Analysing Corpus Callosum Atrophy in Traumatic Brain Injury Patients

Lucas Campbell, Bu Wang, Jasraj, Zhiyuan Lyu

09/04/2021

1. Abstract

The primary purpose of this study was to understand the effect certain variables had on the corpus callosum's rate of atrophy in Traumatic Brain Injury (TBI) patients. In order to make accurate interpretations on variable effects, an investigation was made obtaining a model that best fit the data provided. Given the nature of this study, being exploratory rather than predictive, there was less focus on the predictive power of model(s). A mixed model was found, such that the fixed and random effects explained for approximately 78% of the total variation in the data (conditional R^2); 16% of which was explained by the fixed effect model (marginal R^2). Time was found to be highly significant in the rate of atrophy (p-value ≤ 0.05). The rate of atrophy between the control and patient group was found to be statistically insignificant for the first 925 days post injury. After such period, the patient group rate of atrophy/volume loss was found to be significantly different than the control group. On day 925, a unit change in the group variable (control [0] to patient [1]) resulted in a decrease in the expected total volume of the corpus callosum by a factor of approximately 92.107 units. By day 2000, this decrease became a factor of approximately 180.48 units. The overall effect of sex was found to be statistically insignificant (p-value = 0.092) in the model. However, it was observed that the sexes had significant differences in expected volume loss after day 416. The age group an individual belonged to was considered to be statistically significant in the model (p-value negligible to 0.05). Belonging in the 19–33-year-old age group decreased the expected volume of the corpus callosum by a factor of approximately 123.21 units and belonging in the 65-78-year-old age group, by a factor of approximately 146.27 units. All in all, the model and results obtained in this study were found to be both significant and insightful with respect to the understanding of the variables' effects on corpus callosum atrophy.

2. Introduction

Medical research regarding traumatic brain injuries and specifically, corpus callosum atrophy has many valuable applications. The corpus callosum is a thick bundle of nerve fibers that ensures both sides of the brain are able to communicate and send signals to each other. Damage and loss of volume to this area is known to cause mental disorders, pseudobulbar palsy and speech and movement ataxia. These fibers naturally decay over time, however, for patients who suffer from traumatic brain injuries, the rate of atrophy could be significantly different. Understanding how such injuries affect an individual's natural aging process can aid medical researchers in finding suitable treatments, as well as, preventative measures. This, in turn, has the potential to improve the average quality of life, whilst mitigating medical risk for those with traumatic brain injuries. The following report looks deeper into the corpus callosum's rate of atrophy, as well as, other information pertaining to the traumatic brain injuries affecting this area.

3. Data

3.1 Overview

The Traumatic Brain Injury (TBI) data set was used to fit various Mixed Models in order to analyze the effect of multiple factors on the volume of patient's Corpus Calosum over time. This data set was graciously gifted by an undisclosed research team in North America. It consisted of 775 observations regarding information on TBI patient visits to an undisclosed hospital/research facility. A total of 80 variables were measured on site and an additional 3 derived from the pre-existing values in the set. The orginal practioners/authors of the data set obtained their results from a medical trial. As such, the patients were not all TBI patients but instead some stem from a control group. The following report utilized six main variables in the analysis: "CC_TOT", "elapsed time", "sex", "age", "age group" and "group". For a breakdown of these variables please refer to Table 1.1 at the end of the "Data" section. The research done in this report fits within the larger field of bio-statistical research, trauma studies, data analysis and datamining.

3.2 Transparency in the discretization and creation of variables

3.2.1 Elapsed Time

One of the main focuses of this report was to analyse the atrophy of patients' corpus callosum over time. As such, a variable needed to be constructed to represent the patients' time since injury/last medical visit. A pre-existing variable "days.post.injury" already existed in the data set. However, the previous practitioners did not include any numerical values for those patients in the control group, leaving this variable undesirable for the study. The date and visit variables were utilized to make a universal "elapsed time" variable for all subjects. This variable was constructed by treating the first visit date (in days) as "time zero" for the control group. For the patient group, if available, the elapsed time since injury plus the first visit date was used as a "time zero". Each succeeding visit date was then recorded in relation to this original time. Many individuals, especially those belonging to the control group, had missing recorded dates for their first visit. For these patients, the minimum date (in days) among all other recorded first visits was assigned. The minimum was chosen in lieu of the average in order to mitigate/remove the risk of negative elapsed time values.

