

Cluster analysis for repeated data with dropout: Sensitivity analysis
using a distal event

Peer-reviewed author version

BRUCKERS, Liesbeth; MOLENBERGHS, Geert; Pulinx, Bianca; Hellenthal, Femina & Schurink, Geert (2018) Cluster analysis for repeated data with dropout: Sensitivity analysis using a distal event. In: JOURNAL OF BIOPHARMACEUTICAL STATISTICS, 28(5), p. 983-1004.

DOI: 10.1080/10543406.2018.1428612

Handle: <http://hdl.handle.net/1942/27627>

Cluster Analysis for Repeated Data with Dropout: Sensitivity Analysis Using a Distal Event

Liesbeth Bruckers^{*,1} Geert Molenberghs^{1,2}

Bianca Pulinx³ Femina Hellenthal³ Geert Schurink³

¹, I-BioStat, Universiteit Hasselt, Agoralaan, B3590 Diepenbeek, Belgium.

², I-BioStat, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B3000 Leuven, Belgium.

³, Department of Clinical Chemistry, Maastricht University Medical Center, Maastricht, The Netherlands.

* *E-mail:* liesbeth.bruckers@uhasselt.be

Acknowledgment

The authors gratefully acknowledge support from IAP research Network P7/06 of the Belgian Government (Belgian Science Policy). We are grateful to the Department of Clinical Chemistry, Maastricht University Medical Center (MUMC), Maastricht for kind permission to use the AAA data.

Abstract

Degeneration of the aortic wall becomes life-threatening when the risk of rupture in-

creases. Cluster analysis on repeated measures of the diameter of the artery revealed two subgroups of patients included in a surveillance program. These results were obtained under the assumption of missingness at random. In this article we study the vulnerability of the cluster analysis results - the estimated trajectories, and the posterior membership probabilities - by applying different missing-data models for non-ignorable dropout, as proposed by Muthén (2011), to the growth of the diameter of the artery.

Keywords: Distal event, Incomplete data, Latent class growth models; Pattern-mixture models; Selection models; Sensitivity analysis.

1 Introduction and Motivation

Often the primary endpoint in a clinical trial is the treatment effect at the end of the trial. However, usually the response is collected repeatedly over time and missing data are inevitable when collecting information about patients in a longitudinal way. The pattern and cause for this missingness can vary. Missingness can be non-monotone, i.e. intermittent missing observations, or monotone due to drop-out. The reasons for not obtaining data at all planned visits can be diverse. Adverse events, no response to study treatment, protocol violations, and loss to follow-up are frequently reported in clinical studies as reasons for dropout.

Incomplete records introduce specific challenges in the statistical analysis of the clinical trial data. Naturally, the loss of information has a negative impact on the precision and power. Traditionally, the analysis of clinical trials with incomplete records was a complete case analysis or a last observation carried forward analysis. These methods often result in biased estimators (Gibbons et al., 1993; Molenberghs and Verbeke, 2005; Carpenter et al, 2004; Beunckens et al., 2005). Nowadays, quite a number of methods are available to handle missing data in longitudinal trials (Molenberghs and Kenward, 2007). Likelihood-based analyses are valid when the missing-data mechanism is missing at random (MAR), in the sense that the mechanism describing missingness is allowed to depend on covariates and observed outcomes but, given these, not further on unobserved outcomes or unobserved covariates. However, the possibility of the more general missing not at random (MNAR) mechanism, where there is further dependence of missingness on unobserved outcomes, can never be definitively be excluded. Unfortunately, both MAR and MNAR rely on untestable assumptions.

Estimation of the treatment effect in clinical trials is often further complicated by the fact that the degree patients respond to the treatment is different. Sometimes this heterogeneity between patients can be explained by known factors, e.g. disease severity or genetics. However, often the cause for this variability in treatment response is not known. Patients belong to different subpopulations that can not directly be observed, e.g. patients fully benefiting, partially benefiting or not benefiting at all from the treatment. As a consequence the estimated treatment effect is a mixture of the latent subpopulation's specific treatment effects. Statistical techniques used to reveal latent subgroups of similar patients, are termed cluster analysis. Cluster methods were first studied for cross-sectional multivariate data, but over the last decades a lot of methodological work has been done to extend cluster analysis such that it can cope with complex data structures, in particular repeated measures. Repeated measurements data are often analyzed via linear mixed models (Laird and Ware, 1982; Verbeke and Molenberghs, 2000), that incorporate patient specific evolutions, such as initial stage and rate of change, through random effects. The underlying assumption of a single-component multivariate normal distribution for the random effects implies that the patients belong to one (homogeneous) population. This assumption may be unrealistic when different subpopulations exists. This unobserved heterogeneity can be incorporated into the statistical model. Growth mixture modeling (Section 3), for example, uses categorical as well as continuous latent variables to represent individual differences in evolution (Muthén and Shedden, 1999; Muthén and Muthén 1998, 2007).

As is the case in a likelihood-based analysis of longitudinal data, the standard procedure of using all available data in a growth mixture model when there are missing data is only valid if certain untestifiable assumptions hold. Therefore, it is desirable to perform a sensitivity analysis.

In this paper, we introduce missing-data models when clustering of longitudinal profiles with missing information, by means of a growth-mixture model, is of interest.

Muthén et al. (2011) have provided an excellent overview of these recent innovations. We demonstrate the use of these missing-data models as a tool to study the vulnerability of the cluster analysis results. Where Muthén et al. focussed on the comparison of the results of the different models in terms of the optimal number of clusters and the estimated trajectories; we will complement the model comparison with an evaluation of the sensitiveness of the posterior membership probabilities under the different missing-data models for non-ignorable dropout.

The paper is organized as follows. The AAA data, motivating this research and used to illustrate

the different models, are presented in Section 2. Growth mixture models are introduced in Section 3. Section 4 introduces the models for analyzing longitudinal data in the presence of missingness, the missing at random and the missing not at random growth mixture models are briefly introduced. A sensitivity analysis using a distal event is presented in Section 4.3.3. A distal event is an outcome that relates to the trajectories, that can be viewed as a indicator of the clusters, that contains information about the cluster membership and information about the reasons for dropout. For the AAA study the operational status of a patient is used as a distal outcome. Section 5 discusses the results when fitting these models to the AAA data.

2 Introduction to the Data

Abdominal aortic aneurysm (also known as AAA) is a localized dilatation of the abdominal aorta, caused by degeneration of the aortic wall. Abdominal aortic aneurysms occur most commonly in individuals between 65 and 75 years old and are more common among men and smokers.

As abdominal aortic aneurysms expand, they may become painful and lead to pulsating sensations in the abdomen or pain in the chest, lower back, or scrotum. The risk of rupture is high in a symptomatic aneurysm, which is therefore considered an indication for surgery. Rupture of the artery can be life-threatening as large amounts of blood spill into the abdominal cavity. The mortality of AAA rupture is up to 90%. Sixty-five to 75% of the patients die before they arrive at the hospital and up to 90% die before they reach the operating room. Therefore, symptomatic and large aneurysms are considered for repair by surgical methods. An intervention is often recommended if the aneurysm grows more than 1 cm per year or when it is bigger than 5.5 cm.

In 2006, the academic hospital of Maastricht (the Netherlands) started a follow-up study in patients with an abdominal aorta aneurysm. Between January 2006 en January 2009, all patients with AAA admitted to the department of Vascular Surgery of the academic hospital were invited to participate in the study. Two hundred and eighty-seven AAA patients provided written informed consent. Patients that had a large aneurysm (≥ 55 mm) or (symptoms of imminent) AAA rupture and patients with either an inflammatory or a mycotic aneurysm were excluded from the follow-up study. Patients with an aneurysm diameter between 30 and 55 mm ($n=110$) were invited to participate in an imaging surveillance program. A total of 100 patients formally entered the follow-up program. These patients were seen every 6 months. During these follow-up visits

the diameter of the artery, a number of patient characteristics and blood measurements were collected. Figure 1 shows the diameter curves for the patients in the study. The objective of the surveillance program was two-fold. Firstly, to see if it is possible to predict the diameter of the artery that could be expected at the next follow-up visit? And secondly, to investigate whether subgroups (clusters or classes) of patients whose arteries grow in a similar way can be detected. To study the evolution of the diameter over time and to find clusters of patients with similar growth, the study researchers applied conventional growth models and growth mixture models (Hellenthal et al., 2014). The analysis revealed two subgroups of patients. This growth mixture model is presented in Section 3.

However, for a number of patients the diameter of the artery is not available at all follow-up visits (Table 1). Only 7 patients had complete data for all 7 visits, 11 dropped out after the first visit (at 6 months of follow-up), 15 after the second visit, 21 after the third visit, 7 after the fourth visit, 27 after visit five and finally 12 patients dropped out after visit 6. The diameter of the artery at visit j can be missing because the patient was not yet $j \times 6$ months in the study, or because the patient did not show up at the visit, or because the patient underwent repair of the artery by surgery, etc. It cannot be excluded that the reason for missing a follow-up visit depends on the diameter of the artery. The dropout rate is very comparable among the patients that had surgery of the artery and those who did not (92% versus 93%). Fifty-six percent of the planned measurements were observed.

In the presence of missing data, the growth model and growth mixture model give valid results only when the mechanism describing missingness is missing at random. The possibility of more general missing data mechanisms can never be definitively excluded. It is thus advisable to perform a sensitivity analysis for the results presented in Hellenthal et al. (2014), by fitting different missing-data models to the data.

In this paper, we study the vulnerability of the cluster analysis results, i.e., the estimated trajectories, and the posterior membership probabilities, by applying different missing-data growth-mixture models for non-ignorable dropout to the evolution of the diameter of the artery. These MNAR-models will be extended by including the information whether or not the patient needed repair of the artery by means of a surgical method. The models are introduced in Sections 3 and 4. All analyses were performed with the MPLUS software.

