BACHELOR'S DEGREE THESIS

Degree in Statistics

Title: Factors associated with the increase of infarct volume growth after mechanical thrombectomy in stroke patients

Author: Carlota Martín Pardina

Advisor: Jordi Cortés Martínez and María Hernández Pérez

Department: Statistics and Operations Research

Academic year: 2020-2021

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Abstract

The objective of this end-to-degree project is to find the factors related to the growth of cerebral infarction after mechanical thrombectomy in patients who have suffered a stroke. Different types of statistical modelling (general linear models, generalized linear models and mixed models) are used to evaluate the association between those factors and the infarct growth. The sample of the study is composed of 98 stroke patients undergone mechanical thrombectomy. Measurements of infarct volumes are taken from the neuroimages obtained in Magnetic Resonance Imaging (MRI) scans, which are performed in three different time points: pre-procedure on arrival at the hospital, immediately after the procedure and after five days.

Keywords

Linear Model, Generalized Linear Model, Repeated measures analysis, Infarct growth, Acute Stroke

Resumen

El objetivo de este trabajo de final de grado es conocer los factores relacionados con el crecimiento del infarto cerebral tras una trombectomía mecánica en pacientes que han sufrido un ictus. Diferentes tipos de modelos estadísticos (modelos lineales generales, modelos lineales generalizados y modelos mixtos) se usan para evaluar la asociación entre estos factores y el crecimiento del infarto. La base de datos utilizada tiene 98 pacientes que han sufrido un ictus y, consecuentemente, se les ha realizado una trombectomía mecánica. Las medidas de los diferentes tamaños del infarto se extraen de las neuroimágenes obtenidas mediante resonancias magnéticas, que se realizan en tres intervalos de tiempo distintos: pre-procedimiento a la llegada al hospital, justo después de la realización de la trombectomía mecánica y al cabo de 5 días.

Palabras clave

Modelo Lineal, Modelo Lineal Generalizado, Análisis de medidas repetidas, Crecimiento del infarto, Ictus agudo

Mathematics Subject Classification (MSC)

62P10 - Applications to biology and medical sciences

62J05 - Linear regression

62J12 - Generalized Linear models

Acknowledgment

I would like to express my special thanks to my advisor Jordi Cortés Martínez for his support, exceptional guidance and valuable advice. I would like to thank Dr. María Hernández Pérez, for her passion and dedication to continue investigating stroke patients and saving lives. Also, thank both of them for their confidence in the project. Finally, I would like to thank my family, for their unconditional support.

Contents

1 List of acronyms **3**

2 Introduction **4**

2.1 Background 4

2.2 Definition of stroke 4

2.3 Objectives 5

3 Methodology **7**

3.1 Variable definition 7

3.2 Descriptive analysis 7

3.3 Imputation for missing data 7

3.4 Statistical modelling 8

3.4.1 General linear models 9

3.4.2 Generalized linear models 9

3.4.3 Mixed model 10

3.5 Variable selection 10

3.6 Model outputs 11

3.7 Model validation 11

3.8 Software 11

4 Results **12**

4.1 Descriptive analysis 12

4.1.1 Univariate descriptive analysis 12

4.1.2 Bivariate descriptive analysis 14

4.1.3 Missing data 18

4.2 General linear models 19

4.2.1 Univariate models 19

4.2.2 Multivariate models with infarct growths 20

4.2.3 Multivariate models with infarct volumes 28

4.3 Generalized linear models 39

4.3.1 Multivariate models with infarct volumes 39

4.4 Mixed model 47

5 Discussion **50**

5.1 Main findings 50

5.2 Limitations 50

5.3 Conclusion 51

6 Appendix **52**

6.1 Appendix 1: Descriptive analysis 52

6.2 Appendix 2: General linear models 53

6.2.1 Multivariate models with infarct growths 53

6.2.2 Multivariate models with infarct volumes 55

6.3 Appendix 3: Generalized linear models 57

6.4 Appendix 4: Mixed model 59

Bibliography **61**

1 List of acronyms

Acronym	Description
CI	Confidence interval
EVT	Endovascular Therapy
FEI	The Spanish Stroke Federation
FURIAS	Futile Recanalization in Ischemic Acute Stroke
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random
MRI	Magnetic Resonance Imaging
NIHSS	National Institute of Health Stroke Scale
SEN	The Spanish Neurology Society
TICI	Thrombolysis in Cerebral Infarction
Tmax	Maximum time
VIF	Variance Inflation Factor

2 Introduction

2.1 Background

Stroke has a huge impact on society. In Spain, there are 71,780 new strokes every year. Indeed, it is currently the second cause of death in the general population and the first cause of woman's death, according to FEI [19]. It is also the largest cause of disability and generates a very high cost for health and social services. As reported by the SEN [1], stroke-related costs of a patient in Spain come to nearly 27,711 euros per year (including medical and indirect costs). Consequently, stroke-related costs come to approximately 1,989 million euros per year. Compared with other diseases (stroke is a clinical syndrome rather than a single disease), the consumption of health care resources is higher among stroke patients.

Moreover, two out of every three people who survive a stroke have some type of sequels, in many cases disabling, which significantly increases the need for assistance or care.

Basically, it is the leading cause of long term disability in developed countries and one of the top causes of mortality worldwide. Correct management relies on rapid diagnosis and focused treatment, thorough investigation and rehabilitation programs.

2.2 Definition of stroke

Like all organs, the brain needs the oxygen and nutrients provided by blood to function properly. If the supply of blood is restricted or stopped, brain cells begin to die, which is what happens when strokes occur. This can lead to brain injury, disability and, in the worst scenario, death [15].

Strokes can be classified into two major categories: ischemic (due to lack of blood flow, accounting for 85% of all cases) and hemorrhagic (where a weakened blood vessel supplying the brain bursts). Both cause parts of the brain to stop functioning properly.

Acute ischemic strokes, as mentioned before, result in loss of blood flow, nutrients and oxygen because one of the arteries in the brain is blocked. Consequently, the brain receives less blood than it needs to work normally. There are two revascularization procedures to restore blood flow: the intravenous fibrinolysis, which consists of dissolving the thrombus (blood clot) that occludes the artery with a medicinal product applied intravenously, and the mechanical thrombectomy, which is based on removing the thrombus with the help of a catheter. A successful treatment allows the blood to re-nourish the affected part of the brain and increases the chances that the patient will have a favorable situation after the stroke [5].

Furthermore, the severity of a stroke depends on which blood vessels are affected. A blockage or a rupture in one of the main arteries that supplies the brain will cause widespread damage and severe loss function. A blockage in a minor branch of an artery will cause less damage and a full recovery is more likely, although this can depend on the importance of the brain cells that have been damaged.

The degree of recovery that a patient can achieve after a stroke is variable and depends on the initial size of the brain lesion, the success of the procedure (evaluated using TICI grading system [8]), which parts of the brain were affected, what type of blood vessel was involved and other individual variables that are currently not well known.

It is important to mention that there are two groups of TICI grades; TICI grades less than 2b and TICI grades greater than or equal to 2b. On the one hand, TICI grades less than 2b are TICI 0 (no reperfusion), TICI 1 (perfusion past the initial obstruction, but a limited distal branch, filling with little or slow distal perfusion) and TICI 2a (perfusion of $< 50\%$ of the vascular distribution of the occluded artery). These ones are used when unsuccessful recanalization. On the other hand, TICI grades greater than or equal to 2b are TICI 2b (perfusion of $\geq 50\%$ of the vascular distribution of the occluded artery), TICI 2c (near-complete perfusion without clearly visible thrombus but with delay in contrast run-off) and TICI 3 (complete perfusion). These ones are used when successful recanalization. In conclusion, greater TICI grades mean more success of the procedure (less severe brain injury).

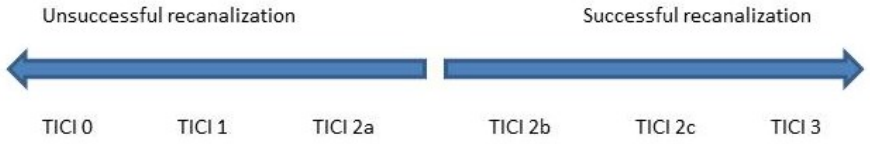


Figure 1: TICI grading system

Recanalization is sometimes not achieved even after multiple passes of the catheter. Whether revascularization becomes futile or harmful with an increasing number of passes as well as criteria for when to halt attempting recanalization remain unknown. Despite this, the final conclusion of a study of the American Heart Association was that the likelihood of successful recanalization got sequentially lower as the number of passes increased. Besides, it is known that the number of passes needed to achieve target vessel recanalization modifies the outcome after mechanical thrombectomy and successful recanalization after a single pass is associated with favorable outcome.

2.3 Objectives

In order to describe the goals of this end-of-degree project, it is necessary to explain what the FURIAS project consists of. The FURIAS project was designed to study the radiological factors associated with futile recanalization after endovascular treatment in acute stroke. Its general objective is to find signs in the neuroimages obtained in MRI scans that can help predict the success of the procedure and the degree of clinical recovery of stroke patients.

Consequently, a prospective observational study is designed in which 3 serial MRI scans are performed on 98 stroke patients undergone mechanical thrombectomy. From the MRI scans a detailed picture of the brain is obtained. It is important to mention that all the neuroimages are de-identified, pseudomized and stored for analysis. An expert reader, blinded to the clinical data, is in charge of segmenting by hand hyperintensities on diffusion weighted imaging creating lesions marks on MRI scans (Figure 2 [14]).

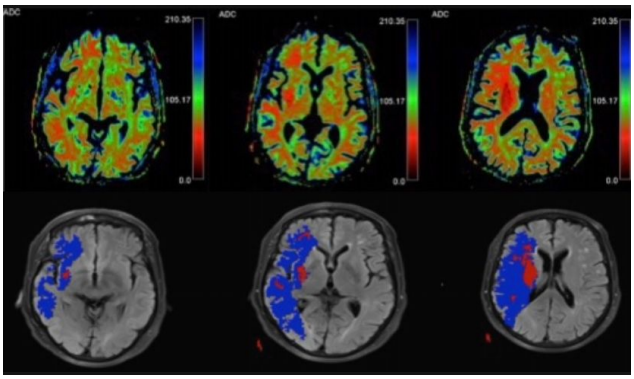


Figure 2: MRI scans of acute stroke patients

In the first MRI scan, done on arrival at the hospital (Infarct volume pre EVT) it can be assessed the size of the initial infarction. The second MRI is performed after mechanical thrombectomy where it can be seen the size of the infarct after the procedure (Infarct volume post EVT). In addition, a third MRI performed 5 days after the mechanical thrombectomy (Infarct volume on day 5), shows the final infarction.

- Initial infarct size = Infarct volume pre EVT
- Post-procedure infarct size = Infarct volume post EVT
- Final infarct size = Infarct volume on day 5

First, the difference between infarct volume on day 5 and infarct volume pre EVT is called *total infarct growth*. Second, the difference between infarct volume post EVT and infarct volume pre EVT is called *early infarct growth*. Last, the difference between infarct volume on day 5 and infarct volume post EVT is called *late infarct growth*.

- Early infarct growth = post - pre
- Late infarct growth = 5d - post
- Total infarct growth = 5d - pre = Early infarct growth + Late infarct growth

The specific objective of this end-of-degree project, which is a substudy framed in the FURIAS project, is to study the evolution of the size of the infarct (evaluated as diffusion lesion) at different time points (pre-procedure on arrival at the hospital, immediately after the procedure and after five days) and to understand which factors are associated with infarct volume growth.

3 Methodology

3.1 Variable definition

Variable	Description
Good collaterals	Good collateral circulation
HIR pre	Hypoperfusion intensity ratio
Infarct volume pre EVT	Initial infarct size
Infarct volume post EVT	Post-procedure infarct size
Infarct volume on day 5	Final infarct size
Number of passes	Number of passes of the catheter needed to achieve target vessel recanalization
TICI grade	TICI grading system
TMax6 volume post	Time to maximum > 6 volume

3.2 Descriptive analysis

A descriptive analysis of the baseline patient's characteristics is performed. For categorical variables, frequencies and percentages are provided. For symmetric continuous variables, mean and standard deviation are shown. However, if the variable is skewed, median and IQR (first and third quartiles) are provided. In addition, some visual representations (box plots, scatter plots, to name a few) of the distribution of infarct volumes are presented.

Furthermore, a numerical and graphical descriptive analysis are performed for the infarct and hypoperfusion volumes and infarct growths stratifying by TICI grade.

The correlation between infarct growths and relevant continuous variables are calculated using the (non-parametric) Spearman's correlation coefficient. On the contrary, for dichotomous variables, Pearson's correlation coefficient is shown (as there are only two categories, polyserial and Pearson's correlations match).

3.3 Imputation for missing data

Leaving out subjects with some missing data deprives of some amount of information. For this reason, a multiple missing imputation, which is a technique for dealing with missing data, is used. Basically, it imputes the data with plausible values based on the information provided by the other variables.

It is important to mention that missing data are typically grouped into three categories [10]:

- MCAR: if the probability of being missing is the same for all cases. This effectively implies that the causes of the missing data are unrelated to the data. When data are MCAR, the complete case analysis performed on the data is unbiased; however, is generally regarded as a strong and often unrealistic assumption because data are rarely MCAR.
- MAR: if the probability of being missing is the same only within groups defined by the observed data. Modern missing data methods generally start from the MAR assumption. The complete data can provide, but not necessarily, asymptotically unbiased estimates.

- MNAR: if the probability of being missing varies for reasons that are unknown; the missingness is related to factors which are not measured by the researcher. Basically, it is data that is neither MAR nor MCAR.

There are multiple ad-hoc solutions, such as a listwise deletion (complete case analysis), pairwise deletion, mean imputation, regression imputation and multiple imputation, to name a few. Assuming the data is not MCAR, multiple imputation was applied in our work.

Multiple imputation always follows three steps: imputation, analysis and pooling (Figure 3, inspired by [13]). Multiple imputation creates several complete versions of the data by replacing the missing values by plausible data values. These plausible values are drawn from a distribution specifically modeled for each missing entry, this step is known as *imputation*. At the end of this step, there are m completed datasets. The m imputed datasets are identical for the observed data entries, but differ in the imputed values. Then these datasets are analyzed in the *analysis* step. Then estimates the parameters of interest from each imputed dataset and pools those parameters into one estimate in the *pooling* step. Under the appropriate conditions, the pooled estimates are unbiased and have the correct statistical properties [4].

In this particular case, assuming the data is not MNAR, MICE [3] (Multiple imputation by chained equations), which is a particular multiple imputation technique, is used. It is important to mention that this robust method, imputes missing data in a dataset through an iterative series of predictive models. In each iteration, each specified variable in the dataset is imputed using the other variables in the dataset. It is a very flexible method and can handle variables of varying types.

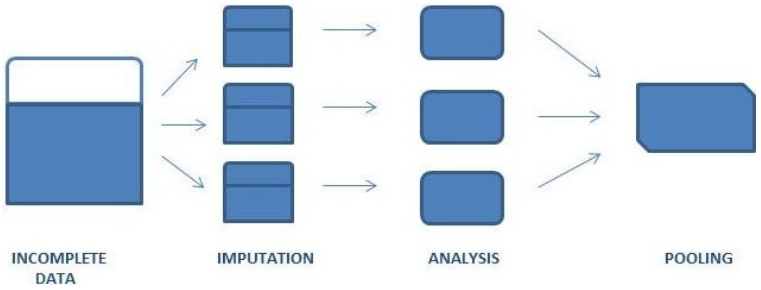


Figure 3: Steps of multiple imputation

3.4 Statistical modelling

To begin with, to study the evolution of the size of the infarct at different time points, general linear models are first applied due to their simplicity. Within these, the infarct growths are first modeled, as requested by the researcher, but the analysis adjusting by baseline values is more efficient than the one that considers the difference over time as the response [17]. For this reason, the infarct volumes, applying the logarithmic transformation, are modeled. Then, the generalized linear models are used to see if the logarithmic transformation could be spared to make the models easier to interpret. Finally, a mixed model is implemented to consider the correlations between the measures corresponding to the same patient. Also, a logarithmic transformation to the infarct volumes is done. Mixed models provide an analytical power that cannot be achieved using the previously indicated methods.

Since logarithms are applied in general linear models using infarct volumes and in the mixed model, the effects on the response variable are multiplicative. In contrast, for the remaining models, the effects are additive.

3.4.1 General linear models

The general linear model usually refers to conventional linear regression, which is a statistical method used to describe the relationship between a continuous random variable Y, known as the response variable, and a group of k non-random variables X, known as predictors. The general equation for a linear model is:

$$Y = X\beta + \epsilon$$

$$Y = (y_1, ..., y_n)'$$

$$\beta = (\beta_0, \beta_1, ..., \beta_k)'$$

$$\epsilon = (\epsilon_1, ..., \epsilon_n)'$$

Where Y is a vector of observations, X is a matrix of known covariates, β is a vector of the coefficients and ϵ is a vector of error terms.

In the general linear model, the following is assumed:

$$\epsilon_i \sim N(0, \sigma^2)$$

Basically, the general linear model is used when the response variable and error term follow a Normal distribution, the relationship between response and explanatory variable is linear, errors are homoscedastic and non-correlated and, finally, independent explanatory variables are not correlated.

3.4.2 Generalized linear models

The generalized linear model is a flexible generalization of the general linear model so that the dependent variable is linearly related to the factors and covariates via a specified link function. The model also allows for the dependent variable to have a non-normal distribution. Basically, a generalized linear model consists of three components [12]:

- A linear predictor which is a linear function of regressors:

$$\eta = X\beta$$

- A link function (g) which is a bijective function that relates the expected value to the linear predictor:

$$g(\mu) = \eta = X\beta$$

$$\mu = g^{-1}(\eta) = g^{-1}(X\beta)$$

- An exponential family of probability distributions:

$$f_{Yi} = (y; \theta_i; \phi) = e^{\frac{\theta_i y - b(\theta_i)}{a(\phi)} + c(y,\phi)}$$

Both statistical methods, general and generalized linear models, relate some predictors to a single outcome variable. The main difference between them is that the general linear models assume that the residuals follow a conditionally normal distribution, while the generalized linear models loosens this assumption and allows for a variety of other distributions from the exponential family for the residuals, so the sampling distribution of the residuals need to be specified. In this study, Gamma family and link identity are used. In the particular case that the link function is the identity function and the distribution of errors is normal generalized and general linear models provide the same fit.

3.4.3 Mixed model

Linear mixed models are an extension of simple linear models to allow both fixed and random effects. These models are particularly useful when analyzing repeated measures in which the response is continuous and measured in multiple occasions at several time points. As observations are collected from the same individuals over time, it may be reasonable to assume that there is some kind of correlation among the observations from the same individual. Linear mixed models are useful for modeling the correlations among the observations of the same patient. Therefore, a random intercept mixed model is done because time is considered to be categorical (time points depend on the duration of each patient's procedure). The general equation for a linear mixed model is [2]:

$$Y = X\beta + Zu + \epsilon$$

Where Y is a vector of observations, X is a matrix of known covariates associated with the fixed effects, β is a vector of the fixed effects, Z is a matrix associated with the random effects, u is a vector of random effects and ϵ is a vector of errors.

3.5 Variable selection

In the multivariate linear models with infarct growths or volumes, relevant variables in the univariate models are candidates to be included in the multivariate models. Each of these variables and their interactions with the TICI grade are tested using the F test (function anova, package stats for R), which provided the final model.

For the selection of variables in the multivariate generalized linear models with infarct volumes, same procedure mentioned above is carried out, but using the deviance test. The variables that caused convergence problems with the model are not included.

Furthermore, for the selection of variables in the mixed model, relevant variables from the other models are included. Nevertheless, those variables post-procedure for which baseline values are not available are not included in the model.

Automatic selection methods such as stepwise are avoided because of their propensity to provide biased results [7].

It is worth mentioning that the number of passes is dichotomized in general linear models using infarct growth, as requested by the researcher, and for the generalized linear models. In contrast, for the remaining models, the number of passes is considered numerical.

Finally, the Variance Inflation Factor is calculated for each variable in each model to measure the collinearity. For the VIF, a cutoff point of three was determined to detect high collinearities.

3.6 Model outputs

For each model, a table with all the coefficients obtained and their 95% CI are provided. To ease the interpretation, an analogue table with these coefficients untransformed according to the data response is also given, if pertinent. For continuous predictors, expected response values are given for the minimum, the median and the maximum values of the predictor's value in the sample. For the factors, expected response values are presented for each category. In both variable types, predictions are provided for each category of the response. In addition, forest plots for the 95% CI of the associations between predictors and response are shown. Finally, graphics pointing out the relationship of the response with each predictor are displayed.

3.7 Model validation

To validate all the models, the graphic of residuals versus fitted values is provided. This plot is used to detect non-linearity, unequal error variances and outliers. First, it can be assumed that the relationship is linear if the residuals are randomly distributed around the residual line. Second, it can be accepted that the variances of the error terms are equal if the residuals roughly (raw residuals in general linear models and Pearson residuals in the generalized linear models) form a horizontal band along the axis of fitted values. Finally, it can be regarded that there are no outliers if any residual stands out from the basic random pattern of residuals.

In case of the presence of outliers, the Cook's distance; which is a measure to detect if there are observations that have a real influence on the fit, the studentized residuals; which are a measure to detect outliers, and the hat-values; whose aim is to highlight a priori influential observations, are provided. Besides, an outlier test is done to verify that these observations are indeed outliers.

Additionally, the coefficient of determination (R-squared) is provided for each model. This coefficient measures the proportion of the total variability of the endogenous variable is explained by the model. Basically, it gives some information about the goodness of fit of a model. For the generalized linear models, the Cragg-Uhler R-squared is provided instead.

3.8 Software

To carry out the analysis, the statistical software that is used in this end-of-degree project is R (R version 3.6.2). In order to turn the analyses into a report, R Markdown is used. Also, different packages, such as *jtools* [11], *mice* [18], *emmeans* [16] and *effects* [9], related to the modelling and visualization of data are used and specified in the code. The code used is uploaded in GitHub and is provided in following link:

https://github.com/carlotapmartin/TFG_FURIAS_study

4 Results

First, an univariate and a bivariate descriptive analysis are done. Besides, the summary of the general linear models, the generalized linear models and mixed models, that are done to evaluate the evolution of infarct growth or volumes, are shown.

