STATISTICS IN MEDICINE
Statist. Med. **18**, 2465–2478 (1999)

ON SUMMARY MEASURES ANALYSIS OF THE LINEAR MIXED EFFECTS MODEL FOR REPEATED MEASURES WHEN DATA ARE NOT MISSING COMPLETELY AT RANDOM

RODERICK J. LITTLE* AND TRIVELLORE RAGHUNATHAN

Department of Biostatistics and Institute for Social Research, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.

SUMMARY

Subjects often drop out of longitudinal studies prematurely, yielding unbalanced data with unequal numbers of measures for each subject. A simple and convenient approach to analysis is to develop summary measures for each individual and then regress the summary measures on between-subject covariates. We examine properties of this approach in the context of the linear mixed effects model when the data are not missing completely at random, in the sense that drop-out depends on the values of the repeated measures after conditioning on fixed covariates. The approach is compared with likelihood-based approaches that model the vector of repeated measures for each individual. Methods are compared by simulation for the case where repeated measures over time are linear and can be summarized by a slope and intercept for each individual. Our simulations suggest that summary measures analysis based on the slopes alone is comparable to full maximum likelihood when the data are missing completely at random but is markedly inferior when the data are not missing completely at random. Analysis discarding the incomplete cases is even worse, with large biases and very poor confidence coverage. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

Many longitudinal studies suffer from attrition, that is, subject dropping out prematurely. Examples include panel surveys or cohort studies, and clinical trials with designs that involve repeated measurements to chart the course of a disease. The resulting data are unbalanced with unequal numbers of measures for each subject. Mixed linear effects models with normal errors[1–3] provide a flexible tool for analysing unbalanced longitudinal data, and software for maximum likelihood estimation under these models is now widely available to practitioners.[4,5] These analysis tools are valuable in that they incorporate all the available information in the data, and they can reduce or even eliminate the bias resulting from an analysis confined to the complete cases. However, estimates from these models assume that the missing data are missing at

* Correspondence to: Roderick J. Little, Department of Biostatistics and Institute for Social Research, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A. E-mail: rlittle@umich.edu

Contract/grant sponsor: National Science Foundation

CCC 0277–6715/99/182465–14$17.50
Copyright © 1999 John Wiley & Sons, Ltd.

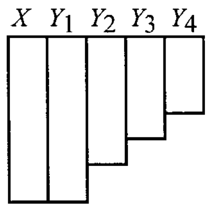

Figure 1. Schematic of a monotone missing data pattern, with X representing covariates, $Y_1, \dots, Y_4$ repeated measures at four time points, and blocks representing data

random, in Rubin's[6] sense. A number of methods that provide for drop-out processes that are not missing at random have also been proposed.[7-15] Little[16] reviews these methods within a unified framework based on likelihood inference for models for the data and the drop-out mechanism.

A drawback of these approaches is that they require specification of a full model for the vector of repeated measures for each individual, and a model for the drop-out mechanism if it is not missing at random. A simple and intuitive alternative approach is summary measures analysis; one or more measures of interest are formulated, these measures are estimated for each subject from their set of repeated measures, and the resulting summary measures are then regressed on subject-level characteristics. This approach may be less efficient than an analysis based on a full statistical model for the repeated measures, but it has potential advantages in terms of simplicity and reduced modelling assumptions, and it is very commonly applied in practice. Naïve applications of the summary measures approaches do not take into account the fact that the measures have differential precision when based on variable numbers of repeated measures. We study here a relatively refined version of the approach that takes this differential precision into account in the context of a linear mixed effects model.

We compare the summary measures approach with inference based on a full model for the repeated measures ignoring the missing-data mechanism, using both maximum likelihood and the method of moments to estimate the variances. Our main interest is in comparisons when the data are not missing completely at random, in the sense that missingness depends on values of the repeated measures. For concreteness we consider linear models where the repeated measures are regressed on time, and the summary measures are an estimated slope and intercept for each individual. Specific questions we address are:

(a) What assumptions about the missing data mechanism are implied by the summary measures approach?
(b) What are the relative merits of multivariate and univariate summary measures analyses when the data are not missing completely at random?
(c) Under what circumstances is a full likelihood-based analysis based on a model for the vector of repeated measures preferable to a summary measures analysis? Are there practically interesting situations where the summary measures analysis dominates an analysis based on the full model?
(d) How important is it to estimate the variance parameters by maximum likelihood, rather than by simpler approaches based on the method of moments?

Copyright © 1999 John Wiley & Sons, Ltd.

2. MODELS FOR THE REPEATED MEASURES AND DROP-OUT MECHANISM

We adopt the following general model for repeated-measures data with drop-outs.[16] Suppose that the data for subject i consist of up to K repeated measures on outcome variables $Y_1, \ldots, Y_K$. The data form a monotone pattern as displayed in Figure 1, so that if Y_j is missing then Y_k is missing for all $k > j$. For subjects $i = 1, \ldots, n$, let $y_i = (y_{i1}, \ldots, y_{iK})$ be a $(1 \times K)$ complete-data vector of outcomes for subject i, possibly incompletely observed. We write $y_i = (y_{\text{obs},i}, y_{\text{mis},i})$, where $y_{\text{obs},i}$ = observed part of y_i, $y_{\text{mis},i}$ = missing part of y_i. Note that lower-case y_i denotes rows (cases) and upper-case Y_k denotes columns (variables).

X_i denotes fixed covariates or design matrices, including times of measurement $(t_{i1}, \ldots, t_{iK})$ measured from some meaningful baseline, for example, calendar time or start of treatment. These variables are assumed fully observed. R_i is the missing-data indicator, indexing complete and incomplete patterns of data. Specifically, let $R_i = 0$ for complete cases, and let $R_i = k$ if a subject drops out between the $(k-1)$th and kth observation time, that is, $y_{i1}, \ldots, y_{i,k-1}$ are observed and $y_{ik}, \ldots, y_{iK}$ are missing. β_i is the $(q \times 1)$ vector of unobserved random coefficients characterizing the mean of y_i. γ is the set of fixed model parameters for the distribution of y_i and ψ is the set of fixed model parameters for the distribution of R_i given y_i.

