Research Report

Department of Statistics

No. 2008:3

Longitudinal, Model-Based Clustering with Missing Data

Jessica Franzén

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Longitudinal, Model-Based Clustering with Missing Data

Jessica Franzén*
Department of Statistics
University of Stockholm

April 2008

Abstract

Non-response is a frequent problem when analysing repeated measurements of multidimensional data. We evaluate a multiple imputation method used in the process of clustering an incomplete longitudinal data set. A model-based, longitudinal clustering method is used. At each time, we assume data to be generated from a mixture model of multivariate normal distributions. Each component of the distribution corresponds to a cluster with cluster-specific parameters. We show that a model-based MCMC clustering approach, can easily and effectively be extended to deal with missing data. Instead of imputing values before analysing them, which is the common imputation procedure, we impute missing values within the model. Missing values are imputed as an iterative step in the Gibbs sampler algorithm, used to estimate the model parameters. The method is applied to two simulated data sets. The method is shown to handle non-response rates up to 40-50 percent without serious loss of precision in estimates. A comparison is made with the mean imputation method, with favourable results for the method of this paper. A real data set consisting of school childrens' attitudes towards three school subjects and their marks in the same subjects is the object of the last part of this paper. The results show more stable estimates, with lower simulation variance when all 1206 individuals are included in the estimation process through imputation, compared to when only the 720 individuals with a complete data set were included..

Keywords: Missing data, Longitudinal, Multiple imputation, Transition matrix, Cluster analysis, Mixture model, Gaussian, Bayesian inference, Clustering, Classification, Gibbs sampler.

*The support from the Swedish Research Council (Grant no 2005-2003) is gratefully acknowledged. Gratitudes to professor Lars Bergman for sharing the IDA data base.

1 Introduction

Non-response is a frequent problem in multivariate data. This is particularly problematic in longitudinal studies. Franzén (2008) studied a model-based approach of longitudinal data over two time points where about 40 percent of the individuals had to be discharged because of one or more missing values. Here, we develop the method to use all available data. The underlying method of this paper is a model-based approach to find group patterns in longitudinal data in several dimensions. Within this method, missing data is imputed as a step in the clustering/estimation process. The aim of this paper is to test the performance of this imputation method. We analyze how the method handles different levels of non-response. The performance is compared to mean-imputation and also to the basal alternative of reducing data by deleting individuals with missing values.

Finite mixture-models are powerful and flexible tools in various classification problems, as they are capable of modelling a wide range of densities. Cross-sectional clustering under the assumption of a mixture-model has been the focus of many papers such as McLachland and Peel (2000), Banfield and Raftery (1993), Bensmail et al. (1997) and, as mentioned above, Franzén (2008). Longitudinal clustering with the mixed-model approach is much more rare in the literature. One example among a few is Scott et al. (2005), where data is clustered at several time points. Transition patterns between clusters at different time points are studied as well as the development of single individuals. Despite method, repeated multivariate measures are often subject to incompleteness. Item non-response and/or partial non-response within items will mostly complicate, both the data analysis and the statistical inference, and threaten the validity of a study.

Most standard statistical methods require complete data. Incomplete data is often dealt with by deletion, where all individuals with missing values are simply excluded. In a longitudinal study, this means that an individual with one or more missing variables for at least one time point, is removed from the data set. This may drastically reduce the data set and worsen the result of the analysis. Valuable information is wasted when individuals with an almost complete variable set are removed, which may easily result in biased estimates. A well-functioning imputation method may improve the result considerably. Little and Rubin (2002) give a comprehensive description of missing data and possible measures.

Many popular methods for imputing missing data in longitudinal studies are based on the assumption of a linear growth curve model for the whole data set. Such a model assumes that data is a linear function of covariates and design variables: see for example Laird (1988), Little (1995), Liu et al. (2000), and Gilks et al. (1993). In the first three papers, the estimates are done using maximum likelihood, where Gilks et al. use Bayesian inference. In this paper, we are not trying to find a linear development pattern to use when imputing values. Instead we make a classification of data at each time point and use each individual's group membership in the imputation process.

Imputation of missing data in longitudinal studies may be done cross-sectionally or longitudinally. The first approach imputes values based on other values from that particular time point, while the latter also uses information from previous and/or future times. Twisk and de Vente (2002) and Engels and Diehr (2003) compare different cross-sectional methods (mean of serie, hot-deck, linear regression) with longitudinal methods (last value carried forward, linear interpolation, longitudinal linear regression). Both papers conclude that longitudinal imputation is preferred over cross-sectional for the methods tested. In this paper, a longitudinal approach is used. When an individual is allocated to a cluster, this is done simultaneously for all time points. Information from all times are taken into consideration when allocating an individual to its clusters throughout times and when imputing missing values.

Missing data imputation under the assumption of a multivariate normal model is well studied: see for example Schafer (1997), Liu (1999), and Gahramani and Jordan (1994). These papers all use the *Expectation-Maximization* (EM) algorithm to estimate the parameters of the cluster model. The EM algorithm finds the maximum likelihood estimates of the model parameters. An alternative to the EM algorithm is Bayesian inference. The Gibbs sampler is a Bayesian simulation technique which iteratively draws samples from the full conditional distributions of the parameters of interest. The posterior distributions are expressed conditional on the other parameters in the model. The parameter value simulated from its distribution in one iterative step, is used as a conditional value in the next step. Replicating the process, generates a random sample from each parameter distribution. Lin et al. (2006) compare *Mean Imputation* (MI) with imputation methods using EM, and *Data Augmentation* (DA), where DA is a special form of Gibbs sampler. The MI method is outperformed by the EM and DA methods. Furthermore, the DA imputation shows promising accuracy in the prediction of missing values when compared to the EM imputation, especially when the missing value rate becomes high.

In this paper, we combine two goals: Classification of longitudinal data and handling of missing values in the data set. Each individual is classified at each time point and in the longitudinal analyses, one learns how subjects move between groups over time, and how group structures change as time passes. We take the missing data into account at the time of the analysis. The technique simultaneously estimates the model parameters and imputes missing values. At each time point, data is assumed to be generated from a mixture of multivariate normal distributions. We cluster data in a longitudinal manner by taking information from all time points into account. Our Bayesian approach to cluster analysis provides a good method for handling missing data, provided the data is *missing at random* (MAR) or *missing completely at random* (MCAR). Under these circumstances, it is fairly easy to include imputation into the analysis as a step in the Gibbs sampler algorithm.

In Section 2, the mixture model is explained and the model notations are introduced. Section 3 deals with the missing value issue, the mechanism behind it, and how the missing values of an individual are distributed conditional on the observed values and its cluster membership. The Gibbs sampler is well suited to simultaneously estimating model parameters and imputing missing values within the algorithm. In Section 4, an explanation of Gibbs sampler and its simulation steps are given. In Section 5, we test the method on two simulated data sets generated from three time points. For each data set, imputation is made for different non-response rates, and the estimation accuracy is the focus when evaluating the results. In the same section, a comparison is made with the mean imputation method as well as the approach of deleting all individuals with missing values. In Section 6, we apply the method on a real data set consisting of 1206 school students' attitudes and grades, collected when they were in third grade and then again in sixth grade. We do the analysis with and without imputation. When deleting individuals with missing variables, we reduce the data set to 720 individuals. Asides from the Appendices, this paper ends with Section 7, where concluding remarks on the study are given.

2 Model Specification

In this section, we describe the situation with complete data, further described in Franzén (2008). Developments for missing data are given in the next section.

We base the cluster analysis on a probability model, where each cluster is represented by a distribution with its specific parameters. Given a certain time point, the population of interest consists of a known number of subpopulations. This can be described as data coming from a mixture distribution. We give the formal notations for the model below.

A sample with n individuals is observed at T different time points. The vector $\mathbf{y}_i^{(t)}$ denotes the true values for individual i at Time t. At each time, each individual is assumed to belong to one of $J^{(t)}$ groups or clusters. If the individual belongs to group j, his values on the variables are assumed to follow a normal distribution with mean $\boldsymbol{\mu}_j^{(t)}$ and covariance matrix $\boldsymbol{\Sigma}_j^{(t)}$. In other words, the data at Time t comes from a mixture of $J^{(t)}$ multivariate normal distributions, each with its specific mean vector and covariance matrix. Each distribution represents a cluster. We introduce one vector $\mathbf{V}^{(t)}$ for each time point, containing indicator variables such as $v_i^{(t)} = j$ if individual i at Time t is a member of Cluster j. The model for an arbitrary individual i at Time t, conditional on its cluster membership, may be expressed as:

$$\mathbf{y}_i^{(t)} \Big| \left\{ v_i^{(t)} = j \right\} \sim N_M \left(\boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)} \right)$$

The true membership of individuals are unknown, i.e. the $v_i^{(t)\prime}s$ are not observed.

We assume random cluster membership, where the probability of belonging to a Cluster j at Time t is the same for all individuals, formally expressed as,

$$p\left(v_i^{(t)} = j\right) = \omega_j^{(t)}$$

Conditional on the time point, we may interpret data as being a sample from a mixed population with proportions $\omega_1^{(t)}, ..., \omega_{J^{(t)}}^{(t)}$. An individual may potentially be a member of any cluster, and this is expressed through the mixture distribution

$$y_i^{(t)} \sim \sum_{j=1}^{J^{(t)}} \omega_j^{(t)} N_M\left(\boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)}\right)$$

Going from one time point to another, individuals remain in the same cluster or move to another according to a Markov process with transition matrix $\mathbf{Q}_t$. The transition matrix $\mathbf{Q}_t$ contains the transition probabilities $q_{jk} = p\left(v_i^{(t+1)} = k \middle| v_i^{(t)} = j\right)$ between Time t and $t+1$. Given a classification in Cluster j at Time t, the probability of being classified into Cluster k at Time $t+1$ is q_{jk}. The size of the $\mathbf{Q}_t$ matrix is $\left(J^{(t)}, J^{(t+1)}\right)$, i.e. the number of rows in $\mathbf{Q}_t$ are the same as the number of clusters at Time t, and the number of columns are equal to the number of clusters at Time $t+1$. Each row in $\mathbf{Q}_t$ sums to 1.

The cluster probabilities at Time $t+1$, $\boldsymbol{\Omega}^{(t+1)} = \left[\omega_1^{(t+1)}, ..., \omega_{J^{(t+1)}}^{(t+1)}\right]$, are direct functions of the probabilities at the previous time $\boldsymbol{\Omega}^{(t)}$, and the transition probabilities in $\mathbf{Q}_t$ according to

$$\boldsymbol{\Omega}^{(t+1)} = \left[\omega_1^{(t+1)}, ..., \omega_{J^{(t+1)}}^{(t+1)}\right] = \boldsymbol{\Omega}^{(t)} \cdot \mathbf{Q}_t$$

In the analysis to follow, we are to estimate the model parameters $\boldsymbol{\mu}_j^{(t)}$, $\boldsymbol{\Sigma}_j^{(t)}$, and $\omega_j^{(t)}$ for all j within all t as well as $\mathbf{Q}_t$ for $t = 1, ..., T-1$. The collection of these four kinds of parameter will be given the catch-all denotation $\boldsymbol{\theta}$. The classification vectors $\mathbf{V}^{(t)}$ for $t = 1, ..., T$ play an active part in the estimation process described in Section 4.2. When an individual is classified to a cluster, this is done simultaneously for all time points. Instead of making the classification for each time point separately based on data from that time point only, we take data from all time points into consideration in the classifying process. An individual's cluster memberships are decided simultaneously for all time points. We use the indicator $\delta_{i,j^{(1)},j^{(2)}...,j^{(T)}}$ to describe individuals development over time. $\delta_{i,j^{(1)},j^{(2)},...,j^{(T)}} = 1$ when observation i belongs to Cluster $j^{(1)}$ at Time 1, and Cluster $j^{(2)}$ at Time 2, until the last Time point T, when it belongs to Cluster $j^{(T)}$. The indicator probabilities are the basis for the simulation of the classification matrix $\mathbf{V}$. According to Bayes' rule we may express the conditional probability for a specific development pattern for individual i given the data and the parameters as:

$$P\left(\delta_{i,j^{(1)},...,j^{(T)}} = 1 \left| \mathbf{y}_i^{(1)}, ..., \mathbf{y}_i^{(T)}, \boldsymbol{\theta}\right.\right) =$$

$$\frac{\omega^{(1)}_{j^{(1)}} \cdot \prod_{t=1}^{T-1} q_{j^{(t)},j^{(t+1)}} \cdot \prod_{t=1}^{T} f_j^{(t)}\left(\mathbf{y}_i^{(t)} \left| \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)}\right.\right)}{\sum_{j^{(1)},...,j^{(T)}} \left(\omega^{(1)}_{j^{(1)}} \cdot \prod_{l=1}^{T-1} q_{j^{(t)},j^{(t+1)}} \cdot \prod_{t=1}^{T} f_j^{(t)}\left(\mathbf{y}_i^{(t)} \left| \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)}\right.\right)\right)} \quad (1)$$

for $i = 1, ..., n$ and all possible combinations of $j^{(1)}, ..., \ j^{(T)}$.

3 Missing Values

There may be many reasons for missing data. Refusals and missed or overlooked questions are causes directly connected to the respondent. Other causes may be information not available, inapplicable questions, or errors in data entry. In behavioral longitudinal studies it is unlikely that every individual's variable set will be complete at all prespecified times. The default option for handling missing data is often listwise deletion. Any individual with at least one missing variable is deleted. Listwise deletion can lead to a considerable loss of information and severely biased estimates. In longitudinal studies, one is especially vulnerable. One missing variable at one time point for an individual, excludes all data at all time points for that individual.

To what extent the result of an analysis is influenced by the incomplete data, depends on whether or not there is a pattern in the drop-out. If the individuals with missing variables have special characteristics, this will produce biased estimates. If drop-out is random, listwise deletion produces a random subsample of the original sample and the estimates will be unbiased, although there will generally be loss of information.

3.1 Missing Data Mechanism

It is important to consider the missing data mechanism in all analysis of incomplete data sets. In this paper, we assume an ignorable non-response mechanism, i.e. that data is *missing completely at random (*MCAR) or *missing at random (*MAR). When the conditions hold we may proceed with our method and exclude complicated missing data modeling. If one suspects a non-ignorable response mechanism, all results may be misleading, but one has no way of ascertaining this except through further data collection.

