Longitudinal Data Analysis
Assignment III

An investigation to study the evolution of abdominal aorta over time and probability of developing abdominal aortic aneurysm (AAA) in patients with enlarged abdominal aorta diameter while accounting for missing data.

Student:
Andrew Kamya (1849786)

Supervisors:
Prof. Dr. Geert Molenberghs
Prof. Dr. Geert Verbeke

Abstract

Background: The patient characteristics as well as the diameter (in mm) of the 101 patients were collected every six months after the date of enrollment. There was a massive percentage of dropouts that was observed after Month 6 of follow up. Apart from obtaining summary statistics, individual profile plots were used to gain a general knowledge of what was expected.
Objectives: This is a longitudinal cohort study to investigate the effects of time as well as patient characteristics on the evolution of the diameter of the abdominal aorta for patients with abdominal aortic aneurysm (AAA) while taking the incompleteness of data into account.
This was implemented by analyzing both the continuous outcome (diameter of abdominal aorta (mm)) and dichotomized outcome (diameter $\geq$ 45: 1, and zero otherwise). Conducting sensitivity analysis for the continuous outcome –that is the diameter in mm.
Methodology: Exploratory analysis was done primarily using individual profile plots. Dropout summary pattern and descriptive statistics on percentage dropout per patients' follow up visit. The mean structure, variance structure and correlation plots were used to determine the functional form of the model. Weighted Generalized Estimating Equations was used in the case of the incomplete data with binary outcome while direct likelihood method was used for the continuous case. Multiple imputation was performed to impute missing data points.
Results: Direct Likelihood method for the linear mixed model showed a significant effect of age, while the time, and the interaction of time and BMI on the probability of having a diameter of abdominal aorta greater than 45mm were significant using the weighted generalized estimating equations.
Conclusions: Age appears to be sensitive to the choice of assumed missing data mechanism, since the inference changes under MAR and MNAR.

Keywords : *Abdominal Aortic Aneurysm (AAA), Missing data, Multiple Imputation (MI), Sensitivity analysis, Weighted Generalized Estimating Equation (weighted GEE)*

Contents

1 Introduction

An abdominal aortic aneurysm (AAA) is an enlargement of the abdominal aorta, which is the main artery supplying the blood to the body. One of the most common occurring disorder of the abdominal aorta is AAA. This condition is mainly caused by atherosclerosis, which occurs as a result of hardening of the arteries. The inflammation or degeneration of the aortic walls caused by high blood pressure or an infection of the aorta also results in AAA. The abdominal aortic aneurysm commonly affects individuals who are 65 years and older. The high occurrence has been witnessed among smokers and mostly males. (Voop, 2005).
The symptomatic aneurysm has a high risk of rupture and can be life-threatening as large amounts of blood spill into the abdominal cavity, which is an indication for surgery. The mortality of AAA rupture is up to 90%. 65 to 75% of patients die before they arrive at the hospital, and up to 90% die before they reach the operating room. Therefore symptomatic and large aneurysms are considered for repair by surgical methods. Often, an intervention is required if the aneurysm grows more than 1cm per year or when it is more significant than 5.5cm (Aggarwal, 2011).
Although several clinical studies have been conducted, few reliable data are available on the growth trend on the enlargement of the abdominal aorta. Therefore, the purpose why this study was done was to study the evolution of the artery diameter on patients with AAA as well as the relation with some patient characteristics.

In clinical trials, missing data is one of the challenges that is always expected. There are several reasons that lead to dropouts during the study. The most common reported reasons for dropouts were; lost to follow-up, death, subject decision not to continue with the study, subject turning ineligible during the study causing investigators decision to terminate their participation et cetera. The main problem resulting from dropout as reported by Molenberghs and Verbeke (2005) being low precision in the results. This is due to increasing variability which is resulted from a decrease in sample size. In this respect, AAA study experienced large amount of dropouts observed during the study.
This report is structured in 6 sections. In this section, the background information and study objective were defined. Section 2 explains more about the data. Section 3 discusses methodology used in data exploration and inference. In section 4 and 5 the interpretation of the results, discussion and recommendation were presented.

2 Research questions

- To describe the evolution of the diameter of abdominal aorta in patients using linear mixed model. This was to be done by taking into account missingness in the data, starting with missing at random approach (MAR)
- Describing marginal evolution using generalized estimating equations approach by considering a dichotomized response. This was to be implement by applying missing data correction techniques; weighted generalized equations and generalized estimating equations after multiple imputation.
- Conducting sensitivity analysis on the mechanisms of the missingness for the continuous outcome after accounting for the missing observations in the data.

3 Data Description

The data set contained observations of 101 AAA patients aged between 52 and 86 years old. These patients were followed up and data were obtained after every six months. During these follow-up visits, the diameter of the artery and several patient characteristics were collected. The variables under the study were age of patient at baseline, coronary disease history, smoking status, body mass index (BMI), length (in cm) and weight (in kg). The response was the continuous and binary version of the artery diameter of the patients. Binary response was;

$$D_{ij} = \begin{cases} 1 & \text{if } Diameter \geq 45 \\ 0 & \text{if } Diameter < 45 \end{cases} \qquad \text{where } \; i = 1, 2, ..., 101 \; \text{ and } \; j = 1, 2, ..., 8$$

Furthermore, all the predictors except patient weight and height were used in the analysis. This decision was made because Body Mass Index (BMI) is a function of height and weight and therefore contain information from the two variables.

4 Methods

4.1 Exploratory Data Analysis

Prior to the main analysis, exploration of the data was done. Exploratory analysis was conducted to the data to discover patterns, data structure and spot any existing anomalies and to obtain clear insight on conceivable implication before model building and hypothesis testing. Majorly, graphical plots together with tables were used. The graphical explorations included; the individual profile plots, the mean structure, the variance structure, the covariance and correlation structure of the outcome of interest. A table was used to present the number of completers, dropout and intermittent observations in the data.

4.2 Missing Data Mechanisms

As discussed earlier, usually, clinical trials are a rich studies to generate data to prove or disprove the hypothesis. One challenging issue after the enrollment of study subjects in a clinical trial is attrition and missing data due to any number of reasons (Kaushal S., 2014). The missing data may create bias causing invalidity of results and conclusions. However, if we plainly preclude these patients with missing data, this might cause a reduction in the number of subjects which in return affects the power of the study. Also, it is likely that patients with missing values are the ones with extreme values (treatment failure, toxicity, and good responders). Excluding these patients will lead to underestimation of variability and hence may lead to inaccurate conclusions (European Medicines Agency, 2010). Luckily, to counteract this challenge, mechanisms that take care of missing data are fully developed and can be applied. As formulated by Rubin (1976), the missingness mechanisms can be classified into three categories; missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Missingness is said to be MCAR if the occurrence of the missingness does not depend on either the observed or missing values. Differently, MAR when the missingness conditionally depends only on the observed data whereas missingness is MNAR if the probability of a measurement missing, depends on unobserved data.

