keselman, H. J, & Othman, A. R. (2003). Analyzing group by time effects in longitudinal two-group randomized trial designs with missing data. *Journal of Modern Applied Statistical Methods*, *2*, 50-72.

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. (1995). Modeling the drop-out mechanism in repeated-measures

studies. J*ournal of the American Statistical Association, 90*, 1112-1121.

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., & 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.