Aorta Diameter Growth in Marfan Syndrome Patients

Students - Nancy Dsouza (13803255) and Shamili Kolli (14568659)

The Marfan data was imported and inspected for any missing values or negative values. No missing or negative values were found. The dataset consists of 159 unique rows of values (patient numbers).

1. Data Description

Table 1 gives a description of the repeated measures taken over the whole dataset (not unique patients) and table 2 gives a description of the characteristics of the unique patients in the dataset. The data contains a total of 1036 observations with 159 unique patients. From the table results, it appears that there is an overrepresentation of females since only 25.1% are males.

Table 1: General Characteristics taken over 1036 patients

	Median (Q1 - Q3)
Age	31.56 (25.2 - 40.6)
Diameter	41.0 (37.0 - 45.0)
	n (%)
Female	776 (74.9 %)
Male	260 (25.1 %)

Performed over repeated measurements (not unique patients). Units: Age(years), Diameter(mm)

Table 2: General Characteristics taken over 159 patients

		Median	Q1	Q3
Age	Min	26.47	19.2	34.7
	Mean	30.26	22.5	37.3
	Max	32.96	24.8	41.9
Diameter	Min	39.00	34.0	44.0
	Mean	41.17	36.4	46.2
	Max	43.00	39.0	49.0
Follow up	Mean	5.80	3.0	8.5
Meeting number	Number of msts	6.00	3.0	9.0

	n (%)
Female	108 (67.9 %)
Male	51 (32.1 %)

Performed for each unique patient. Units: Age(years), Diameter(mm), follow-up time(years)

Figure 1 shows the aorta diameter measures over time considering all patients. Figures 3 and 4 show the aorta diameter measures with age considering a selection of patients whose diameter increments over time are clinically significant.

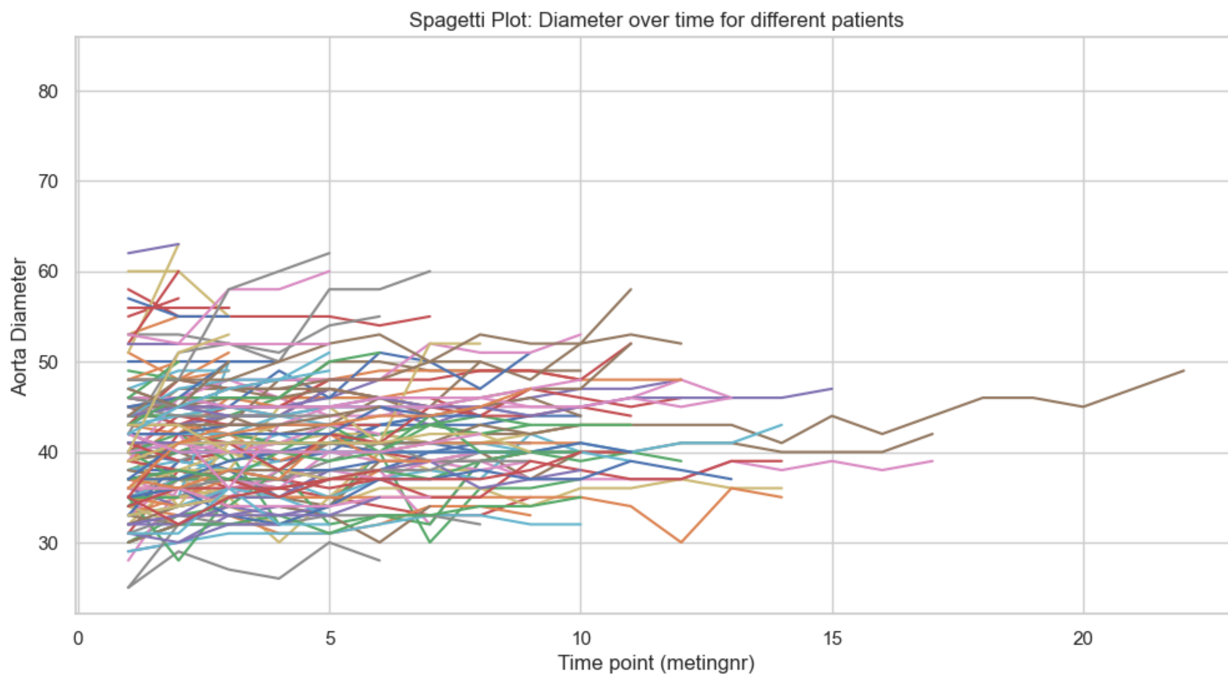


Fig 1: Spaghetti Plot showing the relationship of Aorta Diameter over Time (considering all 159 unique patients)