We use the terminology introduced by Rubin (1976) to distinguish among the three types of missing data mechanism. MCAR means that missingness is not related to the variables under study and MAR means the missingness is related to the observed data but not to the missing data. Suppose we have a variable X which is not subject to non-response and a variable Y which is. For a given data set, X is then recorded for all subjects while Y is incomplete. If the probability that Y is missing has no relationship to X or Y, data is MCAR. If the probability that Y is missing depends only on the value of X, data is MAR. A process that is neither MCAR nor MAR is *missing not at random* (MNAR), and here the missingness depends on unobserved and possibly observed data.

There are no consequences concerning bias when making inferences based on data that are MCAR. In this setting, most analysis will be straightforward. The only issue is how to implement an analysis with missing data. Listwise deletion, which discards all units with missing variables, yields valid inferences, although there may be loss of efficiency.

Rubin (1976) is searching for the weakest simple conditions in the process that cause missing data, such that it is always appropriate to ignore this process when making inference about the distributions of data. This is the case when the missing data is MAR and the parameter of the missing data process is distinct from the parameters in the model. When, as in MAR, the probability of non-response depends on the observed response, but not on the unobserved response, it is not necessary to specify a non-response model or to estimate its parameters in order to obtain valid inference.

3.2 Distribution of Missing Values

The vector $\mathbf{y}_i^{(t)}$ for individual i at Time t can be divided into two parts $\mathbf{y}_i^{(t)} = \left(\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis}\right)^{(t)}$, where $\mathbf{y}_i^{obs}$ is the observed part and $\mathbf{y}_i^{mis}$ is the missing part of $\mathbf{y}_i^{(t)}$. As stated earlier, the distribution for each vector $\mathbf{y}_i^{(t)}$, given the cluster membership j, is multivariate normal with parameters $\boldsymbol{\mu}_j^{(t)}$ and $\boldsymbol{\Sigma}_j^{(t)}$, i.e. $\left(\mathbf{y}_i^{(t)} \middle| v_i^{(t)} = j\right) = \left(\left(\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis}\right)^{(t)} \middle| v_i^{(t)} = j\right) \sim N_M\left(\boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)}\right)$. The $\boldsymbol{\mu}_j^{(t)}$ and $\boldsymbol{\Sigma}_j^{(t)}$ parameters may also be divided according to the missingness in the data vector $\mathbf{y}_i^{(t)}$.

$$\boldsymbol{\mu}_j^{(t)} = \left(\boldsymbol{\mu}_j^{obs}, \boldsymbol{\mu}_j^{mis}\right)^{(t)} \qquad\qquad \Sigma_j^{(t)} = \begin{bmatrix} \boldsymbol{\Sigma}_{j,11} & \boldsymbol{\Sigma}_{j,12} \\ \boldsymbol{\Sigma}_{j,21} & \boldsymbol{\Sigma}_{j,22} \end{bmatrix}^{(t)}$$

The elements in mean vector $\boldsymbol{\mu}_j^{(t)}$ and covariance $\boldsymbol{\Sigma}_j^{(t)}$ are rearranged so that parameters corresponding to the observed values in $\mathbf{y}_i^{(t)}$ are followed by those corresponding to the missing values. The covariance matrix is divided into four parts. $\boldsymbol{\Sigma}_{j,11}$ is the (co)variances for observed dimensions and $\boldsymbol{\Sigma}_{j,22}$ for the corresponding missing dimensions. $\boldsymbol{\Sigma}_{j,12}$ and $\boldsymbol{\Sigma}_{j,21}$ are covariances between missing and observed values. Note that $\boldsymbol{\Sigma}_{j,12} = \boldsymbol{\Sigma}_{j,21}^T$. The matrices $\boldsymbol{\Sigma}_{j,11}$ and $\boldsymbol{\Sigma}_{j,22}$ are always symmetric.

If the missing mechanism is ignorable, i.e. MAR or MCAR, we may express the conditional distribution of the missing values $\mathbf{y}_i^{mis}$, given the observed values $\mathbf{y}_i^{obs}$ and the individuals cluster membership j as:

$$
\begin{aligned}
&\left(\mathbf{y}_i^{mis} \left| \mathbf{y}_i^{obs}, v_i^{(t)} = j \right.\right) \sim N_M\left(\boldsymbol{\mu}_j^{mis} + \boldsymbol{\Sigma}_{j,21}\boldsymbol{\Sigma}_{j,11}^{-1}\left(\mathbf{y}_i^{obs} - \boldsymbol{\mu}_j^{obs}\right), \boldsymbol{\Sigma}_{j,3}\right) \\
&where\ \boldsymbol{\Sigma}_{j,3} = \boldsymbol{\Sigma}_{j,22} - \boldsymbol{\Sigma}_{j,21}\boldsymbol{\Sigma}_{j,11}^{-1}\boldsymbol{\Sigma}_{j,12}
\end{aligned} \tag{2}
$$

Formula (2) is the basis for the imputation process in this paper. The cluster membership carries information about the values of an individual. We use this to impute new values in the Gibbs sampler process which is described in Section 4.2. The imputation is not carried out in a traditional sense, in which the missing values are imputed once before the analysis. Instead, the imputation is here a process where new imputed values are generated in each iteration step in the simulations, labeling it as a form of multiple imputation.

4 Estimation Method

4.1 Bayesian Inference

In classical inference, data is considered random while population parameters are taken as fixed. In Bayesian analysis, parameters themselves follow a probability distribution. Knowledge about a parameter, before data is even considered, is summarized in a prior distribution $p(\theta)$. The likelihood of the observed data y given the parameter θ, denoted $p(y\,|\theta)$, is used to modify the prior belief with the knowledge brought by the data, summarized in a posterior density $p(\theta\,|y)$. According to Bayes theorem, we express the relationship as $p(\theta\,|y) \propto p(\theta)p(y\,|\theta)$. For a thorough explanation of Bayesian inference, see for example Congdon (2007) and Bernardo and Smith (2000).

The unknown parameters in our model are $\boldsymbol{\mu}_j^{(t)}$, $\boldsymbol{\Sigma}_j^{(t)}$, $\omega_j^{(t)}$, $\mathbf{Q}_t$ as well as the latent classification vectors $\mathbf{V}^{(t)}$. We begin by specifying the prior distribution of each parameter.

$$
\begin{aligned}
\boldsymbol{\Sigma}_j^{(t)} &\sim W^{-1}\left(m_j^{(t)}, \boldsymbol{\psi}_j^{(t)}\right) \\
\boldsymbol{\mu}_j^{(t)} \left| \boldsymbol{\Sigma}_j^{(t)} \right. &\sim N_M\left(\boldsymbol{\xi}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)}/\tau_j^{(t)}\right) \\
\left(\omega_1^{(1)}, ..., \omega_{J^{(1)}}^{(1)}\right) &\sim Dir\left(\alpha_1, ..., \alpha_{J^{(1)}}\right) \\
\mathbf{Q}_t(j^{(t)}, \cdot) &\sim Dir(\beta_1^{(t)}, ..., \beta_{J^{(t)}}^{(t)})
\end{aligned}
$$

Except for the first two rows, variables are independent of each other and of different values on t and $J^{(t)}$; i.e. $\left(\boldsymbol{\mu}_1^{(1)}, \boldsymbol{\Sigma}_1^{(1)}\right), ..., \left(\boldsymbol{\mu}_J^{(1)}, \boldsymbol{\Sigma}_J^{(1)}\right), \left(\boldsymbol{\mu}_1^{(2)}, \boldsymbol{\Sigma}_1^{(2)}\right), ...,$ $\left(\boldsymbol{\mu}_1^{(T)}, \boldsymbol{\Sigma}_1^{(T)}\right), ..., \left(\boldsymbol{\mu}_J^{(T)}, \boldsymbol{\Sigma}_J^{(T)}\right), \boldsymbol{\Omega}^{(1)}, \mathbf{Q}_1(1,\cdot), ..., \mathbf{Q}_1(j^{(1)},\cdot), \mathbf{Q}_2(1,\cdot), ..., \mathbf{Q}_{T-1}(j^{(T-1)},\cdot)$ are independent random variables.

The prior for $\boldsymbol{\Sigma}_j^{(t)}$ is the inverse Wishart distribution and for $\boldsymbol{\mu}_j^{(t)}$ the multivariate normal distribution. The Dirichlet distribution is the prior distribution for the population weights $\boldsymbol{\Omega}^{(1)}$ as well as for the probabilities for each row in the transition matrices, $\mathbf{Q}_t(j^{(t)},\cdot)$. The selection of the hyperparameters $(m_j^{(t)}, \boldsymbol{\psi}_j^{(t)}, \boldsymbol{\xi}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)}, \tau_j^{(t)},$ $\alpha_1, ..., \alpha_{J^{(1)}}, \beta_1^{(t)}, ..., \beta_{J^{(t)}}^{(t)})$ is chosen to make the priors weakly informative, in default of any prior information. All the priors above are conjugate distributions, which mean that the posterior distributions are from the same family as the priors, even though they are no longer independent. A full description of the conditional posterior distributions, under the assumption of complete data can be found in Appendix A. The derivations can be found in Franzén (2008).

4.2 Gibbs Sampler

Bayesian inference is often linked to sampling-based estimation methods due to complicated or impossible numerical integration. Gibbs sampler (Geman and Geman, 1984) is a powerful and well suited *Markov Chain Monte Carlo* (MCMC) technique for estimating complex Bayesian statistical models. The Gibbs sampler is an iterative procedure, which generates dependent samples from the joint posterior density of all free parameters in the model. If we can express the distribution of each of the parameters conditional on all the others, then by cycling through these conditional statements, the Markov chain will eventually reach the true joint distribution of interest. The choice of starting values influences the first iterated values. To avoid that these values' influencing the estimates, one removes a suitable number of iterations in the beginning, referred to as the burn-in period.

Before the iteration process is started, one must choose some reasonable starting values for all parameters. These are used as conditional values in the first iteration round. The first step in the iteration involves sampling from the model parameters $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, $\boldsymbol{\Omega}$, and $\mathbf{Q}$, all denoted $\boldsymbol{\theta}$. The first step is in reality four steps where sampling is made from the conditional posterior distribution of each parameter, given in Appendix A. The posterior distributions are given conditional on the other parameters, data including imputed values for those that are missing, and the group classification for each individual, given by $\mathbf{V}$. The second step involves imputing values for the missing data and is used for each individual with at least one missing value. This is done by drawing samples from the distribution in Formula (2). We allow missingness to depend on $\mathbf{V}$, i.e. there may be different nonresponses in different groups. In the last step, the classification vectors are updated in accordance with Formula (1). The classification variables $v_i^{(t)}$ $\{t = 1, ..., T\}$ are simulated according to the posterior probabilities the formula gives for all possible development patterns. We summarize the tree iteration steps as,

1. $p\left(\boldsymbol{\theta}\left|\mathbf{y},\mathbf{V}\right.\right)$
2. $p\left(\mathbf{y}^{mis}\left|\mathbf{y}^{obs},\mathbf{V},\boldsymbol{\theta}\right.\right)$
3. $p\left(\mathbf{V}\left|\mathbf{y},\boldsymbol{\theta}\right.\right)$

In each iteration round, new parameters are generated and the conditional distributions are updated for the next iteration round. For a large enough number of iterations, the process approaches the target posterior distribution.

5 Simulated Data Studies

The simulations are performed in Matlab, version 7.4, by a customized program written by the author. The program is available for downloading together with instructions, on www.statistics.su.se/forskning/MBCA.

5.1 Simulation Procedure

The longitudinal method is applied to two simulated data sets where each contains data from three time points. Both data sets consist of 1000 individuals, which at each time point are generated from multivariate normal distributions with different mean vectors but the same identity covariance matrix. The mean values all lie between -3 and 3. At Time 1, data is generated from six normal distributions in four dimensions, at Time 2 from four normal distributions in five dimensions, and at Time 3 from five normal distributions in six dimensions.

We study different non-response rates η to see how the imputation method manages to improve the clustering results. The non-response is created by deleting η percent of the values randomly over variables and individuals at each time point. The result is a data set with missing values that are *missing completely at random (MCAR)*. A comparison is made with the mean imputation method. We also study how much worse the results become when we exclude individuals with missing values. In the comparisons, we use the performance measures *variance* and *estimation error* as well as *classification accuracy*, all explained further on.

The imputation method is tested on simulated data with well-separated groups as well as overlapping groups. A graphical view of the data sets is given in Figure 1. The three graphs in the first column show the well-separated data for each of the three time points. The second column shows the corresponding graphs for the overlapping data. To be able to present the multidimensional data in two dimensional graphs, we plot data through its first two principal components.

