Longitudinal Data Analysis
Assignment II

An investigation to study the probability of developing abdominal aortic aneurysm (AAA) in patients with enlarged abdominal aorta diameter.

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 is 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).
Methodology: A transitional model, Marginal model with generalized estimating equations (GEE) and generalized linear mixed model (GLMM) were the techniques used to determine the factors that affect the diameter of the patients.
Results: Findings from the study showed no significant influence of time and patient characteristics on the evolution of the diameter of aortic artery nor the probability of having wider diameter of aortic artery than the median value. Nonetheless, there was a strong dependence of the of measurements on the previous measurements.
Conclusions: The findings were in contrary with a number of studies reviewed, most of which reported an association of smoking with AAA. Control on the high number of dropouts by making the interval between measurements shorter would see more insightful results.

Keywords : *Abdominal Aortic Aneurysm (AAA), Covariance Structure, Generalized Estimating Equation (GEE), Generalized Linear Mixed Model (GLMM), Transition models*

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. AAA 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 in 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).

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

The questions of interest for the study are subdivided into marginal, transition and subject specific.

- To determine whether the relationship between the patient characteristics and the aortic diameter of the patient evolve differently over time (months)
- To determine whether the current aortic diameter of the patient depend on previous aortic diameter of the patient while adjusting for patients characteristics?
- To asses how much does the subject specific evolution of aortic diameter of the patient differ from the average evolution while adjusting for patients characteristics

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 patient's 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 binary version of the artery

diameter of the patients, that is;

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

Furthermore, all the predictors except patient weight and height were used in the analysis. This decision was fashioned by the fact 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

Before any further investigation, an exploratory analysis was conducted on the data to data structure and spot any existing anomalies and to obtain clear insight on conceivable implication before model building and hypothesis testing. In this study, descriptive statistics were used to explore the data, that is, graphs and summary tables were used in the section.

4.2 Marginal Models

Marginal models (*population averaged models*) are extensions of generalized linear models for longitudinal data which incorporate the within subject association to model the dependence of repeated measurements of the same individual. These models evaluate mean response using the appropriate link function and describe how the mean response in the population changes over time and its dependence on covariates (Fitzmaurice *etal*., 2008).

Molenberghs and Verbeke (2005) and other different authors, discuss several different marginal model families for both full likelihood and quasi-likelihood based marginal models. Full likelihood models included Bahadur model, Dale Model and Multivariate probit model whilst Generalized estimating equations (GEE) model family was classified as direct extension of the quasi-likelihood based method. However, when the marginal interpretation of the results from analysis of longitudinal data is of interest a full likelihood approach is commonly used because of their efficiency. Nonetheless, there are risks attributed to these methods, especially, when mispecification of the likelihood functions is made. Full likelihood models also are computationally expensive and require excessive time when the number of repeated measurements becomes enormous(Aerts *et al*., 2002). Moreover, some models based on full likelihood suffer from severe restrictions on the parameter space (Molenberghs and Verbeke, 2005).

Based on these reasons, GEE1 from a family of generalized estimating equations (GEE) was proposed especially for this analysis of the correlated measurements, that is to answer the first objective. This approach only requires correct specification of the univariate mean structure and a working assumption about the association structures between pairs of outcomes. This makes GEE1 a valid and appropriate method in case the first-order marginal mean parameters and otherwise, pairwise interactions is of interest(Molenberghs and Verbeke, 2005).

The major advantage of the GEE is that it yields consistent estimates of the model parameters even when the within-subject association among the repeated measurements is mispecified (Fitzmaurice *etal.*, 2008). Also, there is little loss of precision and this makes it efficiently preferred over the full-likelihood approaches. However, severe mispecification may affect the efficiency of estimation (Molenberghs and Verbeke, 2005).
Relying on the theory above, GEE1 was fitted in order to study the effect of treatment on the binary outcome of diameter (0: diameter $<$ *median* diameter and 1: diameter $\geq$ *median* diameter.). For subject i measured at the j^{th} time point, the indicator for diameter of abdominal aorta and the probability of diameter are denoted as Y_{ij} and π_{ij} respectively. Thus, the odds of patient having diameter greater than the median is $\frac{\pi_{ij}}{1-\pi_{ij}}$. For a given time the patient characteristics, the marginal model is given by:

