SORT 37 (2) July-December 2013, 131-152

Selection and pattern mixture models for modelling longitudinal data with dropout: An application study

A. Satty and H. Mwambi*

Abstract

Incomplete data are unavoidable in studies that involve data measured or observed longitudinally on individuals, regardless of how well they are designed. Dropout can potentially cause serious bias problems in the analysis of longitudinal data. In the presence of dropout, an appropriate strategy for analyzing such data would require the definition of a joint model for dropout and measurement processes. This paper is primarily concerned with selection and pattern mixture models as modelling frameworks that could be used for sensitivity analysis to jointly model the distribution for the dropout process and the longitudinal measurement process. We demonstrate the application of these models for handling dropout in longitudinal data where the dependent variable is missing across time. We restrict attention to the situation in which outcomes are continuous. The primary objectives are to investigate the potential influence that dropout might have or exert on the dependent measurement process based on the considered data as well as to deal with incomplete sequences. We apply the methods to a data set arising from a serum cholesterol study. The results obtained from these methods are then compared to help gain additional insight into the serum cholesterol data and assess sensitivity of the assumptions made. Results showed that additional confidence in the findings was gained as both models led to similar results when assessing significant effects, such as marginal treatment effects.

MSC: (MSC 2000) 97K80 or 46N30

Keywords: Identifying restrictions, under-identification, selection models, pattern mixture models, sensitivity analysis.

1. Introduction

In most longitudinal studies where data are collected over a sequence of time points, missing data are caused by individuals dropping out of the study prior to the time

* School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X01 Scottsville 3209, Pietermaritzburg, South Africa. E-mail: alisatty1981@gmail.com

Received: May 2012

Accepted: January 2013

at which the primary endpoint data would be collected. Missingness for longitudinal data often occurs as dropout that is a particular case of missing data. Furthermore, the resulting data obtained from such studies would have a particular type of missing data pattern; that is, a monotone missingness pattern, in which if an individual has missing values for a given time, no data can be obtained for all subsequent times for that individual. In this paper, our focus will be on this type of missing data pattern. Other types of missingness patterns are possible, such as intermittent missingness, but we focus on dropout which occurs most often in longitudinal studies. The mechanisms that lead to missing data are varied. Rubin (1976) and Little and Rubin (1987) classified these mechanisms into three possible categories, namely data missing completely at random, at random, or not at random. For longitudinal data, when data are missed at random or completely at random, available cases analysis, such as mixed models can be used. In contrast, when data are missed not at random, then using a standard mixed model without accounting for the missingness may lead to biased and inconsistent assessment of study results. Standard strategies of analysis currently assess non-random dropout by performing sensitivity analysis using analytical methods that incorporate non-random dropout in longitudinal data with and without a non-random component. Common families of models for data that are subject to dropout are selection and pattern mixture models.

Selection and pattern mixture models are two alternative and important approaches for dealing with longitudinal data when there are dropouts. They make empirically unverifiable assumptions and require extra constraints to identify the parameter estimates. Both models differ in the way the joint distribution of the measurement and dropout processes are factorized. However, other models that drive both the measurement process and dropout process, such shared-parameter models by Wu and Carrol (1988) and Wu and Bailey (1988, 1989) are also available. We restrict ourselves to the selection and pattern mixture models with dropout that falls under the monotone missing data pattern. A selection model factors the joint distribution into the marginal measurement model that describes the distribution of the complete measurements, and the dropout model that describes the conditional distribution of the dropout indicators, given the observed and unobserved measurements (Diggle and Kenward, 1994). However, in many discussions, for example, Diggle and Kenward (1994), Verbeke and Molenberghs (2000) and Molenberghs and Verbeke (2005), the conclusions obtained from selection models depend on the assumptions made some of which cannot be investigated from the data under analysis. Early reference to such models is found in Heckman (1976) in the econometrics area. The use of pattern mixture models, on the other hand, was originally proposed by Little (1993, 1994) as a viable alternative to selection models. In this approach, models are under-identified; that is, for each dropout pattern the observed data does not provide direct information to identify the distributions for the incomplete patterns. Therefore, to overcome this problem, Little (1993, 1994) solves the under-identification problem through the use of identifying restrictions. Early applications concerning selection and pattern mixture models can be found in Marini et al. (1980) and Glynn et al. (1986).

Selection and pattern mixture models are somewhat opposite to each other. That is, these models exploit the conditional probability rule, but they do so in opposite ways. The marginal estimates in selection models can be derived directly, while pattern mixture models estimate the marginal parameters as a weighted average through pattern specific estimates (Little, 1995).

There are several studies in the literature which provide a comprehensive review of these models. The differences between selection models and pattern mixture models have been discussed in many works, for example, Glynn et al. (1986) and Little (1993, 1994). Little (1995) also made an important distinction between selection and pattern mixture models. A comparison of the conclusions based on the selection model with those based on the pattern mixture models have been discussed in Verbeke et al. (2001) and Michiels et al. (2002). Molenberghs et al. (1998a) contrast selection and pattern mixture models. Further discussion of these models can be found in McArdle and Hamagami (1992), Little and Wang (1996), Hedeker and Gibbons (1997), Hogan and Laird (1997), Kenward and Molenberghs (1999), Verbeke and Molenberghs (2000), Molenberghs and Verbeke (2005), Molenberghs and Kenward (2007) and Daniels and Hogan (2008). However, the approach by Daniels and Hogan (2008) is Bayesian based, which is not the focus of the current study.

This paper is primarily concerned with two attractive modelling frameworks to account for non-random dropout, namely selection and pattern mixture models. We demonstrate the application of selection and pattern mixture models for handling dropout in longitudinal data where the dependent variable is missing across time. In particular, we illustrate the application and results of analysis with these models. The underidentification in pattern mixture models is addressed through identifying restrictions, while the use of the selection model is based on Diggle and Kenward's (1994) model. We restrict our attention to the situation in which linear models are used and the outcomes are continuous. The primary objectives are to investigate the potential influence that dropout might exert on the dependent measurement on the considered data as well as how to deal with incomplete sequences. We relate the identified restrictions estimates using a pattern mixture model framework to their corresponding estimates using a selection model framework. We apply the methods to a data set arising from a serum cholesterol study. Section 2 describes the notation and general concepts based on the selection and pattern mixture models. In Section 3, we give a discussion of the two families of models that are used in the analysis, namely selection and pattern mixture models. An application study is provided in Section 4 including the description of the serum cholesterol data to which our methods will be applied. In addition, full analysis and results of the application is also given. Section 5 presents concluding remarks and discussion.