3 Growth Mixture Models

This section introduces the concepts of finite-mixture models and their application to repeated-measurements data. Finite-mixture models (McLachlan and Peel, 2000) are latent-variable models that express the distribution of a variable as a mixture of a finite number of component distributions. These models have been used in a wide range of applications in marketing, social, psychosocial and medical sciences, where the data could be seen as arising from two or more populations. Mixture models provide a framework for clustering by addressing the population heterogeneity in the observed data by means of categorical latent classes, that represent homogeneous subpopulations. Class-membership is latent (not observed) and thus needs to be inferred from the data. Mixture models have been applied to longitudinal settings too. Consider the observed variables \mathbf{x}_i and \mathbf{y}_i where \mathbf{x}_i is a $q \times 1$ vector of covariates and \mathbf{y}_i a $p \times 1$ vector of continuous outcome variables for individual i ($i : 1, \dots, N$). In addition, consider latent variables $\boldsymbol{\eta}_i$, organized as a $m \times 1$ vector of continuous variables and \mathbf{c}_i a latent categorical variable with K classes, $\mathbf{c}_i = (c_{i1}, c_{i2}, \dots, c_{iK})'$, such that $c_{ik} = 1$ if individual i belongs to class k and zero otherwise. In a general latent variable model (Muthén and Shedden, 1999), the latent classes \mathbf{c}_i influence \mathbf{y}_i via the following model:

$$\mathbf{y}_i = \boldsymbol{\Lambda} \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \quad (1)$$

$$\boldsymbol{\eta}_i = \mathbf{A} \mathbf{c}_i + \boldsymbol{\Gamma} \mathbf{x}_i + \boldsymbol{\zeta}_i, \quad (2)$$

where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are $p \times m$ and $m \times q$ matrices of parameters, \mathbf{A} is a $m \times K$ matrix containing intercept parameters for each class, $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\zeta}_i$ are normally distributed residual vectors, $N(\mathbf{0}, \boldsymbol{\Theta}_k)$ and $N(\mathbf{0}, \boldsymbol{\Psi}_k)$, respectively. The latent categorical variable \mathbf{c}_i can be related to the covariates \mathbf{x}_i via a multinomial logistic regression, $\text{logit}[pr(c_{ik} = 1 | \mathbf{x}_i)] = \boldsymbol{\delta}_k + \boldsymbol{\kappa}_k \mathbf{x}_i$. Viewing the \mathbf{c}_i and $\boldsymbol{\eta}_i$ and as missing data, the complete-data log likelihood for a sample of size N is given by $\sum_{i=1}^N [\log f(\mathbf{c}_i | \mathbf{x}_i) + \log f(\boldsymbol{\eta}_i | \mathbf{c}_i, \mathbf{x}_i) + \log f(\mathbf{y}_i | \boldsymbol{\eta}_i)]$, where $f(\cdot)$ represents either a probability or a density function. Parameter estimates for this model can be obtained using maximum likelihood with an EM algorithm (Dempster, Laird and Rubin, 1977). Assignment of the individuals to the subpopulation they belong to, is done based on a individual's posterior probabilities. This is the conditional probability of individual i to belong to subpopulation k , given the observed data,

$pr(c_{ik} = 1 | \mathbf{y}_i, \mathbf{x}_i) = pr(c_{ik} = 1 | \mathbf{x}_i) f(\mathbf{y}_i | \mathbf{c}_i, \mathbf{x}_i) / f(\mathbf{y}_i | \mathbf{x}_i)$. The individual is classified into the subpopulation for which its posterior group-membership probability is largest.

To quantify the accuracy of model-based clustering a standardised summary measure, the entropy, can be used. Entropy ranges from 0 to 1. When the posterior probabilities of class membership do not vary across classes the entropy is 0 (no class separation). A value of 1 points to perfect class separation, this means that for any individual the posterior probability for membership in one specific class equals 1 (Ramaswamy et al., 1993).

4 Incomplete Longitudinal Data

We will use the terminology of Rubin (1976) and Little and Rubin (2002) to distinguish between the different missing-data mechanisms. A process is said to be missing completely at random (MCAR) if the missingness is independent of both the unobserved and observed data. The process is missing at random (MAR) if, conditional on the observed data, the missingness is independent of the unobserved measurements. If the process is neither MCAR nor MAR it is missing not at random (MNAR). The process then depends on the unobserved measurements. Traditional techniques for analyzing incomplete data, such as complete case analysis and last observation carried forward, work well only in a limited set of circumstances and are generally prone to bias. Even when these techniques are unbiased, they tend to be less efficient than modern missing-data methods, such as direct maximum likelihood (DML) and multiple imputation (MI). DML and MI in general make weaker, and probably more realistic, assumptions about the missing data. Consequently, they should produce efficient and broadly unbiased parameters estimates.

4.1 Missing Data Modeling Frameworks

In case of missing data in the repeated measurements \mathbf{y}_i , one needs to specify the full-data likelihood. The full-data likelihood considers not only the repeated measurements, observed and unobserved, as data but also the missing-data indicators for each time point. The missing-data indicators for subject i , \mathbf{m}_i , are binary variables with values for m_{it} equal to 1 if the repeated measurement at time t , y_{it} , is missing and 0 if y_{it} is observed. The observed repeated measurements are grouped in $\mathbf{y}_i^{\text{obs}}$ and the missing measurements in $\mathbf{y}_i^{\text{mis}}$ such that $\mathbf{y}_i = (\mathbf{y}_i^{\text{obs}},$

$\mathbf{y}_i^{\text{mis}}$). The full-data density then becomes $f(\mathbf{y}_i, \mathbf{m}_i | X_i, W_i, \boldsymbol{\theta}, \boldsymbol{\psi})$, with $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ the parameter vectors associated with the measurement process and the missingness process, and X_i and W_i the associated design matrices. Different modeling frameworks for incomplete longitudinal data are obtained by factorizing this full-data density in different ways. This results in selection, pattern-mixture and shared-parameter models. Details about these model can be found in Section 7.1 in the appendix.

The interpretation of the results depends on the chosen framework and different results and conclusions can be obtained. Verbeke et al. (2010) show that always a part of such a model is unidentifiable from the observed data. By replacing these unidentified parts, which rest completely on assumptions, an entire class of models, all producing the same fit to the observed data, can be obtained. More about latency and unidentifiability can be found in the appendix (Section 7.2). To investigate how sensitive the results are to the assumptions made in the different missing-data models, a sensitivity analysis should be conducted. The simplest procedure is to fit a selected number of (MNAR) models which seem plausible. The degree to which conclusions are stable across these models provides an indication of the confidence that is placed in them.

Section 4.2 describes the traditional growth mixture model, assuming MAR, and Section 4.3 presents the missing-data growth-mixture models considered in the sensitivity study for the AAA data.

4.2 Missing at Random Growth Mixture Model for Incomplete Data

In the presence of missing data, the full-data density for a mixture model is given as (where for simplicity of notation dependence of covariates is suppressed): $f(\mathbf{y}_i, \mathbf{m}_i) = \sum_{k=1}^K f(\mathbf{y}_i, \mathbf{m}_i | c_{ik} = 1) f(c_{ik} = 1)$, with $c_{ik} = 1$ when individual i belongs to the k th mixture and zero otherwise. Factorizing this full-data density according to the selection model and assuming MAR for the missing-data mechanism conditional on class, implies the following observed-data likelihood for a growth-mixture model:

$$f(\mathbf{y}_i^{\text{obs}}, \mathbf{m}_i) = \sum_{k=1}^K f(\mathbf{y}_i^{\text{obs}} | c_{ik} = 1) f(\mathbf{m}_i | \mathbf{y}_i^{\text{obs}}, c_{ik} = 1) f(c_{ik} = 1). \quad (3)$$

Thus the standard procedure, when fitting a growth mixture model when facing missingness, to use all available data is only valid when, (1) MAR holds conditional on class (resulting in decomposition (3)), and (2) the missing-data mechanism remains the same for the different classes given the observed data, i.e., $f(\mathbf{m}_i | \mathbf{y}_i^{\text{obs}}, c_{ik} = 1) = f(\mathbf{m}_i | \mathbf{y}_i^{\text{obs}})$, and (3) the first two terms in (3) do not share parameters. Under these assumptions, maximization of the first term of the likelihood is sufficient and the missing-data mechanism can be ignored. When assumption (2) is violated ignorability does not hold for the mixture distribution of $f(\mathbf{y})$. The term $f(\mathbf{m}_i | c_i, *)$ can not be ignored in the EM algorithm. Missingness depends on a latent variable (in the above situation on c_i) leading to a non-ignorable missing-data situation.

The assumption of MAR leads to a convenient simplification of the likelihood, but in general one cannot justify this assumption. Often one is forced to make assumptions about the missing-data mechanism, which cannot be underpinned with the observed data alone.

4.2.1 Missing at Random Model Growth Mixture Model for AAA data

Within the time range of the follow-up, the diameter of the artery exhibits a linear growth (Figure 1). Thus, a linear growth model will be assumed. For class k :

$$y_{it} | c_{ik}=1 = \eta_{0i} + \eta_{1i}t + \varepsilon_{it}, \quad (4)$$

where c_i is the latent-class variable. The random-effects distributions vary as a function of the class k , $\eta_{0i} | c_{ik}=1 = \alpha_{0k} + \zeta_{0i}$ and $\eta_{1i} | c_{ik}=1 = \alpha_{1k} + \zeta_{1i}$.

The residuals are assumed to be normally distributed with zero mean. The ζ_i have a 2×2 covariance matrix D , ε_i is a 7×7 diagonal matrix Σ_k and the error variances are assumed to be constant over time.

Model (4) reflects the first term in (3). This model provides unbiased estimates when in addition to separability, MAR conditional on the latent class holds and missingness is not influenced by the class.

4.3 Missing Not at Random Growth Mixture Models for AAA data

This section presents the MNAR growth mixture models that were applied to the AAA data. Different modeling frameworks have been proposed to jointly model the outcomes and the missing-data process $f(\mathbf{y}_i, \mathbf{m}_i)$. We will in turn discuss selection and pattern-mixture models. The conventional models are presented in the Appendix (Section 7.1) and the models incorporating a latent-class variable c_i are presented here.

