Lund University **Spring 2024**

Blocking the swell

Beta blockers' association with brain edema for glioblastoma patients

Hanna Frederiksen
Amanda Hillström

Contents

Abstract

Glioblastoma is a highly aggressive brain tumor that kills roughly 9,000 Americans yearly. Although a known disease for over a century, little progress has been made in treatment development. A breakthrough may be on its way as recent research suggests glioblastoma patients can be treated with virus injections, oncolytic virotherapy in medical terms. However, this treatment method is somewhat inhibited by steroids. Patients are given steroids around the time of surgery to treat the sometimes fatal fluid imbalance, edema, in the tumor region. Clinical inclinations have shown that beta blockers, normally used to treat heart conditions, could reduce the edema and at the same time allow for virotherapy, potentially prolonging the lives of glioblastoma patients.

This thesis aims to answer the question if glioblastoma patients on beta blockers suffer less from edema at the time around brain tumor surgery. With no direct measurement of the severity of the edema being available in this study, an indirect measure, prescribed steroid (dexamethasone) dosage, is used. Through controlling for confounders in the case and control group, performing significance tests, and several linear regression models, the conclusion was that no such association could be established with the given data. Clinical inclinations seemed to depend more on the other characteristics of the treatment group such as age and extent of resection rather than the effect of the beta blocker. However, more studies need to be done before any firm conclusions can be drawn. Namely, a study of experimental nature that analyses x-rays of the tumor rather than the prescribed steroid dosage to directly observe the severity of the edema.

1 Introduction

1.1 Background

Every year, around 12,000 Americans are diagnosed with glioblastoma (1). Glioblastoma is a type of brain tumor that represents roughly half of all primary malignant brain tumor cases (2). It is known for being notoriously hard to treat, with low survival rate. Currently, the five-year survival rate is 6.9% and the average survival length about eight months.

Cancer is diagnosed in stages ranging from one to four with four being the most severe. Glioblastomas often arise de novo, meaning that they begin as a grade four tumor with no evidence of a lower-grade precursor (1). They are fast-growing and consist of a mixture of cell types which makes them even harder to treat.

When a patient is diagnosed, surgeons attempt to remove the tumor as fast as possible. Chances of survival increase with early removal, as the risk of tumor growth and spread is decreased (1). However, it is nearly impossible to take out the entire tumor given the invasive nature of glioblastoma on nearby brain tissue (3). Parts of the tumor therefore remain on the brain after surgery.

The tumor itself and the trauma of the surgical intervention can cause edema. As explained by Stummer (4), edema in the brain is swelling often caused by accumulation of excess fluid in the brain tissues. As the rigid skull cannot adapt to the fluid change, the intracranial pressure rises, and often cause patients to experience symptoms such as headache. In extension, edema can cause neurological disturbances and reduce blood flow to the brain. Left untreated, brain tissue is damaged and vital functions may be destroyed. In worst case, edema is fatal. To treat it, patients are given steroids that inhibit inflammation. The most common one being dexamethasone (DXM).

Recent research suggests that oncolytic virotherapy may be an efficient way to remove glioma cells (3). By injecting a virus into the patient, anti-tumor immune responses may be triggered that cause the patient's own immune system to attack and destroy the tumor. However, the virotherapy is inhibited whilst the patient is on a high steroid dosage, as high doses of steroids suppress the immune response generated by the virus (5). If virotherapy proves to be a successful treatment method, there is need for an alternative treatment for edema that does not impose restrictions on the virus.

Neurosurgeons at a major neuroscience center on the US east coast have a clinical inclination to believe that beta blockers may reduce brain tumor swelling (3). This project has been performed in collaboration with these doctors, using their data. However, the hospital name is left out of the report.

Beta blockers are used mainly for patients with cardiovascular diseases or hypertension (high blood pressure) (6). It is often in the form of a pill that is ingested daily. Patients may take beta blockers for several years. Using beta blockers instead of steroids to prevent edema would allow for effective oncolytic virotherapy.

This research investigates these initial inclinations to answer the question: Does the use of beta blockers reduce the need of steroid treatment for glioblastoma

tumors in connection with surgery?

1.2 Research question

The hypothesis is that patients on beta blockers will be prescribed a smaller dose of steroids as compared to patients that are not on beta blocker. This is based on the collaborating doctors' clinical experience (3), and there has also been a few published studies that point towards the benefit of using beta blockers for glioblastoma treatment. In particular, a review of ten preclinical studies and one clinical found that beta blockers may reduce tumor proliferation (the rapid cell growth and division in cancer cells) (7). A study specifically carried out on the beta blocker type propanolol found that it suppressed glioblastoma proliferation in a dose-dependent manner (8).

The ambition of this study is to contribute to the field of research by mainly focusing at the relationship between beta blockers use and edema, not tumor growth. Exploring alternatives to steroids for treating swelling is of high importance, as it is an enabling factor for virotherapy. Little has happened in glioblastoma research over the years, and life length has only been prolonged by a couple of months since its discovery in the 1920's (2). Virotherapy shows great potential and could hopefully prolong the lives of the suffering glioblastoma patients (3).

Before diving into the study, it is found necessary to define what a result in favor of the hypothesis would look like. This is to avoid the confirmation bias of looking for a significance until its found. Should 20 significance tests be performed at a 0.05 significance level, it is expected that one would show a false positive.

For this research, it was therefore decided that two different models should be built on the dataset and be based upon the potential confounding factors identified (and presented later). The choice of two rather than one was to explore different mathematical approaches to the same question, but it was noted that the significance levels needed to be stricter than the regular 0.05 in case the results showed an association.

2 The data set

2.1 Data overview

The research was carried out on a data set from a US hospital with a little over 700 patients that underwent surgery for glioblastoma tumors between 2010 and 2023. Personal identifiers were removed from the data beforehand, and every patient was instead assigned a made-up study ID. For every patient, 60 information points were included covering:

- Demographics (e.g age, sex, ethnicity)
- Tumor-specific information (e.g. size, location, spread)

- Treatment (e.g. beta blocker use, other medicines, dosages, chemotherapy, radiation)
- Surgery outcome (e.g. death, months to death post surgery)

Regarding beta blocker usage, the following information was available: if taking a beta blocker (yes/no), beta blocker type (Atenolol, Metoprolol, Nebivolol, Acebutolol, Sotalol, Propranolol, Labetalol) and usage time (months). The specific time periods during which each patient used beta blockers were not available in the data. This issue is problematic and will be addressed later in the report. In the analysis plain beta blockage use or not was looked at, and it was decided to not perform the analysis on the specific subgroups of beta blocker type or usage time. This decision was made due to the fact that further division to subgroups would entail a very small sample size, as well as the fact that the neurological experts involved did not believe it would make a great difference.

The steroid prescribed to the patients in this data is dexamethasone. There are four datapoints regarding prescribed DXM: dose right before surgery, dose right after surgery, dose two weeks after surgery and dose six to eight weeks after surgery. The dexamethasone dose varied between 0 and 60 mg daily per patient and time of measurement. Another measurement was then included; the summed dose of DXM for each patient. After discussing with the neurological experts, it was decided that the half-life of DXM was considered long enough for accumulated dose to be relevant (3).

2.2 Missing values

The research explored the association between beta blocker and prescribed DXM, making these variables especially important. To handle missing data for these variables it was therefore decided that any imputation method could introduce too large of a bias, and that patients with any missing data for these variables were to be removed.

For 222 of the earliest surgeries, there was no information on the patients' beta blocker usage. This was most likely due to this information not being considered relevant at the time, and could possibly be found by manually going through the patients' medical charts. This was however not done for this study, and these patients were therefore excluded. Additionally there was one patient with only null values across all information points that was removed.

Three patients had missing values across all DXM data points. An additional 36 patients had missing values on at least one of their DXM doses. The largest data loss was for DXM dose six to eight weeks post surgery, which constituted 32 out of the 36 patients. All 36 patients were excluded. After the removal, the final data contained 466 surgeries ranging from 2016 to 2023. The steps for removal of patients with missing values are visualized in Figure 1.

It is worth mentioning that the patients that for example were removed for having missing values for DXM dose six to eight weeks post surgery could have been included for the analysis of the earlier stages of DXM dosage. The reason for full exclusion was to be concise in the data, making the analysis of the different stages of DXM dose more comparable as it was done on the same

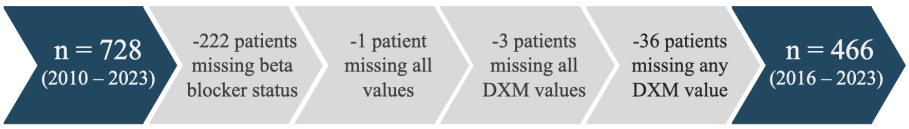

Figure 1: Process of removing missing values to final data

patients. Also, the accumulated DXM dosage could not be retreived for patients with missing DXM values.

19 out of the 32 patients with missing DXM dose six to eight weeks post surgery died within two months post surgery. The reason for the missing data could be that they did not live long enough to receive their final dose. The removal of these implies the removal of a tail of unsuccessful treatments, and probably important samples. As the scope of the research is for the patient to be able to receive DXM (if necessary) at all four occasions, the decision to exclude them is supported, but noted and does affect the distribution of the data.

After the removal of patients with missing information on beta blocker usage and the four different occasions of DXM dosage, there are still patients with missing values for other variables. The removal of these would unnecessarily decrease an already small sample, and as some analysis can be done regardless, these patients were not initially removed. The handling of these will however be addressed in future analysis.