2. Notation and concepts

We introduce modelling incompleteness notation which is largely due to Rubin (1976) and Little and Rubin (1987). Let y_{ij} be the response of interest, for the ith study subject, where $i = 1,\ldots,N$, designed to be measured at occasion t_j, where $j = 1,\ldots,n$. In other words, the original intention was to have n observations per individual. However, due to dropout some individuals end up contributing less than n intended observations. Therefore, generally, we can assume that the ith individual is actually observed n_i times. For subject i and occasion j, define R_{ij}=1, if y_{ij} is observed, and 0, if not. We split y_{ij} into two sub-vectors, y_i^o and y_i^m, representing those y_{ij} for which R_{ij}=1, and R_{ij}=0, respectively. In addition, suppose the missing data occur due to dropout, then the measurements for each subject can be recorded up to a certain time point, after which all data are unobserved. In this case, a dropout indicator can then be defined as D_i, given by $D_i = 1 + \sum_{j=1}^{n} R_{ij}$, denoting the occasion at which dropout first occurs. In modelling a missing data process, it is often necessary to consider a joint model for the measurement process together with the dropout process. Therefore, we assume the full data density is given by

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}), \tag{1}$$

where $\boldsymbol{X}_i$ denotes the design matrix for fixed effects, $\boldsymbol{Z}_i$ denotes the design matrix for random effects, while $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ represent the vectors of parametrization for the joint distribution. In considering the above model in expression (1), we can factorize this joint density function in two possible ways that can facilitate modelling. Specifically, the selection and pattern mixture models mentioned earlier are defined by the conditional factorizations of the joint distribution of $\boldsymbol{Y}$ and $\boldsymbol{R}$, and both are discussed in more detail in Little (1995) and stated briefly below. A selection model is based on the following factorization

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i \mid \boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi}), \tag{2}$$

where the first factor in the above factorization represents the marginal density of the measurement process, while the second factor represents the density of the dropout process, conditional on the measurements. An alternative factorization based on the pattern mixture models (Little 1993, 1994) is of the form

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i \mid \boldsymbol{r}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i \mid \boldsymbol{X}_i, \boldsymbol{\psi}). \tag{3}$$

This factorized density (3) can be seen as a mixture of the conditional distributions, and the model for the measurements depends on the particular missing data pattern. An excellent review of these models is given in Glynn et al. (1986), Little and Rubin (1987), Little (1993, 1994), Hogan and Laird (1997) and Ekholm and Skinner (1998).

The missing data processes have been developed by Rubin (1976) and Little and Rubin (1987) through the selection model framework. They make distinctions among different missing data processes. These processes can be formulated based on the second factor of model (2), i.e,

$$f(\boldsymbol{r}_i \mid \boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\boldsymbol{r}_i \mid \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{X}_i, \boldsymbol{\psi}). \tag{4}$$

Thus, if the distribution of missingness process is reduced to $f(\boldsymbol{r}_i \mid \boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\boldsymbol{r}_i, \boldsymbol{X}_i, \boldsymbol{\psi})$, i.e., the process is independent of the measurements, then the process is defined as missing completely at random (MCAR). If the missingness probability depends on the observed measurement $\boldsymbol{y}_i^o$, but not on the unobserved measurements $\boldsymbol{y}_i^m$, i.e, $f(\boldsymbol{r}_i \mid \boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\boldsymbol{r}_i \mid \boldsymbol{y}_i^o, \boldsymbol{X}_i, \boldsymbol{\psi})$, then the process is termed missing at random (MAR). Finally, data are missing not at random (MNAR) or exhibiting an informative process, when the missingness probability depends on the unobserved measurement, $\boldsymbol{y}_i^m$, and possibly on the observed measurement, $\boldsymbol{y}_i^o$, i.e., $f(\boldsymbol{r}_i \mid \boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\boldsymbol{r}_i \mid \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{X}_i, \boldsymbol{\psi})$. In other words, an informative process in expression (4) cannot be reduced.

3. Selection and pattern mixture models for modelling dropout

We consider the comparison between the selection and pattern mixture models concerning the significant characteristics, such as marginal treatment effects since such a comparison is a useful form of a sensitivity analysis. Specifically, we are interested in parametric selection and pattern-mixture models for modelling dropout. In the following, we briefly review these models.

3.1. Selection model

As mentioned above, a selection model factors the joint distribution into two parts: the marginal measurement model that describes the distribution of the complete measurements and the missingness model that describes the conditional distribution of the response indicators given the observed and unobserved measurements. In other words, in a selection model, we first specify a distribution for the measurement, then propose a manner in which the probability of being observed depends on the data. For continuous outcomes, using a selection model formulation as in equation (2), Diggle and Kenward (1994) combine the multivariate Gaussian linear model together with the dropout model. Similarly, we consider the measurement model to be of the linear mixed effects model (Laird and Ware, 1982). Recall that y_{ij} is the response of interest for the ith study subject, where $i = 1, \ldots, N$, at time point j, where $j = 1, \ldots, n_i$. More generally, the model for $\boldsymbol{y}_i$ the $(n_i \times 1)$ vector of responses for the ith subject can be written as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{5}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are known $(n_i \times p)$ and $(n_i \times q)$ design matrices for fixed and random effects, respectively, $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of fixed effects, $\boldsymbol{b}_i$ is the $(q \times 1)$ vector of the random effects distributed as $N(\boldsymbol{0}, \boldsymbol{G})$, $\boldsymbol{\varepsilon}_i$ is the $(n_i \times 1)$ vector of the residual components distributed independently as $N(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, $\boldsymbol{G}$ is the general $(q \times q)$ covariance matrix with (i, j)th element $d_{ij} = d_{ji}$ and $\boldsymbol{\Sigma}_i$ is the $(n_i \times n_i)$ error covariance matrix. Then, marginally, the responses $\boldsymbol{y}_i$ are distributed as independent normal $\boldsymbol{y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i^{\mathsf{T}} + \boldsymbol{\Sigma}_i)$. Here, $\boldsymbol{\Sigma}_i = \sigma^2\boldsymbol{H}_i + \tau^2\boldsymbol{I}$, where σ^2 denotes the variance of the serially correlated process, $\boldsymbol{H}_i = (h_{jk}) = (\rho(t_j, t_k))$ denotes the associated correlation matrix, τ^2 pertains to the measurement error variability and $\boldsymbol{I}$ is a $(n_i \times n_i)$ identity matrix.

