Published in Statistics in Medicine, Volume 37, Issue 20, Pages 2998-3011

A bi-dimensional finite mixture model for longitudinal data subject to dropout

Alessandra Spagnoli[1], Maria Francesca Marino[2], and Marco Alfò [3]

[1]*Dipartimento di Sanità Pubblica e Malattie Infettive, Sapienza Università di Roma*

[2]*Dipartimento di Statistica, Informatica, Applicazioni, Università degli Studi di Firenze*

[3]*Dipartimento di Scienze Statistiche, Sapienza Università di Roma*

Abstract

In longitudinal studies, subjects may be lost to follow-up and, thus, present incomplete response sequences. When the mechanism underlying the dropout is nonignorable, we need to account for dependence between the longitudinal and the dropout process. We propose to model such a dependence through discrete latent effects, which are outcome-specific and account for heterogeneity in the univariate profiles. Dependence between profiles is introduced by using a probability matrix to describe the corresponding joint distribution. In this way, we separately model dependence *within* each outcome and dependence *between* outcomes. The major feature of this proposal, when compared to standard finite mixture models, is that it allows the nonignorable dropout model to properly nest its ignorable counterpart. We also discuss the use of an index of (local) sensitivity to nonignorability to investigate the effects that assumptions about ignorability of the dropout process may have on model parameter estimates. The proposal is illustrated via the analysis of data from a longitudinal study on the dynamics of cognitive functioning in the elderly.

Keywords: Informative missingness, latent variables, nonparametric maximum likelihood, sensitivity to nonignorability.

1 Introduction

Longitudinal studies are frequently affected by dropout, with some units leaving the study before its planned end. Rubin provided a well-known taxonomy for mechanisms that generate incomplete sequences[1]. Looking at the impact of missing data on parameter estimates in the longitudinal data model, we may define the dropout to be nonignorable in two cases: (i) when the participation to the study still depends on future (potentially unobserved) response values, even after conditioning on the observables (both covariates and responses); (i), when the longitudinal and the dropout model share (completely or partially) model parameters. Such a phenomenon may bias the study design and inference can not be based on a model for the observed responses only. Rather, we need to take into account the mechanism leading to missing data. For this purpose, we focus on the class of Random Coefficient Based Dropout Models[2] (RCBDMs). Two separate (conditional) models are built for the longitudinal response and the dropout indicator; rather than establish a rigid parametric dependence structure, the link between the two processes is assumed to raise due to common or dependent random coefficients, with a given (often continuous) density function. The choice for such a distribution is not obvious and, in the past years, has attracted the attention of several authors. Frequently, the random coefficients are assumed to be Gaussian random variables, and the approach is fully parametric[3,4]. However, this assumption was questioned by several authors[5], since the resulting inference can be highly sensitive, and it may have a strong impact on parameter estimates, especially in the case of short longitudinal sequences. For this reason, the random coefficient distribution may be left unspecified and estimated via a NonParametric Maximum Likelihood (NPML) approach, in a finite mixture context[6–8]. More elaborated approaches have been discussed by Beuckens et al.[9] and Bartolucci and Farcomeni[10], where finite mixtures of regression models with either time-constant Gaussian (partially shared) random effects or time-varying intercepts with Markovian structure are discussed. In the present paper, we start from a simple semi-parametric approach, where the random coefficient distribution is left unspecified and estimated through a discrete distribution. This leads to a finite mixture model which, however, has the substantial drawback that dependence *within* outcomes can not be separated by dependence *between* them. That is, the Missing Not At Random (MNAR) model does not reduce, at least not simply, to the corresponding Missing At Random (MAR) counterpart. Furthermore, if we model

outcomes of mixed type (e.g. Gaussian longitudinal responses and binary missing indicators), they may have a different impact on the *global* log-likelihood. As a consequence, heterogeneity in the different profiles may be fitted with lower/higher precision, based on the corresponding weights.

Starting from such drawbacks, we suggest to consider two separate sets of discrete random coefficients to account for dependence within the longitudinal and the dropout process. Dependence between profiles is modelled by specifying a matrix of prior probabilities connecting each component in the longitudinal data model to each component in the dropout model. when the two are independent, the matrix simply reduces to the product of the corresponding marginals. That is, unlike standard finite mixture models, the proposed MNAR model properly nests the MAR counterpart. As highlighted by Molenberghs et al.[11], for every MNAR model there is a corresponding MAR counterpart that produces exactly the same fit to the observed data; therefore, a sensitivity analysis is always recommended. We propose to explore sensitivity via the *index of local sensitivity to nonignorability* (ISNI)[12–16].

The structure of the paper follows. In section 2, we introduce the motivating application: the Leiden 85+ study. Section 3 discusses general RCBDMs, while our proposal is detailed in Section 4. Sections 5 and 6 present the proposed Expectation Maximization (EM) algorithm for maximum likelihood estimation and the index of local sensitivity we propose. Section 7 shows the application of the proposed model to the Leiden85+ data and the results from the sensitivity analysis. Last section contains concluding remarks.

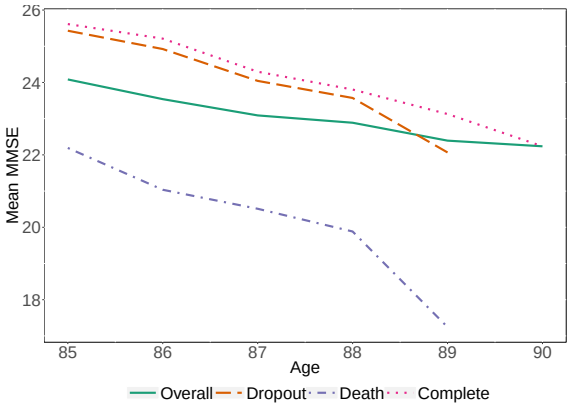
2 Motivating example: Leiden 85+ data

The motivating data come from the Leiden 85+ study, a retrospective study entailing 705 Leiden inhabitants (in the Netherlands), who reached the age of 85 years between September 1997 and September 1999. The study aimed at identifying demographic and genetic determinants for the dynamics of cognitive functioning in the elderly. We considered the following covariates collected at the beginning of the study: gender, educational status (primary/higher education, corresponding to less than/at least 7 years of schooling, respectively), and plasma Apolipoprotein E (APOE) genotype. Three different APOE alleles can be observed: $\epsilon 2, \epsilon 3, \epsilon 4$; these lead to 6 possible combinations (genotypes): $\epsilon 2/\epsilon 2$, $\epsilon 2/\epsilon 3$, $\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 3$, $\epsilon 3/\epsilon 4$, $\epsilon 4/\epsilon 4$. Carrying at least one $\epsilon 2$ allele is

known to reduce the risk of developing dementia, whereas the $\epsilon 4$ allele is known to be linked to an increased risk, see eg[17]. The $\epsilon 2$ allele is he rarest as it appears in about the 7% of the population; allele $\epsilon 4$ is carried by about 14% of the population, while $\epsilon 3$ allele is the most common. Due to prevalence and to their potential use as risk factors, we decided to group the possible combinations in four categories: $APOE_{22-23}$, $APOE_{24}$, $APOE_{33}$ (reference), $APOE_{34-44}$. Only 541 subjects present complete covariate information and will be considered in the analysis.

Study participants were visited yearly until the age of 90 at their place of residence and 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 Mini Mental State Examination index, in the following MMSE[18], is obtained by summing the scores on the items designed to assess potential cognitive impairment. The observed values are integers ranging between 0 and 30 (maximum total score). With the aim at understanding how the MMSE score evolves over time, we show in Figure 1 the corresponding overall mean value across follow-up visits. We also represent the evolution of the mean MMSE stratified by participation in the study (completer, dropout/death before next occasion).

Figure 1: Mean MMSE over time stratified by subjects' participation to the study.



As it is clear, cognitive functioning levels in individuals who die are much lower than those corresponding to subjects who dropout for other reasons or participate until the study end. While the decline in cognitive functioning through time seems to be (at least approximately) constant across groups, the slope for the overall mean tends to the corresponding value for *completers*, as the time passes by. Such a finding highlights a potential dependence between the evolution of the MMSE score over time and the dropout process, which may bias the parameter estimates and the

corresponding inference. In the next section, we introduce a class of models for the *joint* analysis of the longitudinal and the dropout process.

