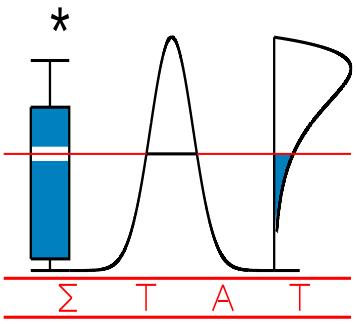
T E C H N I C A L

R E P O R T

0665

SENSITIVITY ANALYSIS IN PATTERN MIXTURE MODELS USING THE EXTRAPOLATION METHOD

RIZOPOULOS, D., VERBEKE, G. and E. LESAFFRE



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Sensitivity Analysis in Pattern Mixture Models using the Extrapolation Method

Dimitris Rizopoulos,[1,*] **Geert Verbeke**[1] **and Emmanuel Lesaffre**[1]

[1]Biostatistical Centre, Catholic University of Leuven,

U.Z. St. Rafaël, Kapucijnenvoer 35, B–3000 Leuven, Belgium

January 10, 2007

Summary. Pattern mixture models are frequently used to analyze longitudinal data where missingness is induced by dropout. When interest lies in the marginal model, inferences are mainly based on assumptions, in which case the two most common model fitting techniques are, the imposition of identifying restrictions and the extrapolation method. Here we consider the extrapolation method under a Bayesian model formulation and propose a suitable prior distribution for the model parameters, which provides a flexible tool for sensitivity analysis. In particular, a parameter that controls the difference between average trends among neighbouring patterns is introduced and sensitivity is assessed by comparing results obtained from analyses with different values for this parameter.

Key words: Pattern Mixture Models; Extrapolation Method; Sensitivity Analysis; Bayesian Approach.

1. Introduction

Many longitudinal studies suffer from attrition, which causes biased inferences if the dropout mechanism is ignored while being informative. To account for informative dropout, a number of model-based approaches have been proposed for the joint modelling of the longitudinal and dropout processes (see Little, 1995; Hogan and Laird, 1997b; Kenward and Molenberghs, 1999 for recent reviews). These approaches can be broadly classified as either *selection models* or *pattern mixture models* (Little, 1995; Little and Rubin, 2002). In both model frameworks inference is based on assumptions and thus

* *email:* dimitris.rizopoulos@med.kuleuven.be

tools for detecting sensitivity to these assumptions are required. Here we propose a new method to assess the sensitivity of the inference under various assumptions in pattern mixture models.

The rationale of the pattern mixture model is to postulate a separate model for the longitudinal measurement process per dropout pattern. For marginal inference two implications arise, namely the conditional distribution for the complete longitudinal vector and a model for the dropout process need to be specified. By definition, except from the pattern of completers, the conditional distribution of the unobserved given the observed components of the measurement process is not identified. The two most common procedures to overcome this underidentification problem, are the imposition of identifying restrictions (Little, 1993; Little, 1994) and the extrapolation method (Verbeke and Molenberghs, 2000). The identifying restrictions constitute a general approach where the conditional distribution of the missing given the observed components of the measurement process, in each pattern, is identified using information from the other patterns. Several sets of restrictions have been proposed including, among others, the available case, complete case and neighbouring case missing value restrictions (Thijs et al., 2002). In this setting, sensitivity can be assessed by checking how much inference changes under a range of plausible missing value restrictions. On the other hand, the extrapolation method is typically applied when we consider simplified models, i.e., when a small number of parameters is used to define a joint distribution for the repeated measurements, and the main interest lies in the marginal average evolutions. In this case the average profiles for the missing components of the measurement process per pattern are obtained by extrapolating the corresponding observed fitted longitudinal profiles. Despite the easy implementation of the extrapolation method, sensitivity analysis is not straightforward. This is one of the reasons why the method has not been used much and why the main focus has been on the identifying restrictions approach.

Here we work under a linear mixed-effects measurement model and develop a Bayesian formulation of the extrapolation method, which under a suitable specification of the prior distributions for the model parameters, provides a flexible tool for sensitivity analysis. In particular, a set of informative priors for the fixed-effects parameters is introduced, which through a sensitivity parameter controls the average profiles in each dropout pattern. Our

suggestion resembles the usage of subjective parameters, introduced by Rubin (1977) to account for nonrespondents in sample surveys. Specifically, the formulation we follow expresses the subjective notion that the average profiles (controlled by the fixed-effects parameters) of neighbouring patterns are more alike than the average profiles of more distant patterns. We further extend this notion and let the sensitivity parameter to depend on the time lag between the dropout times of the neighbouring patterns.

Our research is motivated by a randomized, double-blind parallel group, multi-center study for the comparison of two oral treatments (terbinafine 250mg/day versus itraconazole 200mg/day) for toenail dermatophyte onychomycosis (TDO). The study is described in full by De Backer et al. (1996). TDO is a common toenail infection, difficult to treat, affecting more than 2 out of 100 persons. Antifungal compounds, classically used for the treatment of TDO, need to be taken until the whole nail has grown out healthy. The development of new such compounds, however, has reduced the treatment duration to 3 months. The aim of the present study was to compare the efficacy and safety of 12 weeks of continuous therapy with either terbinafine or itraconazole. In total, 2×189 patients were randomized and distributed in 36 centers. Subjects were followed up during 12 weeks of treatment and followed further, up to a total of 48 weeks. Measurements were taken at baseline, every month during treatment, and every 3 months afterwards, resulting in a maximum of seven measurements per subject. A subset of these data is used here that contains 150 patients taking terbinafine and 148 taking itraconazole. Due to a variety of reasons, 72 (24.2%) out of the 298 patients left the study prematurely, resulting in 7 dropout patterns. Out of the 7 patterns, 3 are very sparse, containing 7 (2.4%), 7 (2.4%) and 3 (1%) subjects, respectively. Table 1 summarizes the number of subjects still in the study at each occasion, for both treatment groups separately as well as the number of subjects per dropout pattern.

[Table 1 about here.]

Further, as an informal exploratory graphical tool, Figure 1 depicts the sample average profiles at each occasion versus time, for each dropout pattern. We can observe that the sparse patterns (especially patterns 2 and 7) present different average evolutions compared with the patterns with more individuals.

[Figure 1 about here.]