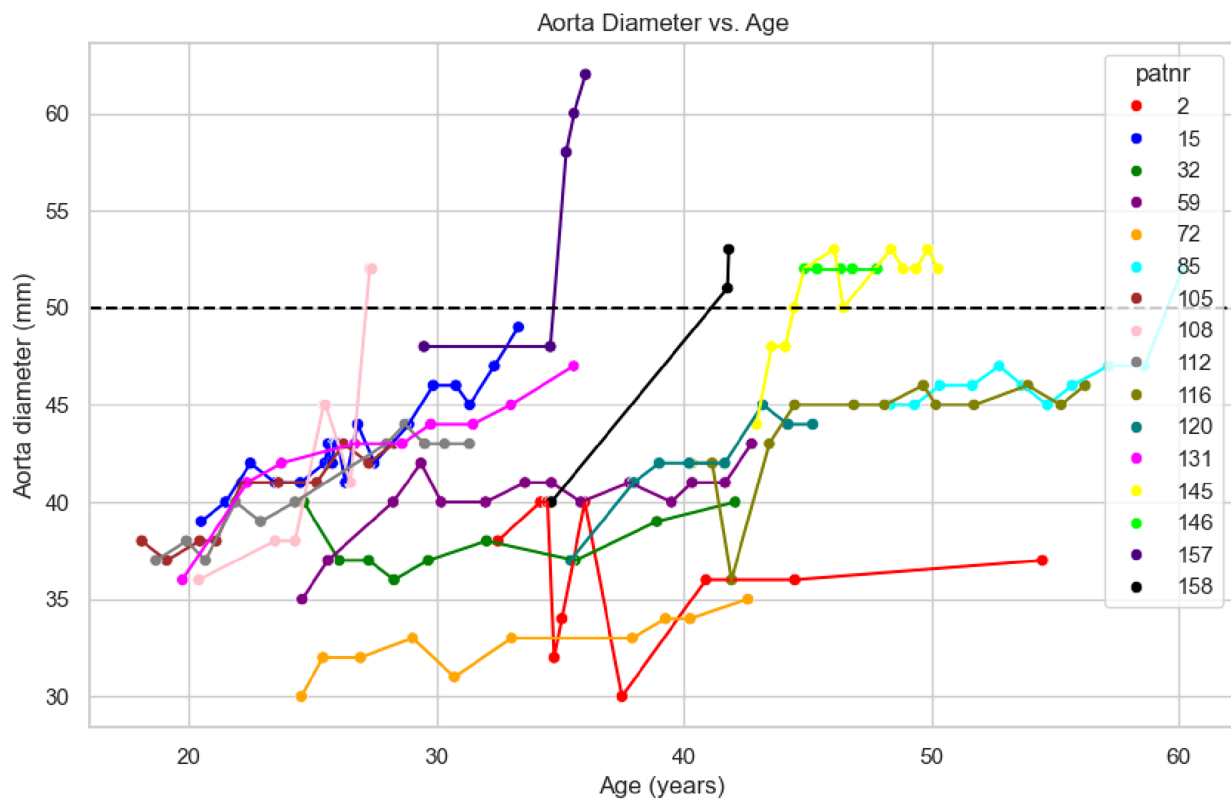


Fig 2: Plot showing the relationship of Diameter with Age (shown in only 16 patients)

Most patients depict increments in diameter with time, either slow (patient 131) or fast (158), while some patients depict stable measures as well (32, 146). There are some that have crossed the 50 mm mark, illustrating an indication for surgery. However, patients 145 and 146 have their diameters stabilizing over time after crossing 50 mm.
Additionally, another surgery indication is a depiction of slightly lower diameter combined with a growth rate of over 5 mm per year. This can be seen for patient 15. There is some measurement variability present, which is evident for patient 2, since the diameter is unlikely to decrease over time, and then increase and decrease so often within a small timespan. This variability may be caused by variations in observers, machines, or measurement locations. For most patients, the relation between diameter and age appears to be linear.