2.3 Selection of case group

In this study, the case group consisted of the patients that were on beta blockers. Beta blocker use was the observed independent variable, and the dependent variable is DXM dosage. The control group naturally became the patients who did not take beta blockers. Note that out of the 466 samples, only 104 patients were on beta blockers. The control group consisted of the remaining 362 patients.

Given the nature of the study - a retrospective, observational study - the case group, patients who take beta blockers, was not randomly selected. These patients were on beta blockers for numerous reasons, and there is certainly a selection bias in the dataset. These patients most likely have, apart from glioblastomas, cardiovascular diseases or hypertension (3), which both could impact surgery outcome in itself, as well as being related to a certain type of patient profile.

There are two types of differences between the case and control group. There are the ones that can be controlled for as they are given in the patient information (one of the 60 data points), and the ones that can only be discussed as it is not seen in the patient data.

The differences that can be adjusted for are sometimes clearly visible in the data, e.g., the case group is over represented by older males. As there are more than 60 information points but only a sample of 466 patients, variables that were thought to potentially affect prescribed DXM had to be selected. In this report, two types of variables amongst these 60 will be discussed; covariates and

confounders (9).

- Covariate: A variable with an impact on the dependent variable, DXM dosage
- Confounding factor: A variable with an impact on the dependent variable that is different between the case and control group and not an effect of the treatment

The differences between the case and control group that could not be distinguished from the given patient information, and therefore not adjusted for, can only be discussed. This will be done further in section 5 - Sensitivity analysis.

2.4 Potential covariates

In every study there are biases. Especially when studying the human anatomy, where there are many hidden causalities that risk masking the true relations. To avoid drawing false conclusions based on confounding effects, it was necessary to identify all covariates and compare their distribution in the case and control group to see if they were also confounders. After discussions with the neurological experts at the hospital, the following six potential covariates (from the 60 initial variables) were brought up for further analysis (3):

- *Extent of resection (EOR)*: How much of the tumor that was removed in surgery; biopsy, sub-total resection, and gross-total resection. Biopsy is a small tumor sample taken to analyze the nature of the tumor - a different surgical intervention compared to the sub-total and gross-total resection that attempts to remove as much of the tumor as possible (3). Full removal is not possible in practice, however gross-total resection implies that all tumor that was visible on the preoperative MR scan was removed.
- *MGMT status*: This is part of the genetic profiling of glioblastoma, and is short for O6-methylguanine-DNA methyltransferase (10). The tumor is classified as unmethylated, weakly methylated or methylated, affecting how well it responds to certain chemotherapy drugs. A higher degree of methylation may imply a more successful treatment.
- *IDH1*: IDH refers to isocitrate dehydrogenase mutation brain tumors, and is also a part of the genetic profiling of glibolatomas (10). Glibolastomas may have mutations in IDH enzymes, including mutation in the gene for the IDH1 isoform, an enzyme involved in cell metabolism (11). They can either be classified as IDH1 wild type or IDH1 mutant.

  IDH1 mutant suggest a secondary glioblastoma tumor, meaning that it was previously a lower grade glioma (10). The mutant status therefore carries a better prognosis than a wild type status. The IDH wild type occurs in about 90 percent of glioblastomas, and IDH mutant represent the remaining 10 percent. In some rare cases, it can not be specified wether the glioblastoma is wild type or mutant, however, no such case exist in this data.
- *Stupp complete*: Was the Stupp protocol followed throughout the patient treatment, yes or no. The Stupp protocol is a standard of care for glioblastomas. Consulted neurological experts (3) meant that a completed Stupp

protocol could indicate an in general healthier patient, as the main reason for not following through is that a patient is too ill to benefit from further treatment.

- *Tumor location*: The brain tumor location in the dataset are specified as supratentorial, infratentorial, and not known. Brain tumors located in the lobes are called supratentorial, while tumors located in the cerebellum or brainstem are called infratentorial (12).

  Clinical inclination among the consulted experts, as well as several studies suggest that headache is more common in patients with infratentorial tumors than with supratentorial tumors (13, 14). As headache is a common symptom of tumor swelling, as well as one of the criteras clinically used for prescribing patients DXM, it is relevant to include.

- *Tumor size*: The size of the tumor is measured preoperatively by taking the largest diameter in cm. Although there was no reliable research found connecting tumor size to headache or swelling, it is not unreasonable that there could be a relation, and it was therefore included.

These were all classified as relevant variables as they could have a potential affect on tumor swelling, headaches or any other factor affecting the prescribed dexamethasone dosage. In addition to this, it was decided to include two other variables:

- *Age at surgery*: Age may be an indication of the overall health of the patient. It was also a known factor that would differ in distribution between the case and control group; beta blocker patients are in general older. When the variable age is mentioned in the report, it is age at surgery that is being referred to.

- *Sex*: Just as age, sex was a variable believed to differ in the case and control group. There are big gender gaps in research, where for example clinical trials are often preformed on men, making them less accurate for females (15). Therefore, it is often wise to include sex in studies even if there are no clinical inclinations to support the decision.

2.5 Overview of all variables

The variables to be explored in this analysis have now been established and presented. Table 1 displays an overview of the dependent, independent and controlled variables, and highlights what type of variable it is, continuous, count or categorical, as well as potential categorical subtype. For further overview of the different categories within the categorical variables, see Table 2.

3 Initial analysis

To gain an initial understanding on whether or not patients on beta blockers are prescribed a statistically significantly lower DXM dosage, the following steps were taken:

Table 1: Overview of all variables

Variable	Role	Type	Subtype
Pre-op DXM dose	Dependent variable	Continuous	
Post-op DXM dose	Dependent variable	Continuous	
2w post-op DXM dose	Dependent variable	Continuous	
6-8w post-op DXM dose	Dependent variable	Continuous	
Total DXM dose	Dependent variable	Continuous	
Beta blocker	Independent variable	Categorical	Nominal
EOR	Controlled variable	Categorical	Ordinal
Age at surgery	Controlled variable	Continuous	
MGMT	Controlled variable	Categorical	Ordinal
Stupp complete	Controlled variable	Categorical	Nominal
IDH1	Controlled variable	Categorical	Nominal
Tumor location	Controlled variable	Categorical	Nominal
Sex	Controlled variable	Categorical	Nominal
Tumor size	Controlled variable	Continuous	

Table 2: Overview of the categorical variables' categories

Variable	Categories
Beta blocker	yes, no
EOR	biopsy, sub-total resection, gross-total resection
MGMT	unmethylated, weakly methylated, methylated
Stupp complete	yes, no
IDH1	wild type, mutant
Tumor location	supratentorial, infratentorial
Sex	female, male

1. Data cleaning
2. Look at the distribution of potential covariates in case and control group
3. Check if covariates actually show correlation with prescribed DXM
4. Control for identified confounders
5. Check if prescribed DXM dosage is significantly different between case and control group

3.1 Data cleaning

In the section 2.2 Missing values, all patients with missing information on beta blocker status and DXM dosage were removed. The possibility to remove additional patients with missing information for the controlled variables was also addressed, but however not performed. This will not be preformed in this section either, but it should be noted that for some of the tests they were automatically be excluded.

Except for handling missing values there are other parts of data cleaning, such as addressing outliers in the data. The reason to remove outliers in this type of analysis would be if they showed to be clearly wrong. Studying the data together with the doctors, no outliers were immediately detected, and therefore no

additional points were removed at this stage. Potential outliers and influential observations will however be discussed and presented later in the report.

3.2 Look at the distribution of potential covariates in case and control group

To decide if the included potential covariates are also confounders, it is necessary to identify whether the distribution of these are equal in the control and case group. In this initial analysis, only confounders are controlled for, while ensuring none of the covariates become confounders in the process.

Table 3 shows an overview of the distribution in the case and control group for the categorical control variables, while Figure 2 and 3 shows the distribution for the continuous control variables age at surgery and tumor size.

Studying the frequency tables below, it is noted that for some of the categories (namely MGMT and tumor location) the number of patients do not sum up to 466 patients. This is due to the fact that there are missing information for some of the patients for these variables. Additionally, for some categories there are very few patients, such as infratentorial, making the result for these categories very unreliable. Both of these error sources will for now be ignored, but addressed in future analysis.

Table 3: Frequency tables for control variables in case and control group (n = 466). Percentage of patients (number of patients) Stupp refers to Stupp complete.