4.3 Techniques for Handling the Missing Data

There are various ways of handling incomplete data ranging from simple and basic techniques to complex ones. Some of the techniques include; complete case analysis (CC), last observation carried forward (LOCF), and direct likelihood based on ignorability where under certain conditions some assumptions are ignored. The first two techniques have been been condemned, since they lead to biased results and hence caution should be taken when applying them. Molenberghs and Verbeke (2005) in their book have argued that LOCF and CC analyses lead to loss of information since only part of the data is considered by this methods and can only be relevant under restricted MCAR assumption of the missing data mechanism. The commonly used methods when dealing with incomplete data, are the direct likelihood, weighted GEE, and multiple imputation. These methods are valid under MAR assumptions (Verbeke and Molenberghs, 2000).

Direct likelihood is technique of handling incomplete data and is valid under the MAR assumption. As discussed by Beunckens *et al* (2005), likelihood-based mixed effects models for continuous responses are valid under MAR assumption. Further in discussion, they reported that for longitudinal studies with missing data, a mixed model only requires that missing data are MAR. These mixed-effects models allow the inclusion of subjects with missing values at some time points (both dropout and intermittent missingness). Alternatively, likelihood-based MAR analysis has been termed as likelihood-based ignorable analysis, or, a direct likelihood analysis.
In a direct likelihood analysis, the observed data are used without deletion nor imputation. In doing so, appropriate adjustments, that is, valid under MAR, are made to parameters at times when data are incomplete, due to the within-patient correlation (Beunckens *et al*, 2005).

Sometimes we are tasked to analyze binary responses. For this type of outcome, Liang and Zeger (1986) proposed the GEE approach whose inferences are valid only under strong assumption that missingness in the data are MCAR since GEE doesn't follow a full likelihood approach. However, weighted GEE has been adopted as a remedy to using normal GEE. Weighted GEE permits missingness in the data to be valid under MAR assumptions. The weights are defined at individual measurement level. Although weighting pertains conducting a complete records analysis, we should put in mind that the complete records are not a representative sample from the population of interest. In longitudinal studies, weights can be subject-level weights or observation-level weights. The observation-level weights are preferable and it involves building separate weight for every measurement within a subject. Indeed, this means that for each occasion within a sequence where dropout occurs, the weight is multiplied by the predictive probability. According to Jansen *et al* (2006), the use of weighted GEE weighting each subject's contribution to the estimating equation by the inverse probability that a subject drops out at the j*th* time.

Finally, multiple imputation (MI) as a strategy is also valid under MAR assumptions. It involves filling the missing values M times to obtain M complete data sets. The M data sets are then analyzed using standard procedures. Finally the results from the M data sets, that is, parameter estimates and standard errors are combined into a single parameter estimate for each covariate of interest, as well as corresponding standard errors which reflects the uncertainty inherent in the imputation of the unobserved responses (Verbeke and Molenberghs, 2019).

4.4 Modelling Framework

Modelling framework adopted in context of missing data analysis include; selection models, pattern mixture models and shared parameter models. In this analysis, centre of attention was on selection models and pattern mixture models. These two framework have identical joint distribution of the observed data and the missing data indicator given the predictor variables. However, upon factorization of the joint distribution, it disintegrates into distinct connotation with respect to parameter estimates (Molenberghs and Kenward, 2007).

4.4.1 Selection Models

This type of modelling framework has been clearly and explicitly defined by Molenberghs *et al*, 2009. Consideration being made on the response variable Y_{ij}. Suppose these response is a sequence of measurement of subject i, $i = 1, ..., N$ at time t_{ij}, The outcomes can conveniently be grouped into a vector $Y_i = (Y_{i1}, ..., Y_{in_i})$. Besides, the vector of the missingness indicator is defined by $R_i = (R_{i1}, ..., R_{in_i})$, with $R_{ij} = 1$ if Y_{ij} is observed and 0 otherwise. Ideally, one would like to consider the density of full data $f(Y_i, R_i|X_i, \theta, \psi)$, where the parameter vectors θ and ψ describe the measurement and missingness processes, respectively. In selection model, the joint distribution of the ith subject's Y_i and R_i is factored as the marginal distribution of Y_i and the conditional distribution of R_i given Y_i, that is;

$$f(Y_i, R_i|Xi, \theta, \psi) = f(Y_i|X_i, \theta)f(Y_i|Y_i, X_i, \psi)$$

where X_i is a measured vector of covariates.

After factorization of the joint distribution, the missing at random (MAR) and missing not at random (MNAR) were the most applicable missing mechanisms for this analysis. For MAR, the probability of a measurement being missing depends only on the observed outcome, say y^0 and/or the covariates. This can be written as;

$$f(Y_i, R_i|X_i, \theta, \psi) = f(Y_i^0|X_i, \theta)f(R_i|Y_i^0, X_i, \psi)$$

The measurement model $f(Y_i^0|X_i, \theta)$ and the dropout model $f(R_i|Y_i^0, X_i, \psi)$ can be fitted separately provided that the parameters (θ, ψ) in both models are functionally disjointed (Verbeke and Molenberghs, 2000).
Sometimes MAR assumption is not fulfilled, then MNAR is assumed where the probability of a measurement being missing depends on unobserved outcome resulting to joint distribution of the observed measurements and missingness data indicator. An elaborate discussion can be found in Verbeke and Molenberghs (2000), Ch. 17.

4.4.2 Pattern mixtures models

Under pattern mixture models, we let Y_{ij} be the jth outcome measurement of the ith patient where $i = 1, ..., 101$ and $j = 0, 6, ..., 42$, R_i be the missingness data indicator for Y_{ij} where $R_i = 1$ if Y_{ij} is observed and zero if Y_{ij} is missing. The joint distribution of the ith subject under the pattern mixture model framework is given by

$$f(Y_i, R_i|X_i, \theta, \psi) = f(Y_i|R_i, X_i, \theta)f(R_i|X_i, \psi)$$

This equation specifies the marginal distribution of r_i and the conditional distribution of y_i given r_i. Pattern mixture models have wide range use in sensitivity analysis. A detailed discussion on pattern mixture models can be found in Verbeke and Molenberghs (2000), Ch. 18.