The several types of models are organized as presented in the following table:

Section	**Type of model**	**Response**
4.2.2	General Linear model	Infarct growths
4.2.3	General Linear model	Infarct volumes
4.3	Generalized Linear model	Infarct volumes
4.4	Mixed model	Infarct volumes (all time points)

4.1 Descriptive analysis

4.1.1 Univariate descriptive analysis

A descriptive analysis of the baseline patient's characteristics is performed. Approximately 58% of the patients are men, which indicates that the patients are quite balanced according to gender. The age of the patients ranges from 38 to 92 years with a mean age of 69.83 years. In this study, the prevalence of diabetes is 16.3% and 67 patients (68.4%) have hypertension. In addition, more than half of the patients have the occlusion located in the MCA M1 segment. It is also important to mention that only 14 (14.29%) patients have TICI grade less than 2b, which means that the recanalization was not successful (Table 1).

Table 1: Baseline characteristics

	Total sample (n = 98)
Age (years)	69.8 ± 13.2
Sex (male)	57 (58.2 %)
Smoking	25 (25.5 %)
Hypertension	67 (68.4 %)
Dyslipidemia	42 (42.9 %)
Diabetes	16 (16.3 %)
Atrial fibrillation	30 (30.6 %)
Stroke etiology	
Atheroembolic	32 (32.7 %)
Cardioembolic	38 (38.8 %)
Undetermined	28 (28.6 %)
Onset to admission (min)	257 [141 , 438]
Baseline NIHSS	18 [12 , 21]
Baseline Glycemia	116 [101 , 142]
Wake up stroke	25 (25.5 %)

Table 1: Baseline characteristics *(continued)*

	Total sample (n = 98)
Site of occlusion	
Intracranial carotid	10 (10.2 %)
Tandem	23 (23.5 %)
MCA M1 segment	51 (52 %)
MCA M2 segment	14 (14.3 %)
Previous intravenous alteplase	48 (49 %)
Onset to groin (min)	320 [236 , 537]
Conscious sedation	94 (95.9 %)
TICI grade	
TICI 0	7 (7.1 %)
TICI 2a	7 (7.1 %)
TICI 2b	20 (20.4 %)
TICI 2c	29 (29.6 %)
TICI 3	35 (35.7 %)

[1] For quantitative variables, ± is used for showing mean ± standard deviation in symmetric variables

[2] For the remaining variables, median [IQR] is provided instead

Infarct volume on day 5 has a higher median and variability than those infarct volumes observed before and immediately after the procedure. For instance, the median is almost 12 mL for infarct volume pre EVT, 20 mL for infarct volume post EVT and 25 mL for infarct volume on day 5 (Figure 4).

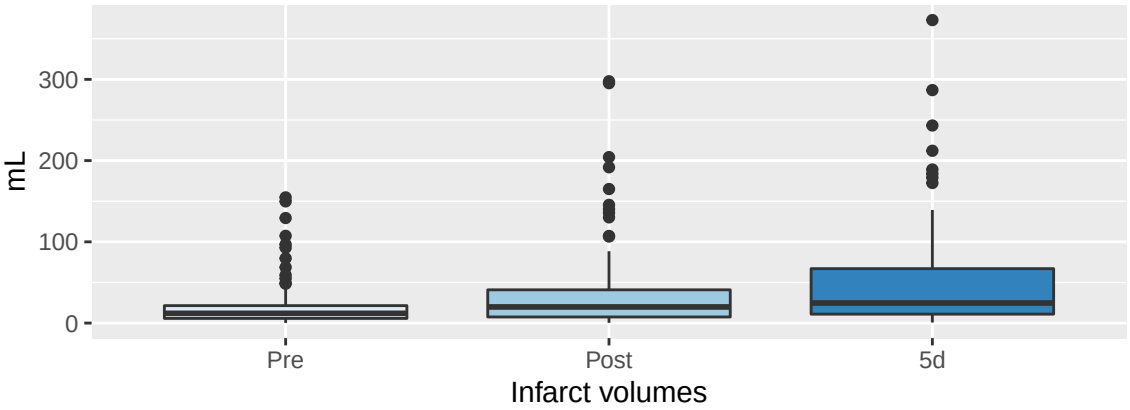


Figure 4: Distribution of infarct volumes (pre-procedure, post-procedure and on day 5)

For patients with TICI grades less than 2b, the volume of the infarct growth tends to grow over the three time points. However, for TICI grades greater than or equal to 2b, the volume of the infarct growth tends to remain constant (Figure 5). An alternative representation can be found in the Appendix 1 (Figure 43), which shows infarct volumes for each TICI grade (not dichotomized).

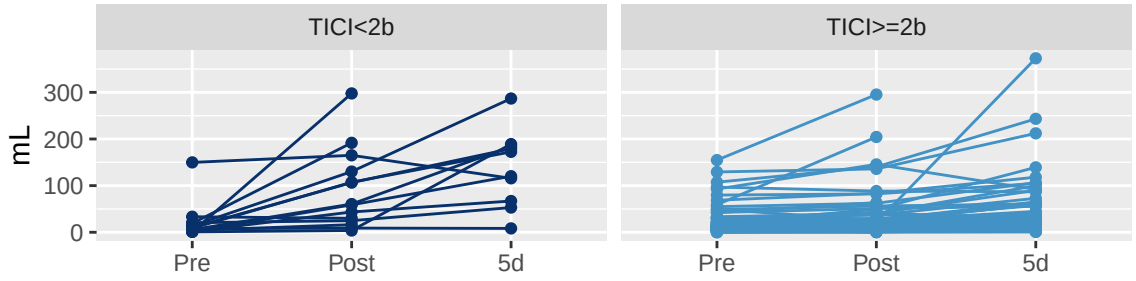


Figure 5: Infarct volumes for each patient stratified by TICI grade

On the top three plots of figure 6, it can be seen that the variability of the points is not constant along the regression line, which means that there is heteroscedasticity. To correct this, infarct volumes are log-transformed (see bottom plots of figure 6) in the general linear models.

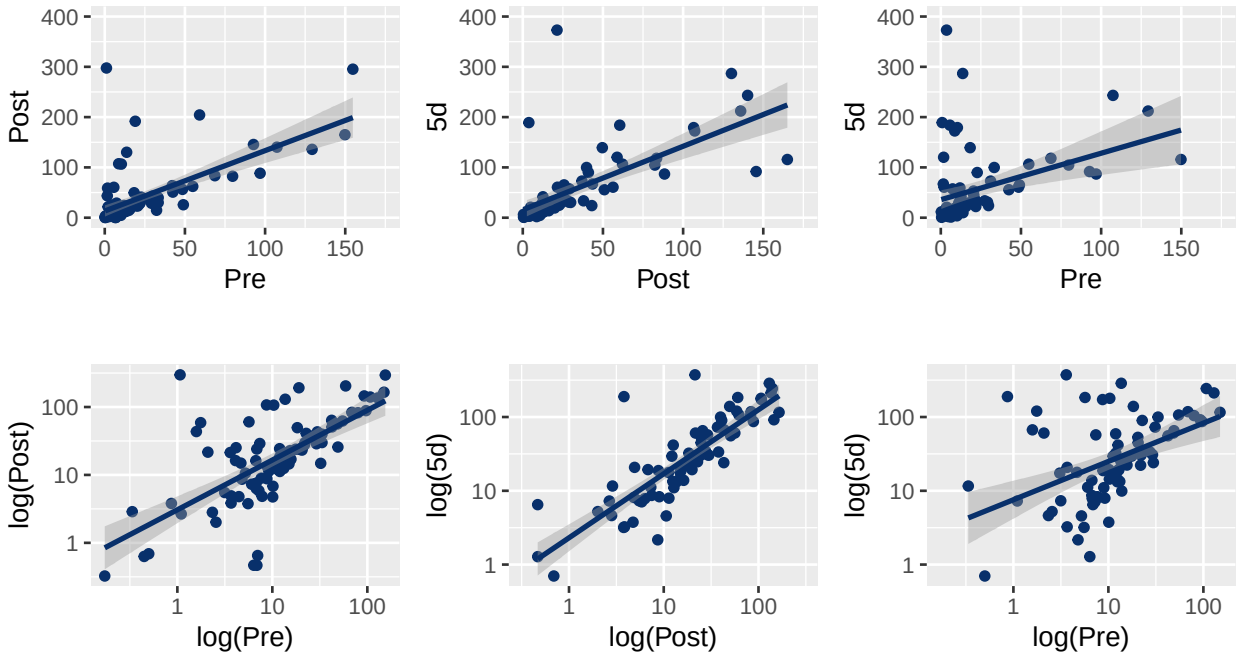


Figure 6: Scatter plots of infarct volumes (pre-procedure, post-procedure and on day 5)

4.1.2 Bivariate descriptive analysis

To start with, it is important to mention that the variable TICI grade is used in table 2 because it is a relevant variable from a clinical point of view because it measures the success of the recanalization.

In order to check if patients in each TICI grade category are comparable at baseline, a non parametric test (Kruskal-Wallis) is made to prove that there are no significant differences in the medians of infarct volume

pre EVT and the Tmax6 volume pre among the TICI grade. The resulted p-values are 0.223 and 0.824, for the infarct volume pre EVT and the Tmax6 volume pre, respectively. In both cases, these results lead to the non-rejection region of the null hypothesis. Consequently, there is no evidence of a different distribution according to TICI grade.

Nevertheless, the last three rows of table 2 show that there is a major difference in infarct growths (mL) for TICI grades less than 2b. Kruskal-Wallis tests resulted significant with p-values equal to 0.001, 0.012, 0.004 for the early, late and total infarct growths, respectively. Patients with TICI grades less than 2b had more severe injuries with higher infarct growth after the procedure.

Table 2: Infarct and hypoperfusion volumes and infarct growths (mL) between the three time points according to TICI grade

	TICI 0 (n=7)	**TICI 2a (n=7)**	**TICI 2b (n=20)**	**TICI 2c (n=29)**	**TICI 3 (n=35)**	**All (n=98)**
Infarct volume pre EVT	8 [1 , 11]	10 [5 , 20]	12 [5 , 22]	18 [7 , 48]	9 [7 , 15]	12 [6 , 21]
Tmax6 volume pre	124 [72 , 136]	88 [70 , 134]	59 [39 , 98]	134 [80 , 174]	92 [48 , 135]	96 [53 , 141]
Infarct volume post EVT	107 [26 , 148]	59 [27 , 84]	13 [4 , 25]	29 [22 , 61]	11 [5 , 21]	20 [8 , 41]
TMax6 volume post	154 [134 , 156]	90 [37 , 164]	2 [0 , 13]	3 [0 , 16]	0 [0 , 8]	3 [0 , 27]
Infarct volume 5 days	144 [79 , 185]	150 [103 , 180]	25 [8 , 40]	56 [16 , 89]	18 [8 , 25]	25 [11 , 67]
Early infarct growth	42 [9 , 108]	55 [8 , 77]	1 [-1 , 7]	7 [3 , 20]	2 [0 , 5]	4 [0 , 12]
Late infarct growth	44 [5 , 134]	67 [53 , 85]	6 [1 , 26]	4 [2 , 37]	4 [0 , 7]	5 [0 , 29]
Total infarct growth	115 [17 , 182]	144 [97 , 171]	10 [2 , 20]	9 [2 , 46]	6 [0 , 12]	9 [1 , 33]

Note:
Median and [IQR] are provided

Moreover, an explanatory analysis is done, prior to the construction of the models, to see if the explicative variables are related to infarct growths by means of a correlation analysis (Table 3).

For instance, a significant positive correlation is found between early infarct growth and the number of passes. As well as, late infarct growth and Tmax volume post. This last variable also is positively correlated with total infarct growth. Another significant positive correlation is found between total infarct growth and occlusion post.

On the contrary, a significant negative correlation is found between all infarct growths and TICI grades greater than or equal to 2b (Table 3).

Table 3: Correlation between infarct growths (mL) and relevant variables

Variable	**Early infarct growth**		**Late infarct growth (All the sample)**		**Late infarct growth (Only TICI>=2b)**		**Total infarct growth**	
	Coefficient	**p**	**Coefficient**	**p**	**Coefficient**	**p**	**Coefficient**	**p**
Age (years)	0.01	0.926	0.158	0.181	0.13	0.311	0.131	0.258
Sex (male)	0.092	0	-0.142	0.012	-0.079	0.003	-0.072	0
Baseline Glycemia	0.302	0.005	0.189	0.116	0.146	0.263	0.255	0.028

Table 3: Correlation between infarct growths (mL) and relevant variables *(continued)*

	Coefficient	**p**	**Coefficient**	**p**	**Coefficient**	**p**	**Coefficient**	**p**
Baseline NIHSS	0.324	0.002	0.198	0.092	0.174	0.172	0.243	0.033
Infarct volume pre EVT			0.074	0.536	0.254	0.045		
TMax6 volume pre	0.15	0.188	0.126	0.326	0.182	0.189	0.12	0.335
HIR pre	0.195	0.088	0.145	0.262	0.191	0.166	0.19	0.127
Good collaterals	-0.366	0.001	0.099	0.459	0.047	0.744	0.031	0.812
Time onset to recanalization	0.131	0.212	0.100	0.399	0.124	0.332	0.148	0.199
Time image to recanalization	0.275	0.008	0.324	0.006	0.256	0.046	0.334	0.004
Number of passes	0.425	<0.001	0.293	0.012	0.264	0.038	0.344	0.002
Complications	0.166	0.112	-0.083	0.486	0.033	0.8	-0.104	0.367
TICI>=2b	-0.49	<0.001	-0.292	0.012			-0.461	<0.001
Occlusion post			0.392	0.001	0.366	0.003	0.436	<0.001
TMax6 volume post			0.548	<0.001	0.505	<0.001	0.523	<0.001
HIR post			-0.271	0.029	-0.226	0.088	-0.214	0.087

Patients with TICI grades less than 2b have a higher median and variability in early infarct growth. For instance, for early infarct growth, the median is almost 42 mL when TICI grade is 0, 55 mL when TICI grade is 2a and not higher than 7 mL for the other greater TICI grades. The same pattern is observed with late and total infarct growths (Figure 7).

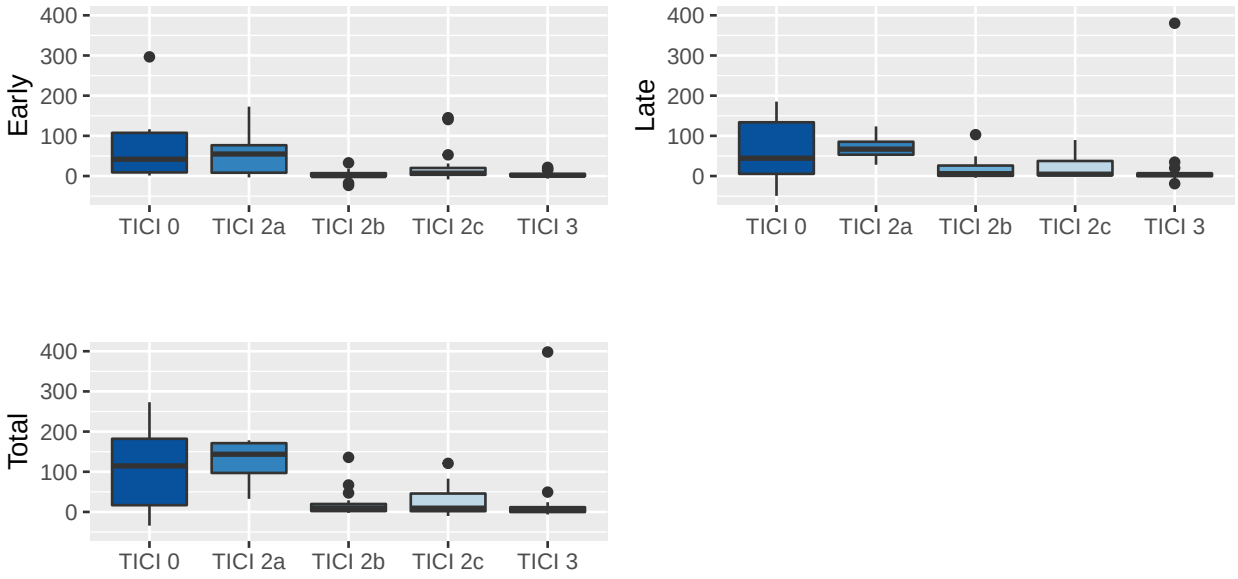


Figure 7: Distribution of infarct growths at the three time points according to TICI grade

The motivation of figure 8 is to assess the contribution of each infarct growth (early and late) to the total infarct growth. This figure shows that for both, early and late infarct growths, the slopes are quite similar between the two groups of TICI grades. However, there is a different intercept; meaning that a patient without differences between the infarct volume pre EVT and post will have a higher late infarct growth if the TICI grade is less than 2b. Alternative representations of figure 8 can be found in the Appendix 1 (Figure 42).

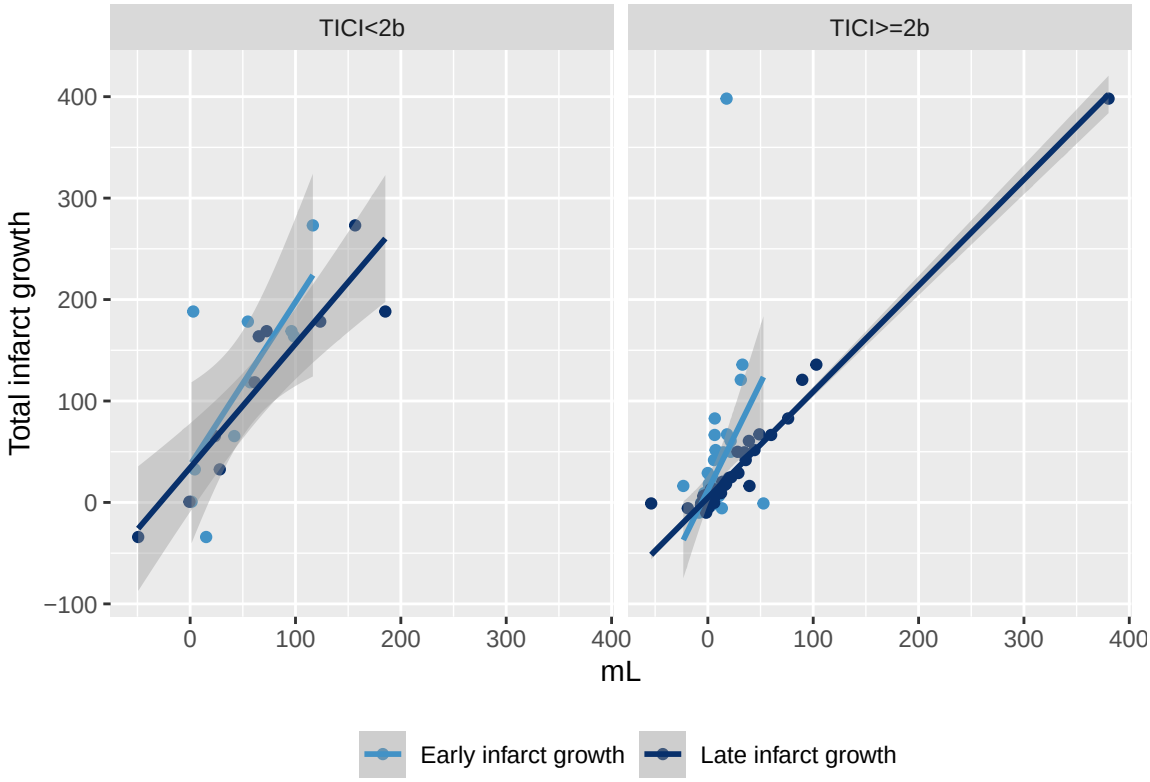


Figure 8: Relationship between the total infarct growth with the early and late infarct growths

In order to know if the slopes, for each TICI grade, in figure 8 can be considered different, a parametric paired t-test is made using the logarithmic transformation. This test proves that there are no significant differences between the slopes stratifying for TICI grades because 1 is not included in either of the two confidence intervals.

For TICI grades less than 2b, the following 95% CI for the ratio of the slopes is obtained:

$$95\%\ CI_{TICI<2b} = [0.219, 1.002]$$

For TICI grades greater than or equal to 2b, the following 95% CI is obtained:

$$95\%\ CI_{TICI\geq 2b} = [0.332, 2.149]$$

4.1.3 Missing data

It is worth to mention that there is missing data on the variables that are used for the models. For this reason, it is interesting to make a descriptive analysis of them (Figure 9). For all models, except for those made using infarct growths, the missing values are imputed.

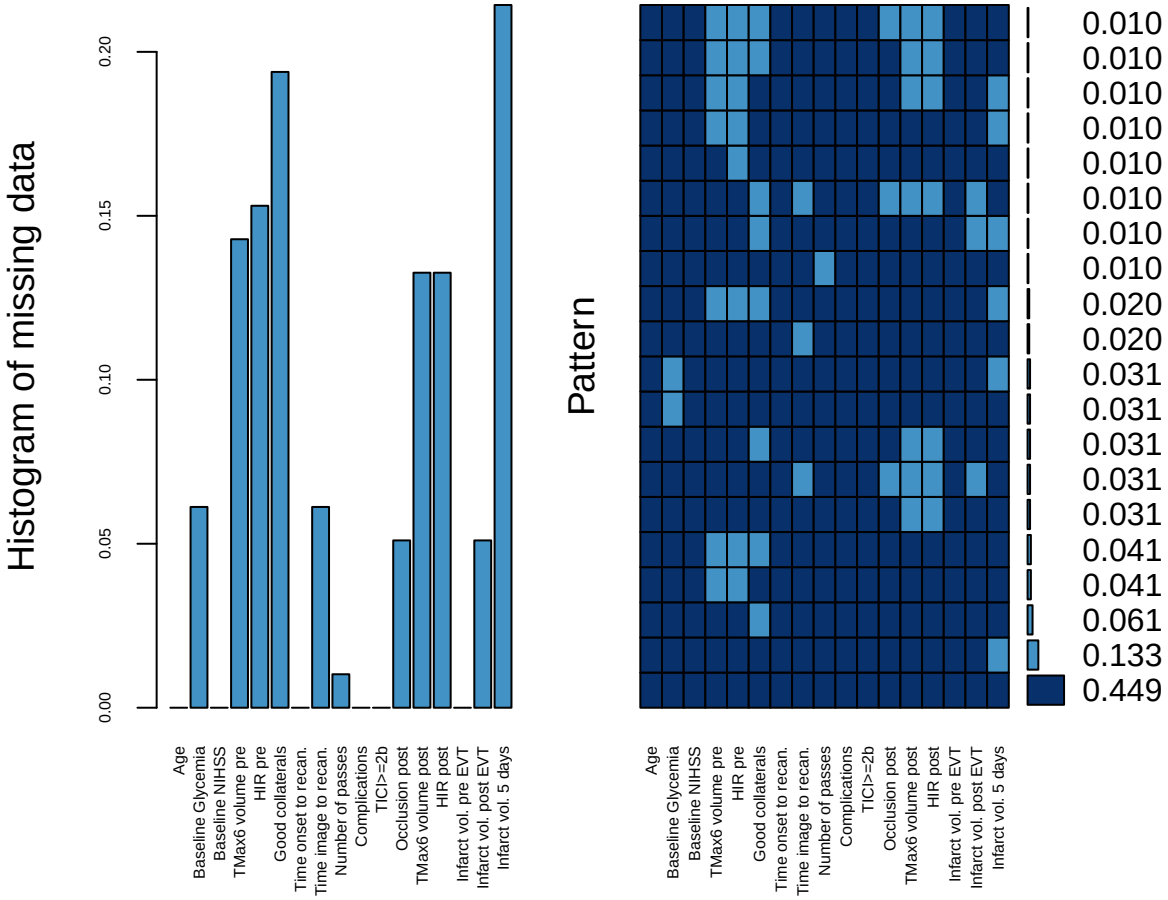


Figure 9: Missing data patterns

Almost 45% of the patients have complete information. Moreover, almost 22% and 19.4% of the values of infarct volume on day 5 and of the variable *Good collaterals* are missing. This last variable also has the highest number of missing values among the relevant predictors (Table 4).

