

GEERT MOLENBERGHS

HERBERT THIJS

BART MICHIELS

GEERT VERBEKE

MICHAEL G. KENWARD

Pattern-mixture models

Journal de la société française de statistique, tome 145, n° 2 (2004),
p. 49-77

http://www.numdam.org/item?id=JSFS_2004__145_2_49_0

© Société française de statistique, 2004, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

PATTERN-MIXTURE MODELS

Geert MOLENBERGHS*, Herbert THIJS*, Bart MICHIELS**

Geert VERBEKE***, Michael G. KENWARD**** (1)

ABSTRACT

Whereas most models for incomplete longitudinal data are formulated within the selection model framework, pattern-mixture models have gained considerable interest in recent years. We outline several strategies to fit pattern-mixture models, including the so-called identifying-restrictions strategies. Multiple imputation is used to apply these strategies to real sets of data. Our ideas are exemplified using quality-of-life data from a longitudinal study on metastatic breast cancer patients and using a longitudinal clinical trial in Alzheimer patients.

Keywords: Delta Method, linear mixed model, missing data, repeated measures, selection model.

RÉSUMÉ

Si la plupart des analyses de données longitudinales incomplètes sont inscrites dans le cadre du modèle par sélection, l'intérêt pour les modèles par mélange a considérablement progressé dans la période récente. Nous donnons un aperçu de plusieurs stratégies d'ajustement de ces derniers modèles, incluant les stratégies dites sous contraintes d'identification. Une imputation multiple est utilisée pour leur mise en œuvre concrète. Nos idées sont illustrées d'une part sur des données de qualité de vie provenant d'une étude longitudinale sur des patients souffrant d'un cancer du sein à métastases, d'autre part sur un essai clinique longitudinal concernant des patients atteints de la maladie d'Alzheimer.

Mots clés: Données manquantes, mesures répétées, méthode delta, modèle linéaire mixte, modèle par sélection.

1. Introduction

It is not unusual in practice for some sequences of measurements to end early for reasons outside the control of the investigator or study scientist, and any unit so affected is often called a dropout. It might therefore be necessary to accommodate dropout in the modelling process to obtain correct inference. In addition, in some cases, this process can itself be of scientific interest. For

* Limburgs Universitair Centrum, Belgium.

** Janssen Pharmaceutical Research and Development, Belgium.

*** Katholieke Universiteit Leuven, Belgium.

**** London School of Hygiene and Tropical Medicine, UK.

(1) Center for Statistics, Universitaire Campus, B-3590 Diepenbeek, Belgium.

Email: geert.molenberghs@luc.ac.be

example, the scientist may want to know which type of subjects are more prone to dropping out than others (*e.g.*, males may be more likely to drop out than females, there may be an age effect in the dropout probability, or less favorable values on previous outcome measures may adversely affect the dropout probability.). In this paper, we will restrict attention to dropout, *i.e.*, monotone missingness. However, some of the terminology will refer to the general case of missingness.

Rubin (1976) and Little and Rubin (1987, Ch. 6) make important distinction between different missing values processes. A dropout or missingness process is said to be MCAR (missing completely at random) if the dropout is independent of both unobserved and observed data and MAR (missing at random) if, conditional on the observed data, the dropout is independent of the unobserved measurements; otherwise the dropout process is termed MNAR (missing not at random). If a dropout process is MAR then a valid analysis can be obtained through a likelihood-based analysis that ignores the dropout mechanism, provided the parameter describing the measurement process is functionally independent of the parameter describing the dropout process, the so-called parameter distinctness condition. This situation is termed ignorable by Little and Rubin (1987). The practical implication is that a software module with likelihood estimation facilities and with the ability to handle incompletely observed subjects manipulates the correct likelihood, providing valid parameter estimates and likelihood ratio values. In many examples, however, the reasons for dropout are many and varied and it is therefore difficult to justify on *a priori* grounds the assumption of random dropout. Arguably, in the presence of non-random dropout, a wholly satisfactory analysis of the data is not feasible. Several approaches have been proposed in the literature (Little 1995, Kenward and Molenberghs 1999).

Many methods are formulated within the selection modelling frame (Little and Rubin 1987) as opposed to pattern-mixture modelling (PMM; Little 1993, 1994a). A selection model factors the joint distribution of the measurement and response mechanisms into the marginal measurement distribution and the response distribution, conditional on the measurements. This is intuitively appealing since the marginal measurement distribution would be of interest also with complete data. Further, Little and Rubin's taxonomy is most easily developed in the selection setting. However, it is often argued that, especially in the context of non-random missingness models, selection models, although identifiable, should be approached with caution (Glynn, Laird and Rubin 1986). Therefore, pattern-mixture models have gained renewed interest in recent years (Little 1993, 1994a, Hogan and Laird 1997). Examples can be found in Molenberghs *et al* (1998, 1999), Ekholm and Skinner (1998), Little and Wang (1996), Hedeker and Gibbons (1997), Cohen and Cohen (1983), Muthén *et al* (1987), Allison (1987), and McArdle and Hamagani (1992).

An important issue is that pattern-mixture models are by construction under-identified. Little (1993, 1994a) solves this problem through the use of identifying restrictions: inestimable parameters of the incomplete patterns are set equal to (functions of) the parameters describing the distribution of the

completers. Identifying restrictions are not the only way to overcome under-identification and we will discuss alternative approaches as well. Some authors perceive this under-identification as a drawback, we believe it is an asset since it forces one to reflect on the assumptions made. We believe that pattern-mixture models can play an important role, arguably next to a selection model approach, when conducting sensitivity analyses to study how sensitive scientific conclusions are to incompleteness encountered in the data.

Two case studies are introduced in Section 2, while their analysis is given in Sections 5 and 6, respectively. A general framework for modelling incomplete data is given in Section 3 and a perspective on pattern mixture models is offered in Section 4.

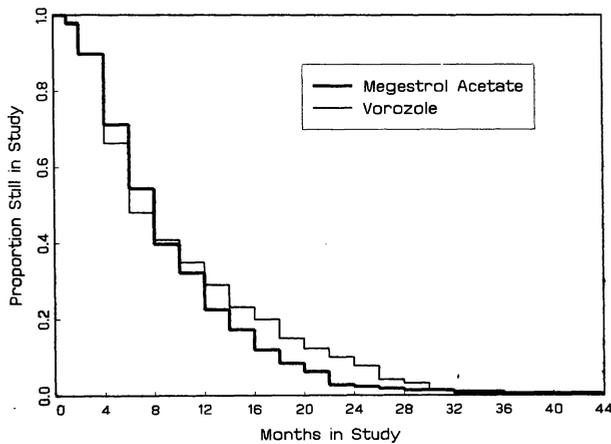


FIG 1. — Vorozole Study. Representation of dropout.

2. Case Studies

2.1. The Vorozole Study

This study was an open-label, multicenter, parallel group design conducted at 67 North American centers (29 Canadian, 38 US). Patients were randomized to either vorozole (2.5 mg taken once daily) or megestrol acetate (40 mg four times daily). The patient population consisted of postmenopausal patients with histologically confirmed estrogen-receptor positive metastatic breast carcinoma. To expedite enrollment, patients with non measurable/non assessable disease at baseline were entered and eligible patients were stratified into three groups according to whether they had measurable, assessable, or nonmeasurable/non assessable disease. All 452 randomized patients were followed until disease progression or death. A graphical representation of the dropout evolution is given in Figure 1. The main objective was to compare the treatment group with respect to response rate while secondary objectives included a comparison relative to duration of response, time to progression,

survival, safety, pain relief, performance status and quality of life. Full details of this study are reported in Goss *et al* (1999). This paper focuses on overall quality of life, measured by the total Functional Living Index in Cancer (FLIC, Schipper *et al* 1984). A higher FLIC score is the more desirable outcome.

Patients underwent screening and for those deemed eligible a detailed examination at baseline (visit 0) took place. Further measurement occasions were months 1, then from months 2 at bi-monthly intervals until month 44. Goss *et al* (1999) analyzed FLIC using a two-way ANOVA model with effects for treatment, disease status (state of tumor), as well as their interaction. No significant difference on any of these effects between the two treatment arms was found. Without being a formal non-inferiority trial, the main conclusion from the primary analysis was that vorozole is well tolerated and as effective as megestrol acetate in the treatment of postmenopausal advanced breast cancer patients with disease progression after tamoxifen treatment. In this paper, we will, apart from treatment, correct for dominant site (the major location of the tumor) of the disease as well as clinical stage (tumor status).

2.2 The Alzheimer Study

The data come from a three-armed clinical trial involving patients with Alzheimer's disease (Reisberg *et al*, 1987), conducted by 50 investigators in 8 countries. The outcome is a dementia score, ranging from 0 to 43. Treatment arm 1 is placebo, with 114 patients, while arms 2, with 115 patients, and 3, with 115 patients, involve active compounds. Of the patient population, 56.4 % are female. There are 341 Caucasians, 2 Orientals and 1 black subject. Age ranges from 56 to 97 years with a median of 81 years. Measurements are taken at baseline, at weeks 1, 2 and then every two weeks until week 12. In agreement with the protocol, we will analyses change versus baseline. This outcome is sufficiently close to normality, unlike the raw score.

Attrition over time is fairly steady for each treatment arm. The sample size per dropout pattern and per treatment arm is displayed in Table 1. Each of the patterns corresponds to the last follow-up visit made. As such, patterns are assumed monotone. An occasional intermediate missing value does not change the membership of a patient to a particular pattern. In each of the arms, about 40 % drop out before the end of the study. Unfortunately, very little is known about the reasons for dropout, in this particular study. While such information is generally important, one also needs to be able to analyse incomplete data in the absence of such knowledge.