We focus here on a special case of this model where the mean of y_{ij} is a linear function of time with a random slope and intercept for each individual. That is

$$(y_{ij} | \beta_i, \gamma) \sim_{\text{ind}} \mathrm{N}(\beta_{i0} + \beta_{i1} t_{ij}, \sigma^2), \qquad j = 1, \ldots, K, \tag{1}$$

$$\beta_i \sim \mathrm{N}_2(\theta, \Lambda); \quad \Lambda = \begin{pmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{pmatrix}$$

where N_k denotes the k-variate normal distribution, the random coefficients $\beta_i = (\beta_{i0}, \beta_{i1})^{\mathrm{T}}$ are a random intercept and slope for subject i, $\theta = (\theta_0, \theta_1)^{\mathrm{T}}$ is the average intercept and slope and the set of fixed model parameters are $\gamma = (\theta_0, \theta_1, \sigma^2, \tau_0^2, \tau_1^2, \rho)$.

We consider models for the joint distribution of y_i, R_i, β_i given X_i, assuming that given $X = (X_1, \ldots, X_n)$, (y_i, R_i, β_i) and (y_j, R_j, β_j) are independent for $i \neq j$. Inference for the fixed parameters γ of the model can be based on the observed likelihood of γ given the data $\{(y_{\text{obs},i}, R_i): i = 1, \ldots, n\}$, obtained formally by integrating the unobserved quantities $\{y_{\text{mis},i}, \beta_i: i = 1, \ldots, n\}$ out of the joint distribution of $\{y_i, R_i, \beta_i \quad i = 1, \ldots, n\}$ for the chosen model:

$$L(\gamma, \psi | Y_{\text{obs}}, R) = \text{const} \times \prod_{i=1}^{n} \int p(y_{\text{obs},i}, y_{\text{mis},i}, R_i, \beta_i | X_i, \gamma, \psi) \, \mathrm{d}y_{\text{mis},i} \, \mathrm{d}\beta_i. \tag{2}$$

Large-sample inference for fixed parameters γ can be based on the method of maximum likelihood with associated asymptotic standard errors, using iterative algorithms such as scoring, EM and extensions.[17–19] Bayesian inference can be implemented via stochastic simulation,[20–22] and provides a better analysis than maximum likelihood for small samples.

This approach requires specification of a model for the distribution of R_i given y_i, X_i, which is often not an easy task. A simpler approach is to base inference on the likelihood ignoring the drop-out mechanism:

$$L(\gamma | Y_{\text{obs}}) = \text{const} \times \prod_{i=1}^{n} \int p(y_{\text{obs},i}, y_{\text{mis},i}, \beta_i | X_i, \gamma) \, \mathrm{d}y_{\text{mis},i} \, \mathrm{d}\beta_i. \tag{3}$$

Copyright © 1999 John Wiley & Sons, Ltd.

Inference based on (3) is valid (although not necessarily fully efficient) when the drop-out mechanism is missing at random,[6,16] in that it does not depend on the missing data $y_{\mathrm{mis},i}$ or unobserved random coefficients β_i after conditioning on the observed variables X_i, $y_{\mathrm{obs},i}$ and the parameters ψ:

$$p(R_i|X_i, y_{\mathrm{obs},i}, y_{\mathrm{mis},i}, \beta_i, \psi) = p(R_i|X_i, y_{\mathrm{obs},i}, \psi). \quad (4)$$

If (4) holds and the parameters γ and ψ are also distinct, then inference based on (3) is efficient and equivalent to inferences based on the full likelihood (2). The drop-out mechanism is then called ignorable.[6,16]

We define two other classes of drop-out mechanisms, one stronger than missing at random and one weaker. As discussed in Little,[16] a number of alternative methods to maximum likelihood, including analysis of the complete cases and unweighted forms of generalized estimating equations, make the stronger assumption of covariate-dependent drop-out, where missingness is allowed to depend only on the covariates X_i:

$$p(R_i|X_i, y_{\mathrm{obs},i}, y_{\mathrm{mis},i}, \beta_i, \psi) = p(R_i|X_i, \psi). \quad (5)$$

Some authors call this condition missing completely at random,[13] but we adopt a terminology more consistent with Rubin's original ideas[6] and confine the latter term to mechanisms that do not depend on the random effects, the repeated measures or the covariates.[16]

We also find it useful to distinguish mechanisms that are missing at random conditional on the values of the unknown random coefficients β_i, and accordingly propose the following new definition. The drop-out mechanism is called subject-specific missing at random if it does not depend on the missing data but is allowed to depend on the observed variables and unobserved random coefficients:

$$p(R_i|X_i, y_{\mathrm{obs},i}, y_{\mathrm{mis},i}, \beta_i, \psi) = p(R_i|X_i, y_{\mathrm{obs},i}, \beta_i, \psi). \quad (6)$$

Subject-specific missing at random is clearly weaker than missing at random. It is relevant in the repeated-measures setting since under this condition maximum likelihood-based summary measures calculated separately for each individual are consistent for the true individual value as the number of repeated measures on an individual tends to infinity. This form of asymptotics is unrealistic for small numbers of repeated measures but has some relevance for longer series of measurements.

We now confine attention to the model (1), and analyses that ignore the drop-out mechanism. We assume for simplicity that all cases have at least two observations. In that case, maximum likelihood estimates for the average slope and intercept under (1) can be expressed in terms of the least squares intercept and slope $\hat{\beta}_i = (\hat{\beta}_{i0}, \hat{\beta}_{i1})^{\mathrm{T}}$ estimated from the observed vector of repeated measurements for subject i. Since the number of repeated measures for each subject varies, these estimates have differential precision. Under the model (1) we have

$$\hat{\beta}_i \sim \mathrm{N}_2(\theta, V_i), \text{ where } V_i = \begin{pmatrix} \tau_0^2 + \sigma^2\{k_i^{-1} + \bar{t}_i^2 u_i^{-1}\} & \rho\tau_0\tau_1 - \sigma^2\bar{t}_i u_i^{-1} \\ \rho\tau_0\tau_1 - \sigma^2\bar{t}_i u_i^{-1} & \tau_1^2 + \sigma^2 u_i^{-1} \end{pmatrix}$$

and $\bar{t}_i$ and $u_i = \sum_{j=1}^{k_i}(t_{ij} - \bar{t}_i)^2$ are, respectively, the mean and sum of squares about the mean of the times of measurement for subject i. We consider four methods of analyses for the average slope θ_1, distinguished by (a) whether the least squares slope and intercept for each individual is

Copyright © 1999 John Wiley & Sons, Ltd.

included in the analysis or the slope alone, and (b) whether the variance parameters are estimated by maximum likelihood or by the method of moments.

