Strategies to fit pattern-mixture models

HERBERT THIJS, GEERT MOLENBERGHS*

*Biostatistics, Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590
Diepenbeek, Belgium*
geert.molenberghs@luc.ac.be

BART MICHIELS

Janssen Research Foundation, Beerse, Belgium

GEERT VERBEKE

*Biostatistical Centre, School of Public Health, Katholieke Universiteit Leuven, Capucijnenvoer 35,
B-3000 Leuven, Belgium*

DESMOND CURRAN

EORTC Data Center, Brussels, Belgium

SUMMARY

Whereas most models for incomplete longitudinal data are formulated within the selection model framework, pattern-mixture models have gained considerable interest in recent years (Little, 1993, 1994). In this paper, we outline several strategies to fit pattern-mixture models, including the so-called identifying restrictions strategy. Multiple imputation is used to apply this strategy to realistic settings, such as quality-of-life data from a longitudinal study on metastatic breast cancer patients.

Keywords: Delta method; Linear mixed model; Missing data; Repeated measures; Sensitivity analysis.

1. INTRODUCTION

It is not unusual in practice for some sequences of measurements to terminate early for reasons outside the control of the investigator, and any unit so affected is often called a dropout. It might therefore be necessary to accommodate dropout in the modeling process, either to obtain correct inferences, or because this process itself is of scientific interest.

Rubin (1976) and Little and Rubin (1987, Chapter 6) make important distinctions between different missing values processes. A dropout process is said to be completely random (MCAR) if the dropout is independent of both unobserved and observed data and random (MAR) if, conditional on the observed data, the dropout is independent of the unobserved measurements; otherwise the dropout process is termed non-random (MNAR). If a dropout process is random 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). This leads to considerable simplification in the analysis. Often, the reasons for dropout are varied and it is

*To whom correspondence should be addressed

difficult to justify 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. Reviews are provided in Little (1995) and Kenward and Molenberghs (1999).

Most methods are formulated within the selection modeling framework (Little and Rubin, 1987), as opposed to pattern-mixture modeling (PMM; Little, 1993, 1994). A selection model factors the joint distribution of the measurement and dropout mechanisms into the marginal measurement distribution and the dropout 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 model setting. However, it is often argued, especially in the context of MNAR models, that selection models, although identifiable, should be approached with caution (Glynn *et al.*, 1986). Therefore, pattern-mixture models have gained renewed interest in recent years (Little, 1993, 1994; Hogan and Laird, 1997). Several authors have contrasted selection models and pattern-mixture models. This is done either (1) to answer the same scientific question, such as marginal treatment effect or time evolution, based on these two rather different modeling strategies, or (2) to gain additional insight by supplementing the selection model results with those from a pattern-mixture approach. Examples include Verbeke *et al.* (2001) or Michiels *et al.* (1999) for continuous outcomes, and Molenberghs *et al.* (1999) or Michiels *et al.* (1999) for categorical outcomes. Further references include Ekholm and Skinner (1998); Molenberghs *et al.* (1998); Little and Wang (1996); Hedeker and Gibbons (1997); Cohen and Cohen (1983); Muthén *et al.* (1987); Allison (1987), and McArdle and Hamagami (1992).

An important issue is that pattern-mixture models are by construction under-identified. Little (1993, 1994) 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. While 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 will indicate how this can serve as a starting point for sensitivity analysis. In Section 2 we will introduce the vorozole study, to which our methods will be applied. Section 3 sketches modeling approaches for incomplete data. Sensitivity analysis strategies in a pattern-mixture context is the topic of Section 4, while in Section 5 the strategy of identifying restrictions is considered in detail. In Section 6, this strategy is spelled out for the simple but insightful case of three outcomes; the vorozole study is used to illustrate the ideas. A full analysis of the vorozole data is presented in Section 7.

2. 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 nonmeasurable/nonassessable disease were entered and eligible patients were stratified into three groups according to whether they had measurable, assessable, or nonmeasurable/nonassessable disease. All 452 randomized patients were followed until disease progression or death. 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: Cancer (FLIC; Schipper *et al.*, 1984). A higher FLIC score is the more desirable outcome. Mean changes (and standard deviations) in FLIC score, per time point (up to two years) and treatment arm are given in Table 1.

Table 1. *Vorozole study. Means (*standard deviations*) per time (*up to two years*) and treatment arm for change in FLIC score versus baseline*

Month	Vorozole			Megestrol acetate		
	N	Mean	Standard deviation	N	Mean	Standard deviation
1	198	0.485	14.162	196	−1.622	15.706
2	176	−1.324	16.343	168	−1.268	16.988
4	130	1.031	17.808	136	0.971	16.825
6	94	4.883	17.425	104	1.808	19.038
8	77	7.519	18.506	76	2.737	19.315
10	68	6.309	16.312	60	2.733	16.808
12	58	4.207	21.079	39	2.821	21.738
14	42	3.857	19.806	32	2.219	20.789
16	37	0.189	18.590	22	1.409	18.940
18	26	1.423	25.942	15	2.533	23.086
20	24	0.750	14.405	11	5.909	21.422
22	20	−1.500	15.426	6	4.500	13.248
24	15	1.733	15.068	5	1.400	8.050

Patients underwent screening, and for those deemed eligible a detailed examination at baseline (occasion 0) took place. Further measurement occasions were months 1, then from months 2 at bi-monthly intervals until month 44. Goss *et al.* (1999) analysed FLIC using a two-way ANOVA model with effects for treatment, disease status and their interaction. No significant treatment effect was found. 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 and time evolution, consider effects of dominant site of the disease as well as of clinical stage.

3. PATTERN-MIXTURE MODELS AND SENSITIVITY ANALYSIS

In modeling missing data one is interested in $f(y_i, d_i | \theta, \psi)$, the joint distribution of the measurements Y_i and the dropout indicators D_i defined by adding 1 to the time of the last measurement. One approach is to use selection models based on the factorization $f(y_i, d_i | \theta, \psi) = f(y_i | \theta) f(d_i | y_i, \psi)$. Another is to use the opposite factorization $f(y_i, d_i | \theta, \psi) = f(y_i | d_i, \theta) f(d_i | \psi)$, this being the basis for pattern-mixture models. Molenberghs *et al.* (1998) showed that pattern-mixture models allow for a natural analog of MAR, hence enabling a similar classification of missing data mechanisms.

