STRATEGIES TO FIT PATTERN-MIXTURE MODELS

Geert Molenberghs[1], Herbert Thijs[1], Bart Michiels[2], Geert Verbeke[3], Desmond Curran[4] *
[1] *Limburgs Universitair Centrum, Belgium,* [2] *Janssen Research Foundation, Belgium,*
[3] *Katholieke Universiteit Leuven, Belgium,* [4] *Icon Clinical Research, Ireland*

Whereas most models for incomplete longitudinal data are formulated within the selection model framework, pattern-mixture models have gained considerable interest in recent years. In this chapter, 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.

Key words: Delta Method; Linear Mixed Model; Missing Data; Repeated Measures.

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 inference, or since this process can itself be of scientific interest. In this paper, we will restrict attention to dropout, i.e., monotone missingness.

Rubin (1976) and Little and Rubin (1987, Ch. 6) make important distinctions between different missing values processes. A dropout process is said to be completely random

*Biostatistics, Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, B–3590 Diepenbeek, Belgium. Email: geert.molenberghs@luc.ac.be
We gratefully acknowledge support from *Fonds Wetenschappelijk Onderzoek-Vlaanderen* Research Project G.0002.98 "Sensitivity Analysis for Incomplete and Coarse Data", and from *Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in Industrie.*

(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. 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).

Most methods are formulated within the selection modeling frame (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 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, 1994, Hogan and Laird 1997). Examples can be found in Verbeke, Lesaffre, and Spiessens (1998), Curran *et al* (1999), 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, 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 as well.

All in all, 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 in this paper how it 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. Finally, in Section 6 we discuss the results of all methods applied to the vorozole set of data.

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). Precisely, a higher FLIC score is the more desirable outcome.

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) analyzed FLIC using a two-way ANOVA model with effects for treatment, disease status, as well as their interaction. No significant difference was found. The main conclusion from the primary analysis was that vorozole is well tolerated and as effective as megestrol aceetate 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 of the disease as well as clinical stage.

3 MODELS FOR INCOMPLETE LONGITUDINAL DATA

In modeling missing data one is interested in $f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi})$ which is the joint distrubution of the measurements Y_i and the dropout indicators D_i defined by adding 1 to the time of the last measurement. A first and most popular approach is by using selection models based on the factorization $f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|\boldsymbol{\theta})f(d_i|\boldsymbol{y}_i, \boldsymbol{\psi})$. In this framework standard missing data concepts such as MCAR, MAR, MNAR (Rubin 1976, Little and Rubin 1987) can be constructed. Recently more interest is put in the opposite factorization $f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|d_i, \boldsymbol{\theta})f(d_i|\boldsymbol{\psi})$ being the basis for pattern-mixture models. Molenberghs, Michiels, Kenward, and Diggle (1998) showed that pattern-mixture models allow for a natural analog of MAR, hence enabling a similar classification of missing data mechanisms. In this paper we will use a popular model for repeated measurements, incorporating random effects (Laird and Ware 1982) and serial correlation (Diggle 1988). Assuming $\boldsymbol{\beta}$ as the p dimensional vector containing the fixed effects, and $\boldsymbol{\varepsilon_i} \sim N(\mathbf{0}, \Sigma)$ as the vector of correlated error terms, we can write this model as follows:

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i. \tag{1}$$

For a detailed description we refer to Diggle, Liang, and Zeger (1994), Verbeke and Molenberghs (1997, 2000). In the pattern-mixture case, the parameters involved in this model will be allowed to depend on dropout pattern.

4 PATTERN-MIXTURE MODELS

Sensitivity analysis for pattern-mixture models can be conceived in many different ways. Crucial aspects are whether pattern-mixture and selection modeling are to be contrasted with one another or rather the pattern-mixture modeling is the central focus of interest.

In the latter case, it is natural to conduct sensitivity analysis *within* the pattern-mixture family. The key area where sensitivity analysis should be focused is on the unidentified components of the model and the way(s) in which this is handled. We will explicitly consider three strategies to deal with under-identification.