To derive the analyses that use the slopes and intercepts, note that the maximum likelihood estimate of θ for known V_i weights the least squares estimates $\hat{\beta}_i$ according to their precision, that is

$$\tilde{\theta} = \left(\sum_{i=1}^{n} V_i^{-1} \hat{\beta}_i \right) \left(\sum_{i=1}^{n} V_i^{-1} \right)^{-1}.$$

With V_i unknown, ignorable maximum likelihood (IML) estimates are obtained by replacing the variance parameters $(\tau_0^2, \tau_1^2, \rho, \sigma^2)$ in V_i by maximum likelihood estimates, obtained by an iterative method such as scoring or EM. A simpler approach is to replace the maximum likelihood estimates of the variance parameters by non-iterative method of moment estimates[23]

$$\tilde{\Lambda} = \frac{1}{n-1} \sum_{i=1}^{n} (\hat{\beta}_i - \bar{\beta})(\hat{\beta}_i - \bar{\beta})^{\mathrm{T}} - \frac{\hat{\sigma}^2}{n} \sum_{i=1}^{n} \begin{pmatrix} \{k_i^{-1} + \bar{t}_i^2 u_i^{-1}\} & -\bar{t}_i u_i^{-1} \\ -\bar{t}_i u_i^{-1} & u_i^{-1} \end{pmatrix}$$

where $\bar{\beta}$ is the unweighted average of the $\hat{\beta}_i$. We call the resulting method generalized least squares (GLS), following Wang-Clow *et al.*[24]

If interest is in the average slope and the intercept is a nuisance parameter, a univariate summary measures approach discards the estimated intercepts and bases inference on the individual slopes $\hat{\beta}_{i1}$ alone, together with a pooled estimate of σ^2. For known V_i, the maximum likelihood estimate of the slope based on these data is

$$\breve{\theta}_1 = \left(\sum_{i=1}^{n} (\tau_1^2 + \sigma^2 u_i^{-1})^{-1} \hat{\beta}_{i1} \right) \left(\sum_{i=1}^{n} (\tau_1^2 + \sigma^2 u_i^{-1})^{-1} \right)^{-1}.$$

The residual variance σ^2 is estimated by the pooled mean square $\hat{\sigma}^2 = \sum_{i=1}^{n} \mathrm{rss}_i/(N-2n)$ where rss_i is the residual sum of squares from the regression for individual i, N is the total number of observations and n is the number of subjects. The between-subjects variance τ_1^2 is estimated by its iterative maximum likelihood estimate conditional on $\sigma^2 = \hat{\sigma}^2$. We call the resulting approach summary measures maximum likelihood (SMML). Estimates and large sample standard errors are readily computed using a maximum likelihood program such as SAS Proc Mixed.[5] More simply, estimating τ_1^2 by the method of moments

$$\tilde{\tau}_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\hat{\beta}_{i1} - \bar{\beta}_1)^2 - \frac{\hat{\sigma}^2}{n} \sum_{i=1}^{n} u_i^{-1}$$

we obtain weighted least squares (WLS) estimates, again using the terminology in Wang-Clow *et al.*[24]

An even simpler approach is to discard the incomplete cases and base inferences on the complete cases. All the methods described above are equivalent when applied to the subset of complete cases, and we call the resulting method complete-case (CC) analysis.

These five methods – IML, GLS, SMML, WLS, CC analysis – yield asymptotically valid inferences when the missing data mechanism is missing completely at random, in that drop-out does not depend on the repeated measures or the unobserved random effects. The efficiency of the estimates varies, however. When data are not missing completely at random, the methods vary

Copyright © 1999 John Wiley & Sons, Ltd.

not only in relative efficiency but also in the degree of bias, as the simulation study described below indicates.

It is well known that IML yields valid inferences when the drop-out mechanism is missing at random, as in equation (4). CC analysis is consistent under covariate-dependent drop-out (5), but is generally biased when drop-out depends on the underlying random coefficients or the repeated measures. The degree of bias depends on the specific form of the mechanism. The method appeared competitive with IML in the simulation of Wang-Clow *et al.*,[24] suggesting that gains in methods that use the incomplete cases may be minor. However, the simulations reported here suggest that this finding was a consequence of the particular choices of parameters in that simulation, and does not have broad validity.

SMML is a maximum likelihood technique, so one might think it is valid for drop-out mechanisms that are missing at random. However, that is only true for maximum likelihood methods like IML that use all the data. For maximum likelihood methods that use only partial summaries of the data, the relevant assumption is that drop-out does not depend on the missing values after conditioning on the observed data in each case *that are included in the analysis*, that is, the summary measure for each case. Since the summary measure is a different combination of the repeated measures for each missing-data pattern, this condition reduces to the stronger assumption of covariate-dependent drop-out (5). To see this, we can write the pertinent missing at random assumption for the summary measures analysis as

$$p(R_i = j | X_i, y_{\text{obs},i}, y_{\text{mis},i}, \beta_i, \psi) = p(R_i = j | X_i, c_j^{\mathrm{T}} y_i, \psi)$$

where c_j is the vector of coefficients defining the summary measure for the pattern $R_i = j$. In particular, for the SMML method $c_j^{\mathrm{T}} y_i = \hat{\beta}_{i1}$. Summing the probabilities over all J patterns yields

$$\sum_{j=1}^{J} p(R_i = j | X_i, c_j^{\mathrm{T}} y_i, \psi) = 1 \text{ for all } y_i \text{ and } X_i.$$

Consider this constraint as y_i and X_i ranges continuously over all their possible values. Since the vector c_j is different for each pattern j, the constraint implies that each $p(R_i = j | X_i, c_j^{\mathrm{T}} y_i, \psi)$ must be a constant function of its argument $c_j^{\mathrm{T}} y_i$, implying a covariate-dependent drop-out mechanism (5).