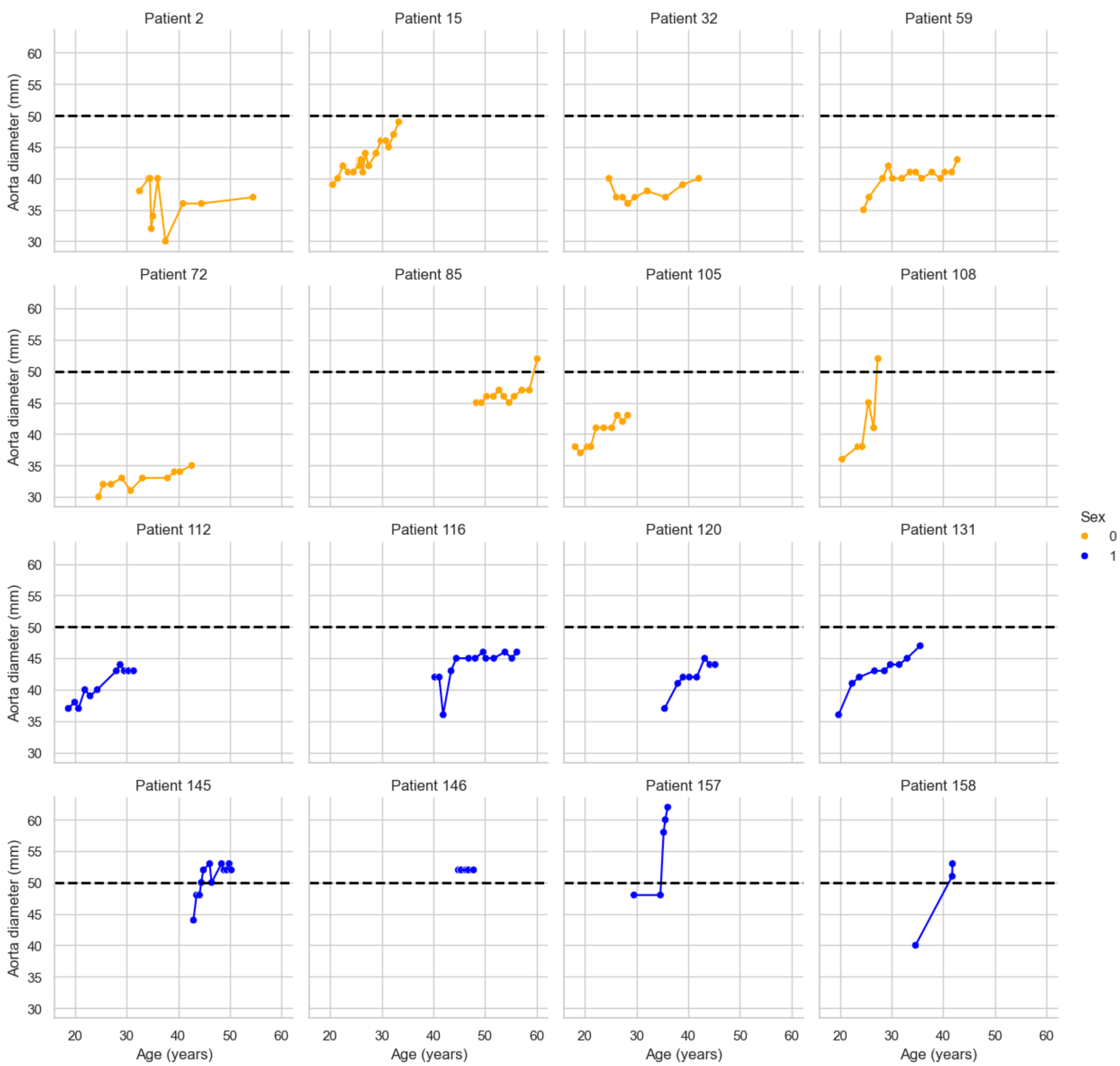


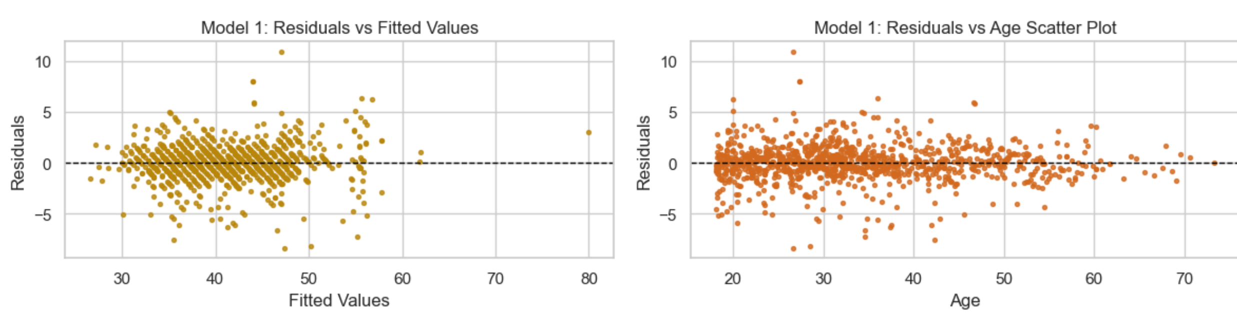
Fig 3: Individual plots of the diameter vs Age in the considered 16 patients