Despite this, there are some relevant variables that have complete information, such as infarct volume pre EVT, age, sex and TICI grade, to name a few.

It is important to mention that the variable that has more missing values is infarct volume on day 5 because some patients died within 5 days after the mechanical thrombectomy. In those cases, one option would have been to assign the worst value of infarct volume to these patients. Nevertheless, the final decision is not to impute the values in the response variables.

Table 4: Number and percentage of missing values by variable (first column) and for each univariate model with infarct growth response

		Early infarct growth	Late infarct growth (All the sample)	Late infarct growth (Only TICI>=2b)	Total infarct growth
	Missings	Missings	Missings	Missings	Missings
Age (years)	0 (0 %)	5 (5.1 %)	25 (25.51 %)	21 (21.43 %)	21 (21.43 %)
Sex (male)	0 (0 %)	5 (5.1 %)	25 (25.51 %)	21 (21.43 %)	21 (21.43 %)
Baseline Glycemia	6 (6.12 %)	11 (11.22 %)	28 (28.57 %)	23 (23.47 %)	24 (24.49 %)
Baseline NIHSS	0 (0 %)	5 (5.1 %)	25 (25.51 %)	21 (21.43 %)	21 (21.43 %)
Infarct volume pre EVT	0 (0 %)		25 (25.51 %)	21 (21.43 %)	
TMax6 volume pre	14 (14.29 %)	19 (19.39 %)	35 (35.71 %)	30 (30.61 %)	31 (31.63 %)
HIR pre	15 (15.31 %)	20 (20.41 %)	36 (36.73 %)	30 (30.61 %)	32 (32.65 %)
Good collaterals	19 (19.39 %)	22 (22.45 %)	40 (40.82 %)	33 (33.67 %)	37 (37.76 %)
Time onset to recanalization	0 (0 %)	5 (5.1 %)	25 (25.51 %)	21 (21.43 %)	21 (21.43 %)
Time image to recanalization	6 (6.12 %)	7 (7.14 %)	27 (27.55 %)	23 (23.47 %)	27 (27.55 %)
Number of passes	1 (1.02 %)	6 (6.12 %)	26 (26.53 %)	22 (22.45 %)	22 (22.45 %)
Complications	0 (0 %)	5 (5.1 %)	25 (25.51 %)	21 (21.43 %)	21 (21.43 %)
TICI>=2b	0 (0 %)	5 (5.1 %)	25 (25.51 %)		21 (21.43 %)
Occlusion post	5 (5.1 %)		26 (26.53 %)	22 (22.45 %)	26 (26.53 %)
TMax6 volume post	13 (13.27 %)		33 (33.67 %)	26 (26.53 %)	33 (33.67 %)
HIR post	13 (13.27 %)		33 (33.67 %)	26 (26.53 %)	33 (33.67 %)

4.2 General linear models

4.2.1 Univariate models

To start with, a univariate analysis is done to detect which are the predictors that are statistically relevant.

In table 5, one can see that in the early infarct growth the values of Occlusion post, TMax6 volume post and HIR post, are not included in the univariate analysis because variables post-procedure cannot be taken into account. Moreover, the variable infarct volume pre EVT it is also not included in the early and total infarct growths because the endogenous variable of these models is calculated based on this variable.

The variable TICI grade resulted significant in all the univariate models because its p-value is lower than the fixed 5% significance level, meaning that the pertinent coefficient is statistically different from zero, indicating that it is a relevant variable.

Also, time image to recanalization has a significant parameter in early and total infarct growth models because the resulted p-values are lower than the fixed 5% significance level. However, for late infarct growth, the resulted p-value is 0.066, which is slightly higher than the significance level but very close to it. Given the above, this variable is considered relevant for these three models.

Table 5: Univariate analysis for early, late and total infarct growth

	Early infarct growth		Late infarct growth (All the sample)		Late infarct growth (Only TICI>=2b)		Total infarct growth	
	Coefficient	**p**	**Coefficient**	**p**	**Coefficient**	**p**	**Coefficient**	**p**
Age (years)	0.49 [-0.21 , 1.18]	0.17	-0.12 [-1.16 , 0.91]	0.812	-0.44 [-1.49 , 0.6]	0.398	0.04 [-1.15 , 1.23]	0.948
Sex (male)	-8.18 [-26.65 , 10.28]	0.381	16.58 [-10.71 , 43.88]	0.23	-8.36 [-35.31 , 18.6]	0.538	10.15 [-21.96 , 42.25]	0.531
Baseline Glycemia	0 [-0.24 , 0.25]	0.986	0.23 [-0.13 , 0.58]	0.207	0.17 [-0.17 , 0.5]	0.326	0.25 [-0.15 , 0.65]	0.225
Baseline NIHSS	1.15 [-0.68 , 2.98]	0.216	0.76 [-1.9 , 3.43]	0.569	0.11 [-2.51 , 2.73]	0.933	1.42 [-1.55 , 4.39]	0.343
Infarct volume pre EVT			-0.12 [-0.57 , 0.33]	0.595	0.19 [-0.29 , 0.67]	0.432		
TMax6 volume pre	0.05 [-0.04 , 0.15]	0.248	0.06 [-0.08 , 0.21]	0.398	0.07 [-0.06 , 0.21]	0.281	0.03 [-0.13 , 0.2]	0.689
HIR pre	55.15 [9.17 , 101.12]	0.019	23.12 [-50.34 , 96.58]	0.531	35.25 [-35 , 105.5]	0.319	13.71 [-67.09 , 94.52]	0.736
Good collaterals	-36.46 [-57.9 , -15.02]	0.001	12.69 [-21.39 , 46.76]	0.459	5.55 [-28.36 , 39.46]	0.744	4.47 [-32.84 , 41.77]	0.812
Time onset to recanalization	0 [-0.03 , 0.03]	0.816	0.01 [-0.03 , 0.06]	0.56	0.03 [-0.02 , 0.07]	0.212	0.02 [-0.03 , 0.07]	0.39
Time image to recanalization	0.29 [0.1 , 0.48]	0.003	0.27 [-0.02 , 0.55]	0.066	0.3 [-0.01 , 0.6]	0.058	0.38 [0.04 , 0.73]	0.028
Number of passes	-19.1 [-37.78 , -0.42]	0.045	-17.1 [-44.85 , 10.65]	0.223	-18.67 [-45.65 , 8.32]	0.172	-30.86 [-62.37 , 0.64]	0.055
Complications	18.93 [-4.5 , 42.35]	0.112	-12.36 [-47.5 , 22.79]	0.486	-1.59 [-110.26 , 107.08]	0.977	-18.33 [-58.55 , 21.89]	0.367
TICI>=2b	-60.42 [-82.8 , -38.05]	<0.001	-48.4 [-85.93 , -10.87]	0.012			-93.06 [-134.22 , -51.89]	<0.001
Occlusion post			46.07 [20.26 , 71.87]	0.001	41.59 [14.28 , 68.89]	0.003	62.35 [31.71 , 92.98]	<0.001
TMax6 volume post			0.69 [0.44 , 0.94]	<0.001	1.31 [0.96 , 1.66]	<0.001	0.94 [0.67 , 1.2]	<0.001
HIR post			-15.22 [-51.63 , 21.18]	0.407	-9.94 [-44.64 , 24.76]	0.568	-19.09 [-61.14 , 22.96]	0.368

4.2.2 Multivariate models with infarct growths

As mentioned in the methodology section, the variables that are statistically relevant in the univariate analysis are candidates to enter in the multivariate models. Each of these variables and their interactions with the TICI grade are tested using the F test, which provided the final models.

Table 6 shows the estimations and the 95% CI of the parameters of each variable. More detailed information of each of these models is included in the Appendix 2.

Table 6: Coefficient point estimates and 95% CI of the multivariate general linear models using infarct growths

	Early infarct growth	**Late infarct growth**	**Total infarct growth**
HIR pre	54.75 [13.36, 96.14]		
Time image to recanalization		-1.59 [-2.44, -0.73]	-1.90 [-2.73, -1.07]
Number of passes	-77.93 [-139.27, -16.59]		
TMax6 volume post		0.23 [-0.27, 0.74]	0.08 [-0.41, 0.57]
TICI>=2B	-70.64 [-98.88, -42.39]	-338.23 [-500.74, -175.73]	-460.3 [-617.48, -303.12]
Time image to recanalization:TICI>=2B		1.70 [0.80, 2.60]	2 [1.13, 2.87]
TMax6 volume post:TICI>=2B		1.05 [0.42, 1.68]	1.39 [0.78, 1.99]
Number of passes:TICI>=2B	76.15 [11.56, 140.74]		

First, for the early infarct growth model, the R-squared obtained is 0.33; this indicator measures the proportion of the total variability of the endogenous variable explained by the model. Besides, 21 observations are deleted because missing values are found. This model has the following relevant variables: HIR pre and the

interaction between the number of passes with the TICI grade. There is no problem of multicollinearity in this model because all the VIFS are lower than 3; the highest VIF found is 1.06, which is from the variable number of passes.

Second, for the late infarct growth model, the R-squared obtained is 0.57. Again, 35 observations are deleted because missing values are found. This model has the following relevant variables: the interactions between time image recanalization and TMax6 volume post, both with the TICI grade. There is no problem of multicollinearity in this model because the highest VIF found is 2.11, which is from the variable TICI grade.

Last, for the total infarct growth model the R-squared obtained is 0.70. Besides, 35 observations are deleted because missing values are found. This model has the same relevant variables than the previous model (the late infarct growth model) and has also no problem of multicollinearity.

Furthermore, an interpretation and graphical representation of all the effects of each regression model using infarct growths is made in table 7 and figure 10.

Table 7: Early, late and total infarct growths predictions

Variable	**Value**	**TICI**	**Early infarct growth**		**Late infarct growth**		**Total infarct growth**	
			Estimate	**Group**	**Estimate**	**Group**	**Estimate**	**Group**
HIR pre	0.855		49.63 [25.1, 74.1]	a				
	0.375		23.37 [7.2, 39.5]	b				
	0.01		3.39 [-19.7, 26.5]	c				
Number of passes	More passes	TICI<2b	80.28 [55.24, 105.3]	a				
	One pass		2.34 [-53.65, 58.3]	b				
	More passes	TICI>=2b	9.64 [-3.45, 22.7]	a				
	One pass		7.86 [-7.22, 22.9]	a				
Time image to rec.	245	TICI<2b			-56.2 [-140.46, 28.1]	a	-10.1 [-91.6, 71.4]	a
	102				170.9 [89.18, 252.6]	b	261.8 [182.7, 340.8]	b
	45				261.4 [139.41, 383.4]	c	370.1 [252.1, 488.1]	c
	245	TICI>=2b			48.9 [10.48, 87.3]	a	54.5 [17.3, 91.6]	a
	102				32.7 [20.96, 44.4]	a	40.4 [29.1, 51.8]	a
	45				26.2 [5.38, 47]	a	34.8 [14.7, 55]	a
TMax6 post	207	TICI<2b			193.58 [110.83, 276.3]	a	253.42 [173.38, 333.5]	a
	3				146.16 [65.28, 227]	a	236.49 [158.26, 314.7]	a
	0.01				145.46 [63.63, 227.3]	a	236.24 [157.1, 315.4]	a
	207	TICI>=2b			266.73 [194.45, 339]	a	308.65 [238.74, 378.6]	a
	3				5.56 [-6.22, 17.3]	b	8.91 [-2.48, 20.3]	b
	0.01				1.73 [-10.5, 14]	c	4.52 [-7.31, 16.3]	c

For the early infarct growth model, that the variable HIR pre has significant differences between the minimum, the median and the maximum. Higher values of HIR pre are associated with higher values of early infarct growth (Table 7).

Moreover, the variable number of passes has significant differences between doing one or more passes if TICI grades are less than 2b; more passes involve higher values of early infarct growth.

For the regression models of late and total infarct growths, the variable time image to recanalization has significant differences between the minimum, the median and the maximum when TICI grades are less than 2b. Consequently, higher values of time image to recanalization come with lower values of late and total infarct growths when TICI grades are less than 2b.

The variable TMax6 post has significant differences between the minimum, the median and the maximum when TICI grades are greater than or equal to 2b. Higher values of TMax6 post show association with higher values of late and total infarct growths when TICI grades are greater than or equal to 2b.

In addition, a graphical representation of all the associations in the regression models using infarct growths is done in figure 10. It is important to mention, again, that there are only 14 patients that have a non successful recanalization, which means that have TICI grades less than 2b.

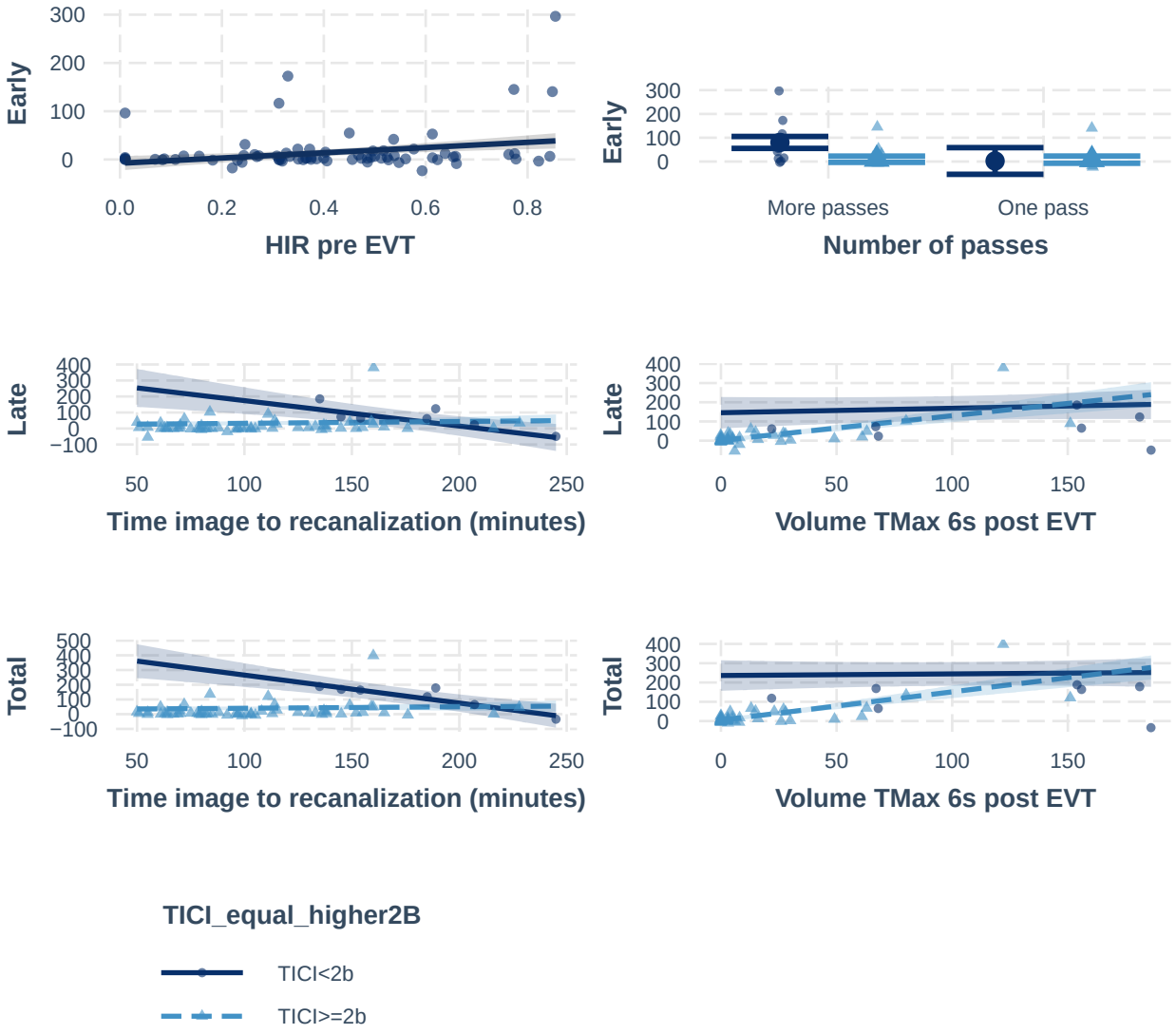


Figure 10: Association plots for the early, late and total infarct growth models

The validation of each regression model using infarct growths is made below.

First, to validate the early infarct growth model, the graphic of residuals versus fitted values serves to assess heteroscedasticity. In this particular case, as shown in figure 11, there seems to be a slightly heteroscedastic pattern. Besides, there is a particular a curvilinear pattern, which indicates that the model may be missing a higher order term to fix the nonlinearity. However, if an older term of HIR pre is applied, the curvilinear pattern is not corrected. For this reason, this older term is not applied.

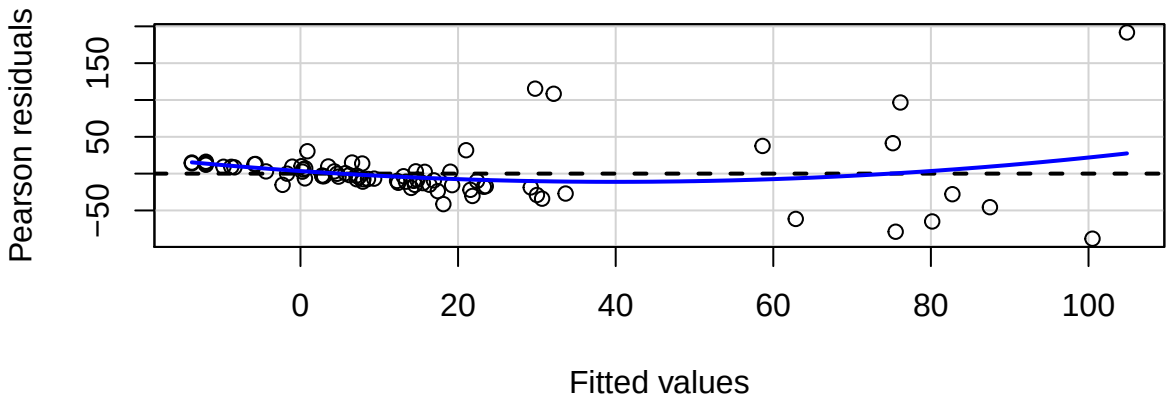


Figure 11: Residuals vs fitted values of early infarct growth model

Figure 12 shows the Cook's distance, which is a measure to detect observations that have a real influence on the fit, the studentized residuals, which are a measure to detect outliers and the hat-values, whose aim is to highlight a priori influential observations. Observation 31 has the highest Cook's distance and studentized residuals in this model and observations 25 and 49 have the highest hat-values.

An outlier test is done to show if some of these observations should be considered outliers. Observation 31 should be considered an outlier because the resulted Bonferroni p-value is lower than the fixed 5% significance level (Table 8).

Table 8: Outlier test for the early infarct growth model

	Rstudent	**Unadjusted p-value**	**Bonferroni p**
31	6.634907	5.447e-09	4.1942e-07

It is important to mention that this patient (observation 31) had no lesion at all (0 mL) when she arrived at the hospital. Nonetheless, the cerebral artery rapidly infarcted right after, creating an abnormally higher early infarct growth.

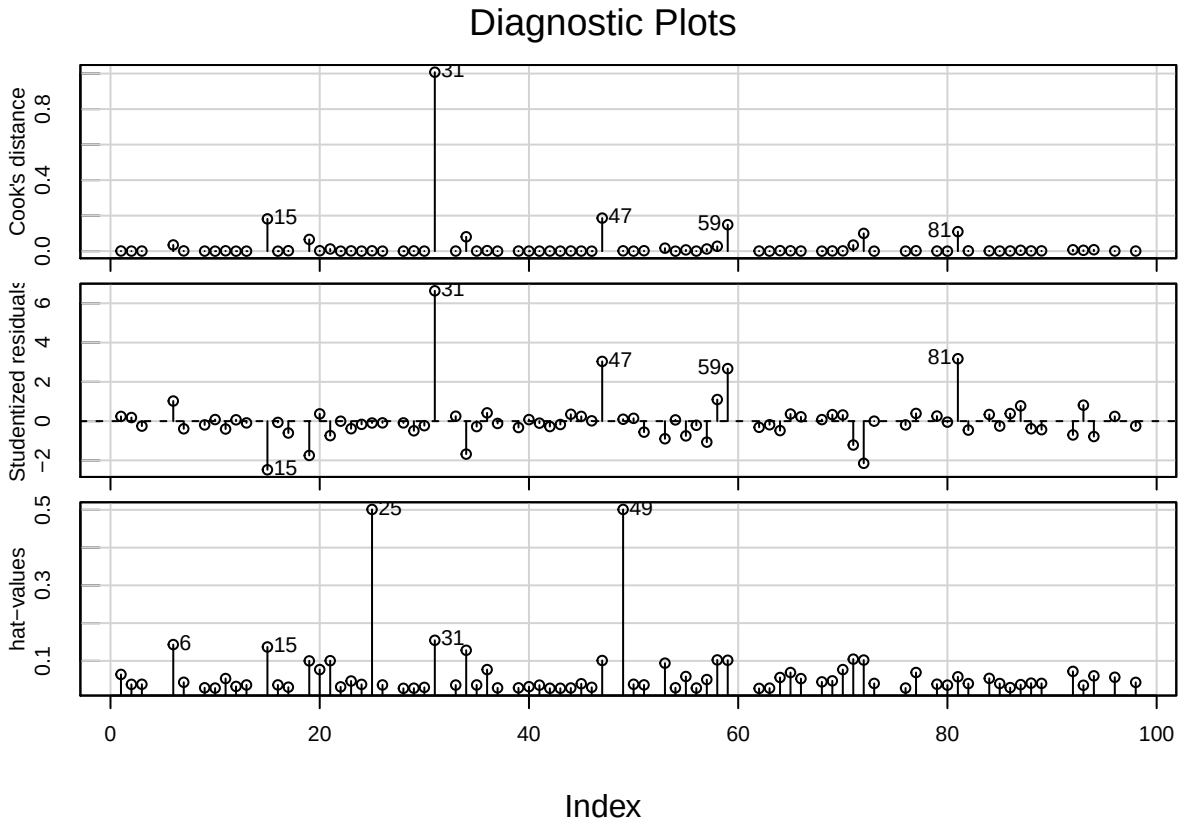


Figure 12: Diagnostic plots for early infarct growth model

Second, for the late infarct growth model, as shown in figure 13, there seems to be a slightly heteroscedastic pattern. Besides, there is a particular a curvilinear pattern, which indicates that there seems to be nonlinearity. However, no variable in this model has a significant parameter that indicates adding a higher order term in order to fix that.