Extent of resection	Case	Control
Biopsy	0.96% (1)	3.59% (13)
Sub-total resection	48.08% (50)	38.67% (140)
Gross-total resection	50.96% (53)	57.73% (209)

MGMT	Case	Control
Unmethylated	57.69% (60)	45.51% (162)
Weakly methylated	1.92% (2)	8.15% (29)
Methylated	40.38% (42)	46.35% (165)

Sex	Case	Control
Female	37.50% (39)	47.24% (171)
Male	62.50% (65)	52.76% (191)

Stupp	Case	Control
Yes	22.12% (23)	24.59% (89)
No	77.88% (81)	75.41% (273)

IDH1	Case	Control
Wild type	98.08% (102)	96.69% (350)
Mutant	1.92% (2)	3.31% (12)

Tumor location	Case	Control
Supratentorial	99.03% (102)	98.88% (352)
Infratentorial	0.97% (1)	1.12% (4)

The following tests were performed to see if the control variables were evenly distributed in the case and control group. Given that some control variables are continuous and some are categorical (ordinal and nominal), different tests had to be used:

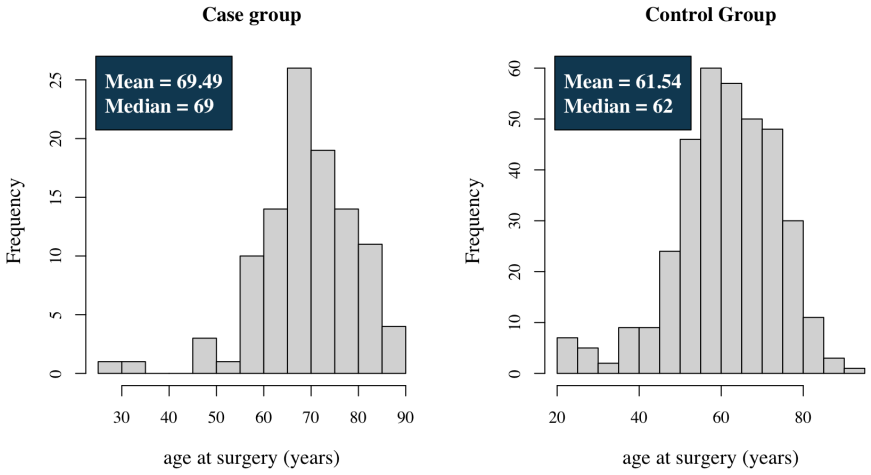

Figure 2: Histograms for age at surgery for case and control group

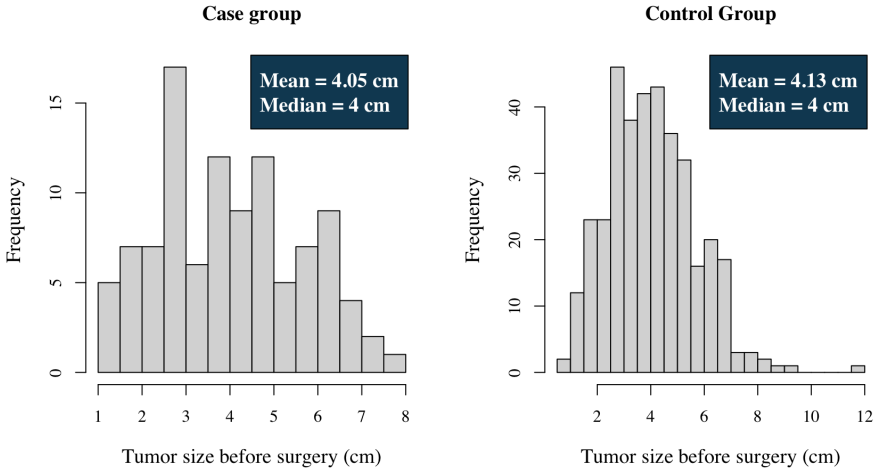

Figure 3: Histograms for tumor size preoperatively for case and control group

- Pearson's χ^2 test: MGMT, Stupp complete, Sex
- Fisher's exact test: EOR, IDH1, Tumor location
- The Mann-Whitney U test: Age at surgery, Tumor size

There are several ways to tackle differences in covariate distributions in case and control group. P-value tests are one of them. Other sources suggest that a ten percentage point difference is a reasonable threshold, as p-value tests are highly sensitive for sample size (16). This knowledge was combined with the p-value tests when deciding what covariates to control for.

3.2.1 Pearson's χ^2 test

The Pearson's χ^2 test (17) is used for categorical variables presented in contingency tables. It should determine if there is a statistically significant difference in the expected frequencies between the categories, e.g., if males are overrepresented in the beta blocker group. That the test statistic follows a χ^2 distribution builds upon the assumption that the samples are independent, which they are in the context of individual patients.

$$H_0: \quad \text{There are no differences between the classes in the population}$$

If the null hypothesis could be rejected at a 95% confidence level, the conclusion was that there was a significant difference between the case and control group with regards to that variable. It should be noted that when H_0 could not be rejected, that does not imply that the groups have an equal distribution. Therefore, a manual overview of the frequency tables was also needed.

The χ^2 p-value is not exact, but an approximation that becomes correct when sample size is infinite. It is therefore suited for large populations, but is not robust to handle categorical frequencies below five (17). This is why it was not suited for EOR, IDH1 or Tumor location, which all had at least one category within each group with fewer than five patients. There are several options for handling these cases, Fisher's exact test was used in this study.

3.2.2 Fisher's exact test

Fisher's exact test (18) is usually performed for smaller sample sizes as it calculates the exact p-value rather than an approximation. The underlying assumption is that the two classification characteristics are not associated, e.g., that a patient on beta blockers does not have a given tumor location.

$$H_0: \text{The probability that a random sample from the case group is from a certain category is equal to the probability from the control group}$$

If the null hypothesis could be rejected at a 0.05 significance level, the conclusion was that there was a significant difference between the case and control group with regards to that variable.

3.2.3 The Mann-Whitney U test

To analyze if the distribution of each continuous variable e.g., age at surgery (A), differed between the case (BB) and control (C) group, the Mann-Whitney U test (19) was used. The test builds upon the hypothesis that for two populations BB and C:

$$\begin{aligned} H_0: &\quad \text{The distribution of } A_{BB} \text{ and } A_C \text{ is identical} \\ H_1: &\quad \text{The distribution of } A_{BB} \text{ and } A_C \text{ is not identical} \end{aligned}$$

If H_0 could be rejected on a 0.05 significance level, it is determined that the distribution of A was different between the groups.

The general assumptions of the test are:

1. All observations are independent of each other
2. The responses are at least ordinal

The benefit of using the Mann-Whitney U test is that no underlying assumption of the distribution needs to be made (19). However, the distributions need to be similar, meaning that one of the groups cannot have too many extreme values compared to the other one. In this case, the data was not normally distributed, which meant that a t-test could not be used, and this was found to be a good alternative.

3.2.4 Results

After the tests were performed, it was determined that the distribution of age at surgery and MGMT status differed significantly between case and control group. In Table 4 the result for the different variables can be seen.

Table 4: Difference in covariate distribution between case and control group

Variable	Test	Significant difference	P-value
EOR	Fisher's	No	0.13
Age at surgery	U-test	Yes	0.00
MGMT	χ^2	Yes	0.02
Stupp complete	χ^2	No	0.70
IDH1	Fisher's	No	0.74
Tumor location	Fisher's	No	1
Sex	χ^2	No	0.10
Tumor size	U-test	No	0.71

It should be highlighted that all assumptions of these tests may not be fulfilled given the relatively low sample size. Therefore, it was necessary to study the significance in combination with the frequency tables shown in Table 3. Although H_0 was not rejected for EOR and sex, the distribution seemed to differ (approaching the 10% limit for some categories), with sub-total resection and men being overrepresented in the case group.

This analysis was conducted to see if a fair comparison could be made between case and control group. The next step was to explore if the potential covariates could be empirically proven to impact DXM dosage. The idea was to control for the confounders - variables that were distributed differently between case and control group *and* had an impact on DXM dosage.

3.3 Check if covariates actually show correlation with prescribed DXM

After identifying the potential skewness in case and control group for the above mentioned covariates, their respective impact on DXM dosage was explored. It

was done as a rough estimate to get an initial sense of the data. The covariates potential interaction with each other was however not explored at this point.

As age at surgery and MGMT status were balanced differently in the case and control group, it was especially interesting to see if they had a significant impact on DXM dosage. If they did, that meant no straight comparison of DXM dosage across case and control could be done without controlling for these first. Otherwise, it would not be possible to know if a difference in prescribed DXM dose was an effect of beta blockers, or e.g., age.

This is a consequence of working with observational studies and with medical data in general. In an experimental study, it would be possible to control for many of these variables beforehand. By randomly assigning patients to either the case or control group, the distribution of both known and unknown covariates would likely be more similar in case and control groups.

To avoid that the effect from beta blockers would impact e.g. age at surgery's correlation with steroid use, the tests were performed separately on all controlled variables within the case and control groups. This was a compromise with sample size, but thought to be the solution that made the most sense.

For the continuous variables (age at surgery and tumor size) and the ordinal variables (EOR and MGMT) linear correlation was asses using Pearson's product moment correlation, and monotonic relations were evaluated using Spearman's rank correlation. The binary variables (Stupp complete, IDH1, tumor location, and sex) were analyzed using the Mann-Whitney U-test.

3.3.1 Pearson's product moment correlation ρ

Pearson's correlation, ρ, is a common way of measuring linear dependence. Given a pair of random variables, e.g., age at surgery (A), DXM dosage before surgery (D), the Pearson correlation coefficient ρ for a sample of pairs (A, D) is given by (9):

$$\rho_{A,D} = \frac{\sum_{i=1}^{n}(A_i - \bar{A})(D_i - \bar{D})}{\sqrt{\sum_{i=1}^{n}(A_i - \bar{A})^2}\sqrt{\sum_{i=1}^{n}(D_i - \bar{D})^2}}$$

If the pairs are from an uncorrelated bivariate normal distribution, the sampling distribution of the studentized Pearson's correlation coefficient follows a t-distribution with $n - 2$ degrees of freedom, n being the sample size (9). A two-sided test was performed at a 95% confidence level to asses if this was the case.

If the pairs are uncorrelated but not normally distributed, the degrees of freedom are reduced. The rather strict assumptions of the Pearson's ρ (linear dependence, normal distribution) encouraged more dependency tests.

3.3.2 Spearman's rank correlation ρ

The Spearman rank correlation (20) between two variables is equal to the Pearson correlation between the rank values of those two variables. Rank is the

sorted order of the nominal or categorical values. For example, with three values for age at surgery: 65, 73, 69, the corresponding rank values would be 1, 3, 2. Spearman's rank asses monotonic relationships (linear and non-linear), whereas Pearson's assesses strictly linear relationships.