3.2.2 Age Group

In many medical related studies, using an individual's age group (teen, adult, elderly etc.) is often a great generalization which yields significant group-wise information. As such, it was desired to descretize the pre-existing "age" variable in such a manner. Referencing the academic paper "Redefining meaningful age groups in the context of disease" (Geifman 2013), an additional age variable, "age.groups" was created, assigning each observed patient into one of the following 9 age groups:

Group								
"0-2"	"3-5"	"6-13"	"14-18"	"19-33"	"34-48"	"49-64"	"65-78"	"79-98+"

Individuals whose age was not recorded at a given visit were assigned the previous age group they belonged to. It was believed that such an assignment assumption would be the most reasonable, given the probability a patient "changes age groups" within a span of 2504 days is negligable (2504 being the maximum range of the elapsed time variable).

3.2.3 Sex

The practioners who gifted this data set did not disclose their variable notation. Sex is a categorical variable whose value takes on either "1" or "2". Given the traditional recording practice of recording Males first versus Females, along with the fact that more patients listed as "1" experienced TBI's, it was assumed that "1" denoted Male patients and "2" Female.

Table 1.1 - The Used Variables

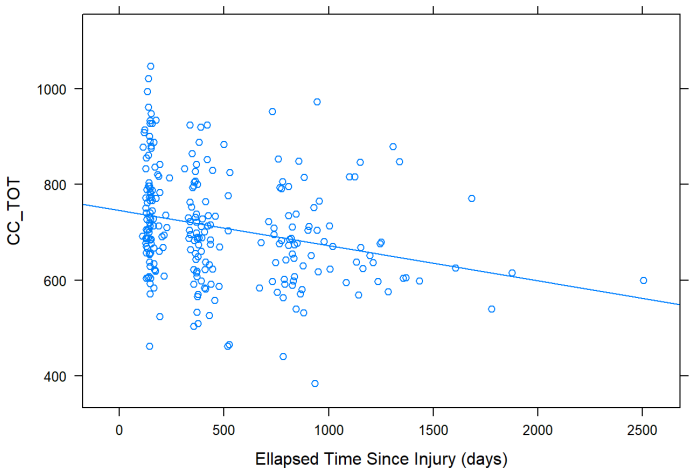
Variable Name	Type	Valid Range	Description
CC_TOT	Continuous	384-1105	Measures the Volume of the Corpus Calsum
elapsed.time	Integer	0-2504	Measures elapsed time since patients first visit
sex	Categorical	1 (Male), 2 (Female)	The sex of the patient in question
group	Categorical	control, patient	Defines which treatment group the patient is in
age	Continuous	Reals	The patient's age as a fraction i.e. 23.94
age.group	Categorical	"14-18", "19-33", "34-48", "49-64", "65-78"	Records the age group an individual belongs to

It should be noted that for reading and mathematical clarity, the variables may be referred to as y_{cc}, x_1, x_2, x_3, x_4, x_5 respectively. Additionally, the exact units used by the data collectors to measure corpus callosum volume etc. was unknown. As such, the word "units" was used as a substitute.

4. Analysis and Results

4.1 General understanding of the data landscape

The first plot below depicts the trend of the pooled data. Specifically, this is the pooled data of all individuals not belonging to the control group. In this plot, we see that as elapsed time increases, the overall volume of the corpus callosum decreases. This was the suspected population trend.