Previous analyses of the same data (Verbeke et al., 2001; Verbeke and Molenberghs, 2000) showed a significant nonignorable dropout mechanism under a selection model. In addition, the use of the extrapolation method in the pattern mixture framework revealed an interesting feature: although the sparse patterns have a very small contribution to the marginal means, the shape of their unreliable (due to sparseness) estimated average evolutions significantly alter the shape of the marginal average evolutions. In this paper and according to the proposed method we use a pattern mixture model framework and allow the average profiles of the sparse patterns to be affected by the average profiles of more frequent but neighbouring patterns, which are expected to show similar average evolutions. A detailed analysis of the Toenail study can be found in Section 5.

The paper is organized as follows: Section 2 briefly reviews the pattern mixture models and their features, and also provides the intuitive justification for the method we propose. Section 3 presents the theoretical framework under which our method is implemented and in Section 4 we discuss the interpretation of the sensitivity parameter and moreover we form a connection with the identifying restrictions approach. Finally, Section 5 illustrates the performance of our method in the Toenail study and in Section 6 we conclude and refer to some extensions.

2. Pattern Mixture Models and the Extrapolation Method

Let $\boldsymbol{Y}$ and R denote the measurement and dropout process, respectively. For an incomplete sequence, R denotes the occasion at which dropout occurs, while for a complete sequence, R equals to the number of the planned measurements plus one. According to R the measurement process is decomposed into $\boldsymbol{Y} = (\boldsymbol{Y}^o, \boldsymbol{Y}^m)$ with $\boldsymbol{Y}^o$ denoting the observed and $\boldsymbol{Y}^m$ the missing response vectors. Under the pattern mixture framework the joint distribution of $\boldsymbol{Y}$ and R is factorized as (covariates are allowed but omitted in the notation)

$$p\left(\boldsymbol{Y}, R\right) = p\left(\boldsymbol{Y}|R\right) p\left(R\right), \tag{1}$$

where $p(\cdot)$ denotes a probability density function. This factorization expresses that subjects with the same dropout times are more alike than subjects with different dropout times. This in turn states that subjects with the same dropout times share the same measurement process and thus a model is fitted for each dropout pattern separately.

Usually in longitudinal studies the main focus of inference is on the marginal average evolutions, i.e., $E[\boldsymbol{Y}]$, while the correlation between the repeated measurements is regarded as nuisance. Under factorization (1) these evolutions take the form

$$\begin{aligned} E[\boldsymbol{Y}] &= \int \boldsymbol{y} \left\{ \int p(\boldsymbol{y}|r)p(r)dr \right\} d\boldsymbol{y} \\ &= \int \left\{ \int \boldsymbol{y} p(\boldsymbol{y} \mid r) d\boldsymbol{y} \right\} p(r)dr, \end{aligned} \tag{2}$$

that is, $E[\boldsymbol{Y}]$ is the expectation of the pattern-specific average evolutions $E[\boldsymbol{Y}|R]$ with respect to $p(R)$. It can be seen from (2) that this marginal expectation requires specification of $E[\boldsymbol{Y}^m \mid \boldsymbol{Y}^o, R]$, i.e., of the expected evolution of the subjects after they dropped out. However, it should be emphasized that the observed data do not contain any information on these average profiles beyond the times of dropout. Hence, assumptions are needed. As noted earlier, the two main strategies to deal with this underidentification problem are, to impose identifying restrictions or to extrapolate beyond the last obtained measurement. Although the first strategy makes explicit assumptions about the unidentified distribution $\boldsymbol{Y}^m|\boldsymbol{Y}^o$ in each dropout pattern, the simplicity of the extrapolation method is apparent, especially when simplified models are considered and the main interest lies only in the specification of $E[\boldsymbol{Y}]$. By simplified models we imply mixed-effects models with simple functional forms to capture the mean longitudinal profiles (e.g., linear or quadratic time trends) and a random-effects structure to model the associations between individual responses.

In the extrapolation method, in accordance with the pattern mixture methodology, we consider a different mean, and possibly covariance, structure per dropout pattern. Then, a full specification of $E[\boldsymbol{Y}]$ is derived by extrapolating the observed fitted profiles $E[\boldsymbol{Y}^o|R]$ to obtain $E[\boldsymbol{Y}^m|R]$. As also Hogan and Laird (1997a) note, the main disadvantage of this simplified approach is that it makes the awkward requirement that each dropout pattern occurs sufficiently often in order to estimate the pattern-specific parameters reliably. Otherwise estimates from sparse patterns could severely affect the marginal average evolutions. The method we propose to overcome this problem and in addition to avoid the restriction of having a separate measurement distribution for each dropout pattern, is to estimate the average evolutions in each pattern by borrowing information from the neighbouring patterns. This can easily be implemented under a Bayesian ap-

proach using properly chosen informative prior distributions. In fact this method can be thought of as an extension of the intuitive idea driving the pattern mixture factorization, i.e., while in ordinary pattern mixture models individuals with the same dropout times share the same average evolutions, here we move to the pattern level and allow neighbouring patterns to have more similar average evolutions than more distant patterns (i.e., with larger time lags in their dropout times). Specifically we introduce a set of sensitivity parameters which tune the degree of resemblance of the average profiles in neighbouring patterns. Thereby we can check the sensitivity of inferences under a range of values for these parameters. The central feature of our proposal is that it explicitly acknowledges the underidentification problem by using extra information from patterns that identify $\boldsymbol{Y}^m|\boldsymbol{Y}^o$, which is in line with the identifying restrictions approach as it is also seen in Section 4.3. In the next section we describe how our proposal can be implemented under the Bayesian approach.

3. Likelihood, Prior Specification and Posterior Conditionals

3.1 *Likelihood*

Let $k = 1, \ldots, K$ denote the dropout pattern, with $k = 1$ corresponding to the completers and consider a random sample of n ($i = 1, \ldots, n$) individuals with n_i available measurements. For the measurement process in each pattern we assume a linear mixed-effects model with pattern-specific fixed-effects

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_k + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{3}$$

$$\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{ik}) \text{ and } \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}_k),$$

where $\boldsymbol{y}_i$ is the $n_i \times 1$ response vector for the ith individual, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are known $n_i \times p$ and $n_i \times q$ design matrices for the fixed- and random-effects, respectively, $\boldsymbol{\beta}_k$ denotes a $p \times 1$ fixed-effects vector for the kth dropout pattern, $\boldsymbol{b}_i$ is the $q \times 1$ random-effects vector with corresponding covariance matrix $\boldsymbol{D}_k$, $\boldsymbol{\varepsilon}_i$ are the random errors with covariance matrix $\boldsymbol{\Sigma}_{ik}$, which depends on i only through its dimensions n_i, and $\boldsymbol{\varepsilon}_i$ is independent of $\boldsymbol{b}_i$. Here we consider only pattern-specific average profiles, whereas the covariance structures $\boldsymbol{D}_k$ and $\boldsymbol{\Sigma}_{ik}$ are assumed to be the same across patterns, i.e., $\boldsymbol{D}_k = \boldsymbol{D}$ and $\boldsymbol{\Sigma}_{ik} = \boldsymbol{\Sigma}_i$, $\forall k$, respectively.

