

Mixed effect models for longitudinal responses: Leiden 85-Plus Study

Marco Alfò, Jeanine J. Houwing-Duistermaat, and Alessandra Spagnoli

Abstract The Mini Mental State Examination index measures the level of the cognitive functioning of older population. We discuss a case study where the analyzed sample includes partial respondents, and a significant source of missing data is death. Such missing data may be non ignorable for likelihood based inference. We propose to adopt a Pattern Mixture model to account for the dependence between missing data mechanism and primary response process. The approach is applied to the Leiden 85-Plus Study.

Key words: longitudinal binomial responses, variance component models, droup-out

1 Introduction

The Mini-Mental Status Examination (MMSE) [7] index is often used to assess the cognitive mental status in older adults. The MMSE is based on a standard questionnaire, and the sum of scores on each (binary) outcome of the questionnaire defines the cognitive mental status of the subjects. In most

Alessandra Spagnoli

Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma, Italy e-mail: alessandra.spagnoli@uniroma1.it

We gratefully acknowledge Dr Ton de Craen and Dr Rudi Westendorp of the Leiden University Medical Centre, for kindly providing the analyzed data.

clinical studies, the progression of cognitive impairment is defined by comparing the MMSE index with a reference cut-off chosen on the basis of population based norms [9]. Studies that involve the older population may have partial respondents and the major cause of nonresponse may be death. To handle partial respondents, in gerontology studies, two approaches are considered, i.e. the complete cases analysis (CC) [4] and the missing-as-incorrect (MAI) approach where missing answers are considered as incorrect. The CC analysis is rarely acceptable and the MAI analysis is an efficient strategy only if the missing mechanism is completely a random (MCAR). In studies where the cognitive function is assessed through the MMSE index, it is difficult to motivate the MCAR assumption since failure to be observed depends on the cognitive function of the subjects; in this case, the missing data can be non-ignorable (NI) [3]. This paper discuss the Leiden 85-Plus study through a variance component (VC) models taking into account dependence between repeated measures and the individual initial conditions. The complex form for the resulting log-likelihood is handled through a simple reparametrization of the baseline model. Further, a pattern mixture (PM) approach is introduced to assess whether missing data may be NI. The paper is organized as follow: in Section 2 we report some details on the “Leiden 85-Plus Study” data; in Section 3 we introduce the general VC model for repeated responses recorded on the same subject handling the initial conditions problem and discuss the PM model presenting the corresponding results. The last section gives concluding remarks and discuss potential future research lines.

2 The motivating example

The data that motived this article are drawn from the Leiden 85-Plus Study carried out by the Leiden University Medical Centre. Between September 1997 and September 1999, 705 inhabitants of Leiden, The Netherlands, reached the age of 85 years, and in the month after their 85th birthday, they were asked to participate to the study. No one was excluded on the basis of health, cognitive functioning or living situations. Of the 705 eligible subjects 14 died before they could be enrolled and 92 refused to participate, resulting in a cohort of 599 enrolled subjects [2]. Given the study characteristics, we

have left-truncated, right-censored survival data on subject with the following covariates collected at entry time: gender, educational status (two levels: primary and higher). Participants were visited at their place of residence, where face-to-face interviews were conducted through a questionnaire whose items are designed to assess orientation, attention, language skills and the ability to perform simple actions. The scores on each outcome was binary (e.g. 1 for a correct answer); the cognitive mental status of a subject was assessed by summing the scores that she/he obtained on the questionnaire items and the corresponding range includes all integer values between 0 and 30 (number of items). MMSE index measurements are missing for some subjects due to the effect of factors such as poor physical health, depression, anxiety or because the subject died.