TABLE 1. – Sample size per treatment arm and dropout pattern in the Alzheimer's study

Pattern	1	2	3	4	5	6	7
Treatment 1	4	5	16	3	9	6	71
Treatment 2	4	9	7	6	3	5	81
Treatment 3	12	4	15	9	5	3	67

3. Data Setting and Modelling Framework

Assume that for subject $i = 1, \dots, N$ in the study a sequence of responses Y_{ij} is designed to be measured at occasions $j = 1, \dots, n$. The outcomes are grouped into a vector $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in})'$. In addition, define a dropout indicator D_i for the occasion at which dropout occurs and make the convention that $D_i = n + 1$ for a complete sequence. It is often necessary to split the vector \mathbf{Y}_i into observed (\mathbf{Y}_i^o) and missing (\mathbf{Y}_i^m) components respectively.

In modelling missing data one is interested in $f(\mathbf{y}_i, d_i | \boldsymbol{\theta}, \boldsymbol{\psi})$ which is the joint distribution of the measurements Y_i and the dropout indicators. The selection model factorization starts from $f(\mathbf{y}_i, d_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \boldsymbol{\theta}) f(d_i | \mathbf{y}_i, \boldsymbol{\psi})$, while the pattern-mixture approach is based upon $f(\mathbf{y}_i, d_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | d_i, \boldsymbol{\theta}) f(d_i | \boldsymbol{\psi})$.

In practice, particular choices for the outcome and missingness factors need to be chosen. Using the vehicle of longitudinal data, we will review some commonly used choices. We focus on parametric models for continuous, normally distributed outcomes as well as for non-Gaussian outcomes.

In principle, one would like to consider the density of the full data $f(\mathbf{y}_i, d_i | \boldsymbol{\theta}, \boldsymbol{\psi})$, where the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describe the measurement and missingness processes, respectively. Covariates are assumed to be measured but, for notational simplicity, suppressed from notation.

Let us first describe the measurement and missingness models in turn, for the continuous outcome case, and then formally introduce and comment on ignorability. The measurement model will depend on whether or not a full longitudinal analysis is done. In case focus is on the last observed measurement or on the last measurement occasion only, one typically opts for classical two- or multi-group comparisons (t test, Wilcoxon, etc.). In case a longitudinal analysis is deemed necessary, the choice made depends on the nature of the outcome. For continuous outcomes, one typically assumes a linear mixed-effects model, perhaps with serial correlation:

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \mathbf{W}_i + \boldsymbol{\varepsilon}_i, \quad (1)$$

(Verbeke and Molenberghs 2000) where \mathbf{Y}_i is the n dimensional response vector for subject i , $1 \leq i \leq N$, N is the number of subjects, X_i and Z_i are $(n \times p)$ and $(n \times q)$ known design matrices, $\boldsymbol{\beta}$ is the p dimensional vector containing the fixed effects, $\mathbf{b}_i \sim N(\mathbf{0}, D)$ is the q dimensional vector containing the random effects, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 I_n)$ is a n dimensional vector of measurement error components, and $\mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N$ are assumed to be independent. Serial correlation is captured by the realization of a Gaussian stochastic process, \mathbf{W}_i , which is assumed to follow a $N(\mathbf{0}, \tau^2 H_i)$ law. The serial covariance matrix H_i only depends on i through the number n of observations and through the time points t_{ij} at which measurements are taken. The structure of the matrix H_i is determined through the autocorrelation function $\rho(t_{ij} - t_{ik})$. This function decreases such that $\rho(0) = 1$ and $\rho(+\infty) = 0$. Finally, D is a general $(q \times q)$ covariance matrix with (i, j) element $d_{ij} = d_{ji}$. Inference is based on the marginal distribution of the response \mathbf{Y}_i which can be expressed as

$$\mathbf{Y}_i \sim N(X_i \boldsymbol{\beta}, Z_i D Z_i' + \Sigma_i). \quad (2)$$

Here, $\Sigma_i = \sigma^2 I_{n_i} + \tau^2 H_i$ is a $(n \times n)$ covariance matrix grouping the measurement error and serial components.

Two popular choices to capture serial correlation are I by means of exponential or Gaussian decay.

Assume that incompleteness is due to dropout only, and that the first measurement Y_{i1} is obtained for everyone. The model for the dropout process is based on, for example, a logistic regression for the probability of dropout at occasion j , given the subject is still in the study. We denote this probability by $g(\mathbf{h}_{ij}, y_{ij})$ in which \mathbf{h}_{ij} is a vector containing all responses observed up to but not including occasion j , as well as relevant covariates. We then assume that $g(\mathbf{h}_{ij}, y_{ij})$ satisfies

$$\text{logit}[g(\mathbf{h}_{ij}, y_{ij})] = \text{logit}[\text{pr}(D_i = j | D_i \geq j, \mathbf{y}_i)] = \mathbf{h}_{ij}\boldsymbol{\psi} + \omega y_{ij}, \quad (3)$$

($i = 1, \dots, N$). When ω equals zero, the dropout model is MAR, and all parameters can be estimated using standard software since the measurement model for which we use a linear mixed model and the dropout model, assumed to follow a logistic regression, can then be fitted separately. If $\omega \neq 0$, the posited dropout process is MNAR. Model (3) provides the building blocks for the dropout process $f(d_i | \mathbf{y}_i, \boldsymbol{\psi})$. This model is often referred to as Diggle and Kenward's (1994) model.

4. Pattern-Mixture Modelling Approach

Fitting pattern-mixture models can be approached in several ways. It is important to decide whether pattern-mixture and selection modelling are to be contrasted with one another or rather the pattern-mixture modelling is the central focus. In the latter case, it is natural to conduct an analysis, and preferably a sensitivity analysis, *within* the pattern-mixture family. We will explicitly consider two strategies to deal with under-identification.

Strategy 1. Little (1993, 1994a) advocated the use of identifying restrictions and presented a number of examples. One of those, ACMV (available case missing values), is the natural counterpart of MAR in the PMM framework.

Strategy 2. As opposed to identifying restrictions, model simplification can be done to identify the parameters. Thijs *et al* (2002) discussed several sub-strategies in detail.

While the second strategy is computationally simple, it is important to note that there is a price to pay. Indeed, simplified models, qualified as "assumption rich" by Sheiner, Beale and Dunne (1997), also make untestable assumptions, just as in the selection model case. From a technical point of view, Strategy 2 only requires to either consider "pattern" as an extra covariate in the model, or to conduct an analysis "by pattern", such that a separate analysis is obtained for each of the dropout patterns. In the identifying restrictions setting on the other hand (Strategy 1), the assumptions are clear from the start.

Pattern-mixture models do not always automatically provide estimates and standard errors of marginal quantities of interest, such as overall treatment effect or overall time trend. Hogan and Laird (1997) provided a way to derive selection model quantities from the pattern-mixture model. An example of such a marginalization is given in Section 5.

4.1. Identifying Restriction Strategies

In line with the results obtained by Molenberghs, Michiels, Kenward, and Diggle (1998), we restrict attention to monotone patterns. In general, let us assume we have $t = 1, \dots, n = T$ dropout patterns where the dropout indicator, introduced earlier, is $d = t + 1$. For pattern t , the complete data density is given by

$$f_t(y_1, \dots, y_T) = f_t(y_1, \dots, y_t) f_t(y_{t+1}, \dots, y_T | y_1, \dots, y_t). \quad (4)$$

The first factor is clearly identified from the observed data, while the second factor is not. It is assumed that the first factor is known or, more realistically, modeled using the observed data. Then, identifying restrictions are applied in order to identify the second component.

While, in principle, completely arbitrary restrictions can be used by means of any valid density function over the appropriate support, strategies which relate back to the observed data deserve privileged interest. One can base identification on all patterns for which a given component, y_s say, is identified. A general expression for this is

$$f_t(y_s | y_1, \dots, y_{s-1}) = \sum_{j=s}^T \omega_{sj} f_j(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (5)$$

We will use ω_s as shorthand for the set of ω_{sj} 's used, the components of which are typically positive. Every ω_s which sums to one provides a valid identification scheme.

Let us incorporate (5) into (4):

$$\begin{aligned} & f_t(y_1, \dots, y_T) \\ &= f_t(y_1, \dots, y_t) \prod_{s=0}^{T-t-1} \left[\sum_{j=T-s}^T \omega_{T-s,j} f_j(y_{T-s} | y_1, \dots, y_{T-s-1}) \right]. \end{aligned} \quad (6)$$

Let us consider three special but important cases, associated with these choices of ω_s in (5). Little (1993) proposes CCMV (complete case missing values) which uses the following identification:

$$f_t(y_s | y_1, \dots, y_{s-1}) = f_T(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (7)$$

In other words, information which is unavailable is always borrowed from the completers. Alternatively, the nearest identified pattern can be used:

$$f_t(y_s | y_1, \dots, y_{s-1}) = f_s(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (8)$$

We will refer to these restrictions as *neighboring case missing values* or NCMV.

The third special case of (5) will be ACMV. Thus, ACMV is reserved for the counterpart of MAR in the PMM context. The corresponding ω_s vectors can be shown to have components:

$$\omega_{sj} = \frac{\alpha_j f_j(y_1, \dots, y_{s-1})}{\sum_{\ell=s}^T \alpha_\ell f_\ell(y_1, \dots, y_{s-1})}, \quad (9)$$

($j = s, \dots, T$) where α_j is the fraction of observations in pattern j (Molenberghs, Michiels, Kenward and Diggle 1998).