3 Random coefficient-based dropout models

Let Y_{it} be a longitudinal response recorded on $i = 1, \ldots, n$, subjects at time occasions $t = 1, \ldots, T$, and let $\mathbf{x}_{it} = (x_{it1}, \ldots, x_{itp})'$ denote an observed p-dimensional covariate vector with $x_{it1} = 1$. Let us assume that, conditional on a q_1-dimensional set of individual-specific random coefficients $\mathbf{b}_{i1} = (b_{i11}, \ldots, b_{i1q_1})'$, the observed responses Y_{it} are independent random variables with density in the Exponential Family. The canonical parameter for this density, θ_{it} is described by the regression model:

$$\theta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{w}'_{it}\mathbf{b}_{i1}. \tag{1}$$

The terms $\mathbf{b}_{i1}$, $i = 1, \ldots, n$, represent sources of (individual-specific and time-constant) unobserved heterogeneity; the vector $\boldsymbol{\beta}$ denotes a $p-$dimensional vector of regression parameters common to all individuals. The covariates whose effects are assumed to vary with i are collected in the vector $\mathbf{w}_{it} = (w_{it1}, \ldots, w_{itq})$, with $\mathbf{w}_{it} \subseteq \mathbf{x}_{it}$. For identifiability purposes, standard assumptions on the random coefficient vector postulate that $\mathrm{E}(\mathbf{b}_{i1}) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{b}_{i1}) = \mathbf{D}$, $i = 1, \ldots, n$. Dropout is a frequent issue with longitudinal studies, since some individuals do not reach the planned end participating only to $T_i \leq T$ measurement occasions, $i = 1, \ldots, n$. Let R_{it} denote the missing data indicator which is equal to 1 if the i-th unit drops-out at any point in the window $(t-1, t)$, and is equal to 0 otherwise; clearly, when we consider dropout, $R_{it} = 1 \Longrightarrow R_{it'} = 1, t' > t = 1, \ldots, T$. Here, we consider a discrete time structure for the study and the time to dropout; however, the following arguments may apply, with a limited number of changes, to continuous time survival process as well. To describe the (potential) dependence between the longitudinal response $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT_i})'$ and the dropout indicator $\mathbf{R}_i = (R_{i1}, \ldots, R_{iT_i+1})'$ for the same subject $i = 1, \ldots, n$, we augment model in eq. 1 with an explicit model for the dropout mechanism, conditional on a set of dropout-specific covariates, say $\mathbf{v}_i$, and a further set of random parameters $\mathbf{b}_{i2} = (b_{i21}, \ldots, b_{i2q_2})$:

$$\begin{aligned} R_{it} \quad &| \quad \mathbf{b}_{i2}, \mathbf{v}_i \sim \mathrm{Bin}\,(1, \phi_{it}) \\ \mathrm{logit}(\phi_{it}) \quad &= \quad \mathbf{v}'_{it}\boldsymbol{\gamma} + \mathbf{d}'_{it}\mathbf{b}_{i2} \end{aligned} \tag{2}$$

where $\mathbf{d}_{it} = (d_{it1}, \ldots d_{itq_2})$ and $\mathbf{d}_{it} \subseteq \mathbf{v}_{it}$. The density of the individual sequence is given by:

$$h(\mathbf{r}_i \mid \mathbf{v}_i, \mathbf{y}_i, \mathbf{b}_{i2}) = h(\mathbf{r}_i \mid \mathbf{v}_i, \mathbf{b}_{i2}) = \prod_{t=1}^{T_i^*} h(r_{it} \mid \mathbf{v}_i, \mathbf{b}_{i2}), \qquad i = 1, \ldots, n \tag{3}$$

where $T_i^\star = \min(T, T_i + 1)$ and conditional independence holds.

Based on such modelling assumptions, the joint density for the couple $(\mathbf{Y}_i, \mathbf{R}_i)$ can be built up as follows:

$$f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{X}_i, \mathbf{V}_i) = \int \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \mathbf{b}_{i1}) \prod_{t=1}^{T_i^*} h(r_{it} | \mathbf{v}_{it}, \mathbf{b}_{i2}) \right] dG(\mathbf{b}_{i1}, \mathbf{b}_{i2}). \tag{4}$$

where $T_i^\star = \min(T, T_i + 1)$ and $G(\mathbf{b}_{i1}, \mathbf{b}_{i2})$ denotes the joint density of $\mathbf{b}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2})$. When $\boldsymbol{b}_{i1} = \boldsymbol{b}_{i2}$, or, a bit more generally, $\mathbf{b}_{i2} = \mathbf{C}\mathbf{b}_{i1}$, with $\mathbf{C} \in \mathcal{M}(q_2, q_1)$, $q_2 \leq q_1$, the model reduces to the so called *shared parameter model*[19,20]. See also Creemers et al.[21] for a more general proposal. Frequently, a Gaussian distribution is considered. However, while the effect of misspecifying $G(\mathbf{b}_{i1}, \mathbf{b}_{i2})$ is relatively small when dealing with a large number of repeated measurements, assumptions on such a distribution may play a crucial role in the presence of short individual sequences, such as those coming from clinical studies[22]. A semi-parametric approach based on discrete, outcome-specific, random coefficients cam be used as an alternative[6–8]. In this case, the joint density for the longitudinal and the missing data process reduces to

$$f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{X}_i, \mathbf{V}_i) = \sum_{k=1}^{K} \left[\prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\zeta}_{1k}) \prod_{t=1}^{T_i^*} h(r_{it} \mid \mathbf{v}_{it}, \boldsymbol{\zeta}_{2k}) \right] \pi_k. \tag{5}$$

As it is clear, the above density resembles that of a finite mixture with K different components characterized by locations $\boldsymbol{\zeta}_{1k}$ and $\boldsymbol{\zeta}_{2k}$ (in the longitudinal and the missing data process, respectively) and masses $\pi_k = \Pr(\mathbf{b}_i = \boldsymbol{\zeta}_k) = \Pr(\mathbf{b}_{i1} = \boldsymbol{\zeta}_{1k}, \mathbf{b}_{i2} = \boldsymbol{\zeta}_{2k})$. In this context, the distribution $G(\mathbf{b}_{i1}, \mathbf{b}_{i2})$ is estimated nonparametrically by a discrete distribution defined on $K \leq n$ locations. While the estimation of a (possibly continuous) distribution via a discrete distribution may seem unsound, it is worth noticing that similar discretizations are used in all approaches that use numerical integration techniques to approximate (non analytically tractable) integrals in eq. (4). The advantage of the finite mixture approach is that locations (integration nodes) and prior probabili-

ties (masses) are not fixed in advance, but are estimated from (and are therefore optimal for) the observed data. As we may notice by looking at eq. (5), the discrete latent variable describing the random coefficient distribution is intrinsically uni-dimensional. That is, while the locations may differ across profiles, their number (K) and the prior probabilities (π_k) are common to both the longitudinal and the missingness process. As a result, the model does not reduce the corresponding MAR counterpart in the case of ignorable dropout, but in very particular cases (e.g. when $K = 1$ or when either $\boldsymbol{\zeta}_{1k} = \boldsymbol{\zeta}_1$ or $\boldsymbol{\zeta}_{2k} = \boldsymbol{\zeta}_2$, $\forall k = 1, \ldots, K$). In the view of *sensitivity* analysis, this may be a crucial drawback. In the next section, we extend the model following an approach similar to that detailed by Alfó and Rocchetti[23] and Lagona[24] with the aim at addressing such issue.