Spearman's rank is high when observations have a similar rank between the two variables. For example, if a high extent of resection is perfectly correlated with the highest rank of DXM dosage, the Spearman ρ would be equal to one. A Spearman's rank of +1 or -1 indicates that the variables are perfect monotone functions of each other. (20)

Spearman's rank is appropriate for both continuous and discrete (ordinal) variables. It was used as a complement to ensure non-linear relationships were also explored. This was also a way to cover up for the breach of the normal distribution assumption when using Pearson's ρ. The Kendall's rank correlation τ was also calculated, but gave the same indication and is therefore not explained further.

3.3.3 The Mann-Whitney U test

For the binary input variables: stupp complete, sex, tumor location, and IDH1, the tests mentioned above could not be used as they require continuous or ordinal data. The Mann-Whitney U test was therefore performed to see if there were any significant difference in prescribed DXM dose between the binary categories across all DXM measurements. Similar to the other tests, this analysis was conducted separately for the case and control group.

3.3.4 Results

As the Pearson's ρ is more strict than Spearman's rank, and given the same indications, these linear correlations are presented in Table 24 and 25 in the Appendix. Age at surgery had a negative correlation with total DXM dosage in both case and control group. This can be seen in Tables 5 and 6, where the Spearman's rank coefficient is displayed for both the case and control group, respectively. Age at surgery's correlation with prescribed DXM dose is especially interesting as the distribution of age at surgery also varied between the case and control group. It seems that DXM dosage is in general lower for older patients. Without controlling for the skewed age distribution, one would risk to get false indications from a comparison of DXM dosage between the beta blocker and non-beta blocker patients. A relation between MGMT status and DXM dosage was not shown, hence the need to control for this variable is smaller.

Given that the distribution of EOR was slightly different in the groups, it should be kept in mind going forward that a larger extent of resection showed a negative correlation with DXM dosage. From a medical perspective, it could be argued that a more successful resection would lower the steroid dosage (3). If a larger part of the tumor is removed, the remains are smaller and will cause less complications. On the other hand, a larger surgical intervention could cause a bigger trauma leading to more severe edema. As for tumor size, a larger tumor correlated with a larger dosage of DXM. Tumor size, however, did not differ significantly between the patient groups, and was therefore not regarded a red flag.

The results were not too different between beta blockers and non-beta blockers as seen in Tables 5 and 6. That there are small variations was expected given the small sample size.

Table 5: Case group: Spearman's rank correlation ρ p-value (ρ if significant)

DXM	Age	EOR	MGMT	Tumor size
Pre-op	**(-0.32)** 0.00	0.97	0.97	**(0.27)** 0.01
Post-op	0.82	**(-0.31)** 0.00	0.81	**(0.21)** 0.03
2w post-op	0.07	**(-0.19)** 0.05	0.75	0.89
6-8w post-op	0.74	0.11	0.98	0.14
Total	**(-0.29)** 0.00	**(-0.24)** 0.02	0.83	**(0.31)** 0.00

Table 6: Control group: Spearman's rank correlation ρ p-value (ρ if significant)

DXM	Age	EOR	MGMT	Tumor size
Pre-op	0.09	0.52	0.62	**(0.28)** 0.00
Post-op	0.12	0.99	0.43	0.59
2w post-op	0.27	**(-0.11)** 0.04	0.71	0.86
6-8w post-op	0.31	**(-0.15)** 0.01	0.92	0.74
Total	**(-0.13)** 0.01	**(-0.10)** 0.05	0.57	**(0.20)** 0.00

For the binary variables there were few variables that had proven correlation with DXM dosage. As seen in Table 7, where Mann-Whitney's U-test is performed on the case group, there was a significant difference in the DXM dose two weeks post surgery for the variable sex. For the control group, Table 8 shows a difference in completion of Stupp protocol, IDH1 and tumor location.

What should be remembered from frequency Table 3 is that for stupp complete, IDH1 and tumor location there was a very small sample of patients that had completed the Stupp protocol, had IDH1 mutant type or infratentorial tumors, respectively. The low frequency makes it more difficult to find a significant difference. However, studying Table 3 the distribution of these covariates in the case and control group do not seem to differ much.

Table 7: Case group: Mann-Whitney's U-test p-value

DXM	Stupp complete	IDH1	Tumor location	Sex
Pre-op	0.71	0.18	0.35	0.92
Post-op	0.15	0.55	0.09	0.46
2w post-op	0.24	0.98	0.46	**0.00**
6-8w post-op	0.73	0.32	0.49	0.15
Total	0.97	0.91	0.13	0.13

Table 8: Control group: Mann-Whitney's U-test p-value

DXM	Stupp complete	IDH1	Tumor location	Sex
Pre-op	0.33	0.98	0.09	0.30
Post-op	0.31	0.81	0.53	0.72
2w post-op	0.22	0.29	0.07	0.79
6-8w post-op	**0.04**	**0.03**	0.34	0.26
Total	0.97	0.53	**0.02**	0.73

In order to answer the research question: does the use of beta blockers reduce the need of DXM, it is now shown that at least the age at surgery had to be

controlled for. EOR and tumor size also showed to be important covariates to keep in mind; however they would not necessarily need to be controlled for if their distributions were the same in the patient groups. Given that EOR was closer to the 0.05 limit and that the distribution tests cannot be fully trusted, it was treated as a confounder as well.

3.4 Control for identified confounders

There are several ways to control for confounders. In the initial analysis part of the report, the focus will be on controlling for the two confounding factors previously identified: age at surgery and extent of resection. To do this, the following two methods were used:

1. Propensity score matching
2. Inverse probability weighting

3.4.1 Propensity score matching

Propensity score matching is a method for addressing selection bias that was introduced by Paul R. Rosenbaum and Donald Rubin in 1983 (15). As this is an observational study, where the treatment group is not randomly selected, biases exists. Propensity score matching attempts to reduce the biases and mimic randomization by creating a sample of the control group that is comparable to the case group in regards to chosen covariates.

Employing propensity score matching usually comprises of four steps (21):

1. Estimating propensity score (outlined below)
2. Matching each patient in case group to one or several in control group based on propensity score
3. Evaluate matching quality
4. Conduct intended outcome analysis

A propensity score is defined as the probability of a unit being assigned to the treatment group given a set of covariates (21). To exemplify this, let B_i define whether or not a patient is in the treatment group, where i = 1,...,N is the total number of patients. $B_i = 1$ means that the patient was on beta blocker, while $B_i = 0$ means that the patient was not on beta blocker. Each patient also has a covariate value vector $\mathbf{X}_i = (\mathrm{X}_{i1}, ..., \mathrm{X}_{ik})$, where k is the number of covariates. The propensity score for unit i (in this case patient i) is thus defined as

$$p(\mathbf{X}_i) = Pr(B_i = 1|\mathbf{X}_i).$$

The propensity score matching method relies on two crucial assumptions regarding the treatment assignment (21):

1. $(Y_{1i}, Y_{0i}) \perp B_i \mid \mathbf{X}_i$
   This assumption states that the the treatment assignment B_i and response (Y_{1i}, Y_{0i}) are conditionally independent given the covariates $\mathbf{X}_i$. Here Y_{1i} denotes the potential outcome of unit i under treatment, and Y_{0i} denotes the potential outcome of unit i under no treatment.

2. $0 < p(\mathbf{X}_i\ ) < 1$
   This ensures that there are units with similar characteristics in both groups.

Under these assumptions, units in the treatment and control groups are balanced in terms of observed covariates, reducing selection bias. Consequently, conducting outcome analysis on the matched data post-matching tends to yield unbiased estimates of treatment effects, enhancing the credibility of treatment effect estimates.

Once the propensity score for each unit has been calculated, there are a choice of numerous matching methods to apply in order to create balanced groups based on the scores. It was decided to use nearest neighbor matching. This method matches each unit i in the treatment group with unit j in the control group based on the closest distance between between the logarithm of their propensity scores. It is calculated as (21):

$$d(i, j) = min_j\{|l(\mathbf{X}_i) - l(\mathbf{X}_j)|\},$$

where $l(\mathbf{X}_i) = \ln\{p(\mathbf{X}_i)/[1 - p(\mathbf{X}_i)]\}$ is the logit of the propensity score, which is commonly used as it has a better property of normality than $p(\mathbf{X}_i)$. There are many pros and cons with the different methods for matching, but in the end nearest neighbor matching was chosen due to its simplicity and effectiveness.

The identified confounders age at surgery and extent of resection were used as covariate vectors. The inclusion of all variables would result in a small size of matched data, therefore it was decided to only include these two, resulting in the covariate value factor $\mathbf{X}_i = (\mathrm{X}_{i,\mathrm{EOR}}, \mathrm{X}_{i,\text{age at surgery}})$. The result will be presented in section 3.5.

3.4.2 Inverse probability weighting

An alternative to propensity score matching that keeps all samples is to use a weighted approach. The step-by-step method was the following (16):

1. Fit a linear regression model for beta blockers against age and extent of resection
2. Use it to calculate the predicted probabilities (propensity score) of being on/off beta blocker based on the covariates
3. Create weights by inverting the propensity score
4. Weight the treatment and control groups using calculated weights so that they are as similar as possible with regards to the defined confounders.

The inverse probability weighting was, just like propensity score matching, based on controlling for age at surgery and extent of resection.