Sensitivity analysis for pattern-mixture models can be done in many different ways. An important distinction is whether pattern-mixture models are to be contrasted with selection models or to be considered in their own right. In the latter case, it is natural to conduct sensitivity analysis *within* the pattern-mixture family, for which the focus is on handling unidentified components. We will explicitly consider two strategies to deal with under-identification.

- *Strategy* 1. Little (1993, 1994) advocated the use of identifying restrictions and presented a number of examples. We will outline a general framework for identifying restrictions in Section 4, with CCMV (introduced by Little 1993), ACMV, and neighboring case missing value restrictions (NCMV) as important special cases. ACMV is the natural counterpart of MAR in the PMM framework (Molenberghs *et al.*, 1998). This provides a way to compare ignorable selection models with their counterparts in the pattern-mixture setting. Molenberghs *et al.* (1999) and Michiels *et al.* (1999) took

up this idea in the context of binary outcomes, with a marginal global odds ratio model to describe the measurement process (Molenberghs and Lesaffre, 1994).

- *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. Alternatively, 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 implemented 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.

While the second strategy is computationally simple, there is a price to pay. Indeed, simplified models, qualified as 'assumption rich' by Sheiner *et al.* (1997), are also making untestable assumptions, just as in the selection model case. Indeed, using the fitted profiles to predict their evolution beyond the time of dropout is based on extrapolation. For example, it is not possible to assume an unstructured time trend in incomplete patterns, except if one restricts attention to the time range from onset until dropout. In contrast, assuming a linear time trend allows estimation in all patterns containing at least two measurements. However, it is less obvious what the precise nature of the dropout mechanism is. An obvious modeling approach, in particular for normally distributed outcomes, is to specify the dropout mechanism as a polytomous regression. In the identifying restrictions setting on the other hand, the assumptions are clear from the start.

A final observation, applying to both strategies, is that 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. Several authors have followed this idea to formally compare the conclusions from a selection model with the selection model parameters in a pattern-mixture model (Verbeke *et al.*, 2001, Michiels *et al.*, 1999).

4. Identifying restriction strategies

In line with the results obtained by Molenberghs *et al.* (1998), we restrict attention to monotone patterns. In general, we have $t = 1, \ldots, T$ dropout patterns where the dropout indicator is $d = t + 1$. For pattern t, the complete data density is given by

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

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 are particularly appealing. One can base identification on all patterns for which a given component, y_s say, is identified. A general expression for this is

$$f_t(y_s|y_1, \ldots y_{s-1}) = \sum_{j=s}^{T} \omega_{sj} f_j(y_s|y_1, \ldots y_{s-1}), \quad s = t+1, \ldots, T. \tag{2}$$

We will use $\boldsymbol{\omega}_s$ as shorthand for the set of ω_{sj} used. Every non-negative $\boldsymbol{\omega}_s$ which sums to one provides a valid identification scheme.

We incorporate (2) into (1) to give

$$f_t(y_1, \ldots, y_T) = f_t(y_1, \ldots, y_t) \prod_{s=0}^{T-t-1} \left[\sum_{j=T-s}^{T} \omega_{T-s,j} f_j(y_{T-s}|y_1, \ldots, y_{T-s-1}) \right]. \tag{3}$$

Expression (3) clearly shows which information is used to complement the observed data density in pattern t in order to establish the complete data density.

We consider three special but important cases. Little (1993) proposes CCMV which uses the following identification:

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

In other words, information which is unavailable is always borrowed from the completers. This strategy can be defended when the bulk of the subjects are complete and only small proportions are assigned to the various dropout patterns. Also, extension of this approach to non-monotone patterns is particularly easy.

Alternatively, the nearest identified pattern can be used:

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

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

The third special case of (2) will be ACMV (Molenberghs *et al.*, 1998). To derive the corresponding $\boldsymbol{\omega}_s$ vectors, we first re-express (2) as

$$f_t(y_s|y_1, \ldots, y_{s-1}) = f_{(\geqslant s)}(y_s|y_1, \ldots, y_{s-1}), \tag{4}$$

for $s = t+1, \ldots, T$. Here, $f_{(\geqslant s)}(.|.) \equiv f(.|., d > s)$, with d an indicator for time of dropout. Note that d is one more than the length of the observed sequence. Now, we can transform (4) as follows:

$$f_t(y_s|y_1, \ldots, y_{s-1})$$
$$= f_{(\geqslant s)}(y_s|y_1, \ldots, y_{s-1})$$
$$= \frac{\sum_{j=s}^{T} \alpha_j f_j(y_1, \ldots, y_s)}{\sum_{j=s}^{T} \alpha_j f_j(y_1, \ldots, y_{s-1})} \tag{5}$$
$$= \sum_{j=s}^{T} \frac{\alpha_j f_j(y_1, \ldots, y_{s-1})}{\sum_{j=s}^{T} \alpha_j f_j(y_1, \ldots, y_{s-1})} f_j(y_s|y_1, \ldots, y_{s-1}). \tag{6}$$

Next, comparing (6) to (2) yields

$$\omega_{sj} = \frac{\alpha_j f_j(y_1, \ldots, y_{s-1})}{\sum_{\ell=s}^{T} \alpha_\ell f_\ell(y_1, \ldots, y_{s-1})}. \tag{7}$$

We have now derived two equivalent explicit expressions of the MAR case. Expression (5) is the conditional density of a mixture, whereas (2) with (7) is a mixture of conditional densities. Clearly, ω defined by (7) is valid. Restrictions (2), with CCMV, NCMV and ACMV as special cases, can be incorporated in a comprehensive strategy to fit PMMs, which we will now sketch. Thereafter, several points will be clarified.