We assume the missingness is due to dropout only, and that the first measurement $\boldsymbol{y}_{i1}$ is observed for each individual. Again, recall that D_i was defined as the dropout indicator which denote the occasion at which dropout first occurs. Now, let $D_i = d_i$ identify the dropout time for subject i, where $D_i = n+1$, if the sequence of measurement is complete. Therefore, the selection models introduced in equation (2) arise when the joint likelihood of the measurement and dropout processes is factorized as following

$$f(\boldsymbol{y}_i, D_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i, D_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(d_i \mid \boldsymbol{y}_i, \boldsymbol{\psi}).$$

The model for dropout process is based on a logistic regression for the probability of dropout at occasion j, given the subject was still in the study at the previous occasion. Let $g_i(y_{ij}, h_{ij})$ denote this probability, where h_{ij} represent the history of the measurement process. Thus, one can assume that $g_i(y_{ij}, h_{ij})$ satisfies the model

$$logit[g(h_{ij}, y_{ij})] = logit[p(D_i = j \mid D_i \geq j, h_{ij}, y_{ij})] = \eta(h_{ij}, y_{ij}), \tag{6}$$

where $\eta(h_{ij}, y_{ij})$ is the linear predictor depending on h_{ij} and y_{ij}. Modelling the dropout mechanism may be simplified in the expression in equation (6) by assuming $\eta(h_{ij}, y_{ij})$ depends only on the current measurement and the previous measurement y_{ij-1}, but not on future measurements or higher order history, with corresponding regression coefficients, ψ_1 and ψ_2. Dependence on future unobserved measurements is not easy to justify therefore it is not modelled here. Higher order history can be included, but we assume first order history for simplicity. This leads to the following logistic expression

$$logit[g(y_{i,j-1}, y_{ij})] = logit[p(D_i = j \mid y_{i,j-1}, y_{ij})] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij}. \tag{7}$$

Note that the linear predictor in equation (7) may depend on other covariates but in the current model we only include the constant ψ_0. According to Little and Rubin's (1987) terminology introduced in the previous section, and based on the expression in equation (7), it is clear that when both parameters ψ_1 and ψ_2 are equal to 0, the dropout

mechanism should be MCAR. However, when ψ_1 is not equal to 0, but ψ_2 equal to 0, the dropout mechanism is referred to as MAR, and finally, when ψ_2 is not equal to 0, dropout mechanism is referred to as MNAR. Here, we note that a likelihood ratio test (LRT) can be used to compare model fit under a model that assumes the missing data due to dropout are MCAR versus MAR (Diggle and Kenward, 1994). The LRT test statistic follows a null asymptotic χ^2_1 distribution. See, Diggle and Kenward (1994) and Molenberghs et al. (1997) for details on the derivation of this statistic. When the LRT test statistic is significant, then it suggests that the least restrictive of the two models is preferred; that is, the model that assumes the dropout is MAR. However, based on the argument of Jansen et al. (2006), we restate that the test for MAR against MNAR is not recommended using the LRT statistic via a model based on the Diggle and Kenward's (1994) type. This is because the behaviour of the LRT statistic for the MNAR parameter ψ_2 is non-standard since the availability of the information on ψ_2 is very rare and interwoven with other features of both measurement and dropout models (Jansen et al., 2006). This is specially true when one considers the model based on Diggle and Kenward type, but it is important to realize that their tests are conditional on the alternative model holding. According to Kenward (1998), such a distinction, between a MAR mechanism or a MNAR mechanism, can only be made using untestable modeling assumptions, such as the distributional form. Molenberghs and Kenward (2007) stated that the assumption giving arise to the dropout in a sample cannot be verified by the observed measurements and any test regarding the dropout process can be invalidated. This can be justified by the fact that parameters of the dropout model are dependent in part on dropout. Furthermore, unless one puts a priori belief in the posited MNAR model, the distinction (MAR/MNAR) is not possible, due to the fact that for any dropout model that assumes dropout are MNAR, there is a MAR model that provides exactly the same fit to the data, but the two models differ in the prediction of what is unobserved (Molenberghs et al., 2008). This problem of model identifiability poses a major complication when considering models for the dropout mechanism. Thus, one recommendation is to conduct a sensitivity analysis of the parameters of the measurement model across models that make different assumptions about the dropout process (see, Molenberghs and Kenward, 2007). Therefore, although the dropout process cannot be known via empirical examination, the analysis can be carried out to study differences in parameters estimates of the measurement process across varying assumptions about the dropout.

3.2. Pattern mixture model

Now, we shift our attention to the pattern mixture models that stratify subjects according to their missingness pattern. Under these models, the thinking is that, a separate model is fit for each pattern and then the results can be combined across the different patterns in order to derive an average estimate of the model parameters. Thus, in these mod-

els the joint distribution of the longitudinal measurements as well as the missing data indicators is divided into response pattern so that the distribution of the longitudinal measurements depends on the pattern of responses. As mentioned earlier, pattern mixture models are under-identified, or possess non-estimable parameters. Therefore, some identifying constraints are required. Little (1993, 1994) proposed the use of the identifying restrictions in which inestimable parameters of the incomplete patterns are set equal to (functions of) the parameters describing the distribution of the completers to deal with under-identifiability of these models. In fact, there is an alternative major strategy simplified to deal with the under-identifiability of pattern mixture models, called model specification in which the different pattern allows for sharing of certain parameters so that the missing pattern can borrow information from patterns with more data points (Verbeke and Molenberghs, 2000). The advantage of this strategy is that the number of parameters decreases which is in general an issue with pattern mixture models. Detailed strategies of pattern mixture modelling are given in Verbeke and Molebergh (2000), Molenberghs et al. (2003) and Molenberghs and Kenward (2007).