4.5 Direct Likelihood: Ignorability

As discussed earlier, direct likelihood is the likelihood based way of using all observed observation in the data set. This method is valid under MAR assumption and the mild separability condition of parameter spaces of the observed data and the missingness mechanism (Molenberghs and Kenward, 2007). Ignorability means that inferences about the measurement mechanism are made without addressing the missingness mechanism. When missingness is ignorable, likelihood based inference does not require specification of the missing data mechanism but does require the full specification of joint distribution of the outcome variable and missing data mechanism regardless of the modelling framework (Fitzmaurice *et al*, 2008).
The inference can be based on the marginal observed data density if and only if θ and ψ are disjoint. This technical requirement is referred to as the separability condition (Verbeke and Molenberghs, 2000). Linear mixed model (LMM) which is a likelihood based method was fitted to the data and is valid analysis under MAR mechanism (Molenberghs and Kenward, 2007). As discussed earlier in section 4.3, Beunckens *al* (2005) reported that fitting a LMM will in return be performing direct likelihood approach of handling missing data.
Consider the LMM below,

$$\begin{aligned} Y_{ij} = (\beta_0 + b_{1i}) + \beta_1 age_i + \beta_2 smoke0_i + \beta_3 smoke1_i + \beta_4 time_{ij} + \beta_5 bmi_i + \beta_6 cd0_i \\ +(\beta_7 smoke0_i + \beta_8 smoke1_i + \beta_9 bmi_i + \beta_{10} cd0_i + b_{2i}) * t_{ij} + \epsilon_{ij} \end{aligned} \tag{1}$$

where $i = 1, ..., 101$, Y_{ij} is the diameter of abdominal aorta (mm) for the ith patient at jth time, $(b_{1i}, b_{2i})' \sim$ N(0, D) and $\epsilon_{ij} \sim$ N(0, Σ_i). The error term ϵ_{ij} is assumed to be independently identically distributed with mean of zero and variance σ^2 with an assumed covariance function which is linear over time. REML was used for parameter estimation, different inference for the marginal model. Marginal testing for the need of the random effects is done using mixtures of chi-squares. In practice, REML is usually used as it allows to estimate the covariance parameters without having to estimate mean first. (Verbeke and Molenberghs, 2000).

4.6 Marginal models

Since our interest was to derive inference on the evolution of the diameter of abdominal aorta and factors affecting this development in the population, weighted generalized estimating equations and GEE after multiple imputation were fitted to the data. Consider the model below;

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij})$$

$$\begin{aligned} \log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + \beta_1 age_i + \beta_2 smoke0_i + \beta_3 smoke1_i + \beta_4 time_{ij} + \beta_5 bmi_i + \beta_6 cd_i \\ + (\beta_7 bmi_i + \beta_8 smoke0_i + \beta_9 smoke1_i + \beta_{10} cd_i) * t_{ij} \end{aligned} \tag{2}$$

Where Y_{ij} is the size of the diameter of abdominal aorta of the patient ith at the jth measurement which is assumed to follow a binary distribution with parameter π_{ij} , $i = 1, 2, ..., 101$. t_{ij} is the time point at which the jth measurement is taken from model parameters and the variables

age_i, $smoke0_i$ *and* $smoke1_i$, bmi_i, cd_i respectively represent the characteristics age, Ex smoker and Smoker (smoking status), BMI, Coronary disease (status) for patient i. Since the time points at which the measurements were taken were equally spaced, the only two working correlation structure that are applicable are first-order autoregressive working correlation structure (AR(1)) and exchangeable working correlation structure (Molenberghs and Verbeke, 2005). The AR(1) working correlation was assumed for the analysis because the current measurement may depend on the outcome of the previous measurement.

4.7 Sensitivity Analysis

Sensitivity analysis is a statistical approach where several (non-random) models are fitted to the data simultaneously. The extend to which conclusions (inferences) are stable across such ranges provides an indication about the belief one can put into them. In doing so, the analyst strengthens their confidence in the MAR model (Molenberghs and Verbeke, 2005). According to Molenberghs and Kenward (2007) book, whichever assumptions are made concerning the missing data mechanism cannot be empirically verified using the observed data. Hence, conducting a thorough sensitivity analysis, where a number of distinct analysis under a variety of plausible (different) assumptions for the missingness mechanism are conducted. Sensitivity analysis assesses the departures of inferences from the MAR assumption because the unknown deviations can carry an unnoticed impact on the conclusion made under the MAR assumption.

According to aforementioned discussions, weighted estimating equations, direct likelihood and multiple imputation methods were employed to take care of the incomplete data. The assumptions of ignorability for direct likelihood, are relaxed permitting the use of the pattern mixture models technique for sensitivity analysis. Pattern mixture model fits the models under different identified restrictions (Verbeke and Molenberghs, 2000). For multiple imputation which is valid under MAR, that is, the missingness in the data relies on the observed outcome and/or covariates only.
However, sensitivity analysis by an extension to MNAR assumption that allows relaxation of MAR assumption was considered. In the book by Verbeke and Molenberghs (2000), they have suggested and discussed methods of conducting sensitivity analysis under selection model and pattern mixture models framework. In these analysis methods, adopted was the use of identifying restrictions to identify all parameters in the model.

4.7.1 Using Identifying Restrictions in the Sensitivity Analysis

Sensitivity analysis within the Pattern mixture model was conducted. This allows us to draw valid statistical inferences on the evolution of the diameter of the abdominal aorta over time. The key issue of Pattern mixture models is that they are chronically under-identified and consequently Little (1995) suggested the so-called identifying restrictions to overcome this under-identification. (Verbeke and Molenberghs, 2000).
For whichever pattern in the data, the density of the complete data can be elucidated. For example, in a monotone pattern, the joint density can be factorized into two parts, that is, the density for the observed data and the density that depends on the unobserved data. The identifying restriction is concerned with the later (Molenberghs and Verbeke, 2005).

In this part of the analysis, complete case missing values (CCMV) and neighbouring case missing values (NCMV) were used. CCMV constrain the unidentifiable conditional distribution of the given missing data pattern given the observed data is equated to its counterparts in the completers, that is, missing information is always derived from completers. On the other hand, NCMV constrain the unidentifiable conditional distributions are equated to those in the next group, that is, information is borrowed from the nearest pattern (Verbeke and Molenberghs, 2000). Other identifying restrictions include available case missing value (ACMV) which is equivalent to MAR in the selection model where unidentifiable conditional distributions of the unobserved outcomes given the observed outcomes in a specific pattern is equated to a combined information from all patterns on which the outcomes are observed (Verbeke and Molenberghs, 2000).

It should be pointed out that, identifying restrictions method can only be applied to monotone data, and therefore the data set has to be transformed to monotone missingness pattern first before applying the identifying restrictions. The AAA data set was already monotone hence no need for transformation. The results of the two identifying restrictions are then compared with the linear mixed model under the assumption of ignorability and most importantly, the stability of the results evaluated.

5 Results

5.1 Exploratory data analysis

5.1.1 Individual Profiles

Figure 1 shows that there seems to be much within-subject variability and between-subject variability among the patients. It can also be observed that the patients have a different diameter at the start of the experiment and this also changes with time. This variation suggests that a random intercept and random slope model could be a plausible starting point.