1. Fit a model to the pattern-specific identifiable densities: $f_t(y_1, \ldots, y_t)$. 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:

$$f_t(y_{t+1}, \ldots, y_T | y_1, \ldots, y_t). \tag{8}$$

4. Using standard multiple imputation methodology (Rubin, 1987; Schafer, 1997; Verbeke and Molenberghs, 2000), draw multiple imputations for the unobserved components, given the observed outcomes and the correct pattern-specific density (8).
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.

We have seen how general identifying restrictions (2), 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 (7) for ACMV, corresponding to items 2 and 3 of the strategy outline. In several cases, the conditional density is a mixture of normal densities. Then, drawing from (2) consists of two steps. First, draw a random uniform variate U to determine which of the $n - s + 1$ components one is going to draw from. Specifically, the kth component is chosen if

$$\sum_{j=s}^{k-1} \omega_{sj} \leqslant U < \sum_{j=s}^{k} \omega_{sj},$$

where $k = s, \ldots, n$. Note that if $k = 1$, the left hand sum is set equal to zero. Next, draw from the kth component.

Note that, except in cases with only a few time points, the number of ω parameters proliferates rapidly. There are several ways to deal with this. First, special but important restrictions such as NCMV, CCMV and ACMV do not suffer from this problem since each of the ω involved is then determined by the choice of restriction. Second, one may set all ω equal to the same constant, chosen from a small set, for example spanning the unit interval. Third, one could put prior distributions on the ω, perhaps governed by simple hyperpriors. The first solution is followed in this paper. The other ones require further exploration.

With Gaussian or categorical outcomes, the conditional distribution of the unobserved outcomes given the observed, is simple to construct. However, for other distributions, it is less straightforward, in which case, the conditional distributions will have to be replaced by the corresponding ratio of marginal distributions. While this changes the algebra, it does not change the methodology fundamentally.

5. A special case: three measurements

In this section, we consider the special but insightful case of three measurements. In Section 5.1, we gain further insight into the general modeling strategy described in the previous section. These methods are then applied to the vorozole study in Section 5.2. The relative simplicity of this method will be employed

to provide further insight into the technical aspects of the proposed procedure. The data analysis of three measurements will also lay the foundations for an efficient *full* analysis of the data over the entire study period, which is the subject of Section 6.

5.1 *Derivations*

In this case, there are only three patterns, and identification (3) takes the following form:

$$f_3(y_1, y_2, y_3) = f_3(y_1, y_2, y_3), \tag{9}$$

$$f_2(y_1, y_2, y_3) = f_2(y_1, y_2) f_3(y_3|y_1, y_2), \tag{10}$$

$$f_1(y_1, y_2, y_3) = f_1(y_1) [\omega f_2(y_2|y_1) + (1 - \omega) f_3(y_2|y_1)] f_3(y_3|y_1, y_2). \tag{11}$$

Since $f_3(y_1, y_2, y_3)$ is completely identifiable from the data, and for $f_2(y_1, y_2, y_3)$ there is only one possible identification, given (2), the only place where a choice has to be made is pattern 1. Setting $\omega = 1$ corresponds to NCMV, while $\omega = 0$ implies CCMV. Using (7) in this particular case, ACMV corresponds to

$$\omega = \frac{\alpha_2 f_2(y_1)}{\alpha_2 f_2(y_1) + \alpha_3 f_3(y_1)}. \tag{12}$$

The conditional density $f_1(y_2|y_1)$ in (11) can be rewritten as

$$f_1(y_2|y_1) = \frac{\alpha_2 f_2(y_1, y_2) + \alpha_3 f_3(y_1, y_2)}{\alpha_2 f_2(y_1) + \alpha_3 f_3(y_1)}.$$

Next, let us illustrate how draws from the conditional densities can be conducted. To this end, we consider identification scheme (9)–(11), initially avoiding the specification of a parametric form for these densities. The following steps are required:

1. Estimate the parameters of the identifiable densities: $f_3(y_1, y_2, y_3)$, $f_2(y_1, y_2)$, and $f_1(y_1)$. Then, for each of the *m* imputations we have to execute the following steps.
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 between three substeps.

    (a) The proportions ω need to be chosen or determined. Every ω in the unit interval is valid. Specific cases are: NCMV ($\omega = 1$); CCMV ($\omega = 0$). For ACMV, ω is calculated from (12). Note that, given y_1, this is a constant, depending on α_2 and α_3.
    In order 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. Then continue with (b) and (c).
    (b) If $U \leqslant \omega$, calculate $f_2(y_2|y_1)$ and draw from it. Otherwise, do the same based on $f_3(y_2|y_1)$.
    (c) Given the observed y_1 and the y_2 which has just been drawn, calculate the conditional density $f_3(y_3|y_1, y_2)$ and draw from it.

Multiple imputation implies repeating steps 2–4 *M* times.

We now expand on steps 1 and 2 and assume that the observed densities are estimated using linear mixed models. Then, $f_3(y_1, y_2, y_3)$, $f_2(y_1, y_2)$, and $f_1(y_1)$ produce fixed-effect and variance parameters.

Group all of them in γ and assume a draw is made from their distribution, γ^* say. To this end, their precision estimates need to be computed. These are easily obtained in most standard software packages, such as SAS, rendering this step a very straightforward one.

We illustrate this for (10). Generally, we would opt for the linear mixed model (Laird and Ware, 1982; Diggle, 1988; Diggle *et al.*, 1994; Verbeke and Molenberghs, 1997, 2000). Writing β for the p-dimensional vector containing the fixed effects, and $\varepsilon_i \sim N(\mathbf{0}, \Sigma)$ as the vector of correlated error terms, the linear model is

$$Y_i = X_i\beta + \varepsilon_i.$$

In the pattern-mixture case, the parameters involved in this model will be allowed to depend on dropout pattern. Now, assume that the ith subject has only two measurements, and hence belongs to the second pattern. Let its design matrices be X_i and Z_i for the fixed effects and random effects, respectively. Its mean and variance for the *third* pattern are

$$\boldsymbol{\mu}_i(3) = X_i\beta^*(3), \tag{13}$$

$$V_i(3) = Z_i D^*(3)Z_i' + \Sigma_i(3), \tag{14}$$

where (3) indicates that the parameters are specific to the third pattern.