4.3.1 Selection Modeling

The Diggle-Kenward model, where the probability of missing data at time t depends directly on the repeated measurements at time t as on the previous measurement, can be extended to a mixture model by specifying the following logistic model for the missing-data indicators, conditional on the data:

$$\text{logit}(m_{it}|y_{it}, y_{i,t-1})|_{c_i=1} = \alpha_{tk} + \beta_{1k}y_{it} + \beta_{2k}y_{i,t-1}, \quad (5)$$

The logistic regression coefficients are allowed to vary across the latent classes. This can be important when patients in high-trajectory classes tend to drop out because of large diameters, and patients in the low-trajectory class tend to drop out because of small diameters of the artery. When the influence of y_{it} and $y_{i,t-1}$ on the dropout probability varies across time, the model can further be generalized:

$$\text{logit}(m_{it}|y_{it}, y_{i,t-1})|_{c_i=k} = \alpha_{tk} + \beta_{1tk}y_{it} + \beta_{2tk}y_{i,t-1}. \quad (6)$$

To lower the number of parameters in the model, α_{tk} , β_{1tk} , and β_{2tk} could be specified to vary as a linear function of time.

An alternative model was proposed by Beunckens et al. (2008). Their model combines features from a selection model and a shared-parameter model. Dropout is influenced by the latent class variable and the random effects. We specified the dependence of dropout on latent class as a linear function of time t and as a function of the random intercept η_{0i} .

$$\text{logit}(m_{it}|c_{ik} = 1, \eta_{0i}) = \alpha_k + \beta_{1k}t + \beta_{2k}\eta_{0i}. \quad (7)$$

The selection-model features of these models are given by the (random effects of the) outcome process influencing the dropout indicators. Given that the dropout indicators and the diameters outcomes are both influenced by the latent class variable (and the random intercept), these models also are of the shared-parameter type. Selection models are inherently inestimable, by imposing assumptions for the repeated measures (in the Diggle-Kenward model) or the random effects (Beunckens model), typically multivariate normal, identification is obtained.

4.3.2 Pattern-mixture Modeling

A pattern-mixture model (Little 1993, 1994) specifies the density as a mixture over different populations, each defined by the observed pattern of missingness. This can be obtained by allowing the random effects of the growth model to vary as a function of the missing data indicators. For each dropout pattern, a linear growth model was specified.

$$y_{it}|m_{i1}, \dots, m_{iT} = \eta_{0i} + \eta_{1i}t + \varepsilon_{it}, \quad (8)$$

and

$$\begin{aligned} \eta_{0i}|m_{i1}, \dots, m_{iT} &= \alpha_0 + \sum_{t=1}^T \beta_{0t}m_{it}, \\ \eta_{1i}|m_{i1}, \dots, m_{iT} &= \alpha_1 + \sum_{t=1}^T \beta_{1t}m_{it}. \end{aligned} \quad (9)$$

An alternative model has been proposed by Roy (2003). Where in the conventional pattern-mixture models the distributions for the different dropout patterns are mixed, Roy's model considers a latent class defined by the missing data indicators, so that $f(\mathbf{y}|\mathbf{m}) = \sum_c f(\mathbf{y}|\mathbf{c})f(\mathbf{c}|\mathbf{m})$. Latent class membership is specified as a multinomial logistic regression model, with the missing data indicators as explanatory variables. The random effects of the measurement outcomes are influenced by this latent class variable.

A generalization to this model was proposed by Muthén (2011). Two latent categorical variables are used. One latent class provides information about the outcome trajectory classes, the second latent class is based on the dropout pattern. We will refer to this model as the Roy-Muthén model. Define a latent class variable c_d for the L dropout classes and a latent class variable c_y

for K trajectory classes for the outcome \mathbf{y} . The model is specified by the following equations.

$$\begin{aligned} y_{it}|(c_{d_{il}} = 1, c_{y_{ik}} = 1) &= \eta_{0i} + \eta_{1i}t + \varepsilon_{it}, \\ \eta_{0i}|(c_{d_{il}} = 1, c_{y_{ik}} = 1) &= \alpha_{0lk} + \zeta_{0i}, \\ \eta_{1i}|(c_{d_{il}} = 1, c_{y_{ik}} = 1) &= \alpha_{1lk} + \zeta_{1i}. \end{aligned} \tag{10}$$

A bivariate-loglinear model is used to specify class membership, $\text{pr}(c_{d_{il}} = 1, c_{y_{ik}} = 1 | m_{i1}, \dots, m_{iT})$. Both latent classes influence the random effects. However, \mathbf{c}_y is not influenced by the missing data indicators. So, \mathbf{c}_y represent different trajectories which are moderated by \mathbf{c}_d within each class of \mathbf{c}_y . Hence, $f(\mathbf{y}|\mathbf{m}) = \sum_{\mathbf{c}_y, \mathbf{c}_d} f(\mathbf{y}|\mathbf{c}_y, \mathbf{c}_d)f(\mathbf{c}_d|\mathbf{m})f(\mathbf{c}_y)$

Pattern-mixture models are over-specified models. The model is identified only for the subgroup of patients with complete data. To solve this problem the inestimable parameters of the incomplete patterns are set equal to parameters describing the distribution of the patients with complete data. In the pattern-mixture models described, the slope of the time effect for patients dropping out after the first visit was set equal to the slope of patients dropping out after the second visit.

4.3.3 Growth Mixture Model Using a Distal Event

In this section, we will study the ability of the MNAR models discussed before, to gauge (possible) non-random missingness by studying the relationship between the latent class membership and a distal event. The AAA study contains useful auxiliary information. For each patient, it is known whether or not s/he underwent surgery. Symptomatic and large aneurysms are considered for repair by surgical methods when the aneurysm grows more than 1 cm per year or when it is bigger than 5.5 cm. This endpoint is called a distal outcome. This outcome contains information about the reason for dropout. The models discussed in Sections (4.2) and (4.3) can be compared in terms of how they relate to this distal outcome. If the models of Sections (4.2) and (4.3) are able to capture non-random missingness one expects congruence between the classes revealed when not including and when including this distal outcome in the model.

Muthén and Shedden (1999) describe an extended finite mixture model that allows joint estimation of (1), a conventional finite growth mixture model where different curve shapes are

captured by class-varying random coefficient means and (2) a logistic regression of the distal event on the classes. Let u_i be a categorical distal outcome. The extended finite mixture model relates the observed variables, x_i , y_i and u_i to each other via the unobserved latent variables η_i and c_i . Equation (1) relates the observed variables y_i to the continuous latent variables η_i and Equation (2) specifies the relation between these continuous latent variables and the categorical ones c_i . The latent categorical variable c_i , representing the mixture components, can be related to the covariates x_i via a multinomial logistic regression, $\text{logit}[pr(c_{ik} = 1|x_i)] = \delta_k + \kappa_k x_i$. A logistic regression model is used to link the latent class c_i to the distal outcome u_i , $\text{logit}[pr(u_i = 1|c_{ik})] = \alpha + \beta_k c_{ik}$. The observed-data likelihood for this mixture is given as: $f(y_i, u_i|x_i) = \sum_c f(c_i|x_i)f(\eta_i|c_i, x_i)f(y_i|\eta_i)f(u_i|c_i)$. Parameter estimates are obtained using the EM algorithm, where the latent variables η_i and c_i are treated as missing data.

The two-class models for the AAA study, discussed in Sections (4.2) and (4.3), are extended with this distal event. A logistic regression model for the operational status was used. So, $u_i=1$ for patients that underwent surgery and $u_i=0$ for patients that did not have an operation. The probability to undergo surgery was allowed to be different for the classes by adding one additional parameter to the models.

5 Results for the AAA study

5.1 Missing Data in the AAA Study

Of the 100 patients included in the AAA study only 7 patients had complete data for all 7 visits. The amount of missingness is described in Section 2. The drop-out rate is high (more than 90%). However, since patients mostly drop out towards the end of the surveillance study 56% of the planned measurements were observed.

Figure 2 shows the mean diameter curve for the patients operated and not operated. The figures also include the mean diameter among those who did and did not drop out at the next visit. Given the small sample sizes and the illustrative purpose of Figure 2 no formal testing for the differences was executed. To supplement the figure, Table 2 displays the average diameter and 95% confidence intervals for operated and not operated patients who do or do not drop out at the

next study visit. The graph suggests that among the group of patients undergoing surgery, at the beginning of the study patients with small diameter drop out (although the number of patients is small, see Table 2). At the start of the study, the average diameter for patients dropping out at visit 2 equals 38.0mm compared to 48.6mm for patients not dropping out. Patients dropping out later in the study are the ones with larger artery diameter. At visits 4 and 5, patients dropping out at the next visit have an average diameter of 60.0/57.7mm compared to 53.8/52.4mm for patients not dropping out at the next visit. For the group of patients not having surgery the mean artery diameter of patients dropping out varies with time in a non systematic way. The difference in sample mean at visit t between patients dropping out at visit $t + 1$ and patients not dropping out is relatively small.

Figure 3 shows, per dropout pattern, the mean diameter curve. To quantify the evolution of the diameter for the different drop-out patterns, a linear mixed model was applied. The average diameter at the first visit and the rate of increase of the artery diameter per drop-out pattern are displayed in Table 3. Patients not dropping out or only at the last visit (visit 7) have smaller diameters than patients dropping out around visits 5 and 6. For example, the average diameter for patients dropping out at visit 7 equals 38.3mm versus 44.0mm for the patients dropping out at visit 5. The increase in the diameter of the artery for patients dropping out of the study at visit 5 seems to be faster (an increase of 1.7mm per 6 months compared to an increase of 0.9mm for the patients dropping out at visit 7). Differences between these drop-out pattern specific evolutions were not formally investigated for statistical significance.

The results when applying the models discussed in Section 4 to the AAA data are presented in the next section. For all models 1–4 classes are studied. However, for the following multi-class models convergence was not obtained: 4-class Diggle-Kenward Model (6), 4-class Beunckens Model and Roy-Muthén models with more than 3 trajectory classes or more than 2 dropout classes. The reason for the non-convergence is the increasing model complexity. The number of parameters in the model increases with increasing number of classes, and at the same time the number of patients per class decreases.

5.2 Missing at Random Model

For the missing at random model, the results are presented in Table 4. It is clear that the 1-class model is performing worse than the multi-class models. Based on the BIC (Schwarz, 1978) the two-class model is preferable. The parameter estimates will be unbiased if MAR, conditional on class, holds, c does not influence the missingness and ignorability holds.

