Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

A Random Pattern-Mixture Model for Longitudinal Data With Dropouts

Wensheng Guo, Sarah J Ratcliffe & Thomas Ten T Have

To cite this article: Wensheng Guo, Sarah J Ratcliffe & Thomas Ten T Have (2004) A Random Pattern-Mixture Model for Longitudinal Data With Dropouts, Journal of the American Statistical Association, 99:468, 929-937, DOI: 10.1198/016214504000000674

To link to this article: https://doi.org/10.1198/016214504000000674

Published online: 31 Dec 2011.

A Random Pattern-Mixture Model for Longitudinal Data With Dropouts

Wensheng GUO, Sarah J. RATCLIFFE, and Thomas T. TEN HAVE

Pattern-mixture models are frequently used for longitudinal data analysis with dropouts because they do not require explicit specification of the dropout mechanism. These models stratify the data according to time to dropout and formulate a model for each stratum. This usually results in underindentifiability, because we need to estimate many pattern-specific parameters even though the eventual interest is usually on the marginal parameters. In this article we extend this framework to a random pattern-mixture model, where the pattern-specific parameters are treated as nuisance parameters and modeled as random instead of fixed. The pattern is defined according to a surrogate for the dropout process. A constraint is then put on the pattern by linking it to the time to dropout using a random-effects survival model. We assume, conditional on the latent pattern effects, that the longitudinal outcome and the dropout process are independent. This model retains the robustness of the traditional pattern-mixture models, while avoiding the overparameterization problem. When we define each subject as a separate stratum, this model reduces to the shared parameter model. Maximum likelihood estimates are obtained using an EM Newton–Raphson algorithm. We apply the method to the depression data from the Prevention of Suicide in Primary Care Elderly Collaborative Trial (PROSPECT). We show when the dropout information is adjusted for under the proposed model, the treatment seems to reduce depression in the elderly.

KEY WORDS: Dropout; EM algorithm; Mixed-effects model; Pattern-mixture mode.

1. INTRODUCTION

Many longitudinal studies suffer from attrition, which can cause bias in the analysis if the dropouts are informative. To account for informative dropout, a number of model-based approaches have been proposed to jointly model the longitudinal outcome and the dropout mechanism (Wu and Carroll 1988; De Gruttola and Tu 1994; Diggle and Kenward 1994; Schluchter 1992; Little 1993; Little 1994; Michiels, Molenberghs, Bijnens, Vangeneugden, and Thijs 2002; Hogan and Laird 1997a). (See Little 1995; Hogan and Laird 1997b; and Kenward and Molenberghs 1999 for recent reviews.) Among these approaches, *pattern-mixture* models (Little 1993) are commonly used because of their robustness in modeling the dropout mechanism. In this article we propose a random pattern-mixture model in which the pattern-specific parameters are treated as nuisance parameters and modeled as random. We first generalize the definition of pattern. The pattern is defined according to a good surrogate for the dropout process, which can be a baseline or time-varying covariate, or time to dropout. A constraint is then imposed on the pattern by linking it to the time to dropout using a random-effects survival model (e.g., Clayton and Cuzick 1985). We assume, conditional on the latent pattern effects, that the longitudinal outcome and the dropout process are independent. This model retains the robustness of the traditional pattern-mixture models, while avoiding the overparameterization problem. In the absence of a good surrogate to define the pattern, we can define each subject as a separate stratum, which reduces the model to a shared-parameter model (Wu and Carroll 1988; De Gruttola and Tu 1994).

Our research was motivated by the PROSPECT (Prevention of Suicide in Primary Care Elderly Collaborative Trial) study (Bruce and Pearson 1999). The aim of this study was to demonstrate that suicidality, hopelessness, and depression can be decreased by improving the recognition and intervention of late-life depression in a representative sample of older primary care patients. For patients diagnosed at baseline with treatable depression, a comprehensive treatment schedule was created. Of interest is whether using antidepressant medication along with counseling improves depression levels compared with only counseling with a depression specialist. The longitudinal response of interest is the Hamilton 23 Depression score (HAMD) obtained at each visit. Subject-specific random effects are needed to account for the heterogeneity in longitudinal profiles. Because patients with a higher depression score are more likely to drop out of the study, the dropout may be informative and thus needs to be taken into account when modeling longitudinal data. More details about the study and the analysis are given in Sections 2 and 5.

Let $\mathbf{Y}$ denote the continuous outcome, R the time until the subject drops out of the study, $\mathbf{X}$ the fixed-effects design matrix for $\mathbf{Y}$, and $\boldsymbol{\beta}$ the vector of unobserved random subject-specific effects. The aim is to model $p(\mathbf{Y}, R|\mathbf{X}) = \int p(\mathbf{Y}, R, \boldsymbol{\beta}|\mathbf{X})\, d\boldsymbol{\beta}$. Existing methods can be considered either selection models or fixed pattern-mixture models.

Selection models use the factorization

$$p(\mathbf{Y}, R, \boldsymbol{\beta}|\mathbf{X}) = p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta})p(R|\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta})p(\boldsymbol{\beta}|\mathbf{X}),$$

where $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta})$ is a subject-specific longitudinal model and $p(R|\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta})$ imposes an assumption on the dropout mechanism. The shared parameter model [also termed the random coefficient selection model by Little (1995)] can be considered a special case of selection models, which assumes that conditional on $\boldsymbol{\beta}$ and $\mathbf{X}$, $\mathbf{Y}$ and R are independent and thus can be modeled separately and linked by $\boldsymbol{\beta}$,

$$p(\mathbf{Y}, R, \boldsymbol{\beta}|\mathbf{X}) = p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta})p(R|\mathbf{X}, \boldsymbol{\beta})p(\boldsymbol{\beta}|\mathbf{X}).$$

The justification for the conditional independence assumption is that dropout is correlated with a latent subject-specific trajectory captured by $\boldsymbol{\beta}$. Calculation of the likelihood requires integration over the random effects $\boldsymbol{\beta}$ and can be difficult to

Wensheng Guo is Assistant Professor (E-mail: *wguo@cceb.upenn.edu*), Sarah J. Ratcliffe is Assistant Professor (E-mail: *sratclif@cceb.upenn.edu*), and Thomas R. Ten Have is Professor (E-mail: *ttenhave@cceb.upenn.edu*), Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104. This research was supported by National Institutes of Health grants CA84438, NIMH62298, NIMH61892, and AG14131. The authors thank the editor, the associate editor, and the referees for their comments, which substantially improved this article.

Journal of the American Statistical Association
December 2004, Vol. 99, No. 468, Applications ans Case Studies
DOI 10.1198/016214504000000674

implement. Schluchter (1992) and De Gruttola and Tu (1994) avoided this difficulty by using a joint normal approach for the longitudinal outcome $\mathbf{Y}$, time to dropout R, and the random effects $\boldsymbol{\beta}$.

Pattern-mixture models stratify the data according to time to dropout (termed dropout pattern), and form a model for each stratum. The final estimate is a weighted average of the stratum-specific estimate. The basic idea of pattern-mixture models is that the dropout pattern provides good surrogate information on the dropout process and, conditional on the pattern, the missing mechanism is ignorable within a stratum, and thus information from the complete cases can be borrowed to predict the incomplete cases. This implies the factorization

$$p(\mathbf{Y}, R, \boldsymbol{\beta}|\mathbf{X}) = p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, R)\, p(\boldsymbol{\beta}|\mathbf{X}, R)\, p(R|\mathbf{X}),$$

where $p(R|\mathbf{X})$ is modeled as a multinomial distribution. However, a full pattern-mixture model is usually underidentified because of the need to estimate many pattern-specific parameters. Constraints are needed to make the model identifiable (Little 1995; Little and Wang 1996; Daniels and Hogan 2000).