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

$$\begin{aligned}\log(\frac{\pi_{ij}}{1-\pi_{ij}}) &= \beta_0 + \beta_1 Age_i + \beta_2 exsmok_i + \beta_3 smok_i + \beta_4 bmi_i + \beta_5 cd_i \\ &\quad + (\beta_6 + \beta_7 bmi_i + \beta_7 exsmok_i + \beta_8 smok_i + \beta_9 cd_i)t_{ij}\end{aligned} \tag{1}$$

Where Y_{ij} is the size of the diameter of aortic artery of the patient i^{th} at the j^{th} 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 j^{th} measurement is taken f odel parameters and the variables Age_i, $exsmok_i$ *and* $smok_i$, 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 the measurements were take is 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.

4.3 Transition Models

The transition model is a conditionally specified model, in which, hierarchical longitudinal correlated data is modelled conditional to set of previous outcomes. This model was used to answer the second question. To easily and efficiently work with this models, Molenberghs and Verbeke (2005) recommended that in this setting, it is advisable to specify apriori modelling the measurement whether part of previous measurements or all the measurements will be used. Therefore, in this analysis, first-order autoregressive model will be considered.

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

$$\begin{aligned}\log(\frac{\mu_{ij}}{1-\mu_{ij}}) &= \beta_0 + \beta_1 Age_i + \beta_2 exsmok_i + \beta_3 smok_i + \beta_4 bmi_i + \beta_5 cd_i \\ &\quad + (\beta_6 + \beta_7 bmi_i + \beta_7 exsmok_i + \beta_8 smok_i + \beta_9 cd_i)t_{ij} + \alpha_1 y_{i,j-1}\end{aligned} \tag{2}$$

Where Y_{ij} and t_{ij} is as defined in section 3.2 while $\beta_0, \beta_1, \beta_2, ..., \beta_8$ are the regression model parameters. Model 2 is referred to as of stationary first-order autoregressive

type. The transitional probabilities between $j-1$ and j were obtained by evaluating Model 1 to $y_{i,j-1}=1$ and $y_{i,j-1}=0$. In this model, covariates are not attached to the previous outcome, and therefore would be constant across the population.

4.4 Generalized Linear Mixed Models

Generalized linear mixed models (GLMM) are extensions of generalized linear models for longitudinal data that accounts for the within-subject association through incorporation of random effects to the model (Fitzmaurice *etal.*, 2009). GLMM are frequently used as random effects model in the context of discrete repeated measurements. Therefore this model was used to answer the third research question. A random effects model is assumed because individual patients might vary at baseline and also in their evolution over time.

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

$$\begin{aligned}\log(\frac{\pi_{ij}}{1-\pi_{ij}}) &= \beta_0 + b_{1i} + \beta_1 Age_i + \beta_2 exsmok_i + \beta_3 smok_i + \beta_4 bmi_i + \beta_5 cd_i \\ &\quad + (b_{2i} + \beta_6 + \beta_7 bmi_i + \beta_8 exsmok_i + \beta_9 smok_i + \beta_{10} cd_i)t_{ij}\end{aligned} \tag{3}$$

where π_{ij} is the probability probability of a success $i=1,2,..,101; j=1,2,...,8$;
$(b_{0i}, b_{1i}) \sim N(0,D)$ where D=$\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$
Adaptive Gaussian quadrature of the integral which is based on maximum likelihood estimation was used to estimate parameters of the models.

4.5 Testing for the need of random effects

Random effects are important in capturing the variability per subject that has not been explained by the independent variables (Molenberghs and Verbeke, 2005).Testing for random effects was performed in order to determine whether they were needed in the GLMM in modelling the subject specific evolution. A mixture of Chi-square's approach was used as a formal test for inclusion of random slope. Mixture Chi-square test that is used when the interested effect is on the boundary of the parameter space. A mixture of $0.5\chi_1^2+0.5\chi_2^2$ was used. This is viewed as weighted average of probability distribution with positive weights that sum to one (Verbeke and Molenberghs, 2005). The null hypothesis was then given as