3.5 Significance tests

After ensuring that the case and control group were comparable, it was time to see how beta blocker usage actually was associated with DXM dose. Distribution and mean of DXM dose prescribed pre surgery, post surgery, two weeks after surgery and six to eight weeks after surgery, was compared for the case group against the control group. The accumulated DXM dose was also compared. In Figure 4 the comparison can be seen before covariate control, in Figure 5 after propensity score matching and in Figure 6 after inverse probability weighting.

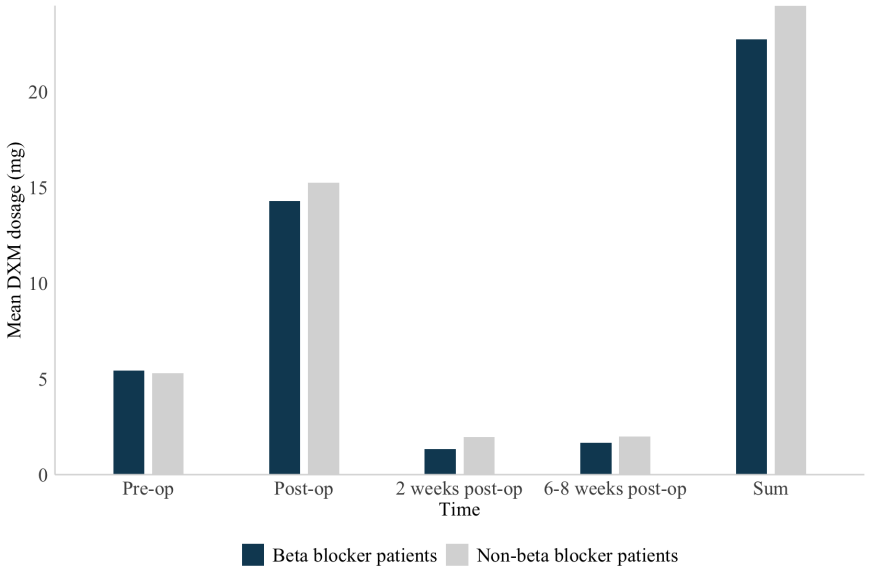

Figure 4: Mean DXM dosage for case and control group without controlling for confounders

As seen in Figure 4, there seems to be a lower mean DXM dosage in the beta blocker group before controlling for covariates. The difference in 2 weeks post-op was close to significantly lower for beta blocker patients (p-value of 0.067) when using the Mann-Whitney U test. See Table 9 for the p-values after performing Mann-Whitney's U-test to compare case and control group.

However, after controlling for age at surgery and extent of resection by propensity score matching and inverse probability weighting, the visual difference and the near-significance between the case and control group was lost, and no significant difference could be found for any stage of treatment.

There are several problems with the covariate control methods metioned above. For both methods it was chosen to focus on only two covariates, investigating the distribution effect of these two after the matching/weighting was done, and disregarding the potential effect this could have on the other covariates. Not only can this imply a disregard of potential biases, but it can also effect the possible generalization of the result. For instance, if the criteria for matching were to exclude individuals over 80 years old, the resulting data would no longer

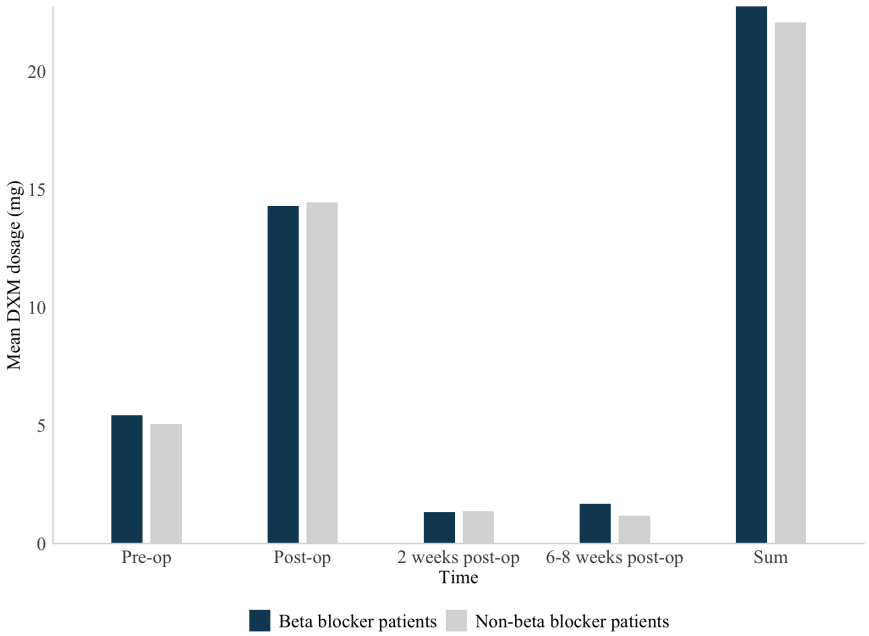

Figure 5: Mean DXM dosage for case and control group after propensity score matching

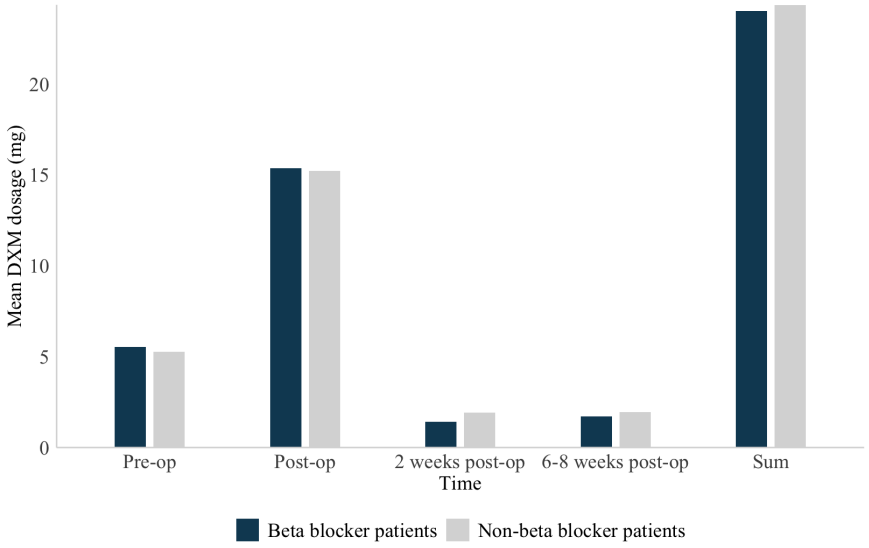

Figure 6: Mean DXM dosage for case and control group after inverse probability weighting

accurately represent those in that age bracket. The propensity score matching heavily reduced the sample size, decreasing the statistical power. Inverse probability weighting does not decrease sample size, but it does increase the variance in the over all patient set, which in turn also decreases statistical power. In order to understand whether significance was lost due to controlling for biases or decreased statistical power, more analysis had to be done.

Table 9: Mann-Whitney's U-test p-values for case and control group comparison by the following methods

DXM	Uncontrolled	Propensity score	Inv prob weight
Pre-op	0.995	0.980	0.985
Post-op	0.091	0.681	0.434
2w post-op	0.067	0.854	0.323
6-8w post-op	0.259	0.483	0.153
Total	0.252	0.582	0.844

4 Multiple linear regression

In this section, the relationship between the variables were explored using regression, building on the insights gained from the initial analysis. Benefits with regression compared to the previous methods used are that sample size is retained and interplay between all variables is taken into account. Regression analysis is a robust method for identifying how different variables effect a specific outcome - in this case how the chosen variables effect the total prescribed daily DXM dose (9). The total DXM dose takes all treatment stages into account and is therefore the most interesting outcome variable given the research question.

There are multiple ways to use regression to see if there is a difference between the total daily DXM dose prescribed for patients who take beta blockers compared to the ones who do not. The following two approaches were used:

1. "The statistical approach": A linear regression used for studying the full data set, with beta blocker usage as an included independent variable.
2. "The predictive approach": A linear regression trained for prediction of non-beta blocker patients, and then tested with beta blocker patients.

The two approaches will be discussed further in their respective section, but are both employing a log-linear model for analyzing the data, with the total DXM dose as the response variable. The first couple of steps in the regression are similar in both cases, and will therefore only be explained thoroughly in the first approach.

4.1 The statistical approach: A linear regression used for studying data

In this first approach of linear regression, the goal of the model was to study and analyze the relationship between the variables, rather than using it for prediction. With this being the main goal, and having limited amount of data, it was decided not to divide the data into training, validation, and test set, but rather to use the entire data set. To prevent the risk of overfitting and for overall model evaluation, statistical methods were used for model reduction. However, these methods do not address the model's performance on new data such as using a test set.

4.1.1 Addressing missing values, low frequencies and sample size

In section 2.2, missing values for beta blocker information as well as DXM dosage was addressed, which led to the removal of all patients with missing values for either of these. Patients with missing values for any of the other variables was also discussed, but was not removed as it was not necessary for the previously conducted analysis. With regression analysis however, the default in all programs is to eliminate any cases with missing data on any of the variables included. So, for this regression analysis part of the report, all patients with missing values for any variable were removed, resulting in a removal of 14 patients.

The previous frequency tables for the distribution of the control variables in the case and control group are after the removal of these 14 patients no longer representative. Instead, the updated version of this can be seen in Table 26 in the Appendix. Studying the frequency tables in Table 26 it is clear that some categories have very few samples, these have been summarized in Table 10 below.