4 A bi-dimensional finite mixture approach

Let us assume that the distribution of the random coefficients in the longitudinal data model $\mathbf{b}_{i1}$ is estimated by a discrete distribution defined on K_1 support points $\{\boldsymbol{\zeta}_{11}, \ldots, \boldsymbol{\zeta}_{1K_1}\}$ with masses $\pi_{g\star} = \Pr(\mathbf{b}_{i1} = \boldsymbol{\zeta}_{1g})$. Similarly, let us assume that the distribution of individual-specific random coefficients $\mathbf{b}_{2i}$ in the missing data model is estimated via a discrete distribution with K_2 distinct support points $\{\boldsymbol{\zeta}_{21}, \ldots, \boldsymbol{\zeta}_{1K_2}\}$ with masses $\pi_{\star l} = \Pr(\mathbf{b}_{i2} = \boldsymbol{\zeta}_{2l})$. That is, we assume that

$$\mathbf{b}_{1i} \sim \sum_{g=1}^{K_1} \pi_{g\star}\, \delta(\boldsymbol{\zeta}_{1g}) \quad \mathbf{b}_{2i} \sim \sum_{\ell=1}^{K_2} \pi_{\star\ell}\, \delta(\boldsymbol{\zeta}_{2\ell})$$

where $\delta(a)$ denotes an indicator function that puts unit mass at a. We then define a joint distribution for the random coefficients, with a mass $\pi_{g\ell} = \Pr(\mathbf{b}_{i1} = \boldsymbol{\zeta}_{1g}, \mathbf{b}_{i2} = \boldsymbol{\zeta}_{2\ell})$ associated to each couple of locations $(\boldsymbol{\zeta}_{1g}, \boldsymbol{\zeta}_{2\ell})$. The associated probability matrix $\boldsymbol{\Pi} = \{\pi_{gl}\}$ describes the association between $\mathbf{b}_{i1}$ and $\mathbf{b}_{i2}$ and, indirectly, between $\mathbf{Y}_i$ and $\mathbf{R}_i$. Obviously, probabilities $\pi_{g\star}$ and $\pi_{\star\ell}$ in the univariate profiles are obtained by marginalizing $\pi_{g\ell}$:

$$\pi_{g\star} = \sum_{\ell=1}^{K_2} \pi_{g\ell}, \qquad \pi_{\star\ell} = \sum_{g=1}^{K_1} \pi_{g\ell}.$$

Under the proposed model specification, the joint density in eq. (5) modifies to:

$$f(\mathbf{y}_i, \mathbf{r}_i \mid \mathbf{X}_i, \mathbf{V}_i) = \sum_{g=1}^{K_1} \sum_{\ell=1}^{K_2} \left[\prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\zeta}_{1g}) \prod_{t=1}^{T_i^*} h(r_{it} \mid \mathbf{v}_{it}, \boldsymbol{\zeta}_{2\ell}) \right] \pi_{g\ell} \tag{6}$$

When compared to a standard finite mixture model, it provides a more flexible (albeit more complex) representation for the random coefficient distribution. By looking at eq. (6), it is immediately clear that the MNAR model directly reduces to its MAR counterpart when $\pi_{g\ell} = \pi_{g\star}\pi_{\star\ell}$, $\forall g = 1, \ldots, K_1, \ell = 1, \ldots, K_2$. Let us consider the logit transform for the joint masses $\pi_{g\ell}$:

$$\xi_{g\ell} = \log\left(\frac{\pi_{g\ell}}{\pi_{K_1K_2}}\right) = \gamma + \alpha_{g\star} + \alpha_{\star\ell} + \lambda_{g\ell}, \tag{7}$$

where $\gamma = \log\left(\left[\pi_{K_1\star}\pi_{\star K_2}\right]/\pi_{K_1K_2}\right)$, $\alpha_{g\star} = \log\left(\pi_{g\star}/\pi_{K_1\star}\right)$, $\alpha_{\star\ell} = \log\left(\pi_{\star\ell}/\pi_{\star K_2}\right)$. Furthermore, the parameter $\lambda_{g\ell} = \log\left(\pi_{g\ell}/\left[\pi_{g\star}\pi_{\star\ell}\right]\right)$ provides a measures of the departure from the independence model. That is, if $\lambda_{g\ell} = 0$, for all $(g, \ell) \in \{1, \ldots, K_1\} \times \{1, \ldots, K_2\}$, then $\pi_{g\ell} = \pi_{g\star}\pi_{\star\ell}$. This corresponds to independence between the random coefficients in the two equations, and, as a by-product, to independence between the longitudinal and the dropout process. In this sense, the vector $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{K_1K_2})$ can be formally considered as a *sensitivity* parameter vector; in fact, when $\boldsymbol{\lambda} = \mathbf{0}$ the proposed MNAR model reduces to the corresponding MAR model. See[9] for a similar proposal in a two-level hierarchical framework.

5 ML parameter estimation

The observed data log-likelihood for the proposed model specification is given by

$$\ell(\boldsymbol{\Upsilon}) = \sum_{i=1}^{n} \log \left\{ \sum_{g=1}^{K_1} \sum_{\ell=1}^{K_2} \left[\prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\zeta}_{1g}) \prod_{t=1}^{T_i^*} h(r_{it} \mid \mathbf{v}_{it}, \boldsymbol{\zeta}_{2\ell}) \right] \pi_{g\ell} \right\},$$

where $\boldsymbol{\Upsilon} = \{\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\pi}\}$ denote the vector of all (free) model parameters, $\boldsymbol{\Phi} = (\boldsymbol{\beta}, \boldsymbol{\zeta}_{11}, \ldots, \boldsymbol{\zeta}_{1K_1})$, $\boldsymbol{\Psi} = (\boldsymbol{\gamma}, \boldsymbol{\zeta}_{21}, \ldots, \boldsymbol{\zeta}_{2K_2})$, and $\boldsymbol{\pi} = (\pi_{11}, \ldots, \pi_{g\ell}, \ldots, \pi_{K_1K_2})$. To maximize the expression above and estimate model parameters, we may rely on an extended EM algorithm[25]. To this purpose, let $\mathbf{z}_i = (z_{i11}, \ldots, z_{ig\ell}, \ldots, z_{iK_1K_2})$ denote the indicator vector, with $z_{ig\ell} = 1$ if the i-th individual

comes from the g-th component in the first and the ℓ-th component in the second profile. Let us further assume the random vector $\mathbf{z}_i$ follows a multinomial distribution, with parameter vector $\boldsymbol{\pi}$. To derive the estimates, we start from the complete-data log-likelihood function:

$$\ell_c(\boldsymbol{\Upsilon}) = \sum_{i=1}^{n}\sum_{g=1}^{K_1}\sum_{\ell=1}^{K_2} z_{ig\ell}\left[\sum_{t=1}^{T_i}\log f(y_{it} \mid \mathbf{x}_{it}, z_{ig\ell}=1) + \sum_{t=1}^{T_i^*}\log h(r_{it} \mid \mathbf{v}_{it}, z_{ig\ell}=1) + \log \pi_{g\ell}\right]. \quad (8)$$

At the r-th iteration of the algorithm, we compute the posterior expectation (E-step) of the complete data log-likelihood, conditional on the observed data $(\mathbf{y}_i, \mathbf{r}_i)$ and the current estimates $\hat{\boldsymbol{\Upsilon}}^{(r-1)}$:

$$Q(\boldsymbol{\Upsilon} \mid \boldsymbol{\Upsilon}^{(r-1)}) = \sum_{i=1}^{n}\sum_{g=1}^{K_1}\sum_{\ell=1}^{K_2} w_{ig\ell}^{(r)}\left[\sum_{t=1}^{T_i}\log f(y_{it} \mid \mathbf{x}_{it}, z_{ig\ell}=1) + \sum_{t=1}^{T_i^*}\log h(r_{it} \mid \mathbf{v}_{it}, z_{ig\ell}=1) + \log \pi_{g\ell}\right].$$