Our primary concern in this study is to apply a pattern mixture model including the identifying restriction strategy. In doing so, we follow Verbeke and Molenberghs (2000) in illustrating the use of this strategy based on the results obtained by Molenberghs et al. (1998b). We are restricting attention to dropout which is a special case of monotone missingness. Let us assume that there are $t = 1,\ldots,T$ dropout patterns, where the dropout indicator, introduced in section 2, is $d = t + 1$. The complete data density, for pattern t, can be expressed as

$$f_t(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T) = f_t(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t) f_t(\boldsymbol{y}_{t+1},\ldots,\boldsymbol{y}_T \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_t). \tag{8}$$

It is clear from equation (8) that the first factor $f_t(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t)$ is identified from the observed data assuming the first factor is known, and modeled using the observed data. Whereas the second factor is not identifiable from the observed data. In order to identify the second component, the identifying restriction can be applied (Verbeke and Molenberghs, 2000). It is often necessary to base identification on all patterns for which a given component is identified. We denote this component by $\boldsymbol{y}_s$. Thus, this can be described as

$$f_t(\boldsymbol{y}_s \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1}) = \sum_{j=s}^{T} \omega_{sj} f_j(\boldsymbol{y}_s \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1}), \qquad s = t+1,\ldots,T. \tag{9}$$

We denote the set of ω_{sj} used by the vector $\boldsymbol{\omega}_s$, components of which are typically non-negative. Every $\boldsymbol{\omega}_s$ that sums to 1 provides a valid identification scheme. Hence, by incorporating equation (9) into (8), we have

$$f_t(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T) = f_t(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t)\prod_{s=0}^{T-t-1}\left[\sum_{j=T-s}^{T}\omega_{T-s,j}f_j(\boldsymbol{y}_{T-s}\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{T-s-1})\right] \tag{10}$$

To establish the complete data density, it is clear in equation (10) whose information can be used to complement the observed data density in pattern t. There are three sets of identifying restrictions associated with such choices of $\boldsymbol{\omega}_s$. Complete case missing values (CCMV) that were proposed by Little (1993) use the following identification

$$f_t(\boldsymbol{y}_s\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1}) = f_T(\boldsymbol{y}_s\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1}), \qquad s = t+1,\ldots,T,$$

corresponding to $\omega_{sT} = 1$ and all others equal 0, which is to say that identification is always done from the completers's pattern. Alternative restrictions are based on so called neighboring case missing values (NCMV). In these restrictions, the nearest identified pattern can be used as follows

$$f_t(\boldsymbol{y}_s\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1}) = f_s(\boldsymbol{y}_s\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1}), \qquad s = t+1,\ldots,T.$$

The NCMV restriction follows from setting $\boldsymbol{\omega}_s = \mathbf{1}$ and all others equal $\mathbf{0}$. Finally, the third case for equation (10) is the available case missing values (ACMV). With regard to the corresponding $\boldsymbol{\omega}_s$ for ACMV, there always is a unique choice. Molenberghs et al. (1998b) show that the corresponding $\boldsymbol{\omega}_s$ can have the following components

$$\omega_{sj} = \frac{\alpha_j f_j(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1})}{\sum_{\ell=s}^{T}\alpha_\ell f_\ell(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{s-1})}, \qquad j = s,\ldots,T, \tag{11}$$

where α_j is the fraction of observations in pattern j. Clearly, ω_{sj} defined by (11) contains positive components and sum to 1. That is, a valid density function is defined. The selection and pattern mixture families can be connected using this MAR-ACMV link. The ACMV is reserved for a counterpart of MAR in the pattern mixture setting. A specific counterpart to MNAR selection models has been studied by Kenward et al. (2003).

4. Application to the NCGS data

4.1. The data

In this section, we describe the application of the selection and pattern mixtures models to data from the National Cooperative Gallstone Study (NCGS). Further background details of this experiment are given in Schoenfield and Lachin (1981) and in its accompanying discussion. In this study, 103 patients were randomly assigned to three

treatment groups corresponding to two doses; that is, high-dose (750 mg per day), low-dose (375 mg per day) and placebo, and were to be treated for four weeks. The current analysis is based on a subset of the data on patients who had floating gallstones and who were assigned to the high-dose and placebo groups. In the NCGS it was suggested that chenodiol would dissolve gallstones but in doing so might increase levels of serum cholesterol. As a result, serum cholesterol (mg/dL) was measured at baseline and at 6, 12, 20 and 24 weeks of follow-up. In this experiment, many cholesterol measurements contain missing values because of missed visits, laboratory specimens were lost or inadequate, or patient follow-up was terminated. In addition, all subjects have observed values at time 6. One group of individuals received study treatment (drug and placebo), but dropped out of the study before the scheduled post-baseline time. These individuals dropped out of the study at time point 12. However, other individuals dropped out of the study either at time point 20 or 24. Therefore, the data presents three possible dropout patterns (dropout at time points 12, 20, or 24). All 103 patients are observed at the first occasion, whereas there are 93, 78 and 67 patients seen at the second, third and fourth weeks, respectively. The percentage of patients that are still in the study after each week is tabulated in Table 1 by treatment arm. Figure 1 represents the means across weeks by treatment group. A primary objective of this trial was to study the safety of the drug chenodiol for the treatment of cholesterol gallstones. In what follows, we restrict our attention to examination of more than just this association between treatment and cholesterol. That is, we investigate the potential influence of dropout on the outcome of interest, the serum cholesterol, as well as the interactive effect of dropout with week and treatment-related influences on the serum cholesterol. The focus here will be on the parameter estimates, standard errors and p-values.

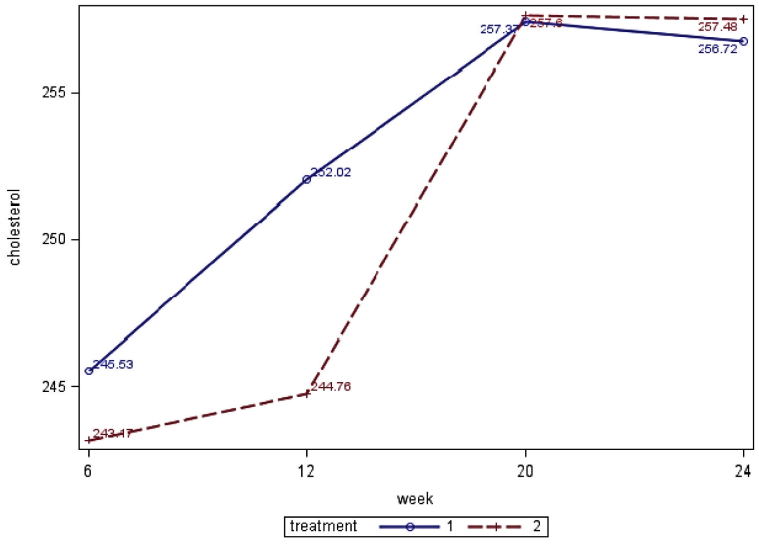

Figure 1: *Serum cholesterol data. Means across weeks by treatment (High dose "1" and Placebo "2").*