Now, based on (13), (14) and the observed values $y_i = (y_{i1}, y_{i2})'$, the parameters for the conditional density follow immediately:

$$\boldsymbol{\mu}_{i,2|1}(3) = \boldsymbol{\mu}_{i,2}(3) + V_{i,21}(3)[V_{i,11}(3)]^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{i,2}(3)),$$

$$V_{i,2|1}(3) = V_{i,22}(3) - V_{i,21}(3)[V_{i,11}(3)]^{-1}V_{i,12}(3),$$

where a subscript 1 indicates the first two components and a subscript 2 refers to the third component. Draws from every other conditional density can be obtained in essentially the same way.

5.2   *Application to the vorozole study*

To provide a specific illustration, we now apply the proposed methodology to those subjects with 1–3 follow-up measurements, respectively. 190 subjects are included, with subsample sizes 35, 86, and 69, respectively. The pattern probabilities are

$$\widehat{\boldsymbol{\pi}} = (0.184, 0.453, 0.363)', \tag{15}$$

and asymptotic covariance matrix

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}) = \begin{pmatrix} 0.000\,791 & -0.000\,439 & -0.000\,352 \\ -0.000\,439 & 0.001\,304 & -0.000\,865 \\ -0.000\,352 & -0.000\,865 & 0.001\,217 \end{pmatrix}. \tag{16}$$

These figures, apart from giving a feel for the relative importance of the various patterns, are needed to calculate the marginal treatment effect and to test for its importance, which was the primary goal of the analysis.

It is of interest to study the treatment-arm-specific pattern probabilities as well. For the vorozole arm, the subsample sizes are 18, 48, and 36, producing probabilities $\widehat{\boldsymbol{\pi}}_v = (0.177, 0.471, 0.354)'$ with asymptotic covariance matrix

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}_v) = \begin{pmatrix} 0.001\,425 & -0.000\,814 & -0.000\,611 \\ -0.000\,814 & 0.002\,442 & -0.001\,628 \\ -0.000\,611 & -0.001\,628 & 0.002\,239 \end{pmatrix}.$$

For the megestrol acetate arm, the subsample sizes are 17, 38, and 33, giving probabilities $\widehat{\pi}_m = (0.193, 0.432, 0.375)'$ and asymptotic covariance matrix

$$\widehat{\mathrm{Var}}(\widehat{\pi}_m) = \begin{pmatrix} 0.001\,771 & -0.000\,948 & -0.000\,823 \\ -0.000\,948 & 0.002\,788 & -0.001\,840 \\ -0.000\,823 & -0.001\,840 & 0.002\,663 \end{pmatrix}.$$

The treatment-arm-specific probabilities are not significantly different from each other. A classical χ^2 test produces $p = 0.864$. Hence, we will employ expressions (15) and (16).

5.2.1 *Model fitting.* 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 more easily 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 probability of death, then identifying restrictions (i.e. extrapolation beyond the time of death) is meaningful. If one does not want to extrapolate beyond the moment of death, one can restrict modeling to the observed data only. The former viewpoint refers to strategy 1, while the latter refers to strategy 2. 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 modeling solution no matter which point of view is adopted. This contrasts with selection models where extrapolation always occurs, at least implicitly.

For strategy 1, we start with fitting a model to the observed data, including linear and quadratic time effects, as well as their interactions with baseline value. Further, time by treatment interaction is included, for consistency with the original analysis plan. 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 2, in the 'initial' column. Obviously, not all effects are estimable in this initial model.

We summarize this model 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 minimum, average, and maximum values (Figure 1). 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 2. In all of the plots (Figures 2–4), the same mean response scale as in Figure 1 was retained, illustrating that the identifying restriction strategies extrapolate much more closely to the observed data mean responses. There are some differences among the identifying restriction methods. Roughly speaking, CCMV extrapolates rather towards a rise whereas NCMV seems to predict more of a decline, at least for baseline value 53. Further, ACMV indicates rather a steady state. For the other baseline levels, a status quo or a mild increase is predicted. 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. Hence, it is likely that the dropout mechanism is not CCMV, since this strategy always refers to the best 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 improvement. Should one want to explore the effect of assumptions beyond the range of (2), one can allow ω_s to include components outside the unit

Table 2. *Vorozole study. Multiple imputation estimates and standard errors for CCMV, NCMV, and ACMV restrictions*

Effect	Initial	CCMV	NCMV	ACMV
Pattern 1				
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$^2 *$ base		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)
Pattern 2				
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$^2 *$ base	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)
Pattern 3				
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$^2 *$ base	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)

interval. In this situation, one has to ensure that the resulting density is still non-negative over its entire support.