3 The modeling approaches

Variance component model In the present study, the data consists of a set of longitudinal measurements, y_{it} , recorded to the i th subject at time $t = (1, \dots, T)$ and a row vector of p observed covariates $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})$. The MMSE index is defined as $y_{it} = \sum_{j=1}^{30} y_{ijt}$ where y_{ijt} represents the MMSE score on the j th item for the i th subject at time t . We assume that the item-specific complexity is constant across items and time and that, conditional on subject-specific random effect $\mathbf{b}_i$ and on the baseline value (the MMSE index at age 85) say Y_{i1} , the responses y_{it} independent Binomial rv with $n = 30$ and parameter π_{it} . The corresponding canonical parameter is defined by the following linear function:

$$\theta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i + \alpha y_{i1} \quad (1)$$

The terms $\mathbf{b}_i$, $i = 1, \dots, n$, represent unobserved individual-specific sources of heterogeneity common to each lower-level unit (time) within the same i th upper-level unit (individual), while $\boldsymbol{\beta}$ is a p -dimensional vector of fixed regression parameters. Those effects that vary across individuals are collected in the design vector $\mathbf{z}_{it} = (z_{it1}, \dots, z_{itq})$. The distribution of $\mathbf{b}_i$ may be defined by adopting a particular parametric specification, for example, $\mathbf{b}_i \sim$

MVN($\mathbf{0}, \Sigma$), or not. Since the $\mathbf{b}_i$'s are shared by all subject i 's responses, the likelihood function results from the following integral:

$$L(\cdot) = \prod_{i=1}^n \int \prod_{t=2}^T f(y_{it}|y_{i1}, \mathbf{b}_i) f(y_{i1}|\mathbf{b}_i) g(\mathbf{b}_i) d\mathbf{b}_i$$

The term $f(y_{i1}|\mathbf{b}_i)$ is not specified by the model assumptions and we have to modify the standard VC approach to account for this modeling structure. The model for $t \geq 2$ is defined by eq(1) while, for the first occasion, we assume:

$$\theta_{i1} = \mathbf{x}_{i1}^\top \beta_1 + \mathbf{z}_{it}^\top \mathbf{b}_{i1}$$

with $\mathbf{b}_{i1} = \Lambda \mathbf{b}_i$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$, $\text{cov}(\mathbf{b}_{i1}) = \Lambda \Sigma \Lambda^\top$. The random effect covariance matrix may be different on later occasions and the same applies to the fixed parameter vector. This can be handled by defining a dummy variable $d_t = 1$, $t \geq 2$, and interacting it with appropriate terms. Table I displays the results obtained by fitting the VC model, where the standard errors have been calculated by inverting the observed information matrix [1]. All fitted models lead to similar conclusions: subjects with at most primary school level are more cognitively impaired than the subjects with higher educational level; at the same time age negatively influences the cognitive function while gender effect is not significant. M_1 represents a standard VC model while in M_2 and M_3 the baseline effect is included. The link between models for $t \geq 2$ and $t = 1$ is handled through a shared (proportional) random effect or correlated random effects in M_2 and M_3 respectively. In both models the effect of baseline MMSE is significant. When the baseline is considered as a covariate in the model, the effect of the educational factor decreases.

As far as the random effect variance is concerned, we may notice that, moving from M_1 to M_3 , the corresponding estimate significantly decreases, showing a certain degree of overlapping between $\mathbf{b}_i$ and Y_{i1} .

Pattern mixture model Longitudinal studies may suffer from the problem of attrition or drop out, i.e. some individuals "drop out" from the trial before its completion. In the present case study, 46.9% (276 out of 599) of the enrolled subjects have incomplete sequences; thus, we may question whether those missing data are ignorable. For each individual, let T_i represent the

		VC model		Baseline VC model		Baseline VC model	
		M_1		M_2		M_3	
Variable		Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
y_{it}	Constant	1.352	0.078	-3.347	0.118	-3.221	0.081
$t \geq 2$	age	-0.194	0.007	-0.190	0.007	-0.195	0.009
	gender	0.229	0.121	0.011	0.060	0.008	0.074
	educational	1.024	0.121	0.201	0.061	0.272	0.072
	y_{i1}			0.209	0.004	0.203	0.003
$t = 1$	gender			-0.025	0.069	-0.051	0.047
	educational			0.314	0.074	0.117	0.052
	σ_b^2	1.781		0.326		0.545	
	$\sigma_{b_1}^2$					0.461	
	Log-likelihood	-2756		-2293		-1761	
	BIC	5551		4650		3600	