Table 1: *NCGS data: Percentage of patients still in study, by treatment arm (Drug = high-dose (750 mg per day)).*

week	drug	placebo
6	100	100
12	45	62
20	57	63
24	46	69

4.2. Fitting selection model

First, we consider fitting the selection model. In line with Diggle and Kenward (1994), we fit the selection models to the serum cholesterol data by combining the measurement model with the logistic regression for dropout model. The combined model for measurement/dropout will be fitted to the serum cholesterol by maximum likelihood using a generic function maximization routine. We use the linear mixed effects model of the form in equation (5) in order to obtain initial values for the parameters of the measurement model. In the fitted model, we assume different intercepts and treatment effects for each of the four time points, with a (4×4) unstructured variance-covariance matrix. Specifically, we consider a multivariate normal model, with unconstrained time trend under placebo and an occasion-specific treatment effect. Since serum cholesterol data consist of 103 subjects ($i = 1, \ldots, 103$) on four time points ($j = 6, 12, 20$ and 24), the model can be written as

$$Y_{ij} = \beta_{j1} + \beta_{j2} G_i + \varepsilon_{ij}, \tag{12}$$

where $G_i = 0$ for placebo and $G_1 = 1$ for active drug. In this way, the parameter estimates and standard errors as well as p-values for the eight mean model parameters can be obtained. To fit this model, we use SAS procedure MIXED with REPEATED statement. Next, we consider the dropout model. The dropout will be allowed to be independent of covariates. We fit the model with an intercept, an effect for previous outcome and an effect for the current unobserved measurement, corresponding to MCAR, MAR and MNAR, respectively. Dependence on future unobserved measurements is theoretically possible, but for simplicity, we model dependence on the current unobserved measurements. The probability of serum cholesterol is assumed to follow the logistic regression model (a commonly used model for dropout process, see, Molenberghs and Kenward, 2007) in equation (7). Therefore, the logistic regression model consists of three parameters; that is, an intercept (ψ_0), the effect of the measurement prior to dropout (ψ_1) and the effect of the measurement at the time of dropout (ψ_2). Consequently, for the four time points, the model can be expressed as follows

$$logit[g(y_{ij-1}, y_{ij})] = logit[p(D_i = j \mid y_{ij-1}, y_{ij})] = \psi_0 + \psi_1 y_{ij-1} + \psi_2 y_{ij}, \quad j = 2,3,4,5. \tag{13}$$

Estimation of a selection model for MNAR can be seen as major complication as the dropout indicators depend on the unobserved measurement. For example, in the selection model mentioned above, the dropout indicators depend in part on the unobserved longitudinal measurements at the time of dropout. This leads to complexity in assessing the likelihood function, however, one that can be handled (Diggle and Kenward, 1994). Virtually, the parameters were estimated using a code written in SAS provided by Dmitrienko et al. (2005) that maximizes the log-likelihood for the model using PROC IML.

Table 2 shows the parameter estimates, standard errors and p-values of the fixed effects for the selection model, including the eight mean model parameters, all into the marginal measurement model as well as in the logistic dropout model. Interestingly, the comparison of the MCAR and MAR produces the same results when compared to those of the complete case analysis, except for negligible differences, as seen in the standard errors. These results are in line with theoretical findings, see, for example, Molenberghs and Kenward (2007). In the context of the assumed model, when examining the statistical significance of the results in the dropout model, the LRT test statistic for comparing the MAR and MCAR models is 17.1. The corresponding tail probability from χ^2 with 1 degree of freedom is $p < 0.001$ which is significant. This indicates that there is a significant evidence for MAR. In other words, dropout completely at random can be ruled out in the context of the assumed model. However, great care has to be taken with such a conclusion (Molenberghs et al., 1997; Molenberghs and Verbeke, 2005). To assess the mechanism that the dropout are MNAR, a problem occurs in that neither an LRT statistic between the models that assume the dropout is MAR against MNAR nor an assessment of ψ_2 relative to its standard error is reliable (Jansen et al., 2006). Consequently, it is not possible to verify the mechanism that the dropout is MNAR (see, Molenberghs, et al., 2008). One of our interests lies in the marginal treatment effect. There is no overall treatment effect and p-values between the three models do not vary too much.

However, the situation is different for the occasion-specific treatment effects considered here. For all weeks, all four p-values for the treatment effects indicate non-significance, whereas for all cases the p-values are certainly highly significance ($p < 0.0001$) for all intercepts. Now, we discuss factors which influence dropout. In doing so, in the full selection models, the logistic regression for dropout is modeled based on (13). As can be seen in Table 2, the maximum likelihood estimates for ψ_1 (0.04) and ψ_2 (-0.16) are not necessarily equal, however, their signs are different. This finding is not surprising. It confirms the argument put forward by Molenberghs et al. (2001a). They pointed out that since two subsequent measurements are usually positively correlated, the dropout model can depend on the increment, i.e., $y_{ij} - y_{i,j-1}$. The dropout estimated from the MNAR model is as follows:

Table 2: *Maximum likelihood for the parameter estimates-Est. (standard errors-s.e.) and p-values, resulting from the selection model under complete cases analysis, MCAR, MAR and MNAR.*

		Complete cases		MCAR		MAR		MNAR	
Effect	Parameter	Est.(s.e.)	p-value	Est.(s.e.)	p-value	Est.(s.e.)	p-value	Est.(s.e.)	p-value
Measurement model									
intercept$_6$	β_{11}	243.17 (6.74)	< 0.0001	243.17 (6.74)	< 0.0001	243.17 (6.74)	< 0.0001	243.17 (6.74)	< 0.0001
intercept$_{12}$	β_{21}	244.93 (6.46)	< 0.0001	244.93 (6.46)	< 0.0001	244.93 (6.45)	< 0.0001	243.98 (6.46)	< 0.0001
intercept$_{20}$	β_{31}	258.92 (6.70)	< 0.0001	258.92 (6.69)	< 0.0001	258.91 (6.70)	< 0.0001	258.13 (6.70)	< 0.0001
intercept$_{24}$	β_{41}	257.08 (8.03)	< 0.0001	257.08 (8.02)	< 0.0001	257.08 (8.03)	< 0.0001	256.28 (8.99)	< 0.0001
treatment effect$_6$	β_{12}	2.36 (8.69)	0.786	2.36 (8.69)	0.784	2.36 (8.68)	0.788	2.36 (8.69)	0.732
treatment effect$_{12}$	β_{22}	6.41 (8.36)	0.445	6.41 (8.32)	0.478	6.41 (8.36)	0.461	6.54 (8.36)	0.441
treatment effect$_{20}$	β_{32}	–5.77 (8.78)	0.512	-5.77 (8.78)	0.535	–5.77 (8.79)	0.427	–5.83 (8.79)	0.456
treatment effect$_{24}$	β_{42}	–2.06 (10.70)	0.847	–2.06 (10.70)	0.883	–2.06 (10.74)	0.856	–2.59 (10.79)	0.912
dropout model process									
intercept	ψ_0			–1.88(0.11)		–1.73(0.14)		–1.64(0.27)	
previous measurement	ψ_1					–0.20(0.05)		0.04(0.02)	
current measurement	ψ_2							–0.16(0.08)	
–2 log-likelihood		3313.3		3346.4		3329.3		3327.7	

$$logit[p(D_i = j \mid y_{ij-1}, y_{ij})] = -1.64 - 0.12y_{i,j-1} - 0.16(y_{ij} - y_{i,j-1}). \quad (14)$$