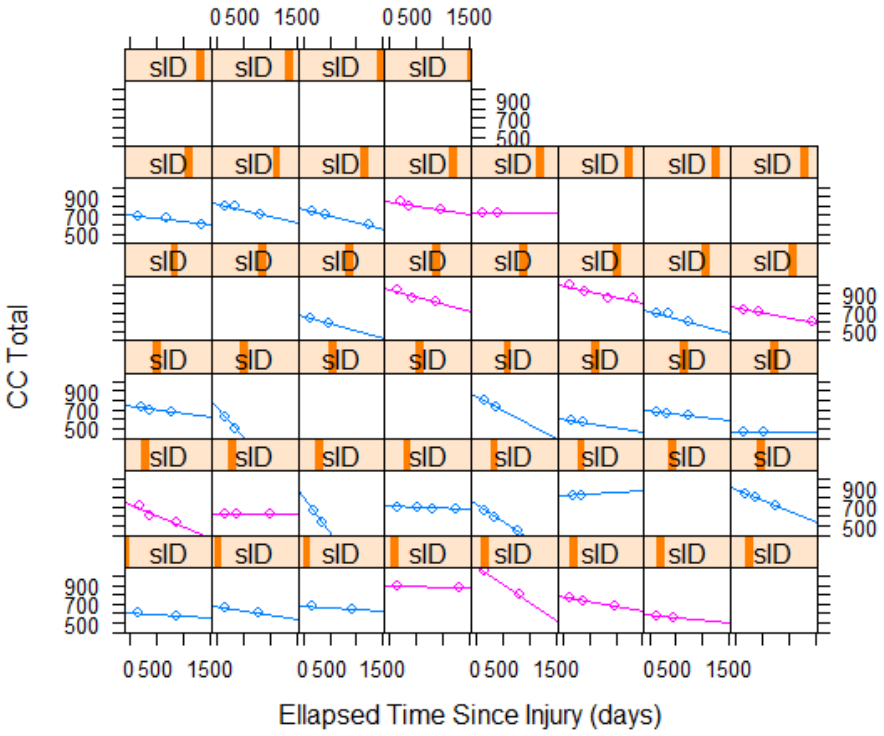


The second plot depicts more valuable information. This plot looks at the volume of an individual's corpus callosum over time for a subset of 37 TBI patients. Each cell represents an individual and the colour denotes the gender (blue for males and pink for females).

As can be seen from this plot, there appears to exist variability among the slopes and intercepts of each individual patient. Thus, as expected, this suggests the data set would be more appropriate for mixed

modelling analysis, allowing for random effects. Additionally, as seen, sex appeared to be a potentially significant variable. Within the female indivudials, the plotted lines seem to be more concisely grouped (less variable), as well as, on average, higher in initial volume than men.

It is commonly known that bone denisty differs between the sexes. Given that TBI's typically occur due to force applied on the cranial structure, this difference in variability and intercept grouping is thought to be representative of an underlying biological effect relating to this phenomenon.



4.2 Obtaining a suitable Mixed Model for interpretation

The practioners of this study were first tasked with finding a model which had a good fit of the data. A total of four models were obtained and compared according to their AIC value. It should be noted, that given the nature of this report being an exploratory and non-predictive study, the p-value and significance of variables in the model were not highly focused on. That is, models with insignificant variables were still valued over sparse ones with all significant variables because of their possible suppressing and confounding nature. Additionally, doing this prevented Dylan's paradox from being a significant issue.

A summary of the four models, along with their respective AIC values, can be seen in the table below.

Model	Independent Variables	Random Effects	AIC
1	$x_1, x_2, x_3, x_5, x_1 : x_2, x_1 : x_3, x_1 : x_5$	intercept & elapsed.time(x_1)	3470.105
2	$x_1, x_2, x_3, x_1 : x_2, x_1 : x_3$	intercept & elapsed.time(x_1)	3475.432
3	$x_1, x_2, x_3, x_4, x_1 : x_2, x_1 : x_3$	intercept & elapsed.time(x_1)	3474.378
4	$x_1, x_2, x_3, x_5, x_1 : x_2, x_1 : x_3$	intercept & elapsed.time(x_1)	3440.518

Model 4 obtained the smallest AIC value, meaning it obtained the best fit of the four. The model and its components can be mathematically represented as follows:

$$\beta_0 = \gamma_{000} + \gamma_{001}X_2 + \gamma_{010}X_3 + u_{00} \ , \quad \beta_1 = \gamma_{100} + \gamma_{101}X_2 + \gamma_{110}X_3 + u_{10}$$

$$Y_{cc} = \gamma_{000}+\gamma_{001}X_2+\gamma_{010}X_3+\gamma_{100}X_1+\gamma_{101}X_2X_1+\gamma_{110}X_3X_1+\beta_3X_{5(1)}+\beta_4X_{5(2)}+\beta_5X_{5(3)}+\beta_6X_{5(4)}+u_{00}+u_{10}X_1+\epsilon$$

Where the γ's are the fixed effects, u's the random effects, and $X_{5(i)},\ i = 1, 2, 3, 4$ are binary dummy variables for the different age groups (1=19-23, 2=33-48, 3=49-64, 4=65-78). Note here that the group 14-18 was used as the baseline group and no patients in the data set belonged to the 0-2, 3-5, 6-13, and 79-98+ age groups.