Strategy 1. Little (1993, 1994) advocated the use of identifiying restrictions and presented a number of examples. We will outline a general framework for identifying restrictions in Section 5, with CCMV (introduced by Little 1993), ACMV, and neighboring case

missing value restrictions (NCMV) as important special cases. Recall that ACMV is the natural counterpart of MAR in the PMM framework. This provides a way to compare ignorable selection models with their counterpart in the pattern-mixture setting. Molenberghs, Michiels, and Lipsitz (1999) and Michiels, Molenberghs, Lipsitz (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. 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 asssume 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.

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), are also making untestable assumptions, just as in the selection model case. 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, Lesaffre, and Spiessens 1998, Curran, Pignatti, and Molenberghs 1998, Michiels *et al* 1999).

5 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 $t = 1, \ldots, T$ dropout patterns where the dropout indicator is $d = t + 1$. For pattern t, the complete data density is given by

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

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

$$f_t(y_s | y_1, \ldots y_{s-1}) = \sum_{j=s}^{T} \omega_{sj} f_j(y_s | y_1, \ldots y_{s-1}), \quad s = t+1, \ldots, T. \tag{3}$$

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

Let us incorporate (3) into (2):

$$f_t(y_1, \ldots, y_T) = f_t(y_1, \ldots, y_t) \prod_{s=0}^{T-t-1} \left[\sum_{j=T-s}^{T} \omega_{T-s,j} f_j(y_{T-s} | y_1, ..., y_{T-s-1}) \right]. \tag{4}$$

Expression (4) clearly shows which information is used to complement the observed data density in pattern t in order to establish the complete data density.

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

$$f_t(y_s | y_1, \ldots y_{s-1}) = f_T(y_s | y_1, \ldots y_{s-1}), \quad s = t+1, \ldots, T. \tag{5}$$

In other words, information which is unavailable is always borrowed from the completers. This strategy can be defended in cases where 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:

$$f_t(y_s|y_1,\ldots y_{s-1}) = f_s(y_s|y_1,\ldots y_{s-1}), \qquad s = t+1,\ldots,T. \tag{6}$$

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

The third special case of (3) will be ACMV. Thus, ACMV is reserved for the counterpart of MAR in the PMM context. The corresponding $\boldsymbol{\omega}_s$ vectors can be shown to have components:

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

5.1 Strategy Outline

We will briefly a general strategy. Several points which require further specification will be discussed in subsequent sections. (1) Fit a model to the pattern-specific identifiable densities: $f_t(y_1,\ldots,y_t)$. This results in a parameter estimate, $\hat{\boldsymbol{\gamma}}_t$. (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 (Rubin 1987, Schafer 1997, Verbeke and Molenberghs 2000).

5.2 Drawing from the Conditional Densities

In the previous section, we have seen how general identifying restrictions (3), 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 $\boldsymbol{\omega}$, such as in (7) for ACMV. This endeavor corresponds to items 2 and 3 of the strategy outline (5.1). In order to carry out item 4, we need to draw imputations from these 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: $f_3(y_1, y_2, y_3)$, $f_2(y_1, y_2)$, 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 three substeps.

1. The proportions ω need to be chosen or determined. Every ω in the unit interval is valid. Specific cases are:

   - For NCMV, $\omega = 1$.
   - For CCMV, $\omega = 0$.
   - For ACMV, ω is calculated from (7). 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).

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, then 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.

A few comments are in place. Except for in cases with only a few time points, the number of ω parameters proliferates quite rapidly. There are several ways to deal with it. First, special but important restrictions such as NCMV, CCMV, and ACMV do not suffer from this problem since each of the ω's involved is then determined by the choice of restriction. Second, one might envisage partial but important sensitivity analysis by letting all ω's be equal to a fixed quantity, which is chosen as, for example, a member of a grid filling the unit interval. Third, one could put prior distributions on the ω's, perhaps governed by simple hyperpriors. The first solution is followed in this paper. The other ones require further exploration.