Before the algorithm is run, we specify the priors for the parameters. They are chosen to be vague in the sense of not being very specific in the prior belief, in order

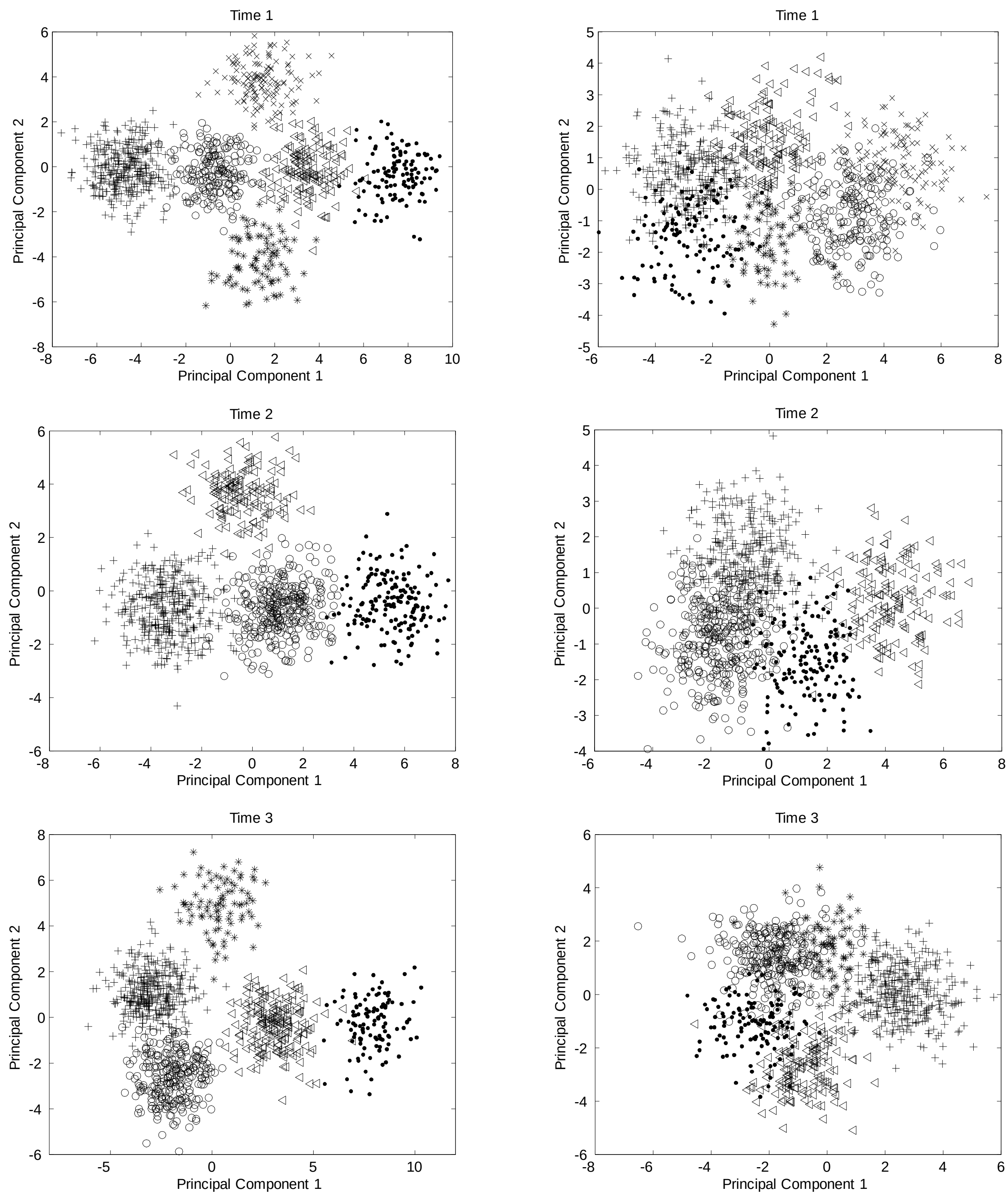

Figure 1: Graphs in column 1 showi data generated from separated groups, and in column 2 from overlapping groups. Data is projected onto the first two principal components at each of the three time points. The principal components stand for 92, 81, and 83 percent of the total variance for the first data material (column 1) and 76, 71, and 55 percent of the total variance for the second data material (column 2).

to let data have the main influence on the results. The estimates are based on 95 000 iterations. The first 5 000 of the total 100 000 iterations where discharged as a burn-in period. This is common procedure since the algorithm usually takes some number of iterations to converge to its right states.

5.1.1 Performance Measures

When comparing the performance of the method for different non-response rates, we use two disparate performance measures. The *Variance* (Var) provides information concerning the spread in the estimate, and the *Estimation Error* (EE) provides information on how far the estimate is from the true value of the constructed data set. The *variance* is a precision measure based on I iterations in the Gibbs sampler. We call it the *simulation variance*. The EE can be seen as a bias measure, however only based on the one sample from one estimation round. Let any of the estimated variables be denoted ξ. The true value of ξ is denoted ξ^T. We express the performance measures as

$$Var_\xi = \frac{\sum_{i=1}^{I} \left(\xi_i - \overline{\xi}\right)^2}{I} \qquad EE_\xi = \left(\overline{\xi} - \xi^T\right)^2 \qquad where\ \overline{\xi} = \frac{\sum_{i=1}^{I} \xi_i}{I} \tag{4}$$

In a Bayesian manner our true value ξ^T follows a certain distribution. If the method works properly, each ξ_i is a draw from the same distribution. This means the expected value of ξ^T, ξ_i, and $\overline{\xi}$, which we denote m_ξ, should be the same, and so should $E\left[\left(\xi^T - m_\xi\right)^2\right]$ and $E\left[\left(\xi_i - m_\xi\right)^2\right]$. We approximate m_ξ with $\overline{\xi}$ which means the expectations $E\left[\left(\xi^T - \overline{\xi}\right)^2\right]$ and $E\left[\left(\xi_i - \overline{\xi}\right)^2\right]$ should be approximately the same for large values of I, or equivalent $Var_\xi = E\left(EE_\xi\right)$. Thus if our estimation method works properly equality would be a confirmation of a functional estimation method. Var_ξ can, in other words, be used to predict the value of EE_ξ. In our simulations, we can in fact give both these values. In real data studies, where we do not know the true values, we are left with only the Var values. If large differences appear between Var_ξ and EE_ξ, it must be due to chance or to the fact that the Markov chain has not yet converged, which can be solved with larger I.

It would be too extensive to account for Var and EE values for all estimated parameters. Instead, we gather values in groups and present mean values of the performance measures for each group. $Mean$ 1 is the mean of EE $\left(\text{or } Var^{(1)}_{\mu kj}\right)$, calculated for all estimated cluster mean values at the first time point. With 6 clusters and 4 variables for each cluster at Time 1, this will be a mean over 24 values. We let $EE^{(1)}_{\mu kj}$ $\left(\text{or } Var^{(1)}_{\mu kj}\right)$ denote the performance measure for the k:th estimated μ variable in Cluster j at Time 1. The EE or Var values are calculated according to (4) for each variable. We make the expression $Mean$ $1 = \left(\sum_{j=1}^{6}\sum_{k=1}^{4} EE_{\mu kj}^{(1)}\right)/24$ which gives us the mean of the performance measure for all variables and clusters at Time 1. $Mean$ 2 and $Mean$ 3 are calculated in a similar way to receive the mean values at Times 2 and 3. $Omega$ is the mean of the performance measure for all cluster probabilities. $Omega$ 1 is the mean over the 6 values at Time 1, i.e. $Omega\ 1 = \left(\sum_{j=1}^{6} EE_{\omega j}^{(1)}\right)/6$, where $EE_{\omega j}^{(1)}$ denotes the performance measure for Cluster j at Time 1. $Omega$ 2 is in the same way the mean over the 4 values at Time 2, and $Omega$ 3 over the 5 values at Time 3. The transition matrices Q_1 and Q_2 are composed of 24 and 20 values respectively. We express the mean of the performance measure over all transition probabilities within Q_1 as $Trans\ 1 = \left(\sum_{j=1}^{4}\sum_{i=1}^{6} EE_{Q_1 ij}\right)/24$, where $EE_{Q_1 ij}$ is the performance measure for the transition probability in Q_1, going from Cluster i at Time 1 to Cluster j at Time 2. $Trans$ 2 is calculated in a similar manner.

5.2 Simulation Results

5.2.1 Estimation Precision

In the columns in Table 1, performance measures for mean, cluster probabilities and transition probabilities are presented for each of the non-response rates η. The performance measures for variances and covariances are given in Tables 10 and 11 in Appendix B.

All 1000 individuals are included in the calculations. Within each table the same data set is used for all non-response rates. The missing values, on the other hand vary for the different levels of non-response. The missing values for one percent level are deleted, unconditional on other levels of non-response.

As expected, the method generates smaller performance measures in the separated groups than in the overlapping groups. In general the EE and Var are higher for the overlapping groups in comparison with the separated groups for the same non-response rate. In the same way the values increase in general within each table when the non-response rate gets higher. There are exceptions from the generalizations. Variations between two estimation runs, may in addition to the different η and group structure (overlapping and separated), depend on the different data sets we get when randomly eliminating missing values. These random fluctuations together with smaller fluctuations in the Gibbs sampler estimation method may cause estimates to deviate from the expected pattern.

Even though the performance measures for single parameters do not show in the tables, the summarized mean values give a good compressed answer on the performance of the method. Our estimation method works well for all sample sizes, even though the true errors (EE), on average, are slightly larger than the predicted (Var). Equal magnitude for Var and EE confirms that the method works properly, as discussed in subsection 5.1. However, from a practical point of view, one should not have more than 40 percent missing values. One reason is a slower convergence rate, which might demand intolerably long iteration chains. The other reason is that the variance increases rapidly for non-response rates higher than 40-45 percent.

η (%)		*Mean* 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
0	EE	0.00773	0.00728	0.01265	0.00015	0.00017	0.00005	0.00172	0.00085
	Var	0.00905	0.00551	0.00816	0.00014	0.00018	0.00016	0.00114	0.00068
5	EE	0.00985	0.00836	0.01376	0.00015	0.00028	0.00006	0.00195	0.00088
	Var	0.01008	0.00592	0.00871	0.00015	0.00019	0.00016	0.00119	0.00069
10	EE	0.01431	0.00755	0.01682	0.00018	0.00019	0.00005	0.00185	0.00086
	Var	0.01120	0.00627	0.01090	0.00016	0.00018	0.00018	0.00122	0.00074
25	EE	0.02530	0.01728	0.02235	0.00026	0.00056	0.00009	0.00281	0.00100
	Var	0.01632	0.00905	0.01421	0.00020	0.00022	0.00022	0.00151	0.00085
40	EE	0.00844	0.02548	0.04545	0.00022	0.00028	0.00030	0.00429	0.00122
	Var	0.02718	0.01311	0.02152	0.00025	0.00042	0.00032	0.00201	0.00108
45	EE	0.04339	0.03432	0.05125	0.00051	0.00115	0.00024	0.00497	0.00140
	Var	0.03098	0.01482	0.02748	0.00028	0.00030	0.00035	0.00216	0.00124
50	EE	0.39139	0.03505	0.10476	0.00074	0.00135	0.00070	0.00438	0.00136
	Var	0.45287	0.01829	0.04653	0.00066	0.00034	0.00055	0.00528	0.00159
55	EE	0.62092	0.05466	0.08890	0.00032	0.00167	0.00079	0.00809	0.00151
	Var	0.87568	0.02374	0.04244	0.00081	0.00041	0.00047	0.00635	0.00170
60	EE	1.20574	0.06336	0.47683	0.00032	0.00254	0.00177	0.02832	0.00565
	Var	2.07609	0.02895	0.60500	0.00331	0.00043	0.00286	0.01682	0.00549

η (%)		*Mean* 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
0	EE	0.00869	0.01062	0.00791	0.00010	0.00053	0.00037	0.00268	0.00084
	Var	0.01424	0.00967	0.00770	0.00025	0.00033	0.00016	0.00166	0.00079
5	EE	0.01015	0.01233	0.00831	0.00006	0.00067	0.00033	0.00341	0.00073
	Var	0.01700	0.01158	0.00840	0.00030	0.00039	0.00017	0.00187	0.00089
10	EE	0.01856	0.00803	0.00946	0.00036	0.00048	0.00030	0.00211	0.00095
	Var	0.01959	0.01354	0.00920	0.00012	0.00041	0.00018	0.00200	0.00087
25	EE	0.02908	0.07202	0.03230	0.00030	0.00134	0.00051	0.00369	0.00192
	Var	0.03527	0.04180	0.01682	0.00055	0.00075	0.00027	0.00314	0.00149
40	EE	0.02454	0.02526	0.02718	0.00028	0.00075	0.00075	0.00499	0.00141
	Var	0.03842	0.02763	0.02005	0.00058	0.00064	0.00030	0.00334	0.00148
45	EE	0.63000	0.04647	0.04367	0.00109	0.00102	0.00040	0.00885	0.00137
	Var	1.36245	0.02976	0.04396	0.00258	0.00060	0.00058	0.00922	0.00199
50	EE	0.27673	0.35252	0.02109	0.00081	0.00398	0.00045	0.02779	0.00380
	Var	0.58651	0.28753	0.03077	0.00125	0.00278	0.00041	0.01947	0.00453
55	EE	0.89819	0.52773	0.17191	0.00059	0.00662	0.00255	0.03320	0.00581
	Var	1.22986	0.60365	0.05474	0.00131	0.00275	0.00080	0.01777	0.00690

TABLE 1: Performance measures. Top table - separated groups, bottom table - overlapping groups

5.2.2 Classification Accuracy

It may be of interest to study the method's performance in the sense of classification accuracies for individuals. Table 2 presents the percent of correctly classified individuals as a function of the number of missing variables and the non-response rate. For each non-response rate and time point, we separate the individuals according to how many missing values they got. The percentage of correctly classified individuals are then calculated for each category.

The Bayesian clustering method generates cluster probabilities for each individual belonging to each cluster. The allocations in the tables below are performed by assigning an individual to the cluster for which that individual has the highest cluster probability estimate. The column furthest to the right gives the overall classification accuracies for all 1000 individuals, independently of the number of missing variables. Data has four variables at Time 1, five at Time 2, and 6 at Time 3. The number of possible missing values is therefore different for the three time points.

The classification accuracies show a promising result for the method. There are high percentages of correctly classified individuals even for high rates of non-response. The overall classification accuracies are above 75 percent for non-response rates as high as 50 percent for the separated data and 40 percent for the overlapping data.

When the number of missing variables for an individual increase, the percentage of correctly classified individuals decreases. Still, for the separated groups, around or above 90 percent of individuals with at most 2 missing values are correctly classified with some exception for Time 1, where the number of variables is only 4. The majority of individuals are correctly classified even with only one observed variable. For the overlapping groups, the result is not as good. Still, there are around or above 70 percent of the individuals with up to 2 missing values which are correctly classified, with a couple of exceptions.

5.2.3 Imputation Contra No Imputation

Given a data set with a certain percent of random non-response, how much better is our method compared to other methods in handling non-response? We will compare our method with two common methods. In this section we remove all individuals without complete variable sets. In the next subsection we use the mean imputation method.

With a fairly low non-response rate of 5 percent a comparison is made between imputing missing values contra running the method with a data set of only "complete" individuals. The remaining data set, after randomly deleting individuals with at least one missing value, consists of 464 individuals for the separated data set and 458 for the overlapping data set.