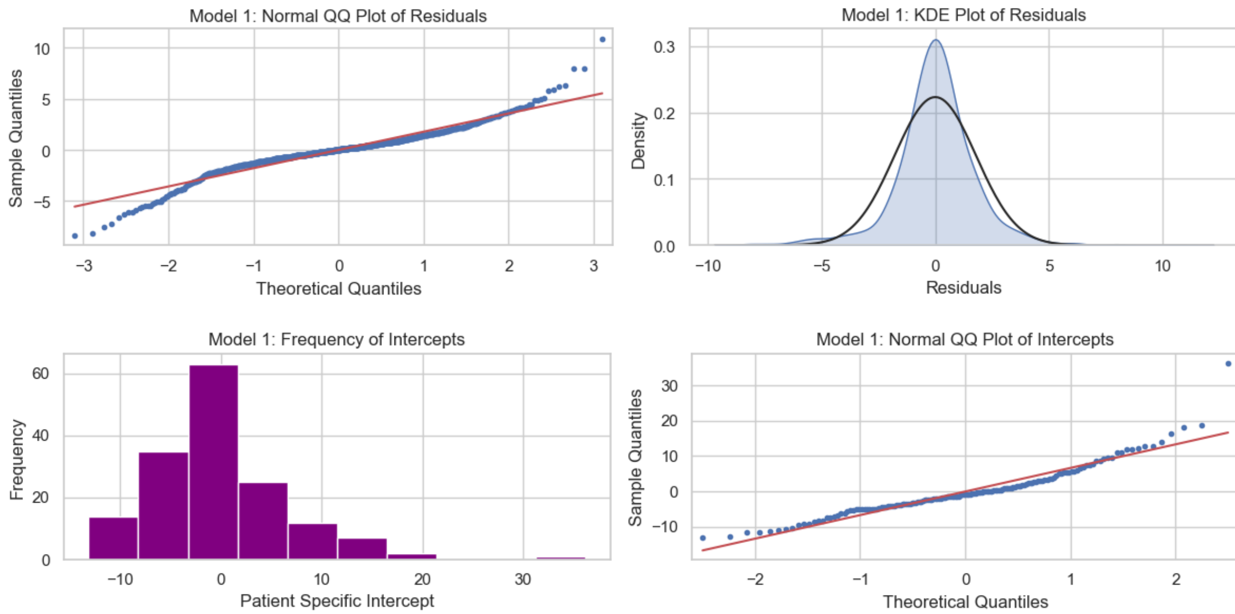
2. Linear Mixed Effects Regression Model

Since each patient has multiple measures over time of aorta diameters, we decided to go for a Linear Mixed Effects Model. Also, different patients have different starting points in their diameter measures with or without varying rates of increase.
In order to use a Linear Mixed Effects model, several assumptions like data linearity, data normality and random effect normality have to be verified. We tested 3 different models and selected the best one based on their AIC and BIC scores. In the description of the models below, we fit the model with the respective characteristics and testing for the assumptions. The results of these assumptions are shown below.