This argument does not prove that covariate-dependent drop-out is a necessary condition for SMML to yield consistent estimates, since Rubin's theory[6] shows that the missing at random assumption is sufficient to ignore the mechanism, not necessary. The size of bias of SMML for particular missing-at-random mechanisms is not clear, and it is also of interest to study the properties of IML and SMML for non-MAR mechanisms. It seems possible that SMML could have less bias than IML if the estimated intercepts for each case that are input into the IML method are biased by the mechanism, but the estimated slopes are relatively free of bias.

The GLS and WLS methods can be expected to be more vulnerable to deviations from missing completely at random than their maximum likelihood counterparts IML and SMML, since the method-of-moments estimates of the variance parameters are not consistent under mechanisms other than covariate-dependent drop-out. However, GLS and IML, and WLS and SMML, have not been studied side by side in previous simulations, so the quantitative effects of using moment estimators rather than maximum likelihood estimators of the variance parameters have not been assessed.

Copyright © 1999 John Wiley & Sons, Ltd.

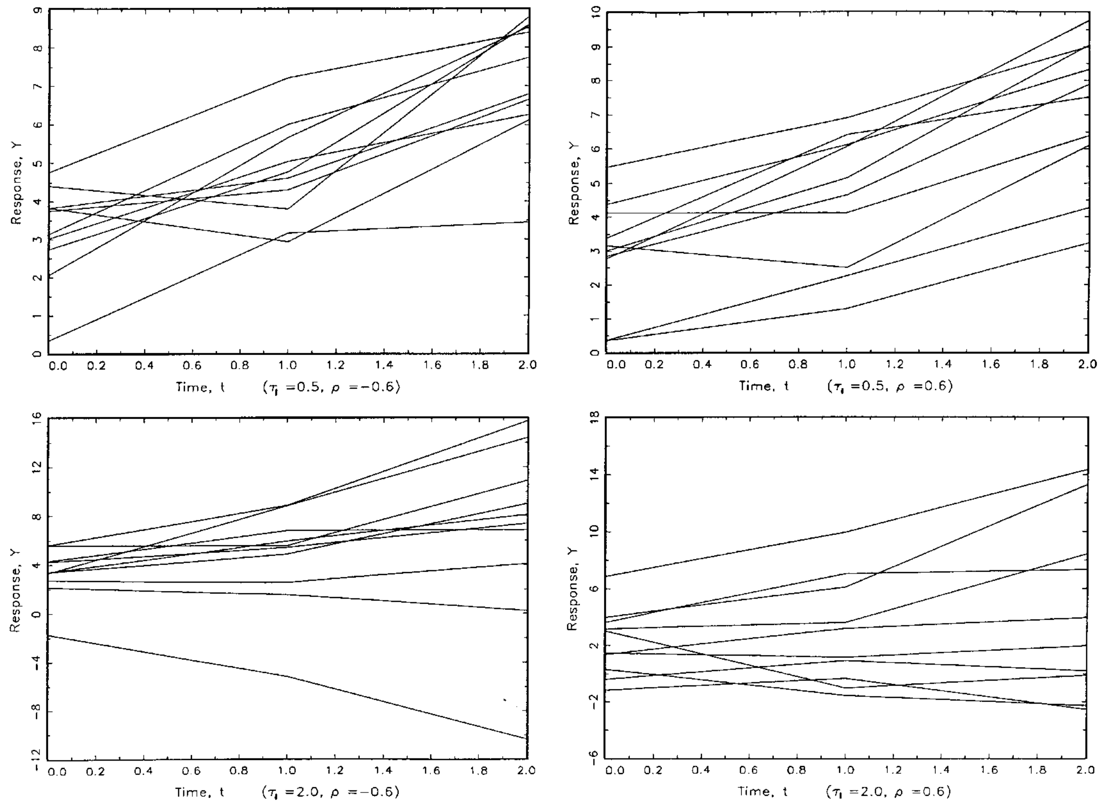

Figure 2. Profile plots of ten randomly chosen subjects from each of the four simulated populations

A simulation study was conducted to address these questions, under a broader range of drop-out mechanisms than have been considered in earlier studies.

3. SIMULATION STUDY

The five methods (IML, GLS, WLS, SMML, CC) were implemented on incomplete data sets with three time points and missing values confined to the third time point. That is, the observed data had only two patterns: complete cases and cases with the third observation missing. The complete data were generated using the following model. First, the subject-specific regression coefficients were generated as

$$\begin{bmatrix}\beta_{i0}\\ \beta_{i1}\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}1\\ 2\end{bmatrix}, \begin{bmatrix}1 & \rho\tau_1\\ \rho\tau_1 & \tau_1^2\end{bmatrix}\right)$$

then n observations were generated independently from univariate normal distributions $[y_{ij}|\beta_{i0}, \beta_{i1}] \sim \mathrm{N}(\beta_{i0} + \beta_{i1}j, \sigma^2)$. Four sets of parameter values were generated according to a factorial design with $\sigma^2 = 4$, $\tau_1^2 = 0{\cdot}25, 4$, $\rho = -0{\cdot}6, 0{\cdot}6$. For each parameter set 1000 data sets were generated for sample sizes $n = 200$ and 1000, yielding 8000 data sets in all. Figure 2 displays

Copyright © 1999 John Wiley & Sons, Ltd.