Figure 14 shows that observations 10 and 87 have the highest Cook's distance and studentized residuals in this model. For this reason, outlier tests are done to show if these observations should be considered outliers. The resulted Bonferroni p-values are lower than the fixed 5% significance level. This means, that observation 10 and 87 should be considered outliers (Table 9).

Table 9: Outlier test for the late infarct growth model

	Rstudent	**Unadjusted p-value**	**Bonferroni p**
10	11.527795	2.0633e-16	1.2999e-14
87	-3.827615	3.2846e-04	2.0693e-02

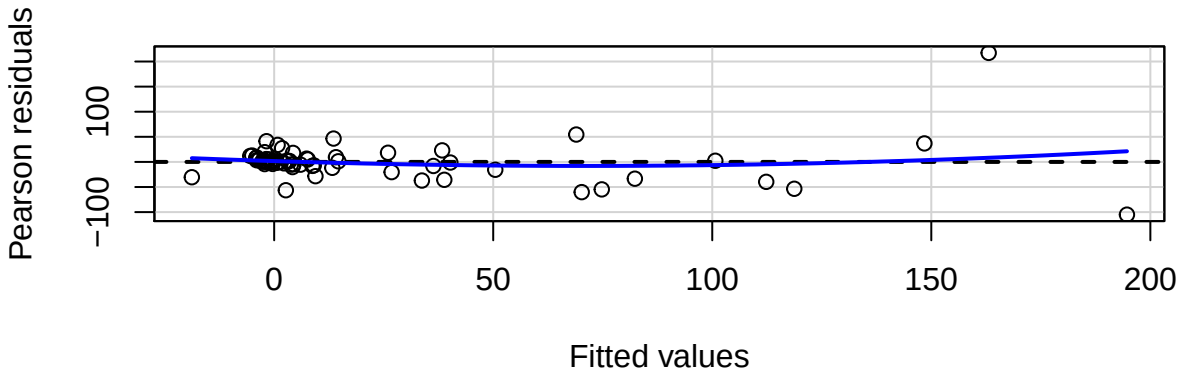


Figure 13: Residuals vs fitted values for late infarct growth

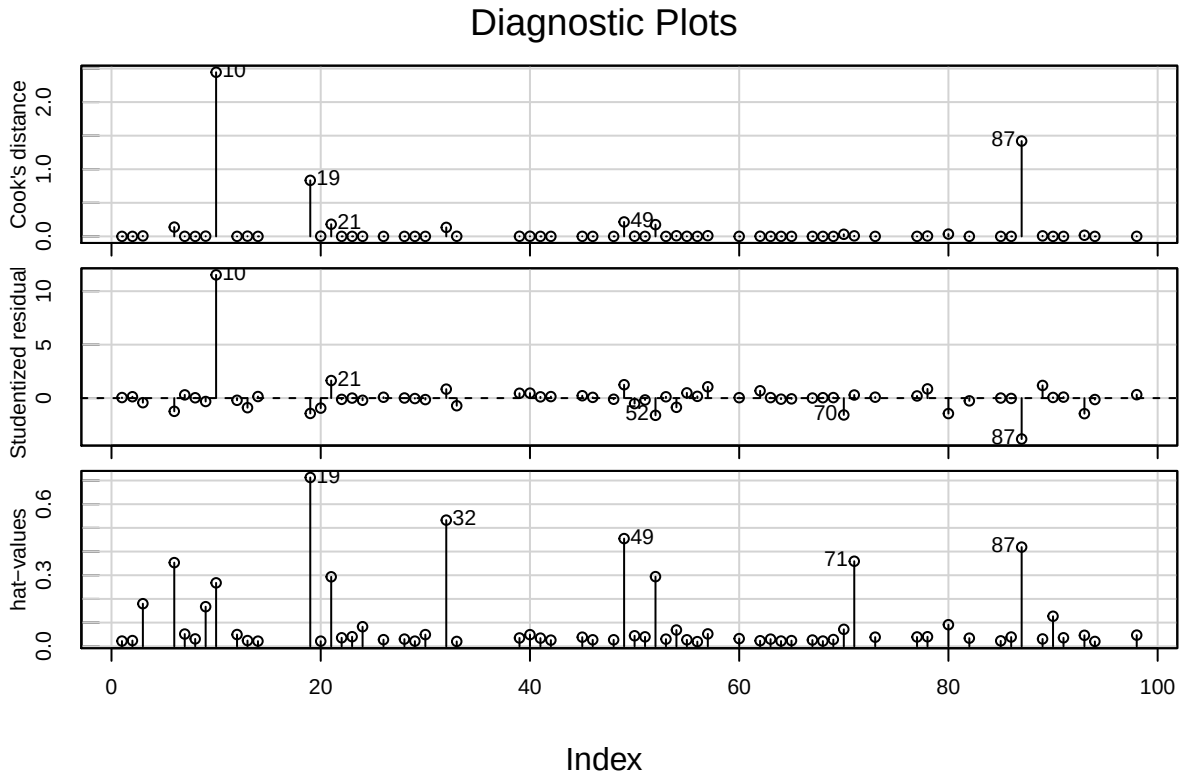


Figure 14: Diagnostic plots for late infarct growth

Last, to validate the total infarct growth model, there is a graphical tool that shows if there is heteroscedasticity. In this case, the residuals have no obvious patterns or systematic structure, which means that linearity might be assumed. Despite that, it seems there is a slightly heteroscedastic pattern (Figure 15).

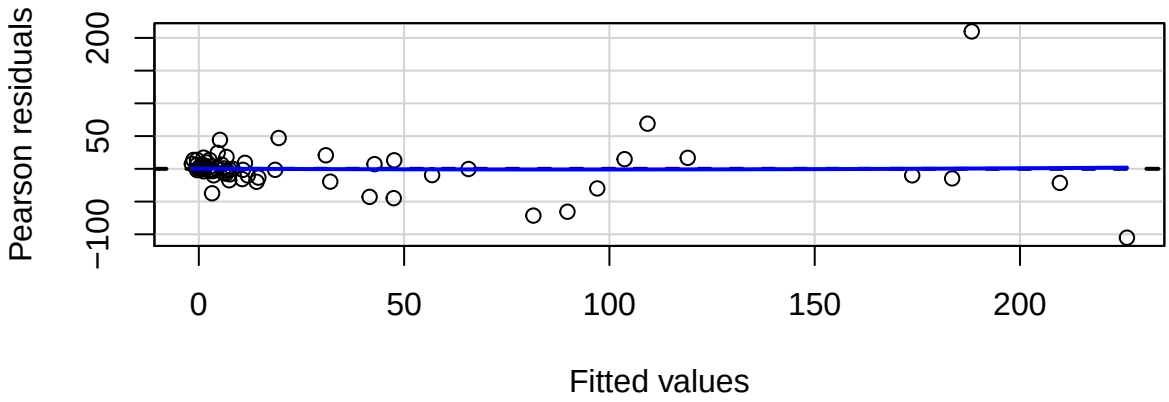


Figure 15: Residuals vs fitted values for total infarct growth

There is an outlier (extreme Y value) which may be underlying a special cause. Figure 16 and table 10 shows that observations 10 and 87 have the highest Cook's distance and studentized residuals in this model. For this reason, outlier tests are done to show if these observations should be considered outliers. The resulted Bonferroni p-values are lower than the fixed 5% significance level. This means, that observation 10 and 87 should be considered outliers (Table 10).

Table 10: Outlier test for the total infarct growth model

	Rstudent	**Unadjusted p-value**	**Bonferroni p**
10	11.479285	2.4387e-16	1.5364e-14
87	-3.998065	1.8882e-04	1.1896e-02

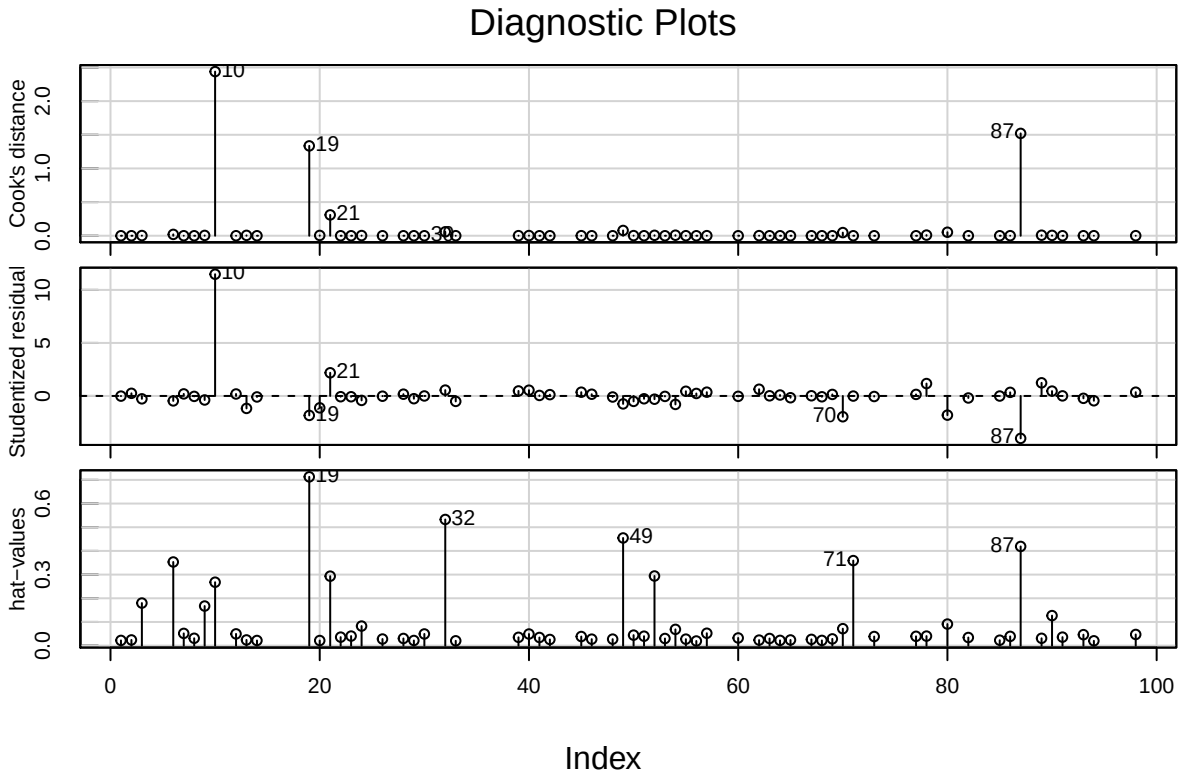


Figure 16: Diagnostic plots for the total infarct growth

To sum up, the forest plots, which draw the association between factors and each infarct growth, of the early, late and total models are shown in figures 17, 18 and 19.

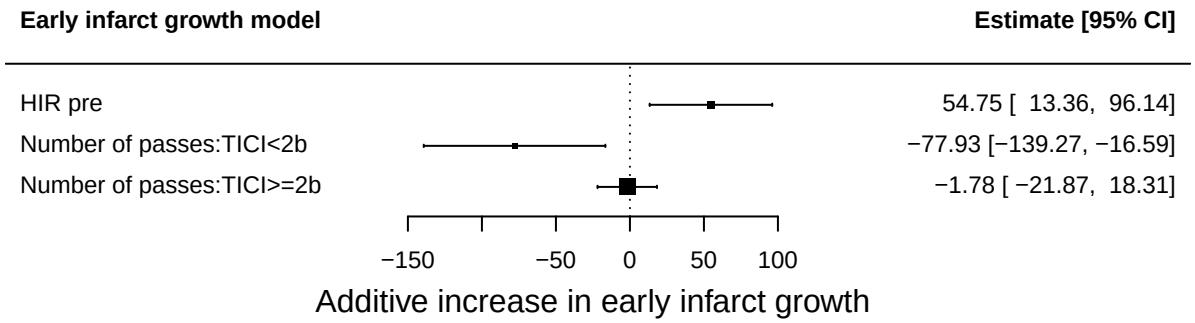


Figure 17: Forest plot of early infarct growth

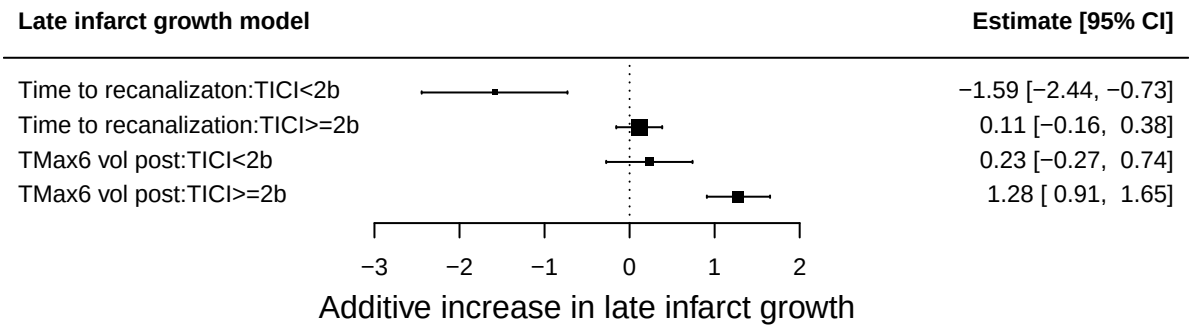


Figure 18: Forest plot of late infarct growth

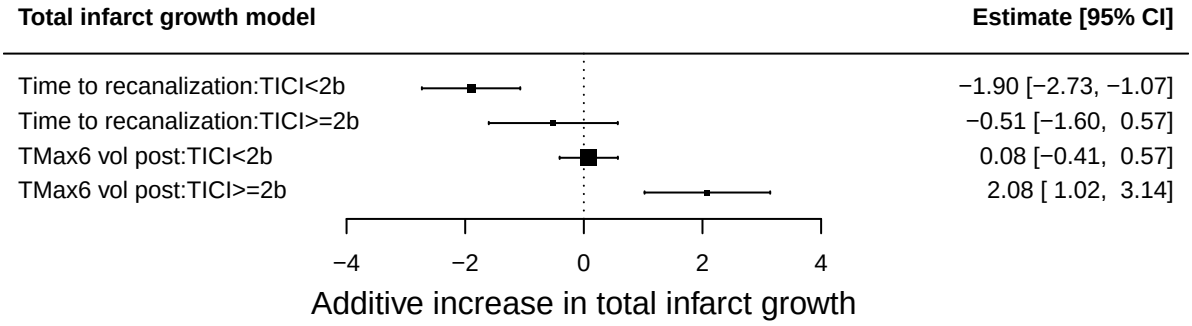


Figure 19: Forest plot of total infarct growth

4.2.3 Multivariate models with infarct volumes

First of all, univariate models are made to explore relationships between infarct volumes (Table 11).

Table 11: Coefficient point estimates and 95% CI of the univariate models using only infarct volumes

	Post vs Pre	**5d vs Post**	**5d vs Pre**
log(Infarct volume pre EVT)	0.73 [0.56, 0.90]		0.52 [0.3, 0.74]
log(Infarct volume post) EVT		0.87 [0.73, 1.00]	

Table 12 shows the coefficient point estimates and 95% CI of the parameters of each variable. More detailed information about each of these models is included in the Appendix 2.

Table 12: Coefficient point estimates and 95% CI of the multivariate models using infarct volumes

	Post vs Pre	**5d vs Post**	**5d vs Pre**
log(Infarct volume pre EVT)	0.18 [-0.11, 0.47]		-0.01 [-0.39, 0.38]
log(Infarct volume post EVT)		0.36 [0.01, 0.72]	
TICI>=2B	0.06 [-1.64, 1.77]	-2.82 [-4.76, -0.88]	-3.77 [-5.05, -2.49]
Number of passes	0.47 [0.26, 0.69]		0.18 [0.08, 0.28]
Time image to recanalization	0.011 [0.002, 0.019]		
TMax6 volume post		-0.003 [-0.01, 0.005]	-0.003 [-0.012, 0.007]
log(Infarct vol. pre EVT):TICI>=2B	0.71 [0.39, 1.02]		0.70 [0.29, 1.12]
Number of passes:TICI>=2B	-0.34 [-0.58, -0.09]		
Time image to recanalization:TICI>=2B	-0.012 [-0.022, -0.003]		
TMax6 volume post:TICI>=2B		0.014 [0.005, 0.023]	0.021 [0.009, 0.033]
log(Infarct vol. post EVT):TICI>=2B		0.45 [0.06, 0.83]	

First, for the model of *Post vs Pre*, the R-squared obtained is 0.75. Compared with the previous model for the early infarct growth, which had an R-squared of 0.33, this one seems to have a better proportion of the total variability of the endogenous variable explained by the model. Besides, 5 observations are deleted because missing values in the infarct volumes post-procedure are found. However, this model has lesser missing values because a previous imputation of missing values for the predictors is performed. This model has the following relevant variables: the interactions between the logarithm of the infarct volume pre EVT, the number of passes and the time image to recanalization, each one with the variable TICI grade.

Second, for the model of *5d vs Post*, the R-squared obtained is 0.79. Compared with the previous model for the late infarct growth, which had an R-squared of 0.57, this one seems to have a better proportion of the total variability of the endogenous variable explained by the model. Besides, 25 observations are deleted because missing values in the infarct volumes, post EVT and on day 5, are found (imputation of missing values for the predictors is also performed). This model has the following relevant variables: the interactions between the logarithm of the infarct volume post EVT and the TMax6s volume post, both with the variable TICI grade.

Last, for the model of 5d vs Pre the R-squared obtained is 0.71. Compared with the previous model for the total infarct growth, which had an R-squared of 0.7, they seem to have the same proportion of the total variability of the endogenous variable explained. Besides, 21 observations are deleted because missing values are found in the infarct volumes on day 5 (imputation of missing values for the predictors is also performed). This model has the following relevant variables: the number of passes and the interactions between the logarithm of the infarct volume pre EVT and the variable TMax6 volume post EVT, both with the variable TICI grade.

Neither of these three models has problems of multicollinearity (VIFs are calculated).

In order to facilitate the interpretation, table 13 contains the exponentiated coefficients.

Table 13: Exponentiated coefficient point estimates and 95% CI of the multivariate models using infarct volumes

	Post vs Pre	**5d vs Post**	**5d vs Pre**
log(Infarct volume pre EVT)	1.19 [0.89, 1.59]		0.99 [0.68, 1.46]
log(Infarct volume post EVT)		1.44 [1.01, 2.06]	
TICI>=2B	1.06 [0.19, 5.86]	0.06 [0.01, 0.41]	0.02 [0.01, 0.08]
Number of passes	1.61 [1.29, 2]		1.2 [1.09, 1.32]
Time image to recanalization	1.011 [1.002, 1.02]		
TMax6 volume post		0.997 [0.99, 1.005]	0.997 [0.988, 1.007]
log(Infarct vol. pre EVT):TICI>=2B	2.03 [1.48, 2.78]		2.02 [1.33, 3.05]
Number of passes:TICI>=2B	0.71 [0.56, 0.91]		
Time image to recanalization:TICI>=2B	0.988 [0.978, 0.998]		
TMax6 volume post:TICI>=2B		1.014 [1.005, 1.023]	1.021 [1.009, 1.033]
log(Infarct vol. post EVT):TICI>=2B		1.56 [1.06, 2.29]	

For regression model *Post vs Pre*, an increase of 10% in infarct volume pre EVT is related to a punctual increase of 1.7%, with a 95% CI of [-1.06%, 4.54%], in infarct volume post EVT for TICI grades less than 2b. Even so, an increase of 10% in infarct volume pre EVT corresponds to an increase of 8.78% [7.45%, 10.14%] in infarct volume post EVT if TICI grades are equal or greater than 2b. The point estimates are obtained as follows:

$$1.19^{log(1.1)} = 1.017$$

$$exp(0.18 + 0.71)^{log(1.1)} = 1.0878$$

The 95% CI are calculated taking into account the information of the variance-covariance matrix of the coefficients and assuming normality.

Likewise, an increase of one pass is associated with an increase of 60.64% [29.06%, 99.96%] in the infarct volume post but only for patients with TICI less than 2b. However, it is related to an increase of 14.51% [2.33%, 28.14%] in infarct volume post EVT if TICI grades are equal or greater than 2b. The point estimates are obtained as follows:

$$exp(0.47) = 1.6064$$

$$exp(0.47 - 0.34) = 1.1451$$

Also, an increase of 5 minutes in time image to recanalization is associated with an increase of 5.54% [1.13%, 10.15%] in the infarct volume post but only for patients with TICI less than 2b. On the contrary, it is related to a decrease of 0.73% [-1.65%, 3.05%] in infarct volume post EVT if TICI grades are equal or greater than 2b. The point estimates are obtained as follows:

$$exp(0.011 \cdot 5) = 1.0554$$

$$exp[5 \cdot (0.011 - 0.012)] = 0.9927$$

For regression model *5d vs Post*, an increase of 10% in infarct volume post EVT is associated with an increase of 3.54% [0.05%, 7.15%] in infarct volume on day 5 if TICI grades are less than 2b. Although, it is related to an increase of 8.04% [6.66%, 9.42%] in infarct volume on day 5 if TICI grades are equal or greater than 2b.