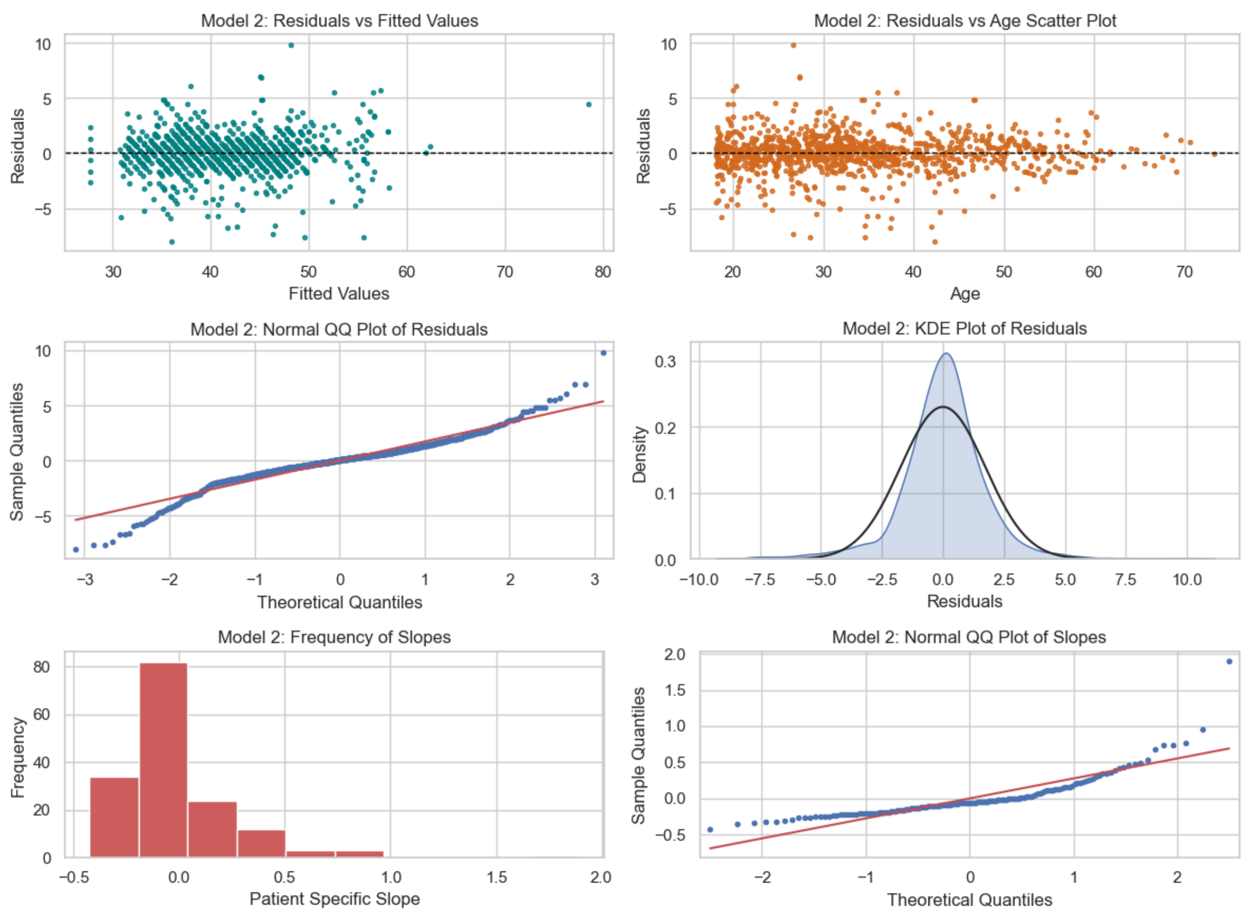
The results of Model 1 and 2 consist of 6 graphs. The first 2 graphs are the residual plots. . If the points in a residual plot are randomly dispersed around the horizontal axis, the variance of the residuals is constant, meaning that the linearity assumption is met for the data. In this case, the model meets the linearity assumption. The 3rd graph plots the data in sorted order versus quantiles from a standard normal distribution. Points on the Normal QQ plot provide an indication of the univariate normality of the dataset. If the data is normally distributed, the points will form a 45º straight line. The 4th graph is the Kernel Density Estimation plot (KDE) of the residuals against a normal distribution (black line). If the residuals density is normally distributed, then this suggests that the residuals meet the normality assumption.
The 5th graph, on the lower left, plots the frequency of the individual intercepts in the model. If the data is normally distributed with a mean of 0, the histogram will show a normally distributed plot centered at 0. The 6th graph, on the lower right, plots the individual intercepts in sorted order versus quantiles from a standard normal distribution. Points on the Normal QQ plot provide an indication of univariate normality of the intercepts. If the intercepts are normally distributed, the points will form a 45º straight line. In all the models, all of the assumptions are met.