In the above equation, $w_{ig\ell} = \mathbb{E}(z_{ig\ell} \mid \mathbf{y}_i, \mathbf{r}_i, \hat{\boldsymbol{\Upsilon}}^{(r-1)})$ is the posterior probability of component membership. Conditional on such weights, in the M-step, we maximize $Q(\boldsymbol{\Upsilon} \mid \boldsymbol{\Upsilon}^{(r-1)})$ with respect to model parameters. That is we find the zeros of the following score functions:

$$S_c(\boldsymbol{\Phi}) = \sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\Phi}}\sum_{g=1}^{K_1}\sum_{\ell=1}^{K_2} w_{ig\ell}^{(r)}\left[\log(f_{ig\ell}) + \log(\pi_{g\ell})\right] = \sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\Phi}}\sum_{g=1}^{K_1} w_{ig\star}^{(r)}\left[\log(f_{ig\star})\right], \quad (9)$$

$$S_c(\boldsymbol{\Psi}) = \sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\Psi}}\sum_{g=1}^{K_1}\sum_{\ell=1}^{K_2} w_{ig\ell}^{(r)}\left[\log(f_{ig\ell}) + \log(\pi_{g\ell})\right] = \sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\Psi}}\sum_{\ell=1}^{K_2} w_{i\star\ell}^{(r)}\left[\log(f_{i\star\ell})\right], \quad (10)$$

$$S_c(\pi_{g\ell}) = \sum_{i=1}^{n}\frac{\partial}{\partial \pi_{g\ell}}\sum_{g=1}^{K_1}\sum_{\ell=1}^{K_2} w_{ig\ell}^{(r)}\pi_{g\ell} - \kappa\left(\sum_{g=1}^{K_1}\sum_{\ell=1}^{K_2}\pi_{g\ell} - 1\right). \quad (11)$$

where $f_{ig\ell} = \prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{it}, z_{ig\ell}=1)\prod_{t=1}^{T_i^*}\log h(r_{it} \mid \mathbf{v}_{it}, z_{ig\ell}=1)$, while $w_{ig\star}^{(r)}$, $w_{i\star l}^{(r)}$, $f_{ig\star}$ and $f_{i\star\ell}$ represent the marginals for the posterior probability $w_{ig\ell}$ and the joint density $f_{ig\ell}$, respectively. The updated estimates for the mixture probabilities can be derived analytically as

$$\hat{\pi}_{g\ell}^{(r)} = \frac{\sum_{i=1}^{n} w_{ig\ell}^{(r)}}{n}.$$

The remaining model parameters are updated via standard Newton-type algorithms. The E- and the M-step are alternated until convergence, that is, until the (relative) difference between two subsequent likelihood values is smaller than a given quantity $\varepsilon > 0$. Given that this criterion may

indicate lack of progress rather than true convergence[26] and the log-likelihood may suffer from multiple local maxima, we usually suggest consider different starting points, for fixed (K_1, K_2), retain the solution providing thee best fit; in the following analyses, we will use $B = 50$ starting points. As it is typical of finite mixture models, the number of locations K_1 and K_2 is treated as fixed and known; the optimal solution for (K_1, K_2) is chosen via model selection techniques, such as AIC[27] or BIC[28].

Standard errors for model parameter estimates are obtained at convergence of the EM algorithm by the standard sandwich formula[29,30]:

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\Upsilon}}) = \mathbf{I}_o(\hat{\boldsymbol{\Upsilon}})^{-1}\widehat{\text{Cov}}(\mathbf{S})\mathbf{I}_o(\hat{\boldsymbol{\Upsilon}})^{-1},$$

where $\mathbf{I}_o(\hat{\boldsymbol{\Upsilon}})$ denotes the observed information matrix computed via the Oakes' formula[31], $\mathbf{S}$ denotes the score vector evaluated at $\hat{\boldsymbol{\Upsilon}}$, and $\widehat{\text{Cov}}(\mathbf{S}) = \sum_{i=1}^{n} \mathbf{S}_i(\hat{\boldsymbol{\Upsilon}})\mathbf{S}_i'(\hat{\boldsymbol{\Upsilon}})$ denotes the estimate for $\text{Cov}(\mathbf{S})$, with $\mathbf{S}_i$ being the individual contribution to the score vector.

6 Sensitivity analysis: definition of the index

The proposed bi-dimensional finite mixture model allows to account for potentially nonignorable dropout. However, as highlighted by Molemberghs et al.[11], for every MNAR model there is a corresponding MAR counterpart that produces exactly the same fit to the observed data; therefore, a sensitivity analysis is always appropriate. In this section, we consider the index of local sensitivity developed by Troxel et al.[12].

Let $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{K_1K_2})$ denote the vector of non ignorability parameters, with $\boldsymbol{\lambda} = \mathbf{0}$ corresponding to the MAR model. Let $\hat{\boldsymbol{\Phi}}(\boldsymbol{\lambda})$ denote the maximum likelihood estimate for model parameters in the longitudinal data model, conditional on a given value for $\boldsymbol{\lambda}$. According to Troxel et al.[12], the *Index of Sensitivity to Non-Ignorability (ISNI)* measures the displacement of model parameter estimates from their MAR counterpart, in the direction of $\boldsymbol{\lambda}$:

$$ISNI_{\boldsymbol{\Phi}} = \left.\frac{\partial\hat{\boldsymbol{\Phi}}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right|_{\boldsymbol{\Phi}(\mathbf{0})} \simeq -\left(\left.\frac{\partial^2\ell(\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\pi})}{\partial\boldsymbol{\Phi}\boldsymbol{\Phi}'}\right|_{\boldsymbol{\Phi}(\mathbf{0})}\right)^{-1}\left.\frac{\partial^2\ell(\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\pi})}{\partial\boldsymbol{\Phi}\boldsymbol{\lambda}}\right|_{\boldsymbol{\Phi}(\mathbf{0})}. \tag{12}$$

Following Xie[14], it can be shown that the following approximate expression holds:

$$\hat{\boldsymbol{\Phi}}(\boldsymbol{\lambda}) = \hat{\boldsymbol{\Phi}}(\mathbf{0}) + ISNI_{\boldsymbol{\Phi}}\boldsymbol{\lambda}, \tag{13}$$

so that the *ISNI* can also be interpreted as the linear impact that changes in $\boldsymbol{\lambda}$ have on $\hat{\boldsymbol{\Phi}}$. It is worth to highlight that, in the present case, $ISNI_{\boldsymbol{\Phi}}$ denotes a matrix rather than a vector as in the standard formulation. In particular, $ISNI_{\boldsymbol{\Phi}}$ is a $[D \times (K_1K_2 - 1)]$-dimensional matrix, where each column provides a measure of the effect that the $K_1K_2 - 1$ elements in $\boldsymbol{\lambda}$ have on the D elements in $\boldsymbol{\Phi}$. To derive a global measure of local sensitivity for $\hat{\Phi}_d$, $d = \ldots, D$, when we move far from the MAR assumption, we need consider a summary of the corresponding rows in the *ISNI* matrix, see the proposals by[14,16,32,33].

7 Analysis of the Leiden 85+ data

In this section, the proposed model is applied to the Leiden 85+ data. We aim at understanding the effects of genetic and socio-economic factors on the dynamics of cognitive functioning in the elderly, while controlling for potential bias in parameter estimates due to potentially nonignorable dropouts. In section 7.1, we provide an exploratory analysis of participants' features. Afterwards, we analyze the effect of these factors on the dynamics of the MMSE score. Results are reported in sections 7.2–7.3. Last, in section 7.4, a sensitivity analysis is performed.

7.1 Exploratory analysis

In Table 1, we summarize the individual features, in terms of covariates and MMSE scores, stratified by the observed pattern of participation. From this table, we may observe that females represent 66.73% of the whole sample and that the 64.88% of the sample has a low educational level. As expected, the most frequent APOE category is $APOE_{33}$ (58.96%), far from $APOE_{34-44}$ (21.08%) and $APOE_{22-23}$ (17.74%), while only few individuals (2.22%) are characterized by $APOE_{24}$. We may notice that more than half of the study participants (50.83%) leave the study before the scheduled end. This proportion is relatively higher for males (58.89%), lower educated participants (52.71%), and for the $APOE_{34-44}$ group (61.40%).

