

Pattern-mixture models with proper time dependence

Peer-reviewed author version

Kenward, Michael G.; MOLENBERGHS, Geert & THIJS, Herbert (2003)

Pattern-mixture models with proper time dependence. In: *Biometrika*, 90(1). p. 53-71.

DOI: 10.1093/biomet/90.1.53

Handle: <http://hdl.handle.net/1942/416>

Pattern-Mixture Models With Proper Time Dependence

BY M.G. KENWARD

*Medical Statistics Unit, London School of Hygiene and Tropical Medicine
Keppel Street, London WC1E 7HT, United Kingdom
mike.kenward@lshtm.ac.uk*

G. MOLENBERGHS AND H. THIJS

*Biostatistics, CenStat, Limburgs Universitair Centrum,
Universitaire Campus, B-3590 Diepenbeek, Belgium
geert.molenberghs@luc.ac.be, herbert.thijs@luc.ac.be*

SUMMARY

Recently, pattern-mixture modelling has become a popular tool for modelling incomplete longitudinal data. Such models are under-identified in the sense that, for any drop-out pattern, the data provide no direct information on the distribution of the unobserved outcomes, given the observed ones. One simple way of overcoming this problem, ordinary extrapolation of sufficiently simple pattern-specific models, often produces rather unlikely descriptions; several authors consider identifying restrictions instead. Molenberghs et al. (1998) have constructed identifying restrictions corresponding to missing at random. In this paper, the family of restrictions where drop-out does not depend on future, unobserved observations is identified. The ideas are illustrated using a clinical study of Alzheimer patients.

Some key words: Drop-out; Longitudinal data; Missing at random; Missing data; Repeated measurements; Selection model.

1. INTRODUCTION

In this paper, we will restrict missingness to be of monotone, drop-out type, and our main focus will be on longitudinal studies. Rubin (1976) and Little & Rubin (1987, Ch. 6) make important

distinctions between different missing-values processes. A drop-out process is said to be completely random if the drop-out is independent of both unobserved and observed data and random if, conditional on the observed data, the drop-out is independent of the unobserved measurements; otherwise the drop-out process is termed nonrandom. If a drop-out process is random then it is to use a likelihood-based analysis that ignores the drop-out mechanism, provided the parameter describing the measurement process is functionally independent of the parameter describing the drop-out process, the so-called parameter distinctness condition. This situation is termed ignorable by Little & Rubin (1987). For reviews see Little (1995) and Kenward & Molenberghs (1999).

Historically, many methods have been formulated within the selection modelling frame (Little & Rubin, 1987) as opposed to pattern-mixture modelling (Little, 1993, 1994). A selection model factorises 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 when the marginal measurement distribution of the complete data is of interest. Furthermore, Little and Rubin's taxonomy is most easily developed in the selection setting. However, it is often argued that, especially in the context of nonrandom missingness models, selection models, although identifiable, should be approached with caution (Glynn et al., 1986). Therefore, there has been renewed interest in pattern-mixture models in recent years (Little, 1993, 1994; Hogan & Laird, 1997). A small literature review is given in Thijs et al. (2002). Another family of models, where random effects are driving both the outcome and missingness processes, was given attention by Wu & Carroll (1988) and Wu & Bailey (1988, 1989). These models are also referred to as shared-parameter models. Little (1995) combines the concepts of pattern-mixture models and shared parameters.

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. Thijs et al. (2002) contrasted simplified models, that can be extrapolated beyond the time of drop-out with identifying-restrictions-based models. We will discuss these in turn.

If models are sufficiently simple, then extrapolation beyond the time of drop-out can be very

easy. One such example arises when simple polynomial time trends within patterns are considered (Hogan & Laird, 1997). Alternatively, ‘pattern’ can be used as a covariate. For example, rather than estimating a separate time trend within each pattern, one could assume that the time evolution within a pattern is unstructured, but time evolutions are parallel across patterns. While these strategies are computationally simple, there a nontrivial price must be paid: the form of the drop-out mechanism, implied by extrapolating a simple mean profile, comes as an arbitrary by-product that may or may not be appropriate or sensible and is usually left unassessed. Thijs et al. (2002) deemed the identification-restrictions strategy to be the more promising one. Little (1993, 1994) presented a number of examples.

Since we are restricting attention to monotone missingness, we can easily indicate a drop-out pattern by the numbers of observations made. In that sense, pattern t collects all individuals with the first t measurements taken ($t = 1, \dots, T$). Thijs et al. (2002) constructed a general identifying-restrictions framework in which the distribution of the $(t + 1)$ th measurement, given the earlier measurements, in pattern t , y_{t+1} say, is set equal to a linear combination of the corresponding distributions in patterns $t + 1$ to T . Since this family is characterised by the use of observable distributions to identify the unobservable ones, we term it the ‘interior’ family of identifying-restrictions. Three members of this family are studied in detail by Thijs et al. (2002): complete-case missing value restrictions (Little, 1993), where information is borrowed from the completers only, available-case missing values, equivalent to missing at random (Molenberghs et al., 1998), for which one particular linear combination needs to be considered, and neighbouring-case missing value restrictions, where information is borrowed from the closest available pattern.

The equivalence of available-case missing values and missing at random is important in that it enables us to make a clear connection between the selection and pattern-mixture frameworks. By implication, the other members of the interior family are of missing not at random type, but there do exist missing not at random type restrictions that are not captured by this family. Whatever framework chosen, the collection of missing not at random mechanisms is huge. The selection modeller often restricts attention to the set of missing not at random mechanisms in which missingness is allowed to depend on past measurements and on the present, possibly unobserved measurement,

but not on future ones. We will term such mechanisms non-future dependent. While they are natural and easy to consider in a selection model context, there exist important examples of mechanisms that are not non-future dependent, such as shared-parameter models (Wu & Bailey, 1989; Little, 1995).

In this paper, we will characterise the class of pattern-mixture models that corresponds to missing non-future dependent. It will be shown that this class intersects with the interior family, available-case missing values being part of the intersection, but neither is a subset of the other. The formulation of this characterisation does not imply prejudice against those pattern-mixture modelling strategies that do allow for dependence on future outcomes. Rather, it should be viewed as a step in refining the missingness mechanism taxonomy. Indeed, the classification based on missing completely at random, missing at random, i.e. available-case missing values, and missing not at random is expanded by further splitting missing not at random into non-future-dependent and future-dependent mechanisms. Thus, our characterisation gives us more control, within the pattern-mixture framework, over the implied missing-value mechanism, lack of which has in the past been a major source of criticism.

2. DATA SETTING

Let us assume that we have $r = 1, \dots, T$ drop-out patterns, where $r = t$ indicates the number of measurements actually obtained. A selection model factorisation is given by

$$f(y_1, \dots, y_T, r = t) = f(y_1, \dots, y_T) f(r = t | y_1, \dots, y_T).$$

Pattern-mixture models are based on the reverse factorisation, given by

$$\begin{aligned} f(y_1, \dots, y_T, r = t) &= f(y_1, \dots, y_T | r = t) f(r = t) \\ &= f_t(y_1, \dots, y_T) f(r = t) \\ &= f_t(y_1, \dots, y_t) f_t(y_{t+1} | y_1, \dots, y_t) f_t(y_{t+2}, \dots, y_T | y_1, \dots, y_{t+1}) f(r = t), \quad (1) \end{aligned}$$

where $f_t(y_1, \dots, y_T) = f(y_1, \dots, y_T | r = t)$. The first three factors in (1) are referred to as the distributions of past, present and future measurements, respectively. Only the first and the fourth factors are identifiable from the data.