Based on the intercept and slope factors (Table 8), these classes could be labeled: (1) 'small diameter and slowly growing' (intercept 38.3mm and slope 1.3mm) and (2) 'large diameter and faster growing'. (intercept 46.0mm and slope 2.9mm) The variances of the intercepts are relatively large, indicating that even within a class there still is heterogeneity. This heterogeneity seems to be more pronounced in class 2 (variance of random intercept equals 30.1 for class 2 versus 23.1 for class 1). The same is seen in terms of the random slope (variance of random slope equals 4.4 for class 2 versus 0.6 for class 1). So, patients of class 2 differ more in their growth than patients of class 1. For both classes, the covariance between the random intercept and slope is not statistically significant, indicating that, conditionally on class-membership, the value of the diameter at the start of the study and the growth are independent.

The two-class GMM solution estimated under MAR is shown in Figure 7.2. The entropy is 0.64. Forty percent of the patients are in the low class showing slow growth of the artery. Sixty percent are in the high class with faster growing arteries. The median posterior membership probability (Table 6) equals 0.98 (0.93) for patients in the low (high) class.

5.3 Missing Not at Random Models

Results of the selection models are presented in Table 4. Multi-class models were not always converging. The number of parameters in the model becomes large in comparison with the number of patients in the study. Diggle-Kenward Model (6) gave problems for the parameters in the drop-out model. The Beunckens model specifying 4 classes, results in a class with only one patient. Based on the BIC statistic the 1-class Diggle-Kenward (DK) model is best. The two-class Beunckens model is comparable. The estimated mean trajectories for the two-class models are shown in Figure 7.2. The class of patients with small diameters is estimated to be 70% under the Diggle-Kenward Model (5), 43% under the Diggle-Kenward Model (6) and 66% in the Beunckens model.

Results of the pattern-mixture models are presented in Tables 4 and 5. In terms of BIC, the two and three class Roy and Roy-Muthén models outperform the traditional pattern-mixture model. The two-class Roy model is preferable. The estimated mean trajectories for the two-class Roy model are shown in Figure 7.2. The class of patients with small diameters is estimated to be 44% under the Roy Model. Table 8 presents for the different two-class models the parameter estimates for the measurement model.

5.4 Model Comparison: Sensitivity of the Results

5.4.1 Posterior Probabilities

Table 6 displays, for the 2-class models presented in Sections 5.2 and 5.3, the median, 10th, and 25th percentile of the posterior membership probabilities. For all models the median posterior membership probability is at least 0.90, the P25 at least 0.80. These high values indicate that for a given model there is good class separation.

Figure 4 shows the posterior probabilities to be classified into the group with the largest and fastest increasing diameters. The graph clearly shows that for a number of patients this probability can change substantially, depending on the dropout model that was chosen.

Table 7 shows the agreement in classification in low versus high class for the different models. The table presents the Kappa statistic. Cohen's kappa coefficient is a statistical measure of inter-rater agreement for categorical items. This measure takes into account the agreement occurring by chance. According to Fleiss (1981), kappas over .75 characterize excellent agreement, values between .40 to .75 indicate fair to good agreement, and kappas below .40 point to poor agreement. These guidelines are, however, arbitrary. Based on these guidelines, we conclude that the classification according to MAR and Roys model are very similar ($\kappa = 0.877$); also Diggle-Kenward Model (5) and Beunckens' model ($\kappa = 0.862$) result in very similar classifications of the patients.

Figure 4 also reveals how certain the classification in the low or high group is. Posterior probabilities between 0.45 and 0.55 indicate that it is uncertain to which group the patient should be classified. Some patients have posterior probabilities in this region, and therefore it is expected that they can switch group when changing dropout model. However, for a number of patients,

the classification according to MAR, Diggle-Kenward Model (6) and Roy's model is clearly into the high group (these patients have posterior probabilities > 0.6 to belong the high group) where, according Diggle-Kenward Model (5) and Beunckens model, the same patients are clearly in the low group (as indicated by posterior probabilities to belong to the high group < 0.3).

As a result of the susceptibility of the posterior probability to the chosen dropout model, it is seen that the number of patients classified in the low or high class, based on their most likely latent-class membership, strongly depends on the assumptions made. Under the MAR assumption 40% of the patients are classified in the lowest latent class, under the Diggle-Kenward Model (5) this increases to 70% (Table 10).

Sixty patients out of a total of 100 are always classified into the same latent class, irrespective of the assumptions made for the missing-data mechanism. Using the classification resulting from the MAR model, we see that the effect of the chosen missing-data mechanism on the posterior probability is smaller for the patients in the low group as compared to the ones in the high group. Of the 60 patients belonging to the group with larger diameter, only 26 (43%) are always classified into the high group for the selection and pattern-mixture models under consideration. Patients classified in the low group almost always (34 out of 40, 85%) end up in the low group.

Figure 5 graphically shows the relationship between the random slope and random intercept. Empirical Bayes estimates for the random effects were obtained for the two-class MAR model. The figure shows this relationship by dividing the patients into two groups. A group of patients that, irrespective of the chosen dropout model, is always classified into the high or low class and a second group of patients whose classification does depend on the assumption made for the missing-data mechanism. For the patients always classified in the same latent class, i.e., always in the low or always in the high class, the Spearman correlation coefficient between the random slope and intercept equals 0.73 ($p < 0.001$). For the patients whose classification depends on the chosen dropout model the correlation equals -0.17 ($p=0.285$). It is thus seen that the relation between a patient's intercept and slope goes in the reverse direction for the two groups. The second group consists of patients with small and fast growing arteries and patients with large but slowly growing arteries.

5.5 Parameter Estimates

Table 8 presents for the different two-class models the parameter estimates for the measurement model. The choice of the dropout model also influences the parameter estimates for the fixed effects. Compared with the model assuming MAR, the largest effect on the intercept and slope is seen in Diggle-Kenward Model (5) and the Beunckens Model (Table 8). The differences in the fixed effects between the models is most pronounced for the latent class with largest and fastest increasing diameters. The Diggle-Kenward Model (5) and the Beunckens Model result in an average diameter at the start of the study for the highest class equal to about 50mm compared to 46mm in the MAR model. The slope for the highest class for Diggle-Kenward Model (5) is 3.9mm per 6 months compared to 2.9mm in the MAR model.

The class-specific mean evolution for the two latent classes, under the different missing-data mechanisms, are graphically presented in Figure 7.2. The variance of the random intercept for the high class in the Diggle-Kenward Model (5) and in the Beunckens model is much smaller than this variance component in the other models, indicating that these models result in more homogeneous latent classes in terms of the arteries diameter at the start of the study. The other variance components (variance of the random slope, covariance of random intercept and slope, and the residual variance) under the different models did not change substantially.

Choosing between these models is difficult. All these models contain unidentifiable parts, which rely completely on assumptions that are not testable from the observed data. The log-likelihoods of MAR, pattern-mixture and selection models cannot be compared. Only comparison of log-likelihoods and BIC within the same family is informative. The models can, however, be compared in terms of they relate to this distal event. This is studied in Section 5.6.

5.6 Models Incorporating a Distal Event

The two-class models for the AAA study discussed in Sections 4.2.1 and 4.3 were extended with a logistic regression model for the operational status. So, $u_i=1$ for patients that underwent surgery and $u_i=0$ for patients that did not have an operation. Overall, 26% of the patients underwent surgery. This probability was allowed to be different for the 2 classes: $\text{logit}[pr(u_i = 1)] = \alpha_1$ for class 1 and $\text{logit}[pr(u_i = 1)] = \alpha_2$ for class 2. This distal event can be incorporated in a sensitivity

analysis since it contains not only information on the observed diameters but also on the reason for dropout. If the original models are able to capture possible non-random missingness then the estimated mean trajectories and the class assignments obtained for the original and extended models will be similar.

Figure 7.2 shows the mean curves for the two-class models as discussed in Sections 5.2 and 5.3. The trajectories estimated under the original model and the extended versions, including the distal event, are presented in the same graph. In general, the extended versions retain the trajectory shapes. The Beunckens model performs worse. For this model, it is also noted that the number of patients classified in the highest class decreases from 34% under the original model to 20% under the extended version (Table 10). With the exception of the Diggle-Kenward Model (5), the proportion of patients classified in the high class is lower for the extended model as compared with the original model. Table 9 shows the transitions between low and high classes for the original and extended models. The percentage of patients that underwent surgery is also given. As noted before, transition from the low to the high class is not as frequent as the opposite transition. The few transitions from the low class in the original model to the high class in the extended model occur for patients that underwent surgery. None of the patients being classified in the high class in the original model and in the low class in the extended model, except one, underwent an operation.

The predictive power of these extended models is presented in Table 10. For all models, the lowest class is clearly a group of patients not needing surgery. In terms of the probability to undergo surgery, for the patients being classified in the high class, Diggle-Kenward Model (5) and Beunckens' models perform best. When classified in the high class according to the Beunckens' extended model, the probability the patient needs surgery is 86%.

Table 9 also displays the Kappa statistics measuring the agreement between the classification according to the original and extended versions of the MAR and MNAR models. This allows a comparison of the models in terms of how they relate to the distal outcome, which contains information about the reason for dropout. If the original model succeeds in capturing non-random missingness, correspondence is expected between the classes uncovered by the original and extended MNAR model. Diggle-Kenward Model (5) and Roy's model have κ values of about 0.9, pointing to very good agreement. The κ values for the other models are not too bad, but clearly Diggle-Kenward Model (5) and Roy's model stand out. Thus, of the MNAR

models considered, the original Diggle-Kenward Model (5) and Roy's model are best capturing the potential for non-ignorable missingness.

6 Discussion and Conclusion

In this paper, a number of models for non-random missingness were fitted to the abdominal aorta aneurysm (AAA) data. Repeated measurements data for 100 patients was available. The complexity of the model rapidly increases with increasing number of presumed classes. But given the longitudinal nature of the data, rather complex models (e.g. Diggle-Kenward model with 62 parameters) can be estimated. This may result in overfitting and non-reproducible results. To guard against overfitting, the optimal model was selected based on the BIC, using a penalty for the number of parameters in the model. It is clear that in the current study the chance to disclose small clusters and the power to choose a more complex model over a simple model is low and that larger sample sizes are needed to address these issues.