In this article we present a random pattern-mixture model that combines the features of the selection models and fixed pattern-mixture models. We first generalize the definition of pattern. First, based on a surrogate for the dropout process, such as a baseline covariate, a time-changing covariate, or time to dropout, we stratify the data into m strata (termed patterns). The pattern effects are explicitly modeled as random effects and used to link $\mathbf{Y}$ and R. This model implies the factorization

$$\begin{aligned} &p(\mathbf{Y}, R, \boldsymbol{\beta}|\mathbf{X}) \\ &\quad = \int p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}, R) p(\boldsymbol{\beta}|\mathbf{X}, \mathbf{u}, R) p(R|\mathbf{X}, \mathbf{u}) p(\mathbf{u}|\mathbf{X})\, d\mathbf{u} \\ &\quad = \int p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}) p(\boldsymbol{\beta}|\mathbf{X}, \mathbf{u}) p(R|\mathbf{X}, \mathbf{u}) p(\mathbf{u}|\mathbf{X})\, d\mathbf{u}, \end{aligned} \tag{1}$$

where $\mathbf{u}$ is the random pattern effects and $p(R|\mathbf{X}, \mathbf{u})$ imposes a constraint on the distribution of $\mathbf{u}$. The assumptions $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}, R) = p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \mathbf{u})$ and $p(\boldsymbol{\beta}|\mathbf{X}, \mathbf{u}, R) = p(\boldsymbol{\beta}|\mathbf{X}, \mathbf{u})$ imply that $\mathbf{Y}$ and $\boldsymbol{\beta}$ depend on R through the random pattern effects $\mathbf{u}$. Although this factorization resembles the shared-parameter model, the model borrows the fundamental idea of stratification from the fixed pattern-mixture model. That is, conditional on the latent pattern effects, the missing mechanism is ignorable within a stratum, and final parameters of interest are the marginal estimates averaging over the latent pattern effects subject to certain constraints. When we define the pattern according to time to dropout and impose a diffuse prior on the pattern, this model reduces to the fixed pattern-mixture model. If we stratify to the finest level (i.e., each subject forms its own stratum), this model reduces to the shared parameter model.

Unlike the traditional pattern-mixture model that always defines the patterns according to time to dropout, the proposed method emphasizes the definition of patterns according to a good surrogate of the dropout process. A good surrogate needs to be related to both the outcome and the time to dropout, and stratifying on the surrogate should result in conditional independence of the two. The time to dropout itself is usually a surrogate for the dropout process and can be used to define the pattern. However, in the presence of censoring, the time to dropout becomes less informative on the underlying dropout process, and it is desirable to define the pattern based on other covariates not affected by the censoring. The initial choice of the surrogate can be based on the information from previous studies, and the conditional independence assumption can be checked from the data. The main advantage of the proposed model over the shared-parameter model is that it allows information to be borrowed across subjects within a pattern because the longitudinal component and the dropout component are linked at the pattern level instead of at the subject level.

This model has similar computational difficulty as the shared-parameter models, because of the need to integrate over $\mathbf{u}$ and $\boldsymbol{\beta}$. To avoid this computational difficulty, we use a joint normal approach. We model the random effects as normally distributed and the outcome and dropout times as multivariate normal, while adjusting for censoring. The EM algorithm (Dempster, Laird, and Rubin 1977; Laird and Ware 1982) can be used to calculate maximum likelihood estimates. To speed up the calculations and to provide inference, we have also investigated a combined EM Newton–Raphson algorithm.

We begin by describing more details of the PROSPECT study in Section 2. We then introduce the random pattern-mixture model in Section 3. In Section 4 we describe an estimation procedure using an EM Newton–Raphson algorithm, and in Section 5 we apply these methods to the data from the PROSPECT study. We report simulations based on these results in Section 6, and provide concluding remarks in Section 7.

2. THE PROSPECT STUDY

The PROSPECT study is an ongoing National Institute of Mental Health-funded collaborative study conducted over the past 4 years by the late-life mood disorders Intervention Research Centers at Cornell University, University of Pennsylvania, and University of Pittsburgh. The original design was given by Bruce and Pearson (1999). The overall aim of the study is to demonstrate that the risk of suicidal ideation in late life can be decreased by improving the recognition and treatment of its primary risk factor, depression. Depression is strongly associated with the risk of suicide and is relatively pervasive, ranging from 1% to 10% in the community-dwelling elderly. In addition, effective treatments for depression based on best-practice guidelines exist but are not yet used adequately in most cases of late-life depression occurring in the community. Hence the intervention was based on a collaborative care model for increasing patient and provider adherence to best-practice guidelines for treating depression and related symptoms such as suicidal ideation. The intervention integrated population-based methodology with clinically sensitive assessment in elderly patients from 18 diverse primary care practices. Because the linchpin of the PROSPECT intervention is the addition of a health specialist to the primary care setting, randomization was performed via the primary care practices rather than at the subject level.

The primary goal of the analysis described in this article is to understand the effect of taking antidepressant medication on a continuous depression outcome when the dropout information is accounted for through stratification on physical functioning. Because medication data are not yet available on the participants in the nonintervention group, the data for this analysis are

limited to those patients attending one of the 10 intervention practices. Further, only patients diagnosed with major depression [as diagnosed by the structured clinical interview for the *Diagnostical and Statistical Manual*, edition 3 revised (DSM-III-R)] are included, because these subjects are primarily the target sample for treatment. Once a patient is diagnosed, the physician can prescribe antidepression medication therapy. If the patient does not want to take an antidepressant or does not respond to one, then the physician can recommend interpersonal therapy or new medication provided by the study. The focus of the analysis is on the difference in depression levels for patients undergoing either of the two types of therapy.

For each of the 157 patients included in this analysis, a comprehensive assessment of depression status was completed at baseline and possibly reassessed after 4, 8, 12, 16, and 24 months. Depression was measured at each visit using the Hamilton 23 Depression (HAMD) score. Patients with a higher score are considered to be more depressed. The HAMD scores over the study period ranged from 6 to 34. A total of 68 patients (43.31%) dropped out of the study due to either death (suicide), institutionalization, or failure to adhere to the prescribed therapy. Two-thirds of the patients dropped out for the last reason, with 20 of these dying. The remaining 89 patients are censored at different follow-up appointments due to the staged entry into the study. Figure 1 shows the HAMD scores over time for the two groups. The patients who died may be sicker than those lost to follow-up, and it is desirable to model the two types of dropouts separately. However, given the relatively small number of dropouts due to death, we could not estimate with sufficient precision the impact of the different types of drop out on the results.

The motivation for our dropout model arises from the relationship among dropout, depression outcome, physical functioning, and treatment. The ambulatory nature of the sample (primary care elderly patients) makes physical function limitations the major obstacle to visiting the primary care physician

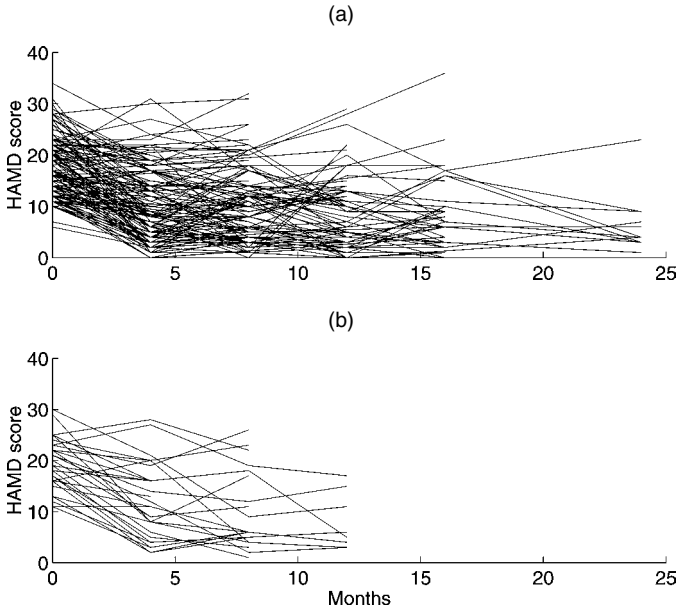

Figure 1. Subject-Specific HAMD Profiles Over Time for the Two Groups: (a) Interpersonal Therapy Patients; (b) Antidepressant Medication Therapy Patients.