However, some insight into this fitted model can be obtained by the re-parameterizing the dropout parameters with respect to increment and the sum of the successive measurements. Therefore, we re-parameterize the dropout probabilities from the dropout model as in equation (13) to obtain

$$logit[p(D_i = j \mid y_{ij-1}, y_{ij})] = \vartheta_0 + \vartheta_1(y_{i,j} + y_{i,j-1}) + \vartheta_2(y_{ij} - y_{i,j-1}), \qquad j = 2,3,4,5. \quad (15)$$

Here, $\vartheta_1 = (\psi_1 + \psi_2)/2$ and $\vartheta_2 = (\psi_1 - \psi_2)/2$. These parameters represent dependence on level and increment in the serum cholesterol, and these quantities are likely to be much less strongly correlated than are y_{ij} and $y_{i,j-1}$. Rewriting the fitted MNAR model as in equation (15),

$$logit[p(D_i = j \mid y_{ij-1}, y_{ij})] = -1.64 - 0.06(y_{i,j} + y_{i,j-1}) + 0.10(y_{ij} - y_{i,j-1}), \quad (16)$$

suggests that the probability of dropout increases with larger negative increments. In other words, those patients with a greater increase in the overall level of the serum cholesterol from the previous week have a higher probability of dropping out of the experiment.

4.3. Fitting pattern mixture models

Now, we turn our attention to fitting the pattern mixture models using the strategy outlined in section 3, making CCMV, NCMV and ACMV identifying restrictions. To fit pattern mixture models through identifying restrictions, three steps in the analysis procedure are needed (For details of implementation, see Molenberghs and Kenward, 2007). First, fit the initial model to the observed data within each of the patterns

$$f_t(y_1, \ldots, y_t), \quad (17)$$

where $t = 1, \ldots, T$ indicate the observed dropout times in the data set. In this step, we fit a separate model within each pattern, then the resulting parameter estimates and their estimated variance-covariance matrices were used to extrapolate the patterns. Second, select an identification scheme to determine the conditional distributions of the unobserved measurements, given the observed ones

$$f_t(y_{t+1}, \ldots, y_T \mid y_1, \ldots, y_t). \quad (18)$$

Table 3: *Multiple imputation parameter estimates-Est. (standard errors-s.e.) and p-values resulting from the pattern mixture model using identifying restrictions ACMV, CCMV and NCMV.*

		ACMV		CCMV		NCMV	
Effect	Parameter	Est.(s.e.)	p-value	Est.(s.e.)	p-value	Est.(s.e.)	p-value
intercept$_6$	β_{11}	243.17 (6.74)	—	243.17 (6.74)	—	243.17 (6.74)	—
intercept$_{12}$	β_{21}	245.44 (7.06)	< 0.0001	245.36 (6.51)	< 0.0001	245.86 (6.55)	< 0.0001
intercept$_{20}$	β_{31}	255.78 (6.71)	< 0.0001	255.88 (6.83)	< 0.0001	257.99 (6.78)	< 0.0001
intercept$_{24}$	β_{41}	256.59 (8.10)	< 0.0001	256.99 (8.21)	< 0.0001	256.99 (8.08)	< 0.0001
treatment effect$_6$	β_{12}	2.36 (8.69)	—	2.36 (8.69)	—	2.36 (8.69)	—
treatment effect$_{12}$	β_{22}	6.23 (8.39)	0.540	6.16 (8.37)	0.539	5.41 (6.45)	0.716
treatment effect$_{20}$	β_{32}	–5.98 (8.85)	0.484	–5.13 (8.81)	0.475	–6.73 (8.82)	0.290
treatment effect$_{24}$	β_{42}	–2.18 (11.03)	0.627	–2.12 (11.64)	0.565	–1.87 (10.13)	0.629

As stated earlier, each of such conditional distributions is a mixture of known normal densities for continuous repeated measures. According to the weights w_s introduced in equation (9), an easy way to simulate values from the mixture distribution is to randomly select a component of the mixture and then draw from it. In this regard, we choose an identifying restriction, mentioned earlier, to define the conditional distributions of the unobserved measurements, conditional upon the observed ones. Third, fit a model to the so-augmented data. Multiple imputation (MI) can be used to fit such models by aiding to draw values for the unobserved components, conditional upon the observed outcomes and correct pattern-specific density in model (18). Here, we notice that MI is a simulation-based technique that imputes the missing values multiple times in order to construct multiple complete data sets. For more detail of this technique, we recommend Rubin's (1987) book. Analytically, MI involves three steps, imputation, analysis and combination. Thus, the identifying step corresponds to the so-called imputation step, and the final model corresponds to the analysis step. Finally, the combination step, is where the inferences from a number of imputations are drawn together and combined into a single one. The goal being to pool the simplicity of imputation strategies, without bias in both point estimates and measures of precision. After applying each of the three identifying restrictions, as above introduced, the same model as before being fitted (12) is analyzed. The model is parameterized as follows: different intercepts and treatment effects for each of the four time points, with a 4×4 unstructured covariance matrix for each pattern. We draw multiple imputations five times. The choice of five times imputations is considered adequate as the efficiency of a parameter estimate based on the number of imputations is $(1+\varsigma/M)^{-1}$, where ς is the rate of missing data and M is the number of imputations (Rubin, 1987). Rubin's (1987) simulation studies indicate that the number of imputations can generally be constrained to fewer than 10. Also, many statistical practices tend to support Rubin's heuristics of 3 to 10 imputations. In general, Schafer and Olsen (1998) recommended the use of M=5 before the results are combined. By this rationale, we achieve at least 97% efficiency as in our case the missing data rate is almost 17%. In this way, we ended up with totally five multiply-imputed data sets for each choice of identifying restriction strategy which can be analysed, using several possible models. Once the imputations have been generated, the final analysis model from each completed data sets is fitted and MI inference conducted. The parameter and precision estimates can be obtained using classical MI machinery. In particular, the asymptotic covariance matrix of the form