Table 10: Summary of low frequency covariates low frequencies

Variable: category	Case	Control
EOR, Biopsy	0.97% (1)	3.44% (12)
MGMT, Weakly methylated	1.94% (2)	8.02% (28)
IDH1, Mutant	1.94% (2)	3.44% (12)
Tumor location, Infratentorial	0.97% (1)	0.86% (3)

In statistics, the "one in ten rule" is a rule of thumb that stipulates how many variables that can be estimated from a data set (22). The rule states that for every ten events, one variable can be included in the model. As of now there are two continuous variables and seven categorical variables to be included in the regression. For the categorical variables, dummy variables will be created as well as reference categories set. This will entail that every categorical variable with two categories will count as one "variable" (a need of ten events) and every categorical variable with three categories will count as two "variables" (a need of 20 events). The concept of dummy variables and reference categories will be explained further later, but what this implies is that the regression would include 11 variables, requiring a sample of at least 110 patients.

At this point the sample size was 452 patients, which satisfied the "one in ten rule". This is however only a guideline, and for investigating the effect of one specific variable there should preferably be a sufficient number of observations within each category of that variable. Table 10 shows that there are only four patients that have tumor location infratentorial, indicating that the result for this variable will be very unreliable, and that the exclusion of these patients are probably better than the inclusion of infratentorial. It was therefore decided to exclude these four patients, resulting in a total of 448 patients, out of which 102 patients are in the case group. The process of removing patients for the final data used for the initial regression model is illustrated in Figure 7.

It should be noted that as patients with tumor location infratentorial were removed, the data set only included patients with supratentorial tumors. The conclusion of this report would therefore be reduced to hold for only supratentorial glioblastomas.

Studying Table 10, the number of patients in the case group for the other

Figure 7: Process of removing patients to initial regression data

categories are also very small, although the total number of patients is at least larger than it was for tumor location infratentorial. With a trade-off between a decreased sample size and unreliable estimates, the decision was to still include the patients for the other categories. While the general guideline mentioned is to have at least ten observations per variable, this report specifically focuses on the effect of the beta blocker usage variable. To ensure enough variability withing each category to estimate their effect accurately, it would therefore be preferred to have at least ten in the case group for all of the variables as well. As this was not fulfilled for biopsy, weakly methylated tumors and mutant IDH1 genes, an extra eye are kept on these estimates.

The last thing to mention before moving along to the next step, 4.1.2 Model selection, is the fact that some of the patients total DXM dosage is zero. In total there are seven patients that were never prescribed any DXM, and therefore have a zero as their value for the total DXM dosage. For the model selection part of this report, two different approaches will be presented: a lin-linear and a log-linear approach. For the log-linear model, the patients with a total DXM value of 0 had to be handled, and it was decided to remove these patients. The process of removing patients for the final data set used for the log-linear regression model can be seen in Figure 8. The removal of these patients will be further discussed in the sensitivity analysis part of this report.

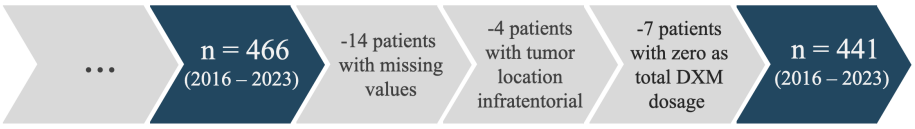

Figure 8: Process of removing patients to log-linear regression data

4.1.2 Model selection

As the modelling is between a continuous variable Y (total DXM), and several variables $x_1, x_2...$ (age at surgery, tumor size...), a multiple linear regression was chosen. The hypothesis is that Y has a linear relationship with $x_1, ..., x_p$ on average, and follows a linear model (9). The model for a generic observation i is

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon_i \qquad i = 1, .., n$$

where ϵ_i is the measurement error that contains all the random variation that cannot be explained by the model, and β_j is a partial regression coefficient reflecting the change in the dependent variable per unit change in the j-th

independent variable. The multiple linear model can also be expressed in matrix notation:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

or

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

where p denotes the number of independent variables, n the sample size, and it is here also assumed that $n > p$ (9). The vector $\boldsymbol{\beta}$ is a vector of unknown constants that are to be estimated from the data, resulting in the final model.

Besides the linearity, the following assumptions are also made for all observations i = 1,...,n (9):

$$\begin{aligned} E(\epsilon_i) &= 0 & E(Y_i) &= \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} = \mu_i \\ V(\epsilon_i) &= \sigma^2 & V(Y_i) &= \sigma^2 \\ \epsilon_i &\sim N(0, \sigma^2) & Y_i &\sim N(\mu_i, \sigma^2) \\ \epsilon_i &\text{ are pairwise independent} & Y_i &\text{ are pairwise independent} \end{aligned}$$

These assumptions imply that the conditional distribution of $Y_i|X = x_i$ is normally distributed, but this can not be known until the model is fitted. To get an initial idea of how well these assumptions holds and whether or not transformation of the data is needed for a better linearity, the first step was to explore how the total daily DXM dosage varied with one of the other variables, here age at surgery was chosen. Age at surgery was chosen as the first variable in the regression as it is one of the variables that showed the strongest correlation with prescribed DXM dosage in the initial analysis.

$$DXM = \beta_0 + \beta_1 * x_{\text{age at surgery}} \quad (1)$$

$$log(DXM) = \beta_0 + \beta_1 * x_{\text{age at surgery}} \quad (2)$$

The data was initially fitted to a lin-linear model, equation (1), and then a log-linear model, equation (2). The reason for trying the log-linear model as well as the lin-linear model is that transformation of data can sometimes help with previous presented assumptions on linear regression. In Figure 9, there are four different plots, two for each model. In the first two plots, the residuals for the two models are plotted against the predicted DXM values, and in the two lower plots, the residuals are plotted against age at surgery. These plots are for visually inspecting if the models lack linearity.

The residuals should preferably be randomly scattered around zero with no trends, and have roughly constant variance. Studying the plots in Figure 9, the

log-linear model seems like the slightly better choice out of the two as an initial approach.

In Figure 10 below, the residuals for the two models are plotted in a QQ-plot as well as a histogram of the residuals. These plots are for visually inspecting the normality of the measurement error. The residuals should preferably lie on a straight line in the QQ-plot, and by visual inspection it was not completely clear which model that was the best choice. However, given the previous inspection of linearity, the log-linear model was chosen to proceed with.

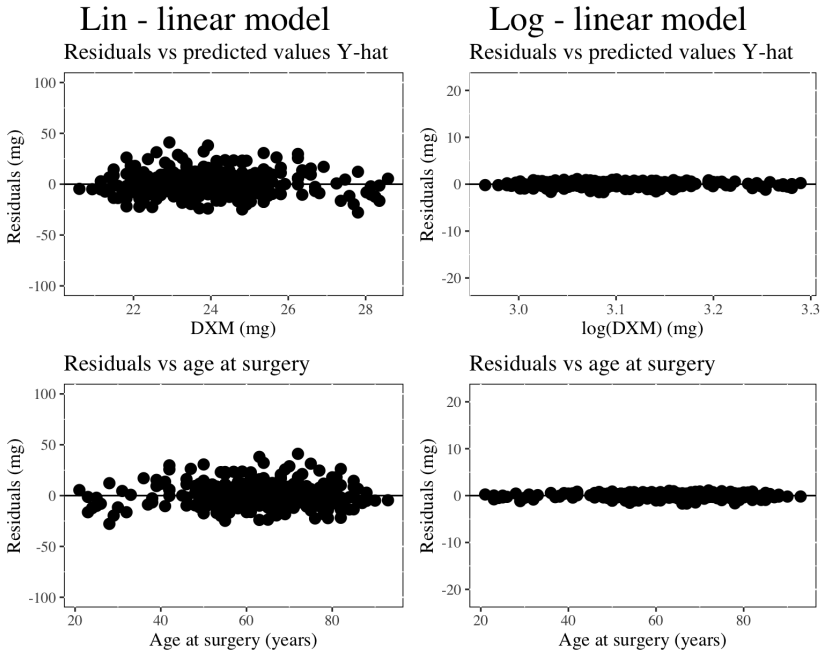

Figure 9: Residuals plotted against age at surgery and predicted DXM

4.1.3 Some background theory

In multiple linear regression, the vector $\boldsymbol{\beta}$ represents unknown parameters which are to be estimated from the available data. If there were no random error present in the dependent variable Y_i, any set of $p+1$ data points could straightforwardly yield explicit values for the $\boldsymbol{\beta}$ vector (β_0, β_1, ..., β_p). However, due to the inherent random variation in $\mathbf{Y}$, a measurement error vector $\boldsymbol{\epsilon}$ emerges. Consequently, there are no direct solutions for calculating the β-parameters. Instead, a methodology is required to combine the information from each data point, ultimately deriving a single solution that optimally fits the model.