Moreover, an increase of one unit in TMax6 volume post is associated with a decrease of 0.25% [-0.49%, 0.99%] in infarct volume on day 5 if TICI grades are less than 2b. However, it is related to an increase of 1.15% [0.59%, 1.74%] in infarct volume on day 5 if TICI grades are equal or greater than 2b.

For regression model *5d vs Pre*, an increase of 10% in infarct volume pre EVT associated with a decrease of 0.06% [-3.66%, 3.65%] in infarct volume on day 5 if TICI grades are less than 2b. Even so, it is related to an increase of 6.85% [5.28%, 8.45%] in infarct volume on day 5 if TICI grades are equal or greater than 2b.

Likewise, an increase of one pass is associated with an increase of 19.80% [8.53%, 32.35%] in infarct volume post EVT when infarct volume pre EVT remains constant. It is important to mention that this can only be accepted for patients that have between one and three passes.

Furthermore, an increase of one unit in TMax6 volume post is related to a decrease of 0.27% [-0.69%, 1.22%] in infarct volume on day 5 if TICI grades are less than 2b. However, it is associated with an increase of 1.84% [1.17%, 2.51%] in infarct volume on day 5 if TICI grades are equal or greater than 2b.

Besides, an interpretation and graphical representation of all the effects of each regression model using infarct volumes is made in tables 14 and 15 and figure 20.

As the number of passes in the *5d vs Pre* model is the only effect that does not interact with the TICI grade, is shown in a separate table (Table 14).

Table 14: Volume predictions on day 5 for the 5d versus Pre model according to some different number of passes

Variable	**Value**	**5d vs Pre**	
		Estimate	**Group**
Number of passes	10	230.44 [96.54, 555.573]	a
	2	54.6 [29.67, 99.48]	b
	1	45.6 [24.05, 85.63]	c

For the regression model *5d vs Pre*, the variable number of passes has significant differences between the minimum, the median and the maximum.

The volume predictions for the three models according to the TICI grade for those variable that presented interaction are shown in table 15.

For the *Post vs Pre* and *5d vs Pre* models, the variable infarct volume pre EVT has significant differences between the minimum, the median and the maximum when TICI grades are greater than or equal to 2b (Table 15).

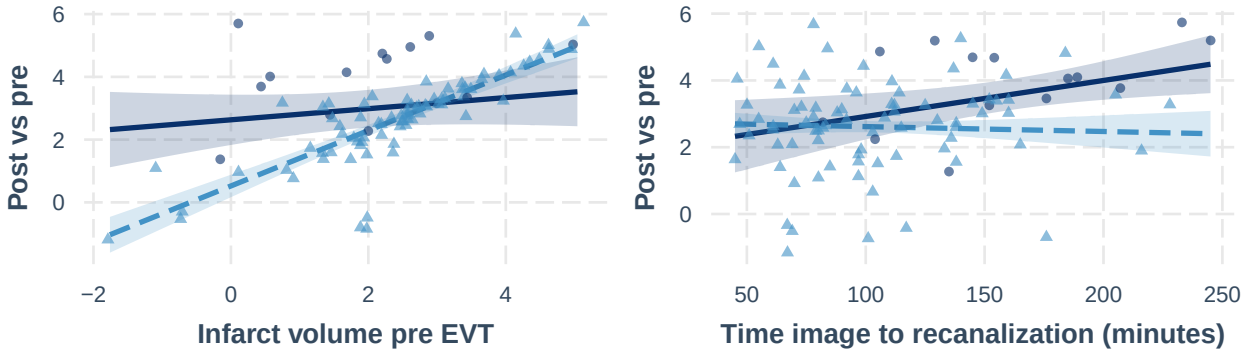
Also, for the *Post vs Pre*, the variable number of passes has significant differences between the minimum, the median and the maximum. Moreover, the variable time image to recanalization has significant differences between the minimum, the median and the maximum when TICI grades are less than 2b.

Moreover, for the *5d vs Post* and *5d vs Pre* models, the variable TMax6 volume post has significant differences between the minimum, the median and the maximum when TICI grades are greater than or equal to 2b.

Finally, for the *5d vs Post* model, the variable infarct volume post EVT has significant differences between the minimum, the median and the maximum when TICI grades are greater than or equal to 2b.

Table 15: Volume predictions for all models according to the TICI grade for those variables that presented interaction

Variable	Value	TICI	Post vs Pre		5d vs Post		5d vs Pre	
			Estimate	Group	Estimate	Group	Estimate	Group
Infarct vol. pre EVT	154.47	TICI<2b	33.78 [11.36, 101.8]	a			125.59 [32.01, 492.75]	a
	11.94		21.54 [11.59, 40.33]	a			122.12 [36.97, 403.43]	a
	0.17		10.18 [3.06, 33.82]	a			120.18 [17.41, 828.82]	a
	154.47	TICI>=2b	145.47 [99.48, 211.88]	a			173.82 [110.72, 273.14]	a
	11.94		15.18 [12.81, 17.96]	b			29.31 [23.83, 36.23]	b
	0.17		0.36 [0.2, 0.63]	c			1.53 [0.77, 3.06]	c
Number of passes	10	TICI<2b	788.4 [159.17, 3866.09]	a				
	2		17.64 [9.3, 33.78]	b				
	1		11.02 [5.16, 23.57]	c				
	10	TICI>=2b	37.71 [15.8, 90.92]	a				
	2		12.81 [10.7, 15.33]	b				
	1		11.13 [8.85, 14.15]	c				
Time image to rec.	245	TICI<2b	88.23 [37.34, 210.61]	a				
	102		18.92 [9.68, 37.34]	b				
	45		10.28 [3.49, 30.27]	c				
	245	TICI>=2b	14.88 [10.7, 20.7]	a				
	102		13.6 [11.59, 16.12]	a				
	45		11.02 [5.58, 21.98]	a				
TMax6 volume post	207	TICI<2b			59.15 [29.08, 120.3]	a	131.63 [31.82, 550.04]	a
	3				99.48 [31.39, 317.35]	a	131.63 [32.46, 533.79]	a
	0.01				100.48 [30.88, 327.01]	a	75.19 [33.45, 169.02]	a
	207	TICI>=2b			223.63 [71.52, 699.24]	a	713.37 [196.37, 2565.73]	a
	3				21.76 [18.36, 26.05]	b	17.29 [14.15, 20.91]	b
	0.01				21.12 [17.46, 25.28]	c	16.28 [13.33, 19.89]	c
Infarct vol. post EVT	298.87	TICI<2b			264.54 [100.08, 699.24]	a		
	19.89				98.49 [37.75, 256.98]	a		
	0.33				21.96 [2.62, 183.83]	a		
	298.87	TICI>=2b			302.48 [197.75, 462.2]	a		
	19.89				33.68 [27.94, 40.61]	b		
	0.33				1.2 [0.68, 2.11]	c		



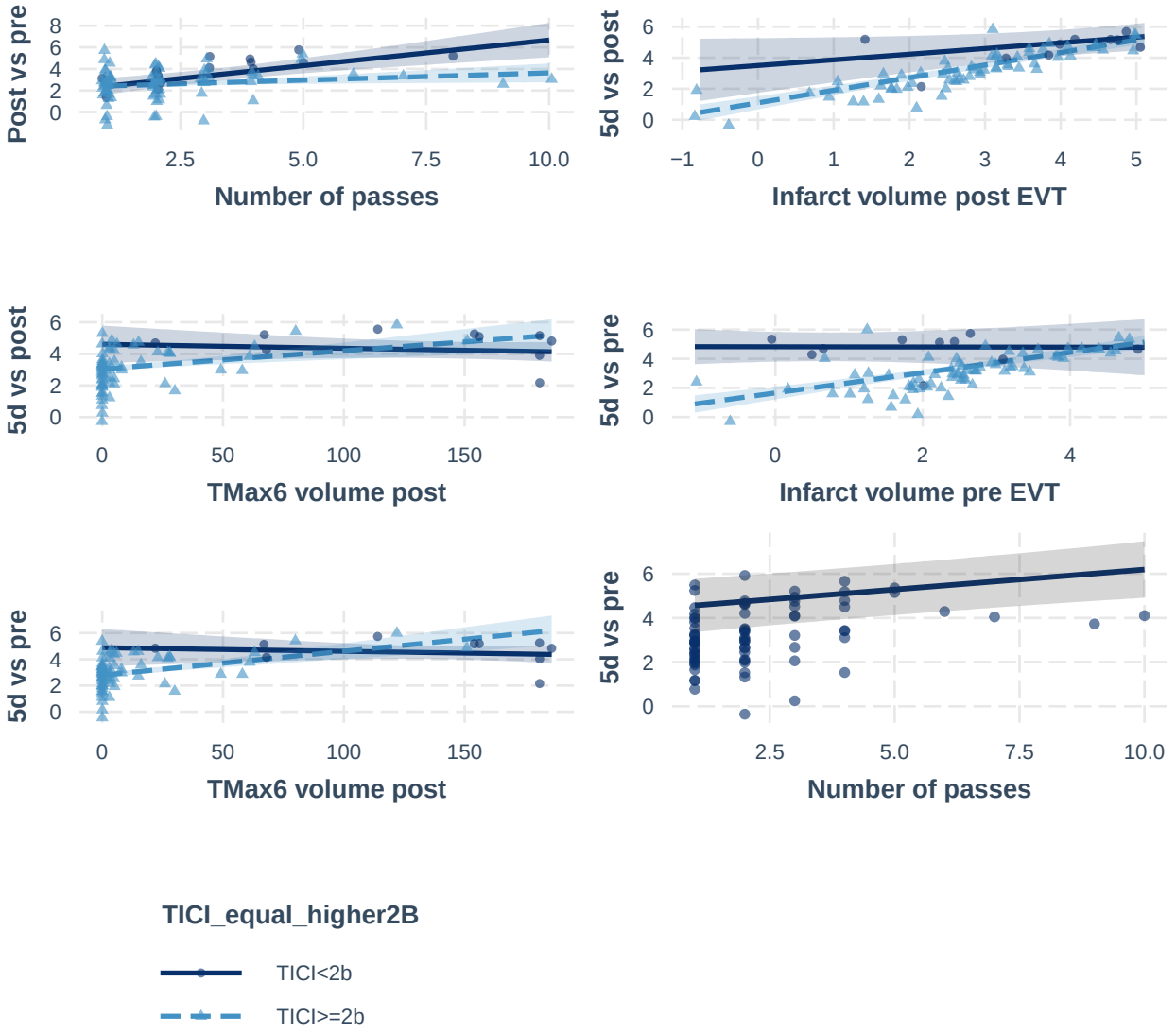


Figure 20: Association plots for Post vs Pre model

The validation of each regression model using infarct volumes is made below.

First, the graphic of residuals versus fitted values serves to validate the *Post vs Pre* model. In this case, as shown in figure 21, there seems to be a slightly heteroscedastic pattern. Besides, there is a curvilinear pattern, which indicates that the model may be missing a higher order term. Although, no variable in this model has a significant parameter into adding a higher order term.

Figure 22 shows that the observation 59 has the highest Cook's distance, observations 19 and 59 have the highest hat-values and observations 24, 14 and 17 have the highest studentized residuals in this model. For this reason, outlier tests are provided, showing that observations 24, 14 and 17 should be considered outliers (Table 16).

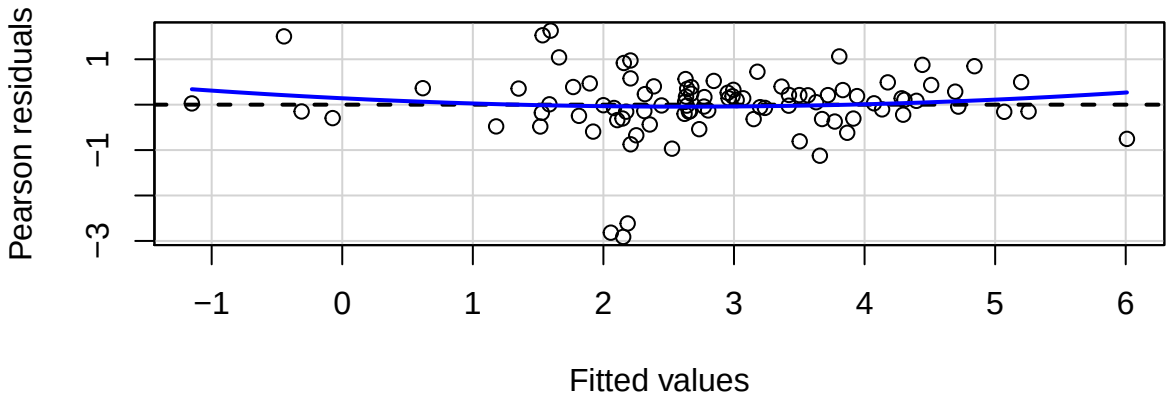


Figure 21: Residuals vs fitted values for the Post vs Pre model

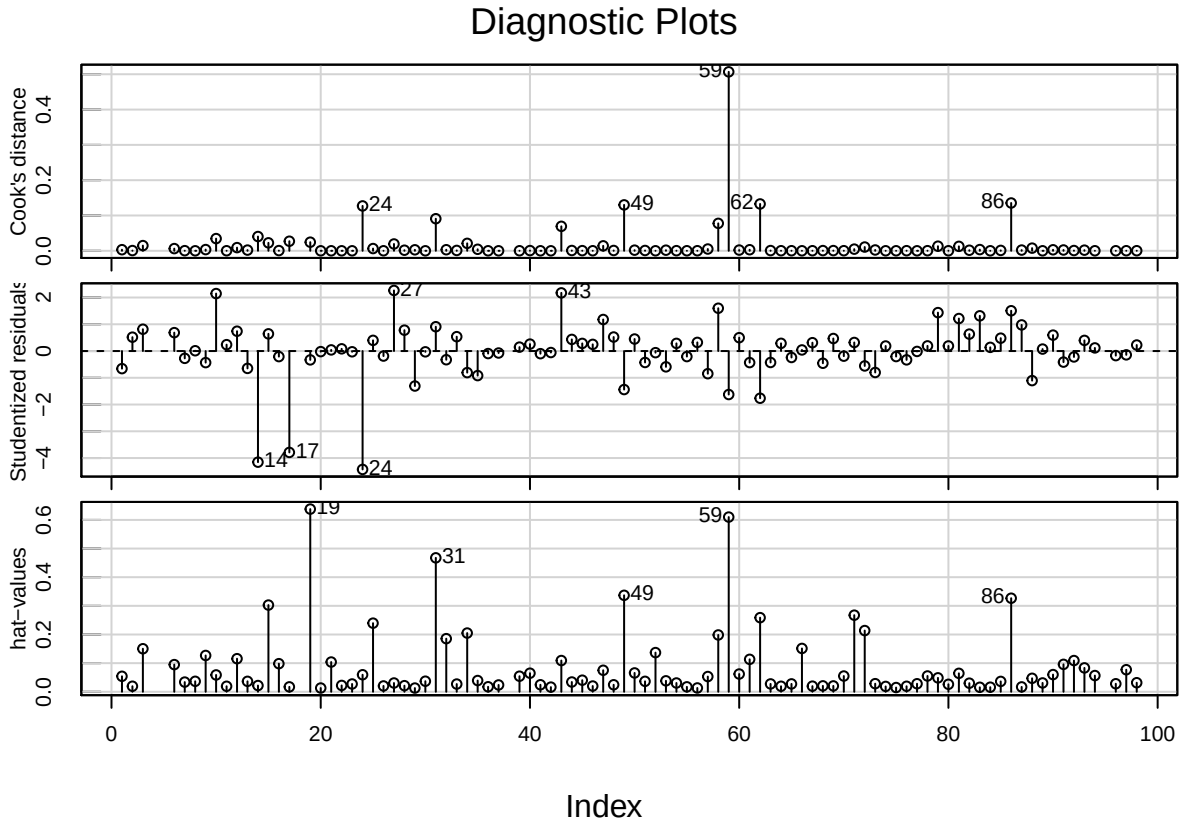


Figure 22: Diagnostic plots for the Post vs Pre model

Table 16: Outlier test for the Post vs Pre model

	Rstudent	Unadjusted p-value	Bonferroni p
24	-4.425948	2.8656e-05	0.0026650
14	-4.154438	7.8038e-05	0.0072576
17	-3.790263	2.8232e-04	0.0262560

Second, the graphic of residuals versus fitted values serves to validate the *5d vs Post* model. As shown in figure 23, the residuals roughly form a horizontal band around the residual line, which means that homoscedasticity might be assumed. Besides, there is a slightly curvilinear pattern, which is not fixed when adding a higher order term.

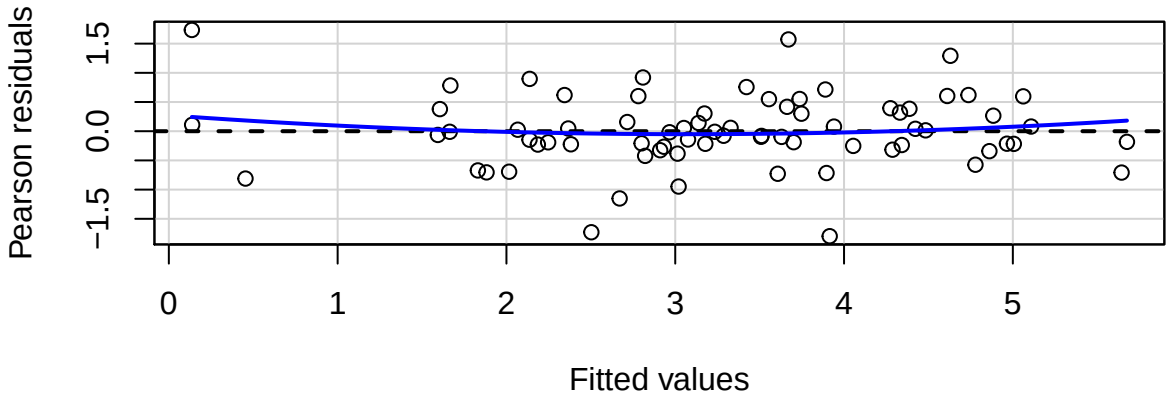


Figure 23: Residuals vs fitted values for the 5d vs Post model

Figure 24 shows that observations 49 and 34 have the highest Cook's distance and that observation 49 has the highest studentized residuals and hat-values in this model. For this reason, outlier tests are provided, showing that observations 49 and 34 should be considered outliers (Table 17).

Table 17: Outlier test for the 5d vs Post model

	Rstudent	Unadjusted p-value	Bonferroni p
49	4.038863	0.00014247	0.010400
34	-3.705374	0.00043382	0.031669

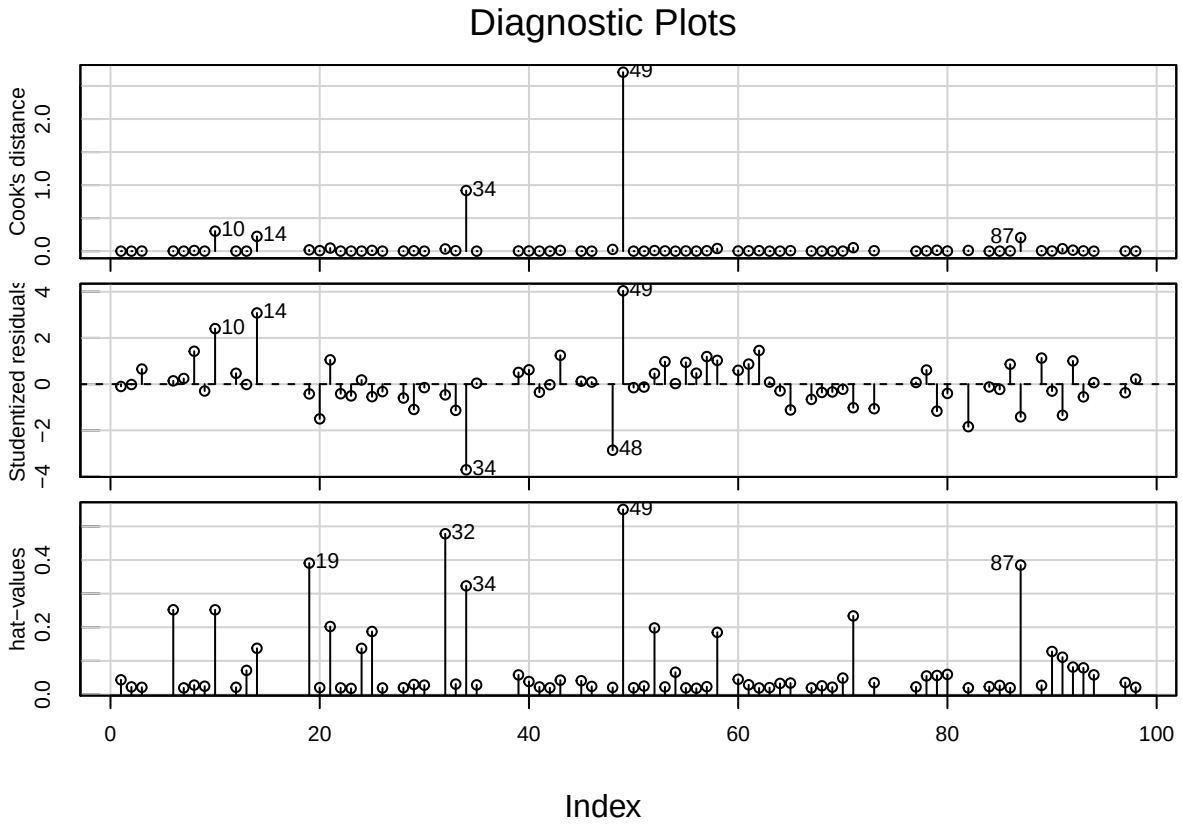


Figure 24: Diagnostic plots for the 5d vs Post model

Last, the graphic of residuals versus fitted values serves to validate the *5d vs Pre* model. In this case, as shown in figure 25, the residuals roughly form a horizontal band around the residual line, which means that homoscedasticity might be assumed. Besides, the residuals have no obvious patterns or systematic structure, which means that linearity might also be assumed.

Figure 26 shows that observation 49 and 34 have the highest Cook's distance and that observation 49 has the highest Cook's distance, observation 34 has the highest studentized residuals and observation 19 has the highest hat-values in this model. For this reason, outlier tests are provided, showing that observation 34 should be considered an outlier (Table 18).