For the missingness process we assume that the dropout indicator $r_i \in \{1, 2, \ldots, K\}$ follows a Multinomial distribution. In particular, the marginal dropout model $p(r_i; \boldsymbol{\pi})$ is of the form

$$\log \frac{\pi_{ik}}{\pi_{iK}} = \alpha_k + \boldsymbol{w}_i^t \boldsymbol{\delta}_k, \quad k = 1, \ldots, K-1, \tag{4}$$

where $\pi_{ik} = Pr(r_i = k)$ denotes the individual's probability for pattern k, $\pi_{iK} = 1 - \sum_{k=1}^{K-1} \pi_{ik}$, α_k are the cut-points, and $\boldsymbol{\delta}_k$ is a vector of regression coefficients for the covariate vector $\boldsymbol{w}_i$. This formulation enables the dropout probabilities to depend on baseline covariates and treatment. Under the Multinomial model, the marginal average evolutions defined in (2) take the form of a weighted sum, i.e., $E[\boldsymbol{Y}_i] = \sum_{k=1}^{K} \pi_{ik} E[\boldsymbol{Y}_i | R_i = k]$.

Under (1), (3) and (4), and suppressing condition on covariates for notational simplicity, the conditional on the random-effects likelihood of the model has the form

$$p\left(\boldsymbol{y}, \boldsymbol{r} | \boldsymbol{b}; \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\pi}\right) = \left\{ \prod_{i=1}^{n} p(r_i; \boldsymbol{\pi}) \right\} \times$$

$$\left\{ \prod_{i \in P_1} p\left(\boldsymbol{y}_i | \boldsymbol{b}_i; \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_i\right) \cdot \ldots \cdot \prod_{i \in P_K} p\left(\boldsymbol{y}_i | \boldsymbol{b}_i; \boldsymbol{\beta}_K, \boldsymbol{\Sigma}_i\right) \right\}, \tag{5}$$

where P_k denotes the set of individuals available in the kth dropout pattern.

3.2 *Prior Specification*

For a Bayesian approach, specification of prior distributions for all the model parameters is required. Proposals for prior distributions for the variance components under model (3) and for the probabilities in the multinomial model $p(r_i; \boldsymbol{\pi})$ have been extensively studied (see e.g., at Gelman, 2006; Gilks et al., 1993; Gelman et al., 1995) and thus we will not further discuss them here. In particular, we use the inverse Wishart distribution for $\boldsymbol{\Sigma}_i$ and $\boldsymbol{D}$ (or inverse Gamma or half-t in the case of scalar variance components) and the Dirichlet distribution for $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^t$. Our main focus now is on specifying a meaningful prior for the fixed-effects $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ which will easily allow sensitivity analysis.

As we have already noted, the pattern mixture factorization (1) postulates that subjects in the same dropout pattern share the same measurement distribution. We extend this belief to the dropout patterns level. In particular, it is assumed that the fixed-effects parameters between the patterns are a priori correlated, and the degree of this correlation is controlled through a set of parameters. This can be formally expressed under the

following set of prior conditional distributions

$$\begin{aligned} \boldsymbol{\beta}_1 &\sim N(\mathbf{0}, \boldsymbol{V}) \\ \boldsymbol{\beta}_2|\boldsymbol{\beta}_1 &\sim N(\boldsymbol{\beta}_1, \gamma_1\boldsymbol{\Delta}_2) \\ \boldsymbol{\beta}_3|\boldsymbol{\beta}_2 &\sim N(\boldsymbol{\beta}_2, \gamma_2\boldsymbol{\Delta}_3) \\ &\vdots \\ \boldsymbol{\beta}_K|\boldsymbol{\beta}_{K-1} &\sim N(\boldsymbol{\beta}_{K-1}, \gamma_{K-1}\boldsymbol{\Delta}_K), \end{aligned} \tag{6}$$

where, we set $\boldsymbol{\beta}_1$ the fixed-effects vector of the completers, $\boldsymbol{\beta}_2$ the fixed-effects vector for subjects with the last measurement missing, and so on. The covariance matrix for the pattern of completers is taken to be $\boldsymbol{V} = 10^6\mathbf{I}_{n_i}$, whereas the $\boldsymbol{\Delta}_k$'s are assumed to be diagonal matrices quantifying prior information about the dispersion of the fixed-effects for the rest of the patterns, with $\gamma_k > 0$ being a sensitivity parameter.

The form of these prior conditionals expresses that, a priori, the fixed-effects parameters of the dropout pattern with $N - s$ available measurements are normally distributed around the fixed-effects parameters of the dropout pattern with $N - s + 1$ available measurements. The $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{K-1})^t$ parameters control the variability of this distribution, indicating how different the fixed-effects parameters of neighbouring patterns are allowed to be. In order each γ_{k-1} to have the same effect on all $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})^t$ of the kth pattern, we use the $\boldsymbol{\Delta}_k$ matrices that provide an indicative estimate for the variability of $\boldsymbol{\beta}_k$. Thus, $\sqrt{\gamma_{k-1}}$ in fact expresses the number of the standard deviations $\sqrt{\boldsymbol{\Delta}_{k,jj}}$, β_{kj} is assumed to vary around $\beta_{j,k-1}$, $\forall j \in \{1, \ldots, p\}$, $k = 2, \ldots, K$. More details regarding choices for the sensitivity parameters and $\boldsymbol{\Delta}_k$'s are given in Section 4.

An extra assumption that we make is that the joint prior for $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$ satisfies a first order Markov property, namely

$$\begin{aligned} p(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K) &= p(\boldsymbol{\beta}_1) \cdot p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) \cdot p(\boldsymbol{\beta}_3|\boldsymbol{\beta}_2, \boldsymbol{\beta}_1) \cdot \ldots \cdot p(\boldsymbol{\beta}_K|\boldsymbol{\beta}_{K-1}, \ldots, \boldsymbol{\beta}_1) \\ &= p(\boldsymbol{\beta}_1) \cdot p(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) \cdot p(\boldsymbol{\beta}_3|\boldsymbol{\beta}_2) \cdot \ldots \cdot p(\boldsymbol{\beta}_K|\boldsymbol{\beta}_{K-1}). \end{aligned} \tag{7}$$

Although this restriction is not strictly required in order to fit the model (i.e., we can construct a joint distribution for $\boldsymbol{\beta}$ for which (6) holds without requiring (7)), it leads to simpler posterior conditionals that enhance interpretability. The required joint distribution for $\boldsymbol{\beta}$, under assumption (7), can be obtained by applying Doob's theorem (Doob, 1942) in a Gauss-Markov process.