This MAR-ACMV link connects the selection and pattern-mixture families. It is further of interest to consider specific sub-families of the MNAR family. In the selection model context, one typically restricts attention to a class of mechanisms where dropout may depend on the current, possibly unobserved, measurement, but not on future measurements. The entire class of such models will be termed missing non-future dependent (MNFD). While they are natural and easy to consider in a selection model context, there exist important examples of mechanisms that do not satisfy MNFD, such as shared-parameter models (Wu and Bailey 1989, Little 1995).

Kenward, Molenberghs, and Thijs (2003) have shown there is a counterpart to MNFD in the pattern-mixture context. The conditional probability of pattern t in the MNFD selection models obviously satisfies

$$f(r = t|y_1, \dots, y_T) = f(r = t|y_1, \dots, y_{t+1}). \quad (10)$$

Within the PMM framework, we define non-future dependent missing value restrictions (NFMV) as follows:

$$\forall t \geq 2, \forall j < t - 1 : \quad (11)$$

$$f(y_t|y_1, \dots, y_{t-1}, r = j) = f(y_t|y_1, \dots, y_{t-1}, r \geq t - 1).$$

NFMV is not a single set of restrictions, but rather leaves one conditional distribution per incomplete pattern unidentified:

$$f(y_{t+1}|y_1, \dots, y_t, r = t). \quad (12)$$

In other words, the distribution of the “current” unobserved measurement, given the previous ones, is unconstrained. Note that (11) excludes such mechanisms as CCMV and NCMV. Kenward, Molenberghs, and Thijs (2003) have shown that, for longitudinal data with dropouts, MNFD and NFMV are equivalent.

For pattern t , the complete data density is given by

$$f_t(y_1, \dots, y_T) \quad (13)$$

$$= f_t(y_1, \dots, y_t) f_t(y_{t+1}|y_1, \dots, y_t) f_t(y_{t+2}, \dots, y_T|y_1, \dots, y_{t+1}).$$

PATTERN-MIXTURE MODELS

It is assumed that the first factor is known or, more realistically, modeled using the observed data. Then, identifying restrictions are applied in order to identify the second and third components. First, from the data, estimate $f_t(y_1, \dots, y_t)$. Second, the user has full freedom to choose

$$f_t(y_{t+1}|y_1, \dots, y_t). \quad (14)$$

Substantive considerations can be used to identify this density. Or a family of densities can be considered by way of sensitivity analysis. Third, using (11), the densities $f_t(y_j|y_1, \dots, y_{j-1})$, ($j \geq t+2$) are identified. This identification involves not only the patterns for which y_j is observed, but also the pattern for which y_j is the current, the first unobserved measurement. An overview of the connection between selection and pattern-mixture models is given in Figure 2.

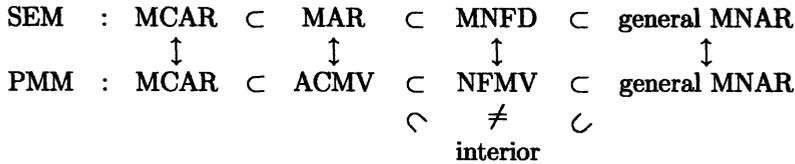


FIG 2. — Relationship between nested families within the selection model, SEM, and pattern-mixture model, PMM, families. MCAR: missing completely at random; MAR: missing at random; MNAR: missing not at random; MNFD: missing non-future dependence; ACMV: available-case missing values; NFMV: non-future missing values; interior: restrictions based on a combination of the information available for other patterns. The “⊂” symbol here indicates “is a special case of”. The “↓” symbol indicates correspondence between a class of SEM models and a class of PMM models.

Two obvious mechanisms, within the MNFD family but outside MAR, are NFD1 (NFD standing for “non-future dependent”), *i.e.*, choose (14) according to CCMV, and NFD2, *i.e.*, choose (14) according to NCMV. NFD1 and NFD2 are strictly different from CCMV and NCMV.

4.2 How to Use Restrictions ?

We will briefly outline a general strategy. Several points which require further specification will be discussed in what follows. (1) Fit a model to the pattern-specific identifiable densities: $f_t(y_1, \dots, y_t)$. This results in a parameter estimate, $\hat{\gamma}_t$. (2) Select an identification method of choice. (3) Using this identification method, determine the conditional distributions of the unobserved outcomes, given the observed ones:

$$f_t(y_{t+1}, \dots, y_T|y_1, \dots, y_t). \quad (15)$$

(4) Using standard multiple imputation methodology (Rubin 1987, Schafer 1997, Verbeke and Molenberghs 2000, Minini and Chavance 2004ab), draw multiple imputations for the unobserved components, given the observed

outcomes and the correct pattern-specific density (15). (5) Analyze the multiply-imputed sets of data using the method of choice. This can be another pattern-mixture model, but also a selection model or any other desired model. (6) Inferences can be conducted in the standard multiple imputation way (Rubin 1987, Schafer 1997, Verbeke and Molenberghs 2000).

We have seen how general identifying restrictions (5), with CCMV, NCMV, and ACMV as special cases, lead to the conditional densities for the unobserved components, given the observed ones. This came down to deriving expressions for ω , such as in (9) for ACMV. In addition, we need to draw imputations from the conditional densities.

Let us proceed by studying the special case of three measurements first. To this end, we consider an identification scheme and we start off by avoiding the specification of a parametric form for these densities. The following steps are required: (1) Estimate the parameters of the identifiable densities: from pattern 3, $f_3(y_1, y_2, y_3)$; from pattern 2, $f_2(y_1, y_2)$; and from pattern 1, $f_1(y_1)$. (2) To properly account for the uncertainty with which the parameters are estimated, we need to draw from them as is customarily done in multiple imputation. It will be assumed that in all densities from which we draw, this parameter vector is used. (3) **For pattern 2.** Given an observation in this pattern, with observed values (y_1, y_2) , calculate the conditional density $f_3(y_3|y_1, y_2)$ and draw from it. (4) **For pattern 1.** We now have to distinguish three sub steps.

1. There is now only one ω involved: for pattern 1, in order to determine $f_1(y_2|y_1)$, as a combination of $f_2(y_2|y_1)$ and $f_3(y_2|y_1)$. Every ω in the unit interval is valid. Specific cases are: for NCMV, $\omega = 1$; for CCMV, $\omega = 0$; for ACMV, ω identifies a linear combination across patterns. Note that, given y_1 , this is a constant, depending on α_2 and α_3 . For NFD1 and NFD2, the first unidentified conditional density can be chosen freely, thereafter a system of ω 's has to be chosen as well.
To pick one of the two components f_2 or f_3 , we need to generate a random uniform variate, U say, except in the boundary NCMV and CCMV cases.
2. If $U \leq \omega$, calculate $f_2(y_2|y_1)$ and draw from it. Otherwise, do the same based on $f_3(y_2|y_1)$.
3. Given the observed y_1 and given y_2 which has just been drawn, calculate the conditional density $f_3(y_3|y_1, y_2)$ and draw from it.

All steps but the first one have to be repeated M times, to obtain the same number of imputed datasets. Inference then proceeds as outlined Rubin (1987), Schafer (1997) and Verbeke and Molenberghs (2000).

In case the observed densities are assumed to be normal, the corresponding conditional densities are particularly straightforward.

In several cases, the conditional density is a mixture of normal densities. Then an additional and straightforward draw from the components of the mixture is necessary.

The algorithmic detail outlined in this section is necessary to implement pattern-mixture models under Strategy 1. Once done, parameter estimates can be studied and hypotheses tests conducted in a conventional way.

5. Analysis of the Vorozole Study

The mean profiles per treatment arm, as well as their 95 % confidence intervals, are plotted in Figure 3. The average profiles indicate an increase over time which is slightly stronger for the vorozole group until month 14, and afterwards, the megestrol acetate group shows a slightly higher FLIC score. As can be seen from the confidence intervals, these differences are clearly not significant.

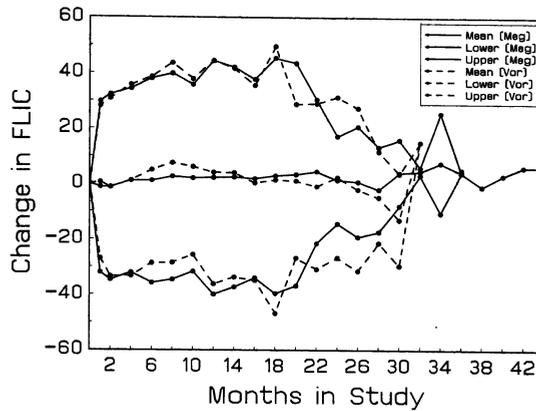


FIG 3. — Vorozole Study. Mean profiles and 95 % confidence intervals

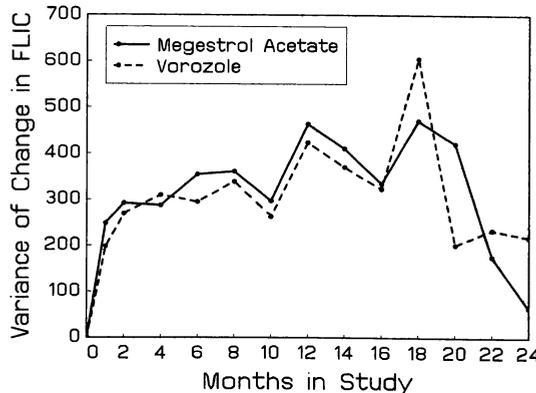


FIG 4. — Vorozole Study. Variance function

The individual profiles augment the averaged plot with a suggestion of the variability seen within the data. The thinning of the data towards the later

study times (only 13 remaining in month 26) suggests that trends at later times should be treated with caution. Therefore, we decided to restrict attention to the first 2 years only rather than the full 4 year period. This leads to a maximum of 13 observations per subject (month 1, 2, 4, 6, . . . , 24). While these plots also give us some indications about the variability at given times and even about the correlation between measurements of the same individual, it is easier to base such considerations on residual profiles and standardized residual profiles.