In strategy 2b, *pattern* is included as a covariate. 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, a quadratic in time is included, as well as its interaction
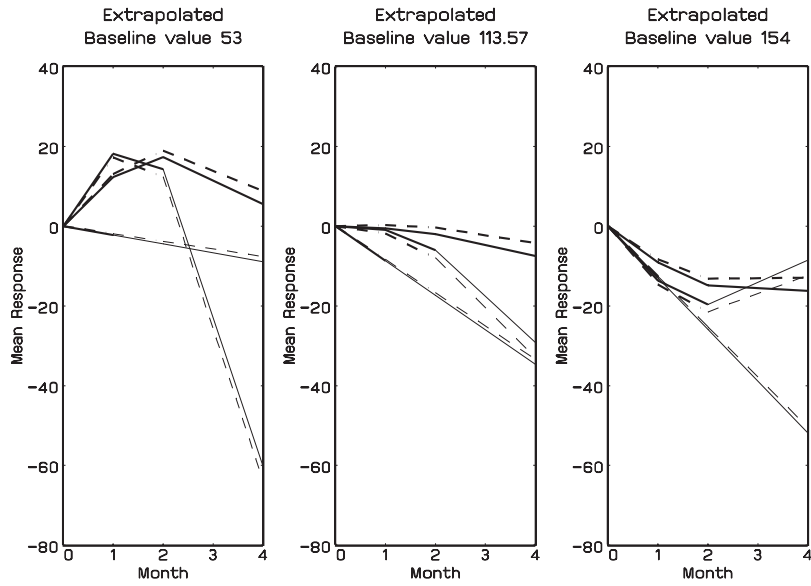
Fig. 1. Vorozole study. Extrapolation based on model fitted to observed data (strategy 2a). For three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented. The bold portion of each curve runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type. The dashed line refers to megestrol acetate, the solid line is the vorozole arm.
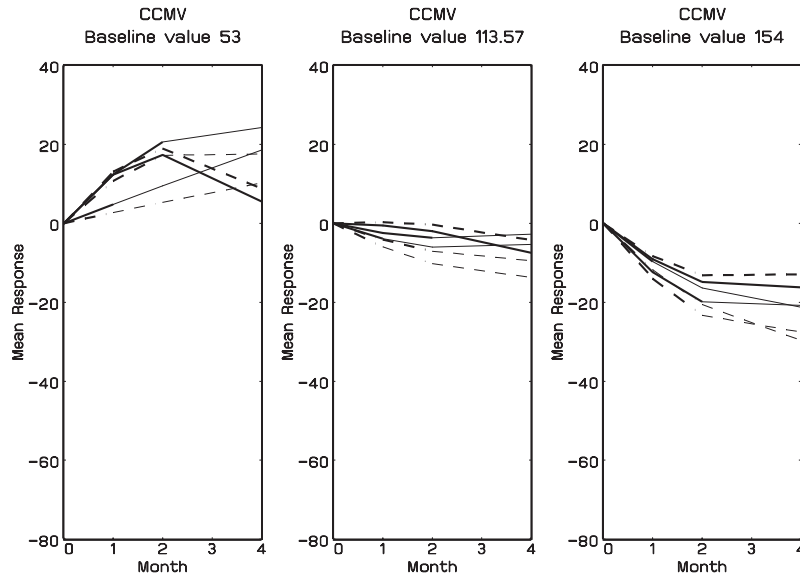

Fig. 2. Vorozole study. Complete case missing value restrictions analysis. For three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented. The bold portion of each curve runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type. The dashed line refers to megestrol acetate, the solid line is the vorozole arm.
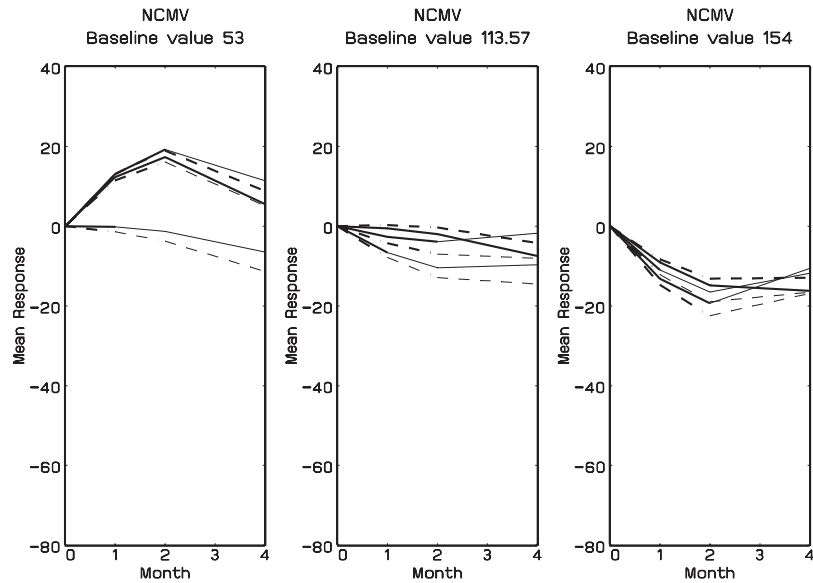
Fig. 3. Vorozole study. Neighboring case missing value restrictions analysis. For three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented. The bold portion of each curve runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type. The dashed line refers to megestrol acetate, the solid line is the vorozole arm.
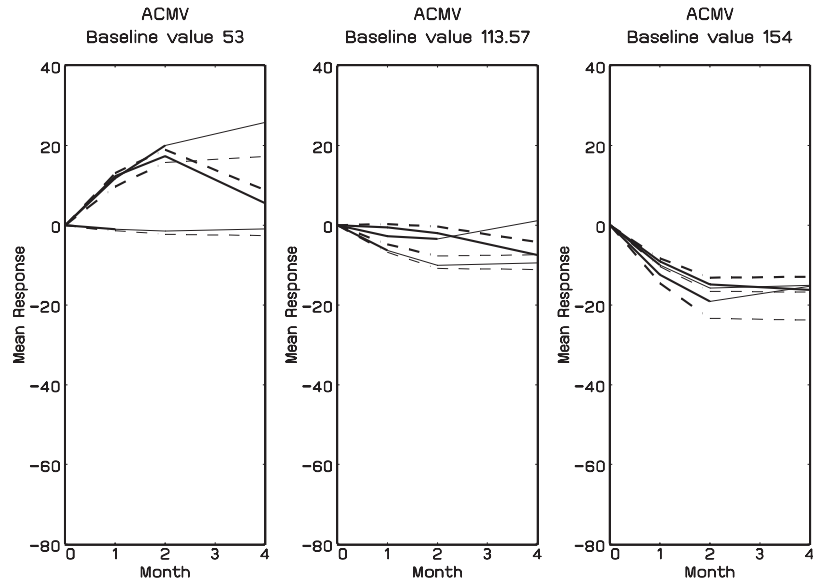

Fig. 4. Vorozole study. Available case missing value restrictions analysis. For three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented. The bold portion of each curve runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type. The dashed line refers to megestrol acetate, the solid line is the vorozole arm.