3.3 *Full Posterior Conditionals*

Taking advantage of the conjugacy between the likelihood and the priors, the full posterior conditionals of the fixed-effects parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ take the form

$$\boldsymbol{\beta}_k|. \sim N\left(\mathcal{V}_k \mathcal{M}_k, \mathcal{V}_k\right), \quad k = 1, \ldots, K, \tag{8}$$

where $\boldsymbol{\beta}_k|.$ denotes the conditional distribution of the parameter vector $\boldsymbol{\beta}_k$ given the observed data and all the other parameters, including the random-effects, and

$$\begin{aligned}
\mathcal{M}_1 &= \left[\sum_{i \in P_1} \mathcal{A}_i + \boldsymbol{\Delta}_2^{-1}\boldsymbol{\beta}_2/\gamma_1\right] \text{ and } \mathcal{V}_1 = \left[\sum_{i \in P_1} \mathcal{B}_i + \boldsymbol{V}^{-1} + \frac{1}{\gamma_1}\boldsymbol{\Delta}_2^{-1}\right]^{-1} \\
\mathcal{M}_2 &= \left[\sum_{i \in P_2} \mathcal{A}_i + \boldsymbol{\Delta}_2^{-1}\frac{\boldsymbol{\beta}_1}{\gamma_1} + \boldsymbol{\Delta}_3^{-1}\frac{\boldsymbol{\beta}_3}{\gamma_2}\right] \text{ and } \mathcal{V}_2 = \left[\sum_{i \in P_2} \mathcal{B}_i + \frac{1}{\gamma_1}\boldsymbol{\Delta}_2^{-1} + \frac{1}{\gamma_2}\boldsymbol{\Delta}_3^{-1}\right]^{-1} \\
&\vdots \\
\mathcal{M}_K &= \left[\sum_{i \in P_K} \mathcal{A}_i + \boldsymbol{\Delta}_K^{-1}\boldsymbol{\beta}_{K-1}/\gamma_{K-1}\right] \text{ and } \mathcal{V}_K = \left[\sum_{i \in P_K} \mathcal{B}_i + \frac{1}{\gamma_{K-1}}\boldsymbol{\Delta}_K^{-1}\right]^{-1},
\end{aligned}$$

where $\mathcal{A}_i = \boldsymbol{X}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{Z}_i \boldsymbol{b}_i)$, $\mathcal{B}_i = \boldsymbol{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i$, and $\boldsymbol{X}_i^t$ denotes the transpose of $\boldsymbol{X}_i$. Further, the conjugate form of the priors chosen for $\boldsymbol{\Sigma}_i$ and $\boldsymbol{D}$ leads to known posterior conditionals and thus to a straightforward implementation of the Gibbs sampler (see Gilks et al., 1993; Carlin, 1996).

Examining the form of the above posterior conditionals, we observe that the priors (6) also protect against overfitting in sparse patterns. In particular, in patterns with few individuals both terms $\sum_{i \in P_k} \boldsymbol{X}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{Z}_i \boldsymbol{b}_i)$ and $\sum_{i \in P_k} \boldsymbol{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i$ are downplayed, while the effect of the fixed-effects of the neighbouring patterns becomes of more importance. Another approach to take into account sparse patterns in the pattern mixture framework has been proposed by Roy (2003), using a latent class indicator.

4. Features of the Proposed Method

4.1 *Interpretation of the Sensitivity Parameters*

Our proposal to associate the $\boldsymbol{\beta}_k$'s through the $\boldsymbol{\gamma}$ parameters is in the spirit of the subjective parameters proposed by Rubin (1977). Rubin considered the univariate case using a simple linear model and postulated a prior conditional distribution for the parameters of the nonrespondents, with mean equal to the parameters of the respondents

and with variance controlled by what he called subjective parameters. This formulation provided a prior intuitive range for the plausible values of the parameter of the nonrespondents. Analogously, in our case the vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{K-1})^t$ of sensitivity parameters expresses a subjective relationship between the fixed-effects parameter vectors of the different dropout patterns, which implicitly also controls the fixed-effects parameter vector for the nonrespondents. Based on formulation (6) and under an approximate Normal or Student's-t distribution for the posterior of the fixed-effects, a plausible range for $\sqrt{\gamma_{k-1}}$ is from zero to four or five, since it reflects the number of $\sqrt{\Delta_{k,jj}}$ standard errors around β_{kj}. As a practical guideline a sensitivity analysis for increasing $\sqrt{\gamma_{k-1}}$ by steps of one is suggested; in case of very sensitive results this could be refined by steps of 0.5.

The whole methodology we have developed so far considers a different γ parameter for each dropout pattern, except from the pattern of completers. Unfortunately, this impedes practicality since a simultaneous consideration of different values for all the γ parameters is required. A functional form of these parameters could be used instead, to simplify matters. In particular, in order to take into account the time lag between the dropout times we use here

$$\gamma_k = \gamma(t_{k+1} - t_k), \tag{9}$$

where $\gamma \in (0, \infty)$ and $t_{k+1} - t_k$ denotes the time lag in the dropout times between pattern $k+1$ and pattern k. Other functional forms could be used as well.

Moreover, under (9) the γ parameter can also be used to partially determine the dropout mechanism, under Rubin's taxonomy (Rubin, 1976; Little and Rubin, 2002). In fact, depending on the choice of the value of γ, we can see from (6) that we move from a *Missing Completely At Random* (MCAR) to a *Missing Not At Random* (MNAR) direction. More specifically, for $\gamma \to 0$ we force all the dropout patterns to share the same fixed-effects parameters which, combined with the fact that we assume the same variance components across patterns, implies MCAR. On the other hand for $\gamma \to \infty$ we let each dropout pattern to have its own fixed-effects parameter vector. However, we note that there is no value of γ leading to a *Missing At Random* (MAR) dropout mechanism. Even though this fact could be regarded as a potential drawback of the proposed method, the sensitivity analysis can be easily augmented by fitting the linear mixed model to the available data which provides a valid fit under MAR. Further, by setting $\gamma \approx 0$

we explicitly restrict the model to MCAR, and consequently discrepancies between the fitted average evolutions and the sample average evolutions for $\gamma \to 0$ can be regarded as evidence against MCAR. However, since both MAR and MNAR are assumptions that cannot be tested based on the data at hand, we should note that the observed data containing only information to reject MCAR. Thus, focus is given on sensitivity of the results under various choices for γ and not in estimating it.