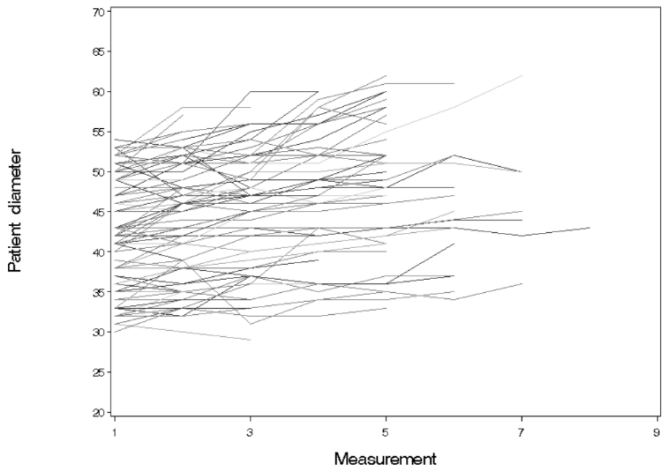

Figure 1: *Individual patients profile*

5.1.2 The Average Evolution

To describe the average evolution of artery diameter (in mm) of patients over time for the overall population, average evolution was plotted and displayed in Figure 2. The average evolution gives the marginal relationship of the response with time. The structure indicates a linear trend with time except for the sharp decrease from the sixth measurement. The standard error bar on the mean also becomes wider. The inflated standard errors bars may be as a result of many dropouts or by the variability of the patients. Although this variation exists, the linear mean model might be appropriate to model the mean structure, but this is subject to a formal test.

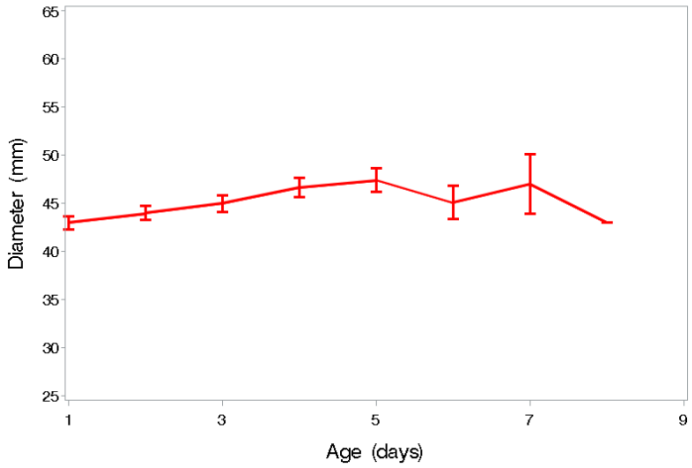

Figure 2: *The mean structure with standard error of the average*

5.1.3 The Variance Structure

After studying the mean structure, the evolution of variance is essential in building a longitudinal model (Verbeke and Molenberghs, 2000). Standardized squared residuals obtained from the mean structure were used to construct the variance function. The figure 3 suggests a stable variance over time with an increase in standard error at each time point; this trend is expected because of the attrition that exists in the data.

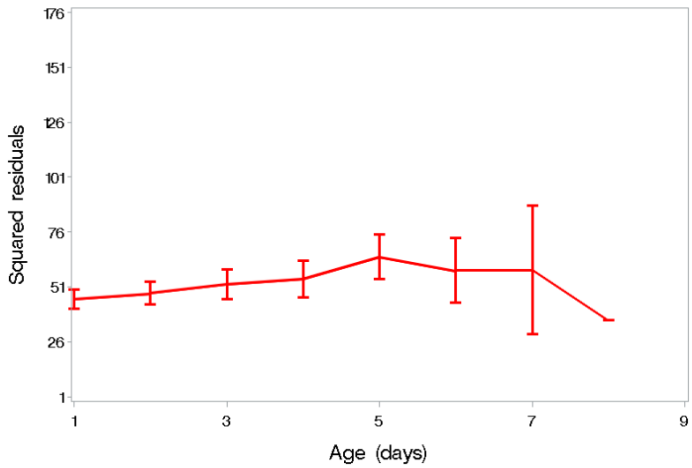

Figure 3: *Variance structure*

5.1.4 The Correlation Structure

Usually, correlation is used to describe how the records of the subjects are correlated. In this study, correlation structure was studied using both the correlation matrix and the scatter plot matrix. But the focus here will be by use of the scatter plot matrix (Figure 4). From this plot, the off-diagonal elements obtained from the pairs of measurement occasions showed the decaying of the correlation with time. Also, the stationarity assumption suggests that the schemes remain within the diagonal bands since the measurements were collected at an equally spaced time interval, that is, measurement 1 for Month 0, measurement 2 for month 6 and the proceeding measurement obtained in that order of sequence.

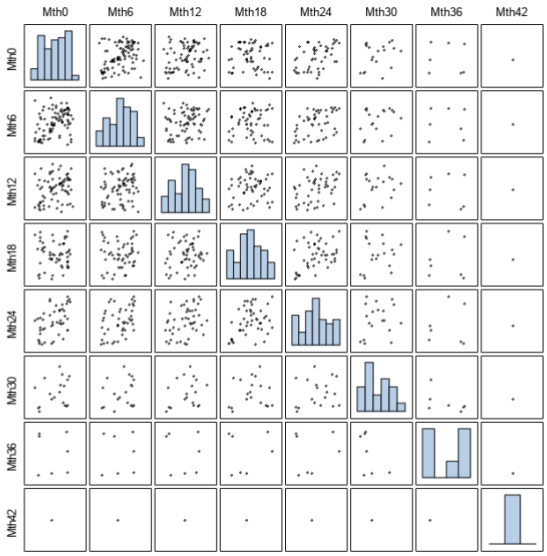

Figure 4: *Correlation structure*

5.1.5 Exploring the missingness patterns

Table 1 summarizes the measurements taken at each follow up time. It can be observed that a total of 101 patients were followed up every six months. It can be observed that the number of patients in the study continues to reduce as time increases, which indicate dropout from the study. Further, only one patient had eight complete measurements. The mean and the variance of the diameter seems to increase with time steadily but reduces on the $6th$ measurement.

Table 1: *General overview on dropouts*

Measurement	Average	Variance	No. of Patients	No. Missing
1	42.97	45.79	101	0
2	43.94	48.61	90	11
3	44.95	52.90	74	27
4	46.62	54.85	53	48
5	47.39	65.31	46	55
6	45.05	56.39	19	82
7	47.00	67.00	7	94
8	43.00	0	1	100

As part of exploratory analysis, assessing the patterns group in the missing data, is the first step that should be looked into. Table 2 shows that approximately 1% of the patient were completers and have all the eight measurements. On the other hand, approximately 5.9% had seven measurements,

roughly 11.9% had 6 measurements. Further, the was 26.7%, 6.9%, 20.8%, 15.8% and 10.9% for 5 measurements through only 1 measurements.