η (%)		0 missing	1 missing	2 missing	3 missing	4 missing	5 missing	6 missing	Overall
0	Time 1	0.9780							0.9780
	Time 2	0.9840							0.9840
	Time 3	0.9760							0.9760
5	Time 1	0.9791	0.9598	0.9091*	-	-			0.9750
	Time 2	0.9859	0.9444	0.9167*	-	-	-		0.9760
	Time 3	0.9742	0.9784	0.9286*	1.0000**	-	-	-	0.9740
10	Time 1	0.9763	0.9377	0.8776*	0.5000**	-			0.9600
	Time 2	0.9802	0.9653	0.9054	0.7500**	-	-		0.9690
	Time 3	0.9759	0.9590	0.9875	0.6923*	1.0000**	-	-	0.9670
25	Time 1	0.9898	0.9463	0.8682	0.6296	0.5000**			0.9230
	Time 2	0.9812	0.9495	0.9228	0.8830	1.0000*	0.0000**		0.9420
	Time 3	0.9874	0.9468	0.9485	0.8889	0.7105*	0.5000**	-	0.9350
40	Time 1	0.9701	0.9511	0.8323	0.6496	0.5556*			0.8620
	Time 2	0.9615	0.9635	0.9086	0.8571	0.7778	0.5000*		0.9040
	Time 3	0.9302*	0.9175	0.9451	0.9163	0.8261	0.7241*	0.2000**	0.9050
45	Time 1	0.9468	0.9406	0.8376	0.6456	0.4348*			0.8210
	Time 2	0.9583*	0.9528	0.9006	0.8367	0.7281	0.4000*		0.8710
	Time 3	0.9730*	0.9308	0.9389	0.8622	0.7919	0.5091	0.1429*	0.8570
50	Time 1	0.8679	0.8840	0.7973	0.6202	0.4063			0.7520
	Time 2	0.9706*	0.9712	0.8956	0.8397	0.7310	0.3929*		0.8490
	Time 3	1.0000*	0.9079	0.9221	0.8550	0.7897	0.6122	0.4211*	0.8310
55	Time 1	0.7813*	0.8367	0.7109	0.5923	0.4257			0.6750
	Time 2	1.0000*	0.9250	0.8862	0.8843	0.7040	0.4528		0.8290
	Time 3	1.0000*	0.9825	0.9312	0.8814	0.8041	0.5956	0.2692*	0.8200
60	Time 1	0.8095*	0.7545	0.6735	0.5694	0.4206			0.6220
	Time 2	0.9091*	0.8955	0.9295	0.8212	0.7371	0.4722		0.8070
	Time 3	1.0000**	0.9130*	0.8881	0.8084	0.7027	0.6141	0.3500*	0.7400

η (%)		0 missing	1 missing	2 missing	3 missing	4 missing	5 missing	6 missing	Overall
0	Time 1	0.9280							0.9280
	Time 2	0.9090							0.9090
	Time 3	0.9620							0.9620
5	Time 1	0.9163	0.9045	0.6000*	-	-			0.9110
	Time 2	0.8976	0.8889	0.8333*	-	-	-		0.8940
	Time 3	0.9598	0.9256	0.8824*	1.0000**	-	-	-	0.9500
10	Time 1	0.9200	0.8524	0.7755*	0.6000**	-			0.8930
	Time 2	0.9089	0.8844	0.8529	0.7000*	1.0000**	1.0000**		0.8960
	Time 3	0.9560	0.9384	0.8692	0.8571*	-	-	-	0.9400
25	Time 1	0.8967	0.8458	0.7248	0.5455*	0.1000*			0.8140
	Time 2	0.8734	0.8395	0.7921	0.7340	0.5294*	0.0000**		0.8180
	Time 3	0.9667	0.9331	0.8553	0.8085	0.7353*	0.4000**	-	0.8875
40	Time 1	0.9478	0.8085	0.7287	0.5541	0.3429*			0.7470
	Time 2	0.8919	0.8745	0.7994	0.7602	0.6471	0.6000**		0.8040
	Time 3	0.9787*	0.9391	0.8815	0.8465	0.6985	0.7773*	0.5000**	0.8578
45	Time 1	0.9011	0.8294	0.6741	0.5529	0.2791*			0.6990
	Time 2	0.9512*	0.8447	0.8075	0.7348	0.7593	0.5556*		0.7910
	Time 3	0.9615*	0.9104	0.8780	0.8633	0.6901	0.5441	0.3333*	0.8220
50	Time 1	0.8442	0.7848	0.6992	0.5759	0.3800			0.6710
	Time 2	0.8065*	0.7384	0.7735	0.7090	0.6096	0.5366*		0.7140
	Time 3	1.0000*	0.9388	0.8987	0.8278	0.7167	0.6292	0.5909*	0.8080
55	Time 1	0.8000*	0.6915	0.6467	0.5164	0.3300			0.5900
	Time 2	0.6000*	0.5000	0.5078	0.5198	0.4147	0.3385		0.4820
	Time 3	0.9000*	0.9231	0.8247	0.7491	0.6968	0.5600	0.3333*	0.7250

TABLE 2: Percentage of individuals that are classified into the right cluster as a function of the number of variables missing and the total non-response rate. Values with one star are based on 1 to 5 individuals and values with two stars on 6 to 50 individuals. A dash indicates no individuals in that specific category. Top graph: separated groups, bottom graph: overlapping groups.

In Table 3, we once again give the performance measures for $\eta = 5$, together with the new, corresponding values when imputation is not used. This would be the same as having a non-response rate of 0 but a data set of about half the size of the original. Both estimates are based on the same data. The performance measures for the (co)variances are found in Table 12 in Appendix B.

The consequences of deleting the missing values are higher (worse) values of the performance measures. The largest differences appear in the *Mean* categories where a few of the measures are as much as about 20 times higher without imputation. An apparently low non-response rate, results in a large reduction of the data set and large increases in the performance measures.

Separated Clusters

Without Imputation

η (%)		*Mean* 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
5	EE	0.11885	0.04531	0.05886	0.00027	0.00119	0.00068	0.00615	0.00233
	Var	0.23211	0.01780	0.02266	0.00040	0.00041	0.00037	0.00284	0.00144
		With Imputation							
		Mean 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
5	EE	0.00985	0.00836	0.01376	0.00015	0.00028	0.00006	0.00195	0.00088
	Var	0.01008	0.00592	0.00871	0.00015	0.00019	0.00016	0.00119	0.00069

Overlapping Clusters

Without Imputation

		Mean 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
5	EE	0.07826	0.22881	0.02063	0.00008	0.00090	0.00054	0.00343	0.00122
	Var	0.31716	0.24414	0.02009	0.00068	0.00060	0.00034	0.00427	0.00292
		With Imputation							
		Mean 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
5	EE	0.01015	0.01233	0.00831	0.00006	0.00067	0.00033	0.00341	0.00073
	Var	0.01700	0.01158	0.00840	0.00030	0.00039	0.00017	0.00187	0.00089

TABLE 3: Performance measures. Comparison study between imputing missing variables contra discharging individuals with one or more missing variables.

5.2.4 Comparison with Mean Imputation

The mean imputation method is a commonly used method with a straightforward application: see for example Little and Rubin (2002). The missing values are simply replaced by an overall mean, based on the non-missing values. For multivariate data, a missing value for variable k is replaced by

$$\overline{y}_k = \frac{\sum_{i=1}^{N_k} y_i^{(k)}}{N_k},$$

where N_k is the number of non-missing values for variable k, and the i:th non-missing value for variable k is denoted $y_i^{(k)}$.

The mean imputation method is applied to the data for non-response rates up to 40 percent. Our clustering algorithm is then applied as if there were no missing values. The results are shown in Table 4 (and in Table 13 in Appendix B for (co)variances). Compared to the corresponding rates in Table 1, the estimates are not as good, especially not for the estimation error (EE). This is however not much of a surprise. In the mean imputation process, data is deformed towards an overall mean and away from cluster-specific values. This causes the estimation error EE to be large. The variance (Var) does not increase as much, but is higher for the mean imputations than our imputation method. When imputing mean values, the overall variation in the data set decreases compared to a full data set. This, in turn, makes it harder to identify clusters since they become more similar to each other. In the iteration process, this causes more jumps for individuals between clusters and therefore a larger variance.

η (%)		*Mean* 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
5	EE	0.02201	0.02981	0.04058	0.00010	0.00037	0.00010	0.00237	0.00085
	Var	0.01220	0.00659	0.00981	0.00017	0.00020	0.00018	0.00126	0.00072
10	EE	0.06236	0.00754	0.01323	0.00025	0.00032	0.00014	0.00279	0.00074
	Var	0.01550	0.06406	0.10134	0.00019	0.00021	0.00021	0.00138	0.00079
25	EE	0.85948	1.38570	0.36259	0.00257	0.00832	0.00078	0.01826	0.00248
	Var	0.05636	0.13839	0.01742	0.00031	0.00103	0.00032	0.00288	0.00142
40	EE	2.03553	2.33434	2.05706	0.01440	0.01395	0.01182	0.04234	0.01684
	Var	0.39551	0.01390	0.01913	0.00029	0.00038	0.00029	0.00242	0.00116

η (%)		*Mean* 1	*Mean* 2	*Mean* 3	*Omega* 1	*Omega* 2	*Omega* 3	*Trans* 1	*Trans* 2
5	EE	0.01935	0.02596	0.01648	0.00011	0.00070	0.00034	0.00399	0.00083
	Var	0.02168	0.01637	0.00920	0.00035	0.00048	0.00019	0.00214	0.00099
10	EE	0.18590	0.04599	0.04347	0.00033	0.00047	0.00057	0.00493	0.00118
	Var	0.24820	0.03818	0.01397	0.00075	0.00088	0.00028	0.00441	0.00138
25	EE	0.75682	0.56054	0.58311	0.00455	0.00611	0.00492	0.03001	0.00880
	Var	0.06344	0.03200	0.02306	0.00039	0.00075	0.00045	0.00539	0.00194
40	EE	1.21787	1.07643	1.25841	0.01219	0.00715	0.02056	0.03659	0.02428
	Var	0.17061	0.01569	0.05647	0.00030	0.00048	0.00072	0.00308	0.00180

TABLE 4: Performance measures when using Mean Imputation. Top table - separated groups, bottom table - overlapping groups

Mean imputation gives fairly good results up to a non-response rate of 10 percent, even though the values in Table 1 are better. For higher non-response rates than 10 percent, the mean imputation method does not manage to estimate the cluster parameters and find the origin of individuals. At these higher levels, the mean imputation does not seem to work. It works rather the opposite way by gradually eliminating cluster specific values, making clustering more difficult. The different magnitude of the EE and Var values is also an indication of a badly functioning estimation process.

Even though mean imputation is not efficient for high non-response rates, for lower rates, it seems better to use it than to exclude individuals with missing values. The mean imputation method shows much better result than the approach of deleting missing values, even when the non-response rate is only 5 percent. However, compared to the values in Table 1, the mean imputation method is outperformed by the imputation method of this paper. This is further confirmed by the classification accuracies for mean imputation given in Table 5, which can be compared to the corresponding values in Table 1. For the mean imputation, the classification accuracies drastically drop for non-response rates higher than 10 percent.

η (%)		0 missing	1 missing	2 missing	3 missing	4 missing	5 missing	6 missing	Overall
5	Time 1	0.9718	0.9195	0.6364*	-	-			0.9590
	Time 2	0.9807	0.9596	0.7917*	-	-	-		0.9720
	Time 3	0.9661	0.9397	0.8923*	1.0000**	-	-	-	0.9580
10	Time 1	0.9645	0.9011	0.7143*	0.0000**	-			0.9330
	Time 2	0.9686	0.9590	0.9054	0.7500**	-	-		0.9600
	Time 3	0.9596	0.9508	0.9500	0.7692*	1.0000**	-	-	0.9530
25	Time 1	0.8061	0.5864	0.4045	0.2593	0.0000**			0.5910
	Time 2	0.6197	0.6414	0.6421	0.6383	0.4545*	0.0000**		0.6340
	Time 3	0.9371	0.9272	0.9038	0.7582	0.6053*	0.5000**	-	0.8830
40	Time 1	0.1343	0.1606	0.2012	0.1959	0.2000*			0.1770
	Time 2	0.4189	0.3321	0.3401	0.3620	0.3529	0.4000**		0.3500
	Time 3	0.3617*	0.2944	0.3161	0.2835	0.3088	0.2121*	0.2500**	0.3010

η (%)		0 missing	1 missing	2 missing	3 missing	4 missing	5 missing	6 missing	Overall
5	Time 1	0.9126	0.7809	0.5000*	-	-			0.8850
	Time 2	0.8924	0.8677	0.8667*	-	-	-		0.8870
	Time 3	0.9612	0.9442	0.8824*	1.0000**	-	-	-	0.9550
10	Time 1	0.9052	0.7491	0.6122*	0.8000**	-			0.8480
	Time 2	0.9058	0.8571	0.8080	0.7000*	1.0000**	1.0000**		0.8830
	Time 3	0.9523	0.9062	0.8037	0.8571*	-	-	-	0.9200
25	Time 1	0.5633	0.3949	0.3119	0.2273*	0.0000*			0.4160
	Time 2	0.4323	0.4711	0.5161	0.4787	0.4706	0.0000*		0.4750
	Time 3	0.6200	0.6602	0.6399	0.4823	0.3235*	0.4000**	-	0.6100
40	Time 1	0.5299	0.3042	0.1982	0.2027	0.2857*			0.2840
	Time 2	0.2568	0.2768	0.3430	0.3077	0.3412	0.2000**		0.3100
	Time 3	0.4894*	0.4822	0.3040	0.2323	0.1985	0.0303*	0.0000**	0.3050

TABLE 5: Percentage of individuals that are classified into the right cluster as a function of the number of variables that are missing for each individual and the total non-response rate. Top table - separated groups, bottom table - overlapping groups

6 Real Data Study

We look at a data set consisting of 1206 school children with 6 variables. The variables are their attitudes to three school subjects, Religion, Mathematics, and their mother tongue Swedish and their marks in the same three subjects. We use data collected at two time points, the first when the children were in third grade in 1965 and the second when they had reached sixth grade in 1968. The data set is part of a much larger data base from the longitudinal research project "Individual Development and Adaption" (IDA) at the Department of Psychology at Stockholm University: see Bergman and Magnusson (1997) and Magnusson(1988). The IDA data base covers a whole range of variables related to behavior, social relations, family climate, psychological, mental, and socioeconomic factors. The purpose of the project is to understand and explain individual development processes.

The variables are measured on a discrete scale with values from 1 to 5. The value 1 represents the attitude "dislike it" and 5 "like it very much". In the same manner 1 is the lowest grade and 5 the highest. Despite discrete values, we use our method developed for normally distributed data.