$$H_0 : d_{12} = d_{22} = 0$$

$$H_A : \text{Either } d_{12} \text{ or } d_{22} \text{ is } > 0$$

4.6 Empirical Bayes Estimates

Empirical Bayes (EB) estimates shows how the subject specific profile deviates from the overall. Since EB estimates are assumed to be random variables,they are estimated using Bayesian methods where a posterior distribution $f(b_i|y_i,\beta,D,\phi)$ is formulated with prior distribution $f(b_i|D)$. Further, the random effects b_i which is assumed to follow a multivariate normal distribution with mean vector zero

and covariance matrix D, and does not depend on the observed data Y_i, then the posterior mode $\hat{b_i}$ is calculated which corresponds to the EB estimates (Verbeke and Molenberghs, 2005).
The posterior probability of b_i is defined as;

$$f(b_i|y_i, \beta, D, \phi) = \frac{f(y_i|b_i, \beta, \phi)f(b_i|D)}{f(y_i|b_i, \beta, \phi)f(b_i|D)db_i}$$

where $f(y_i|b_i, \beta, \phi)$ is the conditional distribution of observed data Y_i given the random effectsb_i and (β, ϕ) and β are the marginal parameter estimates. The posterior probability in this case is not of a normal form implying that posterior mode is preferred to posterior mean as a point estimator of b_i. This point estimator is chosen as the value for b_i that maximizes $f_i(b_i|y_i, \beta, D, \phi)$, therefore, since the estimator $\hat{b}$ of b_1 depends on the unknown parameters their estimates should be replaced by the maximum likelihood estimates (Verbeke and Molenberghs, 2005).
In this analysis, we estimated EB estimates and histogram representation was plotted.

4.7 Software

Both the R statistical software and SAS v9.4 were used in the analysis

5 Results

In this section we presents the results of the exploratory data analysis and the results to research questions.

5.1 Exploratory Data Analysis

Table 1: *General Overview*

Measurement	Proportions	No.Patients
1	0.46	101
2	0.52	90
3	0.60	73
4	0.62	53
5	0.63	46
6	0.47	19
7	0.57	7
8	0.00	1

A total of 101 patients were followed up every six months. The table 1 below shows the proportion and number of patients present at each time point. From the table, it can be observed that the number of patients in the study continues to reduce as time increases, which indicate dropout from the study. Also, only one patient had eight complete measurements. The proportion of the diameter seems to increase with time steadily but reduces on the 6^{th} measurement.

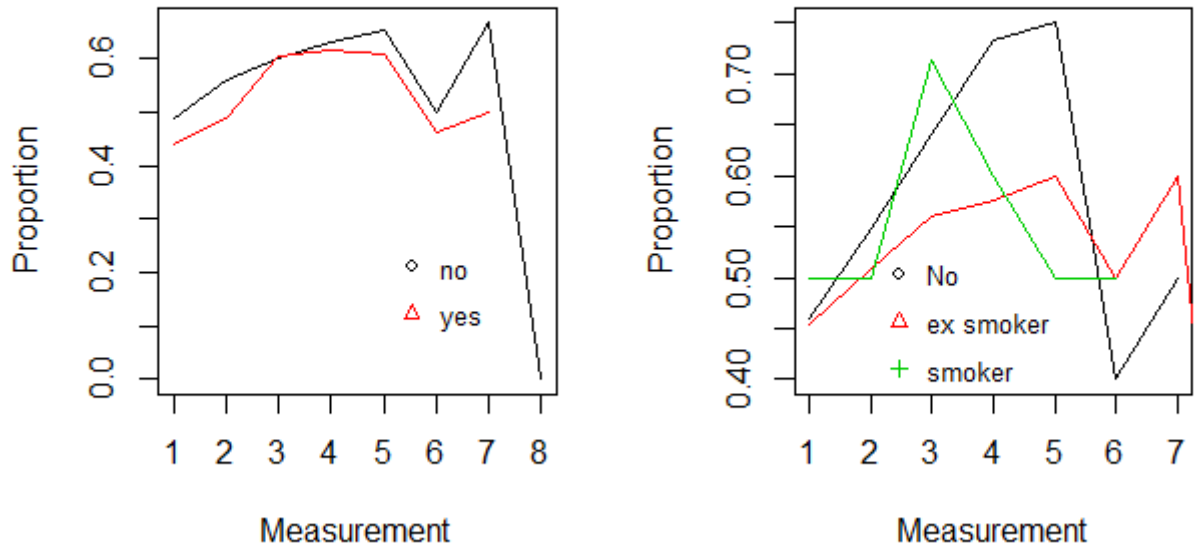

Figure 1: *Proportion plot of diameter of aortic adjusted for Coronary Disease (left) and Smoking(right)*