3. NON-FUTURE DEPENDENCE

In the selection model context, we can formulate missing non-future dependent as

$$f(r = t|y_1, \dots, y_T) = f(r = t|y_1, \dots, y_{t+1}). \quad (2)$$

Note that missing at random is a special case of missing non-future dependent, which in turn is a subclass of missing not at random.

Within the pattern-mixture framework, we define non-future dependent missing value restrictions as follows:

$$f(y_t|y_1, \dots, y_{t-1}, r = j) = f(y_t|y_1, \dots, y_{t-1}, r \geq t - 1), \quad (3)$$

for all $t \geq 2$ and all $j < t - 1$. Non-future missing values is not a comprehensive set of restrictions, but rather leaves one conditional distribution per incomplete pattern unidentified:

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

In other words, the distribution of the ‘current’ unobserved measurement, given the previous ones, is unconstrained. This implies that the non-future missing values class contains members not contained in the interior family, where every restriction takes the form of a linear combination of observable distributions. Conversely, (3) excludes such mechanisms as complete-case missing values and neighbouring-case missing values, showing that there are members of the interior family that are not of non-future missing values type. Finally, choosing (4) of the same functional form as (3) establishes available-case missing values as a member of the intersection of the interior and non-future missing values families. The latter is particularly important since it shows, because of the equivalence of available case missing values and missing at random, that missing at random belongs to both families.

The following theorem establishes the equivalence between missing non-future dependence and non-future missing values, showing the non-future missing values restrictions correspond to non-future dependence, just as available-case missing values corresponds to missing at random.

Theorem 1 . *For longitudinal data with drop-outs, missing non-future dependence is equivalent to non-future missing values.*

The proof of Theorem 1, following arguments similar to those in Molenberghs et al. (1998), is given in the Appendix.

A consequence of using (3) is that the joint distribution will not typically have a simple analytical representation. This is to be understood in the sense that covariate effects would not necessarily be linear on an appropriate scale. However, this is not to say there is no analytical form. Moreover, it does not have to be a major disadvantage, provided the resulting distribution is empirically reasonable. Such a requirement may help guide the choice for (4). As will be seen in the next section, this feature does not prevent the formulation of a tractable analysis strategy.

4. USING IDENTIFYING RESTRICTIONS

For pattern t , the complete data density is given by

$$f_t(y_1, \dots, y_T) = f_t(y_1, \dots, y_t) f_t(y_{t+1} | y_1, \dots, y_t) f_t(y_{t+2}, \dots, y_T | y_1, \dots, y_{t+1}). \quad (5)$$

It is assumed that the first factor is known or, more realistically, modelled using the observed data. Then, identifying restrictions are applied in order to identify the second and third components. The user has full freedom to choose $f_t(y_{t+1} | y_1, \dots, y_t)$. Substantive considerations can be used to identify this density, or a family of densities can be considered by way of sensitivity analysis. Based on (3), the densities $f_t(y_j | y_1, \dots, y_{j-1})$, ($j \geq t+2$) are identified. This identification involves not only the patterns for which y_j is observed, but also the pattern for which y_j is the current, the first unobserved measurement.

From the discussion about the first two factors in (5), it follows that $f_t(y_1, \dots, y_{t+1})$ is identified from modelling and choice. Next, non-future missing values states that

$$f_t(y_s | y_1, \dots, y_{s-1}) = f_{(\geq s-1)}(y_s | y_1, \dots, y_{s-1}), \quad (6)$$

for $s = t+2, \dots, T$. We can transform (6) as follows:

$$\begin{aligned} & f_t(y_s | y_1, \dots, y_{s-1}) \\ &= f_{(\geq s-1)}(y_s | y_1, \dots, y_{s-1}) \\ &= \frac{\sum_{j=s-1}^T \alpha_j f_j(y_1, \dots, y_s)}{\sum_{j=s-1}^T \alpha_j f_j(y_1, \dots, y_{s-1})} \end{aligned} \quad (7)$$

$$= \sum_{j=s-1}^T \frac{\alpha_j f_j(y_1, \dots, y_{s-1})}{\sum_{\ell=s-1}^T \alpha_\ell f_\ell(y_1, \dots, y_{s-1})} f_j(y_s | y_1, \dots, y_{s-1}). \quad (8)$$

Thus, a general expression is

$$f_t(y_s | y_1, \dots, y_{s-1}) = \sum_{j=s-1}^T \omega_{sj}(y_1, \dots, y_{s-1}) f_j(y_s | y_1, \dots, y_{s-1}), \quad s = t+2, \dots, T, \quad (9)$$

with

$$\omega_{sj}(y_1, \dots, y_{s-1}) = \frac{\alpha_j f_j(y_1, \dots, y_{s-1})}{\sum_{\ell=s-1}^T \alpha_\ell f_\ell(y_1, \dots, y_{s-1})}. \quad (10)$$

When ω_{sj} are modelled to differ from those specified in (10) using, for example, techniques such as in Rubin (1977), then missing data mechanisms are obtained that do depend on future observations. In a sensitivity analysis, it can be envisaged that the impact of such departures on substantive conclusions might be explored. Indeed, in the general missing not at random case, the conditional distribution of the unobserved measurements given the observed ones needs to be determined by means of assumptions. Under non-future missing values, only the conditional distribution of the first, i.e. current, unobserved outcome given the observed ones needs to be identified by assumption. Thus, when missing non-future dependence is deemed plausible, one combines the flexibility of a broad class of models with a sensitivity space that is reasonably easily managed. In the special case of missing at random, the conditional distributions of the unobserved outcomes are completely identified by means of available-case missing values and there is no further room for sensitivity analysis.

To get a grasp on the differences and similarities between the interior and non-future missing value families, we focus on the special case of four measurements and four patterns.

The interior family takes the following form, in which, for ease of notation, y_t is replaced by t :

$$f_4(1234) = f_4(1234), \quad (11)$$

$$f_3(1234) = f_3(123)f_4(4|123), \quad (12)$$

$$f_2(1234) = f_2(12) [\omega(12)f_3(3|12) + \{1 - \omega(12)\}f_4(3|12)] f_4(4|123), \quad (13)$$

$$\begin{aligned} f_1(1234) &= f_1(1) \{ \omega_{12}(1)f_2(2|1) + \omega_{13}(1)f_3(2|1) + \omega_{14}(1)f_4(2|1) \} \\ &\quad \times [\omega(12)f_3(3|12) + \{1 - \omega(12)\}f_4(3|12)] f_4(4|123). \end{aligned} \quad (14)$$

To facilitate comparison, we will consider three special cases.

Case 1. Complete-case missing values: $\omega = \omega_{12} = \omega_{13} = 0$ and $\omega_{14} = 1$.

Case 1. Neighbouring-case missing values: $\omega = \omega_{12} = 1$ and $\omega_{13} = \omega_{14} = 0$.