4.2 *Choosing the Dispersion Matrices* $\boldsymbol{\Delta}_k$

The dispersion matrices $\boldsymbol{\Delta}_k$, introduced in (6), are closely related to the γ parameters. In particular, they provide a variability basis such that the γ's have a unified effect on all the fixed-effects parameters. Thus, their choice has a direct effect on γ's interpretation, as it has been described in Section 3.2. In the absence of any prior information, a practical method to obtain $\boldsymbol{\Delta}_k$'s is to fit the linear mixed-effects model in each pattern, under maximum likelihood, and use the diagonal elements of the inverse information matrix for $\hat{\boldsymbol{\beta}}_k$'s.

4.3 *Connection with the Identifying Restrictions Approach*

As we discussed in Section 2, a common criticism of the extrapolation method is that the assumptions about the dropout process in terms of MCAR, MAR and MNAR are not immediately clear. Thus, identifying restrictions for the distribution of $\boldsymbol{Y}^m|\boldsymbol{Y}^o$ are usually regarded as a better choice, since explicit assumptions are made. In fact, the methodology we have developed here is an improvement of the extrapolation method towards the identifying restrictions approach.

To see this we consider the simple scenario of two dropout patterns, namely the completers with a total of N measurements and the dropouts with $N - s$ available measurements. According to (3), the longitudinal measurement model for the two patterns is of the form

$$\begin{bmatrix} \boldsymbol{Y}^o \\ \boldsymbol{Y}^m \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_o & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{do} \\ \boldsymbol{\beta}_{dm} \end{bmatrix} + \boldsymbol{Zb} + \boldsymbol{\varepsilon} \tag{10}$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_c + \boldsymbol{Zb} + \boldsymbol{\varepsilon}, \tag{11}$$

where (10) is the model for the pattern of dropouts, with $\boldsymbol{\beta}_{dm}$ and $\boldsymbol{\beta}_{do}$ denoting the fixed-

effects parameter vectors corresponding to the missing and observed cases, respectively, and $\boldsymbol{X}_o$ and $\boldsymbol{X}_m$ denote the first $N-s$ and the last s rows of the design matrix $\boldsymbol{X}$ for the dropouts. Similarly, (11) is the model for the pattern of completers with $\boldsymbol{\beta}_c$ the corresponding fixed-effects parameter vector. Moreover, for simplicity of exposition, we assume that $cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{n_i}$, where n_i is the number of measurements for the ith individual (either N or $N-s$ here).

Since we work under the normality assumption, the predictive distribution $\boldsymbol{Y}^m|\boldsymbol{Y}^o$ for the dropouts is of the form

$$\boldsymbol{Y}^m|\boldsymbol{Y}^o \sim N(\boldsymbol{X}_m\boldsymbol{\beta}_{dm} + \boldsymbol{V}_1(\boldsymbol{y}^o - \boldsymbol{X}_o\boldsymbol{\beta}_{do}), \boldsymbol{V}_2), \tag{12}$$

where $\boldsymbol{V}_1 = \boldsymbol{Z}_m\boldsymbol{D}\boldsymbol{Z}_o^t[\boldsymbol{Z}_o\boldsymbol{D}\boldsymbol{Z}_o^t + \sigma^2\mathbf{I}]^{-1}$, $\boldsymbol{V}_2 = [\boldsymbol{Z}_m\boldsymbol{D}\boldsymbol{Z}_m^t + \sigma^2\mathbf{I}] - \boldsymbol{V}_1\boldsymbol{Z}_o\boldsymbol{D}\boldsymbol{Z}_m^t$, and $\boldsymbol{Z}_m$, $\boldsymbol{Z}_o$ are defined analogously to $\boldsymbol{X}_m$, $\boldsymbol{X}_o$. Under the restriction of common variance components in the two patterns, the only unidentified component in (12) is the parameter vector $\boldsymbol{\beta}_{dm}$.

Considering the above example, the only difference between the ordinary extrapolation method, an identifying restrictions approach and the method we propose here is in the choice of $\boldsymbol{\beta}_{dm}$ and $\boldsymbol{\beta}_{do}$ in (12). The extrapolation method assumes a common $\boldsymbol{\beta}_e$ in place of $\boldsymbol{\beta}_{dm}$ and $\boldsymbol{\beta}_{do}$. In particular, the ordinary extrapolation method sets $\boldsymbol{\beta}_e \equiv \boldsymbol{\beta}_{do}$ whereas our method assumes that $\boldsymbol{\beta}_e \equiv \boldsymbol{\beta}_*$, where $\boldsymbol{\beta}_*$ is the fixed-effects parameter vector based on (8). On the other hand, the complete case, available case, and neighbouring case missing values restrictions coincide in the case of two patterns (see e.g., Thijs et al., 2002) and assume that $\boldsymbol{\beta}_{dm} \equiv \boldsymbol{\beta}_c$. If we consider the posterior conditional means as point estimators for the fixed-effects parameters and moreover assume negligible prior information for the case that we estimate $\boldsymbol{\beta}_{do}$ and $\boldsymbol{\beta}_c$ separately, we get that

$$\begin{aligned}
\boldsymbol{\beta}_{do} &= \left[\sum_{i\in P_d} \boldsymbol{X}_i^t\boldsymbol{X}_i\right]^{-1}\left[\sum_{i\in P_d} \boldsymbol{X}_i^t(\boldsymbol{y}_i^o - \boldsymbol{Z}_i\boldsymbol{b}_i)\right], \\
\boldsymbol{\beta}_c &= \left[\sum_{i\in P_c} \boldsymbol{X}_i^t\boldsymbol{X}_i\right]^{-1}\left[\sum_{i\in P_c} \boldsymbol{X}_i^t(\boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{b}_i)\right], \\
\boldsymbol{\beta}_* &= \left[\gamma\sum_{i\in P_d} \boldsymbol{X}_i^t\boldsymbol{X}_i + \sigma^2\boldsymbol{\Delta}^{-1}\right]^{-1}\left[\gamma\sum_{i\in P_d} \boldsymbol{X}_i^t(\boldsymbol{y}_i^o - \boldsymbol{Z}_i\boldsymbol{b}_i) + \sigma^2\boldsymbol{\Delta}^{-1}\boldsymbol{\beta}_c\right],
\end{aligned}$$