Table 3. *Vorozole study. Strategy 2b*

Effect	Pattern	Estimate (s.e.)
Time	1	7.29(15.69)
Time	2	37.05(7.67)
Time	3	39.40(9.97)
Time $*$ treat	1	5.25(6.41)
Time $*$ treat	2	3.48(5.46)
Time $*$ treat	3	3.44(6.04)
Time $*$ base	1	$-0.21(0.15)$
Time $*$ base	2	$-0.34(0.06)$
Time $*$ base	3	$-0.36(0.08)$
Time $*$ treat $*$ base		$-0.06(0.04)$
Time^2	1	*−9.18(2.47)*
Time^2	2	$-9.18(2.47)$
Time^2	3	$-7.70(2.29)$
Time$^2 *$ treat		1.10(0.74)
Time$^2 *$ base		0.07(0.02)
σ_{11}		173.63(18.01)
σ_{12}		117.88(17.80)
σ_{22}		233.86(26.61)
σ_{13}		89.59(24.56)
σ_{23}		116.12(34.27)
σ_{33}		273.98(48.15)

with baseline, treatment, and pattern. No interactions beyond the third order are included, and an unstructured covariance matrix is assumed 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. The estimated model parameters are presented in Table 3 and the graphical representation is given in Figure 5. Early dropouts decline immediately, whereas those who stay longer in the study first show a rise and then a decline. However, this is less pronounced for higher baseline values. On the other hand, the extrapolation based on the fitted model is very unrealistic, rendering this deceptively simple approach a bad choice. These findings reinforce the need for careful reflection on the extrapolation method.

5.2.2 *Hypothesis testing.* We focus on treatment effect. A by-product of the pattern-mixture strategy is that a pattern-specific treatment effect can result. This is the case for all five models in Tables 2 and 3. Since here the main scientific interest is on the estimation of the *marginal treatment effect,* we can combine the pattern-specific effects into a pattern-averaged effect, as outlined in the appendix. We apply this approach to the models in Tables 2 and 3. All three pattern-mixture strategies will be considered. Since the identifying restriction strategies are slightly more complicated than the others, we will consider the other strategies first.

For strategy 2a, recall that the parameters are presented in Table 2 as the initial model. The treatment effect vector is $\beta = (0.33, -0.95, 0.82)'$ with $V = \text{diag}(15.28, 3.44, 0.90)$, since the paaterns are analyzed separately. This leads to the test statistic $\beta' V^{-1} \beta = 1.02$ on three degrees of freedom, producing $p = 0.796$.
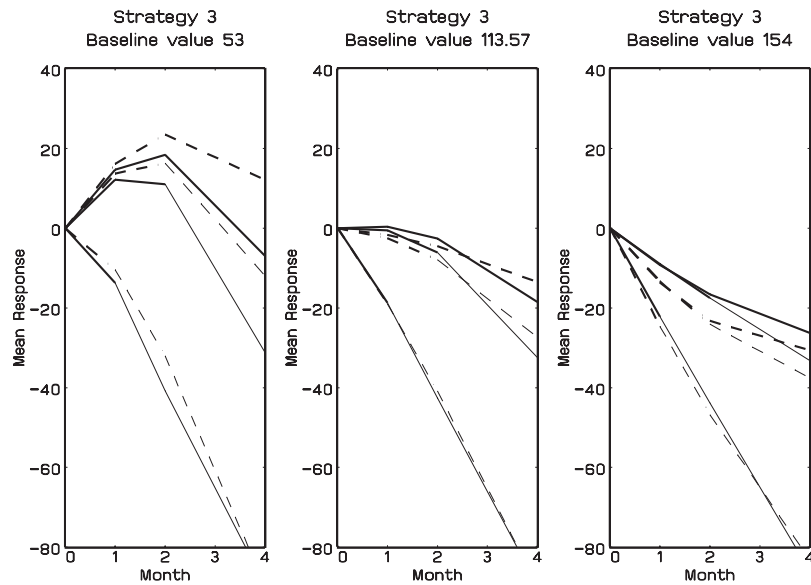
Fig. 5. Vorozole study. Models with *pattern* used as a covariate (strategy 2b). For three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented. The bold portion of each curve runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type. The dashed line refers to megestrol acetate, the solid line is the vorozole arm.

To calculate the marginal treatment effect, we apply (22)–(25). The marginal effect is estimated as $\hat{\beta}_0 = -0.07$ (s.e. 1.16). The corresponding asymptotic p-value is 0.95. Both approaches agree on the non-significance of the treatment effect.

For strategy 2b, the parameters are presented in Table 3. The treatment effect vector is $\beta = (5.25, 348, 3.44)'$ with non-diagonal covariance matrix

$$V = \begin{pmatrix} 41.12 & 23.59 & 25.48 \\ 23.59 & 29.49 & 30.17 \\ 25.48 & 30.17 & 36.43 \end{pmatrix}.$$

The correlation between them is quite substantial. The reason is that some parameters, in particular the other treatment effects (three-way interaction with baseline and time, interaction with quadratic in time), are common to all three patterns, hence inducing dependence across patterns. This leads to the test statistic $\beta' V^{-1} \beta = 0.70$ on three degrees of freedom, producing $p = 0.874$. Calculating the marginalized treatment effect, we obtain $\hat{\beta}_0 = 3.79$ (s.e. 5.44). The corresponding asymptotic p-value is 0.49. The different numerical value of the treatment effect, as compared to those obtained with the other strategies, is entirely due to the presence of a quadratic treatment effect which, for ease of exposition, is left out of the picture in testing here. It is straightforward to add this parameter to the contrast(s) being considered, should one want to do so.