Case 1. Available-case missing values: we set

$$\begin{aligned}\omega(12) &= \frac{\alpha_3 f_3(12)}{\alpha_3 f_3(12) + \alpha_4 f_4(12)}, \\ \omega_{12}(1) &= \frac{\alpha_2 f_2(1)}{\alpha_2 f_2(1) + \alpha_3 f_3(1) + \alpha_4 f_4(1)}, \\ \omega_{13}(1) &= \frac{\alpha_3 f_3(1)}{\alpha_2 f_2(1) + \alpha_3 f_3(1) + \alpha_4 f_4(1)}, \\ \omega_{14}(1) &= \frac{\alpha_4 f_4(1)}{\alpha_2 f_2(1) + \alpha_3 f_3(1) + \alpha_4 f_4(1)}.\end{aligned}$$

For the non-future missing values restrictions, the counterparts to (11)–(14) are

$$f_4(1234) = f_4(1234), \tag{15}$$

$$f_3(1234) = f_3(123)g_3(4|123), \tag{16}$$

$$f_2(1234) = f_2(12)g_2(3|12) [\delta(123)g_3(4|123) + \{1 - \delta(123)\}f_4(4|123)], \tag{17}$$

$$\begin{aligned}f_1(1234) &= f_1(1)g_1(2|1) \{ \delta_{12}(12)g_2(3|12) + \delta_{13}(12)f_3(3|12) + \delta_{14}(12)f_4(3|12) \} \\ &\times [\delta(123)g_3(4|123) + \{1 - \delta(123)\}f_4(4|123)],\end{aligned} \tag{18}$$

with

$$\begin{aligned}\delta(123) &= \frac{\alpha_3 f_3(123)}{\alpha_3 f_3(123) + \alpha_4 f_4(123)}, \\ \delta_{12}(12) &= \frac{\alpha_2 f_2(12)}{\alpha_2 f_2(12) + \alpha_3 f_3(12) + \alpha_4 f_4(12)}, \\ \delta_{13}(12) &= \frac{\alpha_3 f_3(12)}{\alpha_2 f_2(12) + \alpha_3 f_3(12) + \alpha_4 f_4(12)}, \\ \delta_{14}(12) &= \frac{\alpha_4 f_4(12)}{\alpha_2 f_2(12) + \alpha_3 f_3(12) + \alpha_4 f_4(12)}.\end{aligned}$$

The dependence of the ω 's and δ 's on the y follows the same notational conventions. Note that one may or may not want them to depend on covariate levels. To avoid confusion with the interior-

family counterparts, the ω 's have been changed to δ in this case. The freedom of choice is on the distributions g_1 , g_2 and g_3 and not on the δ -weights, which are fixed by assumption (3).

Theorem 1 has made a formal connection between certain selection models, missing non-future dependent, and certain pattern-mixture models, non-future missing values. The models used in the examples of Diggle & Kenward (1994), where the drop-out model is built up from logistic regressions with dependence on the present and previous measurements only, and not on the full history, are members of the missing non-future dependence family. Equations (15)–(18) allow us to derive the missing non-future dependence counterparts of this type of model. Suppressing arguments, assume that d_t is the selection-model probability of drop-out at time $t + 1$. Then the selection model factorisations corresponding to (15)–(18) are

$$f_4(1234)\alpha_4 = f(1234)(1 - d_2)(1 - d_3)(1 - d_4), \quad (19)$$

$$f_3(1234)\alpha_3 = f(1234)(1 - d_2)(1 - d_3)d_4, \quad (20)$$

$$f_2(1234)\alpha_2 = f(1234)(1 - d_2)d_3, \quad (21)$$

$$f_1(1234)\alpha_1 = f(1234)d_2. \quad (22)$$

Taking the ratio of (19) and (20) produces, upon rearranging,

$$g_3(4|123) = \left(\frac{d_4}{1 - d_4} \right) \left(\frac{\alpha_4}{\alpha_3} \right) \left(\frac{f_4(1234)}{f_3(123)} \right). \quad (23)$$

If d_4 has been completely specified, for example a logistic regression in y_4 and y_3 , we can see exactly from (23) how $g_3(4|123)$ depends on d_4 . On the other hand, if d_4 is subject only to the generic constraint that it depends only on y_4 and y_3 , $g_3(4|123)$ can be written generically as

$$g_3(4|123) = \tilde{g}_3(4|3) \frac{f_3(123)}{f_4(1234)}.$$

The other g functions follow from similar logic.

While such links are important in considering how structure in a selection framework is reflected in a pattern-mixture framework, and vice versa, it is obvious that particular models expressed in one framework will typically produce awkward representations in the other and may therefore be hard to implement. It follows that the modeller should not be seeking exact correspondence in the two frameworks, but rather be guided by equivalences between subclasses, such as available case

missing values and missing at random, or missing non-future dependence and non-future missing values.

Next, we will consider a few special cases. This is done by making specific choices for g_1 , g_2 and g_3 .

Case 4. The g functions are selected in the spirit of complete-case missing values; that is they are set equal to their f_4 counterparts. To be explicit, $g_1(2|1) = f_4(2|1)$, $g_2(3|12) = f_4(3|12)$ and $g_3(4|123) = f_4(4|123)$.

Case 5. The g functions are determined in the spirit of neighbouring-case missing values: $g_1(2|1) = f_2(2|1)$, $g_2(3|12) = f_3(3|12)$ and $g_3(4|123) = f_4(4|123)$.

Obviously, one could consider a choice where the g functions are in line with available-case missing values. It is easy to show that available case missing values is then obtained as it should be, since available-case missing values is in the intersection of both families.

Let us now see what the pattern-specific choices reduce to for the five special cases. Of course, there is no room for choice in f_4 and in all five cases $f_3(1234) = f_3(123)f_4(4|123)$. In the second pattern,

$$f_2(1234) = \begin{cases} f_2(12)f_4(34|12) & \text{(Cases 1 and 4),} \\ f_2(12)f_3(3|12)f_4(4|123) & \text{(Cases 2 and 5),} \\ f_2(12) \left\{ \frac{\alpha_3 f_3(123) + \alpha_4 f_4(123)}{\alpha_3 f_3(12) + \alpha_4 f_4(12)} \right\} f_4(4|123) & \text{(Case 3).} \end{cases}$$

While some distributions coincide for the second pattern, all five are different in the first:

$$f_1(1234) = \begin{cases} f_1(1)f_4(234|1) & \text{(Case 1),} \\ f_1(1)f_2(2|1)f_3(3|12)f_4(4|123) & \text{(Case 2),} \\ f_1(1) \left\{ \frac{\alpha_2 f_2(12) + \alpha_3 f_3(12) + \alpha_4 f_4(12)}{\alpha_2 f_2(1) + \alpha_3 f_3(1) + \alpha_4 f_4(1)} \right\} \left\{ \frac{\alpha_3 f_3(123) + \alpha_4 f_4(123)}{\alpha_3 f_3(12) + \alpha_4 f_4(12)} \right\} f_4(4|123) & \text{(Case 3),} \\ f_1(1)f_4(2|1) \left\{ \frac{\alpha_2 f_2(12)f_4(3|12) + \alpha_3 f_3(123) + \alpha_4 f_4(123)}{\alpha_2 f_2(12) + \alpha_3 f_3(12) + \alpha_4 f_4(12)} \right\} f_4(4|123) & \text{(Case 4),} \\ f_1(1)f_2(2|1) \left\{ \frac{\alpha_2 f_2(12)f_3(3|12) + \alpha_3 f_3(123) + \alpha_4 f_4(123)}{\alpha_2 f_2(12) + \alpha_3 f_3(12) + \alpha_4 f_4(12)} \right\} f_4(4|123) & \text{(Case 5).} \end{cases}$$

This illustrates that neighbouring-case missing values and complete case-missing values, while in the interior family, are outside the non-future missing values family. Conversely, cases 4 and 5 do belong

to the intersection of the interior and non-future missing values families, together with available case missing values. Examples of non-future missing values restrictions that are not interior are readily found by making arbitrary choices about at least one of the g functions. In summary, both the interior and non-future missing values families are subclasses of general missing not at random mechanisms, extending the available case missing values class. The interior family is of value when no restriction beyond those produced by the observed data is deemed appropriate. The non-future missing values family is in place when one wants to consider restrictions that avoid the corresponding drop-out mechanism to depend on future, unobserved measurements. Importantly, both families have a nontrivial intersection, in the sense that, apart from available-case missing values, mechanisms such as those in cases 4 and 5 are included.