Table 1. Descriptive Statistics for the PROSPECT Study by Pattern

Pattern	n	Percent dropout	Average survival time (weeks)	Average HAMD score
1	9	67	23.11	16.05
2	16	63	37.00	12.75
3	11	45	29.09	16.03
4	15	53	39.47	13.49
5	9	22	61.33	11.88
6	18	44	70.52	11.35
7	15	40	52.47	12.69
8	23	39	52.45	11.66
9	29	38	54.45	11.13
10	12	25	58.37	9.84

(i.e., dropout). Furthermore, there is a strong relationship between depression and physical functioning, as there is between antidepression treatment and physical functioning. Hence, physical functioning is a good surrogate for the mediating effect of dropout on the effect of treatment on depression outcome. We stratify on physical functioning, such that strata range from very little functional disability to severe functional disability through death.

For these data, the patient's physical ability is measured by the physical functioning (PF) component of the patient's SF-36 score at baseline. The SF-36 is a generic health status measure (Ware, Snow, Kosinski, and Gandek 1993). The motivation for using PF as the surrogate for the dropout process arose from its reported relationship with depression, seeking healthcare, and dropout. There have been reports that elderly patients with limited physical functioning are more likely to be depressed (e.g., Oslin, Streim, Katz, Edell, and Ten Have 2000) and are less likely to seek treatment for their depression (Oslin et al. 2000). In addition, Miller, Rejeski, Reboussin, Ten Have, and Ettinger (2000) showed, based on a national survey of elderly persons, that a strong relationship existed between declining physical function and the propensity to dropout in the form of death and institutionalization in assisted-care living residences. Further, functional disability is seen as one of the major complications in the diagnosis and treatment of depression, because it places competing demands on the physician and can make it more difficult for the patient to follow the recommended therapy. These relationships between physical functioning, depression, seeking treatment, and dropout motivated the use of PF as the surrogate variable.

Given the clinical motivation for the use of PF as the pattern variable, we further checked its use via the data. The PF scores ranged from 0 to 100, with higher scores indicating a more functional patient. Patients with similar scores were stratified together in 10-point increments, resulting in a pattern variable with 10 strata. The average HAMD scores, and survival times (calculated by fitting a Kaplan–Meier survival curve) for each strata are shown in Table 1. It can be seen that the PF score is clearly related to both time to dropout and HAMD score, with an overall correlation coefficient of $-.77$. That is, patients with lower PF scores usually have a higher HAMD score and tend to drop out earlier. Thus the use of PF is further justified by the data. We further check on the conditional independence assumption of our model in Section 5.

3. THE RANDOM PATTERN–MIXTURE MODEL

Based on a surrogate for the dropout process, such as the PF component of the patient's SF-36 score at baseline, we stratify the data into m strata (termed patterns). We consider cases where subject j is nested within the ith stratum. Let $\mathbf{y}_{ij}$ be an $n_{ij} \times 1$ vector of observed outcomes for the jth subject within the ith pattern, $i = 1, \ldots, m, j = 1, \ldots, n_i$. Let r_{ij} be the corresponding dropout time for this subject, with the observed dropout time possibly being censored at some time point c_{ij}.

We model both the outcomes and the dropout times using mixed-effects models (e.g., Schluchter 1992; De Gruttola and Tu 1994), but link the two models by the random pattern effects. Without loss of generality, we write r_{ij}, even though some transformation [e.g., $\log(r_{ij})$] of the dropout times is generally used. Thus,

$$\mathbf{y}_{ij} = \mathbf{X}_{1ij}\boldsymbol{\alpha}_1 + \mathbf{Z}_{ij}\boldsymbol{\beta}_{ij} + \mathbf{W}_{ij}\mathbf{u}_i + \mathbf{e}_{ij} \tag{2}$$

and

$$r_{ij} = \mathbf{x}_{2ij}^T\boldsymbol{\alpha}_2 + \mathbf{b}^T\mathbf{u}_i + \varepsilon_{ij}, \tag{3}$$

where $\mathbf{X}_{1ij}$, $\mathbf{Z}_{ij}$, and $\mathbf{W}_{ij}$ are the known design matrices for the fixed effects, subject-level random effects, and pattern-level random effects for $\mathbf{y}_{ij}$; $\boldsymbol{\alpha}_1$ is the vector of unknown fixed effects; $\boldsymbol{\beta}_{ij} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ is a vector of unknown subject-level random effects; $\mathbf{u}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}})$ is a vector of unknown pattern-level random effects; $\mathbf{e}_{ij} \sim \mathrm{N}(\mathbf{0}, \sigma_{\mathbf{e}}^2\mathbf{I}_{ij})$ is a vector of residuals; $\mathbf{x}_{2ij}$ is the known design matrix linking the unknown parameter vector $\boldsymbol{\alpha}_2$ to r_{ij}; $\mathbf{b}$ is an unknown parameter vector linking $\mathbf{u}_i$ to r_{ij}; and $\varepsilon_{ij} \sim \mathrm{N}(0, s^2)$ is the residual for r_{ij}. Further, δ_{ij} is used to indicate whether r_{ij} is observed ($\delta_{ij} = 1$) or censored ($\delta_{ij} = 0$). The set of parameters to be estimated in this model is $\boldsymbol{\theta} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \mathbf{b}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\mathbf{u}}, \sigma_{\mathbf{e}}^2, s^2)$.

This model assumes that conditioning on the latent pattern effects, the missing-data mechanism is ignorable within a stratum. The informativeness of the dropout is modeled by (3), which imposes a constraint on the random pattern effects $\mathbf{u}_i$. When $\mathbf{b} = \mathbf{0}$, the model reduces to an ignorable dropout model.

4. ESTIMATION

By treating $\mathbf{u}_i$ and $\boldsymbol{\beta}_{ij}$ as missing data, we can use an EM algorithm (Dempster et al. 1977) to calculate the maximum likelihood estimates (MLEs) of the parameters. These are found in two steps. The E-step involves finding the conditional expectation of the sufficient statistics for the complete data log-likelihood. These expected values are then used to estimate the model parameters in the M-step that maximizes the expected log-likelihood. However, a straight EM approach requires calculation of two levels of expectations, pattern-level expectations and subject-level expectations. Instead, we estimate the parameters via a combined EM Newton–Raphson algorithm that contains only pattern-level expectations. This is accomplished by absorbing the subject-level random effects ($\mathbf{Z}_{ij}\boldsymbol{\beta}_{ij}$) into the error term in (2), resulting in a new, structured error covariance matrix for the repeated measures. This new error term is given by $\mathbf{e}_{ij}^* = \mathbf{Z}_{ij}\boldsymbol{\beta}_{ij} + \mathbf{e}_{ij}$, where $\mathbf{e}_{ij}^* \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}_{ij})$ and $\boldsymbol{\Psi}_{ij} = \mathbf{Z}_{ij}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\mathbf{Z}_{ij}^T + \sigma_{\mathbf{e}}^2\mathbf{I}_{ij}$.

In the combined approach, the E-step still finds the conditional expectation of the sufficient statistics. However, in the M-step, there is no closed-form solution for $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\sigma_{\mathbf{e}}^2$. A Newton–Raphson algorithm is used to calculate the estimates for $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, $\sigma_{\mathbf{e}}^2$, and hence $\boldsymbol{\Psi}_{ij}$. Then, conditional on these values, parameter estimates for the fixed effects, $\boldsymbol{\Sigma}_{\mathbf{u}}$, and s^2 are calculated.