The 1206 individuals have different degrees of missingness. All of them have at least one measurement on at least one time point. Table 6 gives a presentation on how many individuals have a certain number of missing variables. The majority of the individuals have zero missing values at Times 1 and 2, and also when taking both times into account. There are, however, 28 percent of the individuals at Time 1 and 18 percent at Time 2 who have at least 1 missing value. There are also quite a few individuals that are short of all variables at either one of the two time points. When mark variables are missing for an individual they are so, almost exclusively, for all three mark variables at a certain time point. Among the attitude variables the same conditions do not apply. Several individuals have one or two missing attitude variables, in addition to those with all attitude variables missing. The total non-response rate, counting all variables at both time points, is 32 percent.

Number of missing variables	0	1	2	3	4	5	6	7	8	9	10	11	Total
Time 1	870	50	10	96	2	0	178						1206
Time 2	992	23	8	78	0	1	104						1206
Time 1 and 2	720	56	18	121	5	0	233	8	2	40	2	1	1206

TABLE 6: Number of individuals represented by how many missing variables they have. One individual may have 0 to 6 missing variables at Time 1 and the same at Time 2. For the two time points together, an individual may have 0 to 11 missing values. If all 12 variables were missing, that individual was removed from the analysis (2 individuals).

It is not possible to determine, from the data alone, if the missing data mechanism is ignorable, i.e. if data fulfill the MAR conditions. We can not check for possible dependencies for missing values, simply because we do not have the missing values. However, here we make the assumption that the missing values fulfill the needed conditions.

As in the simulated examples, the prior distributions are specified, so data has the major influence on the estimates, not the prior distributions. Estimates are based on 95 000 iterations (100 000 minus a burn-in period of 5 000 iterations). The number of clusters is decided after running the algorithm for two clusters and then successively adding one cluster at a time. This is done for the two time

		Time 1									
		Cluster 1		*Cluster 2*		*Cluster 3*		*Cluster 4*		*Cluster 5*	
Analysis		1	2	1	2	1	2	1	2	1	2
Attitude Swedish		2.29	2.57	2.77	2.77	2.22	3.27	2.74	2.57	2.15	1.79
	Var	0.0123	0.0106	0.0081	0.0074	0.0089	0.0044	0.0521	0.0152	0.0309	0.0172
Attitude Math		2.51	2.38	3.99	3.74	2.93	3.05	3.39	3.71	1.85	2.01
	Var	0.0105	0.0134	0.0004	0.0023	0.0093	0.0045	0.0366	0.0046	0.0371	0.0222
Attitude Religion		2.51	2.77	2.76	2.76	2.69	2.59	3.63	3.54	2.10	2.22
	Var	0.0125	0.0103	0.0100	0.0071	0.0107	0.0056	0.0071	0.0069	0.0662	0.0229
Mark Swedish		3.89	4.50	3.79	4.00	2.95	3.00	2.44	1.97	2.23	1.76
	Var	0.0057	0.0034	0.0046	0.0002	0.0030	0.0001	0.0093	0.0019	0.0147	0.0053
Mark Math		4.17	3.97	4.10	3.85	3.00	3.08	2.07	2.56	1.86	2.15
	Var	0.0024	0.0050	0.0042	0.0029	0.0001	0.0015	0.0179	0.0059	0.0068	0.0061
Mark Religion		3.71	3.86	3.53	3.58	3.01	2.99	2.60	2.47	2.45	2.47
	Var	0.0046	0.0046	0.0032	0.0021	0.0029	0.0015	0.0076	0.0041	0.0092	0.0044
Probabilities (%)		18.3	14.9	23.8	21.4	34.4	35.8	12.6	14.3	10.9	13.6

		Time 2									
		Cluster 1		*Cluster 2*		*Cluster 3*		*Cluster 4*		*Cluster 5*	
Analysis		1	2	1	2	1	2	1	2	1	2
Attitude Swedish		2.19	2.23	2.19	2.35	2.14	2.12	2.10	1.87	1.96	1.61
	Var	0.0117	0.0166	0.0054	0.0035	0.0047	0.0029	0.0098	0.0055	0.0454	0.0428
Attitude Math		3.06	2.44	3.06	2.87	2.74	2.73	2.25	2.63	1.64	2.25
	Var	0.0150	0.0242	0.0056	0.0047	0.0053	0.0034	0.0143	0.0098	0.0572	0.0651
Attitude Religion		2.02	2.96	1.97	1.91	1.77	1.79	1.57	1.69	1.74	1.42
	Var	0.0153	0.0173	0.0065	0.0046	0.0052	0.0035	0.0126	0.0078	0.0530	0.0487
Mark Swedish		4.12	4.81	3.73	4.00	3.04	3.00	2.39	2.00	2.09	1.36
	Var	0.0047	0.0054	0.0031	0.0001	0.0021	0.0001	0.0042	0.0002	0.0176	0.0175
Mark Math		4.95	4.37	3.99	3.91	3.00	3.03	2.01	2.30	1.58	1.85
	Var	0.0020	0.0079	0.0002	0.0023	0.0001	0.0014	0.0004	0.0029	0.0322	0.0168
Mark Religion		4.15	4.36	3.70	3.79	2.98	2.96	2.31	2.27	2.08	1.93
	Var	0.0075	0.0078	0.0029	0.0018	0.0024	0.0011	0.0045	0.0026	0.0189	0.0222
Probabilities (percent)		13.8	9.0	25.6	26.4	33.8	37.5	18.7	20.6	8.1	6.5

TABLE 7: Posterior estimates of the mean values for each cluster at the two time points. Proportions between clusters are also given. To the left are the estimates based on the 711 individuals with no missing values and to the right are the estimates based on all 1206 values. Below each estimate is the simulation variance, i.e. the variance in the 95 000 iterations.

point separately. Up until a number of five groups, additional cluster structures appeared for the new cluster at both time points. More than five groups resulted in one or more clusters with almost identical characteristics.

First we run the method for only those individuals with complete data, a total of 720 individuals (Analysis 1). This analysis can be studied in detail in Franzén (2008). The results are compared to the results generated when all 1206 individuals are included, and missing values are imputed within the method (Analysis 2). The estimates of the cluster means and cluster probabilities are given in Table 7, and the transition probabilities in Table 8. The (co)variance estimates are presented in Tables 15-18 in Appendix C. For the estimates in Table 7, the *simulation variance* is presented under its corresponding mean estimate. The *simulation variance* is the variance in the 95 000 iterations. We have arranged the clusters in the order going from better to worse marks.

The mean estimates and their difference in Analysis 1 compared to Analysis 2, can be seen visually in Figure 2. There are no remarkable differences in the cluster patterns. Noticeable is a smaller spread between clusters for the variables "Attitude Math" and "Mark Math" for the two graphs to the right, i.e. when imputation is carried out. The variables "Attitude Swedish" and "Mark Swedish" show opposite results.

The *simulation variance* is lower for a significant part of the estimates, using imputation (Analysis 2). The underlying values of the variables are in the same range for this real data set (1 to 5), as for the simulated studies (-3 to 3). We may therefore make a comparison of the magnitude of the performance measures. In Table 7, the Var values are the variance for one parameter. We calculate the means over all values for a direct comparison to the $Mean$ values. For Analysis 1 $Mean\ 1 = 0.0136$ for Time 1 and $Mean\ 2 = 0.0123$ for Time 2. The corresponding values for Analysis 2 are $Mean\ 1 = 0.0068$ and $Mean\ 2 = 0.0116$. These values are all lower than corresponding values for similar circumstances ($\eta = 30$, overlapping groups) in Table 1. Even without imputation, the method seems to generate good estimates. The above comparison indicates not only that the method works, but also that the variance and estimation errors are relatively low.

Estimates of transition probabilities between Times 1 and 2 are given in Table 8. An expected pattern would be high transition probabilities between clusters of a similar kind. The higher probabilities between the lines in the table confirm the anticipation. Individuals have a tendency to move to clusters of similar characteristics as the cluster they move from. If the clusters are similar at both time points and arranged in the same order, one would expect the highest values in the diagonal of the matrix. In our case, the cluster structures are quite different at the two times. This results in a deviation of the assumption for the first and last line. Cluster 5 has more similar mean estimates to Cluster 4 than to Cluster 5 at Time 2. This explains the higher transition probability from Cluster 5 to Cluster 4 rather than to Cluster 5 at Time 2. The same goes for transition from Cluster 1 at Time 1, where individuals have a higher probability of ending up in Cluster 2 rather than Cluster 1 at Time 2. Analysis 2 shows a more stable estimate of the transition matrix than does Analysis 1. This means the transition probabilities are higher for values in the diagonal and values nearby.

Conclusions regarding the analyses would be that the mean estimates do not differ much when the whole data set is used as compared to data where individuals with missing variables are deleted. The estimates of the cluster- and transition probabilities do differ however. Cluster membership is a little more stable based on the whole data set. In addition, the precision of the estimates are in general better. This suggests that there are advantages using the whole data set in combination with imputation instead of only using complete data.

		Time 2				
		1	*2*	*3*	*4*	*5*
	1	0.25	0.45	0.22	0.04	0.04
	2	0.30	0.43	0.20	0.04	0.03
Time 1	*3*	0.03	0.17	0.54	0.23	0.04
	4	0.05	0.06	0.35	0.39	0.15
	5	0.05	0.06	0.17	0.40	0.31

		Time 2				
		1	*2*	*3*	*4*	*5*
	1	0.37	0.41	0.15	0.03	0.03
	2	0.10	0.57	0.26	0.04	0.02
Time 1	*3*	0.01	0.19	0.63	0.14	0.03
	4	0.03	0.05	0.32	0.48	0.11
	5	0.03	0.05	0.19	0.52	0.21

TABLE 8: Posterior estimate of transition matrices between Time 1 and 2. To the left is the transition matrix estimated without imputation and to the right is the matrix estimated with imputation. Between the demarcations are the three highest probabilities for each row. Given a cluster membership at Time 1, transitions are more probable to clusters of similar charactes at Time 2.

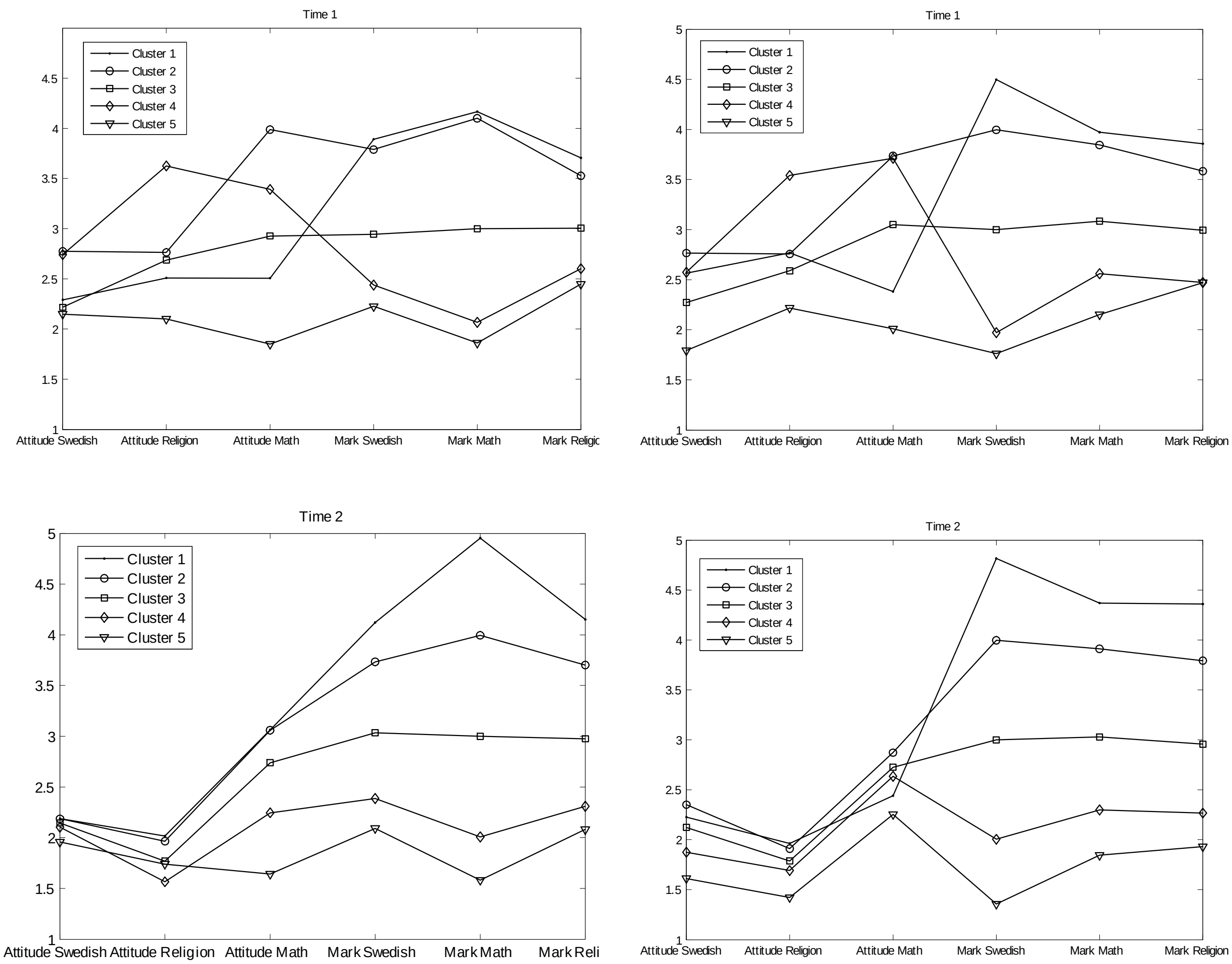

FIGURE 2: Comparison of the mean estimates when individuals with missing values are discharged from the analysis (graphs in left column) and when the whole dataset is included (graphs in right column).

Another comparison between Analysis 1 and 2 is presented in Table 9. We classify each observation to the cluster of which it has the highest cluster probability estimate. Given the cluster classification of all 720 individuals in Analysis 1, the table gives information on how they are classified in Analysis 2, for each time point. The last row shows how the 495 individuals, who were not included in Analysis 1 due to missing values, where classified when they are included in Analysis 2. One may compare the last lines in the two sub-tables below with the cluster probabilities for Analysis 1 in Table 7. It then becomes apparent that the individuals excluded in Analysis 1, have a somewhat different cluster membership when they are included in the Analysis.