The summary for the fixed effect model coefficients can be seen in the table below:

Estimated Value (γ's)	Corresponding Variable	Standard Deviation	P-Value
855.9006	Intercept	71.74727	≤ 0.0001
0.0188	elapsed time	0.01988	0.3453
40.8874	sex	24.16308	0.0933
-16.0624	group	58.56743	0.7844
-123.2141	age group (19-33)	42.30682	0.0041
-138.5281	age group (33-48)	45.81660	0.0029
-137.3872	age group (49-64)	46.04824	0.0033
-146.2683	age group (65-78)	53.19824	0.0066
0.0055	elapsed time : sex	0.01447	0.7065
-0.0822	elapsed time : group	0.02562	0.0016

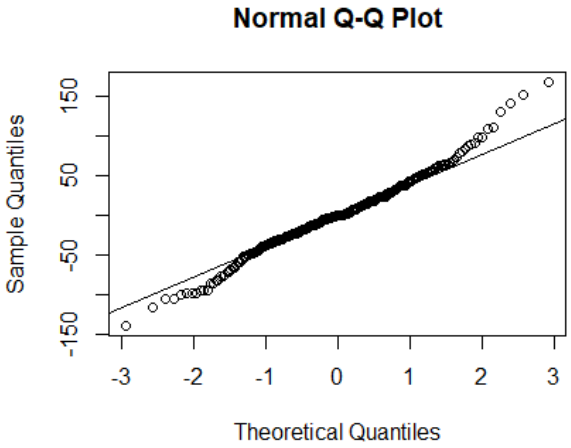
It should be noted that a Wald test for the overall model was run, obtaining a p-value ≤ 0.00001.

4.3 Checking model assumptions

The following section checks whether the Normality, Homoscedasticity/Constant Variance, Linearity and Independence assumptions hold.

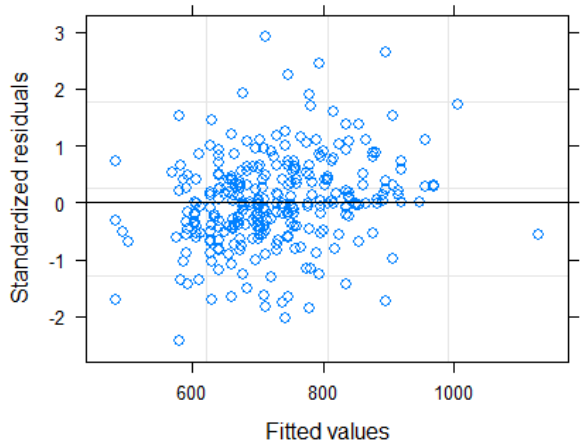
4.3.1 Normality

The following is a Normal QQ plot. While the tails appear to deviate from the line there does not appear to be a serious deviation in general and so, it can be assumed that no serious violation of the normality assumption was made.



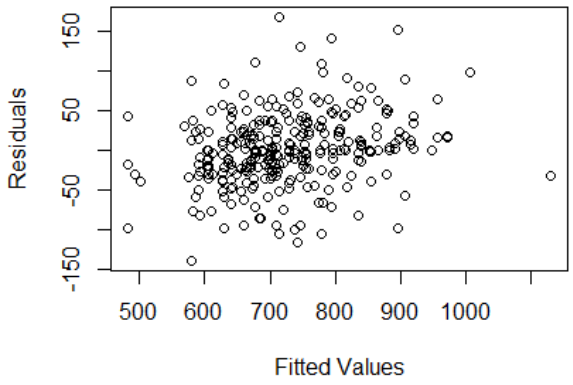
4.3.2 Homoscedasticity

The following residual plot shows no obvious trend and appears to be random about zero. As such, there does not appear to be a strong violation of the homoscedasticity assumption.