where $\boldsymbol{\Delta}$ quantifies the prior variability of $\boldsymbol{\beta}_{do}$, and P_c and P_d denote the sets of completers and dropouts, respectively. Based on the above expressions we can see that our

approach is a compromise between the ordinary extrapolation method and the identifying restrictions approach, since $\boldsymbol{\beta}_*$ can be regarded as a weighted average of both $\boldsymbol{\beta}_c$ and $\boldsymbol{\beta}_{do}$ with weights determined by the sensitivity parameter γ and $\sigma^2\boldsymbol{\Delta}^{-1}$. Thus using $\boldsymbol{\beta}_*$, $\boldsymbol{Y}^m|\boldsymbol{Y}^o$ is identified using information from both the completers and the dropouts, whereby relaxing also the assumption of completely different measurement models per dropout pattern.

5. The Toenail Study

We continue with the analysis of the Toenail study which was introduced in Section 1. For our purposes, we will only consider one of the secondary endpoints, i.e., the unaffected nail length which is measured as follows: initially, the treating physician indicates one of the affected toenails as the target nail; at each occasion, the unaffected nail length of the target nail is measured in millimeters. This response obviously relates to the toe size. Therefore, we will only include here those patients for whom the target nail was one of the two big toenails. This reduces our sample under consideration to 150 subjects for the terbinafine group and 148 subjects for the itraconazole group.

Our primary interest is in the comparison of the average evolutions between the two treatment groups. Previous analyses of the same data (Verbeke et al., 2001; Verbeke and Molenberghs, 2000) using both selection and pattern mixture models, have suggested a potential sensitivity of the results. Namely, there seemed to exist a discrepancy between the marginal average evolutions obtained from the selection model and those from the pattern mixture model using the extrapolation method. A possible explanation was that the sparse patterns significantly alter the shape of the marginal average evolutions obtained from the latter approach. In the light of these findings, here we allow the patterns to affect each other according to our proposal and perform a sensitivity analysis.

The previous analyses also showed that quadratic time trends adequately capture the mean longitudinal profiles, while random-intercepts satisfactorily model the correlation structure. Following these results, we also here assume that $p(\boldsymbol{Y}|R)$ is a linear mixed model of the form

$$Y_{ij}^o = \begin{cases} (\beta_{A0}(k) + b_i) + \beta_{A1}(k)t_{ij} + \beta_{A2}(k)t_{ij}^2 + \varepsilon_{ij}, & \text{group A} \\ (\beta_{B0}(k) + b_i) + \beta_{B1}(k)t_{ij} + \beta_{B2}(k)t_{ij}^2 + \varepsilon_{ij}, & \text{group B} \end{cases} \tag{13}$$

where A denotes the terbinafine group, B denotes the itraconazole group, t_i is the time variable and $\beta(k)$ denotes the coefficient of the kth dropout pattern. For the last two patterns (recall that $k = 1$ corresponds to the completers) it is immediately seen that we only have information to fit constant average trends and linear average trends, respectively. Thus, we restrict the parameters in model (13), according to $\beta_{A1}(7) = \beta_{A1}(6)$, $\beta_{A2}(7) = \beta_{A2}(6) = \beta_{A2}(5)$, and $\beta_{B1}(7) = \beta_{B1}(6)$, $\beta_{B2}(7) = \beta_{B2}(6) = \beta_{B2}(5)$. Moreover, for $k = 7$ there is only one measurement per subject and thus no random-intercepts can be included into the model.

To assess the sensitivity of the results we have considered $\sqrt{\gamma}$ values in the set $\mathcal{S} = \{0.01, 0.5, 1, 2, 5\}$. The two extremes, namely 0.01 and 5, are taken such that they correspond to small and large variance for the prior conditionals (6), respectively. Indicative estimates for the $\mathbf{\Delta}_k$ matrices are obtained by fitting model (13) per dropout pattern under maximum likelihood. In order to supplement the sensitivity analysis with the MAR scenario, we also fit the linear mixed effects model to the available data. For each value in $\mathcal{S}$, 60000 Gibbs sampling iterations are used of which the first 10000 are discarded as burn-in. For each chain common graphical checks (e.g., trace-plots, etc.) were performed to assess convergence.

The dropout probabilities are estimated under the multinomial model presented in Section 3.1. We initially assume different cell probabilities for the two groups; however, the likelihood ratio test with the model that assumes equal cell probabilities was non significant (p-value = 0.443). The estimated dropout probabilities and their corresponding standard errors, for the reduced model, are presented in Table 2.

[Table 2 about here.]

The evaluation of the marginal expectations relies on the extrapolation of the fitted average profiles using the posterior modes obtained from (8). In order to enhance predictability and under the restrictions in model (13), for the patterns with one and two available measurements we borrow the linear and quadratic time effects, or just the quadratic time effect, from the patterns with two and three available measurements, respectively. Figure 2 depicts the marginal average evolutions for the various values of the γ parameter; in each panel, the sample marginal average evolutions are also plotted in order to check the fit of the model on the observed data.

[Figure 2 about here.]

In general, we observe some sensitivity in the shapes of the fitted average evolutions, while MAR and $\sqrt{\gamma} = 2$ seem to provide a good fit to the observed data. However, the fit to the observed should not be considered as a criterion for model selection since the observed profiles are incomplete (e.g., see Diggle et al., 2002, sec. 13.3). A comparison between the fitted MCAR evolutions (i.e., $\sqrt{\gamma} = 0.01$) and the sample evolutions reveals discrepancies of the fitted model, especially in baseline for the itraconazole group and onwards for the terbinafine group. This finding suggests that the MCAR assumption is not valid, which is in line with the findings of Verbeke et al. (2001) using a selection modelling approach.