5. OVERALL STRATEGY

Having outlined how non-future missing value restrictions can be constructed in practice, we can now outline an overall strategy for analysis. A similar route has been followed for the interior family by Thijs et al. (2002). Our strategy will be applied to the case study in the next section.

Step 1. Specify and fit a model to the pattern-specific identifiable densities, $f_t(y_1, \dots, y_t)$. This results in a parameter estimate, $\hat{\gamma}_t$.

Step 2. Choose a model for $f_t(y_{t+1}|y_1, \dots, y_t)$. If this is done in a data-independent way, then a prior can be placed on the corresponding parameters. Otherwise, one can use procedures of the sort described in §4.

Step 3. Using the non-future missing values identification method, determine the conditional distributions of the unobserved outcomes, given the observed ones,

$$f_t(y_{t+2}, \dots, y_T | y_1, \dots, y_{t+1}). \quad (24)$$

Step 4. Once the full distributions are specified, inferences are required based on observed quantities. In principle, this requires some form of integration with respect to the distribution of the unknown quantities. In practice, simulation-based approaches are needed. One familiar route, with the advantage that the simulation and fitted models need not coincide, is multiple

imputation (Rubin, 1987; Schafer, 1997; Verbeke & Molenberghs, 2000). Multiple imputations are drawn for the unobserved components, given the observed outcomes and the correct pattern-specific density (24). For inferences to be correct, imputations have to be proper in the sense of Rubin & Schenker (1987), Rubin (1987, 1996) and Schafer (1997).

Step 5. The multiply-imputed sets of data are then analysed using the method of choice. This can be another pattern-mixture model, but can also be a selection model or any other desired model. However, one should carefully consider the implications of posited differences between the model for imputation and the final analysis model. For example, if the analyst deems the normality assumption to be important, it may be best to retain essentially the same model for the imputed sets of data. The situation will be different for categorical data with, for example, sufficiently rich multinomial models.

Step 6. Inferences can be conducted in the standard multiple imputation way (Rubin, 1987; Schafer, 1997; Verbeke & Molenberghs, 2000), by averaging point estimates and combining within- and between-imputation variability in order to obtain the asymptotic variance-covariance matrix of the estimates. Usually, this will still be a pattern-mixture model whereas substantive interest can be geared towards marginal quantities, such as overall treatment effect. In such a case, marginal effects follow from weighted linear combinations of pattern-specific effects, the variance of which follows from a straightforward application of the delta method. Details can be found in Verbeke & Molenberghs (2000, Ch. 20). We will briefly summarise the main steps.

Let the parameter of interest be denoted by θ . Suppose M imputations are drawn and let $\hat{\theta}^*$ be its estimator from completed dataset m . The M within-imputation estimator for θ are pooled to give the multiple imputation estimator:

$$\hat{\theta}^* = \frac{1}{M} \sum_{m=1}^M \hat{\theta}^{(m)}. \quad (25)$$

Suppose that complete data inference about θ would be based on $(\theta - \hat{\theta}) \sim N(0, U)$. Then one can make normal-based inferences for θ based upon

$$(\theta - \hat{\theta}^*) \sim N(0, V), \quad (26)$$

where

$$V = \hat{W} + \left(\frac{M+1}{M}\right) \hat{B}, \quad (27)$$

$$\hat{W} = \frac{\sum_{m=1}^M U^{(m)}}{M} \quad (28)$$

is the average within-imputation variance, and

$$\hat{B} = \frac{\sum_{m=1}^M (\hat{\theta}^{(m)} - \hat{\theta}^*)(\hat{\theta}^{(m)} - \hat{\theta}^*)'}{M-1} \quad (29)$$

is the between-imputation variance (Rubin, 1987). These formulae allow calculation of multiple imputation based standard errors. At first sight, testing hypotheses could be based on the asymptotic normality results (26) and (27). However, the rationale for using asymptotic results and hence χ^2 reference distributions is not just a function of the sample size, N , but also of the number of imputations, M . Therefore, Li et al. (1991) propose the use of an F reference distribution. To test the hypothesis $H_0 : \theta = \theta_0$, they advocate calculating p -values by

$$p = \text{pr}(F_{k,w} > F), \quad (30)$$

where k is the length of the parameter vector θ , $F_{k,w}$ is an F random variable with k numerator and w denominator degrees of freedom, and

$$F = \frac{(\theta^* - \theta_0)' W^{-1} (\theta^* - \theta_0)}{k(1+r)}, \quad (31)$$

$$\begin{aligned} w &= 4 + (\tau - 4) \left\{ 1 + \frac{(1 - 2\tau^{-1})}{r} \right\}^2, \\ r &= \frac{1}{k} \left(1 + \frac{1}{M} \right) \text{tr}(BW^{-1}), \\ \tau &= k(M-1). \end{aligned} \quad (32)$$

It is interesting to note that, when $M \rightarrow \infty$, the reference distribution of F approaches an $F_{k,\infty} = \chi^2/k$ -distribution, in line with intuition. Good operational characteristics of this procedure are reported in Li et al. (1991), combining nicely with computational ease.

Clearly, procedure (30) can also be used when not the full vector θ , but a single component, a subvector or a set of linear contrasts is the subject of hypothesis testing. When a subvector is of

interest, a single component being a special case, the corresponding submatrices of B and W need to be used in (31) and (32). For a set of linear combinations $L\theta$, one should use the appropriately transformed covariance matrices, $\tilde{W} = LWL'$, $\tilde{B} = LBL'$ and $\tilde{V} = LVL'$. This will be the case for the analysis of the example, where scientific interest is directed towards a pair of treatment contrasts in a three-armed clinical trial.

6. CASE STUDY

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

Figure 1, about here.

Attrition over time is fairly steady for each treatment arm. The sample size per drop-out pattern and per treatment arm is displayed in Table 1. In each of the arms, about 40% drop out before the end of the study. Unfortunately, very little is known about the reasons for drop-out, in this particular study. While such information is generally important, one also needs to be able to analyse incomplete data in the absence of such knowledge.

A linear mixed model (Verbeke & Molenberghs, 2000) was fitted to the outcomes, in which the variance structure was modelled by means of a random subject effect, an exponential serial correlation process and measurement error. The fixed effects considered in the model were, apart from treatment effect, those of age, time, investigator and country, as well as 2- and 3-way interactions. From an initial model-selection exercise, only main effects of age, time, time² and treatment group

were retained.

Scientific interest is in the effect of treatment. Since there are three arms, we consider two treatment contrasts of the experimental arms versus the standard arm. Our focus here will be on estimates and standard errors for these contrasts, as well as on tests for the null hypothesis of no treatment effect.

We first consider the selection model approach. As in Diggle & Kenward (1994), we combine the measurement model with a logistic regression for drop-out with either only an intercept, corresponding to missing completely at random, and also an effect for previous outcome, corresponding to missing at random, or even an effect of the current possibly unobserved measurement, corresponding to missing not at random. The fitted average profiles are plotted in Fig. 2. Parameter estimates and standard errors for the treatment contrasts, as well as the associated test results, are reported in Table 2. Treatment effects are nonsignificant. The likelihood ratio statistic for comparing the missing at random and missing not at random models is 5.4 on 2 degrees of freedom. While this might be taken as some evidence for nonrandom drop-out, such a conclusion is strongly model dependent; see the discussion of Diggle & Kenward (1994) and Verbeke & Molenberghs (2000, Ch. 19). More reasonably it can be taken as some evidence against this particular missing at random model.

Table 2 and Figure 2, about here.