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

where W denotes the average within-imputation variance and B the between-imputation variance (Rubin, 1987). The analysis of identifying restrictions, fitting of imputed data, and a combination of the results into a single inference was implemented using the SAS macro. This SAS macro dealt with the analysis of the three types of identifying

restrictions as follows. First, fit the linear mixed model per pattern using PROCs SORT and MIXED. Second, complete the data using ACMV, CCMV and NCMV restrictions using PROCs IML and MI. Third, analyze the 5 complete data sets using a linear-mixed model using PROC MIXED. Fourth, combine the results from the 5 analyzed models using PROC MIANALYZE.

The results of the three types of identifying restrictions are listed in Table 3. Examining these results we see that the estimates for the corresponding parameters are comparable and their numerical values are indeed very close to each other under the three possible restrictions, namely ACMV, CCMV and NCMV. It can be seen from the analysis that the association p-values for the marginal effect assessments are all nonsignificant, their p-values being all greater than 0.05. However, the association p-values for the intercepts are highly significant ($p < 0.0001$), in line with the p-values obtained from the selection model analysis. In summary, no significant treatment effect is obtained. These findings confirm those obtained from the selection model formulation which gives more weight to this conclusion. These results can be justified by the fact that pattern mixture models using identifying restrictions play a very similar role to the modelling assumptions in the selection model case (Michiels et al., 1999). Furthermore, the parameter estimates and standard errors for the first marginal effect are equal for all the three restrictions CCMV, NCMV and ACMV, see the effects for intercept_6 and treatment_6. Such results should be expected considering the fact that the first outcome contained observed data for all subjects that were considered in the analysis.

As shown in the results in Table 3, the model building using CCMV, NCMV and ACMV restrictions in contrast to selection model did not allow an estimation of whether the dropout process is MNAR or not, because of differences in the modelling assumptions. This agrees with previous studies, see, for example, Molenberghs et al. (1998b), in that the identifying restrictions in a pattern mixture models context can be used only to relate the model to a MAR mechanism. Consequently, an important issue is to equate results for both the ACMV and MAR to make a clear and useful connection between the selection model and the pattern mixture model framework (Verbeke and Molenberghs, 2000; Kenward et al., 2003). With this in mind, the same is true for the selection model, MAR-based ACMV restrictions indicating non-significant treatment effects at all weeks. This can be explained to mean that the treatment effect appear to be independent of the ACMV (MAR) assumption. Although corresponding models include the same effects, the estimates for ACMV are slightly different to those for MAR. These slight differences are to be expected as argued in Kenward et al. (2003) that both models are similar in spirit but not necessarily identical. On the other hand, the parameter estimates and standard errors for the treatment effects obtained by applying NCMV are smaller than those of CCMV and ACMV as seen in some cases. This is to be expected as somewhat CCMV and ACMV pattern mixture models use data from different patterns to multiply impute new values, whereas in NCMV, pattern mixture models take information from the neighboring case patterns only. Further, ACMV and CCMV estimates are closer to each other since many more completers are available than

does NCMV. Therefore, additional variability may be introduced because, depending on the nature of the conditional distributions sampled from, data have been borrowed from more distant patterns.

5. Discussion

In this study, we demonstrated the application of two families of models for analysing incomplete longitudinal data, where the dependent variable is missing across time. In particular, we illustrated the application and compared results of analysis using these models. We focused on the situation in which outcomes are continuous. The models that were considered were the selection model and the pattern mixture model. Many authors have recommended fitting both families of models to be able to gain extra insight into the data to assess sensitivity to the modelling assumptions and to assess the extent of agreement in results as well (see, Molenberghs and Verbeke, 2005). The study focused on the specific cases of selection model and pattern mixture models; that is, a Diggle and Kenward's (1994) model and an identifying restrictions strategy (Little, 1993, 1994), respectively. In applying the selection model, we used logistic regression for modelling dropout, however, a number of other probabilities can be used, for example, using survival analysis techniques, the length of duration on treatment or placebo before dropout can also be modelled. However, in this study, the survival model for dropout cannot be used because the time to event (dropout) is not exactly determined by design. For example, if someone is not seen at week 12, the exact time to dropout could theoretically be any time between week 6 and 12. The objective was to investigate the potential influence that dropout might have or exert on the dependent measurement on the considered data and to deal with incomplete sequences. The results from the pattern mixture models were analogous to those from the selection model to obtain additional insights into the serum cholesterol data. The application was based on an example from a longitudinal clinical trial data.

Findings in general suggested that the conclusion obtained under both modelling frameworks practically coincide. Thus, one can put more confidence in these results as argued by many authors. For example, Michiels et al. (1999, 2002) have argued that greater confidence in a conclusion can be reached when the analysis of joint applications of these models leads to essentially similar inference. Both families of models were compared and noticeable similarities in results were found. Hence, this begs the question as how, depending on the scientific question of interest such as conditional measurement probabilities, to choose between them. Michiels et al. (1999) argued that the selection model can be recommended as a natural choice when the interest is in the population as a whole, i.e., marginal effects. Whereas, pattern mixture models can be considered, when investigating the differences between subgroups that are identified by their measurement patterns, i.e., pattern-specific.

The selection models suggested that the dropout mechanisms were not completely at random. In other words, in the context of the assumed model, there was a lot of evidence in favour of the prevalence of an MAR rather than an MCAR dropout process. However, many authors, Diggle and Kenward (1994) and Molenberghs and Verbeke (2005) for example, stated that careful consideration is necessary with such a conclusion when using only the data under analysis. A problem arises for dealing with dropout that are MNAR. Given this problem in a longitudinal study, it is important to realize that this assumption gives rise to the dropout that is not likely to be known in the application setting. Therefore, any of the different proposed application methods to address dropout that are MNAR cannot easily be verified. For example, one often does not know if the dropout process is accurately captured by a particular method used. Molenberghs and Kenward (2007) suggested that one should apply several approaches to the same data problem. This is the case when the sensitivity of parameters estimates to the different mechanisms about the dropout process may be investigated, for example, by applying models that allow for the dropout to be MNAR. According to Xu and Blozis (2011), if parameter estimates are comparable under different methods, this can indicate that the dropout process may be ignored. However, if different methods give different parameters estimates of the longitudinal model, this can indicate that the dropout process is a vital element for the description of the data in the analysis.