The complete-data log-likelihood for the EM part of the algorithm is given by

$$l_c = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\phi(\mathbf{y}_{ij}|\mathbf{X}_{1ij}, \mathbf{u}_i, \hat{\boldsymbol{\theta}})\phi(\mathbf{u}_i|\hat{\boldsymbol{\theta}})\phi(r_{ij}|\mathbf{x}_{2ij}, \mathbf{u}_i, \hat{\boldsymbol{\theta}}),$$

where $\phi(\cdot)$ is the normal density function. Using this likelihood, the sufficient statistics needed for the M-step include

$$\sum_i\sum_j \varepsilon_{ij}^2, \qquad \sum_i n_i\mathbf{u}_i\mathbf{u}_i^T, \qquad \sum_i\sum_j \mathbf{W}_{ij}\mathbf{u}_i,$$
$$\sum_i\sum_j r_{ij}\mathbf{u}_i, \qquad \sum_i\sum_j \mathbf{x}_{2ijk}\mathbf{u}_i, \qquad 1 \le k \le n_{ij}.$$

Closed-form solutions for their expectations are given in the Appendix.

The expected sufficient statistics calculated in the E-step are then used to calculate the estimates in the M-step. Given $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\sigma_{\mathbf{e}}^2$, and hence $\boldsymbol{\Psi}_{ij}$, we can obtain the estimates

$$\hat{\boldsymbol{\alpha}}_1 = \left(\sum_i\sum_j \mathbf{X}_{1ij}^T\boldsymbol{\Psi}_{ij}^{-1}\mathbf{X}_{1ij}\right)^{-1} \times E\left(\sum_i\sum_j \mathbf{X}_{1ij}^T\boldsymbol{\Psi}_{ij}^{-1}(\mathbf{y}_{ij} - \mathbf{W}_{ij}\mathbf{u}_i)\right),$$

$$\hat{\boldsymbol{\alpha}}_2 = \left(\sum_i\sum_j \mathbf{x}_{2ij}\mathbf{x}_{2ij}^T\right)^{-1} E\left(\sum_i\sum_j \mathbf{x}_{2ij}(r_{ij} - \mathbf{b}\mathbf{u}_i)\right),$$

$$\hat{\mathbf{b}} = \left(E\left(\sum_i n_i\mathbf{u}_i\mathbf{u}_i^T\right)\right)^{-1} E\left(\sum_i\sum_j (r_{ij} - \mathbf{x}_{2ij}^T\boldsymbol{\alpha}_2)\mathbf{u}_i\right),$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{u}} = \left(\sum_i n_i\right)^{-1} E\left(\sum_i n_i\mathbf{u}_i\mathbf{u}_i^T\right),$$

and

$$\hat{s^2} = \left(\sum_i n_i\right)^{-1} E\left(\sum_i\sum_j \varepsilon_{ij}^2\right),$$

where, for simplicity, we write $E(\cdot)$ when we mean the conditional expectations $E(\cdot|\mathbf{y}_i, \mathbf{r}_i, \boldsymbol{\Delta}_i, \hat{\boldsymbol{\theta}})$, with $\boldsymbol{\Delta}_i = \mathrm{diag}(\delta_{i1}, \ldots, \delta_{in_i})$. We continue to use this notation in the following, where no confusion will arise. We then calculate

$$\hat{\mathbf{e}}_{ij}^* = \mathbf{y}_{ij} - \mathbf{X}_{1ij}\hat{\boldsymbol{\alpha}}_1 - E(\mathbf{W}_{ij}\mathbf{u}_i),$$

which can be used to calculate the expected log-likelihood. This expected log-likelihood is a function of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\sigma_{\mathbf{e}}^2$, and a Newton–Raphson algorithm can be used to obtain the estimates for $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\sigma_{\mathbf{e}}^2$ that maximize the expected log-likelihood.

Standard errors for the fixed effects can be obtained via the formula of Louis (1982). The conditional covariance matrix for

the fixed effects $[\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2]$ at convergence is

$$\operatorname{cov}\left(\begin{bmatrix}\hat{\boldsymbol{\alpha}}_1\\ \hat{\boldsymbol{\alpha}}_2\end{bmatrix}\right) = \left(\sum_i \begin{bmatrix}\mathbf{X}_{1i} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_{2i}\end{bmatrix}^T \times \mathbf{C}_{22}^{-1}(i)(\mathbf{I}_i - \boldsymbol{\Delta}_i\mathbf{A}_i\boldsymbol{\Delta}_i\mathbf{C}_{22}^{-1}) \begin{bmatrix}\mathbf{X}_{1i} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_{2i}\end{bmatrix}\right)^{-1}, \quad (4)$$

where $\mathbf{X}_{1i}$, $\mathbf{X}_{2i}$, $\mathbf{C}_{22}^{-1}(i)$, and $\mathbf{A}_i$ are as given in the Appendix. Closed-form solutions for the standard errors of the remaining parameters are difficult to obtain. Consequently, we use parametric bootstrapping to estimate the standard errors.

5. APPLICATION TO THE PROSPECT STUDY

We now return to the PROSPECT study described in Section 2. We applied four models to the data: a standard mixed model assuming no link between dropout and outcome, a shared-parameter model, and two random pattern-mixture models with two different patterns. The first pattern was based on the observed dropout/censoring times of the subjects. We grouped together patients who dropped out or were censored at the same time point, resulting in a pattern with five strata. We also tried to separate the dropouts and censors into different patterns, which resulted in too-sparse a stratification and failure of the model to converge. The other pattern was based on the PF component of the patient's SF-36 score, which was measured only at baseline. Patients with similar scores were stratified together in 10-point increments, resulting in a pattern variable with 10 strata. We also tried other stratifications on the PF scores using different cutpoints and different numbers of strata, but found that the results are not sensitive to the grouping. This is due to the fact that we build the conditional independence on the latent pattern effects instead of conditioning on the patterns themselves, and the normal prior on the latent pattern effects is a weak constraint. The 10-point increments ultimately used were selected for ease of clinical interpretation.

The missing at random mixed-effects model for the longitudinal outcome was used to determine the possibly significant effects for the more-complex shared parameter and random pattern-mixture models. Of the patient demographic information recorded, only gender is significant ($p = .02$). The other variables, such as age ($p = .74$), are highly insignificant. Only the significant demographic variables and the variable of interest (treatment) are presented here for ease of comparison between the models. For the random effects, most covariances are approximately equal to 0. These include random slope effects, based on time, at both the subject and pattern level, as well as a treatment–pattern interaction. Hence, only random intercepts are used in the final models. The results for the four models are given in Table 2 and the following final equations:

Random dropout model:

$$\begin{aligned}\text{HAMD}_{ij} &= \alpha_{10} + \alpha_{11}(\text{Visit time}) + \alpha_{12}(\text{Treatment}) \\ &\quad + \alpha_{13}(\text{Gender}) + \beta_{ij} + u_i + e_{ij},\\ \log(\text{Dropout time})_{ij} &= \alpha_{20} + \alpha_{21}(\text{Treatment}) + \varepsilon_{ij}.\end{aligned}$$

Shared parameter model:

$$\begin{aligned}\text{HAMD}_{ij} &= \alpha_{10} + \alpha_{11}(\text{Visit time}) + \alpha_{12}(\text{Treatment}) \\ &\quad + \alpha_{13}(\text{Gender}) + \beta_{ij} + e_{ij},\\ \log(\text{Dropout time})_{ij} &= \alpha_{20} + \alpha_{21}(\text{Treatment}) + b\beta_{ij} + \varepsilon_{ij}.\end{aligned}$$

Random pattern-mixture models:

$$\begin{aligned}\text{HAMD}_{ij} &= \alpha_{10} + \alpha_{11}(\text{Visit time}) + \alpha_{12}(\text{Treatment}) \\ &\quad + \alpha_{13}(\text{Gender}) + \beta_{ij} + u_i + e_{ij},\\ \log(\text{Dropout time})_{ij} &= \alpha_{20} + \alpha_{21}(\text{Treatment}) + bu_i + \varepsilon_{ij},\end{aligned}$$

where β_{ij} and u_i are the subject and pattern level random intercepts.