Table 1: Leiden 85+ Study: participation to the study by demographic and genetic characteristics of participants

Variable	Total	Completed (%)	Did not complete (%)
Gender			
Male	180 (33.3)	74 (41.1)	106 (58.9)
Female	361 (66.7)	192 (53.2)	169 (46.8)
Education			
Primary	351 (64.9)	166 (47.3)	185 (52.7)
Higher	190 (35.1)	100 (52.6)	90 (47.4)
APO-E			
22-23	96 (17.7)	54 (56.3)	42 (43.7)
24	12 (2.2)	6 (50)	6 (50)
33	319 (59.0)	162 (50.8)	157 (49.2)
34-44	114 (21.1)	44 (38.6)	70 (61.4)
Total	541 (100)	266 (49.2)	275 (50.8)

Figure 2: Leiden 85+ Study: mean of MMSE score stratified by age, and gender(a), educational level (b), APOE (c)

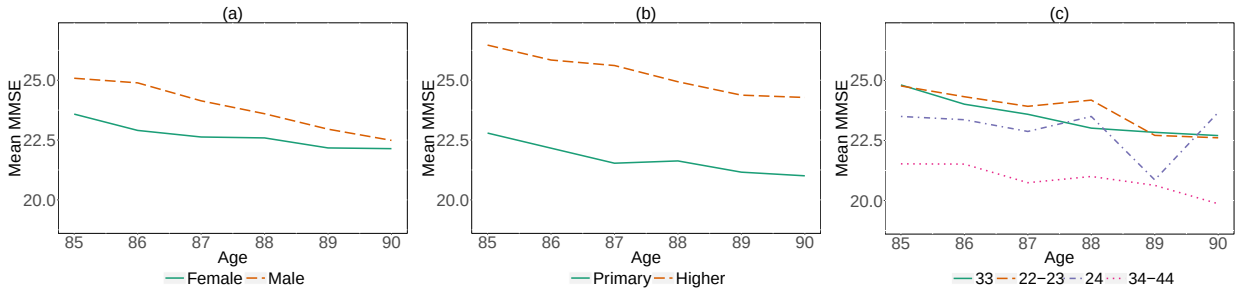


Figure 2 reports the evolution of the mean MMSE score over time, stratified by the available covariates. As it is clear, cognitive impairment is higher for females than males, even if the difference seems to decrease with age, maybe due to the differential dropout propensity by gender (Figure 2a). We may also observe that participants with a higher education are less cognitively impaired at the beginning of the study, and this difference remains persistent over the analyzed time window (Figure 2b). Rather than to a direct effect of education only, this may also reflect the impact of a differential socio-economic status. Last, low MMSE scores are observed for the $APOE_{34-44}$ group, and this is coherent with literature, suggesting that allele $\epsilon 4$ represents a risk factor for dementia [17]. The irregular pattern for $APOE_{24}$ may be due to the reduced size of this group (Figure 2c).

7.2 The MAR model

We start by estimating a MAR model, based on the assumption of independence between the longitudinal and the dropout process. In terms of eq. (6), this is obtained by assuming $\pi_{g\ell} = \pi_{g\star}\pi_{\star\ell}$, $g = 1, \ldots, K_1$, $\ell = 1, \ldots, K_2$; that is, we fit two separate models for the longitudinal response and the dropout indicator, based on a Gaussian and a Bernoulli (conditional) distribution, respectively. We considered as longitudinal response the transform $Y_{it} = \log[1 + (30 - \text{MMSE}_{it})]$ as this is nearly optimal in a Box-Cox sense; as a result, higher values identify more impaired individuals. In both models, we considered individual-specific random intercepts and the same set of covariates. In particular, this includes the variable *age* (centered round 85), the *gender* (ref - females), the *educational level* (ref - primary), and the *APOE* genotype (ref - $APOE_{33}$). As regards the random intercepts, we considered both a parametric (Gaussian) and a semi-parametric (discrete) approach. In this latter case, for each profile, the algorithm was run for varying number of locations; we retained the solution with the lowest BIC, that is with $K_1 = 5$ and $K_2 = 3$ components for the longitudinal and the dropout process, respectively. Parameter estimates are reported in Table 2.

The two (parametric and semi-parametric) approaches lead to similar conclusions on the effect of observed covariates on the longitudinal process. Cognitive impairment increases with age and is lower for males than for females. Furthermore, a strong protective effect seems to be linked to socio-economic status in early life as it may be deduced from the significant and negative effect of higher educational levels. Table 2 also highlights that $APOE_{34-44}$ represents a strong risk factor for cognitive decline. Only few differences may be highlighted when comparing the estimates. These are related to the gender effect, which is not significant in the parametric model, and the effect of higher education, which is much higher under the parametric specification. These difference may suggest that the choice of the random effect distribution may have an impact, in this case.

When the dropout process is considered, we may observe that the results are *qualitatively* the same, but for the size of parameter estimates. This is due to the different scale of the estimated random intercept distribution, with $\sigma_{b_2} = 5.393$ and $\sigma_{b_2} = 1.525$ in the semi-parametric and parametric models, respectively. In the former case, estimated intercepts are quite higher than those predicted by Gaussian quadrature locations, and this leads to inflated effects for the observed covariates as well. However, if we look at the estimated dropout probabilities resulting either

Table 2: Leiden 85+ Study: MAR models. Maximun likelihood estimates, standard errors, log-likelihood, and BIC value

Process		Semi-parametric		Parametric	
	Variable	Coeff.	Std. Err.	Coeff.	Std. Err.
Y	Intercept	1.686		1.792	0.050
	Age	0.090	0.008	0.089	0.005
	Male	-0.137	0.042	-0.085	0.066
	Higher education	-0.317	0.068	-0.623	0.065
	$APOE_{22-23}$	0.062	0.072	0.056	0.083
	$APOE_{24}$	-0.105	0.062	0.096	0.211
	$APOE_{34-44}$	0.347	0.060	0.369	0.079
	σ_y	0.402		0.398	
	σ_{b_1}	0.696		0.684	
R	Intercept	-11.475		-3.877	0.520
	Age	2.758	0.417	0.526	0.131
	Male	0.559	0.467	0.656	0.218
	Higher education	-2.162	0.772	-0.486	0.212
	$APOE_{22-23}$	0.476	0.409	-0.246	0.252
	$APOE_{24}$	-0.026	0.939	0.131	0.618
	$APOE_{34-44}$	0.805	0.461	0.565	0.237
	σ_{b_2}	5.393		1.525	
	$\log L$	-2685.32		-2732.84	
	BIC	5534.26		5572.67	

from the semi-parametric or the parametric models, these are very close to each other, but for few extreme cases which are better recovered by the semi-parametric model. This is a further suggestion that the choice for the random intercept distribution may matter[34].

7.3 The MNAR model

To avoid bias in the parameter estimates due to nonignorable dropout, we fitted the uni-dimensional and the bi-dimensional finite mixture model discussed in Sections 3 and 4. For the former approach, we run the estimation algorithm for $K = 1, \ldots, 10$; thr optimal solution according to the BIC index corresponds to a model with $K = 5$ components. Similarly, for the proposed bi-dimensional finite mixture model, we run the algorithm for $K_1 = 1, \ldots, 10$ and $K_2 = 1, \ldots, 5$ components; the optimal BIC solution corresponds to with $K_1 = 5$ and $K_2 = 3$ components for the longitudinal and the dropout process, respectively. This result is clearly coherent with the evidence from univariate

models. Parameter estimates are reported in Table 3. Looking at the longitudinal data model (left

Table 3: Leiden 85+ Study: MNAR models. Maximun likelihood estimates, standard errors, log-likelihood, and BIC value