The figure 1 shows the plot of the proportion of the aortic diameter of the patients adjusting for whether patients have coronary disease (left) or not and the smoking status (right) of the patients. There seems to be little difference between patients that have and do not have coronary disease while no clear pattern as regards the smoking status of patients.

5.2 Transition model

The fit of the first-order stationary autoregressive model is given in Table 2. It is clear that there is very strong dependence on the previous measurement. Also, further analysis showed that there was no complete dependence on the second previous measurement (*Table not reported*).

Table 2: *Maximum likelihood parameter estimates summary for a first-order stationary autoregressive model*

Effect	Par.	Estimate	SE	P-value
Intercept	β_0	-3.6679	2.9486	0.2135
Age	β_1	0.0326	0.0213	0.1250
Ex-Smokers	β_2	-0.6700	1.0991	0.5422
Smokers	β_3	-0.5313	1.1389	0.6408
BMI	β_4	-0.0223	0.0818	0.7855
CD	β_5	0.5038	0.5812	0.3860
time	β_6	0.1068	0.7628	0.8886
Ex-Smokers*time	β_7	0.1070	0.3284	0.7445
Smokers*time	β_8	0.1506	0.3492	0.6663
CD*time	β_9	-0.1168	0.1774	0.5103
BMI*time	β_{10}	0.0021	0.0262	0.9346
Dep. on $Y_{i,j-1}$	α	3.4913	0.2974	0.0001

Moreover, age, smoking behaviour, Body Mass Index, indication of coronary disease and time were statistically non-significant with p-values greater than 0.05. Similarly, the interaction between these covariates with time were not statistically significant with respect to the level of significance.

5.3 Generalized Estimating Equations

Table 4 summarizes the parameter estimates for the fitted GEE. Both parameter estimates from the model based and empirical approach coincide and hence model based estimates not provided. It can be observed that the standard error from the empirical approach are inflated compared to the model-based ones. This is due to the correlation between the repeated measurements that has been accounted for by use of the empirical approach. Therefore, this implies that ignoring the correlation in these data could lead to invalid conclusions. The correlation matrix provided summarized in Table 3 below should not be interpreted further since it is a representation of the assumed working correlation and could be misleading and therefore no formal inference should be made about the correlation structure (Verbeke and Molembeghs, 2005);

Table 3: *First-order Autoregressive Working Correlation Matrix: Mnt0 to Mnt42 represents the time when first measurement to the eight measurement were collected with time interval of 6 Months*

Working Correlation Matrix								
	Mnt0	Mnt6	Mnt12	Mnt18	Mnt24	Mnt30	Mnt36	Mnth42
Mnt0	1	0.9144	0.8361	0.7646	0.6991	0.6393	0.5845	0.5345
Mnt6	0.9144	1	0.9144	0.8361	0.7646	0.6991	0.6393	0.5845
Mnt12	0.8361	0.9144	1	0.9144	0.8361	0.7646	0.6991	0.6393
Mnt18	0.7646	0.8361	0.9144	1	0.9144	0.8361	0.7646	0.6991
Mnt24	0.6991	0.7646	0.8361	0.9144	1	0.9144	0.8361	0.7646
Mnt30	0.6393	0.6991	0.7646	0.8361	0.9144	1	0.9144	0.8361
Mnt36	0.5845	0.6393	0.6991	0.7646	0.8361	0.9144	1	0.9144
Mnt42	0.5345	0.5845	0.6393	0.6991	0.7646	0.8361	0.9144	1

Table 8 (see appendix section) summarizes type 3 test of fixed effects for GEE1. It can be observed that, no patient characteristics, time and interaction of patient characteristics with time have significant influence on the probability of having wider diameter of aortic artery above the median value. All the p-values were observed to be greater than 0.05 level of significance.