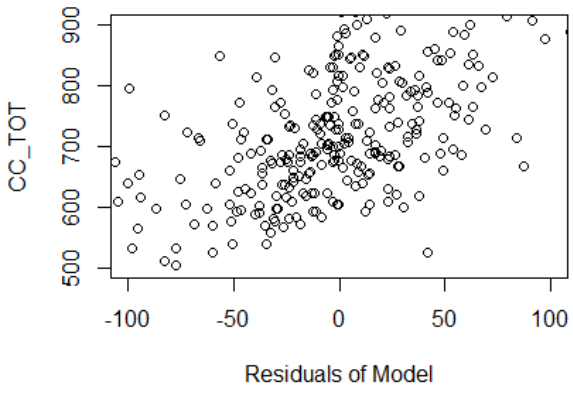
4.3.3 Linearity

The "Residuals vs Fitted" plot can be used to check for the Linearity assumption. As can be seen below, the values appear to be fairly random and do not deviate too far away from zero. As such, there does not appear to be a significant violation of the Linearity Assumption.

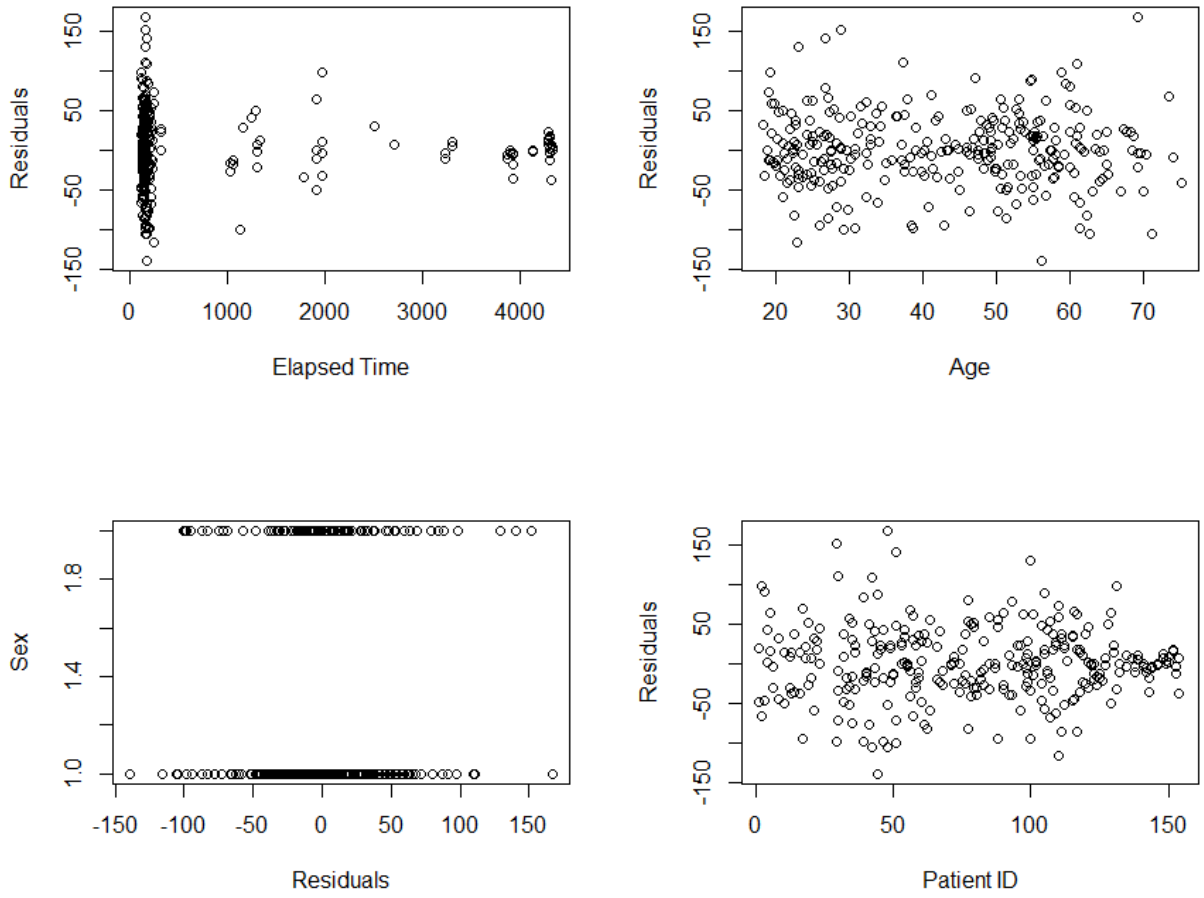


4.3.4 Independence and Autocorrelation

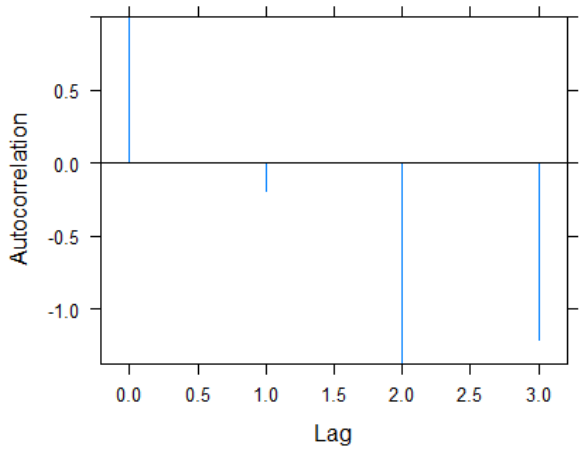
It was known that the data was collected in a time series manner. As such, the observations for each patient's visit was not independent of the others. However, the data was looked into further to see whether or not a serious enough violation of the assumption occured which would require an autocorrelation correction. The Corpus Callosum Total Volume (CC_TOT) was plotted against the residuals of the model and is seen below. While a slight upward linear trend is seen, this does not appear to be "too drastic".



To see whether of not autocorrelation is a significant issue in this model, the residuals were plotted against the predictors. A non-random trend of the data points would suggest autocorrelation should be accounted for. As can be seen in the four plots below, the residual plots appear to be fairly random. While it is recognized that a slight "funnelling in" can be seen in the "Residuals vs. Elapsed Time" plot, this can be thought to be the result of the sparsimonous amounts of data points/observations at the higher elapsed times. Thus, this "funneling in" was not considered a serious violation.



Furthermore, an ACF plot was made for the model. This plot shows the autocorrelation at different lag points. If a trend is found in this plot then this suggests autocorrelation has a big effect in the model. However, even though some of the lag points appear to be autocorrelated, there are not enough lags to suggest/infer any trends regarding the autocorrelation. This is thought to be a consequence of the fact that many patients only had 2-4 recorded visits.