(a) Model 1: Random Intercept and Fixed Slope

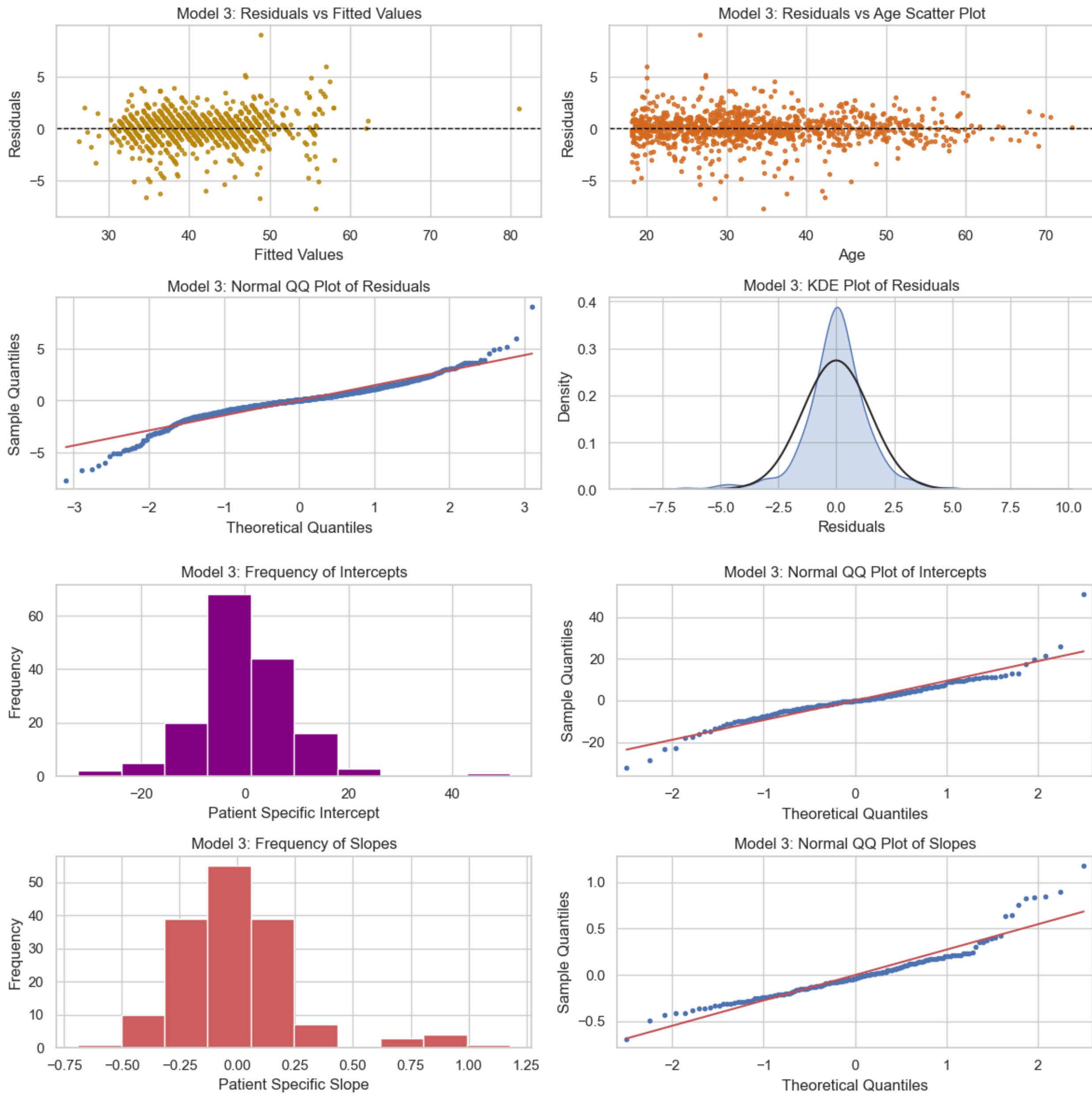




(b) Model 2: Fixed Intercept and Random Slope



(c) Model 3: Random Intercept and Random Slope



We have then calculated the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) of each of the models.
AIC determines the relative information value of the model using the maximum likelihood estimate and the number of parameters (independent variables) in the model. Similarly, BIC penalizes the number of parameters used, however it is recommended to avoid overfitting. Lower AIC/BIC scores are better, and AIC/BIC penalizes models that use more parameters. So if two models explain the same amount of variation, the one with a lower AIC/BIC score will be the better-fitting model. In this case, the best model is

the one with random intercept and slope (model 3) with the lowest AIC and BIC scores (4881.4 and 4916 respectively), shown in Table below.

Table 3: ANOVA Test comparing different models on the AIC & BIC

	Type of Model	AIC	BIC
Model 1	Random intercept & Fixed slope	4988.306	5013.022
Model 2	Fixed intercept & Random slope	5003.886	5028.602
Model 3	Random intercept & Random slope	4881.411	4916.013

We tested the performance of Model 3 by checking how well the model has predicted the diameter measures. This can be seen in the figure below. In this figure as well, we have plotted the fitted values along with the actual diameter measures of the 16 clinically significant patients (considered above as well). The lines in between the points denote the fitted values and the points denote the actual diameter measures of a particular patient.

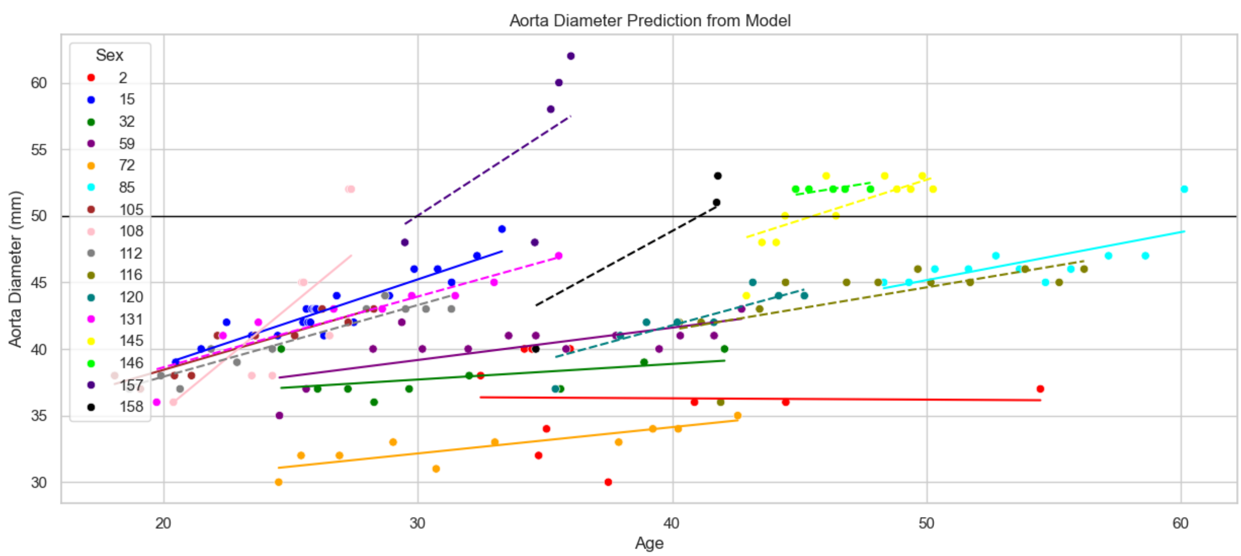


Fig 4: Aorta Diameter Predictions from the model vs the real diameters of different patients according to their age.