For strategy 1, we will consider several approximate methods of inference. The CCMV case will be discussed in detail. The two other restriction types are entirely similar. There are three treatment effects, one for each pattern. Hence, multiple imputation produces a vector of treatment effects and the within,

Table 4. *Vorozole study. Tests of treatment effect for CCMV, NCMV, and ACMV restrictions*

Parameter	CCMV	NCMV	ACMV
Stratified analysis			
k	3	3	3
τ	12	12	12
denominator d.f. w	28.41	17.28	28.06
r	1.12	2.89	1.14
F statistic	0.044	0.022	0.030
p-value	0.988	0.995	0.993
Marginal analysis			
Marginal effect (s.e.)	$-0.85(0.77)$	$-0.63(1.22)$	$-0.74(0.85)$
k	1	1	1
τ	4	4	4
denominator d.f. w	4	4	4
r	1.49	4.57	1.53
F statistic	0.072	0.018	0.054
p-value	0.801	0.900	0.828

between and total covariance matrices:

$$\beta_{CC} = \begin{pmatrix} -2.09, & -1.68, & 0.82 \end{pmatrix}', \tag{17}$$

$$W_{CC} = \begin{pmatrix} 1.67 & 0.00 & 0.00 \\ 0.00 & 0.59 & 0.00 \\ 0.00 & 0.00 & 0.90 \end{pmatrix}, \tag{18}$$

$$B_{CC} = \begin{pmatrix} 2.62 & 0.85 & 0.00 \\ 0.85 & 0.72 & 0.00 \\ 0.00 & 0.00 & 0.00 \end{pmatrix}, \tag{19}$$

and

$$T_{CC} = \begin{pmatrix} 4.80 & 1.02 & 0.00 \\ 1.02 & 1.46 & 0.00 \\ 0.00 & 0.00 & 0.90 \end{pmatrix}. \tag{20}$$

In a stratified analysis, we want to test the hypothesis $H_0 : \beta = \mathbf{0}$. Using (17)–(19), we can apply multiple imputation methodology. Note that even though the analysis is done per pattern, the between and total matrices have non-zero off-diagonal elements. This is because the imputation is based on information from *other* patterns, hence introducing inter-pattern dependence. Results are presented in Table 4. All p-values are non-significant, in line with earlier evidence from strategies 1 and 2, although a bit more extreme.

For the marginal parameter, the situation is more complicated here than with strategies 1 and 2. Indeed, classical theory often assumes inference is geared towards the original vector, or linear contrasts thereof. Formula (21) represents a non-linear transformation of the parameter vector and therefore needs further development. First, consider π to be part of the parameter vector. Since there is no missingness involved

Table 5. *Vorozole study. Strategy 2b.*
Reduced model.

Effect	Pattern	Estimate (s.e.)
Time		33.06(6.67)
Time $*$ treat		0.40(0.84)
Time $*$ base		$-0.29(0.06)$
Time ^{2}	1	$-16.71(3.46)$
Time ^{2}	2	$-8.56(1.90)$
Time ^{2}	3	$-7.09(1.78)$
Time $^2 *$ base		0.06(0.01)
σ_{11}		178.02(18.46)
σ_{12}		121.75(18.30)
σ_{22}		238.31(26.98)
σ_{13}		88.75(24.94)
σ_{23}		121.10(34.70)
σ_{33}		274.58(48.32)

in this part, it contributes to the within matrix, but not to the between matrix. Then, using (22), the approximate within matrix for the marginal treatment effect is

$$W_0 = \pi' W \pi + \beta' \text{Var}(\pi) \beta,$$

with, for the between matrix, simply

$$B_0 = \pi' B \pi.$$

The results are presented in the second panel of Table 4. All three *p*-values are in between those obtained for strategies 2 and 3. Of course, all five agree on the non-significance of the treatment effect. The reason for the differences is to be found in the way the treatment effect is extrapolated beyond the period of observation. Indeed, the highest *p*-value is obtained for strategy 2a and, from Figure 1, we learn that virtually no separation between both treatment arms is projected. On the other hand, wider separations are seen in Figure 5. Finally, we note that all conclusions are conditional upon the unverifiable assumption that the posited restrictions (and hence, dropout mechanisms) are correct. This underscores the need for sensitivity analysis.

6. ANALYSIS OF THE VOROZOLE STUDY

From Section 5 it is clear that the apparently simpler strategies 2a and b are to be avoided. In the current analysis, using all available data, we therefore restrict attention to strategy 1.

The distribution of subjects over patterns is described in Table 6. The largest portion of the subjects remains on study for less than one year, with very few in the subsequent six-month period and only three subjects on trial after 18 months. Based on this information, we chose to remove the three subjects in the last patterns.

The selection of covariates was done by analogy with Section 5.2.1. Covariates kept in the models are time, time interactions with baseline and treatment group, time^2 and the interaction between time^2 and baseline. As before, we opted for NCMV, CCMV and ACMV identifying restrictions.

Assessment of the marginal treatment effect leads to $p = 0.9407$ for NCMV, $p = 0.7570$ for CCMV, and $p = 0.0487$ for ACMV. This needs careful discussion. First, there is a considerable difference with the assessment based on just three measurements, where all three strategies yielded non-significant results.

Table 6. *Vorozole study. Distribution over patterns*

Pattern	Number of measurements	Number of subjects	Proportion	Percentage
3	1	35	0.0873	8.73
4	2	86	0.214	21.45
5	3	69	0.172	17.21
6	4	45	0.112	11.22
7	5	29	0.072	7.23
8	6	33	0.082	8.23
9	7	22	0.055	5.49
10	8	17	0.042	4.24
11	9	19	0.047	4.74
12	10	9	0.022	2.24
13	11	10	0.025	2.49
14	12	6	0.015	1.50
15	13	8	0.020	2.00
16	14	3	0.007	0.75
17	15	4	0.010	1.00
18	16	3	0.007	0.75
20	18	1	0.002	0.25
21	19	1	0.002	0.25
25	23	1	0.002	0.25

Second, the strategies differ dramatically from each other, which is in disagreement with the earlier analysis. The MAR-based ACMV restrictions yield a significant effect, albeit borderline, whereas the others do not lead to evidence in favour of a treatment difference. This suggests that the treatment effect found is conditional on the ACMV (MAR) assumption.

A graphical representation is given in Figure 6. It is, of course, rather difficult to get a clear message from the multitude of patterns, except that for ACMV, as opposed to the other two, the extrapolated profiles lie closer to the observed profiles.