Process		Semipar. "Uni-dim."		Semipar. "Bi-dim."	
	Variable	Coeff.	Std. Err.	Coeff.	Std. Err.
Y	Intercept	1.682		1.687	
	Age	0.094	0.007	0.094	0.007
	Male	-0.129	0.048	-0.135	0.039
	Higher education	-0.31	0.051	-0.317	0.050
	$APOE_{22-23}$	0.091	0.061	0.086	0.058
	$APOE_{24}$	-0.098	0.055	-0.099	0.056
	$APOE_{34-44}$	0.345	0.050	0.344	0.051
	σ_y	0.402		0.402	
	σ_{b_1}	0.701		0.699	
R	Intercept	-3.361		-10.767	
	Age	0.367	0.037	2.406	0.384
	Male	0.504	0.147	1.061	0.850
	Higher education	-0.200	0.151	-1.646	0.530
	$APOE_{22-23}$	-0.090	0.199	0.481	1.090
	$APOE_{24}$	-0.148	0.508	-0.334	0.647
	$APOE_{34-44}$	0.541	0.174	1.365	0.745
	σ_{b_2}	0.577		4.891	
	σ_{b_1,b_2}	0.349		0.985	
	ρ_{b_1,b_2}	0.863		0.288	
	$\log L$	-2686.902		-2660.391	
	BIC	5537.433		5534.758	

panel in the table), we may conclude that estimates are coherent with those obtained in the MAR setting. Small departures can be observed for the effect of age and gender. Males and individuals with high education tend to be less cognitively impaired when compared to the rest of the sample, while subjects carrying $\epsilon 4$ alleles (in the $APOE_{34-44}$ group) present a steeper increase in the observed response (that is, a steeper decrease in the MMSE score). Focusing on the dropout process, we may observe that age, gender and $APOE_{33-34}$ are all positively related to an increased dropout probability. By comparing the estimates obtained under the two models, the above results seem to hold besides the chosen model specification. The only remarkable difference is in the estimated magnitude of the effects for the dropout process and for the random coefficient distribution. For the bi-dimensional finite mixture model, we may observe a stronger impact of the covariates on

the dropout probability. As for the univariate model, this result is likely due to the estimated scale with an intercept value which is much lower than under the uni-dimensional specification. In this framework, the longitudinal response (conditionally Gaussian) may have a higher impact on the likelihood function when compared to the dropout indicator (conditionally Bernoulli). As a result, the estimates for the component-specific locations in the dropout equation substantially differ when compared to those estimated in the univariate model, leading to a correlation estimate equal to 0.865. On the other hand, for the bi-dimensional specification, the estimated correlation is much lower (0.288), and the estimated standard deviations for the random intercepts in the two equations are more in line with those estimated from the univariate models.

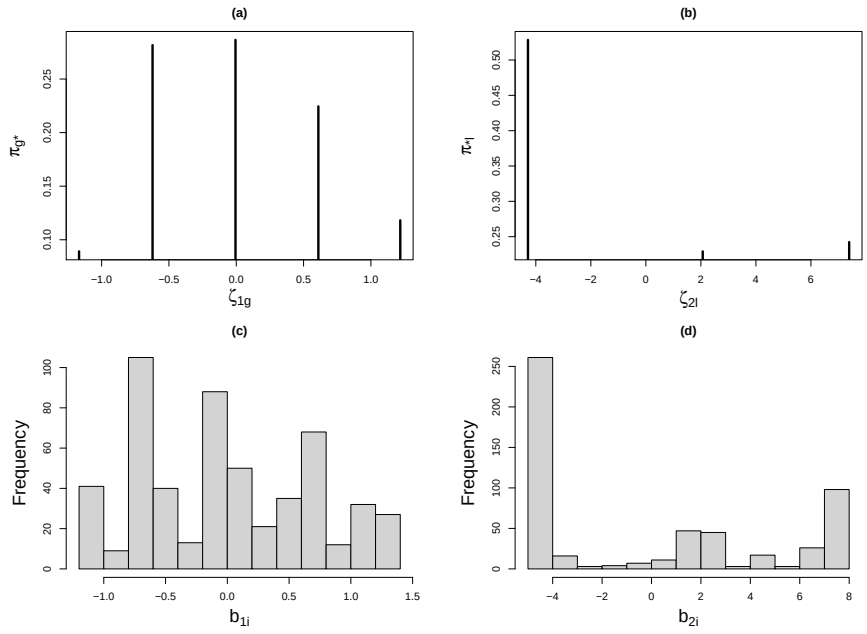
In Table 4 we report the estimated locations for the longitudinal and the dropout process, together with the corresponding conditional distribution $\pi_{\ell|g} = \Pr(b_{i2} = \zeta_{2l} \mid b_{i1} = \zeta_{1g})$. For the longitudinal data process, the estimated locations ζ_{1g} suggest that higher mixture components are associated with a higher cognitive impairment. On the other hand, tor the dropout process, higher estimated locations correspond to a higher chance to dropout from the study. When looking at the estimated conditional probabilities, we may observe a link between lower (higher) values of ζ_{1g} and lower (higher) values of $\zeta_{2\ell}$. That is, participants with better cognitive functioning (lower response values) are usually characterized by a lower probability of dropping out from the study. On the contrary, cognitively impaired participants (i.e. with higher response values) present a higher chance to dropout prematurely, even if there is still some overlapping between the second and the third component in the dropout profile.

Table 4: Maximum likelihood estimates and conditional distribution for the random parameters

	$\zeta_{2\ell}$			
ζ_{1g}	-15.053	-8.701	-3.378	
0.519	0.865	0.090	0.045	1
1.065	0.585	0.170	0.245	1
1.681	0.573	0.227	0.199	1
2.297	0.467	0.289	0.244	1
2.905	0.144	0.364	0.492	1
Tot.	0.528	0.229	0.243	1

To conclude the analysis, we report in Figure 3 the estimated prior (upper panel) and posterior distribution (lower panel) of the random intercept in the longitudinal (left panel) and the dropout

Figure 3: Estimated prior (upper panel)and posterior distribution (lower panel) of the random effects in the longitudinal (left panel) and the missingness data model (right panel)



model (right panel). By looking at this figure, we may conclude that the standard Gaussian assumption is not that appropriate for the Leiden 85+ data. This is particularly evident for the missing data process, where the estimated prior and posterior distributions are quite far from symmetry.

7.4 Sensitivity analysis: results

To investigate robustness to the assumptions on the missing data mechanism, we computed the $ISNI_{\mathbf{\Phi}}$ matrix according to formulas provided in eq. (12). For each estimate $\hat{\Phi}_d$, we derived global measures of its sensitivity to the MAR assumption by computing the norm, the minimum, and the maximum of $|ISNI_{\hat{\Phi}_d}|$. We also considered the ratio between such quantities and the standard errors of the corresponding fixed model parameters derived estimates derived from the MAR model. We also considered the ratio between such quantities and the standard errors of the corresponding parameters derived from the MAR model. Results are reported in Table 5. By looking at this table, we may observe that, as far as the fixed parameters are concerned, the indexes we computed are all close to zero but for the *age* variable. In this case, the *ISNI* takes slightly higher values, especially

for the standardized statistics. Similarly, higher values are observed for the random intercepts, even though this is an expected results, being these parameters directly connected to the missingness process.

Table 5: MAR model estimates: ISNI norm, minimum and maximum (in absolute values), and ratio to the corresponding standard error.