data for samples of ten individuals for each of the combinations of σ^2, τ_1^2 and ρ, to give a sense of what the data look like. For each data set, missing values at the third time point were created by each of the following six drop-out mechanisms:

$$\begin{array}{ll}
\text{(MCAR)} & \text{pr}[y_{i3}\ \text{missing}|y_i,\beta_i] = \Phi\{\lambda_0\} \\
\text{(MAR-SUM)} & \text{pr}[y_{i3}\ \text{missing}|y_i,\beta_i] = \Phi\{\lambda_0 - \lambda_1(y_{i1}+y_{i2})\} \\
\text{(MAR-DIFF)} & \text{pr}[y_{i3}\ \text{missing}|y_i,\beta_i] = \Phi\{\lambda_0 - \lambda_1(y_{i2}-y_{i1})\} \\
\text{(NI-Intercept)} & \text{pr}[y_{i3}\ \text{missing}|y_i,\beta_i] = \Phi\{\lambda_0 - \lambda_1(\beta_{i0}+y_{i1}-y_{i2})\} \\
\text{(NI-Slope)} & \text{pr}[y_{i3}\ \text{missing}|y_i,\beta_i] = \Phi\{\lambda_0 - \lambda_1(\beta_{i1}+y_{i2}-y_{i1})\} \\
\text{(NI-Y3)} & \text{pr}[y_{i3}\ \text{missing}|y_i,\beta_i] = \Phi\{\lambda_0 - \lambda_1 y_{i3}\}
\end{array}$$

where Φ is the cumulative normal distribution function. The coefficients λ_0, λ_1 in the above mechanisms were chosen so that approximately 30 per cent of cases have the third time point missing. The first mechanism is missing completely at random, which is equivalent to covariate-dependent drop-out for this model. The next two mechanisms are missing at random, where in MAR-SUM drop-out depends on the sum of the first two observations and in MAR-DIFF drop-out depends on the difference; the fourth and fifth mechanisms are non-ignorable (NI) elaborations of MAR-SUM and MAR-DIFF where missingness is also allowed to depend on the underlying unobserved intercept (NI-Intercept) or the underlying unobserved slope (NI-Slope); and the last mechanism NI-Y3 is non-ignorable since missingness depends on the outcome at the third repeated measure, which is missing for some cases. All the mechanisms except NI-Y3 are subject-specific missing at random as defined by equation (6).

One reason for studying data sets with just three time points is that the individual estimated intercepts and slopes are easily computed and interpreted. In particular

$$\hat{\beta}_{i1} = \begin{cases} y_{i2} - y_{i1}, & \text{for incomplete cases} \\ (y_{i3} - y_{i1})/2, & \text{for complete cases.} \end{cases}$$

Table I displays the mean empirical bias and empirical standard deviation of the estimated average slopes from each the five methods, calculated over the 1000 data sets, for each parameter set and drop-out mechanism. Table IA shows results for the mechanisms that are missing completely at random and missing at random, and Table IB shows results for the non-ignorable mechanisms. The column label 'Cov' shows the number of 90 per cent confidence intervals that cover the true parameter value. These intervals are based on large-sample standard errors and a normal reference distribution, which are good approximations for the samples sizes simulated. If the intervals have correct nominal coverage we expect 900 to cover the true parameter value. The following conclusions can be drawn from this table:

(a) When data are missing completely at random, there is no evidence of bias for any of the methods, and all the methods have close to nominal coverage, as theory predicts. The CC method is less efficient than the other method when τ_1^2 is large, reflecting the fact that there is more information in the incomplete cases in these settings. The other methods have very similar properties. In particular any loss of efficiency in the summary measures methods SMML and WLS is minor, and method of moments estimates of the variances parameters do not lead to noticeably inferior estimates of the average slope than the maximum likelihood estimates of the variance parameters.

Copyright © 1999 John Wiley & Sons, Ltd.

Table IA. Bias, standard deviation (SD) and confidence coverage of estimates of θ_1 from five methods (missing completely at random and missing at random mechanisms)

N	ρ	τ_1^2	Method	MCAR			MAR-SUM			MAR-DIFF		
				Bias	SD	Cov	Bias	SD	Cov	Bias	SD	Cov
200	− 0·6	0·25	CC	− 1	70	898	− 267	67	12	219	68	56
			WLS	2	68	900	17	75	908	− 66	78	784
			SMML	0	65	930	− 14	69	898	− 35	74	836
			GLS	2	68	894	53	74	832	− 66	77	782
			IML	− 1	70	904	9	74	894	− 8	76	892
200	0·6	0·25	CC	− 1	72	904	− 145	74	346	228	67	40
			WLS	0	68	902	64	76	788	− 58	78	818
			SMML	4	69	912	58	72	770	− 29	76	832
			GLS	1	67	900	48	76	830	− 58	78	816
			IML	2	67	912	19	73	894	− 21	72	896
200	− 0·6	4·0	CC	6	176	902	− 646	176	178	949	138	0
			WLS	7	152	902	175	156	704	− 27	158	902
			SMML	1	154	900	158	154	740	− 24	164	878
			GLS	7	152	906	62	154	892	− 32	156	900
			IML	6	159	892	− 7	150	908	− 4	152	904
200	0·6	4·0	CC	− 6	177	906	41	181	886	956	143	0
			WLS	− 3	154	914	246	159	534	− 11	161	898
			SMML	− 5	155	904	238	167	568	− 20	150	920
			GLS	− 2	154	898	59	160	874	− 17	159	890
			IML	12	148	908	12	152	918	− 17	162	904
1000	− 0·6	0·25	CC	0	33	904	− 264	29	0	218	30	0
			WLS	− 1	31	910	16	32	882	− 67	33	370
			SMML	− 1	31	890	− 18	33	810	− 29	32	738
			GLS	− 1	31	902	52	32	542	− 67	33	372
			IML	0	30	892	1	30	912	− 5	33	899
1000	0·6	0·25	CC	1	32	908	− 143	33	4	220	31	0
			WLS	0	30	902	68	33	330	− 65	34	412
			SMML	1	30	914	56	32	462	− 30	33	724
			GLS	0	30	900	52	33	528	− 64	34	406
			IML	− 3	30	912	18	29	898	− 15	33	896
1000	− 0·6	4·0	CC	4	81	912	− 438	77	0	942	64	0
			WLS	− 1	68	912	169	72	236	− 26	71	876
			SMML	0	68	914	165	71	258	− 15	73	884
			GLS	− 1	67	914	55	71	788	− 29	70	880
			IML	− 3	70	880	7	68	908	13	68	908
1000	0·6	4·0	CC	2	83	892	41	82	846	945	68	0
			WLS	3	72	896	240	70	38	− 22	70	908
			SMML	− 2	73	894	244	67	42	− 13	68	914
			GLS	3	71	896	51	71	810	− 28	70	892
			IML	− 8	70	900	16	72	892	1	69	899

(b) For the MAR-SUM mechanism, IML is the best method, reflecting its large-sample optimality under this mechanism. It has no evidence of bias, precision that is generally better than the other methods, and close to nominal coverage. GLS exhibits moderate bias, reflecting the fact that the method of moments variance estimate is not consistent for mechanisms that are not missing completely at random. This bias translates into poor coverage, particularly for the large sample size. SMML performs comparably to IML when

Copyright © 1999 John Wiley & Sons, Ltd.