It is hard to give guidelines on how much missing data is allowable in a longitudinal analysis. In general the missing data mechanism and missing data patterns are more important than the amount of missing data. The predictive power of the observed measurements for the missing measurements regulates the quality of the statistical results. The predictive distribution for the missing measurements given the observed measurements depends on the correctness of the specified model, the number of observations and the strength of the correlation between the measurements.

The models presented in this manuscript can result in multi-modal likelihood surfaces. To minimize the risk to find a local maximum likelihood solution and not the global likelihood solution different random starting values were specified. Models where latent structures are imposed - such as latent classes, random effects and missing data models- are non-identifiable. The non-identifiable components of these models can be identified only by using unverifiable restrictions. Pattern-mixture models are over-specified models. The model is identified only for the subgroup of patients with complete data. To solve this problem the inestimable parameters of the incomplete patterns were set equal to parameters describing the distribution of the patients with complete data. In a selection model, the regression of m on y is inherently inestimable because y is always missing when m equals 1. By imposing assumptions for the repeated measures or the random effects; typically multivariate normal, identification is obtained. In this manuscript we imposed constraints to make the over-specified models identifiable. In contrast, the non-identifiability can be used to delineate a range of estimates consistent with the observed data, the so-called range of ignorance (Molenberghs et al. 2001a, Kenward et al. 2001, Vansteelandt et al. 2006). A way

to achieve this is to consider models that would be identified for the complete data and fit them to the incomplete data, thereby producing a range of estimates instead of a point estimates.

For the abdominal aorta aneurysm data, similarities in terms of the trajectory shapes are seen, irrespective of the model chosen for dropout. The division of patients in the two latent classes, however, changed substantially. This phenomenon has been studied by a number of authors in specific data and model settings. Verbeke and Molenberghs (2000) applied selection models of the Diggle-Kenward type to the so-called toenail dermatophyte onychomycosis (TDO) data. They considered a measurement model including a serial component. This model was expanded to also include measurement error. Both selection models and a MAR model result in very comparable estimated mean profiles and non-significant treatment effects. The sensitivity of the selection model is, however, seen in the estimated correlation matrix. Inclusion of measurement error results in higher estimated correlations, as compared to only including serial correlation. This directly has an effect on the dropout model. The current observation is less needed for predicting dropout, once the previous observation is known. This analysis clearly shows that finding an appropriate covariance model is important when dropout is present. Kenward (1998) presented maximum likelihood estimates of random and non-random dropout models, deleting a different set of influential observations from the Mastitis in Dairy Cattle data. He noted that the influence of the deletion scheme on the measurement model parameters is small. The estimates for the dropout model, however, changed substantially.

A unifying framework was presented by Verbeke and Molenberghs (2009), bringing together the frameworks of coarse data (e.g., missing data, censoring, grouping) and augmented data (e.g., latent classes, latent variables, random effects, mixture model membership). The term coarse data is reserved for settings where the observed data are less refined than what ideally might have been observed. The authors refer to augmented data when convenience or interpretation-enhancing structures are added to the data, without being observable. These authors term this unification as data enrichment and showed that every model for enriched-data settings can be factored as a product of two components. The first one, the marginal model, is fully identifiable from the observed data; the second one, the predictive model, is the conditional distribution of the enriched data given the observed data. This component cannot be identified from the data without additional, non-verifiable model assumptions. To a large extent, inferences purely in terms of the marginal models (e.g., treatment effects), are unaffected by enrichment, whereas others,

such as empirical Bayes estimates or conclusions based on latent-class membership, strongly depend on assumptions made.

This stresses the need for sensitivity analysis. Rather than fitting a single model and putting belief in it, it is advisable to consider a set of alternative models and assess how vulnerable the results are to the choices made. Such an analysis has been applied to the AAA data. However when results turn out to be sensitive to the assumptions made (e.g. with respect to the missing-data mechanism) it is hard to choose between these models since they all rely on non-verifiable assumptions. In this paper, model validation was done by incorporating a so-called distal event in the model. A distal event is an ultimate outcome that relates to the growth curves studied. The relationship between the latent-class membership and the distal event provides a predictive validity check of the latent classes. Furthermore, the congruence between the latent class formation when not including versus including the distal event in the model is a measure for the ability of the original model to capture non-ignorable missingness. For the AAA study, the fact whether or not a patient had surgery of the artery was used as auxiliary information in the model. We conclude that the Diggle-Kenward Model (5) is preferable. The model has good predictive power for a patients' need to undergo surgery. The trajectories described by the original and extended model Diggle-Kenward Model (5) are very similar; and they result in the same classification of the patients. Given these results we conclude that the Diggle-Kenward Model (5) picked up best the missingness mechanism.

References

- Beunckens, C., Molenberghs, G., and Kenward, M.G. (2005). Tutorial: Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical Trials* **2**, 379–386.
- Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics* **64**, 96–105.
- Carpenter, J., Kenward, M.G., Evans S. and White, I. (2004). Last observation Carry-forward and last observation analysis. Letter to the Editor. *Statistics in Medicine*, **23**, 3241–3244.
- Dempster, A.P., Laird, N.M., and Rubin, D. B. (1977). Maximum likelihood from incomplete

- data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Diggle, P.J. and Kenward, M.G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley & Sons.
- Gibbons, R.D., Hedeker, D., Elkin, I., Waternaux, C., Kraemer, H.C., Greenhouse, J.B., et al. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Arch. Gen. Psychiatry* **50**, 739–750.
- Kenward, M.G. (1998). Selection models for repeated measurements with nonrandom dropout: an illustration of sensitivity. *Statistics in Medicine* **17**, 2723–2732.
- Kenward, M.G., Goetghebeur, E.J.T, and Molenberghs, G. (2001). Sensitivity analysis of incomplete categorical data. *Statistical Modelling* **1**, 31–48.
- Laird, N.M., and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
- Land, K.C., and Nagin, D.S. (1996). Micromodels of Criminal Careers: A Synthesis of the Criminal Careers and Life Course Approaches via Semiparametric Mixed Poisson Regression Models, with Empirical Models. *Journal of Quantitative Criminology* **12**, 163–191.
- 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., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- Molenberghs, G., Kenward, M.G., and Goetghebeur, E. (2001a). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics* **50**, 15–29.

- Molenberghs, G., Verbeke, G., Thijs, H., Lessafre, E., and Kenward, M.G. (2001b). Mastitis in dairy cattle: influence analysis to assess sensitivity of the dropout process. *Computational Statistics and Data Analysis* **37**, 393–113.
- Molenberghs, G., Beunckens, C., Sotito, C., and Kenward, M.G. (2008). Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B* **70**, 371–388.
- Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons.
- Molenberghs, G., Njeru Njagi, E., Kenward, M.G., and Verbeke, G. (2012). Enriched-data problems and essential non-identifiability. *International Journal of Statistics in Medical Research*, **1**, 16–44.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Muthén, L.K. and Muthén, B.O. (1998–2007). *Mplus User's Guide*. Fourth edition. Los Angeles, CA: Muthén and Muthén.
- Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM-algorithm, *Biometrics*, **55**, 463–469.
- Muthén, B., Asparouhov, T., Hunter, A. and Leuchter, A. (2011). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, **16**, 17–33.
- Nagin, D.S. (1999). Analyzing developmental trajectories: a semiparametric, group-bases approach. *Psychological Methods*, **4**, 139–157.
- Nagin, D.S. and Land, K.C. (1993). Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, **31**, 327–362.
- Nagin, D.S. and Tremblay, R.E. (2001). Analyzing developmental trajectories of distinct but related behaviours: a group-based method. *Psychological Methods*, **6**, 18–34.

- Ramaswamy, V., DeSarbo, W., Reibstein, D., and Robinson, W. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, **12**, 103–124.
- Roy, J. (2003). Modeling Longitudinal Data with Nonignorable Dropouts Using a Latent Dropout Class Model. *Biometrics* **59**, 829–839.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of Statistics*, **6**, 461–464.
- Vansteelandt, S., Goetghebeur, E., Kenward, M.G., and Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, **16**, 953–979.
- Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Verbeke, G. and Molenberghs, G. (2009). Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling*, **00**, 000–000.
- Wu, M.C., and Bailey, K.R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939–955.
- Wu, M.C. and Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics*, **44**, 175–188.

7 Appendix

7.1 Missing Data Modeling Frameworks

In case of missing data in the repeated measurements y_i , one needs to specify the full-data likelihood. The full-data likelihood considers not only the repeated measurements, observed and unobserved, as data but also the missing-data indicators for each time point. The missing-data indicator for subject i , m_i , are binary variables with values for m_{it} equal to 1 if the repeated

measurement at time t , y_{it} , is missing and 0 if y_{it} is observed. In case of pure dropout, the vector \mathbf{m}_i can be replaced by a scalar D_i , denoting the time at which the patient drops out from the study.

The observed repeated measurements are grouped in $\mathbf{y}_i^{\text{obs}}$ and the missing measurements in $\mathbf{y}_i^{\text{mis}}$ such that $\mathbf{y}_i = (\mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{mis}})$. The full-data density then becomes:

$$f(\mathbf{y}_i, \mathbf{m}_i | X_i, W_i, \boldsymbol{\theta}, \boldsymbol{\psi}), \quad (11)$$

with $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ the parameter vectors associated with the measurement process and the missingness process, and X_i and W_i the associated design matrices.

Different modeling frameworks for incomplete longitudinal data are obtained by factorizing this full-data density in different ways. This results in selection, pattern-mixture and shared-parameter models. The interpretation of the results depends on the chosen framework and different results and conclusions can be obtained. When interest is also in unobserved latent constructs (such as latent-class membership) one should realise that the different models imply different assumptions for the unobserved data, but that the observed data alone does not allow to choose between these models. Auxiliary information can be insightful to judge the plausibility of the different models (see Section 4.3.3).

A selection model (Diggle and Kenward, 1994) factorizes the full-data density as the product of the marginal density of the measurements and the conditional density of the missingness process, conditioned on the measurements.

$$f(\mathbf{y}_i, \mathbf{m}_i | X_i, W_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | X_i, \boldsymbol{\theta}) f(\mathbf{m}_i | \mathbf{y}_i, W_i, \boldsymbol{\psi}). \quad (12)$$

Whenever $m_{it} = 1$ the measurement is missing, thus the regression of \mathbf{m}_i on \mathbf{y}_i is inherently inestimable. Only by imposing distributional assumptions for the repeated measurements, often multivariate normality, the model can be estimated. Departures from these assumptions can produce biased results.