For the random dropout model, there is an average decrease in the HAMD scores of .57 per month ($p < .0001$). Also, males have a 2-point lower HAMD score ($p = .02$) on average. Although the treatment effect was not significant, it is interesting to note that it was positive. This implies that using the antidepressant medication increases one's HAMD score, and hence the average patient is more depressed on medication. Similar results are also obtained using the shared-parameter model. These results contradict with the expectation.

Table 2. Results From Modeling the HAMD Scores and Log(dropout times) for the PROSPECT Study, Assuming No Link Between the Two Models, a Link at the Subject Level, and Two Different Links at the Pattern Level Based on PF Score and on Dropout Times

Parameter	Random dropout	Shared parameter model	Random pattern-mixture model	
			PF scores	Dropout time
Model for HAMD				
Intercept	$16.41_{(.606)}$	$16.40_{(.53)}$	$16.82_{(.51)}$	$15.72_{(.52)}$
Time	$-.57_{(.04)}$	$-.57_{(.05)}$	$-.61_{(.04)}$	$-.53_{(.05)}$
Treatment	$.24_{(1.10)}$	$.25_{(1.03)}$	$-.25_{(.73)}$	$.24_{(.86)}$
Gender	$-2.07_{(.91)}$	$-2.05_{(.86)}$	$-2.22_{(.62)}$	$-1.82_{(.59)}$
$\boldsymbol{\Sigma}_{\beta}$	$18.75_{(3.24)}$	$18.70_{(2.86)}$	$14.20_{(2.43)}$	$11.98_{(1.83)}$
$\boldsymbol{\Sigma}_{\mathbf{u}}$	$.46_{(.03)}$		$.27_{(.13)}$	$.66_{(.23)}$
$\sigma^2_{\mathbf{e}}$	$29.03_{(2.06)}$	$33.65_{(2.34)}$	$33.71_{(3.08)}$	$36.11_{(3.60)}$
Model for Log(dropout times)				
Intercept	$4.18_{(.08)}$	$4.18_{(.09)}$	$4.06_{(.04)}$	$4.52_{(.01)}$
Treatment	$-.53_{(.16)}$	$-.54_{(.19)}$	$-.44_{(.12)}$	$-.12_{(.04)}$
b		$-.01_{(.02)}$	$-.22_{(.09)}$	$-.34_{(.24)}$
s^2	$.70_{(.06)}$	$.49_{(.05)}$	$.38_{(.06)}$	$.06_{(.07)}$

NOTE: Parameter estimates are given with standard errors in parentheses.

In the random pattern-mixture model with the PF score patterns, the link parameter is significant, which suggests that the dropout is informative. Under this model, most of the estimates are similar to the previous two models. The standard errors are substantially smaller, because the significant link between the two outcomes allow information to be borrowed across the outcomes. There is a large difference in the estimate of the treatment effects. Although it is still not significant, it now has a negative value. Thus, when the dropout information is more adequately incorporated, use of the medication lowers the HAMD scores, implying that antidepressants help the average patient to feel less depressed. This result is as expected medically.

When the patterns are based on the dropout times, the estimate for the treatment effect reverts to being positive. The link parameter is not significant, indicating that the dropout time pattern may not be a good surrogate for the dropout process. The observed dropout distribution is a mixture of non-health-related censoring and dropout related to depression and treatment. Because the distinction between the two is not recorded and there is heavy censoring, stratification based on the observed dropout times may not be able to adequately correct the bias due to informative dropouts. Comparing the results from the proposed model using the PF score as the pattern variable, the positive treatment coefficient found in the random dropout model, shared-parameter model, and random pattern-mixture model using a dropout time pattern may simply be a side effect of not adequately accounting for the dropout process. Because the three models yield different estimates in the application, we investigate the sensitivity through simulations in the next section.

We assessed the assumption of conditional independence in the random pattern-mixture models from the fitted residuals. We calculated the residuals from both the longitudinal and dropout models. The residuals are approximately iid normal. In the model using PF score as the pattern, the residuals for the two outcomes are not correlated, with a correlation coefficient of $-.08$. This indicates that adjusting for their relationship based on PF score resulted in conditional independence. The correlation in the model using dropout time as pattern is greater, with a correlation coefficient of $-.14$, indicating that the stratifying on the time to dropout may not lead to conditional independence.

6. SIMULATION

To study the sensitivity of the model assumptions, we simulate data using different assumptions and fit different models to investigate how much the results changed accordingly. We conducted three simulation studies, generating 500 replicates of the data in each simulation, as follows. Given the known fixed effects, random effects, and link parameter values, plus the error covariances and pattern assignments (for the random pattern-mixture model), we generated dropout times for each of the 157 subjects. Based on these dropout times, we calculated an "exit" time so that approximately 60% of the subjects were censored. The observed dropout times and dropout indicators could then be calculated. We generated the repeated-measures data based on the observed dropout times. That is, we generated the data only at baseline and at follow-up times that were less than the observed dropout times. Once each replicate was generated using the true known parameter values associated with the underlying model, we fitted a number of different models to the data.

In the first simulation, the true underlying model was assumed to be the proposed random pattern-mixture model,

$$\begin{aligned}\text{HAMD}_{ij} &= 16.82 - .606(\text{Visit time}) - 1.33(\text{Treatment}) \\ &\quad + 2.22(\text{Gender}) + \beta_{ij} + u_i + e_{ij}, \\ r_{ij} &= 6.06 - 2.45(\text{Treatment}) - 1.09u_i + \varepsilon_{ij},\end{aligned}$$

where r_{ij} is the log transformation of the dropout time, $\beta_{ij} \sim \text{N}(0,\ 5.15)$, $u_i \sim \text{N}(0,\ 4.5)$, $\text{e}_{ij} \sim \text{N}(0,\ 4.03)$, and $\varepsilon_{ij} \sim \text{N}(0,\ 3.38)$. We fit a shared-parameter model and two random pattern-mixture models, one of which uses the true pattern and the other uses time to dropout to create the pattern. Figure 2 summarizes the results for the fixed effects. Because the true dropout model includes only treatment, only the treatment effect varies substantially across different models. The shared-parameter model underestimates the size of the treatment effect. The random pattern-mixture model using the true pattern produces a very accurate estimate. The one using the time to dropout as the pattern overestimates the size of the treatment effect. This simulation suggests that in the presence of dropout, one should try to collect related covariates, such as SF36-PF, and include this information in the analysis. Simply stratifying the data by subject or by time to dropout does not always correct the bias. The shared-parameter model also overestimated $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ and s^2. The overestimate of $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ is expected, because it is trying to capture both $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ and $\mathbf{\Sigma}_{\mathbf{u}}$ from the true model. Using the time-to-dropout pattern, s^2 is highly underestimated, whereas the remaining covariances are overestimated.