The least squares estimation process relies on the principle that the solution should minimize the sum of squared deviations between the observed dependent variable Y_i and their estimated means, provided by the model (9). In other words, it selects the parameter vector $\hat{\boldsymbol{\beta}}$ that minimizes the sum of squared measurement errors:

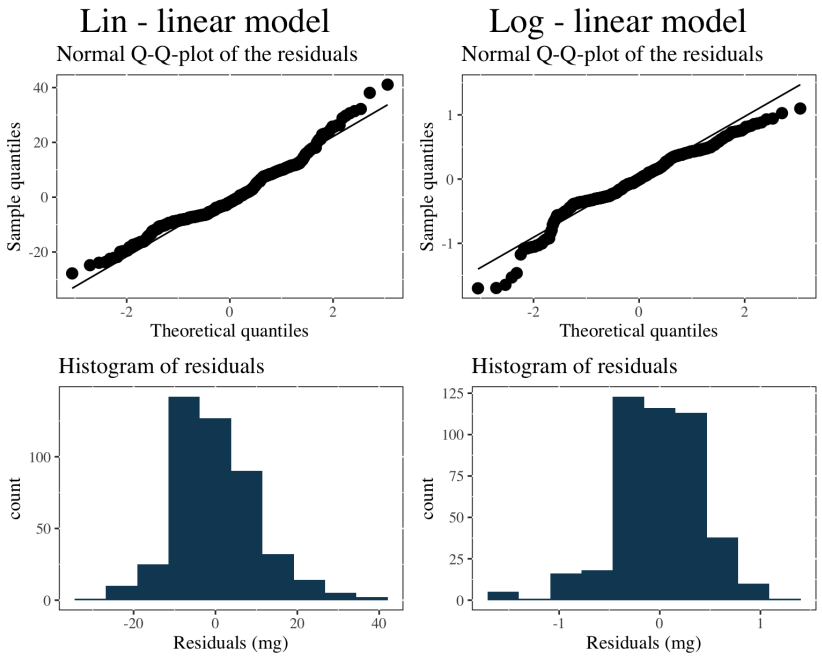

Figure 10: QQ-plot and histogram of residuals

$$\begin{aligned} SS(\text{Measurement errors}) &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^{n} e_i^2. \end{aligned}$$

For optimization, the derivatives of SS(Measurement errors) with respect to $\hat{\boldsymbol{\beta}}$ are set equal to zero, resulting in the so called normal equations. For the multiple linear regression model, the normal equations are expressed in matrix notation as

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{Y}.$$

If $\mathbf{X}^T\mathbf{X}$ has an inverse, then the normal equations have a unique solution given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}).$$

For $\mathbf{X}^T\mathbf{X}$ to have an inverse, the matrix $\mathbf{X}$ has to have full column rank, that is, there can be no linear dependencies among the independent variables (9). When the columns of $\mathbf{X}$ however are linearly dependent, the matrix $\mathbf{X}^T\mathbf{X}$ is singular, $(\mathbf{X}^T\mathbf{X})^{-1}$ has no unique solution, and $\hat{\boldsymbol{\beta}}$ can not be uniquely estimated. With this theory in mind, the addition of the background variables will now be presented.

4.1.4 Adding the background variables

After determining the preferred regression model to proceed with, the next step involved integrating the remaining control variables outlined in section 2.4. Table 1 provides an overview of whether these control variables are continuous or categorical (ordinal or nominal). Ordinal data, such as gender, lacks inherent order among categories, posing challenges in assigning numerical "labels". Conversely, nominal data exhibits some form of order, but assigning numerical "labels" does not necessarily allow for valid mathematical operations (9). For example, consider attempting to assign numerical labels to the MGMT status. Let unmethylated be 1, weakly methylated be 2, and methylated be 3. Then unmethylated + weakly methylated = 3 = methylated, which is illogical.

It is clear that categorical variables can not simply be added to the regression model without careful consideration. Consequently, it was decided to use dummy variables. For each categorical variable, as many new variables were created as there are categories for the variable. For example, the MGMT status variable were replaced with the following three variables:

$$x_{unmethylated} = \begin{cases} 1 & \text{if } x_{MGMT} = \text{unmethylated}, \\ 0 & \text{otherwise} \end{cases}$$

$$x_{weakly\ methylated} = \begin{cases} 1 & \text{if } x_{MGMT} = \text{weakly methylated}, \\ 0 & \text{otherwise} \end{cases}$$

$$x_{methylated} = \begin{cases} 1 & \text{if } x_{MGMT} = \text{methylated}, \\ 0 & \text{otherwise} \end{cases}$$

Using these dummy variables, the $\mathbf{X}$ matrix in the linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

can then be expressed as

$$\mathbf{X} = \begin{bmatrix} 1 & X_{1,\text{unmethylated}} & X_{1,weakly\ methylated} & X_{1,methylated} & \cdots & X_{1p} \\ 1 & X_{2,\text{unmethylated}} & X_{2,weakly\ methylated} & X_{2,methylated} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 1 & X_{n,\text{unmethylated}} & X_{n,weakly\ methylated} & X_{n,methylated} & \cdots & X_{np} \end{bmatrix}$$

$$= [e.g.] = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & X_{1p} \\ 1 & 1 & 0 & 0 & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 1 & 0 & 0 & 1 & \cdots & X_{np} \end{bmatrix}.$$

The second matrix shows an example of how the matrix could look like. It illustrate that, for any row in $\mathbf{X}$, $x_{i,unmethylated} + x_{i,weakly\ methylated} + x_{i,methylated} = 1$, which is the value of the first column in the matrix. That is, one of the columns can be derived as a linear combination of some of the other columns, the matrix does not have full column rank. As mentioned in the previous section, $\boldsymbol{X^T X}$ will then be singular, the inverse has no unique solution, and $\hat{\boldsymbol{\beta}}$ can not be uniquely estimated. A solution is needed.

There are two options to get rid of the linear dependency introduced, and return to an $\mathbf{X}$ with full rank (9):

1. Remove the intercept
2. Remove one of the categories, resulting in one less dummy variable

It was decided to pursue option 2). This approach makes the removed category, for example unmethylated, the so-called reference category. The intercept will then be the expected response for this reference category, and the beta coefficient for the remaining categories shows the systematic effect in relation to the reference category. It is important to remember this for the future interpretation of the beta coefficient, as they have to be interpreted in relation to the reference category.

Now for the next decision, what category to choose as the reference category? If the number of observations in the reference category is small, all $\hat{\beta}$ estimates will be uncertain, and it is therefore best practice to chose the category with the largest sample as the reference category (9). Table 27 in the Appendix includes all the frequency tables for the categorical control variables, and from this the reference category for each of the categorical variables was selected. The result is presented in Table 11 below.

Table 11: The selected reference categories and dummy variables to be included in the regression for each categorical variable

Variable	Reference category	Included dummy variables
EOR	Gross-total resection	x_{biopsy}, $x_{sub\text{-}total\ resection}$
MGMT	Unmethylated	$x_{weakly\ methylated}$, $x_{methylated}$
Sex	Male	x_{female}
Stupp	No	x_{yes}
IDH1	Wild type	x_{mutant}
Beta blocker	No	x_{yes}

Now that the categorical variables were handled, the next step involved exploring the interplay among all of the independent variables. As mentioned it is a problem if the matrix $\mathbf{X}$ is singular, and the same goes for if it is nearly singular (9). If a linear combination of some of the x variables almost equals to one of the other x variables, then $\mathbf{X}$ is nearly singular, $(\boldsymbol{X^T X})^{-1}$ has an unstable solution and $\hat{\beta}$ will have a huge variance. It is therefore important to check that the x variables do not "compete", meaning that there is no collinearity between them. The inclusion of both variables are not only unnecessary, but could be potentially harmful, and a common solution is to only include one of the two.

Collinearity was assessed by visual inspection, variance inflation factor and correlation matrices. For the two continuous variables, age at surgery and tumor size, the scatter plot can be seen in Figure 25 in the Appendix. It looks fairly well scattered and there is no indication of a linear relationship. When studying the scatter plot there is however a potential influential observation, the patient assigned study ID 331. It was not removed, but it will be discussed in the sensitivity analysis. Table 28, also in the Appendix, shows the correlation between the variables (both continuous and dummy) to be included in the regression. The most noticeable correlation emerges between age at surgery and the presence of IDH1 mutation, with a coefficient of approximately 0.304. As illustrated

in Figure 26 in the Appendix, the boxplot comparing IDH1 mutation status with age at surgery reveals that patients with an IDH1 mutation tend to be younger compared to those without. The correlation is however not exceptionally high, and does not suggest an exclusion of either variable. Hence, it was deemed appropriate to include all variables in the study.

To ensure that the inclusion of all variables was indeed appropriate, the variance inflation factor (VIF) was also calculated for each variable. The relationship between VIF and collinearity is based on the idea that collinearity among predictors inflates the variance of parameters estimates (9). The formula for VIF is

$$VIF_j = \frac{1}{1 - R_j^2},$$

where R_j^2 is the coefficient of determination for the regression model for the j-th predictor (X_j) on the other independent variables. If there where to be a near singularity involving the independent variable (X_j) and the other independent variables, R_j^2's value will be close to one, and the VIF would be large. A common guideline is that a VIF above ten is problematic (9). The VIF for all variables can be seen in Table 12, where all values are close to one, which is the expected result if all variables are orthogonal (9). This supports the decision to not exclude any variables at this stage.

Table 12: Variance inflation factor for all regression coefficients to be included

Variables	VIF value
Beta blocker, yes	1.120
Age at surgery	1.269
Tumor size	1.078
Sex, female	1.025
IDH1, mutant	1.152
Stupp complete, yes	1.186
MGMT, methylated	1.181
MGMT, weakly methylated	1.097
EOR, biopsy	1.034
EOR, sub-total resection	1.057

When adding the background variables, potential interaction terms were not explored. If, e.g., change in tumor size (x_{i1}) were to depend on change in age (x_{i2}), then the modeling would have benefited from a so called interaction term ($\beta_{12}x_{i1}x_{i2}$). It would have allowed the independent variables to influence the impact of another. Such relationships were as mentioned however not explored, something that could potentially have made the model better. (9)

4.1.5 The full model

When the variables were added, the full model was fitted as a log-linear regression with total DXM as the response variable. As seen in Table 13, several variables' β-estimates were not shown to be significantly different from zero at a 0.05 level of significance.