In addition, determining the conditional distribution of the unobserved outcomes, given the observed ones, is easy in the Gaussian case. For categorical outcomes this is easy as well since it comes down to determining conditional multinomial probabilities which are again multinomial. However, for other distributional forms, this can be quite burdensome. In that case, the conditional distributions will have to be replaced by the corresponding ratio of marginal distributions. While this will change the algebra a bit, the methodology will not undergo fundamental changes.

6 ANALYSIS OF THE VOROZOLE STUDY

In order to concisely illustrate the methodology described in this chapter, we will apply it to the vorozole study, restricted to those subjects with 1, 2, and 3 follow up measurements, respectively. Thus, 190 subjects are included into the analysis, with subsample sizes 35, 86, and 69, respectively. The corresponding pattern probabilities are

$$\widehat{\boldsymbol{\pi}} = (0.184, 0.453, 0.363)', \tag{9}$$

and asymptotic covariance matrix

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}) = \begin{pmatrix} 0.000791 & -0.000439 & -0.000352 \\ -0.000439 & 0.001304 & -0.000865 \\ -0.000352 & -0.000865 & 0.001217 \end{pmatrix}. \tag{10}$$

These figures, apart from giving a feel for the relative importance of the various patterns, will be needed to calculate marginal effects (such as the marginal treatment effect) from pattern-mixture model parameters.

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

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}_v) = \begin{pmatrix} 0.001425 & -0.000814 & -0.000611 \\ -0.000814 & 0.002442 & -0.001628 \\ -0.000611 & -0.001628 & 0.002239 \end{pmatrix}.$$

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

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}_m) = \begin{pmatrix} 0.001771 & -0.000948 & -0.000823 \\ -0.000948 & 0.002788 & -0.001840 \\ -0.000823 & -0.001840 & 0.002663 \end{pmatrix}.$$

The treatment arm specific probabilities are not significantly different from each other. A classical χ^2 test produces $p = 0.864$. Hence, we will work with expressions (9) and (10).

We will apply each of the three strategies, presented in Section 5.1, to these data. First, a starting model will be fitted (Section 6.1). Second, it will be illustrated how hypothesis testing can be performed, given the pattern-mixture parameter estimates and their estimated covariance matrix (Section 6.2). Third, model simplification will be discussed (Section 6.3).

6.1 Fitting a Model

Strategies 2 and 1

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 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, be cause 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 is always done, be it explicitly by modeling the profile, averaged over all patterns.

In order 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 modeling strategy.

We include time and time^2 effects, as well as their interactions with treatment. Further, time by baseline value interaction is included as well. While we agree such a choice may seem controversial, it is consistent with the analysis plan and therefore we have opted to leave this term in. Alternatively, one could either remove this term or model raw scores rather than change scores. 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 1, 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 time^2 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 1). 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. Using the methodology detailed in Section 5, a GAUSS macro and a SAS macro (available from the authors' web pages), was written

Table 1: **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∗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)
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∗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)
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∗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)

to conduct the multiple imputation, to fit the model to the imputed datasets, and to combinef the results into a single inference. Results are presented in Table 1, 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