Table IB. Bias, standard deviation (SD) and confidence coverage of estimates of θ_1 from five methods (non-ignorable mechanisms)

N	ρ	τ_1^2	Method	NI-Intercept			NI-Slope			NI-Y3		
				Bias	SD	Cov	Bias	SD	Cov	Bias	SD	Cov
200	−0·6	0·25	CC	−254	68	14	219	70	82	136	68	394
			WLS	−22	75	886	7	75	886	100	64	570
			SMML	−43	69	822	17	72	864	97	68	582
			GLS	22	75	878	9	75	886	117	64	470
			IML	−14	74	860	36	65	856	123	67	432
200	0·6	0·25	CC	−65	33	756	215	69	70	206	70	100
			WLS	69	70	748	4	74	896	136	70	390
			SMML	66	66	764	28	77	826	142	66	304
			GLS	49	73	824	−3	73	894	111	71	496
			IML	32	66	880	12	66	916	116	62	444
200	−0·6	4·0	CC	−544	165	48	975	138	0	916	130	0
			WLS	81	151	876	66	150	888	213	142	604
			SMML	81	158	852	92	154	874	212	143	566
			GLS	−19	148	902	25	150	904	127	142	808
			IML	−46	152	904	62	155	864	127	143	796
200	0·6	4·0	CC	209	176	676	985	136	0	1024	140	0
			WLS	217	152	614	69	155	870	218	153	582
			SMML	224	152	570	87	156	862	226	150	530
			GLS	23	154	882	8	154	900	96	154	822
			IML	11	162	892	26	155	888	104	153	796
1000	−0·6	0·25	CC	−252	29	0	214	31	0	136	33	4
			WLS	−23	32	842	7	33	908	99	30	64
			SMML	−51	32	494	27	32	764	100	30	32
			GLS	21	32	858	9	33	898	117	30	12
			IML	−16	32	856	38	29	644	119	31	4
1000	0·6	0·25	CC	−63	33	386	215	30	0	208	32	0
			WLS	67	32	336	6	32	888	139	31	4
			SMML	60	31	404	26	31	782	142	29	2
			GLS	47	33	550	−2	32	908	115	32	30
			IML	32	32	716	13	31	864	121	29	4
1000	−0·6	4·0	CC	−535	77	0	989	58	0	905	62	0
			WLS	92	71	622	76	69	728	195	66	114
			SMML	89	71	662	92	70	670	215	65	42
			GLS	−9	70	906	33	69	864	107	67	546
			IML	−52	65	788	59	64	780	120	72	424
1000	0·6	4·0	CC	214	78	140	994	61	0	1015	60	0
			WLS	216	66	52	76	67	714	214	63	62
			SMML	220	68	68	97	71	636	226	69	52
			GLS	19	68	888	16	67	912	907	64	632
			IML	−0	70	908	36	70	840	97	67	600

$\rho = -0{\cdot}6$, $\tau_1^2 = 0{\cdot}25$ but exhibits substantial bias and poor coverage for other choices of parameters, particularly when $\rho = 0{\cdot}6$, $\tau_1^2 = 4$. WLS is generally similar to SMML for these problems. Finally, CC is seriously biased and has very poor coverage in this setting.

(c) Results for the MAR-DIFF mechanism are broadly similar to results for the MAR-SUM mechanism. IML is again the best method, and GLS is competitive with IML for some choices of parameters but exhibits bias for others. SMML is comparable to IML when

Copyright © 1999 John Wiley & Sons, Ltd.

$\rho = 0{\cdot}6$, $\tau_1^2 = 4$ but inferior to IML for other choices of parameters, where it is moderately biased and has inferior coverage. The bias of SMML is smaller for the MAR-DIFF mechanism than for the MAR-SUM mechanism. WLS is notably inferior to SMML for this set of simulations, and CC is again much worse than all the other methods in terms of bias and coverage.

(d) For the NI-Intercept mechanism, all the methods exhibit bias and below nominal coverage to some degree, but important differences between the methods emerge (Table IB). IML is the best method overall, with small or moderate bias and close to nominal coverage in all but one problem ($n = 1000$, $\rho = 0{\cdot}6$, $\tau_1^2 = 0{\cdot}25$). GLS is similar and nearly as good as IML for these problems. SMML and WLS exhibit more bias and inferior coverage to IML and GLS, and CC is once again much the worst method in terms of bias and coverage.

(e) For the NI-Slope mechanism, IML and GLS are again the best methods, but GLS has somewhat lower bias and better coverage than IML for some problems ($n = 1000$, $\rho = -0{\cdot}6$, $\tau_1^2 = 0{\cdot}25$). Further examination of the estimates indicates that IML tends to underestimate τ_1^2 for these problems and GLS tends to overestimate τ_1^2. These characteristics tend to lead to estimates of the slope that are less biased for GLS than for IML. SMML and WLS are similar and somewhat inferior to IML and GLS for these problems, and CC is again the worst method by a wide margin.

(f) For the NI-Y3 mechanism none of the methods is very satisfactory. IML and GLS again emerge as the best methods, but their coverage in the cases with sample size 1000 is very poor. CC is again the worst of the methods considered in terms of bias and coverage.

Table II displays bias and standard deviation of estimates of the between-subject variance τ_1^2 from the five methods; the bias and precision of estimates of the within-subject variance σ^2 did not vary much between methods. GLS and WLS produce the same estimates of τ_1^2 and hence have the same results in Table II. When data are missing completely at random, IML shows considerable gains in efficiency over alternative methods, with some evidence of bias that is diminished in the larger sample size problems. For other mechanisms, IML generally produces the best estimates of τ_1^2, and the superiority of the maximum likelihood methods (IML, SMML) over method of moments counterparts (GLS, WLS) is evident.

4. DISCUSSION

Our simulation study includes a broader range of drop-out mechanisms and parameter values than previous studies. Nevertheless it is important to be cautious in generalizing conclusions, since no simulation can cover all the situations that arise in practice. With that caveat, a number of interesting findings emerge from our results.