Table 1 Parameter estimates for Variance component models

drop out time, if $T_i < T$ then the i th individual is a drop out, otherwise is a study completer. In many trials, drop out is known only to occur in a interval between t_j and t_{j+1} for some j . Then, T_i is often considered to be discrete and occurring (arbitrarily) at t_j . Let R_i denote a missing data indicator, defined as $R_{it} = 1$ if the i th unit drop outs at any point within $(t-1, t)$, $t = 1, \dots, T$ and $R_{it} = 0$ otherwise. The main question is whether the subject that drop out differ from the subjects that still participate to the study. We may introduce a model for the drop out mechanism process, assuming conditional independence given the (shared) random effects and conditioning on a set of drop out specific covariates $\mathbf{v}_i$: $h(\mathbf{r}_i | \mathbf{v}_i, \mathbf{y}_i, \mathbf{b}_i) = h(\mathbf{r}_i | \mathbf{v}_i, \mathbf{b}_i)$, $i = 1, \dots, n$. If we denote by $\mathcal{T} = \min(T, T_i + 1)$, the joint density for the primary response and the missing-data is:

$$f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_{it}, \mathbf{v}_i; \mathbf{b}_i) = \int \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \mathbf{b}_i) \prod_{t=1}^{\mathcal{T}} h(r_{it} | \mathbf{v}_{it}, \mathbf{b}_i) \right] dG(\mathbf{b}_i) \quad (2)$$

Supposing that the primary outcome and the dropout mechanism are affected by different (possibly correlated) sources of heterogeneity, the joint density for the primary response and the missing-data indicator becomes:

$$f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_{it}, \mathbf{v}_i; \mathbf{b}_i) = \int \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \mathbf{b}_{i1}) \prod_{t=1}^{\tau} h(r_{it} | \mathbf{v}_{it}, \mathbf{b}_{i2}) \right] dG(\mathbf{b}_i) \quad (3)$$

where $\mathbf{b}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2})$ denote a subject and outcome-specific vector of random coefficients. When (3) is adopted and $(\mathbf{b}_{i1}, \mathbf{b}_{i2})$ are dependent, Little [8] shows that drop out depends on the unobserved component of $\mathbf{Y}_i$, i.e. $\mathbf{R}_i$ depends on $\mathbf{Y}_i^{mis}$ through $\mathbf{b}_i$. Therefore, the corresponding likelihood can not be factorized and the missing data mechanism is said to be NI. In this context, statistical inference is mostly based on selection or PM model approach. To define a PM, we summarize the missing-data indicator by:

$$S_i = T - \sum_{t=1}^T R_{it} = \sum_{t=1}^T (1 - R_{it})$$

indicating the number of observed responses for the i -th individual, $S_i \leq T$. We consider model (1) and assume $\mathbf{z}_{it} \equiv 1$; according to Wu and Bailey [6], we assume the following model holds: $\zeta[E(S_i | b_i)] = \mathbf{v}_{it}^T \boldsymbol{\phi} + b_i$ where $\boldsymbol{\phi}$ represents the corresponding fixed model parameter vector and $\mathbf{b}_i$ is a shared random effect. The joint distribution for the observed data and the time to drop out is therefore:

$$f(y_{i1}, \dots, y_{iT}, s_i) = \int \prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, b_i) h(s_i | \mathbf{v}_i, b_i) dG(b_i)$$

To avoid misspecification of the missing-data mechanism, Alf  and Aitkin [5] proposed to focus on the conditional distribution $f(\mathbf{y}_i | s_i)$. The corresponding model is a PM model, with the loglikelihood function defined by:

$$\ell(\cdot | s_i) = \sum_{i=1}^n \log \left\{ \int \prod_{t=1}^T f(y_{it} | \mathbf{x}_{it}, b_i) dG(b_i | s_i) \right\}$$