3. Predictions for New patients

Figure 4 depicts the aorta diameter predictions for new fictional patients with randomly selected ages varying between 20 to 50 years. The predictions are made using model 3 (random intercept and random slope). Men are predicted to have significantly higher diameters than women. The confidence intervals are smaller for women (<3 mm) than for men (<5 mm). This suggests that on average, a male with Marfan's syndrome of which we have no prior measurements is expected to have an indication for surgery at 36 years of age (95% CI: 31-46 years). However, the diameter growth pattern varies widely between patients, which can be seen in figures 3 & 4, and by the fact that the random slope and random intercept model fits the data best. Therefore, taking one or more prior measurements is likely to majorly increase the prediction's accuracy.

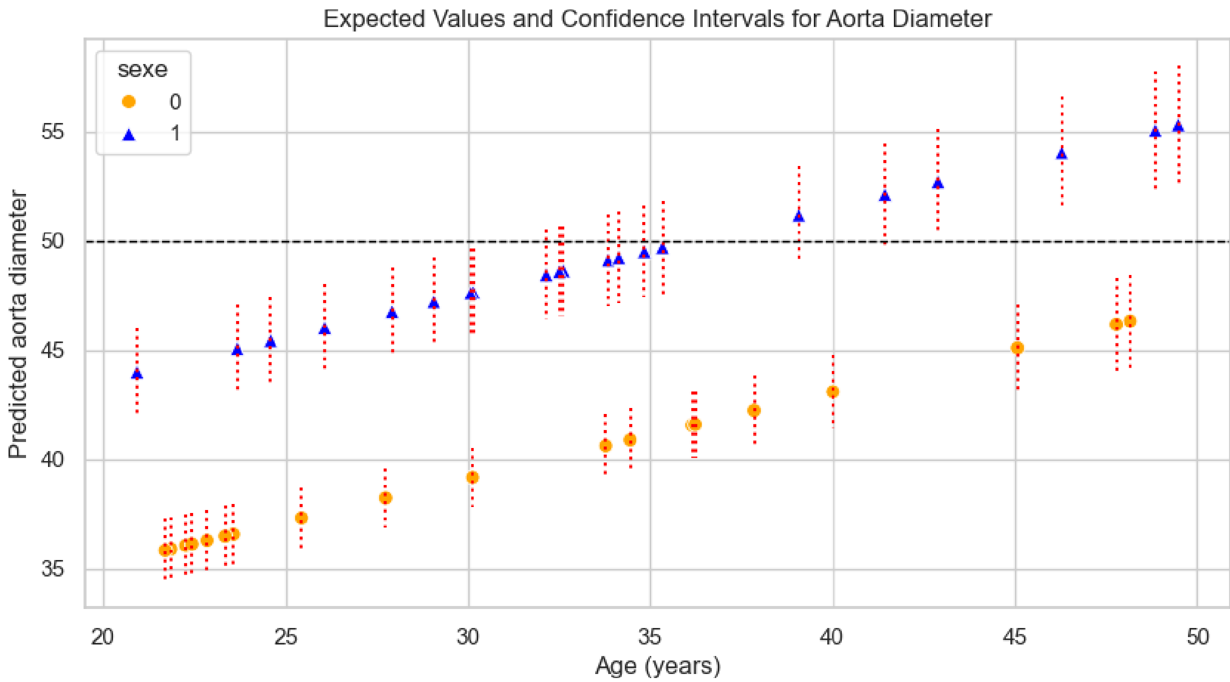


Fig 5: Aorta Diameter Predictions for New patients using Model 3

4. Predictions for Current Patients

Figures 6 & 7 depicts the predicted aorta diameters for the patients in our data up to one year after their last available measurement plotted against their future age and their patient numbers respectively. In both the figures, the predictions are represented as coloured data points (according to male or female) with their respective 95% CI bounds in red dotted lines.

From both figures, males are predicted to have slightly higher aorta diameters than females, since most males than females cross the 50 mm threshold in a year. The confidence intervals are slightly smaller for young women than young males, however they tend to increase with age for both sexes. These observations are in line with figure 5 where predictions were made for fictional patients with no prior measurements.