		Time 1					
		Analysis 2					
		1	*2*	*3*	*4*	*5*	
	1	45	27	25	3	0	100
	2	16	55	26	3	0	100
Analysis 1	*3*	8	10	57	15	11	100
	4	0	6	36	51	7	100
	5	4	1	32	1	61	100
	Excluded	10	23	37	14	16	100

		Time 2					
		Analysis 2					
		1	*2*	*3*	*4*	*5*	
	1	28	60	11	0	0	100
	2	13	53	29	5	0	100
Analysis 1	*3*	2	19	59	19	0	100
	4	0	5	34	54	7	100
	5	0	0	19	53	28	100
	Excluded	4	27	42	22	4	100

TABLE 9: Illustration of the difference in classification when comparing Analyses 1 and 2, i.e. inclusive or exclusive of individuals with missing values. Given the cluster classification for Analysis 1, each row gives the percentage of the same individuals' being classified into the 5 different clusters for Analysis 2. The last line of each Table gives the classification of individuals excluded in Analysis 1, but included when imputing values.

7 Concluding Remarks

Non-response is a frequent problem in longitudinal studies of multivariate data. Multiple imputation is carried out as an integrated step in a longitudinal, model-based clustering method. At each data collection point, data is assumed to be generated from a mixture model of multivariate, normal distributions. Each distribution represents a cluster with its specific characteristics. Model parameters which include mean vectors, (co)variances, cluster probabilities and transition probabilities between clusters at two consecutive time points, are estimated using Bayesian inference.

The method is tested on real and simulated data with various rates of non-response. Studies with simulated data show a well functioning imputation method which handles non-response rates up to 40-45 percent without serious loss of precision in estimates. It outperforms the common solution which deletes observations with one or more missing values, and it also outperforms the results of the mean imputation method. For the real data study, comparisons are made between our integrated imputation/estimation method and the analysis using data with only a complete variable set. No major differences in the cluster means occurred, but when using the whole data set, the variances of the estimates are lower and the cluster membership is more stable.

Although this paper is presented with a longitudinal approach in mind, our methodology is equally applicable to cross-sectional imputation. The longitudinal approach may however help in the classification. An individual with no or very few observed values at one time point may yet have a high probability of being classified into the right cluster. Its classification at other time points, and the transition matrices in between, increase the probability of a correct classification.

References

Banfield, J. D. and Raftery, A. E. (1993). "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, 49, 3, 803-821.

Bensmail, H., Celeux, G., Raftery, A. E. and Robert, C. P. (1997). "Inference in Model-Based Cluster Analysis," *Statistics and Computing*, 7, 1-10.

Bergman, L. R. and Magnusson, D. (1997). "A person-oriented approach in research on developmental psychopathology," *Development and Psychopathology*, 9, 291-319.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*, Chichester: John Wiley and Sons.

Congdon, P. (2007). *Bayesian Statistical Modelling,* West Sussex: Wiley.

Engels, J. M., and Diehr, P. (2003). "Imputation of Missing Logitudinal Data: A Comparison of Methods," *Journal of Epidemiology*, 56, 968-976.

Franzén, J. (2008). "Successive Clustering of Longitudinal Data - A Bayesian Approach," Research Report 2008:2:, Department of Statistics, Stockholm University.

Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transaction on pattern Analysis and Machine Intelligence,* 6, 721, 741.

Ghahramani, Z. and Jordan, M. I. (1994). "Supervised learning from incomplete data via an EM approach," in: Cowan, J. D., Tesarro, G., and Alspector, J. (Eds.). *Advances in Neural Information Processing Systems*, vol. 6, 120-127. Morgan Kaufmann, San Francisco.

Gilks, W. R. and Wang, C. C. Yvonnet, B. and Coursaget, P. (1993). "Random-Effects Models for Longitudinal Data using Gibbs Sampling," *Biometrics*, 49, 2, 441-453.

Laird, N. M. (1988). "Missing Data in Longitudinal Studies," *Statistics in Medicine*, 7, 305-315.

Lin, T. I., Lee, J. C., Ho, H. J. (2006). "On Fast Supervised Learning for Normal Mixture Models with Missing Information," *Pattern Recognition*, 39, 1177-1187.

Little, R. J. A. (1995). "Modeling the Drop-Out Mechanism in Repeated-Measures Studies," *Journal of the American Statistical Association*, 90, 431, 1112-1121

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data,* New York: Wiley.

Liu, C. (1999). "Efficient ML Estimation of the Multivariate Normal Distribution from Incomplete Data," *Journal of Multivariate Analysis*, 69, 206-217.

Liu, M., Taylor, J. M. G., and Belin, T. R. (2000). "Multiple Imputation and Posterior Simulation for Multivariate Missing Data in Longitudinal Studies," Biometrics, 56, 1157-1163.

Magnusson, D. (1988). *Individual Development from an Interactional Perspective - A Longitudinal Study*, Hillsdale, NJ: Lawrence Erlbaum.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models,* New York: Wiley.

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

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London.

Scott, S. L., James, G. A., and Sugar, C. A. (2005). "Hidden Markov Models for Longitudinal Comparisons," *Journal of the American Statistical Association*, 100, 470, 359-369.

Twisk, J, and de Wente, W. (2002). "Attrition in Longitudinal Studies: How to Deal with Missing Data," *Journal of Clinical Epidemiology*, 55, 329-337.

Appendix A

Posterior distribution of the covariance matrices is the inverse Wishart

$$\mathbf{\Sigma}_j^{(t)} \left| \mathbf{y}^{(t)}, \mathbf{V}^{(t)} \sim W^{-1} \left(n_j^{(t)} + m_j^{(t)}, \boldsymbol{\psi}_j^{(t)} + \mathbf{\Lambda}_j^{(t)} + \frac{n_j^{(t)} \tau_j^{(t)}}{n_j^{(t)} + \tau_j^{(t)}} (\overline{\mathbf{y}}_j^{(t)} - \boldsymbol{\xi}_j^{(t)})(\overline{\mathbf{y}}_j^{(t)} - \boldsymbol{\xi}_j^{(t)})' \right) \right.$$

where $n_j^{(t)}$ is the number of observations from Cluster j, $\overline{\mathbf{y}}_j^{(t)}$ is the sample mean in Cluster j, and $\mathbf{\Lambda}_j^{(t)} = \sum_{i \in j} (\mathbf{y}_i^{(t)} - \overline{\mathbf{y}}_j^{(t)})(\mathbf{y}_i^{(t)} - \overline{\mathbf{y}}_j^{(t)})'$, for $t = 1, ..., T$.

Posterior distribution of the mean vectors is the normal distribution

$$\boldsymbol{\mu}_j^{(t)} \left| \mathbf{y}^{(t)}, \mathbf{\Sigma}_j^{(t)}, \mathbf{V}^{(t)} \sim N_M \left(\overline{\boldsymbol{\xi}}_j^{(t)}, \mathbf{\Sigma}_j^{(t)} / (\tau_j^{(t)} + n_j^{(t)}) \right) \right.$$

where $\overline{\boldsymbol{\xi}}_j^{(t)} = \frac{\tau_j^{(t)} \boldsymbol{\xi}_j^{(t)} + n_j^{(t)} \overline{\mathbf{y}}_j^{(t)}}{\left(n_j^{(t)} + \tau_j^{(t)} \right)} \qquad t = 1, ..., T$

Posterior distribution of the cluster probabilies is the Dirichlet distribution

$$\omega_1^{(1)}, ..., \omega_{J^{(1)}}^{(1)} \mid \mathbf{V}^{(1)} \sim Dir \left(\left(\alpha_1 + \sum_{i=1}^{n} I \left(v_i^{(1)} = 1 \right) \right), ..., \left(\alpha_{J^{(1)}} + \sum_{i=1}^{n} I \left(v_i^{(1)} = J^{(1)} \right) \right) \right)$$

where $\sum_{i=1}^{n} I \left(v_i^{(j)} = j \right)$ counts the number of observations classified into Cluster j.

The posterior distributions for each row in the transition matrices is the Dirichlet distribution

$$\mathbf{Q}_t(j^{(t)}, \cdot) \left| \mathbf{V}^{(t)} \sim Dir \left(\beta_1^{(t)} + n^{(t)} \left(j^{(t)}, 1 \right), ..., \beta_{J^{(t)}}^{(t)} + n^{(t)} \left(j^{(t)}, J^{(t+1)} \right) \right) \right.$$

where $n^{(t)}(j^{(t)}, j^{(t+1)})$ counts the number of transitions from Cluster $j^{(t)}$ to Cluster $j^{(t+1)}$, between Time t and $t + 1$.

Appendix B

η (%)			*Var 1*	*Cov 1*	*Var 2*	*Cov 2*	*Var 3*	*Cov 3*	*Var 4*	*Cov 4*	*Var 5*	*Cov 5*	*Var 6*	*Cov 6*
0	Time 1	EE	0.0041	0.0017	0.0147	0.0038	0.0090	0.0077	0.0114	0.0244	0.0624	0.0149	0.0030	0.0016
		Var	0.0068	0.0033	0.0132	0.0068	0.0176	0.0095	0.0251	0.0132	0.0387	0.0189	0.0169	0.0084
	Time 2	EE	0.0147	0.0006	0.0193	0.0022	0.0159	0.0042	0.0206	0.0070				
		Var	0.0072	0.0036	0.0086	0.0043	0.0155	0.0076	0.0169	0.0088				
	Time 3	EE	0.0035	0.0027	0.0087	0.0062	0.0133	0.0035	0.0770	0.0266	0.0777	0.0163		
		Var	0.0063	0.0032	0.0089	0.0045	0.0124	0.0060	0.0323	0.0167	0.0365	0.0180		
5	Time 1	EE	0.0038	0.0020	0.0211	0.0028	0.0073	0.0083	0.0184	0.0313	0.0744	0.0212	0.0035	0.0025
		Var	0.0074	0.0038	0.0150	0.0081	0.0210	0.0117	0.0291	0.0157	0.0425	0.0213	0.0182	0.0092
	Time 2	EE	0.0126	0.0016	0.0224	0.0024	0.0147	0.0041	0.0186	0.0096				
		Var	0.0075	0.0039	0.0092	0.0049	0.0168	0.0087	0.0181	0.0097				
	Time 3	EE	0.0033	0.0041	0.0075	0.0057	0.0129	0.0022	0.0908	0.0310	0.0833	0.0200		
		Var	0.0069	0.0037	0.0095	0.0050	0.0127	0.0065	0.0354	0.0186	0.0392	0.0195		
10	Time 1	EE	0.0034	0.0037	0.0077	0.0052	0.0215	0.0379	0.0072	0.0067	0.0030	0.0036	0.1197	0.0307
		Var	0.0077	0.0042	0.0159	0.0084	0.0300	0.0166	0.0240	0.0137	0.0196	0.0104	0.0524	0.0261
	Time 2	EE	0.0165	0.0014	0.0085	0.0028	0.0141	0.0065	0.0286	0.0064				
		Var	0.0081	0.0044	0.0091	0.0050	0.0182	0.0096	0.0196	0.0109				
	Time 3	EE	0.0064	0.0030	0.0054	0.0064	0.0174	0.0044	0.1263	0.0423	0.1441	0.0249		
		Var	0.0074	0.0040	0.0102	0.0055	0.0166	0.0088	0.0451	0.0256	0.0561	0.0288		
25	Time 1	EE	0.0112	0.0094	0.0218	0.0096	0.0054	0.0058	0.0475	0.0795	0.0453	0.0303	0.0098	0.0039
		Var	0.0110	0.0067	0.0263	0.0158	0.0380	0.0238	0.0470	0.0288	0.0579	0.0322	0.0285	0.0172
	Time 2	EE	0.0197	0.0032	0.0170	0.0075	0.0273	0.0053	0.0727	0.0285				
		Var	0.0106	0.0068	0.0145	0.0092	0.0250	0.0145	0.0328	0.0211				
	Time 3	EE	0.0076	0.0036	0.0128	0.0098	0.0146	0.0030	0.2287	0.1150	0.1063	0.0426		
		Var	0.0098	0.0064	0.0161	0.0098	0.0186	0.0109	0.0675	0.0430	0.0664	0.0382		
40	Time 1	EE	0.0180	0.0090	0.0239	0.0185	0.0028	0.0202	0.1971	0.1495	0.1203	0.0790	0.0204	0.0181
		Var	0.0144	0.0098	0.0338	0.0231	0.0792	0.0509	0.0884	0.0563	0.1111	0.0681	0.0456	0.0291
	Time 2	EE	0.0372	0.0170	0.0215	0.0109	0.0418	0.0143	0.0722	0.0479				
		Var	0.0152	0.0108	0.0206	0.0151	0.0412	0.0247	0.0462	0.0300				
	Time 3	EE	0.0032	0.0060	0.0243	0.0099	0.0178	0.0104	0.3150	0.2133	0.2416	0.0833		
		Var	0.0144	0.0107	0.0214	0.0151	0.0281	0.0174	0.0869	0.0590	0.1244	0.0725		
45	Time 1	EE	0.0154	0.0125	0.0026	0.0050	0.0254	0.0377	0.1678	0.1477	0.0685	0.0429	0.0271	0.0125
		Var	0.0156	0.0102	0.0349	0.0251	0.0990	0.0594	0.1083	0.0659	0.1011	0.0590	0.0572	0.0360
	Time 2	EE	0.0167	0.0207	0.0366	0.0055	0.0868	0.0364	0.0993	0.0293				
		Var	0.0151	0.0115	0.0193	0.0140	0.0541	0.0359	0.0441	0.0322				
	Time 3	EE	0.0145	0.0082	0.0117	0.0163	0.0252	0.0088	0.4676	0.2182	0.2459	0.0554		
		Var	0.0140	0.0101	0.0237	0.0159	0.0520	0.0340	0.1264	0.0862	0.1376	0.0783		
50	Time 1	EE	0.0019	0.0172	0.0140	0.0086	0.0896	0.0231	0.3505	0.3829	0.0952	0.0758	0.1030	0.0338
		Var	0.0150	0.0105	0.1015	0.0595	0.2400	0.1379	0.1253	0.0816	0.1535	0.0906	0.2106	0.1199
	Time 2	EE	0.0326	0.0123	0.0306	0.0281	0.1201	0.0531	0.1108	0.0441				
		Var	0.0186	0.0153	0.0362	0.0233	0.0640	0.0390	0.0592	0.0456				
	Time 3	EE	0.0168	0.0113	0.0174	0.0181	0.0526	0.0198	0.2756	0.2850	0.4256	0.1709		
		Var	0.0231	0.0186	0.0309	0.0223	0.0733	0.0473	0.1109	0.0775	0.3002	0.1678		
55	Time 1	EE	0.0069	0.0108	0.0301	0.0161	0.1351	0.1012	0.1016	0.1090	0.2771	0.1910	0.1589	0.0505
		Var	0.0207	0.0159	0.1504	0.0964	0.2324	0.1685	0.2397	0.1502	0.1921	0.1177	0.1975	0.1310
	Time 2	EE	0.0411	0.0195	0.0891	0.0094	0.1896	0.0723	0.1277	0.0637				
		Var	0.0236	0.0184	0.0478	0.0303	0.0926	0.0624	0.0695	0.0481				
	Time 3	EE	0.0045	0.0074	0.0110	0.0225	0.0450	0.0227	0.2646	0.2781	0.5159	0.1160		
		Var	0.0225	0.0186	0.0319	0.0216	0.1062	0.0636	0.1266	0.0931	0.2443	0.1368		
60	Time 1	EE	0.0961	0.1170	0.0834	0.0129	0.2323	0.0763	0.2028	0.1268	0.0828	0.0134	0.1438	0.0311
		Var	0.1181	0.0939	0.2184	0.1382	0.2903	0.2720	0.2797	0.2369	0.2867	0.1738	0.1303	0.0704
	Time 2	EE	0.0513	0.0246	0.0183	0.0150	0.1941	0.0487	0.2538	0.0408				
		Var	0.0283	0.0231	0.0505	0.0409	0.1484	0.0772	0.0877	0.0613				
	Time 3	EE	0.0187	0.0112	0.0323	0.0118	0.0682	0.0222	0.4151	0.4330	0.4036	0.1427		
		Var	0.1407	0.0717	0.0654	0.0378	0.1953	0.1142	0.1505	0.1149	0.2271	0.1263		

TABLE 10: Performance measures for separated groups. Estimation deviations of (co)variances presented for each non-response rate, time point, and cluster separately. The Var columns give the mean estimation deviation for the diagonal in the covariance matrix for each cluster, i.e. the variances. The Cov columns give the same values for the non-diagonal elements in each matrix, i.e. the covariances.