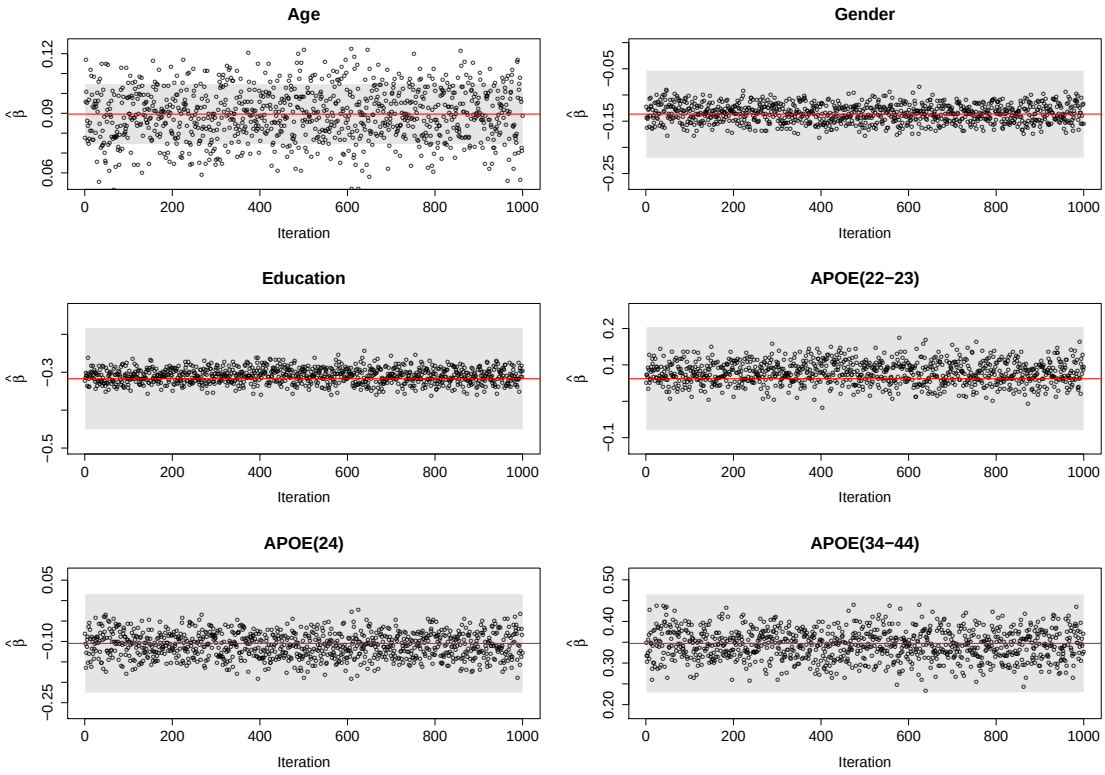
	se	$\|\|ISNI\|\|$	$\frac{\|\|ISNI\|\|}{se}$	$\min\|ISNI\|$	$\frac{\min\|ISNI\|}{se}$	$\max\|ISNI\|$	$\frac{\max\|ISNI\|}{se}$
$\boldsymbol{\zeta}_{11}$	0.117	0.0414	0.354	0.0014	0.012	0.0204	0.174
$\boldsymbol{\zeta}_{12}$	0.074	0.0580	0.784	0.0016	0.022	0.0303	0.409
$\boldsymbol{\zeta}_{13}$	0.074	0.044	0.595	0.0002	0.003	0.0255	0.345
$\boldsymbol{\zeta}_{14}$	0.083	0.1044	1.258	0.0005	0.006	0.0527	0.635
$\boldsymbol{\zeta}_{15}$	0.071	0.0088	0.124	0.0009	0.013	0.0045	0.063
Age	0.008	0.0089	1.113	0.0001	0.013	0.0054	0.675
Male	0.042	0.0058	0.138	0.0003	0.007	0.0028	0.067
Higher education	0.068	0.0075	0.110	0.0001	0.001	0.004	0.059
$APOE_{22-23}$	0.072	0.0111	0.154	0.0001	0.001	0.0074	0.103
$APOE_{24}$	0.062	0.0123	0.198	0.0005	0.008	0.0051	0.082
$APOE_{34-44}$	0.06	0.012	0.200	0.0009	0.015	0.0061	0.102
σ_y	0.194	0.1123	0.579	0.0001	0.001	0.0824	0.425

To further study the potential impact that assumptions on the dropout mechanism may have on longitudinal model parameters, we followed also a simulation-based approach. We considered the following two scenarios: (i) we simulated $B = 1,000$ values for the sensitivity parameters $\boldsymbol{\lambda}$ from a Uniform distribution, $\lambda_{g\ell}(b) \sim \mathrm{U}(-3, 3)$ for $g = 1, \ldots, K_1$ and $\ell = 1, \ldots, K_2$ ($\lambda_{K_1 K_2} = 0$); (ii) we simulated $B = 1,000$ values for a constant c from a uniform distribution, $c(b) \sim \mathrm{U}(-3, 3)$, and computed $\lambda_{g\ell}(b) = c(b)\,\hat{\lambda}_{g\ell}$, $g = 1, \ldots, K_1$, $\ell = 1, \ldots, K_2$, where $\hat{\lambda}_{g\ell}$ denotes the MLe for $\lambda_{g\ell}$ under the MNAR model. To analyze how changes in $\boldsymbol{\lambda}$ impact the estimates $\hat{\boldsymbol{\Phi}}$, we computed, for both simulated scenarios

$$\hat{\boldsymbol{\Phi}}(b) = \hat{\boldsymbol{\Phi}}(\mathbf{0}) + ISNI_{\boldsymbol{\Phi}} \times \boldsymbol{\lambda}(b). \tag{14}$$

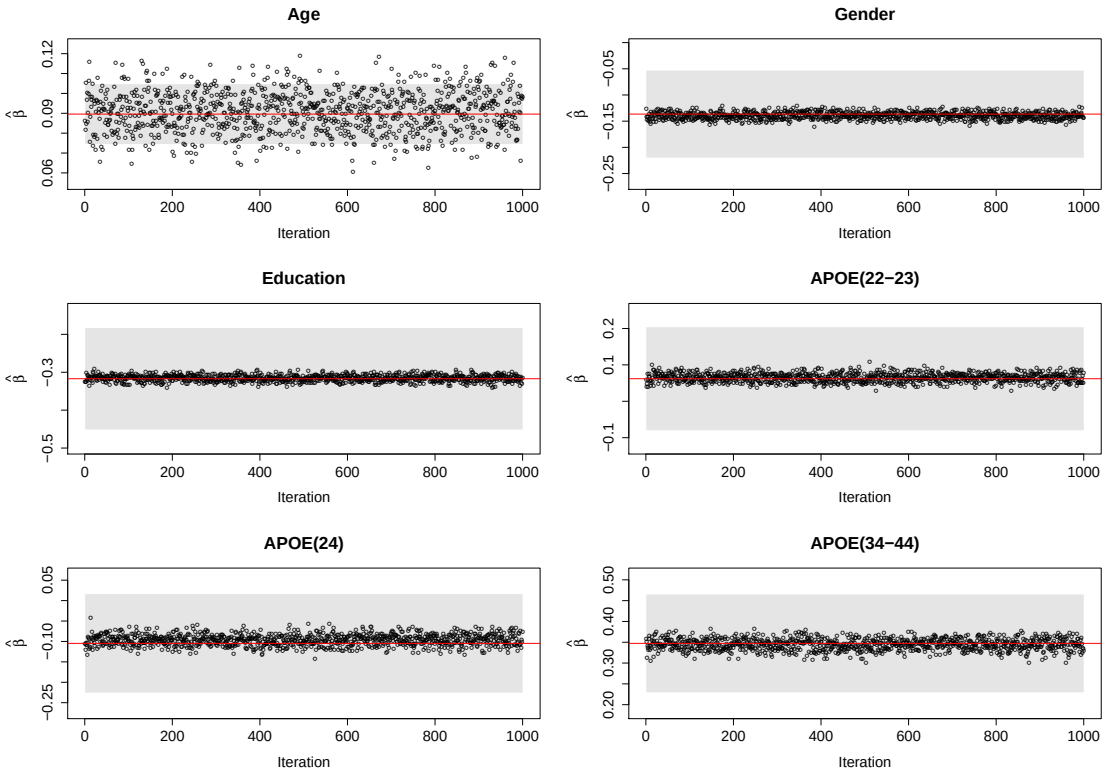
Scenario (i) allows to study general sensitivity of parameter estimates; that is, it allows to investigate how estimates vary with unstructured changes in $\boldsymbol{\lambda}$. On the other hand, under Scenario (ii) starts from the estimated pattern of dependence between the random intercepts in the longitudinal and the dropout models and allows to analyze how parameter estimates change when the correlation between the two processes increases (in absolute value) with respect to the estimated one. The proposed approach for sensitivity analysis could be seen as a particular version of local influence

Figure 4: Leiden 85+ Study: Sensitivity analysis according to Scenario (i)



diagnostics developed to check for influential observations by perturbing individual weights[35–37]. Here, we perturb weights associated to the group of subjects allocated to a given component. Obviously, a *global* influence approach could be adopted as well[38]. We report in Figures 4 and 5 the estimates obtained under Scenario (i) and (ii), respectively. The red line and the grey bands correspond to the point and the 95% interval estimates calculated under the MAR assumption for model parameters. From the former figure, we can observe that only the effect of *age* is slightly sensitive to assumptions on ignorability. Strong *local* changes in the random coefficient probability matrix may cause positive (respectively negative) changes in the *age* effect. In particular, the changes with the major impact are those in the upper left or intermediate right corners of the matrix in Table 4, where low values of both random coefficients or high values for $\zeta_{2\ell}$ and intermediate values for ζ_{ig} are stored. Overall, the relative frequency of points within the corresponding MAR confidence interval is equal to 0.737. This suggests that, even if some sensitivity to assumptions about ignorability of the dropout process is present, estimates always remain within a reasonable set.