The variance function is displayed in Figure 4. The variance function seems to be relatively stable, except for a sharp decline near the end (at which point there are large dropout rates), and hence a constant variance model is a plausible starting point.

5.1. A Selection Model Formulation

Let us consider Diggle and Kenward's (1994) selection model as introduced in Section 3. For the measurement model, we start by ignoring the dropout mechanism. This choice will turn out to be justified at the end of this section. Since we are modelling change versus baseline, all models are forced to pass through the origin. This is done by allowing the main covariate effects, but only through their interactions with time. The following covariates were considered for the measurement model: baseline value, treatment, dominant site, and time in months (up to a cubic time trend). Second order interactions were considered as well. Then, a backwards selection procedure was performed. For design reasons, treatment was kept in the model in spite of its non-significance. An F test for treatment effect produces a p value of 0.5822. Apart from baseline, no other time-stationary covariates were kept. A quadratic time effect provided an adequate description of the time trend. Based on the variogram, we confined the random-effects structure to random intercepts, and supplemented this with a spatial Gaussian process and measurement error. The final model is presented in Table 2. The total correlation between two measurements, one month apart, equals 0.696. The residual correlation, which remains after accounting for the random effects, is still equal to 0.491. The serial correlation, obtained by further ignoring the measurement error, equals $\rho = (-1/7.22^2) = 0.981$.

Fitted profiles are displayed in Figure 5 and Figure 6. In Figure 6, empirical Bayes estimates of the random effects are included whereas in Figure 5 the purely marginal mean is used. For each treatment group, we obtain three sets of profiles. The fitted complete profile is the average curve that would be obtained, had all individuals been completely observed. If we use only those predicted values that correspond to occasions at which an observation was made, then the fitted incomplete profiles are obtained. The latter are somewhat above the former when the random effects are included, and somewhat below when they are not, suggesting that individuals with lower measurements are more likely to disappear from the study. In addition, while the fitted complete curves are very close (the treatment effect was not significant), the fitted incomplete curves are not, suggesting that there is more

PATTERN-MIXTURE MODELS

dropout in the standard arm than in the treatment arm. This is in agreement with the observed dropout rate and should not be seen as evidence of a bad fit. Finally, the observed curves, based on the measurements available at each time point, are displayed. These are higher than the fitted ones, but this should be viewed with the standard errors of the observed means in mind (see Figure 3).

TABLE 2. – Vorozole Study. Selection Model

Effect	Estimate (s.e.)
<i>Fixed-Effect Parameters:</i>	
time	7.78 (1.05)
time*baseline	-0.065 (0.009)
time*treatment	0.086 (0.157)
time ²	-0.30 (0.06)
time ² *baseline	0.0024 (0.0005)
<i>Variance Parameters:</i>	
random intercept (δ^2)	105.42
serial variance (τ^2)	77.96
serial association (ϕ)	7.22
measurement error (σ^2)	77.83

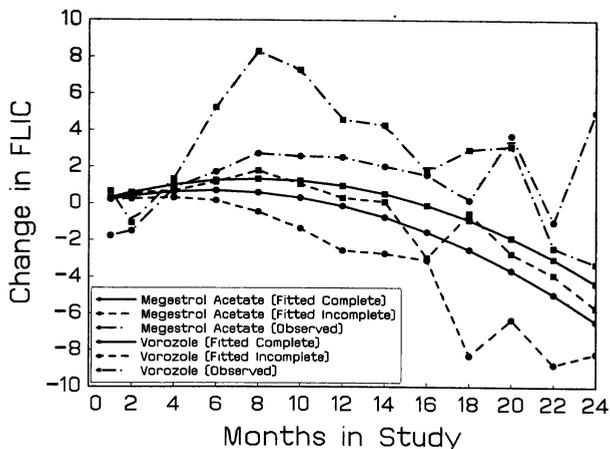


FIG 5. — Vorozole Study. Fitted profiles (averaging the predicted means for the incomplete and complete measurement sequences, without the random effects)

Next, we will study factors which influence dropout. A logistic regression model, described by (3) is used. To start, we restrict attention to MAR processes. The first model includes treatment, dominant site, baseline value,

PATTERN-MIXTURE MODELS

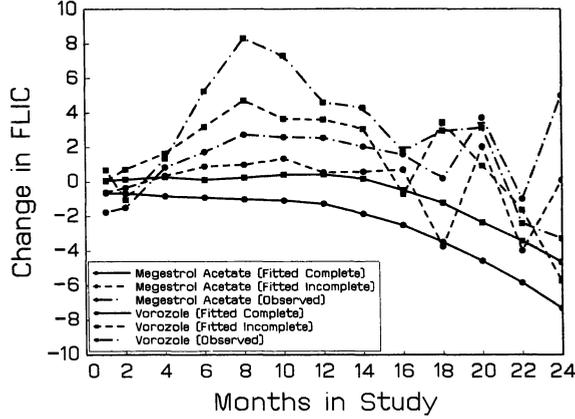


FIG 6. — Vorozole Study. Fitted Profiles (averaging the predicted means for the incomplete and complete measurement sequences, including the random effects)

and the previous measurement but only the last two are significant, producing

$$\text{logit}[g(\mathbf{h}_{i,j})] = 0.080(0.341) - 0.014(0.003)\text{base}_i - 0.033(0.004)y_{i,j-1}. \quad (16)$$

Diggle and Kenward (1994) and Molenberghs, Kenward, and Lesaffre (1997) considered non-random versions of this model by including the current, possible unobserved measurement, such as in (3). This requires more elaborate fitting algorithms, since the missing data process is then non-ignorable. Diggle and Kenward used the simplex algorithm (Nelder and Mead 1965), while Molenberghs, Kenward, and Lesaffre (1997) fitted their models with the EM algorithm. Both Diggle and Kenward and Molenberghs, Kenward, and Lesaffre observed that in informative models, dropout tends to depend on the increment, *i.e.*, the difference between the current and previous measurements $Y_{ij} - Y_{i,j-1}$. Clearly, a very similar quantity is obtained as $Y_{i,j-1} - Y_{i,j-2}$, but a major advantage of such a model is that it fits within the MAR framework. In our case, we obtain

$$\begin{aligned} \text{logit}[g(\mathbf{h}_{i,j})] &= 0.033(0.401) - 0.013(0.003)\text{base}_i \\ &\quad + 0.012(0.006)y_{i,j-2} - 0.035(0.005)y_{i,j-1} \\ &= 0.033(0.401) - 0.013(0.003)\text{base}_i \\ &\quad - 0.023(0.005)\frac{y_{i,j-2} + y_{i,j-1}}{2} \\ &\quad - 0.047(0.010)\frac{y_{i,j-1} - y_{i,j-2}}{2} \end{aligned} \quad (17)$$

indicating that both size and increment are significant predictors for dropout. In (17), j starts from 2 onwards and it is assumed that the first follow-up

measurement is always obtained. We conclude that dropout increases with a decrease in baseline, in overall level of the outcome variable, as well as with a decreasing evolution in the outcome.

Both dropout models (16) and (17) can be compared with their non-random counterparts, where y_{ij} is added to the linear predictor. The first one becomes

$$\text{logit}[g(\mathbf{h}_{ij}, y_{ij})] = 0.53 - 0.015\text{base}_i - 0.076y_{i,j-1} + 0.057y_{ij} \quad (18)$$

while the second one becomes

$$\begin{aligned} \text{logit}[g(\mathbf{h}_{ij}, y_{ij})] \\ = 1.38 - 0.021\text{base}_i - 0.0027y_{i,j-2} - 0.064y_{i,j-1} + 0.035y_{ij}. \end{aligned} \quad (19)$$

Formal testing of dropout models (18) versus (16) and for (19) versus (17) are possible in principle, but will not be carried out for two reasons. First, the likelihood function tends to be very flat for non-random dropout models and therefore the determination of the likelihood ratio is often computationally non-trivial. More fundamentally, Rubin (1994), Little (1994b), and Laird (1994), and Molenberghs, Kenward and Lesaffre (1997) point out that formal testing for non-random dropout faces philosophical objections. Indeed, non-random dropout models are identified only due to strong but unverifiable assumptions. Hogan and Laird (1997) suggest pattern-mixture models as a viable alternative.

5.2. Pattern-Mixture Model Formulations

We will first analyze the data using basic pattern-mixture models and then switch over to pattern-mixture models based on identifying restrictions.

5.2.1. Initial Pattern-Mixture Models

The dropout process (3) simplifies to $f(d_i|S_i, \psi)$ (S_i being covariates) which is a, possibly covariate-corrected, model for the probability to belong to a particular pattern. Its components, $g(h_{ij})$, containing only covariates now, describe the dropout rate at each occasion.

The measurement model has to reflect dependence on dropout. In its most general form, this implies that (1) is replaced by

$$\begin{cases} \mathbf{Y}_i &= X_i\boldsymbol{\beta}(d_i) + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \\ \mathbf{b}_i &\sim N(\mathbf{0}, D(d_i)), \\ \boldsymbol{\varepsilon}_i &\sim N(\mathbf{0}, \Sigma_i(d_i)). \end{cases} \quad (20)$$

Thus, the fixed effects as well as the covariance parameters are allowed to change with dropout pattern and *a priori* no restrictions are placed on the structure of this change.