Next, we turn attention to the pattern-mixture models, using the strategy outlined in §5. We briefly describe each of the six steps. First, pattern-specific linear mixed models are fitted to the data. The fixed-effects structure comprises treatment indicators, time and its square, and age. The covariance structure is captured by means of a random subject effect, an exponential serial correlation process and measurement error. Secondly, the five special cases outlined at the end of §4 are used to define the identifying restrictions. In §4, the focus was on four measurements since this enabled us to highlight differences between all of them; here we will use all seven measurements for analysis. Thirdly, from these choices, the conditional distributions of the unobserved outcomes, given the observed ones, are constructed. Fourthly, five imputations are drawn from these conditionals.

Fifthly, the resulting multiply imputed datasets are analysed, using the same pattern-mixture model as was applied to the incomplete data. Finally, multiple-imputation inference is conducted. Admittedly, the set of restrictions considered here is far from exhausting the entire space of possible restrictions and hence leaves room for additional sensitivity analyses. However, it allows us to compare future and non-future dependent mechanisms, as well as a comparison with the selection models fitted earlier. Also there, different routes of sensitivity analysis could be chosen, such as replacing the parameter describing the dependence of drop-out on the current, possibly unobserved measurement by a sensitivity parameters.

Multiple imputation inference can be based upon the average estimate (25), the asymptotic ‘total’ covariance matrix (27), and the associated testing procedure (30). However, we have to reflect carefully on the implications of fitting a pattern-mixture model. Since there are seven drop-out patterns, there are also seven pairs of treatment contrasts. As indicated in Thijs et al. (2002) and Verbeke & Molenberghs (2000), either one can focus on a stratified analysis, in which the null hypothesis is directed to all 7×2 contrasts simultaneously, or one can construct the marginal effects, here a single pair of marginal treatment contrasts. For the first approach, the theory provided in the previous section is sufficient. For the marginal route, some extension is needed. Let $\theta_{\ell t}$ represent the treatment contrasts $\ell = 1, 2$ in pattern $t = 1, \dots, 7$, and let π_t be the proportion of patients in pattern t , easily derived from Table 1. Then the marginal contrasts are

$$\theta_{\ell} = \sum_{t=1}^7 \theta_{\ell t} \pi_t, \quad \ell = 1, 2. \quad (33)$$

Of course, covariates can be assumed to be present or absent in the mixing distribution. For example, if drop-out differs among treatment arms, it seems imperative to let the mixing distribution depend on treatment. In the current example, the effect of treatment on the mixing probabilities is not significant. The marginalised within-imputation variance is obtained using the delta method:

$$\tilde{W} = (A_1 | A_2) \left(\frac{W}{0} \middle| \frac{0}{\text{var}(\pi_t)} \right) \begin{pmatrix} A'_1 \\ A'_2 \end{pmatrix}, \quad (34)$$

where

$$A_1 = \frac{\partial(\theta_1, \theta_2)}{\partial(\theta_{11}, \theta_{12}, \theta_{21}, \dots, \theta_{72})}, \quad A_2 = \frac{\partial(\theta_1, \theta_2)}{\partial(\pi_1, \dots, \pi_7)}.$$

Similarly, the between-imputation variance is

$$\tilde{B} = A_2 B A'_2. \quad (35)$$

Expressions (34) and (35) are combined to produce the total marginal variance \tilde{V} as in (27). These quantities can be used for precision estimation and hypothesis testing, as outlined at the end of §5.

The results of our analysis are reported in Table 2. The marginal treatment effect assessments are all nonsignificant, in line with the results from the selection model analysis. However, all stratified treatment assessments produce significant p values, although to various levels of strength. Strong evidence is obtained from the available-case missing values model. Of course, the complete-case missing values analysis provides even stronger evidence, but this assumption may be unrealistic, since even patterns with few observations are completed using the set of completers, corresponding to pattern 7. Both of the other non-future missing values mechanisms, corresponding to Cases 4 and 5, where drop-out does not depend on future unobserved values, provide mild evidence for treatment effect. Importantly, we are in a position to consider which patterns are responsible for an identified treatment effect. Note that the contrasts are nowhere near significant in the complete pattern 7, while patterns 4 and 6 seem to contribute to the effect, consistently across patterns. The first contrast of pattern 3 is significant only under complete-case missing values, perhaps explaining why this strategy yields the most significant result.

Figure 3 graphically summarises the fit of these models for the first treatment arm; very similar displays for the other arms have been omitted. Clearly, the chosen identifying restrictions have a strong impact, especially for the patterns with earlier drop-out. Of course, from Table 1 it is clear that the earlier patterns are rather sparsely filled. It is striking to see that the missing non-future dependence patterns are not all grouped together. An important and perhaps counterintuitive feature is that the fitted averages depend on the identification method chosen, even at time points prior to drop-out. The reason for this is that, after imputation, a parametric model is fitted to the completed sequences as a whole, as opposed to, for example, change point models with change point at the time of drop-out. Hence, the smoothing induced by the parametric model applies across the entire sequence, before and after drop-out.

Figure 3, about here.

7. CONCLUDING REMARKS

Thijs et al. (2002) argued that pattern-mixture models, supplemented by identifying restrictions that specify the conditional distribution of the unobserved measurements given the observed ones in a given pattern, is a potentially useful way of modelling incomplete longitudinal data. Such restrictions allow one to reflect carefully on the nature of the assumptions made. Thijs et al. (2002) put forward one particular class of such restrictions, termed here the interior family. By looking at the implications of pattern-mixture models for the nature of the drop-out mechanism, we have been able both to answer criticisms of restrictions of the interior type and to formulate a new class of pattern-mixture models, called non-future missing values, corresponding to those drop-out mechanisms in which drop-out does not depend on future unobserved measurements. Such a class is useful since in practice one may often want to prevent drop-out from depending on the future. At the same time, the non-future missing values family provides a sensible restriction on the vast collection of missing not at random models, while containing the important subfamily of missing at random models. Such classes are important in the construction of sensitivity analyses for incomplete data.

While this is useful, undoubtedly in cases where substantive information is available on the nature of and reasons for drop-out, it is equally true when such information is not available, but general considerations can be made using knowledge about the nature of the problem. Indeed, parallel to the selection model case, the developments made here permit one to consider a taxonomy within the pattern-mixture model family as well, as presented in Fig. 4.

Figure 4, about here.

Hence, one can decide to limit the restrictions to whatever class is deemed appropriate. This map also enables us, to a large extent, to compare selection and pattern-mixture models. For example, available case missing values and missing at random models are similar in spirit, but they do not have to be the same since the corresponding models may include different effects; for example, effects for the drop-out pattern are typically absent in a selection model. Also, Cases 4 and 5 are at the same level as the typical Diggle & Kenward (1994) model, which is missing non-future dependent. External information about the reason for drop-out varies greatly from study

to study. The taxonomy in Fig. 4 allows the researcher to reflect both what is known about the drop-out mechanism and the patterns of behaviour that will be allowed in the sensitivity analysis. In particular, we noted that restriction to either the non-future missing values family, in which drop-out does not depend on future measurements, or to the interior family, in which restrictions depend on the observed patterns only, or to both may be desirable, depending on the context. We have shown they have a nontrivial intersection, including not only available case missing values, but also Cases 4 and 5.

In addition, pattern-mixture models allow for an easy distinction between marginalised inference and inference stratified over drop-out pattern. While the primary analysis of a clinical trial will usually be the marginal analysis, the stratified view allows a more detailed consideration of the treatment effect. In our case study, it was seen that patients with an intermediate length of follow-up are primarily responsible for the observed treatment effect.