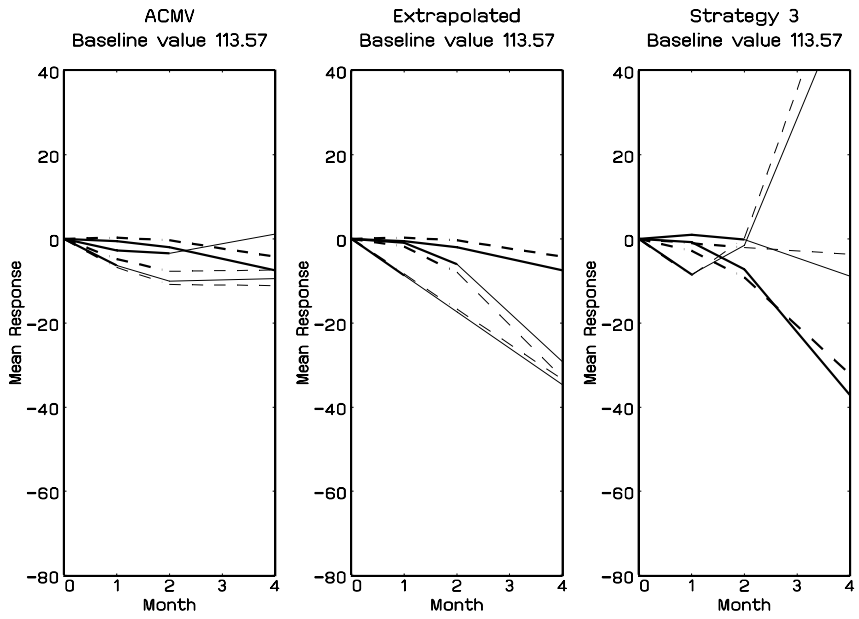

Figure 1: **Vorozole Study. For average level of baseline value, extrapolation based on initial model (strategy 2), ACMV, and strategy 3 is shown. The bold portion of the curves runs from baseline until the last obtained measurement, and the extrapolated piece is shown in thin type. The dashed line refers to megestrol acetate; the solid line is the Vorozole arm.**

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. Hence, it is well possible 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 her 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 (3), one can allow $\boldsymbol{\omega}_s$ to include components outside of the unit interval. In that situation, one has to ensure that

Table 2: **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)

the resulting density is still non-negative over its entire support.

Strategy 2b

In this strategy, *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 time^2 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. In order to achieve this goal, every effect would have to be made pattern-dependent. The estimated model parameters are presented in Table 2.

A graphical representation is given in Figure 1. 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.

6.2 Hypothesis Testing

For ease of exposition, let us assume we are interested in a single effect, e.g., time∗treatment interaction. For simplicity, we will generically refer to the parameter of interest as *treatment effect*. In the simplest case of a single parameter for the effect of interest, the corresponding selection model would contain exactly this single treatment effect parameter, turning the hypothesis testing task into a very straightforward one. If there would be several treatment-effect parameters, such as in a three-armed trial or in an analysis where interactions between treatment and other effects are included, standard hypothesis testing theory, can be applied.

In pattern-mixture models, it is possible to have a treatment-effect parameter for each pattern separately. This is the case for all five models in Tables 1 and 2. Let us note in passing that this does not need be the case. For example, in the final Strategy 2b analysis in Section 6.3 treatment effect is reduced to a single parameter. In such cases, the assessment of treatment effect is no more difficult than in a corresponding selection model. Therefore, this section will focus on the situation where there are pattern-dependent treatment effects.

It is useful to point out a strong analogy with post-hoc stratification, where pattern plays the role of a stratifying variable. A selection model corresponds to a pooled analysis, where data from all patterns (strata) are combined, without correction for the "confounding effect" stemming from the dropout patterns. A pattern-mixture model on the other hand does correct for pattern and hence, in a sense, for the confounding effect arising from pattern. If treatment effect does not interact with pattern, such as in the Strategy 2b analysis in Section 6.3, then a simple, so-called *corrected*, treatment effect estimate is obtained. Finally, if treatment effect interacts with pattern such as in all five models above (although it is not significant in this case), there is heterogeneity of treatment effect across patterns (cf. heterogeneity of the relative risks in epidemiological studies).

In the latter case, two distinct routes are possible. The more "epidemiologic" view-point is to direct inferences towards the vector of treatment effecs. In our case, this implies, for example, testing for the treatment by time interaction to be zero in all three patterns simultaneously. Alternatively, one can calculate the same quantity as would be obtained in the corresponding selection model. Let us consider the latter option in more detail.