Table 18: Outlier test for the 5d vs Pre model

	Rstudent	**Unadjusted p-value**	**Bonferroni p**
34	-3.587642	0.00061905	0.047667

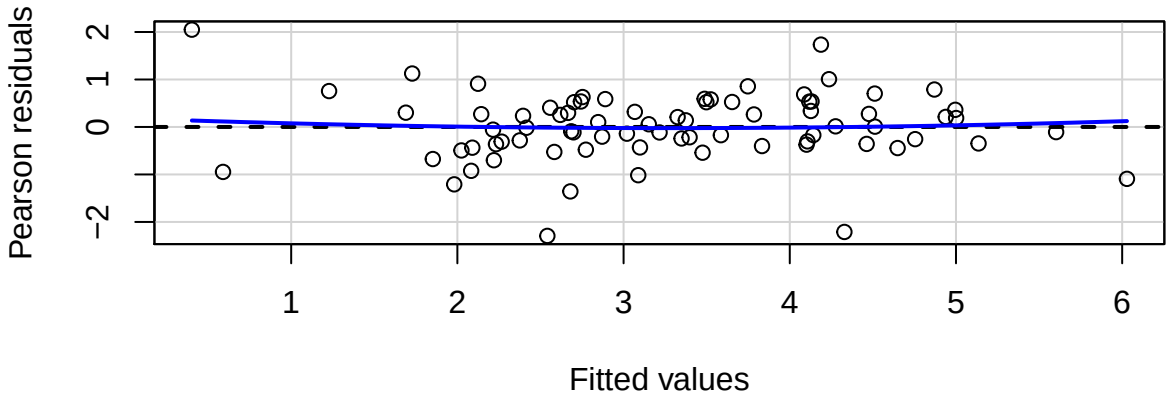


Figure 25: Residuals vs fitted values for the 5d vs Pre model

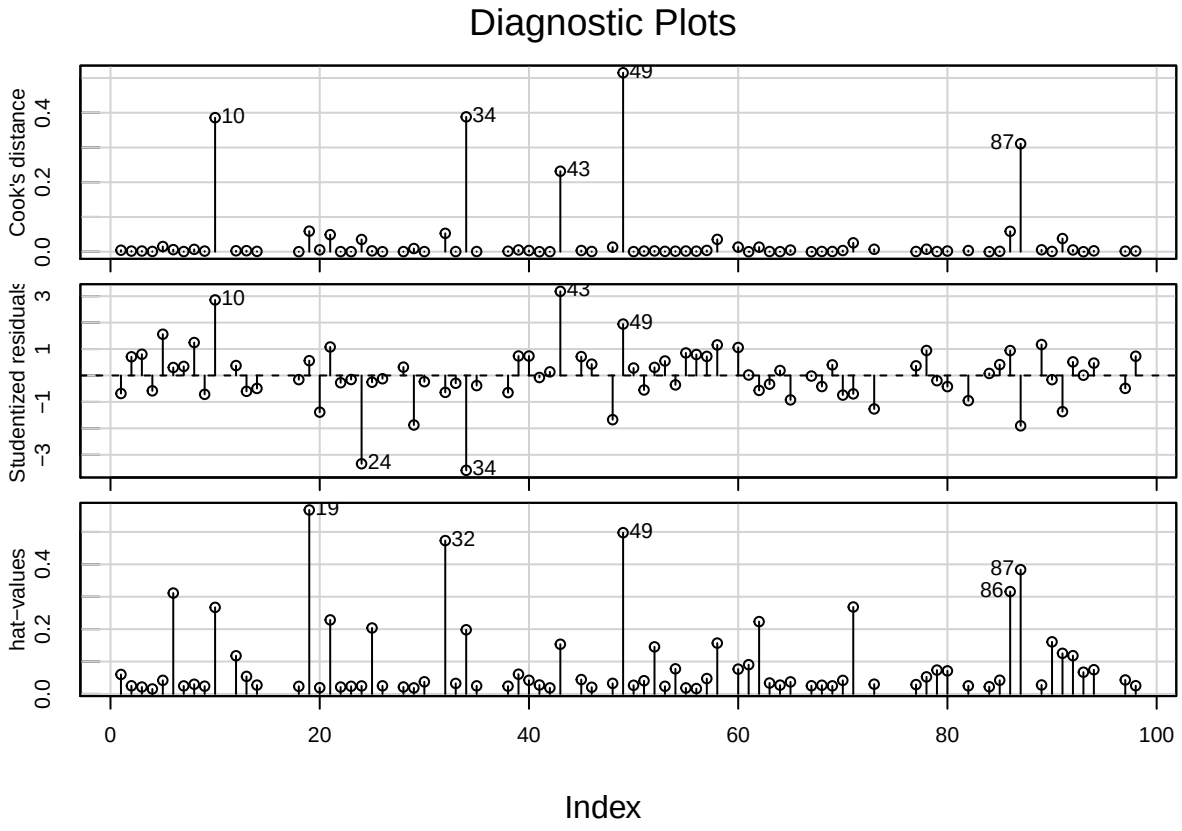


Figure 26: Diagnostic plots for the 5d vs Pre model

To sum up, the forest plots of the *Post vs Pre*, the *5d vs Post* and *5d vs Pre* models are shown in figures 27, 28 and 29. It is important to mention that for the *Post vs Pre* model an increase of 10% is shown for the logarithm of the infarct volume pre EVT and an increase of 5 minutes is shown in the variable time image to recanalization.

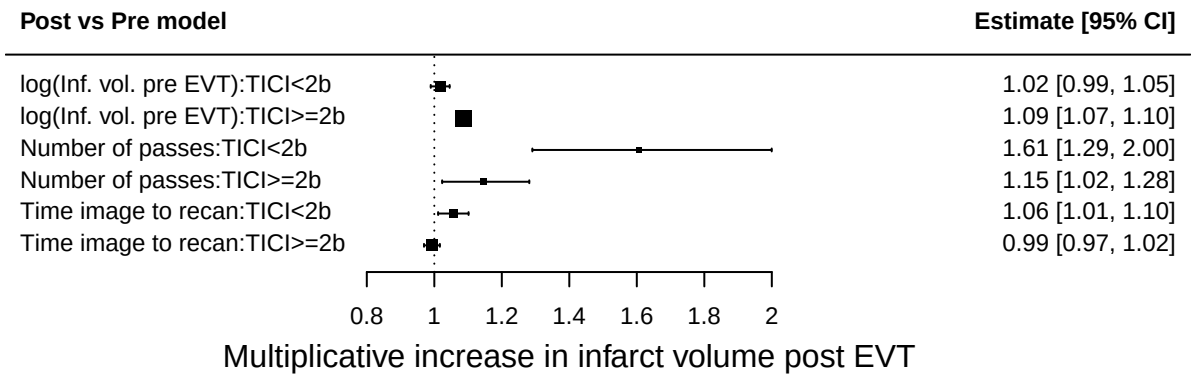


Figure 27: Forest plot of Post vs Pre model

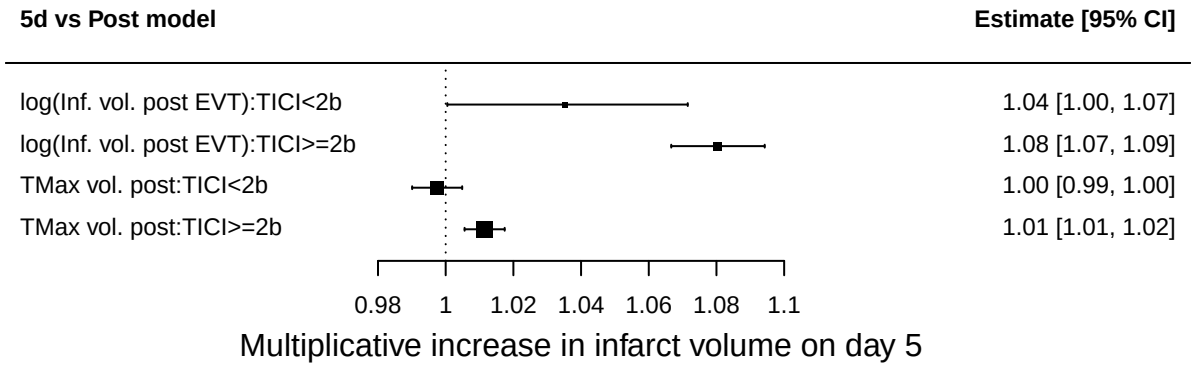


Figure 28: Forest plot of 5d vs Post model

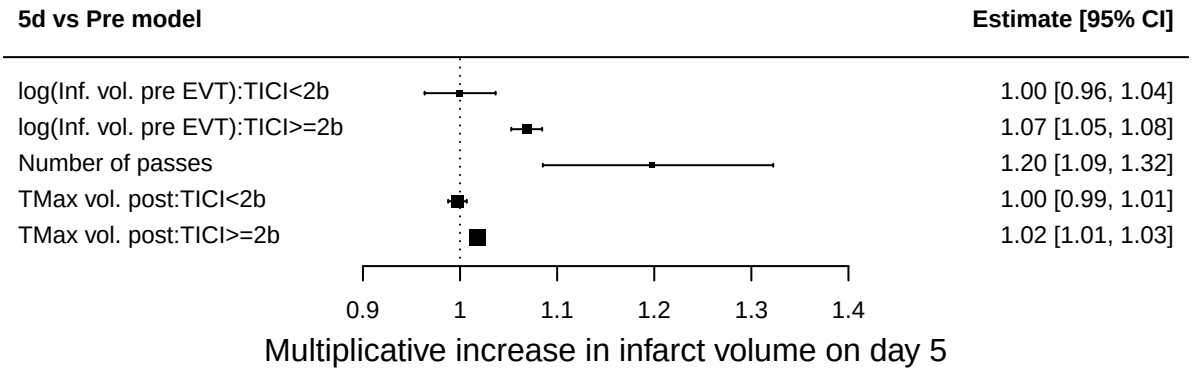


Figure 29: Forest plot of 5d vs Pre model

4.3 Generalized linear models

4.3.1 Multivariate models with infarct volumes

Generalized linear models are applied to correct the problem of heteroscedasticity, to avoid logarithmic transformation and to have more flexibility when modeling, since these models allow a greater number of families. In this case, Gamma family and link identity are used.

Table 19 shows the estimations and the 95% CI of the parameters of each variable. More detailed information on each model is included in the Appendix 3.

Table 19: Coefficient point estimates and 95% CI of the multivariate generalized linear models using infarct volumes

	Post vs Pre	**5d vs Post**	**5d vs Pre**
Infarct volume pre EVT	1.28 [1.01, 1.55]		-0.26 [-1.83, 1.30]
Infarct volume post EVT		1.20 [0.82, 1.59]	
Number of passes (One pass)	-82.86 [-125.05, -40.66]		
TICI>=2b	-81.97 [-123.86, -40.07]		-103.95 [-197.28, -10.63]
Occlusion post		38.81 [15.59, 62.03]	48.43 [20.55, 76.31]
Infarct vol. pre EVT:TICI>=2b			1.41 [-0.23, 3.05]
Number of passes:TICI>=2b	79.55 [37.32, 121.79]		

The Cragg-Uhler R-squared obtained are 0.76, 0.71 and 0.66, for the *Post vs Pre*, *5d vs Post* and *5d vs Pre*, respectively.

A previous imputation of missing values for the predictors is performed. Nevertheless, missing values are found in these models because infarct volumes post-procedure are not included in the imputation.

First, for the *Post vs Pre* model, the following relevant variables are found: infarct volume pre EVT and the interaction between the number of passes and TICI grades. Second, for the *5d vs Post* model, the infarct volume post EVT and the occlusion are relevant. Last, for the *5d vs Pre* model, the occlusion and the interaction between infarct volume pre EVT and the variable TICI grade are relevant.

Neither of these three models has problems of multicollinearity (VIFs are calculated).

In addition, an interpretation and graphical representation of all the effects in each of the generalized linear models using infarct volumes is made in table 20 and figure 30.

Table 20 shows the volume predictions for the three generalized linear models, according to the TICI grade for those variable that presented interaction.

Table 20: Volume predictions for the generalized linear models

Variable	Value	TICI	Post vs Pre		5d vs Post		5d vs Pre	
			Estimate	Group	Estimate	Group	Estimate	Group
Infarct volume pre EVT	154.734		221.8 [180, 263.6]	a				
	11.983		38.3 [27.3, 49.2]	b				
	0.172		23.1 [12.3, 33.8]	c				
Infarct volume post EVT	297.658				380.4 [268.6, 492.2]	a		
	19.877				45.9 [34.3, 57.5]	b		
	0.325				22.3 [10.4, 34.3]	c		
Occlusion	Yes				81.4 [58.6, 104.1]	a	114.3 [71.3, 157]	a
	No				42.6 [30.6, 54.5]	b	65.8 [24.1, 108]	b
Number of passes	More passes	TICI<2b	113.9 [71.3, 156.4]	a				
	One pass		31 [23.2, 38.9]	b				
	More passes	TICI>=2b	31.6 [26.2, 37.1]	a				
	One pass		28.6 [22.8, 34.4]	b				
Infarct volume pre EVT	154.734	TICI<2b					91.9 [-120.5, 304.4]	a
	11.983						129.2 [45.8, 212.5]	a
	0.172						132.3 [41.4, 223.1]	a
	154.734	TICI>=2b					206.1 [132.3, 279.9]	a
	11.983						42.1 [28.2, 56.1]	b
	0.172						28.5 [14.1, 43]	c

For the *Post vs Pre* model, the variable infarct volume pre EVT has significant differences between the minimum, the median and the maximum. The variable number of passes also has significant differences between these three values (Table 20).

Also, for the *5d vs Post* model, the variable infarct volume post EVT has significant differences between the minimum, the median and the maximum. There are also significant differences between doing or not doing an occlusion.

Finally, for the *5d vs Pre* model, there are significant differences between doing or not doing an occlusion. The variable infarct volume pre EVT has also significant differences between the minimum, the median and the maximum when TICI grades are greater than or equal to 2b.

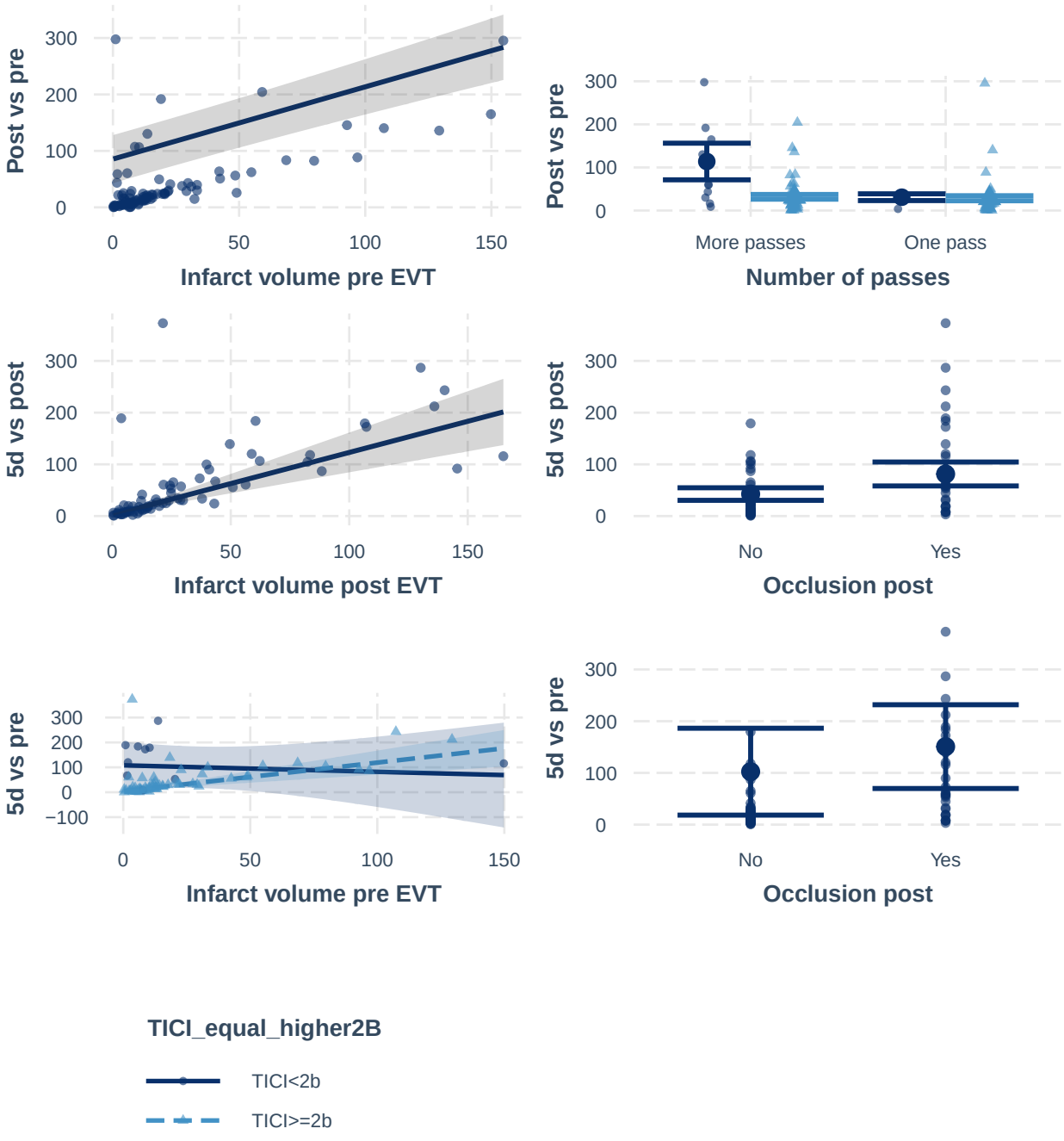


Figure 30: Association plots of the interactions

The validation of each generalized linear model using infarct volumes is made below.

First, to validate the *Post vs Pre* model, the graphic of residuals versus fitted values serves to assess heteroscedasticity. In this case, there seems to be a slightly heteroscedastic pattern (Figure 31). To ensure this fact, the Levene test is done. The resulted p-value is 0.16; this result lead to the non-rejection region of the null hypothesis. Consequently, homoscedasticity can be assumed.

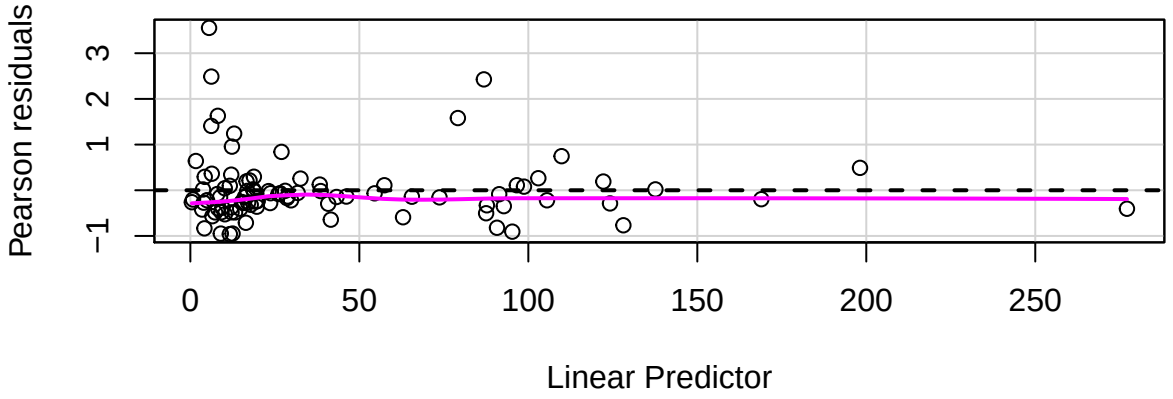


Figure 31: Residuals vs fitted values for Post vs Pre model

Figure 32 shows that observation 45 has the highest Cook's distance and hat-value and observation 21 has the highest studentized residuals in this model. Consequently, an outlier test is performed, showing that observation 21 should not be considered an outlier (Table 21).

Table 21: Outlier test for the Post vs Pre model

	Rstudent	**Unadjusted p-value**	**Bonferroni p**
21	-3.106003	0.0018963	0.17636

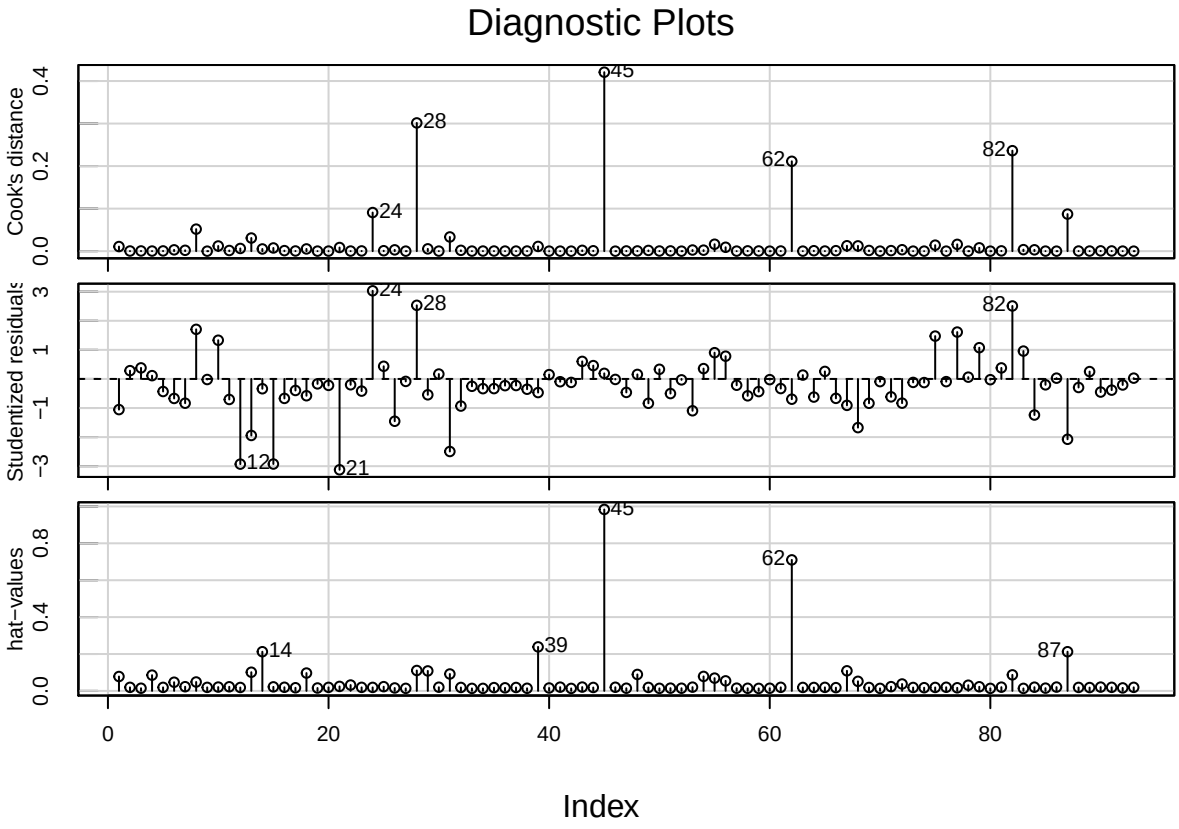


Figure 32: Diagnostic plots for the Post vs Pre model

Second, to validate the *5d vs Post* model, the graphic of residuals versus fitted values serves to assess heteroscedasticity. In this case, there seems to be a slightly heteroscedastic pattern (Figure 33). Despite that, a Levene test is done. The resulted p-value is 0.40; this result lead to the non-rejection region of the null hypothesis. Consequently, homoscedasticity can be assumed.