The structure of the selection dropout model adopted that dropout increases with a unit change in the serum cholesterol; that is, the dropout is related to the larger negative increments $(y_{ij} - y_{i,j-1})$ rather than to any actual observation $(y_{ij} + y_{i,j-1})$, which implies that patients with a greater decrease in the overall level of the serum cholesterol from the previous week have a higher probability of dropping out of the experiment. This situation is very common in practice within a model of the Diggle and Kenward type, and we refer to Molenberghs and Kenward (2007), Diggle and Kenward (1994) and Molenberghs et al. (1997) as examples. Under the modeling scheme applied in this study, it can be seen from the analysis that the treatment effects over all weeks under all ACMV, CCMV and NCMV restrictions were non-significant, and the same is true for the selection model analysis. Therefore, it is clear that there is a strong evidence for no significant treatment in the context of serum cholesterol data. It appeared that the non-significant treatment effects were not conditional upon any dropout mechanism holding. As a results, the conclusions obtained from CCMV, NCMV and NCMV restrictions did not differ considerably. As argued in Molenberghs et al. (2008), the choice between them is not always clear. Although they fit the observed data equally well, the difference between them only becomes clear with respect to estimation of the missing data, conditional upon the observed data.

On the other hand, the use of different models in which the data were analysed, can be considered as a sensitivity analysis. In particular, the use of pattern mixture models including identifying restrictions can be seen as a first tool for assessing the sensitivity of the assumptions made. Further, other more complex or flexible sensitivity analysis are also possible, under new models for the probability of dropout. The

analysis conducted here is a typical sensitivity analysis as the serum cholesterol data were analyzed using different assumptions concerning the longitudinal measurements and dropout mechanisms. In particular, both models compared well concerning some aspects, for example, marginal treatment effects. Such comparisons as these can play a vital role in sensitivity analysis by providing additional motivation, for example, when considering the choice between selection and pattern mixture models. In conclusion, because the true model and measurement process as well as dropout process are often unverifiable, the recommendation that in many settings, multiple strategies or models such as selection and pattern mixture models be applied to the same data set in order to investigate the impact of assumption on dropout or missingness is supported.

References

Daniels, M. and Hogan, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. CRC: Chapman and Hall.

Diggle, P. J. and Kenward, M. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, 43, 49–94.

Dmitrienko, A., Offen, W. W., Faries, D., Christy Chuang-Stein, J. L. and Molenberghs G. (2005). *Analysis of Clinical Trial Data Using the SAS System*, Cary, NC: SAS Publishing.

Ekholm, A. and Skinner, C. (1998). The muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics*, 47, 251–263.

Glynn, R. J., Laird, N. M. and Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In: *Drawing Inferences from Self-Selected Samples*. New York: Springer-Verlag.

Heckman, J. J. (1976). The common structure of statistical models of trucation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5, 475–492.

Hedeker, D. and Gibbons, R. D. (1997). Application of random-effects pattern mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 64–78.

Hogan, J. and Laird, N. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, 16, 259–272.

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G. and Kenward, M. G. (2006). The nature of sensitivity in missing not at random models. *Computational Statistics and Data Analysis*, 50, 830–858.

Kenward, M. and Molenberghs, G. (1999). Parametric models for incomplete continuous and categorical longitudinal data. *Statistical Methods in Medical Research*, 8, 51–83.

Kenward, M., Molenberghs, G. and Thijs, H. (2003). Pattern mixture models with proper time dependence. *Biometrika*, 90, 53–71.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.

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. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112–1121.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Little, R. J. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, 52, 98–111.

Marini, M. M., Olsen, A. R. and Rubin, D. B. (1980). Maximum likelihood estimation in panel studies with attrition. *Sociology Methodology*, 11, 314–357.

McArdle, J. J. and Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using laten growth structural models. *Experimental Aging Research*, 18, 145–166.

Michiels, B., Molenberghs, G. and Lipsitz, S. R. (1999). Selection models and pattern mixture models for incomplete data with covariates. *Biometrics*, 55, 978–983.

Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T. and Thijs, H. (2002). Selection models and pattern-mixture models to analyze longitudinal quality of life data subject to dropout. *Statistics in Medicine*, 21, 1023–1042.

Molenberghs, G., Michiels, B. and Kenward, M. (1998a). Pseudo-likelihood for combined selection and pattern-mixture models for missing data problems. *Biometrical Journal*, 40, 557–572.

Molenberghs, G., Michiels, B., Kenward, M. and Diggle, P. J. (1998b). Missing data mechanism and pattern mixture models. *Statistica Neerlandica*, 52, 135–161.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

Molenberghs, G. and Kenward, M. (2007). *Missing Data in Clinical Studies*. New York: Wiley.

Molenberghs, G., Thijs, H., Kenward, M. G. and Verbeke, G. (2003). Sensitivity analysis of continuous incomplete longitudinal outcomes. *Statistics Neerlandica*, 57, 112–135.

Molenberghs, G., Kenward, M. G. and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with non-random dropout. *Biometrika*, 84, 33–44.

Molenberghs, G., Verbeke, G., Thijs, H., Lesaffre, E. and Kenward, M. (2001). Mastitis in dairy cattle: influence analysis to assess sensitivity of the dropout process. *Computational Statistics and Data Analysis*, 37, 93–113.

Molenberghs, G., Beunckens, C., Sotto, C. and Kenward, M. (2008). Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of Royal Statistical Soceity*, 70, 371–388.

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

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

Schafer, J. L. and Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analysts perspective. *Multivariate Behavioral Research*, 33, 545–571.

Schoenfield, L. J. and Lachin, J. M. (1981). The steering committee, and the NCGS group, "Chenodiol (Chenadeoxycholic Acid) for Dissolution of Gallstones: The National Cooperative Gall-stone Study". *Annals of Internal Medicine*, 95, 257–282.

Verbeke, G., Lesaffre, E. and Spiessens, B. (2001). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, 35, 419–439.

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

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44, 175–188.

Wu, M. C. and Bailey, K. R. (1988). Analysis changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 7, 337–346.

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.

Xu, S. and Blozis, S. A. (2011). Sensitivity analysis of mixed models for incomplete longitudinal data. *Journal of Educational and Behavioral Statistics*, 36, 237–256.