For the second simulation, we assumed the true model to be

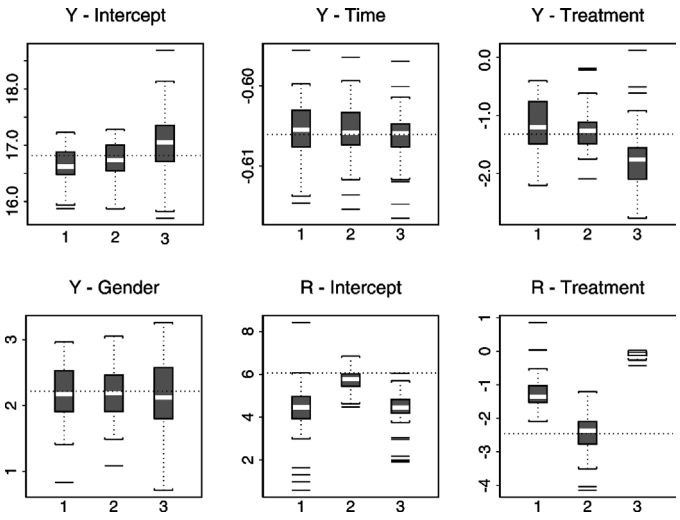

Figure 2. Simulation Results Assuming That the True Model Is a Random Pattern-Mixture Model Linked at the Pattern Level. The true values are indicated by the dashed lines. Y indicates estimates from the longitudinal model, R indicates estimates from the dropout model. 1, shared parameter model; 2, random pattern-mixture model linked at the true pattern; 3, random pattern-mixture model using dropout time pattern.

a shared-parameter model,

$$\begin{aligned}\text{HAMD}_{ij} &= 16.82 - .606(\text{Visit time}) - 1.33(\text{Treatment}) \\ &\quad + 2.22(\text{Gender}) + \beta_{ij} + e_{ij}, \\ r_{ij} &= 6.06 - 2.45(\text{Treatment}) - 1.09\beta_{ij} + \varepsilon_{ij}.\end{aligned}$$

The first "pattern" is created by randomly grouping the subjects into 10 strata. This allows us to investigate the effect on the estimates obtained from the random pattern-mixture model using a noninformative pattern variable when the two outcomes are linked at the subject level. The second pattern is based on time to dropout. The results from fitting a shared-parameter model and two random pattern-mixture models are shown in Figure 3. As expected, the shared-parameter model produces accurate estimates when it is the true underlying model. The model using time to dropout as the pattern variable results in biased estimates for the treatment effects, as well as the covariance estimates. When a noninformative pattern variable is used, the fixed-effects estimates from the random pattern-mixture model are underestimated by an average of 4.5%, and the standard errors are inflated by an average of 7.3%. The 95% confidence interval coverages of the fixed effects are also slightly below the nominal level (90.6% on average). Also, s^2 is over-estimated, whereas $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ is slightly underestimated due to the addition of the pattern random effect. This suggests that even when the true underlying model is the shared-parameter model, fitting a random pattern-mixture model using a noninformative pattern variable can still correct some bias, but the estimates are not as efficient as those obtained from the true model.

For the third simulation, we assumed that the longitudinal outcome and the dropout process were linked by the underlying true dropout times (without censoring). That is, we assumed that we knew the true underlying dropout time and used this information in generating the longitudinal outcomes,

$$\begin{aligned}\text{HAMD}_{ij} &= 6.82 - .606(\text{Visit time}) - 1.33(\text{Treatment}) \\ &\quad + 2.22(\text{Gender}) + \beta_{ij} + .05r_{ij} + e_{ij}, \\ r_{ij} &= 6.06 - 2.45(\text{Treatment}) + \varepsilon_{ij},\end{aligned}$$

where $\beta_{ij} \sim \text{N}(0, 6.15)$, $\text{e}_{ij} \sim \text{N}(0, 1.53)$, and $\varepsilon_{ij} \sim \text{N}(0, 3)$. A random censoring time c_{ij} is drawn from a uniform distribution on $[0, r_{ij}/.6]$. The subject is labelled as censored if $r_{ij} > c_{ij}$, and as dropout if otherwise. This led to approximately 60% of the subjects being censored. We also created a surrogate of the true underlying dropout time: x^*_{ij} such that $\text{corr}(x^*_{ij}, r_{ij}) = .75$. We then defined three patterns. The first two patterns were based on the observed dropout/censoring times, with one pattern ignoring the dropout/censoring indicator and the other pattern having separate patterns for the dropout and censored subjects. The third pattern was based on the surrogate x^*_{ij}. We created 10 strata in each setting according to the percentiles of the observed variable. This resulted in 20 strata for the model with separate patterns for the dropout and censored subjects. We fitted a shared-parameter model and three random pattern-mixture models. Figure 4 summarizes the results for the fixed effects from the HAMD model. The estimates for most of the parameters are biased under the shared parameter model and the model linked by the pattern based on the observed dropout time (indicator ignored). The random pattern-mixture model using the pattern based on the surrogate x^*_{ij} produces satisfactory results. The model with separate patterns for the censored and dropout subjects also produces satisfactory results for all parameters except the intercept. The one using the observed time to dropout, ignoring the indicator, to define patterns leads to biased results. This is due to the fact that we treated the time to censoring and time to dropout the same in creating the patterns. Subjects with quite different true underlying dropout times are

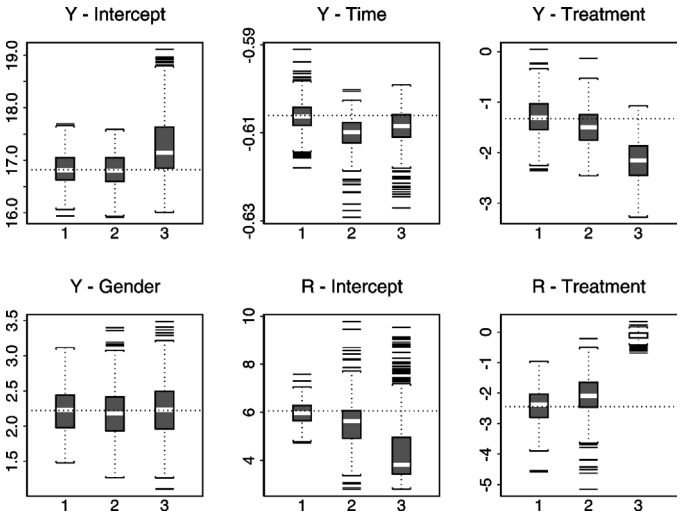

Figure 3. Simulation Results Assuming That the True Model Is a Shared Parameter Model Linked at the Subject Level. The true values are indicated by the dashed lines. Y indicates estimates from the longitudinal model, R indicates estimates from the dropout model. 1, shared parameter model; 2, random pattern-mixture model using noninformative pattern; 3, random pattern-mixture model using dropout time pattern.

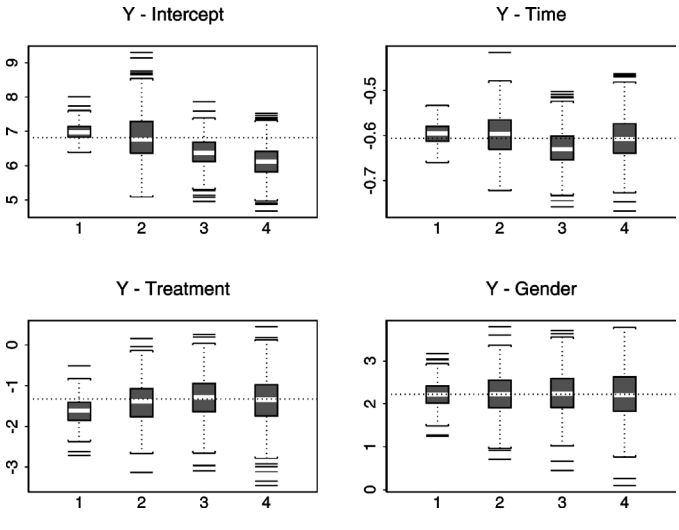

Figure 4. Simulation Results Assuming That the Longitudinal Outcomes Are Directly Linked to the True Dropout Times Before Censoring Occurred. The true values are indicated by the dashed lines. Y indicates estimates from the longitudinal model. 1, shared parameter model; 2, random pattern-mixture model using surrogate patterns; 3, random pattern-mixture model using patterns based on dropout/censoring time; 4, random pattern-mixture model with 20 patterns based on dropout and censoring times separately.

grouped together. Thus this stratification does not adequately model the true underlying link.