As discussed in Section 4, model family (20) contains under identified members since it describes the full set of measurements in pattern d_i , even though

there are not measurements after occasion $d_i - 1$. To avoid this problem, simplified (identified) models can be considered. The advantage is that the number of parameters decreases, which is generally an issue with pattern-mixture models. Hogan and Laird (1997) noted that in order to estimate the large number of parameters in general models, one has to make the awkward requirement that each dropout pattern is sufficiently “filled”, in other words one has to require large numbers of dropouts.

In analogy with the exploration in the selection model context, it is natural to explore the data from a pattern-mixture point of view. To this end, plots per dropout pattern can be constructed. Figure 7 display pattern-specific profiles. Figure 7 clearly shows that pattern-specific profiles are of a quadratic nature with in most cases a sharp decline prior to dropout. Note that this is in line with the fitted dropout mechanism (17). Therefore, this feature needs to be reflected in the pattern-mixture model. In analogy with our selection model, the profiles are forced to pass through the origin. This is done by allowing only time main effect and interactions of other covariables with time in the model.

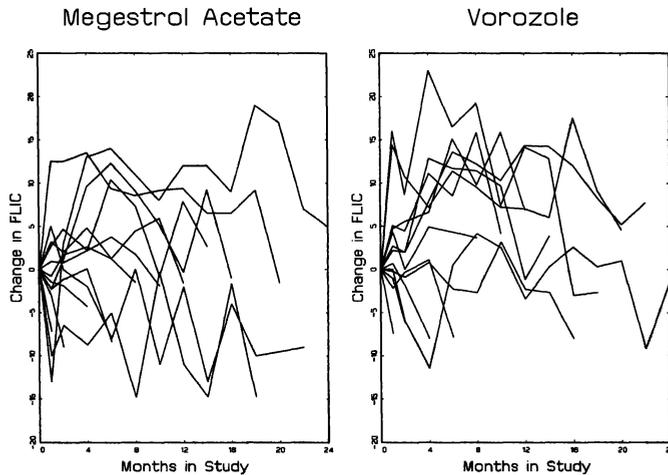


FIG 7. — Vorozole Study. Mean profiles, per dropout pattern

The most complex pattern-mixture model we consider includes a different parameter vector for each of the observed patterns. This is done by including the interaction of all effects in the model with *pattern*, a factor variable calculated as 2+ the number of observations after baseline. We then proceed by backward selection in order to simplify the model. First, we found that the covariance structure is common to all patterns, encompassing random intercept, a serial exponential process, and measurement error.

For the fixed effects we proceeded as follows. A backward selection procedure, starting from a model that includes a main effect of time and time², as well as interactions of time with baseline value, treatment effect, dominant site and pattern, and the interaction of pattern with time². This procedure revealed

PATTERN-MIXTURE MODELS

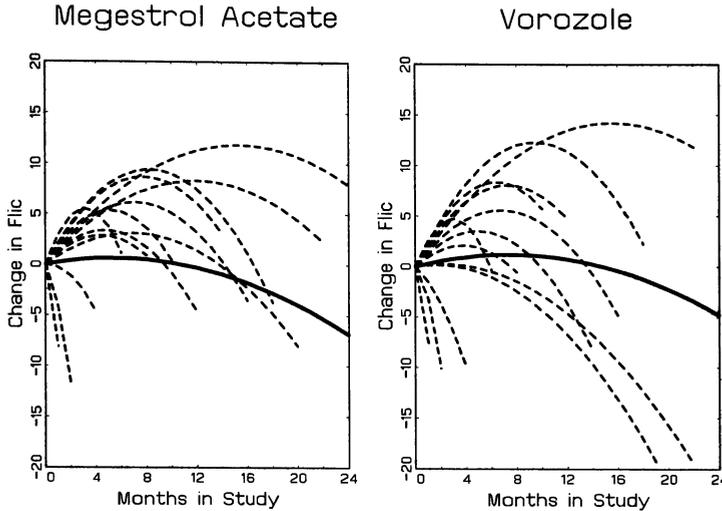


FIG 8. — Vorozole Study. Fitted selection and first pattern-mixture models

main effects of time and time^2 , as well as interactions of time with baseline value, treatment effect, and pattern, and the interaction of pattern with time^2 . As was the case with the selection model in Table 2, treatment effect is non-significant. Indeed, a single degree of freedom F test yields a p value of 0.6868. Note that such a test is possible since treatment effect does not interact with pattern, in contrast to the model which we will describe later. The fitted profiles are displayed in Figure 8. We observe that the profiles for both arms are very similar. This is due to the fact that treatment effect is not significant but perhaps also because we did not allow a more complex treatment effect. For example, we might consider an interaction of treatment with the square of time and, more importantly, an treatment effect which is pattern-specific. Some evidence for such an interaction is seen in Figure 7. Our second, expanded model, allowed for up to cubic time effects, the interaction of time with dropout pattern, dominant site, baseline value and treatment, as well as their two- and three-way interactions. After a backward selection procedure, the effects included are time and time^2 , the two-way interaction of time and dropout pattern, as well as three factor interactions of time and dropout pattern with (1) baseline, (2) group, and (3) dominant site. Finally, time^2 interacts with dropout pattern and with the interaction of baseline and dropout pattern. No cubic time effects were necessary, which is in agreement with the observed profiles in Figure 7. The model is graphically represented in Figure 9.

Because a pattern-specific parameter has been included, we have several options for the assessment of treatment. Since there are 13 patterns (remember we cut off the patterns at 2 years), one can test the global hypothesis, based on 13 degrees of freedom, of no treatment effect. We obtain $F = 1.25$, producing $p = 0.2403$, indicating that there is no overall treatment effect. Each of the treatment effects separately is at a non-significant level. Alternatively, the

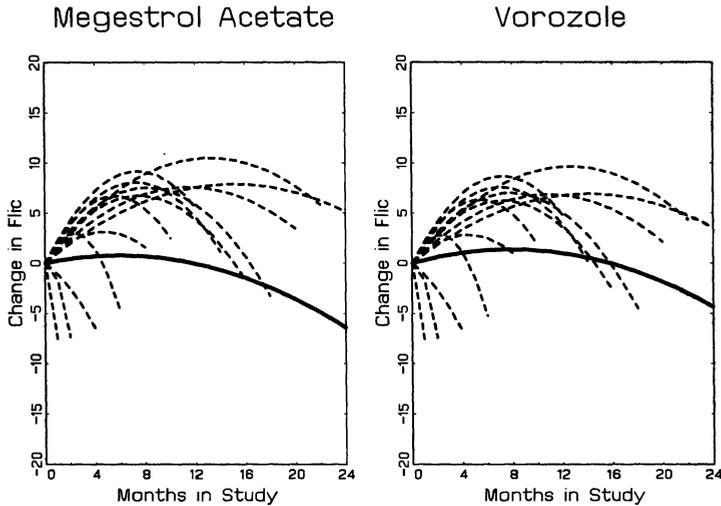


FIG 9. — Vorozole Study. Fitted selection and second pattern-mixture models

marginal effect of treatment can be calculated, which is the weighted average of the pattern-specific treatment effects, with weights given by the probability of occurrence of the various patterns. Its standard error is calculated using a straightforward application of the delta method. This effect is equal to $-0.286(0.288)$ producing a p value of 0.3206, which is still non-significant.

In summary, we obtain a non-significant treatment effect from all our different models, which gives more weight to this conclusion. Further, the pattern-mixture model formulation has lead to important additional insight, primarily regarding the evolution of a patient's profile in connection with her dropout time, which in a selection model would go unnoticed.

5.2.2. Models Based on Identifying Restrictions

Consider those subjects with 1, 2, and 3 follow-up measurements, respectively, i.e, we restrict attention to three patterns only. Thus, 190 subjects are included into the analysis, with subsample sizes 35, 86, and 69, respectively. The corresponding pattern probabilities are $\hat{\pi} = (0.184, 0.453, 0.363)'$. The asymptotic variance-covariance matrix can be derived without difficulty. We will now apply each of the three strategies. We recognize a full analysis, using all patterns, is both interesting and feasible.

The patients in this study drop out mainly because they relapse or die. This in itself poses specific challenges that can be addressed within the pattern-mixture framework much easier than in the selection model framework. Indeed, if one is prepared to make the assumption that a patient who dies is representative of a slice of the population with the same characteristics, and with a certain probability to die, then identifying restrictions (*i.e.*, extrapolation beyond the time of death) is meaningful. In case one does not want to extrapolate beyond the moment of death, one can restrict modelling to the observed data only. An intermediate approach would be to allow for

extrapolation beyond relapse and not beyond death. (For the current dataset, the information needed in order to do so is unavailable.) Note that, while this may seem a disadvantage of pattern-mixture models, we believe it is an asset, because this framework not only forces one to think about such issues, it also provides a modelling solution, no matter which point of view is adopted. This contrasts with selection models where extrapolation is always done, be it explicitly by modelling the profile, averaged over all patterns.

To apply the identifying restriction **Strategy 1**, one first needs to fit a model to the observed data. We will opt for a simple model, with parameters specific to each pattern. Such a model can be seen as belonging to the second modelling strategy.

We include time and time² effects, as well as their interactions with treatment. Further, time by baseline value interaction is included as well, since it is believed to have an important impact on a patient's profile and chance of dropping out. All effects interact with time, in order to force profiles to pass through the origin, since we are studying change versus baseline. An unstructured 3×3 covariance matrix is assumed for each pattern.

Parameter estimates are presented in Table 3, in the "initial" column. Of course, not all parameters are estimable. This holds for the variance components, where in patterns 1 and 2 the upper 1×1 block and the upper 2×2 block are identified, respectively. In the first pattern, the effects in time² are unidentified. The linear effects are identified by virtue of the absence of an intercept term.