Additionally, when autocorrelation was adjusted for in the Mixed Model, the conditional R^2 of the fixed and random effects model decreased significantly when compared to Model 4 (Conditional $R^2 \approx 0.16\ldots$ R^2 will be discussed further in the next section). While the AIC value for the adjusted model decreased when compared to Model 4, the decrease was less than 40. Given Model 4 has a negilgable AIC and also having a significant increase in the proportion of explained variance in the data, this model was preferred over the adjusted one.

It should be noted that the practioners recognize that not accounting for the possible autocorrelation between points may affect the accuracy of the standard deviations of the estimated effects, as well as, inject bias into the estimated coefficients. However, after observing the plots and significant difference in R^2 values, this bias and lack of conservativeness was not thought to be a significant issue for this specific study.

4.4 Marginal and Conditional R^2's of the Model

The R^2 values of the model were calculated according to the method expressed by Shinichi Nakagawa and Holger Schielzeth in their 2012 paper, "A general and simple method for obtaining R^2 from generalized linear mixed-effects models" (Nakagawa 2012). For those unfamiliar, this method calculates two R^2 values, one marginal and another conditional. The marginal R^2 denotes the proportion of the total variance explained by the fixed effects model, and the conditional R^2, the proportion of the total variance explained by the fixed and random effects model.

Let f denote the fixed effects, r the random effects, and ϵ the model residuals. Then,

$$R^2_{marginal} = \frac{V(f)}{V(f) + V(r) + V(\epsilon)} \quad , \quad R^2_{conditional} = \frac{V(f) + V(r)}{V(f) + V(r) + V(\epsilon)}$$

Model 4 obtained a marginal R^2 of approximately 0.1696233 and a conditional R^2 of approximately 0.7803927. Thus, the model's fixed and random effects explained for approximately 78% of the total variation in the data; 16% of which was explained by the fixed effect model.