The proposed random pattern-mixture model provides a flexible framework to account for informative dropouts in longitudinal data analysis. The key modeling assumption is that conditional on the random pattern effects, the longitudinal outcome, and the time to dropout are independent. This assumption can be checked from the data. Under this framework, the shared-parameter model can be viewed as a special case, with each subject being its own pattern. The three simulations together suggest the importance of finding a good pattern variable in the modeling of longitudinal data with informative dropouts, which can be a baseline or time-varying covariate, time to dropout, or the subjects themselves. An inappropriate choice of the pattern variable can lead to biased and inefficient estimates.

7. CONCLUSION

We have proposed a random pattern-mixture model for longitudinal data with dropouts. This model borrows the fundamental idea of stratification from the traditional pattern-mixture models, and thus does not require explicit specification of the dropout mechanism. Unlike the traditional pattern-mixture models that fit a model for each pattern, we treat the pattern-specific parameters as nuisance parameters and explicitly model them as random. A constraint is put on the random pattern effects by linking them to the time to dropout. The parameters of interest are the marginal parameters after integrating out the random pattern effects. This model avoids the overparameterization problem of the fixed pattern-mixture model while retaining its robustness.

We have also extended the definition of "dropout pattern" to be "stratification according to a surrogate." In defining the pattern, one should explore covariates that can provide information on the underlying dropout process, because the key assumption of the model is that, conditional on the latent pattern effects, the longitudinal outcome and the time to dropout are independent. Violating this assumption can lead to biased estimates and invalid inferences. Although time to dropout is usually a good surrogate for the underlying dropout process, in the presence of heavy censoring, the observed dropout time may no longer be a good surrogate for defining the pattern.

We fitted the proposed method to data from the PROSPECT study, in which the dropouts were thought to be informative. The PF component of the SF-36 was found to be a good surrogate for the dropout process, because it was related to both the time to dropout and the HAMD scores. We showed that stratifying on the PF score led to conditional independence of time to dropout and the longitudinal outcome. When dropout information is adjusted for under our proposed model, the treatment seems to reduce depression in the elderly.

APPENDIX: COMPUTATIONAL FORMULAS FOR THE EM NEWTON–RAPHSON ALGORITHM

The expectations of the sufficient statistics involving the pattern effect, $\mathbf{u}_i$, can be derived from the multivariate normal distribution $\phi(\mathbf{u}_i, \mathbf{y}_i, \mathbf{r}_i)$ and the assumption that $\mathbf{y}_{ij}$ and r_{ij} are conditionally independent. Combining the various matrices and vectors for all subjects within pattern i, we obtain $\mathbf{y}_i = [\mathbf{y}_{i1}^T, \ldots, \mathbf{y}_{in_i}^T]^T$, $\mathbf{X}_{1i} = [\mathbf{X}_{1i1}^T, \ldots, \mathbf{X}_{1in_i}^T]^T$, $\mathbf{W}_i = [\mathbf{W}_{1i1}^T, \ldots, \mathbf{W}_{1in_i}^T]^T$, $\mathbf{r}_i = [r_{i1}, \ldots, r_{in_i}]^T$, $\mathbf{X}_{2i} = [\mathbf{x}_{2i1}, \ldots, \mathbf{x}_{2in_i}]^T$, and $\tilde{\boldsymbol{\Psi}}_i = \operatorname{diag}(\Psi_{1i1}, \Psi_{1i2}, \ldots, \Psi_{1in_i})$. Thus

$$\boldsymbol{\Sigma}_{\mathbf{y}_i} = \operatorname{cov}(\mathbf{y}_i, \mathbf{y}_i) = \mathbf{W}_i \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{W}_i^T + \tilde{\boldsymbol{\Psi}}_i,$$

$$\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{r}_i} = \operatorname{cov}(\mathbf{y}_i, \mathbf{r}_i) = \mathbf{W}_i \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{b} \mathbf{1}_{n_i}^T,$$

$$\boldsymbol{\Sigma}_{\mathbf{r}_i} = \operatorname{cov}(\mathbf{r}_i, \mathbf{r}_i) = (\mathbf{b}^T \boldsymbol{\Sigma}_u \mathbf{b}) J_{n_i} + s^2 \mathbf{I}_{n_i},$$

$$\mathbf{C}_{22}(i) = \operatorname{cov}\begin{pmatrix} \mathbf{y}_i \\ \mathbf{r}_i \end{pmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{y}_i} & \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{r}_i} \\ \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{r}_i}^T & \boldsymbol{\Sigma}_{\mathbf{r}_i} \end{bmatrix}.$$

We can then use the properties of the multivariate normal distribution to derive

$$\hat{\mathbf{u}}_i = E(\mathbf{u}_i | \mathbf{y}_i, \mathbf{r}_i, \hat{\boldsymbol{\theta}}) = [\, \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{W}_i^T \quad \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{b} \mathbf{1}^T \,] \mathbf{C}_{22}^{-1}(i) \begin{bmatrix} \mathbf{y}_i - \mathbf{X}_{1i}\boldsymbol{\alpha}_1 \\ \hat{\mathbf{r}}_i - \mathbf{X}_{2i}\boldsymbol{\alpha}_2 \end{bmatrix}, \tag{5}$$

$$\Longrightarrow$$

$$\operatorname{var}(\mathbf{u}_i | \mathbf{y}_i, \mathbf{r}_i, \hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}_{\mathbf{u}} - [\, \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{W}_i^T \quad \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{b} \mathbf{1}^T \,] \mathbf{C}_{22}^{-1}(i) \times [\, \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{W}_i^T \quad \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{b} \mathbf{1}^T \,]^T$$

$$E(\mathbf{u}_i \mathbf{u}_i^T | \mathbf{y}_i, \mathbf{r}_i, \hat{\boldsymbol{\theta}}) = [\, \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{W}_i^T \quad \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{b} \mathbf{1}^T \,] \mathbf{C}_{22}^{-1}(i) \mathbf{A}_i \mathbf{C}_{22}^{-1}(i) [\, \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{W}_i^T \quad \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{b} \mathbf{1}^T \,]^T + \operatorname{var}(\mathbf{u}_i | \mathbf{y}_i, \mathbf{r}_i, \hat{\boldsymbol{\theta}}) \tag{6}$$

where

$$\mathbf{A}_i = \begin{bmatrix} (\mathbf{y}_i - \mathbf{X}_{1i}\boldsymbol{\alpha}_1)(\mathbf{y}_i - \mathbf{X}_{1i}\boldsymbol{\alpha}_1)^T \\ (\mathbf{y}_i - \mathbf{X}_{1i}\boldsymbol{\alpha}_1)(\hat{\mathbf{r}}_i - \mathbf{X}_{2i}\boldsymbol{\alpha}_2)^T \\ (\hat{\mathbf{r}}_i - \mathbf{X}_{2i}\boldsymbol{\alpha}_2)(\mathbf{y}_i - \mathbf{X}_{1i}\boldsymbol{\alpha}_1)^T \\ \widehat{\mathbf{r}_i \mathbf{r}_i^T} - \hat{\mathbf{r}}_i \boldsymbol{\alpha}_2^T \mathbf{X}_{2i}^T - \mathbf{X}_{2i}\boldsymbol{\alpha}_2 \hat{\mathbf{r}}_i^T + \mathbf{X}_{2i}\boldsymbol{\alpha}_2 \boldsymbol{\alpha}_2^T \mathbf{X}_{2i}^T \end{bmatrix}$$