7. CONCLUDING REMARKS

In this paper, we have illustrated three distinct strategies to fit pattern-mixture models. In this way, we have brought together several existing practices. Little (1993, 1994) has proposed identifying restrictions, which we here formalized as strategy 1 using the connection with MAR and multiple imputation. Strategies 2 and 3 refer to fitting a model per pattern and using pattern as a covariate, respectively.

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. Moreover, this approach offers the possibility of considering several forms of restrictions. Special attention should be given 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 modeling approaches is useful to obtain additional insight into the data and/or to assess sensitivity.

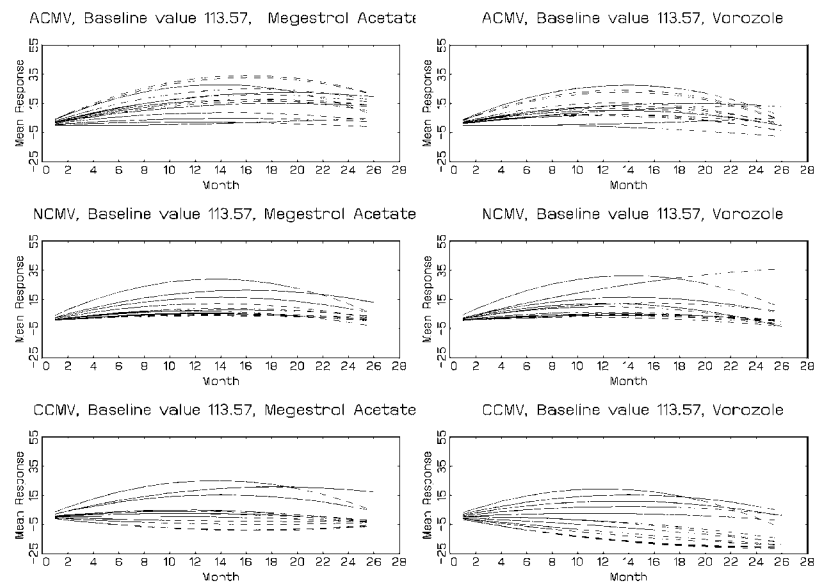The identifying restrictions strategy provides further opportunity for sensitivity analysis. Indeed, since

Fig. 6. Vorozole study. Models using all patterns. Strategy 1. For the average level of baseline value and for both treatment arms, plots of mean profiles over time are presented. The solid portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in dashed type.

CCMV and NCMV are extremes for the ω_s vector in (2), 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, provided that tehy correspond to valid conditional densities.

We believe that our approach can play a useful role, as one 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.

SAS and GAUSS codes, for both the three-pattern and general versions, are available from the authors' web pages.

APPENDIX: ESTIMATING MARGINAL EFFECTS

Let $\beta_{\ell d}$ represent the treatment-effect parameter estimates $\ell = 1, \ldots, g$ (assuming there are g groups) in pattern $d = 1, \ldots, T$ and let π_d be the proportion of subjects in pattern d. Then, the estimates of the marginal treatment effects β_ℓ are

$$\beta_\ell = \sum_{d=1}^{T} \beta_{\ell d} \pi_d, \qquad \ell = 1, \ldots, g. \tag{21}$$

The variance is obtained using the delta method. Precisely, it assumes the form

$$\text{Var}(\beta_1, \ldots, \beta_g) = A V A', \tag{22}$$

where

$$V = \left(\begin{array}{c|c} \text{Var}(\beta_{\ell d}) & 0 \\ \hline 0 & \text{Var}(\pi_d) \end{array} \right) \tag{23}$$

and

$$A = \frac{\partial(\beta_1, \ldots, \beta_g)}{\partial(\beta_{11}, \ldots, \beta_{Tg}, \pi_1, \ldots, \pi_T)}. \tag{24}$$

The estimate of the variance–covariance matrix of the $\hat{\beta}_{\ell d}$ is obtained from statistical software. The multinomial quantities are easy to obtain from the pattern-specific sample sizes. In the case of the Vorozole data, these quantities are presented in (15) and (16). A Wald test statistic for the null hypothesis $H_0 : \beta_1 = \cdots = \beta_g = 0$ is then given by

$$\beta_0' A V A' \beta_0, \tag{25}$$

where $\beta_0 = (\beta_1, \ldots, \beta_g)'$.

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 edn. Hillsdale, NJ: Erlbaum.

DIGGLE, P. J. (1988). An approach to the analysis of repeated measures. *Biometrics* **44**, 959–971.

DIGGLE, P. J., LIANG, K.-Y. AND ZEGER, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Oxford Science, Clarendon.

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 modeling with nonignorable nonresponse. In Wainer, H. (ed.), *Drawing Inferences from Self-Selected Samples*, New York: Springer, pp. 115–142.

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.

LAIRD, N. M. AND WARE, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.

LI, K. H., RAGHUNATHAN, T. E. AND RUBIN, D. B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an *F* reference distributions. *Journal of the American Statistical Association* **86**, 1065–1073.

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

LITTLE, R. J. A. (1995). Modeling the drop-out 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). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research* **18**, 145–166.

MICHIELS, B., MOLENBERGHS, G. AND LIPSITZ, S. R. (1999). A pattern-mixture adds ratio model for incomplete categorical data. *Communication in Statistics: Theory and Methods* **28**, 2893–2870.

MOLENBERGHS, G. AND LESAFFRE, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* **89**, 633–644.

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 modeling with data that are not missing completely at random. *Psychometrika* **52**, 431–462.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

RUBIN, D. B. (1978). Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse. *Imputation and Editing of Faulty or Missing Survey Data*, US Department of Commerce, pp. 1–23.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

RUBIN, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**, 473–489.

RUBIN, D. B. AND SCHENKER, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association* **81**, 366–374.

SCHAFER, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

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.

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.

TANNER, M. A. AND WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–550.

VERBEKE, G. AND MOLENBERGHS, G. (1997). *Linear Mixed Models in Practice: A SAS Oriented Approach*, Lecture Notes in Statistics 126. New York: Springer.

VERBEKE, G. AND MOLENBERGHS, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

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.