The results in Table 4 summarized the parameter estimates, both model based and empirical standard errors. It can be observed that there is no statistically significant effect of smoking behavior, age, BMI, coronary disease, time in evolution of the diameter of aortic artery. Also interaction between these covariates with time were non-significant at 0.05 level of significance.

Table 4: *Parameter estimates from GEE analysis, provided is the Empirical and model-based Par. and SE.*

Effects		Par.	Empirical SE	Model-based SE	P-value
Intercept	β_0	-2.395	2.8046	2.7739	0.3931
age	β_1	0.0318	0.0279	0.0270	0.2555
Ex-Smokers	β_2	-0.4118	0.8578	0.8303	0.6312
Smokers	β_3	-0.3161	0.8625	0.8442	0.7140
BMI	β_4	-0.0016	0.0541	0.0574	0.9765
CD	β_5	0.3330	0.4515	0.4481	0.4608
time	β_6	0.4383	0.4096	0.3735	0.2845
Ex-Smokers*time	β_7	-0.0538	0.2731	0.1850	0.8439
Smokers*time	β_8	-0.0496	0.2746	0.1921	0.8567
CD*time	β_9	-0.0531	0.1056	0.0962	0.6152
BMI*time	β_{10}	-0.0060	0.0114	0.0128	0.6009

5.4 Generalized Linear Mixed Model (GLMM)

The GLMM model with random intercept and random slope was fitted. Parameter estimates for fixed effects are summarized in Table 5. Similar conclusion that there

is no significance of covariates given the random effect in the model at 5% level of significance.

Table 5: *Parameter estimates, standard errors and corresponding p-values from GLMM analysis*

	Estimate	Std. Error	z value	P-value
(Intercept)	-13.2948	12.0332	-1.105	0.269
age	0.1714	0.1087	1.576	0.115
bmi	0.0094	0.2745	0.034	0.973
exsmok	-2.5106	3.3218	-0.756	0.450
smok	-2.5218	3.5154	-0.717	0.473
cd	2.2295	1.8395	1.212	0.226
time	4.1595	3.7854	1.099	0.272
bmi*time	-0.0861	0.1380	-0.624	0.532
exsmok*time	-0.4870	1.4350	-0.339	0.734
smok*time	-0.4859	1.6430	-0.296	0.767
cd*time	0.0201	0.8669	0.023	0.982

The random intercept and slope with their standard error are 26.258 (5.124) and 3.218 (1.794) respectively. The correlation between the random intercept and slope is 0.49.

Since we were interested in testing whether only random intercept is needed in the model, a likelihood ratio statistics $2ln(\lambda_N)$ which should be compared by the mixture of two chi-squared distributions with 1 and 2 degrees of freedom all having equal weights of $0.5(\chi^2_{1:2})$ was conducted. The $2ln(\lambda_N)$ is given as 0.6 and the p-value obtained from $0.5P(\chi^2_1 > 2ln(\lambda_N) + 0.5P(\chi^2_2 > 2ln(\lambda_N)$ was 0.588. Notably, this implies that the covariance structure could be reduced to only the variance of the random intercept. The reduction of the covariance structure suggested that patients at the begin of the study starts from different points but evolve in the similar manner over time. Henceforth, we adopted the model with only the random intercept for further inference and interpretation.

Table 6: *Parameter estimates, S.E and corresponding p-values for the Random Intercepts model*

	Estimate	Std. Error	z value	P-value
Intercept	-36.4697	23.7754	-1.534	0.125
Age	0.3443	0.2128	1.618	0.106
BMI	0.54442	0.5026	1.083	0.279
Ex smoker	-6.7958	6.1947	-1.097	0.273
Smoker	-6.5540	6.3902	-1.026	0.305
Coronary disease	1.1019	3.0808	0.358	0.721
Time	3.2615	2.9343	1.112	0.266
BMI*time	-0.0733	0.1011	-0.725	0.469
Ex smoker*time	0.3322	1.1704	0.284	0.777
Smoker*time	0.5773	1.3681	0.422	0.673
Coronary disease*time	-0.2046	0.6786	-0.302	0.763

Table 6 shows the results for the random intercept model. It can also be observed than none of the covariates used in the random intercept model had a significant effect on the probability of having diameter of aortic artery greater than the median diameter.