Table 2: *Missingness pattern in the data*

Measurement									
1	2	3	4	5	6	7	8	Number	Percent.
Completers									
O	O	O	O	O	O	O	O	1	0.9
Drop out patterns									
O	O	O	O	O	O	O	M	6	5.9
O	O	O	O	O	O	M	M	12	11.9
O	O	O	O	O	M	M	M	27	26.7
O	O	O	O	M	M	M	M	7	6.9
O	O	O	M	M	M	M	M	21	20.8
O	O	M	M	M	M	M	M	16	15.8
O	M	M	M	M	M	M	M	11	10.9

O represents observe, M represents missing

5.2 Analysis of continuous response

5.2.1 Direct Likelihood-LMM

Based on the aforementioned theory on direct likelihood, we saw that linear mixed models correspond to direct likelihood under MAR. For this analysis LMM was fitted to determine whether several patient characteristics indeed affect the diameter of the abdominal aorta while taking into account the missing observations. This model contained random intercept and random slope, the interaction of time and other covariates was fitted. The use of random effects was based on the exploration of individual profiles. It was evident that at the beginning of the study, diameter of the aortic artery varied between patients. Similarly, the evolution between the patients was different over time. This was done because random effects in LMM represent variability in subject-specific intercepts and slopes not explained by the covariates in the model. Table 4 summarizes the result of the fitted model. Also, the effect of age and the interaction of smoke with time significantly affect the diameter of the abdominal aorta in patients with AAA.

Table 3: *Solutions for fixed effects: Linear Mixed Model*

Effect	Estimate	SE	t	P-value
Intercept	22.6469	9.4082	2.41	0.0180
age	0.1988	0.0945	2.1	0.0367
smoke0	-1.2239	2.1434	-0.57	0.5687
smoke1	0.3976	0.7279	0.55	0.5855
bmi	0.2454	0.1623	1.51	0.1321
cdisease0	-0.9990	1.3339	-0.75	0.4548
time	0.1832	0.1512	1.21	0.2289
time*smoke0	0.0697	0.0502	1.39	0.1667
time*smoke1	-0.0502	0.0411	-1.22	0.2239
bmi*time	0.0017	0.0059	0.29	0.7701
time*cdisease0	-0.0258	0.0419	-0.62	0.5377

Table 4: *Type III test for Fixed Effect*

	D.o.F			
Effect	Num.	Den.	F Value	p-value
age	1	194	4.43	0.0367
smoke	2	194	0.35	0.7023
bmi	1	194	2.29	0.1321
cdisease	1	194	0.56	0.4548
time	1	87	1.35	0.2484
time*smoke	2	194	4.15	0.0173
bmi*time	1	194	0.09	0.7701
time*cdisease	1	194	0.38	0.5377

The random effects in a LMM represent the variability in subject specific intercepts and slopes that are not explained by the fixed covariates (Verbeke and Molenberghs, 2000). Consequently, in

the fitted LMM, the random effects were included. As discussed earlier, mixture of chi-square test was conducted. In doing so, we are able to establish whether the evolution of the diameter of the abdominal patients over time is significantly different either at the beginning of the study, and/or during the study. This test is simply testing the need of random effects in the model. Ideally, mixture of chi-square test mixture chi-square test is often used when the interested effect is on the boundary of the parameter space. A mixture of $0.5\chi^2_1 + 0.5\chi^2_2$ was used. This can be viewed as weighted average of probability distribution with positive weights that sum to one (Verbeke and Molenberghs, 2005)

The hypothesis of interest in the first case was $H_0 : d_{12} = d_{22} = 0$ where d_{12} was the covariance between random slope and random intercept, and d_{22} was the variance of random slope. The obtained p-value is statistically significant at 5%. Hence, the null hypothesis was rejected, and we concluded that the model with both random slope and intercept was appropriate. For the next model comparison, the other null hypothesis of interest was $H_0 : d_{11} = d_{12} = 0$ where d_{11} was the variance of random Intercept. Interestingly, both hypotheses lead to the same conclusion with $\chi^2_{1:2}$ resulting into p $< .0001$, and therefore, it was necessary to include both random slope and random intercept in the final LMM model (see Table 6).

Table 5: *Summary of Variance components of different random effect models*

Effect		LMM Model
Covariance of b_1		Estimate (SE)
Var(b_{1i})	d_{11}	42.8943 (6.5597)
Var(b_{2i})	d_{22}	0.0225 (0.0063)
Cov(b_{1i}, b_{2i})	d_{12}= d_{21}	0.3697 (0.1482)
Residual variance		
Var (ϵ_{ij})	σ^2	2.3029 (0.4402)
ReML		1921.3

Table 6: *Test for random intercept and slope*

Random effects	ReML
Model 1: Intercepts, time (slope)	1921.3
Model 2: time (slope)	1948.5
Model 3: Intercepts	1946.8
-2ln(Model2:1)	27.2
-2ln(Model3:1)	25.5
p-value (Both tests)	<.0001

5.3 Analysis of Binary Outcome

As pointed earlier, weighted estimation equations was employed to the binary outcome. Specifically, weighted GEE which is valid under MAR was used. Table 7 summarizes the parameter estimates, standard errors and the corresponding p-values. The parameter estimates, standard errors with related p-values of the missingness model based on weighted GEE were reported in Table 8. From the missingness model, it was discovered that the missingness of data does not depend on the previous outcome (p=0.1130). Additionally, no patient characteristics was noted to have influence on the missing observation, since all the effects were not statistically significant at 5% significance level (all $p > 0.05$). From Table 7, it can be observed that, time and interaction of BMI with time has significant influence on the probability of having a wider diameter above the median diameter of the abdominal artery with p=0.0084 and p=0.0424 respectively. Other patient characteristics, such as age, smoking behaviour, coronary disease with their interaction with time proved to have no significant effect on the probability of having wider diameter of abdominal aorta above the median value (all $p > 0.05$ level of significance).