η (%)			*Var 1*	*Cov 1*	*Var 2*	*Cov 2*	*Var 3*	*Cov 3*	*Var 4*	*Cov 4*	*Var 5*	*Cov 5*	*Var 6*	*Cov 6*
0	Time1	EE	0.0038	0.0064	0.0226	0.0093	0.0291	0.0055	0.1680	0.0063	0.0175	0.0070	0.0261	0.0100
		Var	0.0111	0.0056	0.0192	0.0090	0.0245	0.0115	0.0534	0.0248	0.0405	0.0202	0.0280	0.0140
	Time 2	EE	0.0072	0.0053	0.0022	0.0042	0.0554	0.0096	0.0045	0.0039				
		Var	0.0089	0.0046	0.0089	0.0045	0.0236	0.0111	0.0331	0.0153				
	Time 3	EE	0.0099	0.0033	0.0126	0.0055	0.0126	0.0046	0.0111	0.0066	0.0636	0.0063		
		Var	0.0070	0.0035	0.0100	0.0048	0.0165	0.0078	0.0211	0.0105	0.0312	0.0142		
5	Time 1	EE	0.0080	0.0087	0.0412	0.0169	0.0368	0.0056	0.1804	0.0078	0.0090	0.0089	0.0295	0.0124
		Var	0.0124	0.0064	0.0259	0.0126	0.0277	0.0131	0.0616	0.0285	0.0470	0.0251	0.0347	0.0182
	Time 2	EE	0.0065	0.0044	0.0037	0.0063	0.0656	0.0109	0.0065	0.0065				
		Var	0.0096	0.0051	0.0097	0.0050	0.0262	0.0127	0.0462	0.0208				
	Time 3	EE	0.0081	0.0034	0.0125	0.0054	0.0167	0.0059	0.0119	0.0078	0.0692	0.0087		
		Var	0.0074	0.0039	0.0109	0.0054	0.0178	0.0088	0.0244	0.0127	0.0330	0.0153		
10	Time 1	EE	0.0138	0.0090	0.0398	0.0201	0.0292	0.0056	0.2463	0.0087	0.0207	0.0058	0.0169	0.0061
		Var	0.0150	0.0079	0.0248	0.0126	0.0315	0.0153	0.0868	0.0367	0.0542	0.0259	0.0324	0.0172
	Time 2	EE	0.0089	0.0051	0.0055	0.0040	0.0526	0.0157	0.0131	0.0123				
		Var	0.0106	0.0058	0.0102	0.0055	0.0283	0.0150	0.0489	0.0213				
	Time 3	EE	0.0109	0.0046	0.0169	0.0104	0.0199	0.0070	0.0111	0.0097	0.0706	0.0124		
		Var	0.0082	0.0045	0.0121	0.0063	0.0200	0.0101	0.0266	0.0140	0.0376	0.0176		
25	Time 1	EE	0.0136	0.0124	0.0321	0.0065	0.0647	0.0134	0.4649	0.0408	0.1179	0.0756	0.0104	0.0034
		Var	0.0267	0.0161	0.0337	0.0178	0.0538	0.0278	0.1617	0.0803	0.1365	0.0328	0.0364	0.0212
	Time 2	EE	0.0101	0.0078	0.0089	0.0071	0.1085	0.0674	0.1108	0.0635				
		Var	0.0137	0.0080	0.0145	0.0093	0.0538	0.0328	0.1357	0.0656				
	Time 3	EE	0.0076	0.0047	0.0242	0.0102	0.0432	0.0153	0.0796	0.0243	0.1029	0.0171		
		Var	0.0110	0.0069	0.0185	0.0102	0.0324	0.0176	0.0529	0.0297	0.0704	0.0375		
40	Time 1	EE	0.0084	0.0262	0.1181	0.0068	0.0254	0.0174	0.3289	0.0315	0.0553	0.0253	0.0473	0.0234
		Var	0.0261	0.0170	0.0469	0.0235	0.0548	0.0293	0.1559	0.0664	0.1135	0.0623	0.0530	0.0360
	Time 2	EE	0.0144	0.0191	0.0041	0.0149	0.0820	0.0337	0.0438	0.0261				
		Var	0.0216	0.0149	0.0222	0.0141	0.0580	0.0368	0.1023	0.0510				
	Time 3	EE	0.0058	0.0075	0.0229	0.0080	0.0136	0.0116	0.0268	0.0279	0.2272	0.0392		
		Var	0.0153	0.0108	0.0205	0.0125	0.0386	0.0226	0.0463	0.0298	0.0969	0.0680		
45	Time 1	EE	0.0590	0.0167	0.0417	0.0134	0.0814	0.0513	0.2628	0.0448	0.2763	0.0411	0.0591	0.0068
		Var	0.0565	0.0388	0.0569	0.0281	0.2319	0.0825	0.3211	0.1281	0.3599	0.1666	0.0727	0.0504
	Time 2	EE	0.0235	0.0202	0.0127	0.0036	0.1229	0.0269	0.1683	0.0695				
		Var	0.0237	0.0167	0.0241	0.0174	0.0673	0.0416	0.1455	0.0705				
	Time 3	EE	0.0157	0.0103	0.0398	0.0226	0.0415	0.0328	0.0842	0.0261	0.2458	0.0575		
		Var	0.0191	0.0146	0.0413	0.0241	0.0633	0.0395	0.1129	0.0605	0.2106	0.1109		
50	Time 1	EE	0.0123	0.0294	0.0269	0.0108	0.1308	0.0357	0.3286	0.0596	0.2069	0.0407	0.1176	0.0261
		Var	0.0450	0.0301	0.1288	0.0679	0.2418	0.1385	0.2486	0.1266	0.3236	0.1672	0.0977	0.0600
	Time 2	EE	0.0398	0.0198	0.0370	0.0194	0.2067	0.0738	0.1240	0.0452				
		Var	0.1256	0.0613	0.1324	0.0612	0.0926	0.0635	0.1897	0.0959				
	Time 3	EE	0.0059	0.0114	0.0391	0.0205	0.0183	0.0173	0.1644	0.0458	0.0437	0.0315		
		Var	0.0173	0.0124	0.0411	0.0279	0.0461	0.0273	0.1355	0.0759	0.0932	0.0539		
55	Time 1	EE	0.0398	0.0381	0.0958	0.0084	0.0977	0.0120	0.3614	0.0318	0.2929	0.0353	0.2511	0.0185
		Var	0.0474	0.0342	0.2023	0.1290	0.2170	0.1117	0.3742	0.1562	0.4132	0.1718	0.4998	0.1230
	Time 2	EE	0.0514	0.0173	0.0329	0.0119	0.1331	0.0436	0.0732	0.0207				
		Var	0.1501	0.0823	0.1303	0.0708	0.1627	0.1094	0.1678	0.1001				
	Time 3	EE	0.0270	0.0155	0.0526	0.0132	0.0749	0.0433	0.1811	0.0450	0.5154	0.2118		
		Var	0.0451	0.0328	0.0471	0.0285	0.0831	0.0562	0.1437	0.0807	0.1896	0.1534		

TABLE 11: Performance measures for overlapping groups. Estimation deviations of (co)variances presented for each non-response rate, time point, and cluster separately. The Var columns give the mean estimation deviation for the diagonal in the covariance matrix for each cluster, i.e. the variances. The Cov columns give the same values for the non-diagonal elements in each matrix, i.e. the covariances.

Separated Groups

η (%)			*Var 1*	*Cov 1*	*Var 2*	*Cov 2*	*Var 3*	*Cov 3*	*Var 4*	*Cov 4*	*Var 5*	*Cov 5*	*Var 6*	*Cov 6*
5	Time 1	EE	0.0049	0.0073	0.0174	0.0102	0.0213	0.0411	0.0641	0.1040	0.3097	0.0906	0.0122	0.0089
		Var	0.0155	0.0076	0.0264	0.0137	0.1203	0.0734	0.1139	0.0664	0.1451	0.0719	0.0420	0.0209
	Time 2	EE	0.0376	0.0029	0.0273	0.0072	0.4846	0.0620	0.0497	0.0252				
		Var	0.0180	0.0091	0.0235	0.0122	0.1193	0.0487	0.0379	0.0197				
	Time 3	EE	0.0060	0.0037	0.0173	0.0218	0.0206	0.0101	0.3977	0.2027	0.1046	0.0363		
		Var	0.0137	0.0068	0.0266	0.0130	0.0463	0.0220	0.1220	0.0732	0.0720	0.0351		

Overlapping Groups

η (%)			*Var 1*	*Cov 1*	*Var 2*	*Cov 2*	*Var 3*	*Cov 3*	*Var 4*	*Cov 4*	*Var 5*	*Cov 5*	*Var 6*	*Cov 6*
5	Time 1	EE	0.0094	0.0111	0.0284	0.0124	0.0127	0.0127	0.6554	0.0459	0.2588	0.0820	0.0820	0.0174
		Var	0.0229	0.0114	0.0484	0.0216	0.0782	0.0311	0.6407	0.0824	0.2259	0.1306	0.1542	0.0434
	Time 2	EE	0.0025	0.0121	0.0090	0.0121	0.0365	0.0215	0.0260	0.0184				
		Var	0.0181	0.0096	0.0213	0.0107	0.0902	0.0387	0.0912	0.0383				
	Time 3	EE	0.0227	0.0137	0.0216	0.0108	0.0407	0.0217	0.0070	0.0101	0.1787	0.0373		
		Var	0.0173	0.0087	0.0195	0.0095	0.0454	0.0210	0.0457	0.0227	0.1077	0.0533		

TABLE 12: Performance measures when eliminating individuals with missing values. Estimation deviations of (co)variances presented for each non-response rate, time point, and cluster separately. The Var columns give the mean estimation deviation for the diagonal in the covariance matrix for each cluster, i.e. the variances. The Cov columns give the same values for the non-diagonal elements in each matrix, i.e. the covariances.

η (%)			*Var 1*	*Cov 1*	*Var 2*	*Cov 2*	*Var 3*	*Cov 3*	*Var 4*	*Cov 4*	*Var 5*	*Cov 5*	*Var 6*	*Cov 6*
5	Time 1	EE	0.0555	0.0026	0.0155	0.0020	0.0308	0.0120	0.3034	0.0293	0.0975	0.0209	0.0366	0.0045
		Var	0.0104	0.0052	0.0143	0.0069	0.0326	0.0205	0.0573	0.0244	0.0468	0.0468	0.0250	0.0122
	Time 2	EE	0.0248	0.0017	0.0151	0.0015	0.0366	0.0053	0.1601	0.0180				
		Var	0.0079	0.0040	0.0089	0.0044	0.0191	0.0097	0.0260	0.0138				
	Time 3	EE	0.0053	0.0041	0.0280	0.0036	0.0120	0.0022	0.4645	0.0602	0.1416	0.0159		
		Var	0.0070	0.0036	0.0098	0.0049	0.0132	0.0063	0.0548	0.0299	0.0441	0.0216		
10	Time 1	EE	0.2030	0.0057	0.0110	0.0041	0.0730	0.0407	0.4770	0.0120	0.3039	0.0500	0.0607	0.0207
		Var	0.0142	0.0072	0.0201	0.0078	0.0417	0.0301	0.0643	0.0305	0.0831	0.0408	0.0272	0.0139
	Time 2	EE	0.0469	0.0011	0.0103	0.0014	0.0860	0.0050	0.4121	0.0139				
		Var	0.0085	0.0043	0.0076	0.0037	0.0241	0.0122	0.0351	0.0190				
	Time 3	EE	0.0059	0.0038	0.0470	0.0046	0.0201	0.0035	1.2250	0.0911	0.3307	0.0224		
		Var	0.0073	0.0038	0.0107	0.0053	0.0197	0.0094	0.0848	0.0467	0.0679	0.0346		
25	Time 1	EE	0.4948	0.0499	1.6946	0.0878	7.5574	2.5291	2.3107	0.2454	1.1142	0.1072	1.0130	0.1338
		Var	0.0289	0.0145	0.0996	0.0178	0.1280	0.0530	0.5753	0.1320	0.4369	0.1939	0.1305	0.0496
	Time 2	EE	0.0895	0.0040	0.1544	0.0026	1.4898	0.4845	4.2239	0.7507				
		Var	0.0133	0.0076	0.0427	0.0065	0.2319	0.1254	0.4536	0.1602				
	Time 3	EE	0.0423	0.0038	0.1688	0.0123	0.0140	0.0054	3.0694	0.0857	0.6207	0.0733		
		Var	0.0102	0.0052	0.0152	0.0070	0.0180	0.0088	0.1310	0.0722	0.1056	0.0573		
40	Time 1	EE	0.5088	0.0594	6.2618	1.2229	9.1801	2.1000	6.4056	1.1659	4.2800	0.9468	0.5438	0.0789
		Var	0.0962	0.0421	0.1573	0.0531	0.1726	0.0647	0.3341	0.1357	0.1298	0.0414	0.2240	0.1305
	Time 2	EE	0.2412	0.0513	1.1319	0.1151	4.2177	0.6971	0.9517	0.2413				
		Var	0.0334	0.0194	0.0391	0.0087	0.1185	0.0470	0.0340	0.0110				
	Time 3	EE	0.7903	0.1097	0.0927	0.0783	4.9252	1.5897	5.2283	0.6791	0.8769	0.1736		
		Var	0.0418	0.0124	0.0289	0.0151	0.1498	0.0589	0.2226	0.1147	0.0449	0.0150		