The main interest of the study was to compare marginal average differences between the two treatment groups. Taking into account the randomization, i.e., no treatment difference at baseline, and under (13) this can be formally expressed through the following hypothesis,

$$H_0 : \begin{cases} \sum_{k=1}^{7} \pi_k \beta_{A1}(k) - \sum_{k=1}^{7} \pi_k \beta_{B1}(k) = 0 \\ \sum_{k=1}^{7} \pi_k \beta_{A2}(k) - \sum_{k=1}^{7} \pi_k \beta_{B2}(k) = 0 \end{cases},$$

versus the alternative that H_0 does not hold. Taking advantage of the MCMC output we estimate directly the probability of at least one marginal fixed-effect parameter being equal for the two treatments, i.e.,

$$\mathcal{P} = 1 - Pr\left(\{|\tilde{\beta}_1| < \epsilon\} \cap \{|\tilde{\beta}_2| < \epsilon\}\right), \tag{14}$$

where $\tilde{\beta}_m = \beta_m / \hat{\sigma}_{\beta_m}$, $\hat{\sigma}_{\beta_m}$ is an estimate of the standard deviation for the posterior distribution of β_m, and $\beta_m = \sum_{k=1}^{7} \pi_k (\beta_{Am}(k) - \beta_{Bm}(k))$, with $m = 1, 2$. Based on an approximate Normal or Student's-t assumption for the posterior distribution of the marginal fixed-effects $\tilde{\beta}_m$, we take ϵ equal to 2. The estimates of $\mathcal{P}$ are presented in Table 3.

[Table 3 about here.]

In all cases there is not enough evidence of a significant treatment effect, which is in agreement with the results of the previous analyses. However, we observe small fluctuations, especially for the two extreme γ values and the MAR setting. This reveals

a potential sensitivity of inference to the chosen value of γ, which could be of major importance if the treatment effect was more significant.

6. Conclusion

We have proposed a new method for sensitivity analysis in pattern mixture models under the extrapolation method and demonstrated its usage through a real data example. The two main strengths of this procedure are that it specifies a flexible relationship between the fixed-effects parameters which is in line with the identifying restrictions approach and also shares the simplicity of the ordinary extrapolation method.

The linear mixed-effects measurement model we have considered in Section 3.1 is a constrained version of the full pattern mixture model, assuming that the covariance structures $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_i$ are the same across patterns. This simplified version of the model could have an effect on the precision estimates, especially when there are big discrepancies between the covariance structures of different patterns. In cases where there are indications for different covariance structures among different patterns, our approach could be extended to the covariance parameters as well. In particular, a set of prior conditionals, in the spirit of (6), could be postulated for the variance components and appropriate modifications could be applied to the posterior conditionals. However, two potential difficulties with such an extension arise; firstly, Metropolis-Hastings steps would be required in the MCMC algorithm since the posterior conditionals are not of known forms, and secondly, there would be a much larger set of sensitivity parameters to be considered.

References

Carlin, B. (1996). Hierarchical longitudinal modelling. In Gilks, W., Richardson, S. and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

De Backer, M., De Keyser, P., De Vroey, C. and Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/day – a double-blind comparative trial. *British Journal of Dermatology* **134**, 16–17.

Diggle, P., Heagerty, P., Liang, K.-Y. and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press, New York, 2nd edition.

Doob, J. (1942). The brownian movement and stochastic equations. *Annals of Mathematics* **43**, 351–369.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, to appear.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.

Gilks, W., Wang, C., Yvonnet, B. and Coursaget, P. (1993). Random-effects models for longitudinal data using Gibbs sampling. *Biometrics* **49**, 441–453.

Hogan, J. and Laird, N. (1997a). Mixture models for the joint distribution of repeated measurements and event times. *Statistics in Medicine* **16**, 239–258.

Hogan, J. and Laird, N. (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 data. *Statistical Methods in Medical Research* **8**, 51–83.

Little, R. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.

Little, R. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483.

Little, R. (1995). Modeling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.

Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley, New York, 2nd edition.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* **59**, 829–836.

Rubin, D. (1976). Inference and missing data (with discussion). *Biometrika* **63**, 581–592.

Rubin, D. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* **72**, 538–543.

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G. and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3**, 245–265.

Verbeke, G., Lesaffre, E. and Spiessens, B. (2001). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal* **35**, 419–439.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