Figure 5: Leiden 85+ Study: Sensitivity analysis according to Scenario (*ii*)



When looking at the results obtained under Scenario (*ii*) (see Figure 5), we may observe that changes in the parameter estimates are more clearly linked to correlation between the random effects in the two profiles. As before, some sensitivity to departures from the MAR assumption is observed for the *age* effect only. The relative frequency of points within the corresponding MAR confidence interval is equal to 0.851, which suggests a lower sensitivity to assumptions about ignorability of the dropout process, when compared to Scenario (*i*). In this case, high positive/negative correlation between the random intercepts leads to estimates that are higher/lower than the corresponding MAR counterparts.

8 Conclusions

In this paper, we define a random coefficient based dropout model where the association between the longitudinal and the dropout process is modelled through discrete, outcome-specific, latent effects. A bi-dimensional representation for the random coefficient distribution is used, which

allows for a (possibly) different number of locations in each model. A full probability matrix connecting the locations in a model to those in the other one is considered. The main advantage of this flexible representation is that the resulting MNAR model properly nests a model where the dropout mechanism is ignorable. This allows us to consider a (local) sensitivity analysis, based on the ISNI index, to check changes in model parameter estimates as we move far from the MAR assumption. The data application show good robustness of all model parameter estimates. A slight sensitivity to assumptions on the dropout mechanism is observed for the *age* effect which, however, is always restricted to a reasonable set, and may be reasonably due to the differential participation of more cognitively impaired individuals.

Acknowledgement

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

References

[1] Rubin D. Inference and missing data *Biometrika.* 1975;63:581–592.

[2] Little RJA. Modeling the drop-out mechanism in repeated-measures studies *Journal of the American Statistical Association.* 1995;90:1112–1121.

[3] Verzilli CJ, Carpenter JR. A Monte Carlo EM algorithm for random-coefficient-based dropout models *Journal of Applied Statistics.* 2002;29:1011–1021.

[4] Gao S. A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data *Statistics in Medicine.* 2004;23:211–219.

[5] Scharfstein D, Rotnitzky A, Robins J. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models *Journal of the American Statistical Association.* 1999;94:1096–1120.

[6] Alfò M, Aitkin M. Random coefficient models for binary longitudinal responses with attrition *Statistics and Computing.* 2000;10:279–287.

[7] Alfò M, Maruotti A. A selection model for longitudinal binary responses subject to non-ignorable attrition *Statistics in Medicine.* 2009;28:2435–2450.

[8] Tsonaka R, Verbeke G, Lesaffre E. A Semi-Parametric Shared Parameter Model to Handle Nonmonotone Nonignorable Missingness *Biometrics.* 2009;65:81–87.

[9] Beunckens C, Molenberghs G, Verbeke G, Mallinckrodt C. A Latent-Class Mixture Model for Incomplete Longitudinal Gaussian Data *Biometrics.* 2008;64:96–105.

[10] Bartolucci F, Farcomeni A. A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates *Biometrics.* 2015;71:80–89.

[11] Molenberghs G, Beunckens C, Sotto C, Kenward MG. Every missing not at random model has got a missing at random counterpart with equal fit *Journal of the Royal Statistical Society, Series B.* 2008;70:371–388.

[12] Troxel AB, Ma G, Heitjan DF. An index of local sensitivity to non-ignorability *Statistica Sinica.* 2004;14:1221–1237.

[13] Ma G, Troxel AB, Heitjan DF. An index of local sensitivity to non-ignorability in longitudinal modeling *Statistics in Medicine.* 2005;24:2129–2150.

[14] Xie H. A local sensitivity analysis approach to longitudinal non-Gaussian data with non-ignorable dropout *Statistics in Medicine.* 2008;27:3155–3177.

[15] Viviani S, Rizopoulos D, Alfò M. Local sensitivity to non-ignorability in joint models *Statistical Modelling.* 2014;14:205–228.

[16] Gao W, Hedeker D, Mermelstein R, Xie H. A scalable approach to measuring the impact of nonignorable nonresponse with an EMA application *Statistics in Medicine.* 2016;35:5579-5602.

[17] Hauser P.S., Ryan R.O.. Impact of apolipoprotein E on Alzheimer's disease *Current Alzheimer Research.* 2013;10:809–817.

[18] Folstein M, Folstein S, McHig P. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician *Journal of Psychiatry Research.* 1975;12:189–198.

[19] Wu M, Carroll R. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process *Biometrics.* 1988;44:175–188.

[20] Wu M, Bailey K. Estimation and comparison of changes in the presence of informative right censoring: conditional linear models *Biometrics.* 1989;45:939–955.

[21] Creemers A, Hens N, Aerts M, Molenberghs G, Verbeke G, Kenward M. Generalized shared-parameter models and missingness at random *Statistical Modelling.* 2011;11:279–310.

[22] Rizopoulos D, Verbeke G, Lesaffre E, Vanrenterghem Y. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros *Biometrics.* 2008;64:611–619.

[23] Alfò M, Rocchetti I. A flexible approach to finite mixture regression models for multivariate mixed responses *Statistics and Probability Letters.* 2013;83:1754–1758.

[24] Lagona F. *Model-Based Classification of Clustered Binary Data with Non-ignorable Missing Values*:155–165. Berlin, Heidelberg: Springer Berlin Heidelberg 2013.

[25] Dempster A.P., Laird N.M., Rubin D.B.. Maximum Likelihood from Incomplete Data via the EM Algorithm *Journal of the Royal Statistical Society, series B.* 1977;39:1–38.

[26] Karlis D. A cautionary note on the EM algorithm for finite exponential mixtures Tech. Rep. 150Department of Statistics, Athens University of Economics and Business 2001.

[27] Akaike H. Information theory and an extension of the maximum likelihood principle in *Second International Symposium on Information Theory* (Petrov B. N., Csaki F.. , eds.):267–281Akadémiai Kiado 1973.

[28] Schwarz G. Estimating the dimension of a model *The Annals of Statistics.* 1978;6:461–464.

[29] White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity *Econometrica: Journal of the Econometric Society.* 1980:817–838.

[30] Royall RM. Model robust confidence intervals using maximum likelihood estimators *International Statistical Review.* 1986;54:221–226.

[31] Oakes D. Direct calculation of the information matrix via the EM algorithm *Journal of the Royal Statistical Society, Series B.* 1999;61:479–482.

[32] Xie H, Heitjan DF. Sensitivity analysis of causal inference in a clinical trial subject to crossover *Clinical Trials.* 2004;1:21–30.

[33] Xie H. Analyzing longitudinal clinical trial data with nonignorable missingness and unknown missingness reasons *Computational Statistics and Data Analysis.* 2012;56:1287–1300.

[34] Neuhaus JM, McCulloch CE. Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter *Statistical Science.* 2011;26:388–402.

[35] Jansen I, Molenberghs G, Aerts M, Thijs H, Van Steen K. A local influence approach applied to binary data from a psychiatric study *Biometrics.* 2003;59:410–419.

[36] Rakhmawati TW, Molenberghs G, Verbeke G, Faes C. Local influence diagnostics for hierarchical count data models with overdispersion and excess zeros *Biometrical journal.* 2016;58:1390–1408.

[37] Rakhmawati TW, Molenberghs G, Verbeke G, Faes C. Local influence diagnostics for generalized linear mixed models with overdispersion *Journal of Applied Statistics.* 2017;44:620–641.

[38] White I, Carpenter J, Horton NJ. A mean score method for sensitivity analysis to departures from the missing at random assumption in randomised trials *Statistica Sinica.* 2017;to appear.