Figure 34 shows that observations 45 and 8 have the highest Cook's distance and studentized residuals and the observations 12 and 21 have the highest hat-values. Because of this, an outlier test is provided, showing that observation 21 should be considered an outlier (Table 22).

Table 22: Outlier test for the 5d vs Post model

	Rstudent	**Unadjusted p-value**	**Bonferroni p**
8	3.786826	0.00015258	0.011139

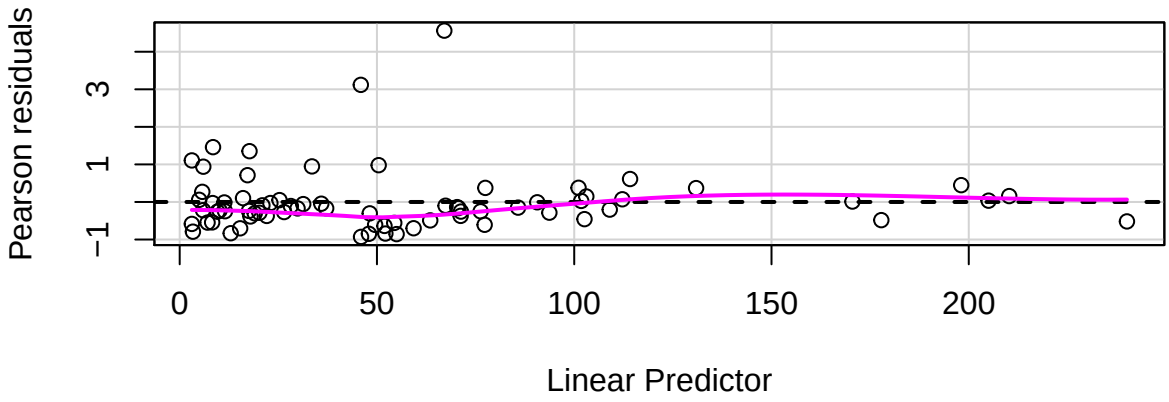


Figure 33: Residuals vs fitted values for 5d vs Post model

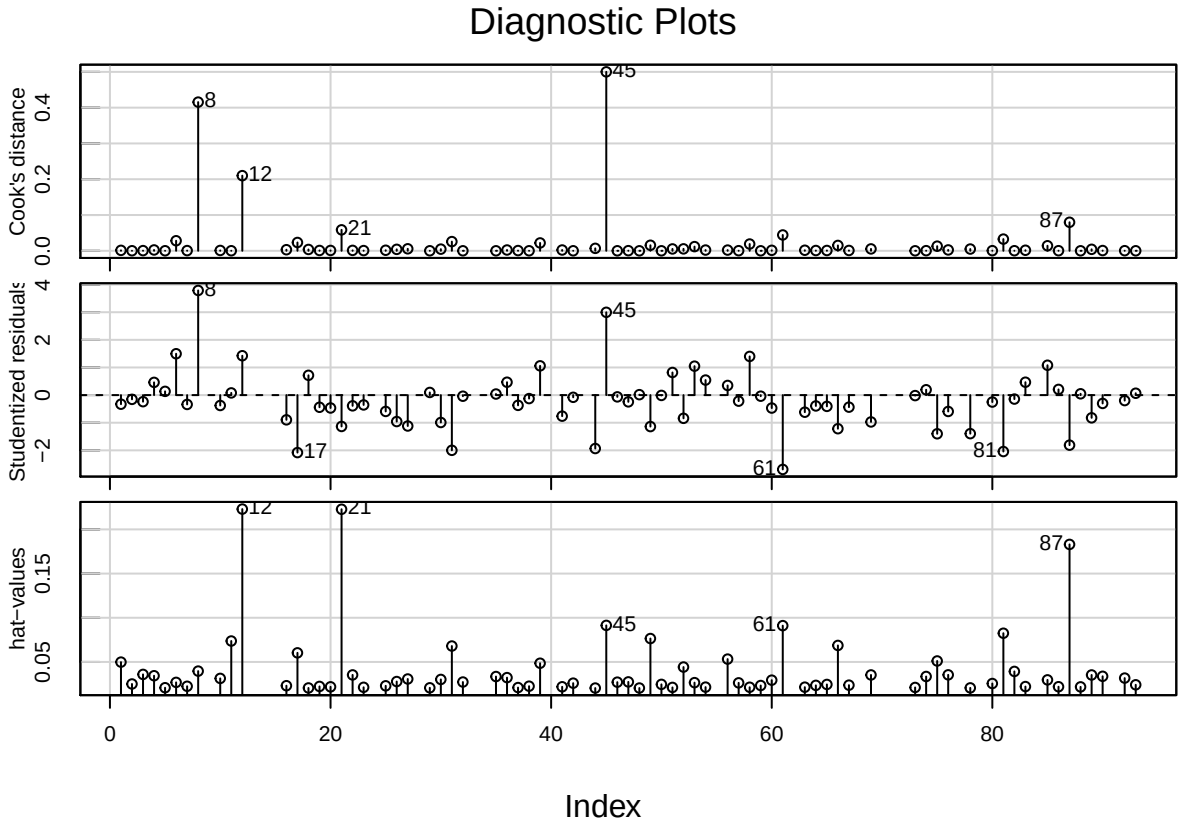


Figure 34: Diagnostic plots for the 5d vs Post model

Last, to validate the *5d vs Pre* model, the graphic of residuals versus fitted values serves to assess heteroscedasticity. In this case, there seems to be a slightly heteroscedastic pattern (Figure 35). Although, the Levene test is done. The resulted p-value is 0.43. This result lead to the non-rejection region of the null hypothesis.

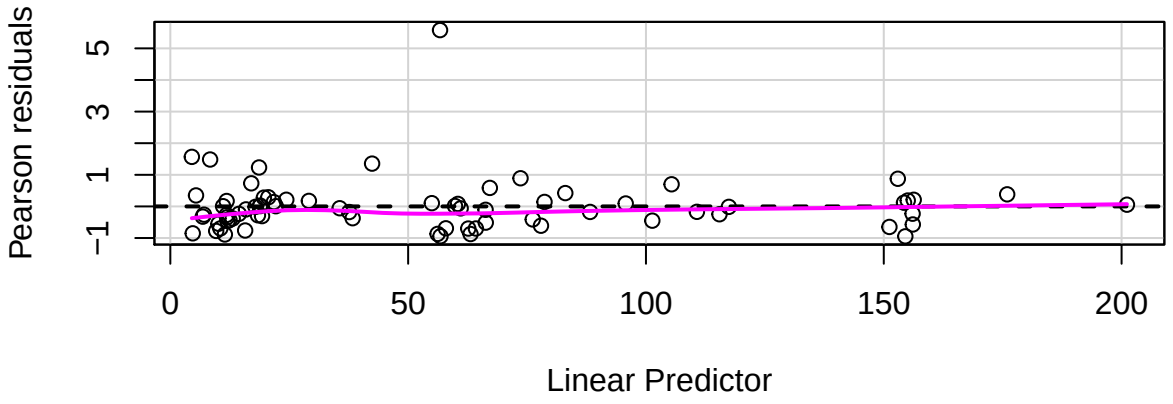


Figure 35: Residuals vs fitted values for 5d vs Pre model

Figure 36 shows that the observation 8 has the highest Cook's distance and studentized residuals and observation 16 has the highest hat-values in this model. Consequently, an outlier test is done, showing that observation 8 should be considered an outlier (Table 23).

Table 23: Outlier test for the 5d vs Pre model

	Rstudent	**Unadjusted p-value**	**Bonferroni p**
8	4.335399	1.455e-05	0.0010621

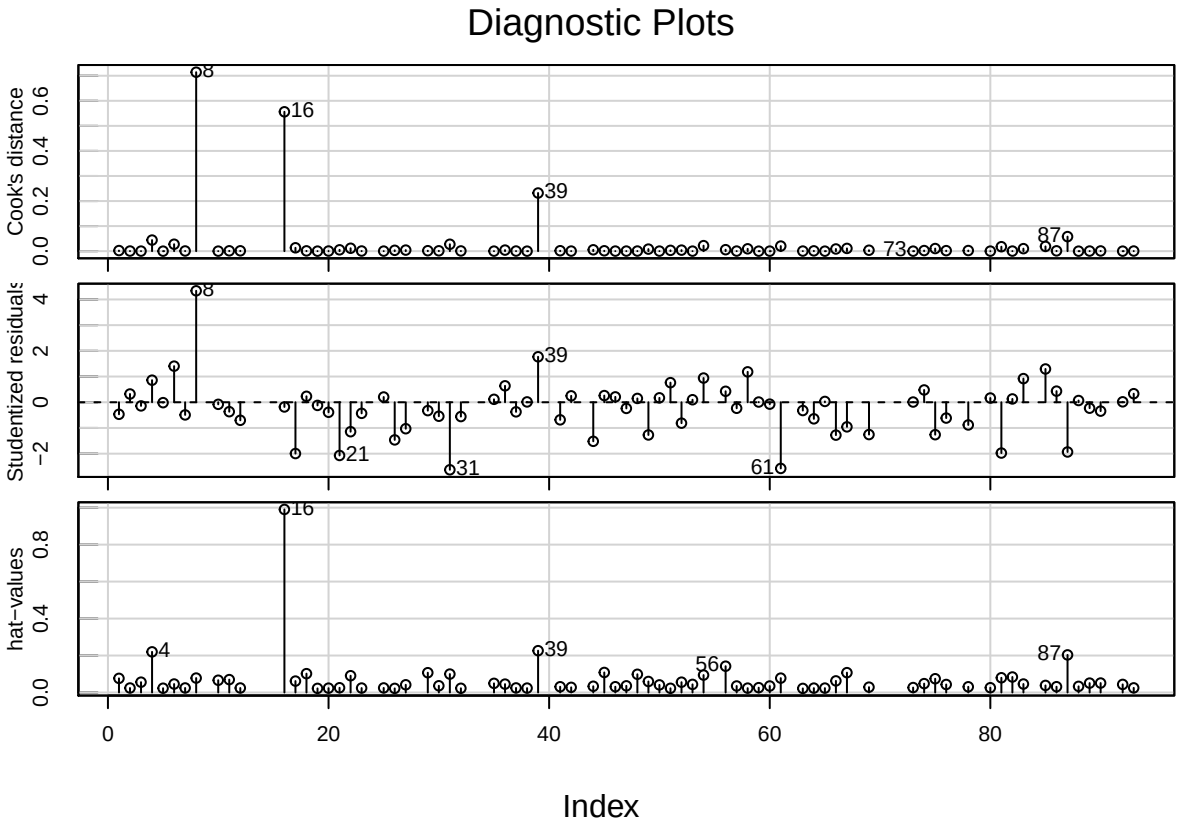


Figure 36: Diagnostic plots for the 5d vs Pre model

To sum up, the forest plots of the *Post vs Pre*, the *5d vs Post* and *5d vs Pre* models are shown in figures 37, 38 and 39.

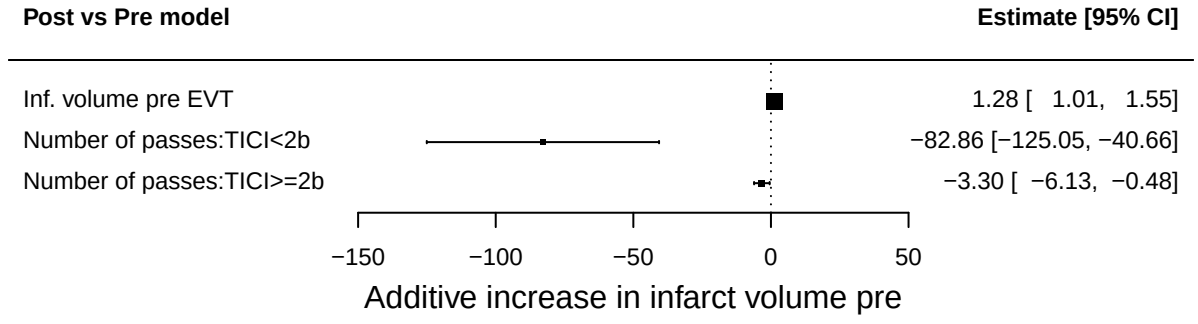


Figure 37: Forest plot of Post vs Pre model

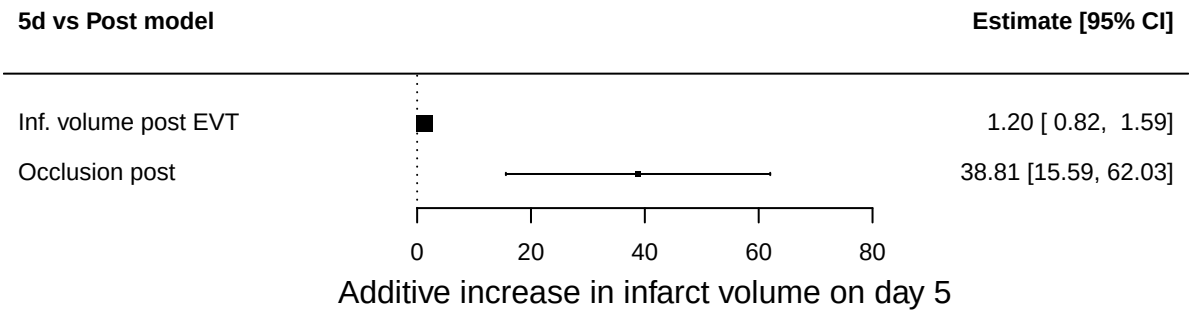


Figure 38: Forest plot of 5d vs Post model

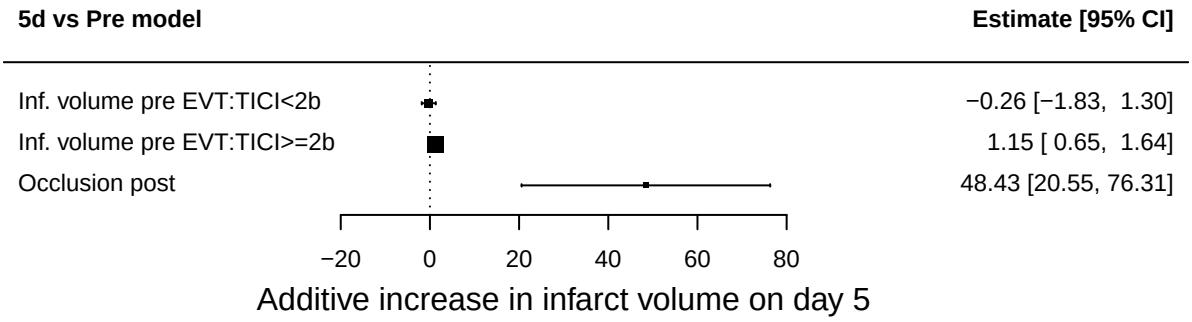


Figure 39: Forest plot of 5d vs Pre model

4.4 Mixed model

Mixed models provide an analytical power that cannot be achieved using general or generalized linear models because they consider the correlations between the measure corresponding the same patient [6]. A mixed model with random intercepts was fitted to the data (showed below) and more detailed information about this model is included in the Appendix 4.

Table 24 shows the estimations and the 95% CI of the parameters of fixed effects. Even though, for random intercepts, it shows the standard deviation.

Table 24: Coefficient point estimates and 95% CI of the mixed model

	Estimate	**Std. Dev.**
Fixed Effects		
Number of passes	0.17 [0.03, 0.31]	
Time Log(Infarct volume post EVT)	1.96 [1.45, 2.47]	
Time Log(Infarct volume 5d)	2.72 [2.14, 3.29]	
TICI>=2B	0.63 [-0.12, 1.38]	
Time Log(Infarct volume post EVT):TICI>=2B	-1.73 [-2.28, -1.19]	
Time Log(Infarct volume 5d):TICI>=2B	-2.1 [-2.72, -1.48]	
Random Effects		
Subject		1.09

For instance, a one unit increase in the predictor number of passes corresponds to a 18.5% [3.41%, 35.79%] increase in the outcome infarct volume. Likewise, an increase of 10% in the infarct volume post EVT is related to a punctual increase of 20.53% [14.85%, 26.5%] in the outcome infarct volume if TICI grades are less than 2b. However, it is associated with an increase of 2.16% [0.12%, 4.25%] if TICI grades are equal or greater than 2b.

Additionally, an increase of 10% in the infarct volume on day 5 is related to a punctual increase of 29.56% [22.66%, 36.86%] in the outcome infarct volume if TICI grades are less than 2b. Even so, it is associated with an increase of 6.06% [3.81%, 8.36%] if TICI grades are equal or greater than 2b.

The correlations of the fixed effects can also be seen in the Appendix 4.

To validate this model, the graphic of residuals versus fitted values is used. There seems to be a homoscedastic pattern and linearity because the residuals have no obvious patterns or systematic structure (Figure 40).

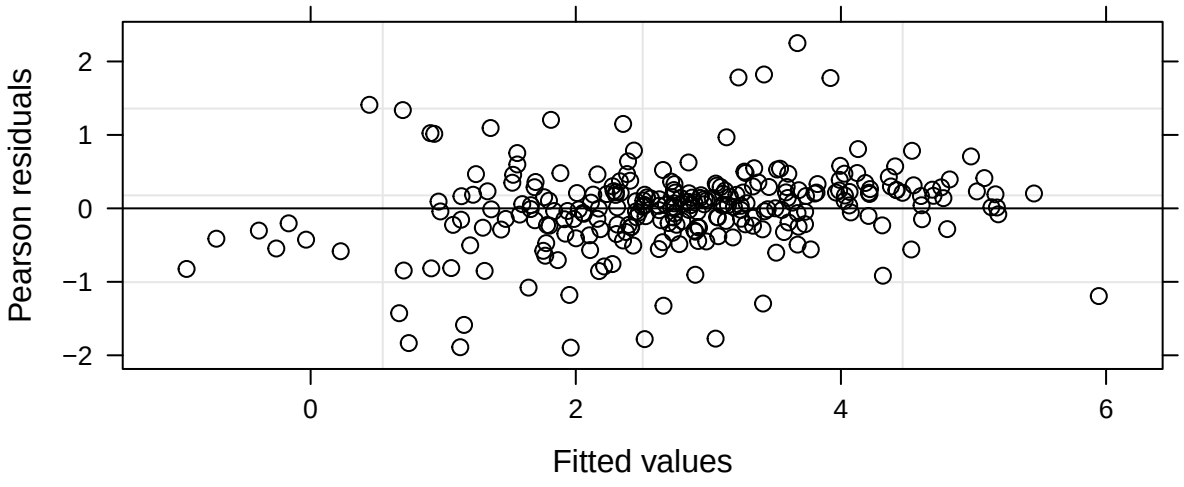


Figure 40: Residuals vs fitted values for the mixed model

To sum up, the forest plot of this model is shown in figure 41. It is important to mention that an increase of 10% is shown for the logarithm of the infarct volumes in this plot.

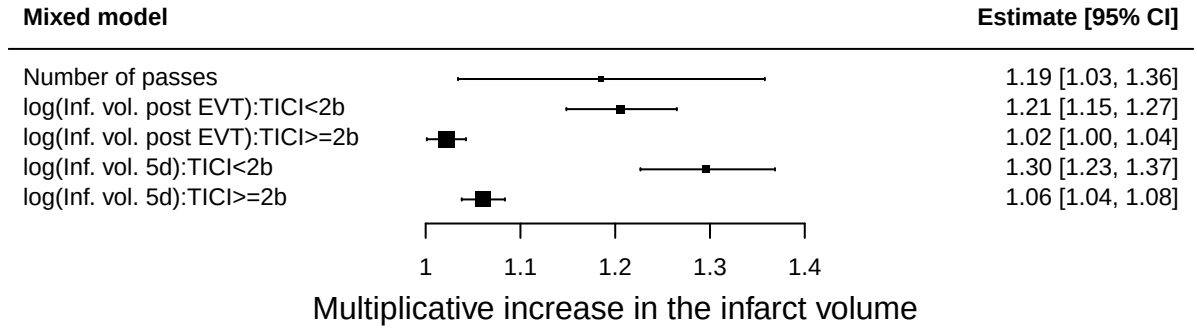


Figure 41: Forest plot of the mixed model

5 Discussion

5.1 Main findings

The relevant factors found in each model are provided in the following table:

Variables	Early	Post vs Pre (lm)	Post vs Pre (glm)	Late	5d vs Post (lm)	5d vs Post (glm)	Total	5d vs Pre (lm)	5d vs Pre (glm)	Mixed
HIR pre	x									
Infarct volume pre EVT	R	X	x	-	-	-	R	X	X	R
Infarct volume post EVT	-	-	-	R	X	x	-	-	-	R
Number of passes	X	X	X					x		x
Occlusion post						x			x	
Time image to recanalization		X		X			X			
TMax6 volume post				X	X		X	X		

[1] Uppercase X indicates when interaction with TICI grade is relevant

[2] Uppercase R indicates that the corresponding infarct volume is used in the response variable, not as a covariate

[3] - indicates that the corresponding infarct volume was not considered

The main findings were obtained through the general linear model using infarct volumes and the mixed model because they are the most consistent and reliable models done in this end-of-degree project.

The most relevant factors associated with the increase of the infarct volume post EVT are the infarct volume pre EVT, the number of passes and the time image to recanalization. Increasing the infarct volume pre EVT, is always related to an increase in the infarct volume post EVT. However, it tends to increase more for patients with TICI grade greater than or equal to 2b than for the rest of the patients.

Furthermore, rising one pass of the catheter to achieve target vessel recanalization in the mechanical thrombectomy, is always related to a significant increase of the infarct volume post EVT. Nevertheless, it tends to increase more for patients with TICI grade less than 2b rather than for those with TICI grade greater than or equal to 2b.