Rather than assuming a parametric specification for $G(b_i | s_i)$, the conditioning is handled through a simple parametric change in the b_i 's mean: $b_i = b_i^* + E[b_i | s_i]$ where: $E[b_i | s_i] = \alpha_0 + \alpha_1 s_i$. Corresponding results have to be interpreted as conditional on the observed time to drop out, and s_i enters the linear predictor as an additional covariate. Table II displays the results obtained by fitting the PM model to the Leiden 85+ study data. These results agree with those displayed in Table I with respect to age, gender and

		Pattern mixture M_1		Pattern mixture M_2	
Variable		Coeff.	Std. Err.	Coeff.	Std. Err.
y_{it} $t \geq 2$	Constant	-3.484	0.124	-3.262	0.086
	age	-0.199	0.008	-0.203	0.009
	gender	0.012	0.060	-0.024	0.076
	educational	0.187	0.060	0.220	0.0752
	s	0.059	0.016	0.037	0.014
	y_{i1}	0.204	0.005	0.201	0.004
$t = 1$	gender	0.032	0.071	-0.021	0.048
	educational	0.318	0.075	0.131	0.052
	s	0.051	0.017	0.016	0.011
σ_b^2		0.317		0.535	
$\sigma_{b_1}^2$				0.456	
Log-likelihood		-2287		-1758	
BIC		4653		3609	

Table 2 Parameter estimates for Pattern mixture models

baseline MMSE parameter estimates. As before, models M_1 and M_2 in Table II consider shared and correlated random effects in the first and subsequent occasions, respectively. The introduction of the variable that represents the propensity to remain into the study influences the MMSE value. Thus, the subject who continue to participate to the study have higher MMSE index and are in a better health status; at the same time, estimates of the effect that represents the educational status decreases when we introduce S . This means that the effect of educational status likely includes a part of the effect of S , that is, subjects with higher MMSE index values and a higher educational level tend to remain longer in the study. This evidence may suggest that the missing data are NI. As can be evidenced model M_2 provides a better fit in terms of both loglikelihood and BIC values than model M_1 suggesting that a shared random effect model is questionable.

4 Concluding remarks

Motivated by a specific case study, we have presented a model PM for longitudinal binomial responses to account for potentially NI drop out mechanism. The results show that ignoring the influence of drop out may lead to biased estimates in the standard VC models. The shared random effects hypothesis in the PM model may be not well supported. Further research lines will focus on selection models handling discrete and continuous time survival data, and an additional genetic covariates.

References

1. M Alfö and M Aitkin. Random coefficient models for binary longitudinal responses with attrition. *Statistics and Computing*, 10:279–287, 2000.
2. RM Crum, JC Anthony, SS Bassett, and MF Folstein. Population-based norms for the Mini-Mental State Examination by age and educational level. *Journal of American Medical Association*, 63:81–92, 1993.
3. AB der Wiel, E van Exel, AJ de Craen, J Gussekloo, AM Lagaay, DL Knook, and RGJ Westendorp. A high response is not essential to prevent selection bias: Results from the Leiden 85-plus study. *Journal of Clinical Epidemiology*, 55:1119–1125, 2002.
4. HH Dodge, K Meguro, H Ishii, S Yamaguchi, JA Saxton, and M Ganguli. Cross-cultural comparisons of the Mini-mental State Examination between Japanese and U.S. cohorts. *International Psychogeriatrics*, 21:113–122, 2009.
5. MF Folstein, SE Folstein, and PR McHig. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research*, 12:189–198, 1975.
6. F Lagona and Z Zhang. A missing composite covariate in survival analysis: A case study of the Chinese Longitudinal Health and Longevity Survey. *Statistics in Medicine*, 29:248–261, 2009.
7. RJA Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90:1112–1121, 1995.
8. TA Louis. Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 44:226–233, 1982.
9. MC Wu and KR Bailey. Estimation and comparison of changes in the presence of informative censoring by modeling the censoring process. *Biometrics*, 44:175–188, 1989.