Table 7: *Parameter estimates, Standard Errors (S.E) an p-values for WGEE*

Parameter	Estimate	S.E	Z	p-value
Intercept	4.585	2.9594	1.55	0.1213
age	-0.0428	0.0279	-1.53	0.1252
smoke0	-0.5208	0.8243	-0.63	0.5275
smoke1	-0.0824	0.5521	-0.15	0.8814
bmi	-0.0481	0.0632	-0.76	0.4464
time	-1.4541	0.5518	-2.64	0.0084
cdisease0	0.2038	0.4968	0.41	0.6817
time*smoke0	0.1132	0.2961	0.38	0.7022
time*smoke1	0.1273	0.2171	0.59	0.5575
bmi*time	0.0429	0.0221	1.94	0.0424
time*cdisease	0.1002	0.2023	0.5	0.6204

Table 8: *Weighted GEE Missingness model*

Parameter	Estimate	S.E	Z	p-value
Intercept	0.0816	1.7466	0.05	0.9627
diamprev	-0.3921	0.2474	-1.58	0.1130
age	0.0158	0.0177	0.89	0.3709
smoke0	0.2817	0.4615	0.61	0.5416
smoke1	0.1241	0.2695	0.46	0.6453
bmi	0.0073	0.0335	0.22	0.8269
cdisease0	-0.2901	0.2468	-1.18	0.2398

5.4 Multiple Imputation

Multiple imputation was performed to the data set to fill in the missing observations. The outcome variable, together with BMI which was time varying predictor variable were imputed. The number of imputation was chosen arbitrary to be M=10. For stability check, the number was increased sequentially to M=20 and we conclude that indeed M=10 imputation is still a good choice. After imputation, inference was done by averaging the parameter estimates, while the standard errors were accounting both using the within and between subject variability as discussed earlier. Table 9 summarizes the final resulting output, that is, the parameter estimates, standard error (S.E) and corresponding p-values. From the results no variable or any interaction with time were found to be significant at 5% level of significance.

Table 9: *Parameter estimates, standard error(S.E), p-values: MI-GEE*

Parameter	Estimate	S.E	Z	p-value
Intercept	-0.5548	1.7165	-0.32	0.7466
age	-0.0193	0.0126	-1.54	0.1260
smoke0	-0.3382	0.7149	-0.47	0.6362
smoke1	-0.0705	0.3912	-0.18	0.8569
bmi	0.0814	0.0511	1.59	0.1117
time	-0.0586	0.4366	-0.13	0.8939
cdisease0	0.2084	0.3707	0.56	0.5742
smoke0*time	-0.0004	0.1849	-0.00	0.9985
smoke1*time	-0.0254	0.1016	-0.25	0.8031
bmi*time	-0.0028	0.0162	-0.17	0.8663
time*cdisease0	-0.0088	0.0999	-0.09	0.9298

The results shows that all of the patient characteristics except BMI and history of coronary disease had a negative influence on the probability. However, none of these covariates were found to be significantly influencing the probability of having abdominal aorta greater than 45mm. The p-values for the covariates were all greater than 0.05 level of significance.

5.5 Sensitivity analysis

It was of interest to observe how inference was affected by the MAR assumption, hence an MNAR multiple imputation approach was also considered for the model (1) above. The sensitivity analysis was performed using 2 different approach, first Neighbouring Case Missing Value (NCMV), and Complete Case Missing Value (CCMV).

Results from the analysis shows that inference seems to be consistent for direct likelihood, NCMV, and CCMV. Age was significant for direct likelihood and CCMV but not significant for NCMV. These findings therefore show that age appears to be sensitive to the choice of assumed missing data mechanism, since the inference changes under MAR and MNAR.

Table 10: *Sensitivity Analysis Diameter as an outcome Variable*

	NCMV			**CCMV**		
Parameter	Estimate	Std Error	P-value	Estimate	Std Error	P-value
intercept	23.8089	36.7708	0.5207	26.9506	31.0053	0.3954
age	0.2797	0.2224	0.2111	0.4373	0.1647	0.0114
bmi	0.1676	1.4069	0.9064	-0.473	0.9524	0.6259
time	12.6747	18.5295	0.5052	-13.4168	11.4274	0.2574
smoking0	4.5672	11.6058	0.6939	12.1253	10.0225	0.2264
smoking1	-0.1589	6.5298	0.9806	4.0135	5.6347	0.4764
cdisease0	-3.3484	6.1168	0.5841	-4.6323	5.2441	0.3771
time*smoking0	-5.9733	4.9481	0.2274	-8.0896	5.2445	0.1230
time*smoking1	-2.8381	2.8136	0.3133	-4.2900	2.9794	0.1504
bmi*time	-0.4710	0.7095	0.5182	0.5224	0.4320	0.2464
time*cdisease0	1.8199	2.6365	0.4901	1.4311	2.8840	0.6200

6 Discussion and Conclusion

In this study, the diameter of the AAA data was analyzed. The outcome variables of interest was the diameter of the abdominal aorta (in mm). The analysis recognizes and handles the missingness in the data. The models fitted to this data using the linear mixed model for the continuous response and generalized estimating equations for a binary outcome. The selection model and pattern mixture framework were used as a way of handling the missingness. The results for MI-GEE model indicated that there is a significant effect of time and interaction of time and BMI on the probability of having a diameter of abdominal aorta greater than or just equal to the median diameter (in cm). For the direct likelihood-LMM, the results indicated that there is a significant effect of age on the evolution of the diameter of the abdominal aorta over time. Further, with regards to the drop out model, there was no direct link between the previous outcome to the drop out from the trial which may suggest that the missingness in the data is MCAR. However, caution should be applied here because this might be attributed to chance.

As proposed by Molenberghs and Kenward (2007), there are several application of several model approaches that allow for missingness to be MNAR. In order to check for the varying assumption about the missing data mechanism, the pattern mixture framework using NCMV and CCMV were used in this analysis. This was performed only for the continuous response and result showed that

inference were consistent with direct likelihood approach although age appears to be sensitive to the choice of assumed missing data mechanism. Literature reveals a number of patient characteristics used in these analyses showing significant effect on the evolution of the aortic artery which is one of the effect that facilitates development of abdominal aortic aneurysm (AAA). The paper by Normal *et al* (2013), discussed that smoking is causally associated with AAA. In other older research by Willick *et al* (1999) indicated that the duration of exposure to smoking rather than the level of exposure appeared to determine the risk of the development of an AAA especially in men older than 50 years. Similarly, Umebayashi *et al* (2018) discussed that the older population are predisposed to the risk of developing AAA. Aging is one of the dominating factors that is associating with AAA. Also, history of coronary disease (CAD) has been found to have high prevalent among AAA patients, for example among ruptured AAA, all CAD patients were above 60 years old; 80% had AAA diameter 5.5 cm. However, with regards to the results of these analyses, patient characteristics aside from age showed no significant effect on the evolution of the aortic artery. In conclusion, this contradictory result can be due to the choice of assumed missing data mechanism.

7 Recommendations

As a recommendation, future studies should include the gender of the subjects participating in the study since AAA may be affecting individual of the particular gender and the evolution between the different gender might be ultimately different. Furthermore, if the same study is planned, the interval between the two study visits should be reduced to mitigate the high number of dropouts.

References

[1] Aggarwal, S., Qamar, A., Sharma, V., Sharma, A. (2011). Abdominal aortic aneurysm: a comprehensive review. Experimental Clinical Cardiology, 16(1), 11.y.