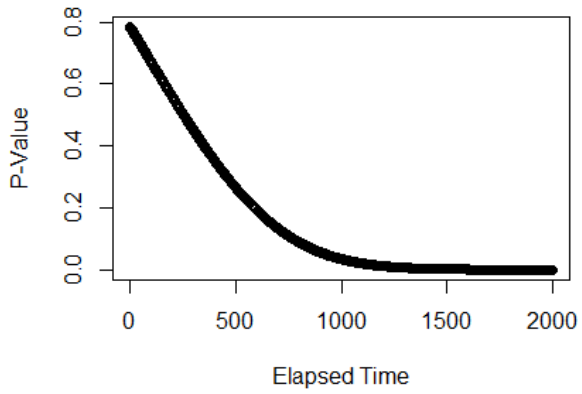
4.5 The significance between the Control and Patient group

It was desired to determine if there was sufficient evidence suggesting the rate of atrophy/decrease in corpus callosum volume differed between the control and patient groups. Taking the partial derivative of the model with respect to the group variable (X_3), yielded the following function of time.

$$\frac{\delta Y_{cc}}{\delta X_3} = \gamma_{010} + \gamma_{110} X_1 = -16.0624 - 0.0822 x_1$$

Due to the fact that the reference level is the control group, this partial derivative shows the significance in the "gap" between the groups and the effect that being in the patient group has over time (in days).

A Wald test was used to see this significance at different time points. The plot below shows the significance of the patient group (the gap) over the time.



Using a significance level of $\alpha = 0.05$ the difference between the groups, as seen above, becomes statistically significant after 925 days "post injury". Additionally, it is seen that this significance is increased after this period as well. For instance, on day 925 the p-value is 0.0497, day 1000, 0.0342, and day 2000, 0.00014.

4.5.1 The effect that being a patient had on atrophy rate

On day 925, a unit change in the group variable (control [0] to patient [1]) resulted in a decrease in the expected total volume of the corpus callosum by a factor of approximately 92.107 units. On day 1000, this decrease became a factor of approximately 98.27 and by day 2000, 180.48 units.

4.6 The effect of Time

It was desired to determine whether or not elapsed time since injury had a significant effect on the atrophy of an individual's corpus callosum. A wald test was conducted to see if the time variable was significant. This test produced a p-value of 0.00833 implying that elapsed time was statistically signficant.

In order to determine the effect time had, the partial derivative of the model with respect to the elapsed time variable (X_1) was taken, resulting in the following function of sex and group.

$$\frac{\delta Y_{cc}}{\delta X_1} = \gamma_{100} + \gamma_{110}X_3 + \gamma_{101}X_2 = 0.0188 - 0.0822x_3 + 0.0055x_2$$

By "plugging in" the different possible combinations for X_3 and X_2, the following effects table was obtained.

Effect of Time for..	Estimated Effect	Standard Deviation	P-value
Female and Patient	-0.057937	0.019622	0.00360
Female and Control	0.024274	0.017900	0.17691
Male and Patient	-0.063395	0.020537	0.00237
Male and Control	0.018815	0.019882	0.34534

From the table above, it is can be seen that, the effect of time for female patients and the effect of time for male patients are highly significant. However, the effect of time for control groups of both genders appear insignificant. One possible reason for this phenomenon: the control group is just like placebo group in the case of vaccine experiment; that means the control group would not be expected to change drastically over an elapsed period of time. The role of the control group was to compare the "natural" decay rate to the patient group in order to test for a significance decrease in expected volume loss. Therefore, this result was considered reasonable for control group.

For females experiencing traumatic brain injuries, an expected decrease by a factor of 0.057937 units in the total volume of the corpus callosum occurs for every unit increase in elapsed time. Males experiencing traumatic brain injuries have an expected decrease by a factor of 0.063395 units for every unit increase in elapsed time.

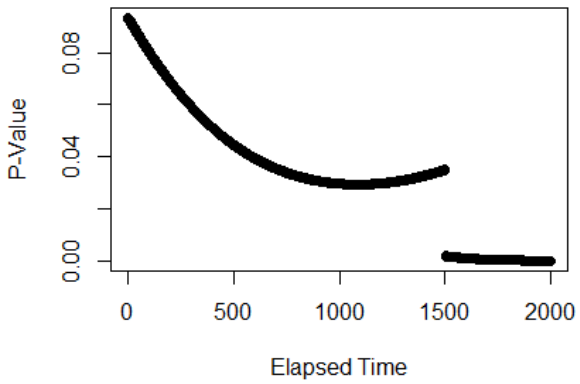
4.7 The effect of Sex

It was desired to determine whether or not any biological differences regarding patients' genders were significant. A Wald test was conducted to see if sex was significant. This test produced a p-value of 0.09232 implying that at $\alpha = 0.05$ level significance, sex is not statistically significant.