One often assumes the following form for the conditional density in the selection model, proposed

by Diggle and Kenward (1994):

$$\text{logit}(m_{it}|y_{it}, y_{i,t-1}) = \beta_{0t} + \beta_1 y_{it} + \beta_2 y_{i,t-1}. \quad (13)$$

The probability of missing data at time t depends directly on the repeated measurements at time t as on the preceding measurement. MAR holds if $\beta_1 = 0$ and $\beta_1 \neq 0$ is an indication of MNAR. A logistic regression could be used to make inference about the missing-data mechanism. However, the association between \mathbf{m}_i and \mathbf{y}_i is estimated under distributional assumptions.

The pattern-mixture model (Little, 1995) specifies a different measurement model for each pattern of missing values, and the full-data density is obtained as the mixture of the models weighted by the probability of each missing value pattern.

$$f(\mathbf{y}_i, \mathbf{m}_i | X_i, W_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \mathbf{m}_i, X_i, \boldsymbol{\theta}) f(\mathbf{m}_i | W_i, \boldsymbol{\psi}). \quad (14)$$

The above specified model is also inestimable without imposing assumptions. A quadratic growth model with only two observed data points is not identified. Therefore, values are assumed for the inestimable parameters. Information from across patterns can be shared (Hedeker and Gibbons, 1997, 2006) or identifying restrictions can be used (Little, 1995; Thijs et al., 2002; Kenward et al., 2003). Implementing different constraints can produce different results and unfortunately there is no way to gauge the credibility of the assumptions. The marginal estimates, obtained via a pattern-mixture model, are function of the model parameters and therefore standard errors are not automatically produced. Approximate standard errors are routinely obtained via the delta method (Hedeker and Gibbons, 1997).

Shared-parameter models (Wu and Carroll, 1988) assume the existence of latent variables, that are shared between both factors in the full-data density. Often conditional independence is assumed, i.e., the measurement process and missingness process are independent conditional on the latent variable. The latent variable can be a random effect \mathbf{b}_i , resulting in the following model:

$$f(\mathbf{y}_i, \mathbf{m}_i | X_i, W_i, \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{b}_i) = f(\mathbf{y}_i | X_i, \boldsymbol{\theta}, \mathbf{b}_i) f(\mathbf{m}_i | W_i, \boldsymbol{\psi}, \mathbf{b}_i). \quad (15)$$

By regressing the missing-data indicators on the random effects (e.g., intercept and slope), the

probability of missingness depends on the entire set of repeated measurements, including the unobserved ones. This model is identified by making assumptions for the distribution of the random effects, and further by assuming that the measurement process and the missingness process are independent conditional on the random effects.

Alternatively, a categorical latent variable can be used (Roy, 2003). The missing-data patterns are assumed to be related to unobserved latent-class membership, where the number of classes is less than the number of observed missing-data patterns. The likelihood for the response is a mixture of latent dropout classes, as opposed to the observed dropout patterns themselves. This approach allows observed patterns to be sparse without needing to specify additional identifying restrictions.

The full-data density is never available and inference is made on the observed data. Depending on the postulated missing-data mechanism different implications are ensued. The implications of the postulated missing-data mechanism are easiest seen in a selection-modeling framework as defined by Rubin (1976). The conditional distribution of the missingness process, conditional on the measurements, plays a central role in the selection model. This conditional distribution simplifies under the assumption of MCAR and MAR, and this has implications for the joint density of the observed data and the missing-data indicators. The selection probability $f(\mathbf{m}_i|\mathbf{y}_i, W_i, \psi)$ under the three missing-data mechanisms is given by:

$$\begin{aligned}\text{MCAR} &: f(\mathbf{m}_i|W_i, \psi), \\ \text{MAR} &: f(\mathbf{m}_i|\mathbf{y}_i^{\text{obs}}, W_i, \psi), \\ \text{MNAR} &: f(\mathbf{m}_i|\mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{mis}}, W_i, \psi).\end{aligned}$$

The joint density of the observed data and missing-data indicators thus simplifies as:

$$\begin{aligned}\text{MCAR} &: f(\mathbf{y}_i^{\text{obs}}|X_i, \boldsymbol{\theta})f(\mathbf{m}_i|W_i, \psi), \\ \text{MAR} &: f(\mathbf{y}_i^{\text{obs}}|X_i, \boldsymbol{\theta})f(\mathbf{m}_i|\mathbf{y}_i^{\text{obs}}, W_i, \psi), \\ \text{MNAR} &: \int f(\mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{mis}}|X_i, \boldsymbol{\theta})f(\mathbf{m}_i|\mathbf{y}_i^{\text{obs}}, \mathbf{y}_i^{\text{mis}}, W_i, \psi)d\mathbf{y}_i^{\text{mis}}.\end{aligned}$$

Under MAR the likelihood factorizes into two components. When the parameter space of (θ, ψ) is given by the product of the individual parameter spaces (separability condition), inference for θ can be obtained on the marginal observed data density alone, ignoring the missingness process. So, in the likelihood framework and when the separability condition is satisfied, ignorability is equivalent to $\text{MAR} \cup \text{MCAR}$. Under MNAR the joint distribution of \mathbf{y}_i and \mathbf{m}_i cannot be simplified. Inference is only possible by making unverifiable assumptions. Whenever $m_{it} = 1$ the measurement is missing, thus the regression of \mathbf{m}_i on \mathbf{y}_i is inherently inestimable. Only by imposing distributional assumptions, often multivariate normality, the model can be estimated. Departures from these assumptions can produce biased results.

7.2 Appendix: Latency and Unidentifiability

Using the terminology of Molenberghs et al. (2010) a growth-mixture model encompasses three types of data. Of course, the observed data (outcomes $\mathbf{y}_i^{\text{obs}}$, \mathbf{m}_i and covariates \mathbf{x}_i) are one type of data. The latent variables, continuous ones (\mathbf{b}_i) and categorical ones (\mathbf{c}_i) are so-called augmented data, referring to the addition of constructs to the observed data. The augmented data are always fully unobserved. The reason for the augmentation of the data is that in general the model development simplifies. The term coarsened data is used to refer to the fact that the observed data are coarser than hypothetically conceived. Ideally, the data are observed fully but in practice this is not the case. In the growth-mixture model, formulated above, the missing observations ($\mathbf{y}_i^{\text{mis}}$) are coarsened data. The augmented and coarsened data taken together are called enriched data.

Molenberghs et al. (2010) show that always a part of the model for enriched data is unidentifiable from the observed data. By replacing these unidentified parts, which rest completely on assumptions, an entire class of models can be obtained. All of these models produce the same fit to the observed data. The authors assume that data \mathbf{z}_i for an independent unit $i = 1, \dots, N$ are augmented with \mathbf{a}_i . The \mathbf{a}_i can take any enriched data form and is thus broader than the latent classes used before.

The joint model $f(\mathbf{z}_i, \mathbf{a}_i | \theta, \psi)$ can be factorized in the following ways:

$$f(\mathbf{z}_i, \mathbf{a}_i | \theta, \psi) = f(\mathbf{z}_i | \mathbf{a}_i, \theta) f(\mathbf{a}_i | \psi) = f(\mathbf{z}_i | \theta, \psi) f(\mathbf{a}_i | \mathbf{z}_i, \theta, \psi). \quad (16)$$

Replacing the posterior density, of the enriched data given the observed data $f(\mathbf{a}_i|\mathbf{z}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$, where $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}$ are the estimates obtained from the original model, by an arbitrary conditional density $f(\mathbf{d}_i|\mathbf{z}_i, \gamma)$ retains the fit to the original model. Here, \mathbf{d}_i is used to indicate that the original and substituted enriched data can be different.

Using the notation of the previous section, in a growth-mixture model, it is the posterior density $f(\mathbf{b}_i, \mathbf{c}_i, \mathbf{y}_i^{\text{mis}}|\mathbf{y}_i^{\text{obs}}, \mathbf{m}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$ that could be replaced without changing the fit to the data. This posterior function could further be factorized in the following three terms :

$$\begin{aligned} f(\mathbf{b}_i, \mathbf{c}_i, \mathbf{y}_i^{\text{mis}}|\mathbf{y}_i^{\text{obs}}, \mathbf{m}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) &= f(\mathbf{y}_i^{\text{mis}}|\mathbf{y}_i^{\text{obs}}, \mathbf{m}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) \\ &\times f(\mathbf{b}_i|\mathbf{c}_i, \mathbf{y}_i^{\text{mis}}, \mathbf{y}_i^{\text{obs}}, \mathbf{m}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) \\ &\times f(\mathbf{c}_i|\mathbf{b}_i, \mathbf{y}_i^{\text{mis}}, \mathbf{y}_i^{\text{obs}}, \mathbf{m}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}). \end{aligned} \quad (17)$$

A simplification, resulting in the conventional growth-mixture model, assumes that the random effects \mathbf{b}_i are influenced by the class; and that prediction of the class is done solely by covariates. In this case (17) becomes:

$$\begin{aligned} f(\mathbf{b}_i, \mathbf{c}_i, \mathbf{y}_i^{\text{mis}}|\mathbf{y}_i^{\text{obs}}, \mathbf{m}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) &= f(\mathbf{y}_i^{\text{mis}}|\mathbf{y}_i^{\text{obs}}, \mathbf{m}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) \\ &\times f(\mathbf{b}_i|\mathbf{c}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) \\ &\times f(\mathbf{c}_i|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}). \end{aligned} \quad (18)$$

Replacing one of these densities will not change the value of the observed data log-likelihood. The first term of (18) could, for example, be replaced by its MAR counterpart $f(\mathbf{y}_i^{\text{mis}}|\mathbf{y}_i^{\text{obs}})$. The posterior density of the random effects given the class, is generally assumed to be the normal distribution. These densities could be replaced by other distributions, e.g., gamma random effects. Finally, the last term in (18) specifies the number of latent classes. A model specifying l classes instead of k could be considered, resulting in the same fit.