A striking result is the very poor performance of CC. This is in contrast with the simulations in Wang-Clow *et al.*,[24] where CC was very competitive with other methods. Those authors simulated a case where the variability of the slopes was very small. Our results underline the fact that CC can be very biased when data are not missing completely at random, and methods that use all the available data can reduce bias, improve precision and improve coverage in these settings.

Methods that use all the available information (IML and GLS) outperformed summary measures methods that discard the intercepts (SMML and WLS) when the mechanism was not missing completely at random, including both missing at random and non-ignorable cases. This is

Copyright © 1999 John Wiley & Sons, Ltd.

Table II. Bias and standard deviation of estimates of τ_1^2 from five methods

N	ρ	τ_1^2	Method	MCAR		MAR-SUM		MAR-DIFF		NI-Intercept		NI-Slope		NI-Y3	
				Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
200	−0·6	0·25	CC	3	107	−88	88	−28	90	−94	96	−45	94	−30	111
			WLS	15	173	279	184	348	162	210	176	162	186	6	169
			SMML	18	107	62	116	138	132	24	111	41	116	−10	97
			GLS	15	173	279	184	348	162	210	176	162	186	6	169
			IML	20	86	27	95	27	95	−3	85	−13	80	−18	75
200	0·6	0·25	CC	−10	110	12	94	−31	91	26	106	−63	97	−78	98
			WLS	−5	167	216	171	353	159	135	166	182	157	−3	165
			SMML	24	110	105	120	129	118	74	109	32	113	−68	92
			GLS	−5	167	216	171	353	159	135	166	182	157	−3	165
			IML	70	78	84	85	84	88	78	81	64	86	16	71
200	−0·6	4·0	CC	21	550	−354	496	−1673	358	−562	489	−1926	331	−1773	327
			WLS	7	523	364	544	454	519	368	531	211	532	−64	519
			SMML	18	492	285	564	130	569	190	528	−186	512	−507	498
			GLS	7	523	364	564	456	519	368	531	211	532	−64	519
			IML	4	297	9	501	−22	513	−101	460	−146	447	−344	494
200	0·6	4·0	CC	−13	541	108	542	−1672	360	−65	530	−1928	329	−1950	320
			WLS	−23	524	174	530	447	533	2	513	247	562	−44	543
			SMML	0	493	51	495	119	526	−75	471	−219	530	−682	525
			GLS	−23	524	174	530	467	533	2	513	247	542	−46	543
			IML	−63	663	62	438	33	487	76	458	−4	526	−252	490
1000	−0·6	0·25	CC	−3	46	−85	41	−31	40	−96	41	−54	43	−36	48
			WLS	−1	78	281	80	352	79	209	83	157	84	−12	83
			SMML	−6	48	50	57	115	56	3	55	24	53	−25	48
			GLS	−1	78	281	80	352	79	209	83	157	84	−12	83
			IML	−5	39	8	39	12	37	−10	33	−30	35	−33	35
1000	0·6	0·25	CC	2	46	13	45	−33	40	28	50	−55	44	−76	42
			WLS	2	77	214	78	352	80	116	82	159	81	−8	81
			SMML	6	49	95	54	115	55	65	50	24	51	−58	48
			GLS	2	77	214	78	352	80	116	82	159	81	−8	81
			IML	6	36	76	34	77	34	74	36	54	36	3	34
1000	−0·6	4·0	CC	7	239	−346	221	−1650	161	−540	218	−1933	143	−1792	150
			WLS	−5	223	368	231	437	238	337	230	247	227	−66	246
			SMML	−11	213	292	211	84	236	185	232	−207	239	−524	221
			GLS	−5	223	368	231	437	238	337	230	247	227	−66	246
			IML	24	216	−11	211	−32	233	−125	205	−197	213	−371	227
1000	0·6	4·0	CC	5	249	28	256	−1662	167	−83	228	−1931	145	−1945	151
			WLS	−6	232	103	236	431	228	−23	219	258	242	−55	247
			SMML	−16	220	107	212	92	234	−39	229	−193	242	−514	249
			GLS	−6	232	103	236	431	228	−23	219	258	242	−55	247
			IML	30	192	22	210	58	213	57	216	−100	211	−217	206

an important finding since the simplicity of methods based on the slopes alone is compelling, and it is generally thought that these methods involve minor losses of efficiency over full maximum likelihood. Our simulations suggest that this is indeed the case when the mechanism is missing completely at random, but not when missingness depends on the data.

A related point is that SMML is generally biased under missing at random mechanisms, as is seen in our simulations under MAR-SUM and MAR-DIFF and predicted by the theoretical argument in Section 2. SMML is a maximum likelihood method, but it is not valid under missing at random since it discards data, namely the intercepts. Further inspection of the results shows that the bias of SMML is generally greater for MAR-SUM and NI-Intercept mechanisms than

Copyright © 1999 John Wiley & Sons, Ltd.

for MAR-DIF and NI-Slope mechanisms. This finding reflects the fact that missingness is more closely associated with the estimated intercepts for MAR-SUM and NI-Intercept mechanisms than for the MAR-DIF and NI-Slope mechanisms. Thus dropping the information about the mechanism carried in the intercepts leads to a greater bias in those cases.

Note that under subject-specific missing at random, the estimated slope for each individual is a consistent estimate of the true slope for that individual as the number of repeated measures tends to infinity. The individual-level summary measures are valid in this weak sense for all the mechanisms in our study except NI-Y3, which is not subject-specific missing at random. However, the overall estimate of the slope in SMML is a weighted average of the individual slopes, with weights that depend on the estimate of τ_1^2. It is bias in the estimate of τ_1^2, as reflected in Table II, that creates the problem in the SMML estimate of θ_1.

The bias of SMML for missing at random mechanisms does not necessarily mean that the method is always inferior to IML for non-ignorable mechanisms. We believe that it is theoretically possible for SMML to have less bias than IML under some circumstances, but we found no real evidence of this in our simulations. IML generally outperformed SMML for the non-ignorable mechanisms we simulated.