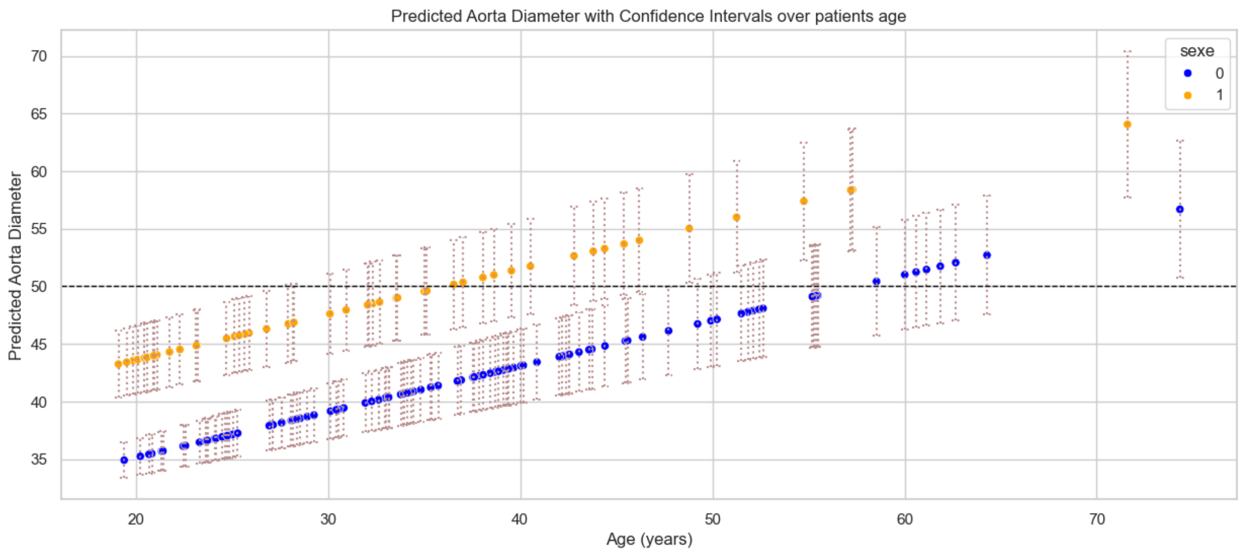


Fig 6: Predicted Aorta diameters with their confidence intervals of different patients with their age (1 year later).

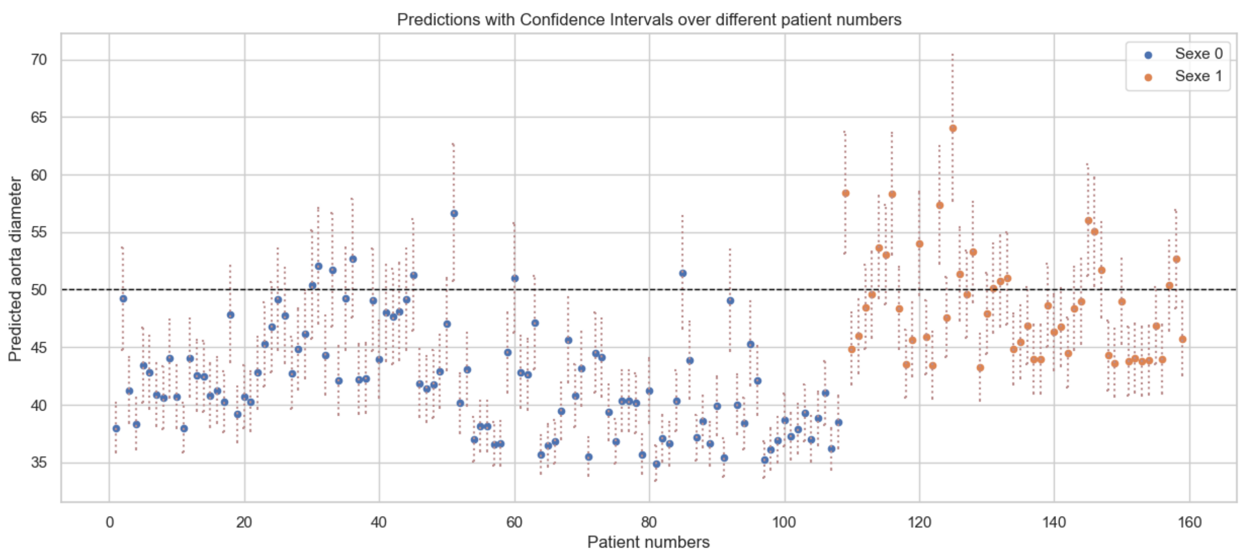


Fig 7: Predicted Aorta diameters with their confidence intervals of different patients with their patient numbers (1 year later).

The prediction model used can give insight into how reliable the aorta measurements are likely to be, while the aorta measurements can indicate whether the growth is deviating from the prediction. The prediction model can be used to monitor whether the patient has a risk of crossing the 50 mm threshold, and an early decision can be made regarding treatment or aortic surgery. Therefore, combining the prediction model and aortic measurements brings the potential to make more accurate estimates of the actual diameter than either of them alone.