None of these densities is identifiable from the observed data, they are solely determined by modeling assumptions. To investigate how sensitive the results are to the assumptions, a sensitivity

analysis should be conducted. In a broad sense, a sensitivity analysis is one in which several statistical models are considered simultaneously and/or a statistical model is further scrutinized using specialised tools (such as diagnostic measures). The simplest procedure is to fit a selected number of (MNAR) models which seem plausible, or one in which a primary analysis is supplemented with a number of modifications. The degree to which conclusions are stable across these models provides an indication of the confidence that is placed in them. Comparison of the models is also possible in terms of their predictive value for an event occurring in the (near) future. By studying the congruence between the latent class formation when including and not including a distal event in the model, the ability of the model to capture non-ignorable missingness can be studied (see Section 4.3.3).

Table 1: Coverage of the AAA study: number of patients still in the study per visit (and the split over the patients undergoing and not undergoing an operation is also given).

Visit	Time (years)	Total	not operated	operated
		N	N	N
1	0	100	73	27
2	0.5	89	65	24
3	1	74	52	22
4	1.5	53	37	16
5	2	46	32	14
6	2.5	19	14	5
7	3	7	5	2

Table 2: Average diameter and 95% confidence interval per study visit, according to patients undergoing and not undergoing an operation and dropping out or not at the next visit.

not operated					operated				
drop out at next visit					drop out at next visit				
no		yes			no		yes		
Visit	N	average (95% CI)	N	average (95% CI)	N	average (95% CI)	N	average (95% CI)	
1	65	40.5 [39.0; 42.0]	8	45.4 [39.5; 51.3]	24	48.6 [46.7; 50.5]	3	38.0 [0.0; 76.1]	
2	52	41.4 [39.6; 43.1]	13	42.2 [38.7; 45.7]	22	50.4 [48.7; 52.1]	2	44.5 [-76.2; 165.2]	
3	37	42.5 [40.5; 44.4]	15	40.7 [36.7; 44.8]	16	51.5 [49.6; 53.4]	6	53.3 [48.1; 58.6]	
4	32	43.1 [41.0; 45.2]	5	43.8 [39.2; 48.4]	14	53.8 [51.4; 56.1]	2	60 -	
5	14	40.6 [37.8; 43.5]	18	46.1 [43.0; 49.2]	5	52.4 [45.2; 59.6]	9	57.7 [55.2; 60.1]	
6	5	43.4 [35.5; 51.3]	9	41.1 [37.4; 44.8]	2	54.5 [10.0; 98.9]	3	53.3 [35.7; 70.9]	

Table 3: *Evolution of the diameter per drop-out pattern. The intercept and slope (and standard error) estimated by means of a linear mixed model are displayed.*

Drop out at visit	Intercept (s.e.)	Slope (s.e.)
2	43.9 (2.2)	-
3	40.4 (1.5)	1.1 (0.5)
4	40.8 (1.8)	1.2 (0.3)
5	41.7 (1.6)	1.7 (0.6)
6	44.0 (1.2)	1.1 (0.2)
7	38.3 (1.7)	0.9 (0.2)
no drop out	41.3 (2.2)	0.8 (0.2)

Table 4: *Summary of the MAR and MNAR models applied to the AAA data: number of classes, log-likelihood, number of parameters, and BIC.*

Model	# classes	log(L)	# par.	BIC
MAR	1	-952.040	6	1932.039
MAR	2	-925.211	13	1910.289
MAR	3	-911.006	20	1914.116
MAR	4	-901.437	27	1927.214
Selection Models:				
Diggle-Kenward	1	-1132.386	14	2329.244
Diggle-Kenward Model (5)	2	-1128.814	29	2391.178
Diggle-Kenward Model (5)	3	-1090.499	44	2383.625
Diggle-Kenward Model (5)	4	-1069.043	59	2409.791
Diggle-Kenward Model (6)	2	-1105.308	41	2399.428
Diggle-Kenward Model (6)	3	-1087.660	62	2460.841
Beunckens Model (7)	2	-1122.674	19	2332.846
Beunckens Model (7)	3	-1107.441	29	2348.432
Pattern-Mixture Models:				
Pattern-mixture	1	-964.797	16	1967.277
Roy	2	-920.584	19	1928.666
Roy	3	-900.032	32	1947.430
Roy	4	-884.476	45	1976.184

Table 5: Summary of the Roy-Muthén models applied to the AAA data: number of classes, log-likelihood, number of parameters, and BIC.

Model	# cy	# cd	log(L)	# par.	BIC
Roy-Muthén	2	2	-917.823	22	1936.960
Roy-Muthén	3	2	-897.564	30	1933.282

Table 6: Percentiles (median, P10 and P25) of the distribution of posterior probabilities to be classified in the low or high class. - AAA data.

Model	Lowest class			Highest class		
	P50	P10	P25	P50	P10	P25
MAR	0.98	0.63	0.85	0.93	0.72	0.80
Diggle-Kenward Model (5)	0.99	0.77	0.91	1.00	0.98	0.99
Diggle-Kenward Model (6)	0.99	0.63	0.89	0.92	0.71	0.80
Beunckens Model (7)	0.93	0.73	0.81	1.00	0.90	0.98
Roy	0.99	0.72	0.91	0.94	0.62	0.76

Table 7: Kappa statistic (κ) measuring the agreement in classification for the two-class models presented in Table 8 - AAA study.

	MAR	Diggle-Kenward Model (5)	Diggle-Kenward Model (6)	Beunckens	Roy
MAR	1.000	0.444	0.600	0.511	0.877
Diggle-Kenward Model (5)		1.000	0.375	0.862	0.504
Diggle-Kenward Model (6)			1.000	0.407	0.695
Beunckens				1.000	0.576
Roy					1.000

Table 8: Summary of the measurement part of the two-class MAR and MNAR models applied to the AAA data: class-specific parameter estimates (standard errors).

		MAR	DK Model (5)	DK Model (6)	Beunckens	Roy
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Class I: Lowest Class						
Measurement Model						
Fixed Effects:						
Intercept	β_0^1	38.294(1.337)	39.819(0.701)	39.686(1.078)	39.512(0.865)	38.494(1.259)
Slope	β_1^1	1.328(0.195)	2.113(0.250)	1.171(0.254)	1.859(0.236)	1.377(0.188)
Covariances (b_i^1):						
$\text{var}(b_{0i}^1)$	d_{11}^1	23.138(6.601)	27.818(3.405)	35.669(6.263)	25.622(3.348)	23.294(5.518)
$\text{var}(b_{1i}^1)$	d_{22}^1	0.554(0.255)	1.383(0.451)	1.175(0.354)	1.173(0.527)	0.556(0.241)
$\text{cov}(b_{0i}^1, b_{1i}^1)$	d_{12}^1	1.301(1.093)	-1.120(1.171)	-1.761(1.752)	1.753(1.342)	1.548(1.040)
Residual variance:						
$\text{var}(\varepsilon_{ij}^1)$	$(\sigma^2)^1$	0.601(0.082)	1.725(0.235)	1.204(0.256)	1.397(0.329)	0.629(0.098)
Class II: Highest Class						
Measurement Model						
Fixed Effects:						
Intercept	β_0^2	46.002(0.974)	49.986(0.473)	45.299(1.060)	49.781(0.761)	46.057(1.081)
Slope	β_1^2	2.917(0.395)	3.867(0.705)	2.427(0.378)	2.950(0.624)	2.931(0.400)
Covariances (b_i^2):						
$\text{var}(b_{0i}^2)$	d_{11}^2	30.089(8.136)	3.232(2.048)	33.199(7.077)	4.538(2.184)	30.827(9.031)
$\text{var}(b_{1i}^2)$	d_{22}^2	4.388(1.393)	4.162(1.275)	3.046(1.288)	6.440(2.049)	4.541(1.410)
$\text{cov}(b_{0i}^2, b_{1i}^2)$	d_{12}^2	-0.091(1.986)	-0.097(1.410)	2.001(2.140)	-0.860(1.942)	-0.267(2.092)
Residual variance:						
$\text{var}(\varepsilon_{ij}^2)$	$(\sigma^2)^2$	3.307(0.452)	3.691(0.710)	3.863(0.616)	3.195(0.552)	3.097(0.447)

Table 9: Number of patients with transitions; percentage of operated patients; κ statistic measuring the agreement in classification for the original (Table 8) and extended model.

		Extended Model				κ statistic ([95% CI])
		Low class		High class		
		n	%	n	%	
		operated		operated		
Original Model						
MAR	Low	36	2.8	4	100.0	0.754 [0.624; 0.884]
	High	8	0.0	52	42.3	
DK Model (5)	Low	68	7.4	2	100.0	0.905 [0.813; 0.996]
	High	2	0.0	28	71.4	
DK Model (6)	Low	38	0.0	5	100.0	0.603 [0.451; 0.756]
	High	15	0.0	42	52.4	
Beunkens	Low	66	12.1	0	-	0.654 [0.496; 0.811]
	High	14	7.1	20	90.0	
Roy	Low	43	7.0	1	100.0	0.879 [0.786; 0.973]
	High	5	0.0	51	45.10	

Table 10: Estimated probability to undergo surgery given latent-class membership under two class MAR and MNAR models - AAA data.