[2] Beunckens, C., Molenberghs, G., Kenward, M. G. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. Clinical Trials, 2(5), 379-386.

[3] European Medicines Agency. (2010). Guideline on missing data in confirmatory clinical trials.

[4] Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2009). Longitudinal data

[5] Gałecki, A., Burzykowski, T. (2013). Linear mixed-effects models using R: A step-by-step approach. Springer Science Business Media.

[6] Izenman A.J. (2013) Multivariate Regression. In: Modern Multivariate Statistical Techniques. Springer Texts in Statistics. Springer, New York, NY

[7] Jansen, I., Beunckens, C., Molenberghs, G., Verbeke, G., Mallinckrodt, C. (2006). Analyzing incomplete discrete longitudinal clinical trial data. Statistical Science, 21(1), 52-69.

[8] Kaushal S. (2014). Missing data in clinical trials: Pitfalls and remedies. International journal of applied basic medical research, 4(Suppl 1), S6–S7.

[9] Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. Journal of the american statistical association, 90(431), 1112-1121.

[10] Molenberghs, G., Kenward, M. (2007). Missing data in clinical studies (Vol. 61). John Wiley Sons.

[11] Molenberghs G, Verbeke G. (2005). Models for Discrete Longitudinal Data, 1 edn.New York: Springer-Verlag New York.

[12] Robert E. Weiss. (2005). Modeling Longitudinal Data-Springer-Verlag New York.

[13] Norman, P. E., Curci, J. A. (2013). Understanding the effects of tobacco smoke on the pathogenesis of aortic aneurysm. Arteriosclerosis, thrombosis, and vascular biology, 33(7), 1473–1477. doi:10.1161/ATVBAHA.112.300158

[14] Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592.

[15] Umebayashi, R., Uchida, H. A., Wada, J. (2018). Abdominal aortic aneurysm in aged population. Aging, 10(12), 3650–3651. doi:10.18632/aging.101702

[16] Verbeke G. and Molenberghs G. (2009). Linear Mixed Models for Longitudinal Data.

[17] Verbeke, G. and Molenberghs, G. (2019), Introduction to Longitudinal Data Analysis: Lecture Notes, Hasselt University.

[18] Vorp, D. A., Geest, J. P. V. (2005). Biomechanical determinants of abdominal aortic aneurysm rupture. Arteriosclerosis, thrombosis, and vascular biology.

[19] Wilmink TB, Quick CR, Day NE,J Vasc Surg. 1999: The association between cigarette smoking and abdominal aortic aneurysms.