Let us present this and later models graphically. Since there is one binary (treatment arm) and one continuous covariate (baseline level of FLIC score), insight can be obtained by plotting the models for selected values of baseline. Precisely, we chose the average value (Figure 10). Bold line type is used for the range over which data are obtained for a particular pattern and extrapolation is indicated using thinner line type. Note that the extrapolation can have surprising effects, even with these relatively simple models. Thus, while this form of extrapolation is simple, its plausibility can be called into question.

This initial model provides a basis, and its graphical representation extra motivation, to consider identifying restriction models. Results are presented in Table 3, for each of the three types of restrictions (CCMV, NCMV, ACMV). For patterns 1 and 2 there is some variability in the parameter estimates across the three strategies, although this is often consistent with random variation (see the standard errors). Since the data in pattern 3 are complete, there is of course no difference between the initial model parameters and those obtained with each of the identifying restriction techniques.

In all of the plots, the same mean response scale was retained, illustrating that the identifying restriction strategies extrapolate much closer to the observed data mean responses. There are some differences among the identifying restriction methods, but this is not graphically represented here. This conclusion needs to be considered carefully. Since these patients drop out mainly because they relapse or die, it seems unlikely to expect a rise in quality of life. This consideration is evidence against CCMV, where missing information is

PATTERN-MIXTURE MODELS

Effect	initial	CCMV	NCMV	ACMV
<i>Pattern 1:</i>				
Time	3.40(13.94)	13.21(15.91)	7.56(16.45)	4.43(18.78)
Time*base	-0.11(0.13)	-0.16(0.16)	-0.14(0.16)	-0.11(0.17)
Time*treat	0.33(3.91)	-2.09(2.19)	-1.20(1.93)	-0.41(2.52)
Time ²		-0.84(4.21)	-2.12(4.24)	-0.70(4.22)
Time ² *treat		0.01(0.04)	0.03(0.04)	0.02(0.04)
σ_{11}	131.09(31.34)	151.91(42.34)	134.54(32.85)	137.33(34.18)
σ_{12}		59.84(40.46)	119.76(40.38)	97.86(38.65)
σ_{22}		201.54(65.38)	257.07(86.05)	201.87(80.02)
σ_{13}		55.12(58.03)	49.88(44.16)	61.87(43.22)
σ_{23}		84.99(48.54)	99.97(57.47)	110.42(87.95)
σ_{33}		245.06(75.56)	241.99(79.79)	286.16(117.90)
<i>Pattern 2:</i>				
Time	53.85(14.12)	29.78(10.43)	33.74(11.11)	28.69(11.37)
Time*base	-0.46(0.12)	-0.29(0.09)	-0.33(0.10)	-0.29(0.10)
Time*treat	-0.95(1.86)	-1.68(1.21)	-1.56(2.47)	-2.12(1.36)
Time ²	-18.91(6.36)	-4.45(2.87)	-7.00(3.80)	-4.22(4.20)
Time ² *treat	0.15(0.05)	0.04(0.02)	0.07(0.03)	0.05(0.04)
σ_{11}	170.77(26.14)	175.59(27.53)	176.49(27.65)	177.86(28.19)
σ_{12}	151.84(29.19)	147.14(29.39)	149.05(29.77)	146.98(29.63)
σ_{22}	292.32(44.61)	297.38(46.04)	299.40(47.22)	297.39(46.04)
σ_{13}		57.22(37.96)	89.10(34.07)	99.18(35.07)
σ_{23}		71.58(36.73)	107.62(47.59)	166.64(66.45)
σ_{33}		212.68(101.31)	264.57(76.73)	300.78(77.97)
<i>Pattern 3:</i>				
Time	29.91(9.08)	29.91(9.08)	29.91(9.08)	29.91(9.08)
Time*base	-0.26(0.08)	-0.26(0.08)	-0.26(0.08)	-0.26(0.08)
Time*treat	0.82(0.95)	0.82(0.95)	0.82(0.95)	0.82(0.95)
Time ²	-6.42(2.23)	-6.42(2.23)	-6.42(2.23)	-6.42(2.23)
Time ² *treat	0.05(0.02)	0.05(0.02)	0.05(0.02)	0.05(0.02)
σ_{11}	206.73(35.86)	206.73(35.86)	206.73(35.86)	206.73(35.86)
σ_{12}	96.97(26.57)	96.97(26.57)	96.97(26.57)	96.97(26.57)
σ_{22}	174.12(31.10)	174.12(31.10)	174.12(31.10)	174.12(31.10)
σ_{13}	87.38(30.66)	87.38(30.66)	87.38(30.66)	87.38(30.66)
σ_{23}	91.66(28.86)	91.66(28.86)	91.66(28.86)	91.66(28.86)
σ_{33}	262.16(44.70)	262.16(44.70)	262.16(44.70)	262.16(44.70)

TABLE 3. – Vorozole Study. Multiple imputation estimates and standard errors for CCMV, NCMV, and ACMV restrictions.

always borrowed from the complete group, *i.e.*, the one with the best prognosis. ACMV, which compromises between all strategies may be more realistic, but here NCMV is likely to be better since information is borrowed from the nearest pattern.

Nevertheless, the NCMV prediction looks more plausible since the worst baseline value shows declining profiles, whereas the best one leaves room for

PATTERN-MIXTURE MODELS

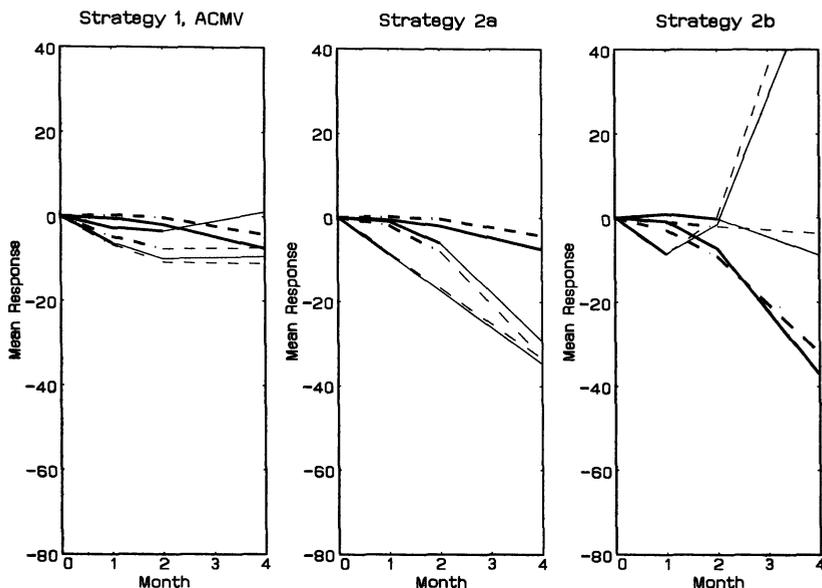


FIG 10. — Vorozole Study. For average level of baseline value 113.57, strategies 1 (ACMV), 2a, and 2b are shown. The bold portion of the curves runs from baseline until the last obtained measurement, and the extrapolated piece is shown in thin type. The dashed line refers to megestrol acetate; the solid line is the Vorozole arm.

improvement. Should one want to explore the effect of assumptions beyond the range of (5), one can allow ω_s to include components outside of the unit interval. In that situation, one has to ensure that the resulting density is still non-negative over its entire support.

Strategy 2. As opposed to identifying restrictions, model simplification can be done in order to identify the parameters. The advantage is that the number of parameters decreases, which is desirable since the length of the parameter vector is a general issue with pattern-mixture models. Indeed, Hogan and Laird (1997) noted that in order to estimate the large number of parameters in general pattern-mixture models, one has to make the awkward requirement that each dropout pattern occurs sufficiently often. Broadly, we distinguish between two types of simplifications.

Strategy 2a. Trends can be restricted to functional forms supported by the information available within a pattern. For example, a linear or quadratic time trend is easily extrapolated beyond the last obtained measurement. One only needs to provide an ad hoc solution for the first or the first few patterns. In order to fit such models, one simply has to carry out a model building exercise within each of the patterns separately.

Strategy 2b. Next, one can let the parameters vary across patterns in a controlled parametric way. Thus, rather than estimating a separate time trend within each pattern, one could for example assume that the time evolution

within a pattern is unstructured, but parallel across patterns. This is effectuated by treating pattern as a covariate. The available data can be used to assess whether such simplifications are supported within the time ranges for which there is information. An initial model is considered with the following effects: time, the interaction between time and treatment, baseline value, pattern, treatment*baseline, treatment*pattern, and baseline*pattern. Further time² is included, as well as its interaction with baseline, treatment, and pattern. No interactions beyond the third order are included, and unstructured covariance matrix is common to all three patterns. This implies that the current model is *not* equivalent to a Strategy 1 model, where all parameters are pattern-specific. To achieve this goal, every effect would have to be made pattern-dependent. A graphical representation is given in Figure 10. Early dropouts decline immediately, whereas those who stay longer in the study first show a rise and then decline thereafter. However, this is less pronounced for higher baseline values. On the other hand, the extrapolation based on the fitted model is very unrealistic, in the sense that for the early dropout sharp rises are predicted, which is totally implausible.

These findings suggest, again, that a more careful reflection on the extrapolation method is required. This is very well possible in a pattern-mixture context, but then the first strategy, rather than the second and third strategies, has to be used.