Model	Original			Extended			
	Lowest class	Highest class	Entropy	Lowest class	Highest class	Lowest class	Highest class
	n	n		n	n	$p(u = 1)$	$p(u = 1)$
MAR	40	60	0.640	44	56	0.05	0.44
Diggle-Kenward Model (5)	70	30	0.890	70	30	0.08	0.74
Diggle-Kenward Model (6)	43	57	0.665	53	47	<0.01	0.52
Beunckens Model (7)	66	34	0.787	80	20	0.11	0.86
Roy	44	56	0.662	48	52	0.06	0.45

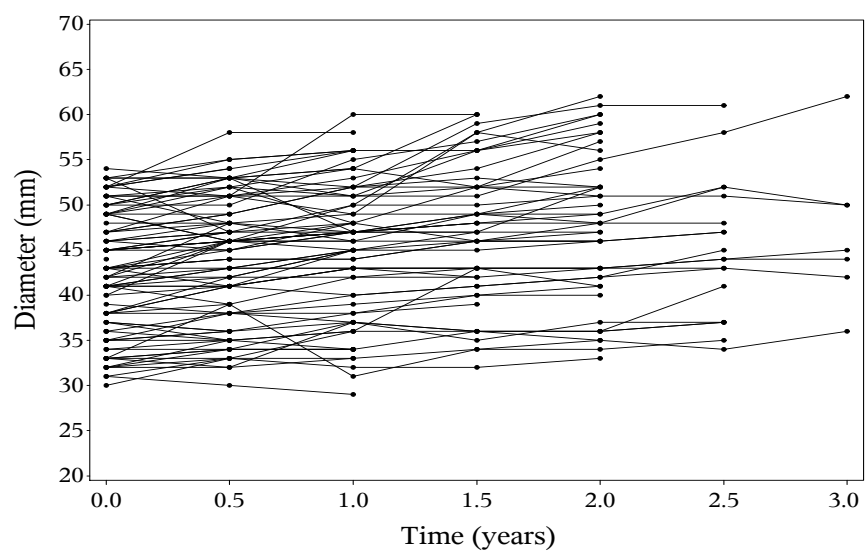


Figure 1: *Evolution of the patients' diameter – AAA study.*

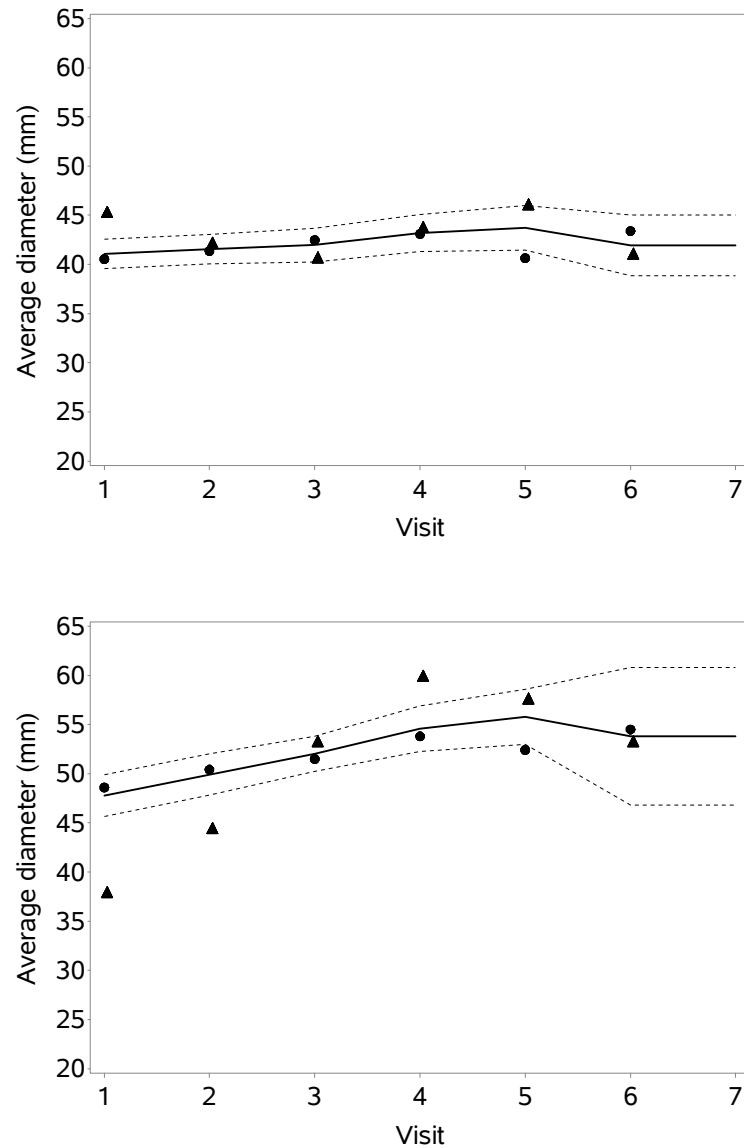


Figure 2: Evolution of the (sample) average diameter (full lines) and a 95 % confidence interval (dashed lines) for patients that were not operated (upper panel) and for patients that under went surgery (lower panel). At each visit the mean diameter is also presented according to dropout status at the next visit. The filled circles present the average diameter for the patients still in the study at the next visit, the filled triangles present the average diameter for the patients dropping out at the next visit — AAA Data.

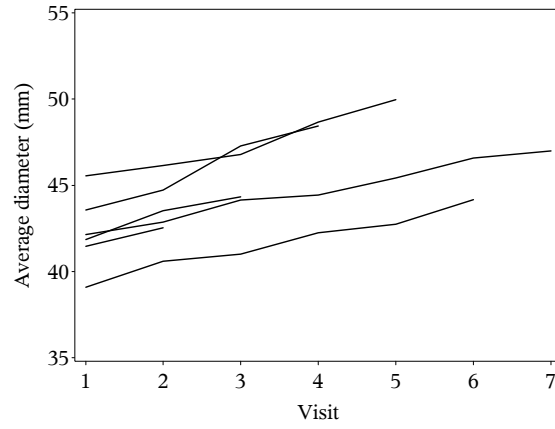


Figure 3: *Evolution of the (sample) average diameter per dropout pattern– AAA Data.*

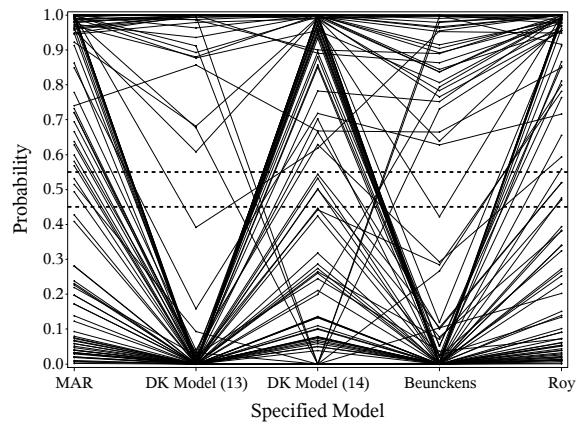


Figure 4: *Posterior probability to be classified into the group with the largest and fastest increasing diameters, for the two-class models presented in Table 8 – AAA study. Posterior probabilities between the dashed horizontal lines (at 0.45 and 0.55) indicate large uncertainty in classification.*

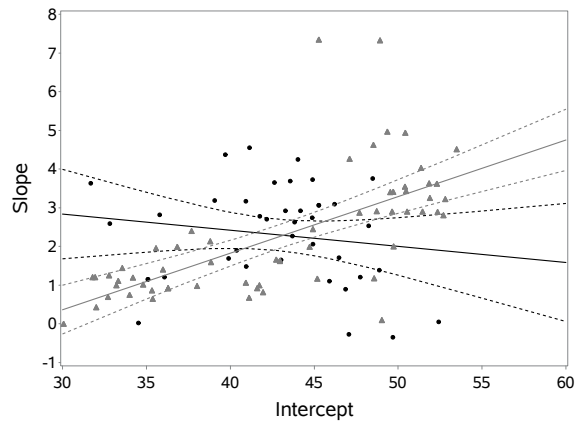
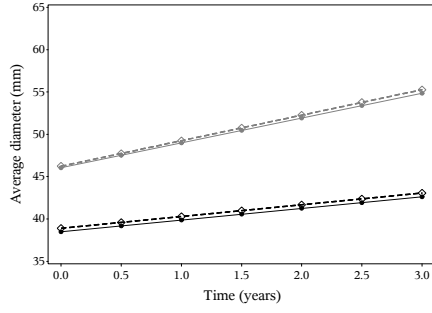
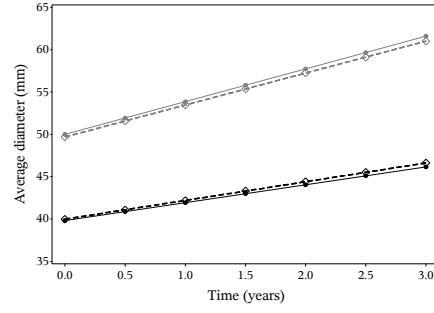


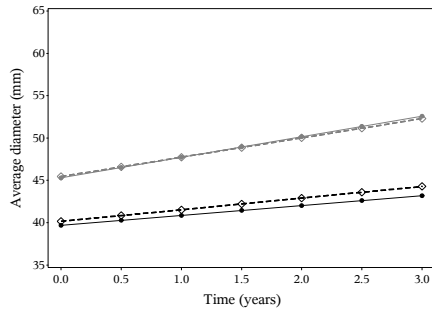
Figure 5: Empirical Bayes estimates for random intercept and random slope, under the two-class MAR – AAA study. Patients always classified in same latent class are represented by triangles and gray lines, patients whose classification depends on the chosen dropout model are represented by dots and black line. The full line represent the regression lines, the dashed lines 95% confidence intervals for the regression lines.



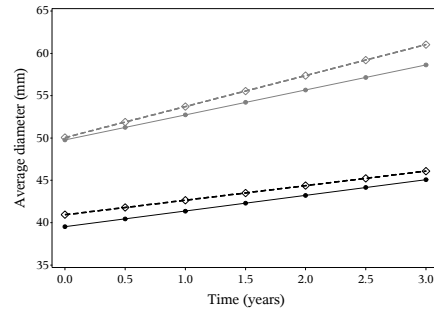
(a) MAR



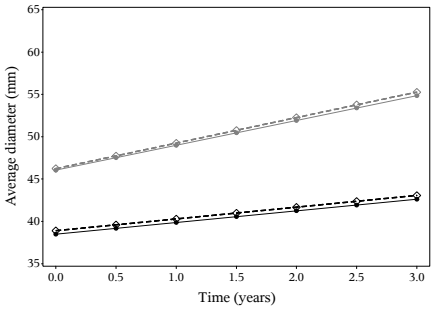
(b) Diggle-Kenward Model (5)



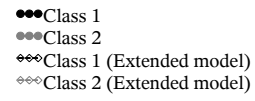
(c) Diggle-Kenward Model (6)



(d) Beunckens



(e) Roy



(f) Legend

Figure 6: Estimated average trajectories under the MAR and MNAR two-class models (original model and the extended version to include a distal event) – AAA data.