Besides, an increase of five minutes in the time image to recanalization is associated with an increase in the early infarct growth in patients with TICI grade less than 2b, but a decrease in patients TICI grade is greater than or equal to 2b. This might be because when patients do not recanalize spending five more minutes in time image recanalization is not going to affect the infarct volume post EVT.

The most relevant factor associated with the increase of the infarct volume on day 5 is the variable TMax6 volume post. An increase of this variable is related to a significant increase in infarct volume on day 5 in patients with TICI grades greater than or equal to 2b.

Nonetheless, all these results must be interpreted with caution and a number of limitations should be borne in mind.

5.2 Limitations

The findings of this study have to be seen in light of some limitations. First, the number of patients with TICI grades less than 2b is too small, only 14 patients did not recanalize. For this reason, it was difficult to find significant relationships from this group and the assumption of Normal distribution required for most of statistical tests cannot be assessed with lower sample sizes.

Second, there are some missing values in the infarct volumes, especially in the infarct volume on day 5 because some patients died within 5 days after the procedure, which is when the third MRI is performed. These missing data may introduce some kind of bias as well as a loss of statistical power and precision. Models done with imputed missing values and adjusting for all volumes decrease the risk of biased results.

Finally, in the multivariate generalized linear models, variables that caused convergence problems with the model are not included, which could imply that some relevant factors might have been ignored.

5.3 Conclusion

To conclude, rising one pass of the catheter to achieve target vessel recanalization in the mechanical thrombectomy, corresponds to a significant increase in the outcome infarct volume. This result seems reasonable since a greater number of passes during surgery is associated with more difficulties in controlling the infarct volume. As the relationship between the number of passes and the growth of the infarct volume is more marked in patients with a TICI less than 2b, in these patients the convenience of increasing the number of passes should be assessed through an experimental study. Also, increasing one unit the variable TMax6 volume post is associated with an increase of the final infarct size but only in patients that do recanalize. Unlike the number of passes, the researcher has no control over the variable Tmax6, since this can be seen as a result of the procedure. This implies that the association found has relevance from a predictive point of view, that is, it allows anticipating the infarcted volume on 5 days, but it cannot be modified to improve the patient's evolution.

Finally, future research should focus on re-analyzing new data with more patients, especially with ones that do not recanalize, in other words, with TICI less than 2b.

6 Appendix

6.1 Appendix 1: Descriptive analysis

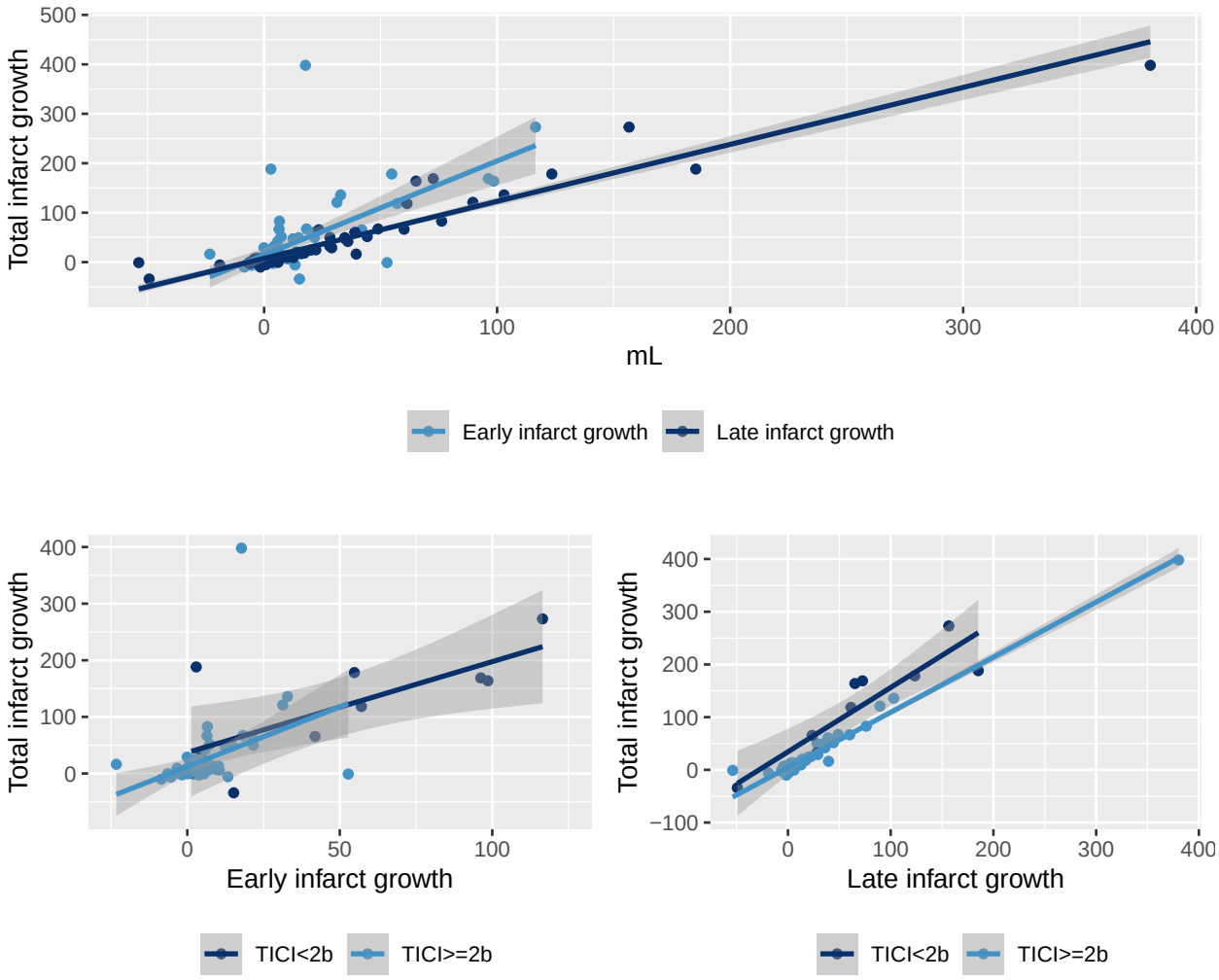


Figure 42: Scatter plots of the infarct growths

The first graph in figure 1 shows that early infarct growth has more influence to the total infarct growth than late infarct growth. That is because early infarct growth has a bigger slope.

Moreover, the left graph shows that the slope is quite similar between these two groups. However, there is a different intercept. This means that a patient without differences between pre and post-procedure infarct volumes will have more difference in the total infarct growth if the TICI grade is less than 2b. Also, the right graph shows that there is a different intercept. This means that a patient without differences between pre and post-procedure infarct volumes will have more difference in the total infarct growth if the TICI grade is less than 2b.

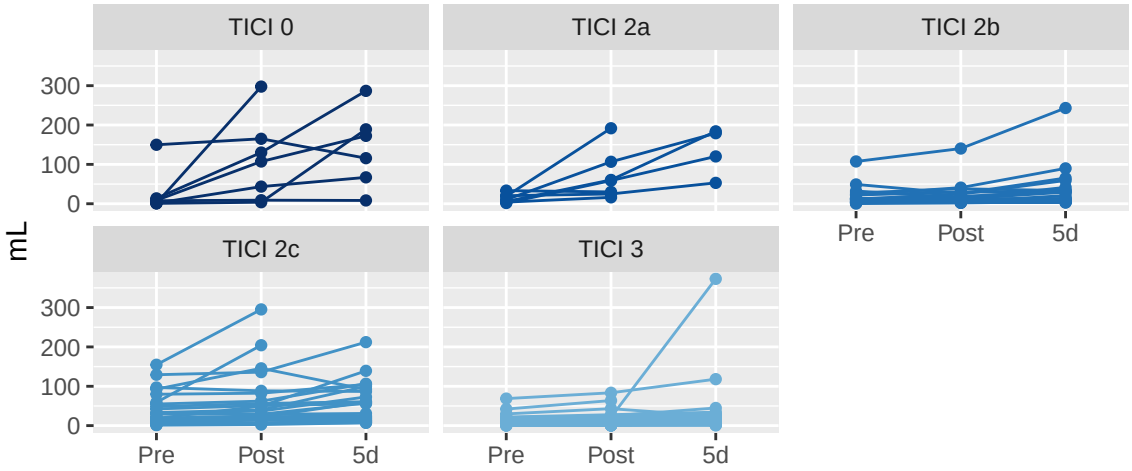


Figure 43: Infarct volumes for each patient stratified by TICI grade

6.2 Appendix 2: General linear models

6.2.1 Multivariate models with infarct growths

Regression Model for early infarct growth

Observations	77 (21 missing obs. deleted)
Dependent variable	Early_infarct_growth
Type	OLS linear regression

F(4,72)	8.94
R^2	0.33
Adj. R^2	0.29

	Est.	2.5%	97.5%	t val.	p
(Intercept)	58.09	27.89	88.28	3.84	0.00
HIR_pre_EVT	54.75	13.36	96.14	2.64	0.01
num_passes_catOne pass	-77.93	-139.27	-16.59	-2.53	0.01
TICI_equal_higher2BTICI>=2b	-70.64	-98.88	-42.39	-4.98	0.00
num_passes_catOne pass:TICI_equal_higher2BTICI>=2b	76.15	11.56	140.74	2.35	0.02

Standard errors: OLS

Table 25: VIFs of regression model for early infarct growth

	HIR_pre_EVT	num_passes_cat	TICI_equal_higher2B
VIF	1.02	1.06	1.04

Regression Model for late infarct growth

Observations	63 (35 missing obs. deleted)
Dependent variable	Late_infarct_growth
Type	OLS linear regression

F(5,57)	15.27
R^2	0.57
Adj. R^2	0.54

	Est.	2.5%	97.5%	t val.	p
(Intercept)	326.97	167.27	486.67	4.10	0.00
imaging_recan	-1.59	-2.44	-0.73	-3.71	0.00
TMax_6s_post_EVT	0.23	-0.27	0.74	0.92	0.36
TICI_equal_higher2BTICI>=2b	-338.23	-500.74	-175.73	-4.17	0.00
imaging_recan:TICI_equal_higher2BTICI>=2b	1.70	0.80	2.60	3.79	0.00
TICI_equal_higher2BTICI>=2b:TMax_6s_post_EVT	1.05	0.42	1.68	3.34	0.00

Standard errors: OLS

Table 26: VIFs of regression model for late infarct growth

	imaging_recan	TMax_6s_post_EVT	TICI_equal_higher2B
VIF	1.38	2	2.11

Regression Model for total infarct growth

Observations	63 (35 missing obs. deleted)
Dependent variable	Total_infarct_growth
Type	OLS linear regression

F(5,57)	26.67
R^2	0.70
Adj. R^2	0.67

	Est.	2.5%	97.5%	t val.	p
(Intercept)	453.58	299.11	608.04	5.88	0.00
imaging_recan	-1.90	-2.73	-1.07	-4.59	0.00
TMax_6s_post_EVT	0.08	-0.41	0.57	0.34	0.74
TICI_equal_higher2BTICI>=2b	-460.30	-617.48	-303.12	-5.86	0.00
TMax_6s_post_EVT:TICI_equal_higher2BTICI>=2b	1.39	0.78	1.99	4.56	0.00
imaging_recan:TICI_equal_higher2BTICI>=2b	2.00	1.13	2.87	4.61	0.00

Standard errors: OLS

Table 27: VIFs of regression model for total infarct growth

	imaging_recan	TMax_6s_post_EVT	TICI_equal_higher2B
VIF	1.38	2	2.11

6.2.2 Multivariate models with infarct volumes

Regression Model for Post vs Pre

Observations	93 (5 missing obs. deleted)
Dependent variable	Log_V_post_EVT
Type	OLS linear regression

F(7,85)	36.99
R^2	0.75
Adj. R^2	0.73

	Est.	2.5%	97.5%	t val.	p
(Intercept)	0.30	-1.31	1.92	0.37	0.71
Log_V_pre_EVT	0.18	-0.11	0.47	1.22	0.23
TICI_equal_higher2BTICI>=2b	0.06	-1.64	1.77	0.07	0.94
num_passes	0.47	0.26	0.69	4.30	0.00
imaging_recan	0.01	0.00	0.02	2.51	0.01
Log_V_pre_EVT:TICI_equal_higher2BTICI>=2b	0.71	0.39	1.02	4.44	0.00
TICI_equal_higher2BTICI>=2b:num_passes	-0.34	-0.58	-0.09	-2.73	0.01
TICI_equal_higher2BTICI>=2b:imaging_recan	-0.01	-0.02	-0.00	-2.49	0.01

Standard errors: OLS

Table 28: VIFs of regression model for Post vs Pre

	Log_V_pre_EVT	TICI_equal_higher2B	num_passes	imaging_recan
VIF	1.02	1.26	1.27	1.49

Regression Model for 5d vs Post

Observations	73 (25 missing obs. deleted)
Dependent variable	Log_V_5d
Type	OLS linear regression

F(5,67)	50.47
R^2	0.79
Adj. R^2	0.77

	Est.	2.5%	97.5%	t val.	p
(Intercept)	3.57	1.67	5.48	3.75	0.00
Log_V_post_EVT	0.36	0.01	0.72	2.02	0.05
TICI_equal_higher2BTICI>=2b	-2.82	-4.76	-0.88	-2.90	0.01
TMax_6s_post_EVT	-0.00	-0.01	0.00	-0.69	0.50
Log_V_post_EVT:TICI_equal_higher2BTICI>=2b	0.45	0.06	0.83	2.32	0.02
TICI_equal_higher2BTICI>=2b:TMax_6s_post_EVT	0.01	0.00	0.02	2.95	0.00

Standard errors: OLS

Table 29: VIFs of regression model for Post vs Pre

	Log_V_post_EVT	TICI_equal_higher2B	TMax_6s_post_EVT
VIF	1.12	2.45	2.47

Regression Model for 5d vs Pre

Observations	77 (21 missing obs. deleted)
Dependent variable	Log_V_5d
Type	OLS linear regression

F(6,70)	28.55
R^2	0.71
Adj. R^2	0.69

	Est.	2.5%	97.5%	t val.	p
(Intercept)	4.48	3.22	5.73	7.12	0.00
Log_V_pre_EVT	-0.01	-0.39	0.38	-0.03	0.97
TICI_equal_higher2BTICI>=2b	-3.77	-5.05	-2.49	-5.87	0.00
num_passes	0.18	0.08	0.28	3.64	0.00
TMax_6s_post_EVT	-0.00	-0.01	0.01	-0.57	0.57
Log_V_pre_EVT:TICI_equal_higher2BTICI>=2b	0.70	0.29	1.12	3.38	0.00
TICI_equal_higher2BTICI>=2b:TMax_6s_post_EVT	0.02	0.01	0.03	3.59	0.00

Standard errors: OLS

Table 30: VIFs of regression model for 5d vs Pre

	Log_V_pre_EVT	TICI_equal_higher2B	num_passes	TMax_6s_post_EVT
VIF	1.04	2.58	1.02	2.51

6.3 Appendix 3: Generalized linear models

Regression Model for Post vs Pre

Observations	93
Dependent variable	V_post_EVT
Type	Generalized linear model
Family	Gamma
Link	identity

$\chi^2(4)$	111.50
Pseudo-R^2 (Cragg-Uhler)	0.76
Pseudo-R^2 (McFadden)	0.15
AIC	744.82
BIC	760.01

	Est.	2.5%	97.5%	t val.	p
(Intercept)	85.49	43.61	127.36	4.00	0.00
V_pre_EVT	1.28	1.01	1.55	9.36	0.00
numOne pass	-82.86	-125.05	-40.66	-3.85	0.00
TICI_equal_higher2BTICI>=2b	-81.97	-123.86	-40.07	-3.83	0.00
numOne pass:TICI_equal_higher2BTICI>=2b	79.55	37.32	121.79	3.69	0.00

Standard errors: MLE

Table 31: VIFs of regression model for Post vs Pre

	V_pre_EVT	num	TICI_equal_higher2B
VIF	1.16	1.16	1.01

Regression Model for 5d vs Post

Observations	73 (20 missing obs. deleted)
Dependent variable	V_5d
Type	Generalized linear model
Family	Gamma
Link	identity

$\chi^2(2)$	74.25
Pseudo-R² (Cragg-Uhler)	0.71
Pseudo-R² (McFadden)	0.12
AIC	653.02
BIC	662.18

	Est.	2.5%	97.5%	t val.	p	VIF
(Intercept)	2.52	0.02	5.02	1.98	0.05	NA
V_post_EVT	1.20	0.82	1.59	6.07	0.00	1.09
oclus_post_EVTYes	38.81	15.59	62.03	3.28	0.00	1.09

Standard errors: MLE

Regression Model for 5d vs Pre

Observations	73 (20 missing obs. deleted)
Dependent variable	V_5d
Type	Generalized linear model
Family	Gamma
Link	identity

$\chi^2(4)$	67.94
Pseudo-R² (Cragg-Uhler)	0.66
Pseudo-R² (McFadden)	0.11
AIC	669.40
BIC	683.14

	Est.	2.5%	97.5%	t val.	p
(Intercept)	108.09	14.84	201.34	2.27	0.03
V_pre_EVT	-0.26	-1.83	1.30	-0.33	0.74
TICI_equal_higher2BTICI>=2b	-103.95	-197.28	-10.63	-2.18	0.03
oclus_post_EVTYes	48.43	20.55	76.31	3.41	0.00
V_pre_EVT:TICI_equal_higher2BTICI>=2b	1.41	-0.23	3.05	1.68	0.10

Standard errors: MLE

Table 32: VIFs of regression model for Post vs Pre

	V_pre_EVT	TICI_equal_higher2B	oclus_post_EVT
VIF	1.02	1.07	1.09

6.4 Appendix 4: Mixed model

Observations	268
Dependent variable	volume
Type	Mixed effects linear regression

AIC	783.85
BIC	816.17
Pseudo-R^2 (fixed effects)	0.19
Pseudo-R^2 (total)	0.77

Fixed Effects

	Est.	2.5%	97.5%	t val.	d.f.	p
(Intercept)	1.42	0.61	2.23	3.44	121.42	0.00
num_passes	0.17	0.03	0.31	2.44	94.03	0.02
timeLog_V_post_EVT	1.96	1.45	2.47	7.57	166.09	0.00
timeLog_V_5d	2.72	2.14	3.29	9.27	170.11	0.00
TICI_equal_higher2BTICI>=2b	0.63	-0.12	1.38	1.66	133.89	0.10
timeLog_V_post_EVT:TICI_equal_higher2BTICI>=2b	-1.73	-2.28	-1.19	-6.19	166.25	0.00
timeLog_V_5d:TICI_equal_higher2BTICI>=2b	-2.10	-2.72	-1.48	-6.67	169.96	0.00

p values calculated using Kenward-Roger standard errors and d.f.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: volume ~ (1 | subject) + num_passes + time * TICI_equal_higher2B
##    Data: d_melted_imp2
##
```

Random Effects		
Group	Parameter	Std. Dev.
subject	(Intercept)	1.09
Residual		0.68

Grouping Variables		
Group	# groups	ICC
subject	98	0.72

```
## REML criterion at convergence: 765.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7702 -0.3394  0.0640  0.3536  3.2871
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 1.1958   1.0935
##  Residual             0.4684   0.6844
## Number of obs: 268, groups:  subject, 98
##
## Fixed effects:
##                                                 Estimate Std. Error t value
## (Intercept)                                      1.42237    0.41350   3.440
## num_passes                                       0.16975    0.06947   2.443
## timeLog_V_post_EVT                               1.95943    0.25868   7.575
## timeLog_V_5d                                     2.71754    0.29311   9.271
## TICI_equal_higher2BTICI>=2b                      0.63063    0.38077   1.656
## timeLog_V_post_EVT:TICI_equal_higher2BTICI>=2b -1.73471    0.28034  -6.188
## timeLog_V_5d:TICI_equal_higher2BTICI>=2b       -2.10019    0.31482  -6.671
##
## Correlation of Fixed Effects:
##             (Intr) nm_pss tmL_V__EVT tmL_V_5 TICI__ tL_V__EVT:
## num_passes  -0.552
## tmLg_V__EVT -0.313  0.000
## timeLg_V_5d -0.284  0.015  0.441
## TICI__2BTIC -0.870  0.209  0.340      0.303
## tL_V__EVT:T  0.290 -0.002 -0.923     -0.407  -0.366
## tL_V_5:TICI  0.267 -0.018 -0.411     -0.931  -0.329  0.442
```

Bibliography

[1] Bates, D., Mächler, M., Bolker, B. and Walker, S. 2019. El atlas del ictus. *Sociedad Española de Neurología*. (2019).

[2] Bates, D., Mächler, M., Bolker, B. and Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*. (2015).

[3] Buuren S, G.-O.K. van 2011. Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software, Articles*. (2011).

[4] Buuren, S. van 2012. *Flexible imputation of missing data*. Chapman; Hall.

[5] Campbell, D.S., B. C.V. 2019. Ischemic stroke. *Nat Rev Dis Primers*. (2019).

[6] Cayuela, L. 2010. *Modelos lineales mixtos en r*. Universidad de Granada.

[7] Frank E. Harrell, J. 2006. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer Series in Statistics.

[8] Higashida RT, R.H., Furlan AJ 2003. Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. *AHA Journals Home*. (2003).

[9] John Fox, B.P., Sanford Weisberg 2020. *Effects: Effect displays for linear, generalized linear and other models*.

[10] Little RJ, R.D. 2019. *Statistical analysis with missing data*. Wiley.

[11] Long, J.A. 2021. *Tools for summarizing and visualizing regression models*. jtools: Analysis; Presentation of Social Scientific Data.

[12] McCullagh, P. and Nelder, J. 1989. *Generalized linear models*. Chapman; Hall.

[13] Nissen, J., Donatello, R. and Van Dusen, B. 2019. Missing data and bias in physics education research: A case for using multiple imputation. *Physical Review Physics Education Research*. 15, (Jul. 2019).

[14] Pérez, M.H. *Neuroimagen en el ictus agudo: Búsqueda de nuevos biomarcadores subrogados del pronóstico clínico*. UAB.

[15] Prof Bruce C V Campbell, P.P.K. 2020. Stroke. *The Lancet*. (2020).

[16] Russell V. Lenth, Paul Buerkner, Maxime Herve, Jonathon Love, Hannes Riebl, Henrik Singmann 2021. *Emmeans: Estimated marginal means, aka least-squares means*. The American Statistician.

[17] Senn, S. 2007. *Statistical issues in drug development*. Wiley.

[18] Stef van Buuren, K.G.-O. *Multivariate imputation by chained equations*. amices.github.io.

[19] *Código ictus*. FEI Federación Española del Ictus.