5.5 Empirical Bayes (EB) estimates

The figure 2 shows the standardized Empirical Bayes of the random intercept. We cannot tell of the normality assumption of the random intercept because of the shrinkage property of the empirical bayes. Some of the patients have few measurements which may lead to high variability. A possible way to check for this normality assumption is to extend the model with finite mixtures, but we do not consider that here.

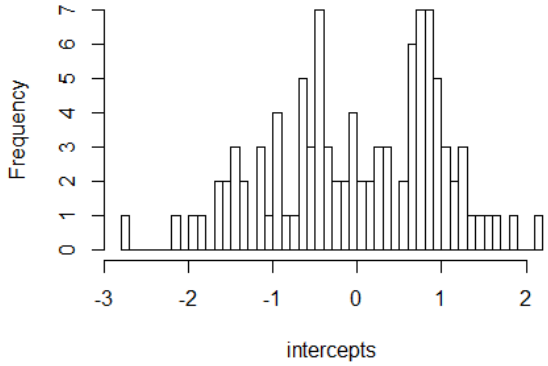

Figure 2: *Empirical Bayes for random intercept*

The major advantage of EB estimates is that they are the best linear unbiased estimator. They are important in checking for outlying observations, and prediction of subject-specific evolution over time. With this advantage, EB estimates are used in outliers detection by expressing subjects whose evolution largely deviates from the average evolution.
The disadvantage of EB estimate is that they are affected by shrinkage property. Shrinkage property is when the observed data are shrunk towards the prior average. This property does not help us to fully observe the heterogeneity in the data. In addition, high within subject variability leads to lot of shrinkage.

5.6 Subject specific predictions

Table 7 shows the observed and predicted values for 3 of the patients. These are not so different from each other, for example, the observed and predicted values for patient 1043 are approximately the same this can be explained by the large between subject variability observed and low within subject variability.

Table 7: Subject Specific Predictions

ID	Time	Age	BMI	Smoking	Coronary Disease	$\hat{b}_{1i}$	Observed	Predicted
1022	0	75	24.8	1	0	-10.9526	0	2.22×10^{-16}
1022	6	75	24.8	1	0	-10.9526	0	2.22×10^{-16}
1022	12	75	24.8	1	0	-10.9526	0	2.22×10^{-16}
1022	18	75	24.8	1	0	-10.9526	0	1.95×10^{-13}
1022	24	75	24.8	1	0	-10.9526	0	8.21×10^{-10}
1022	30	75	24.8	1	0	-10.9526	0	3.45×10^{-6}
1022	36	75	24.8	1	0	-10.9526	0	1.43×10^{-2}
1025	0	81	28.5	0	0	-12.4227	0	2.22×10^{-16}
1025	6	81	28.5	0	0	-12.4227	0	1.11×10^{-14}
1025	12	81	28.5	0	0	-12.4227	0	1.20×10^{-11}
1025	18	81	28.5	0	0	-12.4227	0	1.31×10^{-8}
1025	24	81	28.5	0	0	-12.4227	0	1.43×10^{-5}
1025	30	81	28.5	0	0	-12.4227	0	1.53×10^{-2}
1043	0	78	24.2	1	0	4.8989	1	1.00
1043	6	78	24.2	1	0	4.8989	1	9.95×10^{-1}
1043	12	78	24.2	1	0	4.8989	1	9.99×10^{-1}

6 Discussion and Conclusions

In this report, focus was made on a binary longitudinal outcome. Three analyses were considered to investigate the effect of subject characteristics on the development of AAA disease in the observational longitudinal study. The results obtained from GEE1 shows that the model-based and empirical based standard errors are almost similar although empirical based standard errors were slightly larger, this is an indication that the assumed working correlation is not far from the true underlying correlation structure of subjects (Verbeke and Molenberghs, 2017).

In GEE1, age, Body Mass Index, an indication of coronary disease, smoking behaviour, time as well as there interaction with time have no significant influence on the evolution of the diameter of aortic artery.