We note that IML was noticeably superior to GLS in our missing at random simulations, reflecting the fact that maximum likelihood, unlike the method of moments, yields consistent estimates of variance under MAR-SUM and MAR-DIFF. For the non-ignorable mechanisms neither method consistently dominated the other. Overall the added computation of IML over GLS has little pay-off when the data are missing completely at random, but IML seems preferable to GLS when the data are not missing completely at random.

None of the methods considered here is satisfactory for all mechanisms, as indicated by the disappointing results for the NI-Y3 mechanism. As noted in the introduction, in these simulations we have not studied methods that are tuned to non-ignorable drop-out mechanisms. There is a growing literature of these methods,[7–15] as discussed in Little.[16] In their simulation study, Wang-Clow *et al.*[24] consider unweighted least squares (UWLS), which computes unweighted averages of the least squares slope of individuals in each treatment group. This method is unbiased for mechanisms that depend on the unobserved random slopes and intercepts, but the method had poor precision in their simulations. A more promising approach is the weighted least squares method considered by Wang-Clow *et al.*[24] that incorporates a covariate adjustment on time to drop-out. This ANCOVA method performed well for the non-ignorable mechanisms in Wang-Clow *et al.*, but was inferior to IML for missing at random mechanisms. Wang-Clow *et al.*[24] computed the variance components in this ANCOVA method using the method of moments, as in Wu and Bailey;[8] since method of moments approaches did not do as well as maximum likelihood in our simulations, we suggest that a maximum likelihood version of the method ANCOVA based on the pattern-mixture model in Little[16] might do even better.

Methods tuned to particular non-ignorable mechanisms yield consistent estimates under covariate-dependent drop-out, although there is a loss of efficiency of estimation that might be considerable. It is important to emphasize that these methods are in general biased under missing at random drop out, since the missing at random model is not generally a submodel of the hypothesized non-ignorable mechanism. No method has been found that performs uniformly well for the broad range of missing at random and not missing at random mechanisms that might occur in practice. Whether such methods can be found remains an open question.

Copyright © 1999 John Wiley & Sons, Ltd.

ACKNOWLEDGEMENTS

This research was supported by a grant from the National Science Foundation. We thank a referee for useful suggestions on an earlier draft.

REFERENCES

1. Harville, D. A. 'Maximum likelihood approaches to variance component estimation and to related problems (with discussion)', *Journal of the American Statistical Association*, **72**, 320–340 (1977).
2. Laird, N. M. and Ware, J. H. 'Random-effects models for longitudinal data', *Biometrics*, **38**, 963–974 (1982).
3. Jennrich, R. I. and Schluchter, M. D. 'Incomplete repeated-measures models with structured covariance matrices', *Biometrics*, **42**, 805–820 (1986).
4. Dixon, W. J. *BMDP Statistical Software*, University of California Press, Berkeley, CA, 1988.
5. SAS 'The mixed procedure', in *SAS/STAT Software*: *Changes and Enhancements*, *Release 6.07*, Technical Report P-229, SAS Institute, Inc., Cary, NC, 1992, Chapter 16.
6. Rubin, D. B. 'Inference and missing data', *Biometrika*, **63**, 581–592 (1976).
7. Wu, M. C. and Carroll, R. J. 'Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process', *Biometrics*, **44,** 175–188 (1988).
8. Wu, M. C. and Bailey, K. R. 'Estimation and comparison of changes in the presence of informative right censoring: conditional linear model', *Biometrics*, **45**, 939–955 (1989).
9. Schluchter, M. D. 'Methods for the analysis of informatively censored longitudinal data', *Statistics in Medicine*, **11**, 1861–1870 (1992).
10. Little, R. J. A. 'Pattern-mixture models for multivariate incomplete data', *Journal of the American Statistical Association*, **88**, 125–134 (1993).
11. Little, R. J. A. 'A class of pattern-mixture models for normal missing data', *Biometrika*, **81**, (3), 471–483 (1994).
12. Little, R. J. A. and Wang, Y-X. 'Pattern-mixture models for multivariate incomplete data with covariates', *Biometrics*, **52**, 98–111 (1996).
13. Diggle, P. and Kenward, M. G. 'Informative dropout in longitudinal data analysis (with discussion)', *Applied Statistics*, **43**, 49–94 (1994).
14. Mori, M., Woolson, R. F. and Woodsworth, G. G. 'Slope-estimation in the presence of informative censoring: modeling the number of observations as a geometric random variable', *Biometrics*, **50**, 39–50 (1994).
15. Molenberghs, G., Kenward, M. G. and Lesaffre, E. 'The analysis of longitudinal ordinal data with nonrandom dropout', *Biometrika*, **84**, (1), 33–44 (1997).
16. Little, R. J. A. 'Modeling the drop-out mechanism in longitudinal studies', *Journal of the American Statistical Association*, **90**, 1112–1121 (1995).
17. Dempster, A. P., Laird, N. M. and Rubin, D. B. 'Maximum likelihood from incomplete data via the EM algorithm (with discussion)', *Journal of the Royal Statistical Society*, *Series B*, **39**, 1–38 (1977).
18. Little, R. J. A. and Rubin, D. B. *Statistical Analysis with Missing Data*, Wiley, New York, 1987.
19. Meng, X. L. and Van Dyk, D. 'The EM algorithm – an old folk song sung to a fast new tune, (with discussion)', *Journal of the Royal Statistical Society*, *Series B*, **59**, (3), 511–568 (1997).
20. Gelfand, A. E. and Smith, A. F. M. 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409 (1990).
21. Tanner, M. A. *Tools for Statistical Inference*, 3rd edn, Springer-Verlag, New York, 1996.
22. Gilks, W. R., Wang, C. C., Yvonnet, B. and Coursaget, P. 'Random-effects models for longitudinal data using Gibbs' sampling', *Biometrics*, **49**, 441–453 (1993).
23. Vonesh, E. F. and Carter, R. L. 'Efficient inferences for random-coefficient growth curve models with unbalanced data', *Biometrics*, **43**, 617–628 (1987).
24. Wang-Clow, F., Lange, M., Laird, N. M. and Ware, J. H. 'Simulation study of estimators for rate of change in longitudinal studies with attrition', *Statistics in Medicine*, **14**, 283–297 (1995).

Copyright © 1999 John Wiley & Sons, Ltd.