ACKNOWLEDGEMENT

We would like to thank David Clayton for helpful discussions. We gratefully acknowledge support from Fonds Wetenschappelijk Onderzoek-Vlaanderen and from Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in Industrie. We acknowledge support from Interuniversity Attraction Poles Programme P5/24-Belgian State-Federal Office for Scientific, Technical and Cultural Affairs.

APPENDIX

Technical details

The missing non-future dependence assumption states that

$$f(r = t | y_1, \dots, y_T) = f(r = t | y_1, \dots, y_{t+1}) \quad (\text{A1})$$

and the non-future missing values assumption corresponds to

$$f(y_t | y_1, \dots, y_{t-1}, r = j) = f(y_t | y_1, \dots, y_{t-1}, r \geq t - 1), \quad (\text{A2})$$

for all $t \geq 2$ and all $j < t - 1$.

Lemma A1. *In a longitudinal setting with drop-out, non-future missing values is equivalent to $f(y_t|y_1, \dots, y_{t-1}, r = j) = f(y_t|y_1, \dots, y_{t-1})$, for all $t \geq 2$ and all $j < t - 1$.*

Remark. This lemma assumes strictly one fewer identification than the corresponding lemma in Molenberghs et al. (1998).

Proof of Lemma A1. Take $t \geq 2, j < t - 1$. Then non-future missing values leads to

$$\begin{aligned}
f(y_t|y_1, \dots, y_{t-1}) &= \sum_{i=1}^{t-2} f(y_t|y_1, \dots, y_{t-1}, r = i)f(r = i) + f(y_t|y_1, \dots, y_{t-1}, r \geq t-1)f(r \geq t-1) \\
&= \sum_{i=1}^{t-2} f(y_t|y_1, \dots, y_{t-1}, r = j)f(r = i) + f(y_t|y_1, \dots, y_{t-1}, r = j)f(r \geq t-1) \\
&= f(y_t|y_1, \dots, y_{t-1}, r = j) \left\{ \sum_{i=1}^{t-2} f(r = i) + f(r \geq t-1) \right\} \\
&= f(y_t|y_1, \dots, y_{t-1}, r = j).
\end{aligned}$$

Conversely, again for $t \geq 2, j < t - 1$,

$$\begin{aligned}
f(y_t|y_1, \dots, y_{t-1}, r \geq t-1)f(r \geq t-1) &= f(y_t|y_1, \dots, y_{t-1}) - \sum_{i=1}^{t-2} f(y_t|y_1, \dots, y_{t-1}, r = i)f(r = i) \\
&= f(y_t|y_1, \dots, y_{t-1}) - \sum_{i=1}^{t-2} f(y_t|y_1, \dots, y_{t-1})f(r = i) \\
&= f(y_t|y_1, \dots, y_{t-1}) \left\{ 1 - \sum_{i=1}^{t-2} f(r = i) \right\} \\
&= f(y_t|y_1, \dots, y_{t-1}, r = j) \left\{ 1 - \sum_{i=1}^{t-2} f(r = i) \right\} \\
&= f(y_t|y_1, \dots, y_{t-1}, r = j)f(r \geq t-1).
\end{aligned}$$

Proof of Theorem 1.

(i) Missing non-future dependence implies non-future missing values.

Consider the ratio Q of the complete data density to the density involving only the previous and current measurements. This gives, under the missing non-future dependence assumption,

$$Q = \frac{f(y_1, \dots, y_T)f(r = i|y_1, \dots, y_{i+1})}{f(y_1, \dots, y_{i+1})f(r = i|y_1, \dots, y_{i+1})} = f(y_{i+2}, \dots, y_T|y_1, \dots, y_{i+1}). \quad (\text{A3})$$

Furthermore, one can always write,

$$\begin{aligned} Q &= \frac{f(y_{i+2}, \dots, y_T|y_1, \dots, y_i, r = i)f(y_1, \dots, y_{i+1}|r = i)f(r = i)}{f(y_1, \dots, y_{i+1}|r = i)f(r = i)} \\ &= f(y_{i+2}, \dots, y_T|y_1, \dots, y_{i+1}, r = i). \end{aligned} \quad (\text{A4})$$

Equating (A3) and (A4), we obtain

$$f(y_{i+2}, \dots, y_T|y_1, \dots, y_{i+1}, r = i) = f(y_{i+2}, \dots, y_T|y_1, \dots, y_{i+1}). \quad (\text{A5})$$

To show that (A5) implies the non-future missing values conditions (3), we proceed by induction on t . If $t = 2$, non-future missing values imposes no restriction and the result holds trivially.

Suppose by induction that non-future missing values holds for all $t \leq i$. We will prove the hypothesis for $t = i + 1$. Choose $j \leq i$, arbitrary but fixed. Then, from the induction hypothesis and Lemma A.1, it follows that, for all $j < t - 1 \leq i - 1$,

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

Taking the product over $t = j + 2, \dots, i$ then gives

$$f(y_{j+2}, \dots, y_i|y_1, \dots, y_{j+1}, r = j) = f(y_{j+2}, \dots, y_i|y_1, \dots, y_{j+1}). \quad (\text{A6})$$

After integration over y_{i+2}, \dots, y_T , equation (A5) leads to

$$f(y_{j+2}, \dots, y_{i+1}|y_1, \dots, y_{j+1}, r = j) = f(y_{j+2}, \dots, y_{i+1}|y_1, \dots, y_{j+1}). \quad (\text{A7})$$

Dividing (A7) by (A6) and equating the left- and right-hand sides, we find that

$$f(y_{i+1}|y_1, \dots, y_i, r = j) = f(y_{i+1}|y_1, \dots, y_i).$$

This holds for all $j \leq i - 1$, and Lemma A.1 shows that this is equivalent to non-future missing values.

(ii) Non-future missing values implies missing non-future dependence.

Starting from the non-future missing values assumption and Lemma A.1, we have that, for all $t \geq 2$ and all $j < t - 1$,

$$f(y_t|y_1, \dots, y_{t-1}, r = j) = f(y_t|y_1, \dots, y_{t-1}). \quad (\text{A8})$$

We now factorise the full data density as

$$\begin{aligned} f(y_1, \dots, y_T, r = i) &= f(y_1, \dots, y_{i+1}, r = i) f(y_{i+2}, \dots, y_T | y_1, \dots, y_{i+1}, r = i) \\ &= f(y_1, \dots, y_{i+1}, r = i) \prod_{t=i+2}^T f(y_t | y_1, \dots, y_{t-1}, r = i). \end{aligned}$$

Using (A8), we obtain that

$$\begin{aligned} f(y_1, \dots, y_T, r = i) &= f(y_1, \dots, y_{i+1} | r = i) f(r = i) \prod_{t=i+2}^T f(y_t | y_1, \dots, y_{t-1}) \\ &= f(y_1, \dots, y_{i+1} | r = i) f(r = i) f(y_{i+2}, \dots, y_T | y_1, \dots, y_{i+1}) \\ &= \frac{f(y_1, \dots, y_{i+1} | r = i) f(r = i)}{f(y_1, \dots, y_{i+1})} f(y_1, \dots, y_{i+1}) f(y_{i+2}, \dots, y_T | y_1, \dots, y_{i+1}) \\ &= \frac{f(y_1, \dots, y_{i+1} | r = i) f(r = i)}{f(y_1, \dots, y_{i+1})} f(y_1, \dots, y_T) \\ &= f(r = i | y_1, \dots, y_{i+1}) f(y_1, \dots, y_T). \end{aligned} \quad (\text{A9})$$

An alternative factorisation of $f(y, r)$ gives

$$f(y_1, \dots, y_T, r = i) = f(r = i | y_1, \dots, y_T) f(y_1, \dots, y_T). \quad (\text{A10})$$

It follows from (A9) and (A10) that

$$f(r = i | y_1, \dots, y_T) = f(r = i | y_1, \dots, y_{i+1}),$$

completing the proof of Theorem 1.