Precisely, 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:

$$\beta_\ell = \sum_{d=1}^{T} \beta_{\ell d} \pi_d, \qquad \ell = 1, \ldots, g. \tag{11}$$

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

$$\text{Var}(\beta_1, \ldots, \beta_g) = AVA', \tag{12}$$

where

$$V = \left(\begin{array}{c|c} \text{Var}(\beta_{\ell d}) & 0 \\ \hline 0 & \text{Var}(\pi_d) \end{array} \right) \tag{13}$$

and

$$A = \frac{\partial(\beta_1, \ldots, \beta_g)}{\partial(\beta_{11}, \ldots, \beta_{Tg}, \pi_1, \ldots, \pi_T)}. \tag{14}$$

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 (9) and (10). A Wald test statistic for the null hypothesis $H_0 : \beta_1 = \ldots = \beta_g = 0$ is then given by

$$\boldsymbol{\beta}_0' AVA' \boldsymbol{\beta}_0, \tag{15}$$

where $\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_g)'$.

We will now apply both testing approaches for the models presented in Tables 1 and 2. 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.

Strategy 2a

Recall that the parameters are presented in Table 1 as the initial model. The treatment effect vector is $\boldsymbol{\beta} = (0.33, -0.95, 0.82)'$ with, since the patterns are analyzed separately,

diagonal covariance matrix:

$$V = \begin{pmatrix} 15.28 & & \\ & 3.44 & \\ & & 0.90 \end{pmatrix}.$$

This leads to the test statistic $\boldsymbol{\beta}' V^{-1} \boldsymbol{\beta} = 1.02$ on 3 degrees of freedom, producing $p = 0.796$.

In order to calculate the marginal treatment effect, we apply (12)–(15). The marginal effect is estimated as $\hat{\beta}_0 = -0.07$ (s.e. 1.16). The corresponding asymptotic p value is $p = 0.95$. Both approaches agree on the non-significance of the treatment effect.

bfseries Strategy 2b

The parameters are presented in Table 2. The treatment effect vector is $\boldsymbol{\beta} = (5.25, 348, 3.44)'$ 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 time^2), are common to all three patterns, hence inducing dependence across patterns. This leads to the test statistic $\boldsymbol{\beta}' V^{-1} \boldsymbol{\beta} = 0.70$ on 3 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 $p = 0.49$. The different numerical value of the treatment effects, 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.

Strategy 1

For this case, we will consider several approximate ways 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, between, and total covariance matrices:

$$\boldsymbol{\beta}_{CC} = (-2.09, -1.68, 0.82)', \tag{16}$$

$$W_{CC} = \begin{pmatrix} 1.67 & 0.00 & 0.00 \\ 0.00 & 0.59 & 0.00 \\ 0.00 & 0.00 & 0.90 \end{pmatrix}, \tag{17}$$

$$B_{CC} = \begin{pmatrix} 2.62 & 0.85 & 0.00 \\ 0.85 & 0.72 & 0.00 \\ 0.00 & 0.00 & 0.00 \end{pmatrix}, \tag{18}$$

and

$$T_{CC} = \begin{pmatrix} 4.80 & 1.02 & 0.00 \\ 1.02 & 1.46 & 0.00 \\ 0.00 & 0.00 & 0.90 \end{pmatrix}. \tag{19}$$

In the stratified case, we want to test the hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$. Using (16)–(18), 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 3. All p values are highly 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 (11) represents a non-linear transformation of the parameter vector and therefore needs further development. First, consider $\boldsymbol{\pi}$ to be part of the parameter vector. Since there is no missingness involved in this part, it contributes to the within matrix, but not to the between matrix. Then, using (12), the approximate within matrix for the marginal treatment effect is

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

with, for the between matrix, simply

$$B_0 = \boldsymbol{\pi}' B \boldsymbol{\pi}.$$

The results are presented in the second panel of Table 3. All three p values are in between those obtained for Strategies 2 and 3. Of course, all five agree on the non-significance

Table 3: **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

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 for Strategy 2b. Finally, we note that all conclusions are conditional upon the unverifiable assumption that the posited restrictions (and hence, dropout mechanisms) are correct. Therefore, they should preferrably be used in conjunction, within a sensitivity analysis.