To test for treatment effect, one can follow two strategies. In the first one, the focus is on the *marginal* treatment effect, *i.e.*, one calculates the marginal treatment effect from the pattern-specific effects. Delta-method arguments complete the procedure. We obtain p values 0.801 (CCMV), 0.900 (NCMV), and 0.828 (ACMV). Alternatively, one can consider a 3 d.f. test, stratified for pattern. The resulting p values are 0.988 (CCMV), 0.995 (NCMV), and 0.993 (ACMV). Note also the closeness between the second strategy and the route taken by Shih and Quan (1997).

6. Analysis of the Alzheimer Study

A linear mixed model (Verbeke and Molenberghs, 2000) was fitted to the outcomes, in which the variance structure was modelled by means of a random subject effect, an exponential serial correlation process and measurement error. The fixed effects considered in the model were, apart from treatment effect, those of age, time, investigator and country, as well as 2- and 3-way interactions. From an initial model selection, only main effects of age, time, time² and treatment group were retained. Note that, when required by the study protocol, it is perfectly possible to leave in such effects as investigator, center, or country, regardless of their significance. Scientific interest is in the effect of treatment. Since there are three arms, we consider two treatment contrasts of the experimental arms versus the standard arm. Our focus here will be on estimates and standard errors for these contrasts, as well as on tests for the null hypothesis of no treatment effect.

PATTERN-MIXTURE MODELS

We first consider the selection model approach. Parameter estimates and standard errors for the treatment contrasts, as well as the associated test results, are reported in Table 4. Treatment effects are non-significant. The likelihood ratio statistic for comparing the missing at random and missing not at random models is 5.4 on 2 degrees of freedom. While this might be taken as some evidence for MNAR, care as to be taken (discussion of Diggle and Kenward 1994, Verbeke and Molenberghs 2000, Ch. 19). More reasonably it can be taken as some evidence against this particular missing at random model.

		Selection Models		Pattern-mixture Models				
Pat.	Cont.	MAR	MNAR	ACMV	CCMV	NCMV	NFD1	NFD2
Stratified Analysis								
1	1			9.27(6.42)	5.84(5.16)	-4.19(6.27)	4.90(8.29)	5.44(6.52)
	2			-8.19(6.58)	-6.92(6.15)	2.56(5.12)	-7.78(7.62)	-4.48(7.76)
2	1			2.78(4.75)	-0.00(2.90)	-4.43(3.54)	0.61(4.88)	-1.49(4.07)
	2			-3.57(4.53)	-5.08(3.92)	-1.37(4.12)	-6.48(5.22)	-4.54(5.46)
3	1			6.78(4.20)	6.95(2.66)	0.10(2.40)	4.18(2.64)	0.18(3.65)
	2			-1.75(2.76)	-3.44(2.12)	0.83(2.14)	-2.66(2.29)	-0.10(2.20)
4	1			11.05(3.21)	10.87(2.85)	6.59(3.09)	9.65(3.56)	9.97(2.90)
	2			-3.84(4.09)	-6.55(3.88)	-3.23(4.09)	-6.84(3.78)	-4.30(4.24)
5	1			0.15(5.71)	-2.05(6.29)	-5.60(6.46)	-3.02(5.92)	-6.13(6.42)
	2			-0.74(3.99)	-0.87(4.51)	0.92(4.68)	-0.53(4.24)	1.05(4.57)
6	1			14.16(3.75)	12.91(3.71)	13.44(3.72)	13.28(3.82)	12.72(3.79)
	2			-5.24(3.48)	-4.74(3.69)	-4.95(3.79)	-4.71(3.63)	-4.77(3.70)
7	1			-0.99(0.85)	-0.99(0.85)	-0.99(0.85)	-0.99(0.85)	-0.99(0.85)
	2			1.68(0.88)	1.68(0.88)	1.68(0.88)	1.68(0.88)	1.68(0.88)
<i>F</i> value				2.45	2.96	1.76	1.92	1.77
<i>p</i> value				0.0024	0.0002	0.0407	0.0225	0.0413
Marginal Analysis								
	1	0.55(0.71)	0.45(0.71)	1.97(1.05)	1.47(0.87)	-0.48(0.85)	1.05(1.04)	0.37(0.96)
	2	0.64(0.71)	0.69(0.71)	-0.24(0.81)	-0.56(0.86)	0.91(0.77)	-0.59(1.01)	0.19(0.84)
<i>F</i> value		2.82	2.68	2.15	1.23	0.66	0.52	0.19
<i>p</i> value		0.2446	0.2619	0.1362	0.3047	0.5208	0.6043	0.8276

TABLE 4. – Alzheimer Study. Inference for treatment contrasts for the case study. For the contrasts, parameter estimates and standard errors, in parentheses, are reported.

Next, we turn attention to the pattern-mixture models. In the pattern-specific initial models, linear mixed models are fitted to the data. The fixed-effects structure comprises treatment indicators, time and its square, and age. The covariance structure is captured by means of a random subject effect, an exponential serial correlation process and measurement error. Then, the five special cases outlined at the end of Section 4 are used to define the identifying restrictions. Admittedly, the set of restrictions considered here is far from exhausting the entire space of possible restrictions and hence leaves room for additional sensitivity analyses. However, it allows us to compare future and non-future dependent mechanisms, as well as a comparison with the selection models fitted earlier.

The results of our analysis are reported in Table 4. The marginal treatment effect assessments are all non-significant, in line with the results from the

PATTERN-MIXTURE MODELS

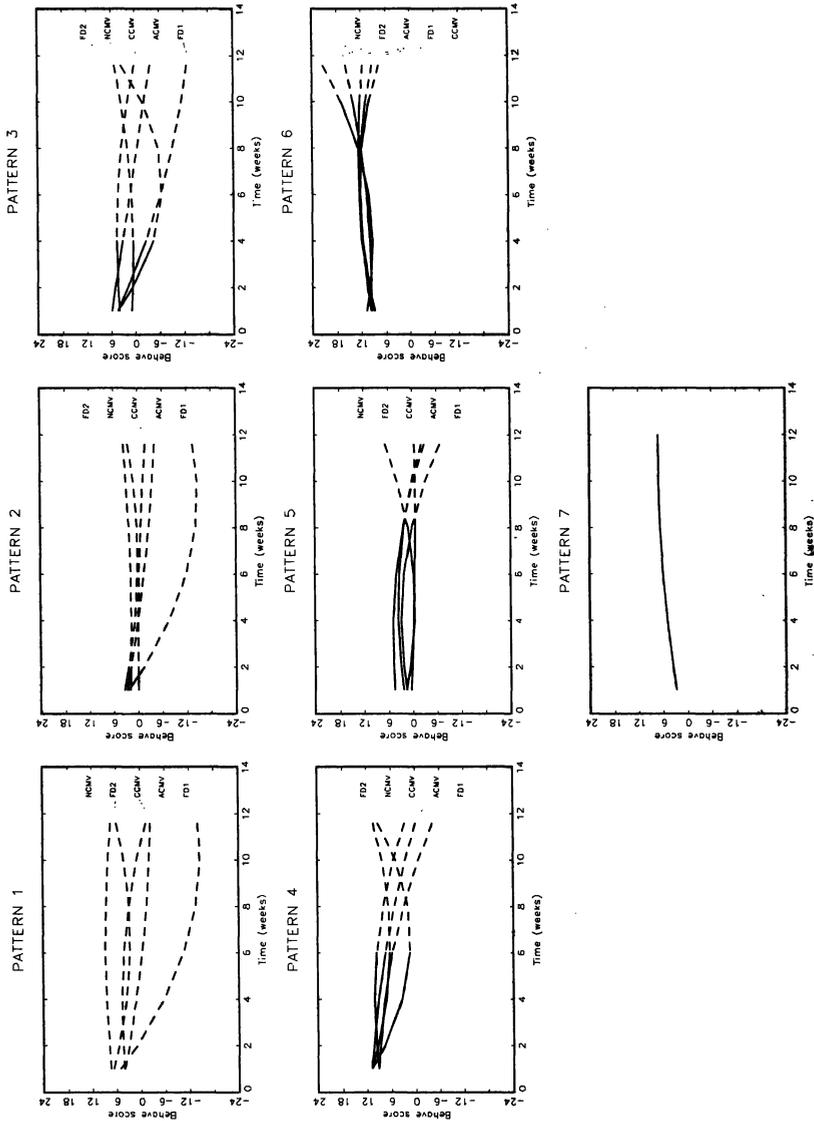


FIG 11. — Alzheimer Study. Pattern-mixture models. Fitted average profiles for each of the five identification strategies. Treatment arm 1. ACMV, available-case missing values; CCMV, complete-case missing values; NCMV, neighbouring-case missing values; FD1, Case 4; FD2, Case 5.

selection model analysis. However, all stratified treatment assessments produce significant p values, although to various levels of strength. Strong evidence is obtained from the available-case missing values model. Of course, the complete-case missing values analysis provides even stronger evidence, but this assumption may be unrealistic, since even patterns with few observations

are completed using the set of completers, corresponding to pattern 7. Both of the other non-future missing values mechanisms, corresponding to Cases 4 and 5, where dropout does not depend on future unobserved values, provide mild evidence for treatment effect. Importantly, we are in a position to consider which patterns are responsible for an identified treatment effect. Note that the contrasts are nowhere near significant in the complete pattern 7, while patterns 4 and 6 seem to contribute to the effect, consistently across patterns. The first contrast of pattern 3 is significant only under complete-case missing values, perhaps explaining why this strategy yields the most significant result. Figure 11 graphically summarises the fit of these models for the first treatment arm; very similar displays for the other arms have been omitted. Clearly, the chosen identifying restrictions have a strong impact, especially for the patterns with earlier dropout. Of course, from Table 1 it is clear that the earlier patterns are rather sparsely filled. It is striking to see that the missing non-future dependence patterns are not all grouped together. An important and perhaps counterintuitive feature is that the fitted averages depend on the identification method chosen, even at time points prior to dropout. The reason for this is that, after imputation, a parametric model is fitted to the completed sequences as a whole, as opposed to, for example, change point models with change point at the time of dropout. Hence, the smoothing induced by the parametric model applies across the entire sequence, before and after dropout.