Interestingly enough, after taking the partial derivative of the model with respect to the sex variable (X_2) the following function of time was obtained.

$$\frac{\delta Y_{cc}}{\delta X_2} = \gamma_{001} + \gamma_{101} X_1 = 40.8874 + 0.0055 x_1$$

This function shows the effect that being female has at different time points. Plotting the significance of this effect at different times, yields the following plot.



As can be seen, the effect that being female has on the atrophy of the corpus callosum is not statistically significant intially (until day 416). Given that the rate of atrophy in the human body tends to not dramatically occur within a short period of time, this initial trend was expected at the beginning. The significance in the rate of atrophy between females and males appears to follows a "quadratic curve" until day 1500 where a "discontinuity" occurs, and the p-value drops to approximately 0.0019 and continues to slowly decrease.

This interesting phenomena is suspected to have occurred because of the sparcity of observations in the data set that occur after day 1000, as well as, due to the fact that males make up most of the TBI patients/recorded visits after this day.

4.8 The effect of Age Group

It was suspected that the older individuals experiencing traumatic brain injuries would experience an increased rate of atrophy; which means the total volume of corpus callosum would "shrink" faster for each increase in age group. A Wald test was conducted to explore whether or not the age group of an individual had a significant effect on the volume of Corpus Callosum. The following table summarizes the results.

Age Group	Estimated Effect	P-Value
19-33	-123.2141	0.00408
33-48	-138.5281	0.00289
49-64	-137.3872	0.00328
65-78	-146.2683	0.00663
Overall Significance of Age Group Variable		0.0508

As can be seen above, every age group obtained a fairly small marginal p-value. Despite the fact that the overall significance of the age group variable obtains a p-value over 0.05, it was considered negligible in this condition. Thus, age group still appears to be significant in the rate of atrophy.

As expected, the estimated coefficients show that as an individuals rises in age group, the more significant the expected effect on the decrease in volume of Corpus Callosum will occur. Specifically, for individuals in the 19-33-year-old age group, a decrease in the expected total volume of the corpus callosum occurs by a factor of approximately 123.21 units. For individuals in the 33-48-year-old age group, approximately 138.52 units, 137.39 units for those 49-64 and finally, 146.27 units for those 65-78.

5. Conclusion

Given all the tests and comparisons performed in this report, the selected model's fixed and random effects explained for approximately 78% of the total variation in the data (conditional R^2); 16% of which was explained by the fixed effect model (marginal R^2). Elapsed time since injury was found to be highly significant in the rate of atrophy (p-value ≤ 0.05). The rate of atrophy between the control and patient group was found to be statistically insignificant for the first 925 days post injury. After such period, the patient group rate of atrophy/volume loss was found to be significantly different than the control group. On day 925, a unit change in the group variable (control [0] to patient [1]) resulted in a decrease in the expected total volume of the corpus callosum by a factor of approximately 92.107 units. By day 2000, this decrease became a factor of approximately 180.48 units. The overall effect of sex was found to be statistically insignificant (p-value = 0.092) in the model. However, it was observed that the sexes had significant differences in expected volume loss after day 416. The age group an individual belonged to was considered to be statistically significant in the model (p-value negligible to 0.05). Belonging in the 19–33-year-old age group decreased the expected volume of the corpus callosum by a factor of approximately 123.21 units and belonging in the 65-78-year-old age group, by a factor of approximately 146.27 units. Further studies can look deeper into other possible factors affecting corpus callosum atrophy rates such as, an individual's fitness level or occupation.

References

Geifman, Nophar, et al. “Redefining Meaningful Age Groups in the Context of Disease.” AGE, vol. 35, no. 6, 2013, pp. 2357–2366., doi:10.1007/s11357-013-9510-6.

Nakagawa, Shinichi, and Holger Schielzeth. “A General and Simple Method for obtainingR2from Generalized Linear Mixed-Effects Models.” Methods in Ecology and Evolution, vol. 4, no. 2, 2012, pp. 133–142., doi: 10.1111/j.2041-210x.2012.00261.x.