REFERENCES

DIGGLE, P.J. & KENWARD, M.G. (1994). Informative drop-out in longitudinal data analysis (with Discussion). *Appl. Statist.* **43**, 49–93.

- GLYNN, R.J., LAIRD, N.M. & RUBIN, D.B (1986). Selection modelling versus mixture modelling with nonignorable nonresponse. In *Drawing Inferences from Self-Selected Samples*, Ed. H. Wainer, pp. 115–42. New York: Springer Verlag.
- HOGAN, J.W. & LAIRD, N.M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statist. Med.* **16**, 239–58.
- KENWARD, M.G. & MOLENBERGHS, G. (1999). Parametric models for incomplete continuous and categorical longitudinal studies data. *Statist. Meth. Med. Res.* **8**, 51–83.
- LI, K.H., RAGHUNATHAN, T.E. & RUBIN, D.B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *J. Am. Statist. Assoc.* **86**, 1065–73.
- LITTLE, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Am. Statist. Assoc.* **88**, 125–34.
- LITTLE, R.J.A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–83.
- LITTLE, R.J.A. (1995). Modelling the drop-out mechanism in repeated-measures studies. *J. Am. Statist. Assoc.* **90**, 1112–21.
- LITTLE, R.J.A. & RUBIN, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- MOLENBERGHS, G., MICHIELS, B., KENWARD, M.G. & DIGGLE, P.J. (1998). Missing data mechanisms and pattern-mixture models. *Statist. Neer.* **52**, 153–61.
- REISBERG, B., BORENSTEIN, J., SALOB, S.P., FERRIS, S.H., FRANSSEN, E. & GEORGOTAS, A. (1987). Behavioral symptoms in Alzheimer’s disease: phenomenology and treatment. *J. Clin. Psychiat.* **48**, 9–13.
- RUBIN, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–92.
- RUBIN, D.B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *J. Am. Statist. Assoc.* **72**, 538–43.

- RUBIN, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- RUBIN, D.B. (1996). Multiple imputation after 18+ years. *J. Am. Statist. Assoc.* **91**, 473–89.
- RUBIN, D.B. & SCHENKER, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J. Am. Statist. Assoc.* **81**, 366–74.
- SCHAFER, J.L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.
- THIJS, H., MOLENBERGHS, G., MICHIELS, B., VERBEKE, G., & CURRAN, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3**, 245–265.
- VERBEKE, G. & MOLENBERGHS, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- WU, M.C. & BAILEY, K.R. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statist. Med.* **7**, 337–46.
- WU, M.C. & BAILEY, K.R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939–55.
- WU, M.C. & CARROLL, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–88.

Table 1: Sample size per treatment arm and drop-out pattern in the Alzheimer’s study.

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

Table 2: Inference for treatment contrasts for the case study. For the contrasts, parameter estimates and standard errors, in parentheses, are reported.

Pat.	Cont.	Selection Models		Pattern-mixture Models				
		MAR	MNAR	ACMV	CCMV	NCMV	FD1	FD2
Stratified Analysis								
1	1			9.27(6.42)	5.84(5.16)	-4.19(6.27)	4.90(8.29)	5.44(6.52)
	2			-8.19(6.58)	-6.92(6.15)	2.56(5.12)	-7.78(7.62)	-4.48(7.76)
2	1			2.78(4.75)	-0.00(2.90)	-4.43(3.54)	0.61(4.88)	-1.49(4.07)
	2			-3.57(4.53)	-5.08(3.92)	-1.37(4.12)	-6.48(5.22)	-4.54(5.46)
3	1			6.78(4.20)	6.95(2.66)	0.10(2.40)	4.18(2.64)	0.18(3.65)
	2			-1.75(2.76)	-3.44(2.12)	0.83(2.14)	-2.66(2.29)	-0.10(2.20)
4	1			11.05(3.21)	10.87(2.85)	6.59(3.09)	9.65(3.56)	9.97(2.90)
	2			-3.84(4.09)	-6.55(3.88)	-3.23(4.09)	-6.84(3.78)	-4.30(4.24)
5	1			0.15(5.71)	-2.05(6.29)	-5.60(6.46)	-3.02(5.92)	-6.13(6.42)
	2			-0.74(3.99)	-0.87(4.51)	0.92(4.68)	-0.53(4.24)	1.05(4.57)
6	1			14.16(3.75)	12.91(3.71)	13.44(3.72)	13.28(3.82)	12.72(3.79)
	2			-5.24(3.48)	-4.74(3.69)	-4.95(3.79)	-4.71(3.63)	-4.77(3.70)
7	1			-0.99(0.85)	-0.99(0.85)	-0.99(0.85)	-0.99(0.85)	-0.99(0.85)
	2			1.68(0.88)	1.68(0.88)	1.68(0.88)	1.68(0.88)	1.68(0.88)
<i>F</i> value				2.45	2.96	1.76	1.92	1.77
<i>p</i> value				0.0024	0.0002	0.0407	0.0225	0.0413
Marginal Analysis								
	1	0.55(0.71)	0.45(0.71)	1.97(1.05)	1.47(0.87)	-0.48(0.85)	1.05(1.04)	0.37(0.96)
	2	0.64(0.71)	0.69(0.71)	-0.24(0.81)	-0.56(0.86)	0.91(0.77)	-0.59(1.01)	0.19(0.84)
<i>F</i> value		2.82	2.68	2.15	1.23	0.66	0.52	0.19
<i>p</i> value		0.2446	0.2619	0.1362	0.3047	0.5208	0.6043	0.8276

Pat., pattern; Cont., contrast; MAR, missing at random; MNAR, missing not at random; ACMV, available-case missing values; CCMV, complete-case missing values; NCMV, neighbouring-case missing values; FD1, Case 4; FD2, Case 5.