7. Concluding Remarks

In the vorozole study, we have concentrated on total FLIC (*i.e.*, change of the score versus baseline), a quality of life score measured in a multi-centric two arm study in postmenopausal women suffering from metastatic breast cancer. Since virtually all patients were followed up until disease progression or death, the amount of dropout is large. A very large group of patients drops out after just a couple of months.

While classically only selection models are fitted, pattern-mixture models can be seen as a viable alternative. We analysed the data using both, leading to a sensitivity analysis. More confidence in the results can be gained if both models lead to similar conclusions.

The average profile in the selection model depends on the baseline value, as well as on time. The latter effect is mildly quadratic. There is no evidence for a treatment difference. However, it should be noted that the average profile found is the one that *would* have been observed, had no subjects dropped out, and under the additional assumption that the MAR assumption is correct. Fitting non-random dropout models, in the sense of Diggle and Kenward (1994) is possible, but computationally difficult for a fairly large trial like this one. A separate study of the dropout mechanism revealed that dropout increases with three elements: (1) an unfavourable baseline score, (2) an unfavourable value at the previous month, as well as (3) an unfavourable change in value from the penultimate to the last obtained value.

A pattern-mixture model is fitted by allowing at first a completely separate parameter vector for each observed dropout pattern, which is then simplified by using standard model selection procedures, by considering whether effects are common to all patterns. A first pattern-mixture model features a common treatment effect, of which the assessment is then straightforward. A second model includes a separate treatment effect for each dropout pattern. This leads to two distinct tests. The first one tests for equality of the whole treatment vector to be zero. The second one first calculates the marginal treatment effect from the vector of effects, by composing a weighted sum, where the weights are the multinomially estimated probabilities of the various patterns. In all cases, there is no treatment effect. However, a graphical display of the fitted profiles per pattern is enlightening, since it clearly confirms the trend detected in the selection models, that patients tend to drop out when their quality of life score is declining. Since this feature is usually coupled to an imminent progression or death, it should not come as a surprise. An important advantage of pattern-mixture models is that fitting them is more straightforward than non-random selection models. The additional calculations needed for the marginal treatment effect and its associated precision can be done straightforwardly using the delta method.

Further, we have illustrated three distinct strategies to fit pattern-mixture models. In this way, we have brought together several existing practices. Little (1993, 1994a) has proposed identifying restrictions, which we here formalized using the connection with MAR and multiple imputation. Strategies 2a and 2b refer to fitting a model per pattern and using pattern as a covariate.

By contrasting these strategies on a single set of data, one obtains a range of conclusions rather than a single one, which provides insight into the sensitivity to the assumptions made. Especially with the identifying restrictions, one has to be very explicit about the assumptions and moreover this approach offers the possibility to consider several forms of restrictions. Special attention should go to the ACMV restrictions, since they are the MAR counterpart within the pattern-mixture context.

In addition, a comparison between the selection and pattern-mixture modelling approaches is useful to obtain additional insight into the data and/or to assess sensitivity.

The identifying restrictions strategy provides further opportunity for sensitivity analysis. Indeed, since CCMV and NCMV are extremes for the ω_s vector in (5), it is very natural to consider the idea of *ranges* in the allowable space of ω_s . Clearly, any ω_s which consists of non-negative elements that sum to one is allowable, but also the idea of extrapolation could be useful, where negative components are allowed, given they provide valid conditional densities.

We believe that our approach can play a useful role, as a member of a collection of sensitivity tools. Of course, a sensitivity analysis can be conducted within different frameworks, and there are times where the setting will determine which framework is the more appropriate one (for example Bayesian or frequentist), in conjunction with technical and computational considerations. Draper (1995) has considered ways of dealing with uncertainty in the very

natural Bayesian framework and developments in the missing value setting are ongoing. A thorough comparison between the various frameworks will be interesting and worth undertaking in the future.

Using the Alzheimer study, we have further shown that one can ensure pattern-mixture models to avoid dependence of missingness on future occasions. This widens the range of various models that can be fitted, for example within the context of a sensitivity analysis, even further.

Those interested in implementing the methods presented in this paper, are welcome to contact the first author to obtain SAS software code.

Acknowledgments: We gratefully acknowledge support from FWO-Vlaanderen Research Project “Sensitivity Analysis for Incomplete and Coarse Data” and Belgian IUAP/PAI network “Statistical Techniques and modelling for Complex Substantive Questions with Complex Data”.

References

- ALLISON P.D. (1987). Estimation of linear models with incomplete data. *Sociology Methodology*, 71–103.
- COHEN J. and COHEN P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Erlbaum.
- DIGGLE P.J. and KENWARD M.G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.
- DRAPER D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B*, **57**, 45–97.
- EKHOLM A. and SKINNER C. (1998). The muscatine children’s obesity data reanalysed using pattern mixture models. *Applied Statistics*, **47**, 251–263.
- GLYNN R.J., LAIRD N.M. and RUBIN D.B. (1986). Selection Modelling versus mixture modelling with nonignorable nonresponse. In *Drawing Inferences from Self-Selected Samples*, Ed. H. Wainer, pp. 115–142. New York: Springer Verlag.
- GOSS P.E., WINER E.P., TANNOCK I.F., SCHWARTZ L.H. and KREMER A.B. (1999). A randomized phase III trial comparing the new potent and selective third-generation aromatase inhibitor vorozole with megestrol acetate in postmenopausal advanced breast cancer patients. *Journal of Clinical Oncology*, **17**, 52–63.
- HEDEKER D. and GIBBONS R.D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, **2**, 64–78.
- HOGAN J.W. and LAIRD N.M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, **16**, 239–258.
- KENWARD M.G. and MOLENBERGHS G. (1999). Parametric models for incomplete continuous and categorical longitudinal studies data. *Statistical Methods in Medical Research*, **8**, 51–83.
- KENWARD M.G., MOLENBERGHS G. and THIJS H. (2003). Pattern-mixture models with proper time dependence. *Biometrika*, **90**, 53–71.

PATTERN-MIXTURE MODELS

- LAIRD N.M. (1994). Discussion to Diggle P.J. and Kenward M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 84.
- LITTLE R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, **88**, 125–134.
- LITTLE R.J.A. (1994a). A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471–483.
- LITTLE R.J.A. (1994b). Discussion to Diggle P.J. and Kenward M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 78.
- LITTLE R.J.A. (1995). Modelling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.
- LITTLE R.J.A. and RUBIN D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- LITTLE R.J.A. and WANG Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, **52**, 98–111.
- MCARDLE J.J. and HAMAGAMI F. (1992). modelling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research*, **18**, 145–166.
- MININI P. and CHAVANCE M. (2004a). Observations longitudinales incomplètes : de la modélisation des observations disponibles à l'analyse de sensibilité. *Journal de la Société française de Statistique*, **145**, 2, 5–18.
- MININI P. and CHAVANCE M. (2004b). Sensitivity analysis of longitudinal binary data with non-monotone missing values. *Biostatistics*, **5**, 531–544.
- MOLENBERGHS G., KENWARD M. G. and LESAFFRE E. (1997). The analysis of longitudinal ordinal data with non-random dropout. *Biometrika*, **84**, 33–44.
- MOLENBERGHS G., MICHIELS B. and KENWARD M.G. (1998). Pseudo-likelihood for combined selection and pattern-mixture models for missing data problems. *Biometrical Journal*, **40**, 557–572.
- MOLENBERGHS G., MICHIELS B., KENWARD M.G. and DIGGLE P.J. (1998). Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica*, **52**, 153–161.
- MOLENBERGHS G., MICHIELS B. and LIPSITZ S.R. (1999). Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics*, **55**, 978–983.
- MUTHÉN B., KAPLAN D. and HOLLIS M. (1987). On structural equation modelling with data that are not missing completely at random. *Psychometrika*, **52**, 431–462.
- NELDER J.A. and MEAD R. (1965). A simplex method for function minimisation. *The Computer Journal*, **7**, 303–313.
- REISBERG B., BORENSTEIN J., SALOB S.P., FERRIS S.H., FRANSSSEN E. and GEORGOTAS A. (1987). Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *Journal of Clinical Psychiatry*, **48**, 9–13.
- RUBIN D.B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.
- RUBIN D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons, New York.
- RUBIN D.B. (1994). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 80–82.
- SCHAFFER J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall.

PATTERN-MIXTURE MODELS

- SCHIPPER H., CLINCH J. and MCMURRAY A. (1984). Measuring the quality of life of cancer patients: the Functional-Living Index-Cancer: development and validation. *Journal of Clinical Oncology*, **2**, 472-483.
- SHEINER L.B., BEAL S.L. and DUNNE A. (1997). Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. *Journal of the American Statistical Association*, **92**, 1235-1244.
- SHIH W.J. and QUAN H. (1997). Testing for treatment differences with dropouts present in clinical trials - A composite approach. *Statistics in Medicine*, **16**, 1225-1239.
- THIJS H., MOLENBERGHS G., MICHIELS B., VERBEKE G. and CURRAN D. (2002). Strategies to fit pattern-mixture models. *Biostatistics*, **3**, 245-265.
- VERBEKE G. and MOLENBERGHS G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- WU M.C. and BAILEY K.R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939-55.