and the values of $\hat{\mathbf{r}}_i = E(\mathbf{r}_i | \mathbf{y}_i, \mathbf{r}_i, \hat{\boldsymbol{\theta}})$ and $\widehat{\mathbf{r}_i \mathbf{r}_i^T} = E(\mathbf{r}_i \mathbf{r}_i^T | \mathbf{y}_i, \mathbf{r}_i, \hat{\boldsymbol{\theta}})$ depend on whether the subjects informatively dropped out of the study or were censored. If a subject dropped out, then his or he expected value is simply the observed dropout time. If the subject was censored at c_{ij}, then we need expressions for $\hat{r}_{ij} = E(r_{ij} | \mathbf{y}_i, r_{ij} > c_{ij}, \hat{\boldsymbol{\theta}})$ and $E(r_{ij}\mathbf{r}_i | \mathbf{y}_i, r_{ij} > c_{ij}, \hat{\boldsymbol{\theta}})$. These can be found using the properties of truncated normal distributions and $\boldsymbol{\Sigma}_{\mathbf{r}_i}^{-1/2}(\mathbf{r}_i - \mathbf{X}_{2i}\boldsymbol{\alpha}_2) \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$.

Using these, we can also calculate $E(r_{ij}\mathbf{u}_i | y, \hat{\boldsymbol{\theta}})$,

$$E(r_{ij}\mathbf{u}_i | y, \hat{\boldsymbol{\theta}}) = \begin{cases} r_{ij}\hat{\mathbf{u}}_i & \text{if } \delta_{ij} = 1 \\ \mathbf{C}_{12}\mathbf{C}_{22}^{-1}(i) \begin{bmatrix} \hat{r}_{ij}(\mathbf{y}_i - \mathbf{X}_{1i}\boldsymbol{\alpha}_1) \\ E(r_{ij}\mathbf{r}_i | \cdots) - \hat{r}_{ij}\mathbf{X}_{2i}\boldsymbol{\alpha}_2 \end{bmatrix} & \text{if } \delta_{ij} = 0. \end{cases}$$

Note also that $\mathbf{C}_{22}^{-1}(i)$ can be calculated efficiently, without the inversion of a matrix of rank $\sum_j (n_{ij} + 1)$ for each i, because of its block compound symmetric structure.

To calculate the expectation of the remaining sufficient statistic, we use the multivariate normal distribution of $\phi(\varepsilon_{ij}, \mathbf{y}_i, \mathbf{r}_i)$. Thus

$$E(\varepsilon_{ij} | \mathbf{y}_{ij}, r_{ij}, \hat{\boldsymbol{\theta}}) = [\, \mathbf{0}^T \quad \Omega_{\boldsymbol{\varepsilon}} \,] \mathbf{C}_{22}^{-1}(i) \begin{bmatrix} \mathbf{y}_i - \mathbf{X}_{1i}\hat{\boldsymbol{\alpha}}_1 \\ \hat{\mathbf{r}}_i - \mathbf{X}_{2i}^T\hat{\boldsymbol{\alpha}}_2 \end{bmatrix},$$

$$\Longrightarrow$$

$$\operatorname{cov}(\varepsilon_{ij} | \mathbf{y}_{ij}, r_{ij}, \hat{\boldsymbol{\theta}}) = s^2 - [\, \mathbf{0}^T \quad \Omega_{\boldsymbol{\varepsilon}} \,] \mathbf{C}_{22}^{-1}(i) [\, \mathbf{0}^T \quad \Omega_{\boldsymbol{\varepsilon}} \,]^T$$

$$E(\varepsilon_{ij}^2 | \mathbf{y}_{ij}, r_{ij}, \hat{\boldsymbol{\theta}}) = [\, \mathbf{0}^T \quad \Omega_{\boldsymbol{\varepsilon}} \,] \mathbf{C}_{22}^{-1}(i) \mathbf{A}_i \mathbf{C}_{22}^{-1}(i) [\, \mathbf{0}^T \quad \Omega_{\boldsymbol{\varepsilon}} \,]^T + \operatorname{cov}(\varepsilon_{ij} | \mathbf{y}_{ij}, r_{ij}, \hat{\boldsymbol{\theta}}),$$

where $\mathbf{A}_i$, $\mathbf{C}_{22}(i)$, and $\hat{\mathbf{r}}_i$ are as given previously.

[Received September 2002. Revised March 2004.]

REFERENCES

Bruce, M., and Pearson, J. (1999), "Designing an Intervention to Prevent Suicide: PROSPECT (Prevention of Suicide in Primary Care Elderly: Collaborative Trial)," *Dialogues in Clinical Neuroscience*, 1, 100–112.

Clayton, D., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 148, 82–117.

Daniels, M., and Hogan, J. W. (2000), "Parameterizing Pattern-Mixutre Model for Sensitivity Analysis Under Informative Dropout," *Biometrics*, 56, 1241–1248.

De Gruttola, V., and Tu, X. (1994), "Modelling Progressing of CD4-Lymphocyte Count and Its Relationship to Survival Time, *Biometrics*, 50, 1003–1014.

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Diggle, P., and Kenward, M. G. (1994), "Informative Dropout in Longitudinal Data Analysis" (with discussion), *Applied Statistics*, 43, 49–94.

Hogan, J., and Laird, N. (1997a), "Mixture Models for the Joint Distribution of Repeated Measures and Event Times," *Statistics in Medicine*, 16, 239–257.

——— (1997b), "Model-Based Approaches to Analysing Incomplete Longitudinal and Failure Time Data," *Statistics in Medicine*, 16, 259–272.

Kenward, M., and Molenberghs, G. (1999), "Parametric Models for Incomplete Continuous and Categorical Longitudinal Data," *Statistical Methods in Medical Research*, 8, 51–83.

Laird, N., and Ware, J. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Little, R. (1993), "Pattern-Mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association*, 88, 125–134.

——— (1994), "A Class of Pattern-Mixture Models for Normal Missing Data," *Biometrika*, 81, 471–483.

——— (1995), "Modeling the Drop-Out Mechanism in Repeated-Measures Studies," *Journal of the American Statistical Association*, 90, 1112–1121.

Little, R., and Wang, Y. (1996), "Pattern-Mixutre Models for Multivariate Incomplete Data With Covariates," *Biometrics*, 52, 98–111.

Louis, T. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 44, 226–233.

Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T., and Thijs, H. (2002), "Selection Models and Pattern-Mixture Models to Analyse Longitudinal Quality of Life Data Subject to Drop Out," *Statistics in Medicine*, 21, 1023–1041.

Miller, M., Rejeski, W., Reboussin, B., Ten Have, T., and Ettinger, W. (2000), "Physical Activity, Functional Limitation, and Activity Restrictions in Older Adults," *Journal of the American Geriatrics Society*, 48, 1264–1272.

Oslin, D., Streim, J., Katz, I., Edell, W., and Ten Have, T. (2000), "Change in Disability Follows Inpatient Treatment for Late-Life Depression," *Journal of the American Geriatrics Society*, 48, 357–362.

Schluchter, M. (1992), "Methods for the Analysis of Informatively Censored Longitudinal Data," *Statistics in Medicine*, 11, 1861–1870.

Schluchter, M., Greene, T., and Beck, G. (2001), "Analysis of Change in the Presence of Informative Censoring: Application to a Longitudinal Clinical Trial of Progressive Renal Disease," *Statistics in Medicine*, 20, 1197–1214.

Ware, J., Snow, K., Kosinski, M., and Gandek, B. (1993), *SF-36 Health Survey Manual and Interpretation Guide*, Boston, MA: The Health Institute, New England Medical Center.

Wu, M., and Carroll, R. (1988), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process," *Biometrics*, 44, 175–188.