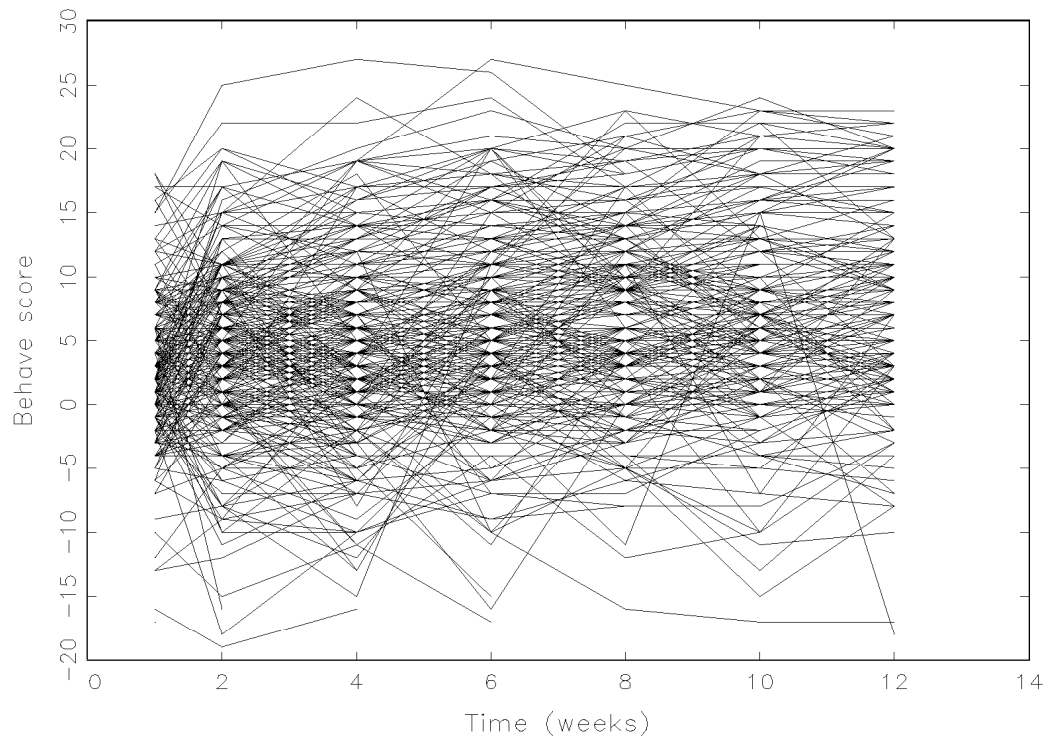


Figure 1: *Individual profiles in the case study.*

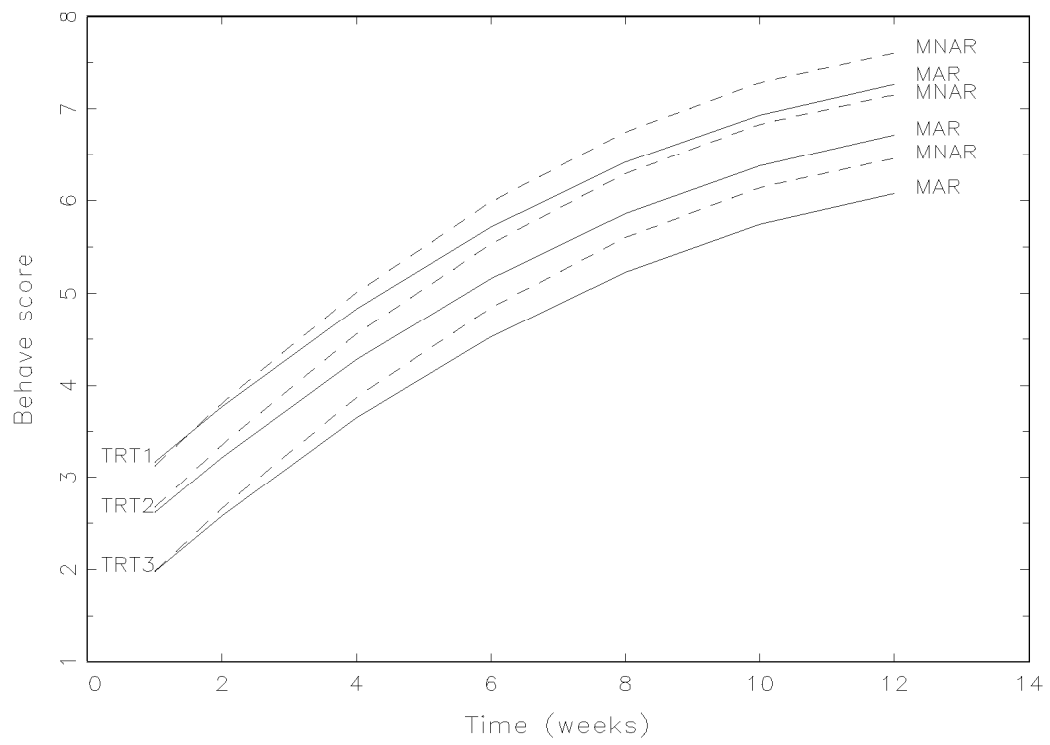


Figure 2: *Selection models in the case study. Fitted average profiles. MAR, missing at random; MNAR: missing not at random; TRT, treatment.*

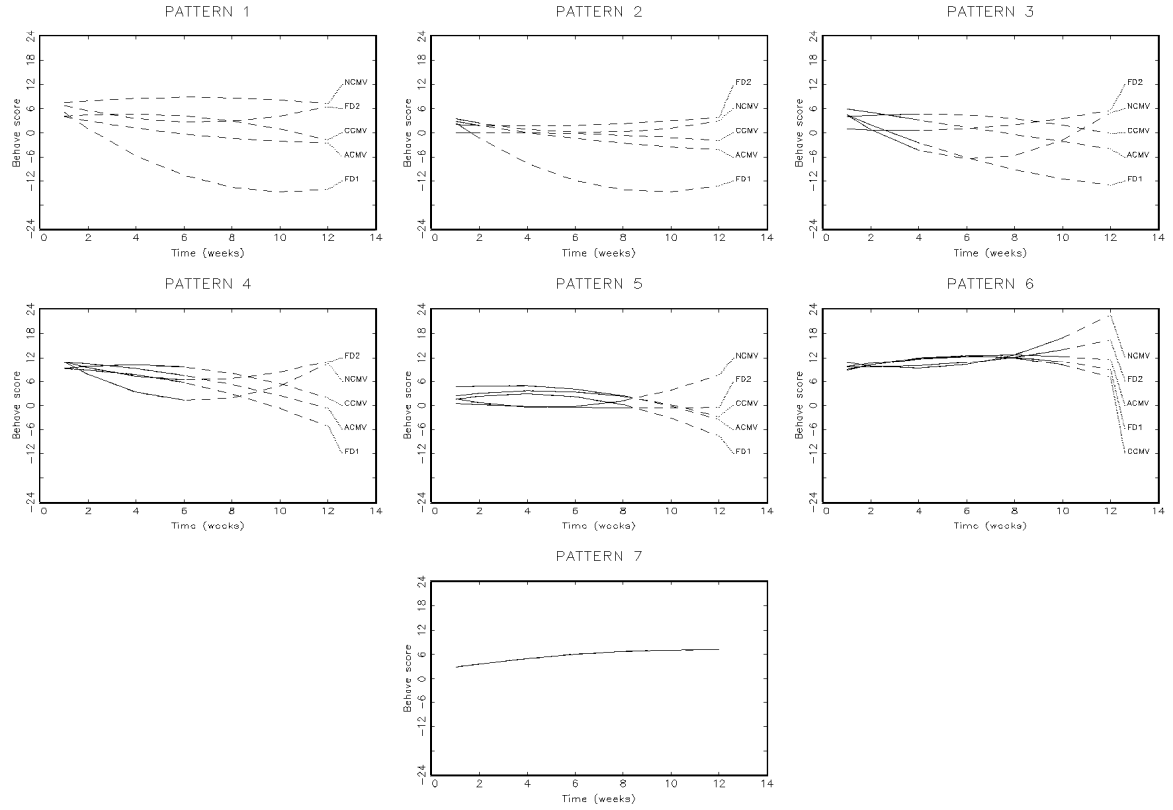


Figure 3: *Pattern-mixture models. Fitted average profiles for each of the five identification strategies. Treatment arm 1. ACMV, available-case missing values; CCMV, complete-case missing values; NCMV, neighbouring-case missing values; FD1, Case 4; FD2, Case 5.*

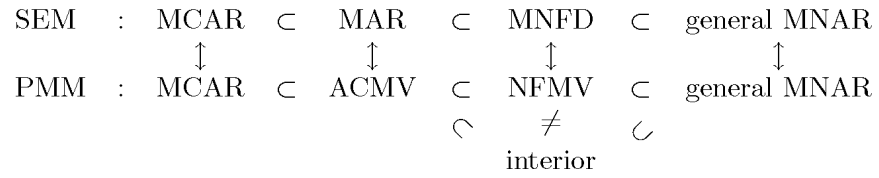


Figure 4: *Relationship between nested families within the selection model, SEM, and pattern-mixture model, PMM, families. MCAR: missing completely at random; MAR: missing at random; MNAR: missing not at random; MNFD: missing non-future dependence; ACMV: available-case missing values; NFMV: non-future missing values.*