8 Codes

```
/*Exploratory data analysis*/
/*****individual plot****/;
goptions reset=all ftext=swiss device=psepsf gsfname=fig1
gsfmode=replace rotate=landscape i=join;
proc gplot data=lda.aaa_1;
plot diameter*measurement=patient / haxis=axis1 vaxis=axis2
nolegend;
axis1 label=(h=2 'Measurement') value=(h=1.3) order=(1 to 10 by 2)
minor=none;
axis2 label=(h=2 A=90 'Patient diameter') value=(h=1.3) order=
(20 to 70 by 5)
minor=none;
title h=2 'Individual profiles';
run;quit;

/****mean structure plot****/
goptions reset=all;
proc gplot data=lda.aaa_2;
plot diameter*measurement=smoke / haxis=axis1 vaxis=axis2
legend=legend1;
symbol1 i=std1mjt w=2 color=red;
axis1 label=(h=1.5 'measurement') value=(h=1.5) order=(1 to 10 by
3)
minor=none;
axis2 label=(h=1.5 A=90 'Diameter (in mm)') value=(h=1.5) order=(20
to 90 by 10)
minor=none;
legend1 label=none position=(top left inside) offset=(5,-3)
value=(H=1.3
"Non smokers" "Ex-smokers" "Smokers") frame down=3;
title h=1.5 'Average evolution'; run;

/* Direct likelihood */

proc mixed data=lda.aaa_3a method=reml nobound empirical covtest;
class patient cdisease smoke timeclass;
model diameter = age smoke bmi cdisease time smoke*time bmi*time
cdisease*time/  s chisq;
random intercept time / type=un subject=patient g gcorr v vcorr
solution ;
repeated timeclass / type=ar(1) subject=patient r rcorr;
run;
```

```
/* Weighted GEE */
/* Categorization */

data lda.miss2;
set lda.miss2;
if diameter >= 45 then diam=1;
else if diameter =. then diam=.;
else diam=0;
diamprev=lag(diam);
timeClass=measurement;
time=measurement;
run;

/*weighted GEE*/
proc gee data=lda.miss2;
class patient smoke cdisease timeClass;
model diam = age smoke bmi time cdisease smoke*time bmi*time
cdisease*time / dist=bin;
repeated subject=patient / withinsubject=timeClass type=ar(1) covb
modelse;
missmodel diamprev age smoke bmi cdisease /type=obslevel;
ods output GEEEmpPEst=lda.gmparms parminfo=lda.gmpinfo
CovB=lda.gmcovb;
run;

/* IMPUTATION */
proc mi data = lda_hor out = lda_hor_a seed = 1 simple nimpute = 10
round = 0.1;
var age smoking cdisease bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7
diameter_1
diameter_2 diameter_3 diameter_4 diameter_5 diameter_6 diameter_7;
freq _FREQ_;
run;

/*Transforming horizontal dataset back to vertical format:*/
data lda_imputed (keep = _imputation_ patient age smoking cdisease
time  bmi diameter);
set lda_hor_a;
array r (7) bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7;
array s (7) diameter_1 diameter_2 diameter_3 diameter_4 diameter_5
diameter_6 diameter_7;
do j = 1 to 7;
time = j;
bmi = r(j);
```

```
diameter = s(j);
timeclss = time;
output;
end;
run;

/* Creating monontone */;
proc mi data=lda_hor seed=17200705 simple out=monotone nimpute=10
round=0.1;
freq _FREQ_;
title "Monotone imputation";
var age smoking cdisease bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7
diameter_1
diameter_2 diameter_3 diameter_4 diameter_5 diameter_6 diameter_7;
mcmc impute=monotone;
run;

/* Manipulation */
data lda.mi21;
set lda.mi2;
if diameter >= 45 then diam=1;
else diam=0;
if diameter eq . then diam=.;
timeClass=time;
cdisease0=0; if cdisease=0 then cdisease0=1;
cdisease1=0; if cdisease=1 then cdisease1=1;
smoke0=0; if smoking=0 then smoke0=1;
smoke1=0; if smoking=1 then smoke1=1;
smoke2=0; if smoking=2 then smoke2=1;
run;

proc genmod data=lda.mi21;
class patient smoking cdisease timeClass;
by _imputation_;
model diam = age smoke0 smoke1 smoke2 bmi time cdisease0 cdisease1 smoke0*time
    smoke1*time smoke2*time bmi*time  cdisease0*time
    cdisease1*time /covb dist=binomial;
repeated subject=patient / withinsubject=timeClass type=ar(1)
modelse;
ods output GEEEmpPEst=lda.gmparms parminfo=lda.gmpinfo
CovB=lda.gmcovb1;
run;

/*Deleting redundant parameters*/
```

```
data lda.Gmpinfo;
set lda.Gmpinfo;
if parameter="Prm1" then delete;
if parameter="Prm9" then delete;
if parameter="Prm12" then delete;
if parameter="Prm15" then delete;
run;
proc mianalyze parms=lda.gmparms covb=lda.gmcovb1
parminfo=lda.gmpinfo wcov bcov tcov;
modeleffects Intercept age smoke0 smoke1 bmi time  cdisease0
smoke0*time smoke1*time
bmi*time cdisease0*time ;
run;

/*#QUESTION 3: SENSITIVITY ANALYSIS FOR THE CONTINUOUS OUTCOME#*/
/** NCMV restriction mechanism **/
*** Creating monontone ***;
proc mi data=lda_hor seed=17200705 simple out=monotone nimpute=10
round=0.1;
freq _FREQ_;
title "Monotone imputation";
var age smoking cdisease bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7
diameter_1 diameter_2 diameter_3 diameter_4 diameter_5 diameter_6
diameter_7;
mcmc impute=monotone;
run;
proc print data = monotone;
run;

/*Imputing for NCMV */
proc mi data = monotone seed = 17200705 simple out = monotone_ncmv
nimpute = 1;
freq _FREQ_;
title 'Model multiple imputation';
var age smoking cdisease bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7
diameter_1 diameter_2 diameter_3 diameter_4 diameter_5 diameter_6
diameter_7;
monotone reg;
mnar model (bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7 diameter_1
diameter_2 diameter_3 diameter_4 diameter_5 diameter_6 diameter_7 /
modelobs = ncmv);
run;

/*Transforming horizontal dataset back to vertical format:*/
data lda_ncmv (keep = _imputation_ patient age smoking cdisease time
```

```
timeclss bmi diameter);
set monotone_ncmv;
array p (7) bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7;
array q (7) diameter_1 diameter_2 diameter_3 diameter_4 diameter_5
diameter_6 diameter_7;
do j = 1 to 7;
time = j;
bmi = p(j);
diameter = q(j);
timeclss = time;
output;
end;
run;


/*creating dummies for smoking and cdisease*/
data lda_ncmv_2;
set lda_ncmv;
cdisease0 = 0; if cdisease = 0 then cdisease0 = 1;
cdisease1 = 0; if cdisease = 1 then cdisease1 = 1;
smoking0 = 0; if smoking = 0 then smoking0 = 1;
smoking1 = 0; if smoking = 1 then smoking1 = 1;
smoking2 = 0; if smoking = 2 then smoking2 = 1;
run;


/* Fitting the pmm under NCMV */
proc mixed data = lda_ncmv_2 method = REML noclprint noitprint
asycov covtest;
title 'Model per pattern';
class patient cdisease smoking timeclss;
by _Imputation_;
model diameter = age bmi time smoking0 smoking1 smoking2 cdisease0
cdisease1 smoking0*time smoking1*time smoking2*time
    bmi*time age*time cdisease0*time cdisease1*time/ s
    chisq ;
random intercept time / type = un subject = patient g gcorr v vcorr solution ;
repeated timeclss / type = ar(1) subject = patient r rcorr;
ods output SolutionF = mixparms;
run;


/*Combining results of ncmv*/
proc mianalyze parms = mixparms ;
modeleffects intercept age bmi time smoking0 smoking1 smoking2
```

```
cdisease0 cdisease1 smoking0*time smoking1*time smoking2*time
    bmi*time age*time cdisease0*time cdisease1*time;
run;

/** CCMV restriction mechanism **/
*** Creating monontone ***;
proc mi data=lda_hor seed=17200705 simple out=monotone nimpute=10
round=0.1;
freq _FREQ_;
title "Monotone imputation";
var age smoking cdisease bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7
diameter_1 diameter_2 diameter_3 diameter_4 diameter_5 diameter_6
diameter_7;
mcmc impute=monotone;
run;
proc print data = monotone;
run;


/*Imputing for CCMV */
proc mi data = monotone seed = 17200705 simple out = monotone_ccmv nimpute = 1;
freq _FREQ_;
title 'Model multiple imputation';
var age smoking cdisease bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7
diameter_1 diameter_2 diameter_3 diameter_4 diameter_5 diameter_6
diameter_7;
monotone reg;
mnar model (bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7 diameter_1
diameter_2 diameter_3 diameter_4 diameter_5 diameter_6 diameter_7 /
modelobs = ccmv);
run;

proc print data = monotone_ccmv;
run;

/*Transforming horizontal dataset back to vertical format:*/
data lda_ccmv (keep = _imputation_ patient age smoking cdisease time
timeclss bmi diameter);
set monotone_ccmv;
array p (7) bmi_1 bmi_2 bmi_3 bmi_4 bmi_5 bmi_6 bmi_7;
array q (7) diameter_1 diameter_2 diameter_3 diameter_4 diameter_5
diameter_6 diameter_7;
do j = 1 to 7;
time = j;
bmi = p(j);
```

```
diameter = q(j);
timeclss = time;
output;
end;
run;


/*creating dummies for smoking and cdisease*/
data lda_ccmv_2;
set lda_ccmv;
cdisease0 = 0; if cdisease = 0 then cdisease0 = 1;
cdisease1 = 0; if cdisease = 1 then cdisease1 = 1;
smoking0 = 0; if smoking = 0 then smoking0 = 1;
smoking1 = 0; if smoking = 1 then smoking1 = 1;
smoking2 = 0; if smoking = 2 then smoking2 = 1;
run;


/* Fitting the pmm under CCMV */
proc mixed data = lda_ccmv_2 method = REML noclprint noitprint
asycov covtest;
title 'Model per pattern';
class patient cdisease smoking timeclss;
by _Imputation_;
model diameter = age bmi time smoking0 smoking1 smoking2 cdisease0
cdisease1 smoking0*time smoking1*time smoking2*time
    bmi*time age*time cdisease0*time cdisease1*time/ s
    chisq ;
random intercept time / type = un subject = patient g gcorr v vcorr
solution ;
repeated timeclss / type = ar(1) subject = patient r rcorr;
ods output SolutionF = mixparms_a;
run;

/*Combining results of ncmv*/
proc mianalyze parms = mixparms_a ;
modeleffects intercept age bmi time smoking0 smoking1 smoking2
cdisease0 cdisease1 smoking0*time smoking1*time smoking2*time
    bmi*time age*time cdisease0*time cdisease1*time;
run;
```