6.3 Model Reduction

Model building guidelines for the standard linear mixed-effects model can be found in Verbeke and Molenberghs (1997, Chapter 2). These guidelines can be used without any problem in a selection model context, but the pattern-mixture case is more complicated. Of course, the same general principles can be applied, taking into account the intertwining between the mean or fixed-effects structure on the one hand and the components of variability on the other hand.

In addition to these principles, one has to reflect on the special status of *pattern* in a

Table 4: **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)

pattern-mixture model. Broadly, we can distinguish between two cases, reflecting Strategy 2a (a per pattern analysis) and Strategy 2b (use pattern as a covariate). In fact, the identifying restriction strategy leaves the method of analysis unspecified, as mentioned in item 6 of the strategy outline (Section 5.1). While we have chosen to conduct a per-pattern analysis (Table 1), as in Strategy 1, it is possible to conduct a global analysis, using pattern as a covariate, or even to use selection modeling, as long as the *proper* nature of the imputation is preserved (Rubin 1987). Therefore, we will discuss and illustrate model reduction using the second and third strategies.

Strategy 2b

Model reduction in a context where pattern is used as a covariate is clearly of the same level of complexity as with complete data or for a selection model. Let us reduce the model presented in Table 2. It is convenient to use a hierarchical representation of the model. The following effects are removed using a hierarchical sequence of models, and using F test statistics: the time by pattern by treatment interaction ($p = 0.934$), the time by pattern interaction ($p = 0.776$), the time by pattern by baseline value interaction ($p = 0.707$), the time by baseline by treatment interaction ($p = 0.165$), and the time^2 by treatment interaction ($p = 0.093$). The reduced model is displayed in Table 4.

Strategy 2a

For strategy 2a, where a per-pattern analysis is conducted, there are several model building decisions to be made.

- In the process of simplifying, one can allow that effects are shared between two or more patterns. However, then this strategy effectively reduces to Strategy 2b and we will not allow it here.
- When simplifying the model, effects are either absent or common to all patterns. Again, this approach is close to Strategy 2b, and can be conducted within that framework without any problem if one start with a model where all effects, including the covariance parameters. For this reason, we will not pursue it further.
- Finally, model reduction is done entirely separately in each of the patterns. This may yield different levels of simplification for each pattern and certainly a pattern-specific set of covariates which is found to influence the response profile. This strategy will be illustrated.

In order to enable treatment-effect assessment, the interaction between time and treatment will not be removed from the models. In pattern 1, there is one simplification possible in the sense that the interaction between time and baseline is not significant ($p = 0.415$). Thus, the only effects that remain in the model are time and the time by treatment interaction. For patterns 2 and 3, there are no non-significant effects to be removed. In conclusion baseline FLIC score influences the follow up scores in patterns 2 and 3, but not in pattern 1.

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 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.

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 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 CCMV and NCMV are extremes for the $\boldsymbol{\omega}_s$ vector in (3), it is very natural to consider the idea of *ranges* in the allowable space of $\boldsymbol{\omega}_s$. Clearly, any $\boldsymbol{\omega}_s$ 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.

The SAS and GAUSS macros which have been used to carry out the multiple imputation related tasks are available from the authors' web pages.

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.

Curran, D., Pignatti, F., and Molenberghs, G. (1997) Milk protein trial: informative dropout versus random drop-in. *Submitted for publication.*

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 Publications, Clarendon Press.

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 *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.

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.

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. In: *Imputation and Editing of Faulty or Missing Survey Data*, U.S. Department of Commerce, pp. 1–23.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons, New York.

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-Verlag.

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

Verbeke, G., Lesaffre, E., and Spiessens, B. (2000) The practical use of different strategies to handle dropout in longitudinal studies. *Submitted for publication.*