The second approach utilizes a transitional model for the data. Modelling the dependence of the diameter of aortic artery on the previously outcome. It was established that there is a very strong dependence on the current outcome given the first previous outcome. Moreover, when checking the effect of patient characteristics on the probability of having relatively higher diameter above the median, it was discovered that similar to results obtained in GEE1 approach, there was no significant effect of age, Body Mass Index, an indication of coronary disease, smoking behaviour, time as well as there interaction with time on the advancement of the diameter of aortic artery.

In the subject-specific approach, a random intercept model was used. In this case, random effect model was reduced to random intercept model. This is an indication that the subjects evolved at the same rate after doing the mixtures of chi-square test. From the results, the all the parameter estimates were observed not to be different from zero. In general, all the patient characteristics used in this analyses

showed no significant effect on the evolution of the aortic artery which is one of the effect that facilitates development of AAA. Contrary to results obtained, in the paper by Normal *etal*, 2013, discussed that smoking is causally associated with abdominal aortic aneurysm. In other older research by Willick *etal*, 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 *etal*, 2018 discussed that the older population are predisposed to the risk of developing AAA. Aging is one of the dominating factor that is associating with AAA. Also, CAD has been studied to have high prevalent amongst AAA patients, for example among ruptured AAA, all CAD patients were above 60 years old; 80% had AAA diameter of $\geq$5.5cm.

7 Recommendation

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] Molenberghs, G. and Verbeke, G. (2005). Models for discrete longitudinal data. New York: Springer.

[2] Molenberghs, G. and Verbeke, G. (2017). Introduction to Longitudinal Data Analysis, course notes. Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat) Uni- versiteit Hasselt and Katholieke Universitiet Lueven, Belgium.

[3] Aerts, M., Molenberghs, G., Ryan, L.M. and Geys, H. eds. (2002). Topics in modeling of clustered data. Chapman Hall/CRC.

[4] Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2009). Longitudinal data analysis. Handbook. Hoboken, NJ: John Wiley Sons.

[5] Verbeke, G. and Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data. Series in Statistics. New York: Springer

[6] 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

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

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

8 The Appendix

Table 8: *Test of Fixed Effects for GEE1 analysis*

Source	DF	Chi-Square	P-value
age	1	1.28	0.2575
smoke	2	0.27	0.8753
bmi	1	0	0.9765
cdisease	1	0.55	0.457
time	1	1.21	0.2717
time*smoke	2	0.05	0.9736
time*cdisease	1	0.26	0.6078
bmi*time	1	0.24	0.6208

9 CODE

```
#*********************************GEE*****************************************#
proc genmod data=ng.aaa_dichotomized descending;
class patient timeClass smoke(ref="0") cdisease(ref="0");
model diameter= age smoke bmi cdisease time smoke*time cdisease*time bmi*time
```

```
                / dist=binomial link=logit type3;
repeated subject=patient / withinsubject=timeClass
type=ar(1) covb corrw modelse;
run;
#***************************Transition Model*********************************#

data ng.aaa_dichotomized_trans;
set ng.aaa_dichotomized;
diameter_1=Lag1(diameter);
diameter_2=Lag1(diameter);
run;

proc genmod data=ng.aaa_dichotomized_trans descending;
class smoke(ref="0") cdisease(ref="0");
model diameter = age smoke bmi cdisease time smoke*time cdisease*time bmi*time
                diameter_1 diameter_2/ dist=binomial;
run;

glmm13 <- glmer(Dia2 ~ AAA$Age.at.study.entry + AAA$BMI + AAA$Smoking.. +
                  AAA$Coronary.disease.. + AAA$Measurement*AAA$Coronary.disease.. +
                  AAA$Measurement*AAA$BMI + AAA$Measurement*AAA$Smoking.. +
                  (1 + AAA$Measurement | AAA$Patient) ,data= AAA,
               family=binomial(link='logit'), nAGQ = 1, method = "REML")


glmm2 <- glmer(Dia2 ~ AAA$Age.at.study.entry + AAA$BMI + AAA$Smoking.. +
                  AAA$Coronary.disease.. + AAA$Measurement*AAA$Coronary.disease.. +
                  AAA$Measurement*AAA$BMI + AAA$Measurement*AAA$Smoking.. +
                  (1| AAA$Patient) ,data= AAA,
               family=binomial(link='logit'), nAGQ = 10)

#***********************************GLMM*****************************************#
```