η (%)			Var 1	Cov 1	Var 2	Cov 2	Var 3	Cov 3	Var 4	Cov 4	Var 5	Cov 5	Var 6	Cov 6
5	Time 1	EE	0.0198	0.0064	0.0483	0.0104	0.0639	0.0170	0.2166	0.0078	0.0492	0.0245	0.0580	0.0120
		Var	0.0141	0.0069	0.0298	0.0138	0.0359	0.0180	0.0884	0.0297	0.0642	0.0347	0.0595	0.0214
	Time 2	EE	0.0091	0.0042	0.0027	0.0055	0.1433	0.0145	0.0155	0.0129				
		Var	0.0092	0.0046	0.0093	0.0048	0.0335	0.0164	0.0556	0.0266				
	Time 3	EE	0.0136	0.0034	0.0178	0.0058	0.0369	0.0070	0.0291	0.0097	0.1013	0.0101		
		Var	0.0073	0.0036	0.0116	0.0055	0.0195	0.0086	0.0287	0.0141	0.0352	0.0163		
10	Time 1	EE	0.0923	0.0188	0.1017	0.0227	0.0758	0.0065	0.4728	0.0064	0.0864	0.0325	0.0338	0.0061
		Var	0.0237	0.0110	0.0311	0.0169	0.1768	0.0306	0.3454	0.0639	0.1048	0.0536	0.0583	0.0193
	Time 2	EE	0.0161	0.0057	0.0033	0.0048	0.1657	0.0249	0.0268	0.0252				
		Var	0.0114	0.0054	0.0101	0.0056	0.0496	0.0255	0.0755	0.0330				
	Time 3	EE	0.0189	0.0040	0.0391	0.0078	0.0714	0.0090	0.0232	0.0057	0.2870	0.0225		
		Var	0.0079	0.0040	0.0129	0.0065	0.0295	0.0113	0.0310	0.0147	0.0686	0.0323		
25	Time 1	EE	2.1150	0.3862	0.2410	0.0266	1.6014	0.4440	1.1865	0.0397	2.1580	0.1595	0.0067	0.0077
		Var	0.0855	0.0238	0.2617	0.1174	0.0696	0.0203	0.1522	0.0840	0.4932	0.1190	0.0471	0.0270
	Time 2	EE	0.1885	0.0252	0.0641	0.0119	0.4422	0.0373	0.4002	0.0516				
		Var	0.0650	0.0078	0.0741	0.0123	0.0425	0.0248	0.0928	0.0328				
	Time 3	EE	0.1272	0.0057	0.7678	0.0843	0.3716	0.0148	0.4634	0.1894	0.9220	0.1994		
		Var	0.0210	0.0080	0.0394	0.0117	0.0493	0.0153	0.0275	0.0131	0.1265	0.0592		
40	Time 1	EE	0.9853	0.0791	2.2376	0.3693	0.8117	0.0922	1.5983	0.7528	2.2048	0.3379	0.6304	0.2100
		Var	0.2237	0.0611	0.0971	0.0194	0.0297	0.0043	1.0506	0.5467	0.1951	0.0371	0.2429	0.1500
	Time 2	EE	0.3680	0.0342	0.3626	0.0130	1.8161	0.5213	0.4934	0.0388				
		Var	0.0388	0.0056	0.0243	0.0041	0.0901	0.0422	0.0435	0.0107				
	Time 3	EE	0.6135	0.1990	0.2827	0.1717	0.1956	0.0140	0.6842	0.0296	0.4920	0.0634		
		Var	0.3282	0.0699	0.0475	0.0228	0.1627	0.0130	0.0289	0.0057	0.0579	0.0080		

TABLE 13: Performance measures when using mean imputation. Top table: separated groups, bottom table: overlapping groups. Estimation deviations of (co)variances presented for each non-response rate, time point, and cluster separately. The Var columns give the mean estimation deviation for the diagonal in the covariance matrix for each cluster, i.e. the variances. The Cov columns give the same values for the non-diagonal elements in each matrix, i.e. the covariances.

Appendix C

Posterior Covariance Estimates at Time 1 without Imputation	
Covariance 1	*Covariance 2*
$\begin{pmatrix} 1.37 & -0.11 & 0.20 & 0.23 & 0.01 & 0.13 \\ & 0.91 & 0.10 & 0.01 & 0.08 & -0.01 \\ & & 1.40 & 0.02 & -0.03 & 0.18 \\ & & & 0.62 & 0.12 & 0.26 \\ & & & & 0.25 & 0.05 \\ & & & & & 0.51 \end{pmatrix}$	$\begin{pmatrix} 1.00 & 0.01 & 0.36 & 0.10 & 0.01 & 0.03 \\ & 0.06 & 0.01 & 0.01 & 0.01 & 0.01 \\ & & 1.22 & 0.03 & 0.12 & 0.13 \\ & & & 0.56 & 0.15 & 0.24 \\ & & & & 0.33 & 0.10 \\ & & & & & 0.44 \end{pmatrix}$
Covariance 3	*Covariance 4*
$\begin{pmatrix} 1.38 & 0.18 & 0.24 & 0.21 & 0.00 & 0.08 \\ & 1.40 & 0.06 & -0.10 & 0.00 & -0.15 \\ & & 1.64 & 0.03 & 0.00 & 0.20 \\ & & & 0.57 & 0.00 & 0.20 \\ & & & & 0.02 & 0.00 \\ & & & & & 0.51 \end{pmatrix}$	$\begin{pmatrix} 1.34 & 0.13 & 0.04 & 0.15 & 0.01 & 0.00 \\ & 0.65 & 0.04 & -0.04 & -0.01 & -0.06 \\ & & 0.36 & -0.06 & -0.01 & -0.01 \\ & & & 0.56 & 0.02 & 0.13 \\ & & & & 0.15 & -0.00 \\ & & & & & 0.52 \end{pmatrix}$
Covariance 5	
$\begin{pmatrix} 1.75 & -0.01 & 0.05 & -0.05 & -0.07 & 0.09 \\ & 1.63 & -0.58 & -0.01 & -0.06 & -0.15 \\ & & 1.84 & -0.20 & -0.09 & 0.27 \\ & & & 0.78 & 0.09 & 0.05 \\ & & & & 0.32 & -0.01 \\ & & & & & 0.50 \end{pmatrix}$	

TABLE 14: Posterior estimates of covariance matrices at Time 1 for Analysis 1.

Posterior Covariance Estimates at Time 2 without Imputation	
Covariance 1	*Covariance 2*
$\begin{pmatrix} 1.03 & 0.26 & 0.33 & -0.03 & 0.01 & 0.00 \\ & 1.27 & 0.36 & -0.13 & 0.04 & -0.12 \\ & & 1.18 & 0.05 & 0.01 & 0.18 \\ & & & 0.41 & 0.03 & 0.24 \\ & & & & 0.13 & 0.03 \\ & & & & & 0.66 \end{pmatrix}$	$\begin{pmatrix} 0.98 & 0.16 & 0.28 & 0.20 & 0.00 & 0.12 \\ & 0.99 & 0.20 & -0.12 & 0.01 & -0.13 \\ & & 1.20 & 0.01 & -0.00 & 0.12 \\ & & & 0.56 & 0.00 & 0.30 \\ & & & & 0.04 & 0.00 \\ & & & & & 0.52 \end{pmatrix}$
Covariance 3	*Covariance 4*
$\begin{pmatrix} 1.17 & 0.12 & 0.33 & 0.17 & -0.00 & 0.15 \\ & 1.19 & 0.03 & -0.19 & 0.00 & -0.16 \\ & & 1.31 & 0.04 & -0.00 & 0.26 \\ & & & 0.51 & 0.00 & 0.26 \\ & & & & 0.02 & -0.00 \\ & & & & & 0.61 \end{pmatrix}$	$\begin{pmatrix} 1.10 & -0.04 & 0.50 & 0.26 & -0.00 & 0.20 \\ & 1.46 & 0.15 & -0.27 & -0.00 & -0.16 \\ & & 1.49 & 0.10 & 0.00 & 0.24 \\ & & & 0.52 & 0.00 & 0.27 \\ & & & & 0.05 & 0.01 \\ & & & & & 0.55 \end{pmatrix}$
Covariance 5	
$\begin{pmatrix} 1.16 & -0.05 & 0.22 & -0.11 & 0.06 & 0.11 \\ & 1.20 & 0.05 & -0.13 & -0.28 & -0.02 \\ & & 1.57 & 0.19 & 0.03 & 0.30 \\ & & & 0.53 & 0.11 & 0.18 \\ & & & & 0.62 & 0.18 \\ & & & & & 0.52 \end{pmatrix}$	

TABLE 15: Posterior estimates of covariance matrices at Time 2 for Analysis 1.

Posterior Covariance Estimates at Time 1 with Imputation	
Covariance 1	*Covariance 2*
$\begin{pmatrix} 1.20 & 0.14 & 0.14 & 0.17 & 0.01 & 0.04 \\ & 1.75 & 0.19 & 0.41 & 0.51 & 0.14 \\ & & 1.25 & 0.02 & 0.05 & 0.25 \\ & & & 0.36 & 0.19 & 0.08 \\ & & & & 0.66 & 0.17 \\ & & & & & 0.60 \end{pmatrix}$	$\begin{pmatrix} 1.24 & 0.07 & 0.34 & 0.00 & -0.12 & 0.00 \\ & 0.27 & 0.06 & 0.01 & 0.01 & -0.02 \\ & & 1.27 & 0.00 & -0.07 & 0.04 \\ & & & 0.05 & 0.00 & 0.00 \\ & & & & 0.58 & 0.10 \\ & & & & & 0.39 \end{pmatrix}$
Covariance 3	*Covariance 4*
$\begin{pmatrix} 1.40 & 0.18 & 0.33 & 0.00 & -0.04 & -0.05 \\ & 1.29 & 0.08 & 0.00 & 0.18 & 0.01 \\ & & 1.70 & -0.00 & -0.05 & 0.13 \\ & & & 0.03 & 0.00 & -0.00 \\ & & & & 0.52 & 0.13 \\ & & & & & 0.46 \end{pmatrix}$	$\begin{pmatrix} 1.42 & 0.02 & 0.15 & 0.00 & -0.14 & -0.09 \\ & 0.33 & 0.04 & -0.02 & -0.01 & -0.03 \\ & & 0.47 & -0.03 & -0.12 & -0.01 \\ & & & 0.16 & 0.05 & 0.03 \\ & & & & 0.63 & 0.11 \\ & & & & & 0.46 \end{pmatrix}$
Covariance 5	
$\begin{pmatrix} 1.73 & 0.07 & 0.19 & 0.00 & -0.10 & 0.04 \\ & 1.81 & -0.41 & -0.06 & 0.18 & -0.02 \\ & & 2.11 & -0.12 & -0.19 & 0.20 \\ & & & 0.40 & 0.12 & 0.06 \\ & & & & 0.57 & 0.07 \\ & & & & & 0.45 \end{pmatrix}$	

TABLE 16: Posterior estimates of covariance matrices at Time 1 for Analysis 2.

Posterior Covariance Estimates at Time 2 with Imputation	
Covariance 1	*Covariance 2*
$\begin{pmatrix} 1.15 & 0.30 & 0.30 & 0.05 & 0.04 & -0.03 \\ & 1.57 & 0.25 & 0.16 & 0.07 & 0.04 \\ & & 1.15 & 0.02 & 0.06 & 0.05 \\ & & & 0.28 & 0.04 & 0.07 \\ & & & & 0.57 & 0.15 \\ & & & & & 0.54 \end{pmatrix}$	$\begin{pmatrix} 0.93 & 0.06 & 0.40 & 0.00 & -0.08 & -0.02 \\ & 1.17 & 0.14 & 0.00 & 0.37 & 0.02 \\ & & 1.22 & -0.00 & 0.05 & 0.14 \\ & & & 0.04 & 0.00 & 0.00 \\ & & & & 0.64 & 0.18 \\ & & & & & 0.52 \end{pmatrix}$
Covariance 3	*Covariance 4*
$\begin{pmatrix} 1.14 & 0.23 & 0.27 & -0.00 & -0.06 & 0.01 \\ & 1.35 & 0.27 & -0.00 & 0.38 & 0.08 \\ & & 1.36 & 0.00 & 0.08 & 0.22 \\ & & & 0.02 & -0.00 & 0.00 \\ & & & & 0.62 & 0.13 \\ & & & & & 0.45 \end{pmatrix}$	$\begin{pmatrix} 1.03 & 0.15 & 0.34 & 0.00 & -0.08 & 0.08 \\ & 1.40 & -0.02 & -0.00 & 0.21 & 0.06 \\ & & 1.47 & 0.00 & 0.06 & 0.20 \\ & & & 0.05 & 0.00 & 0.00 \\ & & & & 0.59 & 0.21 \\ & & & & & 0.51 \end{pmatrix}$
Covariance 5	
$\begin{pmatrix} 1.37 & -0.44 & 0.16 & 0.14 & -0.07 & 0.19 \\ & 1.31 & 0.02 & -0.18 & -0.05 & -0.28 \\ & & 1.66 & 0.06 & -0.03 & 0.31 \\ & & & 0.48 & 0.11 & 0.15 \\ & & & & 0.57 & 0.11 \\ & & & & & 0.76 \end{pmatrix}$	

TABLE 17: Posterior estimates of covariance matrices at Time 2 for Analysis 2.