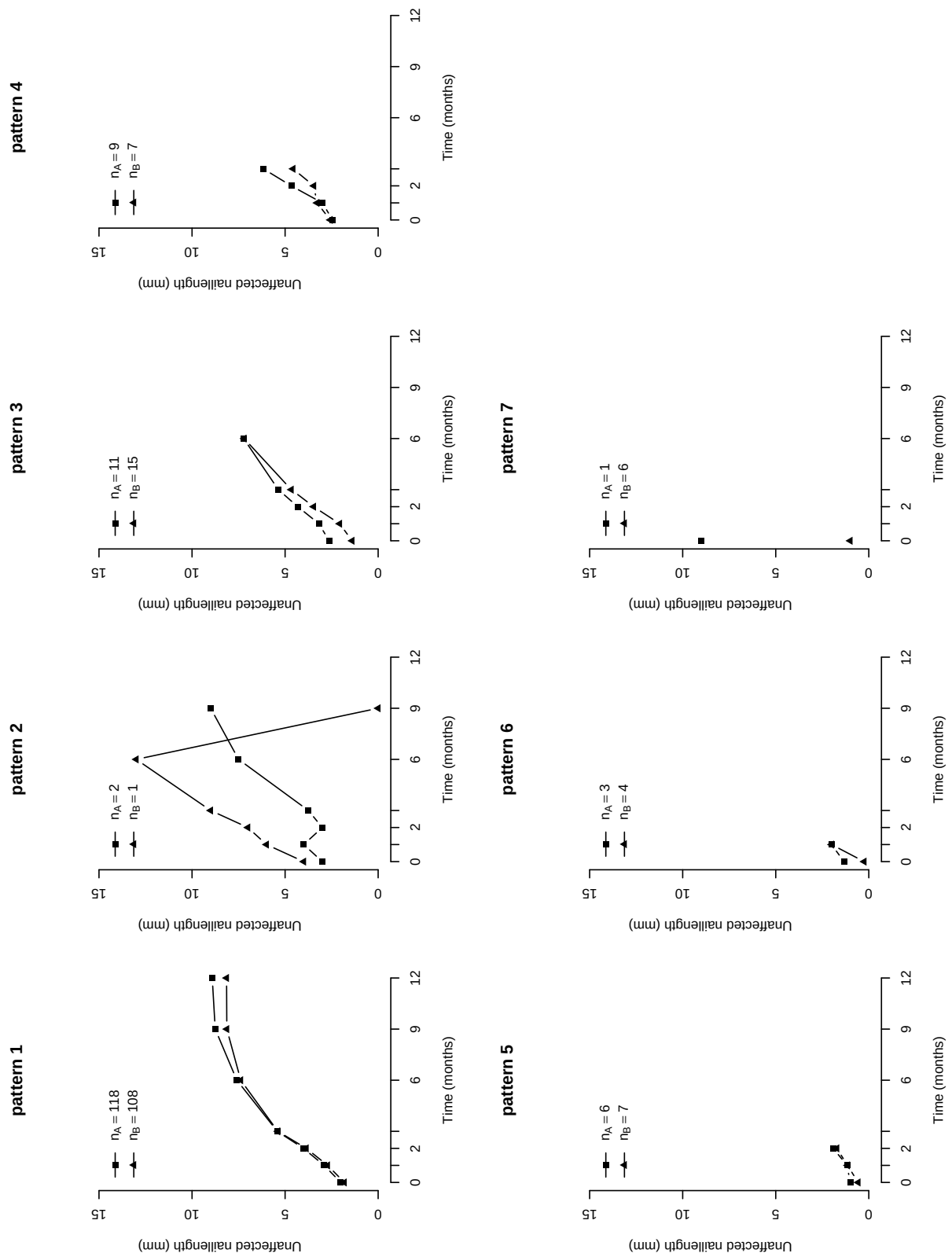


Figure 1. Average profiles for each dropout pattern, based on sample means at each occasion. For each pattern the number of subjects in group A (terbinafine group) and in group B (itraconazole group) is denoted by n_A and n_B, respectively.

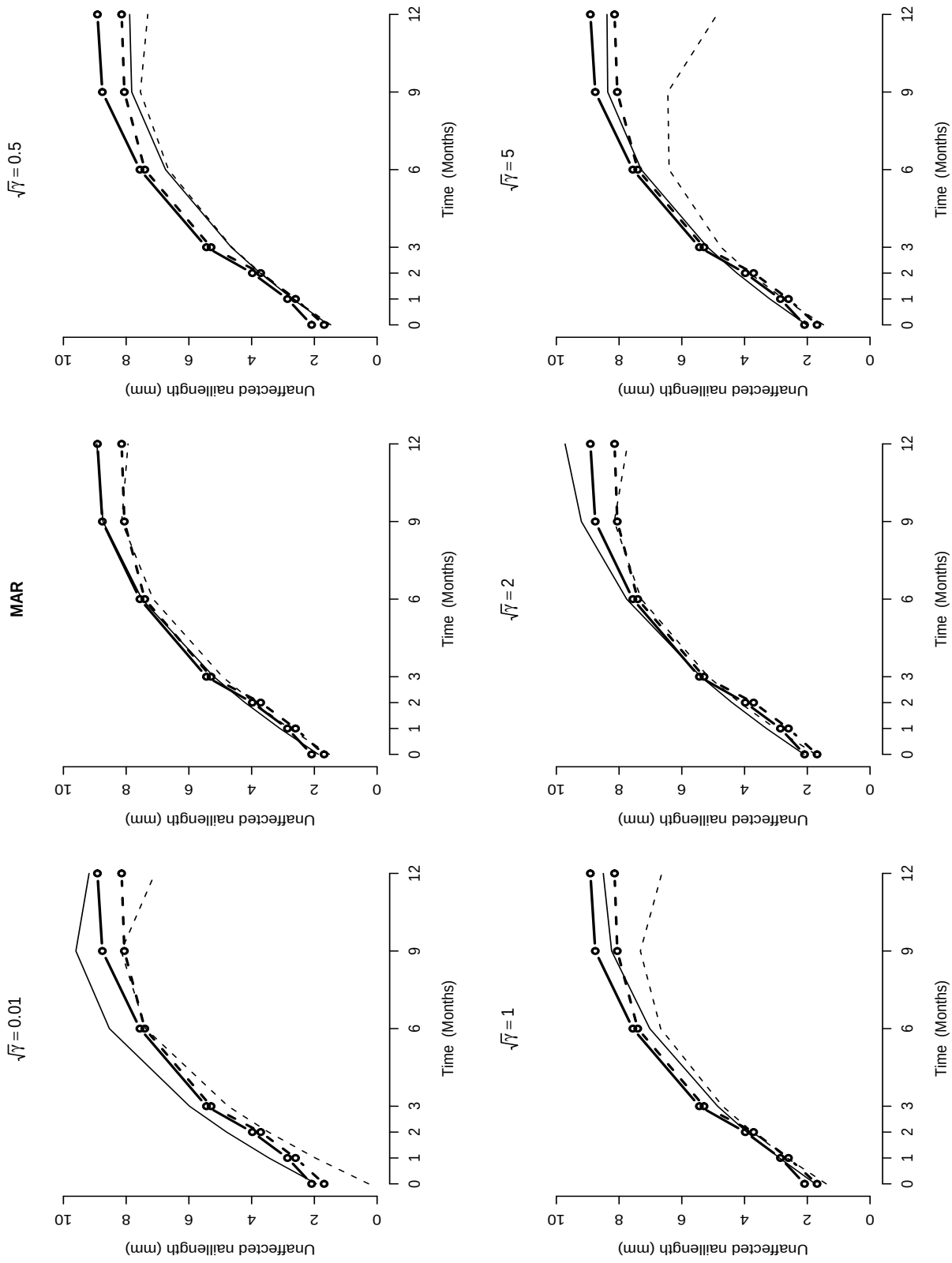


Figure 2. Fitted marginal average evolutions for various values of γ and MAR. The solid line denotes the terbinafine group and the dashed line the itraconazole group. The bold solid and dashed lines with circles depict the sample marginal average evolutions, for the terbinafine group and the itraconazole group, respectively.

Table 1

Toenail Data. Summary of the number of observations taken at each occasion in the TDO study, for each group separately and in total.

	# Observations			# patients per dropout pattern		
Time (months)	terbinafine	itraconazole	Total	terbinafine	itraconazole	Total
0	150	148	298	1	6	7
1	149	142	291	3	4	7
2	146	138	284	6	7	13
3	140	131	271	9	7	16
6	131	124	255	11	15	26
9	120	109	229	2	1	3
12	118	108	226	118	108	226

Table 2

Fitted dropout probabilities and standard errors under the multinomial model.

Dropout pattern k:	1	2	3	4	5	6	7
Fitted prob. $\hat{\pi}_k$	0.758	0.010	0.087	0.054	0.044	0.023	0.023
Stand. error $\hat{\pi}_k$	0.025	0.006	0.016	0.013	0.012	0.009	0.010

Table 3

Estimated probability of at least one marginal fixed-effect being equal for the two treatments, for various values of γ under (9) and MAR.

$\sqrt{\gamma}$	0.01	0.5	1	2	5	MAR
$\mathcal{P}$	26.27%	22.71%	22.48%	22.54%	23.56%	16.82%