Table 13: Full model coefficients with confidence intervals, **significant variables**

Variables	β-value	Lower (2.5%)	Upper (97.5%)
Intercept	**3.139**	2.862	3.415
Beta blocker, yes	-0.014	-0.116	0.087
Age at surgery	**-0.004**	-0.008	-0.001
Tumor size	**0.049**	0.023	0.074
Sex, female	0.009	-0.073	0.090
IDH1, mutant	**-0.280**	-0.525	-0.034
Stupp, yes	-0.019	-0.121	0.083
MGMT, methylated	-0.002	-0.090	0.086
MGMT, weakly methylated	0.104	-0.063	0.271
EOR, biopsy	0.089	-0.162	0.340
EOR, sub-total resection	**0.099**	0.015	0.182

With the research question in mind, no significant contribution of the beta blocker could be proven. However, the full model contained several potentially insignificant variables, and therefore it would need to be optimized before any firm conclusions could be drawn.

4.1.6 Model reduction

To mitigate the risk of insignificant variables affecting the significance of other variables, the model was reduced step-by-step. The following approach was taken to shrink the model without losing valuable information:

1. Test if the full model actually is better than a model with just an intercept (Global F-test)
2. Remove an insignificant variable (t-test) from the model
3. See if the reduced model performed as well as the full model (partial F-test)
4. Repeat steps 2-3, stop when reduced model could be proved to be worse than the full model (partial F-test)
5. Compare information in all model versions with different measurements (AIC, BIC, R^2)

The hypothesis and logic behind the different tests are presented below.

4.1.7 Global F-test

Some models do not manage to capture any structure in the underlying data that actually makes them better than the assumption of a simple intercept. The global F-test tests, for a model with p independent variables and n samples (9):

$$\begin{aligned} H_0: &\quad \beta_k = 0 \quad \text{for all k = 1,2,...,p against} \\ H_1: &\quad \beta_k \neq 0 \quad \text{for at least one } k = 1, 2, ..., p \end{aligned}$$

If H_0 is true, then none of the x-variables are needed and the signal is just noise. Under the assumption that the noise is normally distributed and H_0 true, then

$$F = \frac{MS(Regression)}{MS(Error)} \approx 1 \text{ and} \sim F(p, n-(p+1))$$

with MS being the expected mean squared. Should the model be a good fit, the error of the regression is expected to be similar to the error that cannot be modeled (9).

The global F-test was hence used to determine whether *any* of the included x-variables were significant. Should they not be, there would not be a reason to try to reduce the model either. The full model had an F-value of 3.48 on 10 and 430 degrees of freedom, yielding a p-value of 0.0002. Hence, it was proven to be better than the intercept model at a 0.05 level.

4.1.8 Partial F-test

Partial F-tests can be used for nested models to deem whether a reduced model is as good at explaining the data compared to one with more explanatory variables (9). In this context, the partial F-test was used to see if information was lost from the full model when trying to make it more simple by removing one variable at a time.

$$\begin{aligned} H_0: &\quad \beta_k = 0 \quad \text{for some specific k against} \\ H_1: &\quad \beta_k \neq 0 \quad \text{for all k = 1,2,...,p} \end{aligned}$$

The standard error s^2 is calculated for the full and nested model. Then, the increase in error, Q, has k degrees of freedom. H_0 is rejected on a 0.05 significance level if:

$$F = \frac{Q/k}{s^2_{full}} > F_{0.05,k,n-(p+1)}$$

Just like the global F-test, the partial F-test builds upon the assumption of normally distributed errors (9). The nested model has to have the same data set as the full model, hence the nested models were created based on the full model dataset.

4.1.9 t-test

The final variable selection test used was a t-test. The t-test is used to test if a given β_k is significantly different from zero, given all the other variables in

the model (9). This means that a certain β could show to be insignificant, but if the model would change (e.g., a highly correlated x-variable is also removed) significance could be gained.

$$t = \frac{\tilde{\beta}_j - 0}{d(\tilde{\beta}_j)} \sim t(n - (p + 1))$$

The distribution of the t-value also builds upon normally distributed errors. The null hypothesis is rejected if

$$t > t_{\alpha/2, n-(p+1)}$$

for a two-tailed test with $n - (p + 1)$ degrees of freedom.

4.1.10 R-squared and adjusted R-squared

When choosing what variables to include, there were multiple models created. To compare them with each other the calculated R^2, AIC, and BIC values were used as guidance.

R^2 looks at how much of the variability in DXM that can be explained by the model in question (9). R^2 is computed from the summed squares of residuals $SS(Res)$ and the total uncorrected sum of squares $SS(Total_{corr})$.

$$R^2 = \frac{SS(Regr)}{SS(Total_{corr})} = 1 - \frac{SS(Error)}{SS(Total_{corr})}$$

However, R^2 will favor large models as more parameters often succeed in explaining variance (9). To control for that, it was complemented with the adjusted R^2, given by:

$$R^2 = 1 - \frac{MS(Error)}{MS(Total_{corr})}.$$

4.1.11 Akaike information criterion (AIC)

For a model with $p+1$ parameters, $\hat{\beta}$ being the maximum likelihood estimates, the Akaike information criterion is given as (9):

$$AIC(p + 1) = 2(p + 1) - 2lnL(\hat{\beta}; Y)$$

The best model is the one with the smallest AIC, i.e. the one that manages to maximize the likelihood function without too high parameter cost (9).

4.1.12 Schwarz Bayesian Criterion (BIC)

Schwarz Bayesian Criterion works in a similar way to AIC that penalizes larger models more (9).

$$BIC(p+1) = ln(n)(p+1) - 2lnL(\hat{\beta}; Y)$$

In general, AIC is used to identify the models best for prediction whilst BIC identifies the model that best explains the data (9). Only variables that have a significant contribution should be in the model. Hence, BIC was mainly looked at in this case.

4.1.13 Results

Stepwise exclusion from the full model, Table 13, yielded the results seen in Table 14. All the reduced models passed the partial F-test, which meant that it could not be proven that the removed variables' contribution were significantly different from zero. After MGMT, Sex, Stupp and Beta blocker status was removed there were no other variables that had a t-test p-value below 0.05. Further removal was still tested yet not successful, and thus not shown here.

Table 14: Reduced models and test results, p-value for partial F-test

Excluded variables	R^2	R^2 adj	BIC	AIC	$F_{partial}$
Full model	**0.075**	0.053	566	517	N/A
MGMT	0.071	0.054	556	515	0.440
MGMT, Sex	0.072	0.056	550	513	0.641
MGMT, Sex, Stupp complete	0.071	0.058	544	511	0.769
MGMT, Sex, Stupp complete, Beta blocker	0.071	**0.060**	**538**	**509**	0.848

That beta blocker showed to be insignificant was not too surprising given the results in the initial analysis. It further strengthens the view that beta blocker use may not have an effect on total DXM dosage. As infratentorial patients were removed, this result only holds for supratentorial tumors.

Which reduced version to chose was fairly straightforward in this case, given that adjusted R^2, BIC and AIC indicated the same model. That R^2 favored the full model makes sense as the more variables, the more explanatory power. One could choose to rely solely on the tests to decide the final model, or one could choose the smallest possible model that is still not significantly different from the full model. In this case it was the same model, see Table 15 for the final model. The dummy variable for biopsy in extent of resection was not significant; however potential significance would be hard to find given the small sample size.

4.2 Predictive modeling with control group as baseline

The second approach used was to create a predictive model based on non-beta blocker patients and then try it for beta blocker patients. The plan was to do the following:

Table 15: Final model in relation to log(DXM)

Variables	β-value	Lower (2.5%)	Upper (97.5%)
Intercept	3.141	2.882	3.400
Age at surgery	-0.004	-0.008	-0.001
Tumor size	0.050	0.025	0.075
IDH1, mutant	-0.284	-0.526	-0.042
EOR, biopsy	-0.092	-0.156	0.341
EOR, sub-total resection	0.098	0.016	0.181

1. Create a model for DXM dosage prediction for non-beta blocker patients
2. Test model performance on a test set of non-beta blocker patients
3. Test model performance on beta blocker patients

Should the model give a consistently too large prediction on beta blocker patients compared to the test set, the conclusion would support the hypothesis of the research.

4.2.1 Dividing the data

For simplicity, it was decided to use the data that had removed the patients with a total DXM of zero regardless of a lin-linear or log-linear model fitting the best. See Figure 8 in section 4.1.1 for the process of getting to the final data used for this regression.

The model was supposed to be based solely on non-beta blocker patients and thus the sample size was from the beginning n = 341. The data set was then randomly shuffled, and 70 patients were removed to be saved for the test set (comprising roughly 20% of the total sample). For reproducibility, a seed was set, and the same seed was used throughout the report when needed. The new categorical variable distributions for the training set are presented in Table 16.

The "one in ten rule" mentioned earlier remains fulfilled. Since the regression now focuses solely on patients not using beta blockers, the independent variable for beta blocker usage is eliminated, resulting in a regression with only 9 variables. Following the "one in ten rule", this necessitates a sample size of at least 90 events, which is met as the training set comprises 271 patients. It should be noted that the sample size for biopsy patients and IDH1 mutant patients now is even smaller than before. The small sample size of these categories was not taken into consideration when creating the model, but the effects of this will be further discussed in the sensitivity analysis.

4.2.2 Model selection

Both a lin-linear and log-linear model were explored following the same methodology as in the previous section 4.1.2. Residual plots for the two approaches to predictive regression modelling are presented in the in Figure 27 and 28 in the Appendix. Through visual inspection, it was determined that the log-linear model was a slightly better fit for the predictive regression model as well.

Table 16: Frequency tables for non-beta blocker train data set (n=271)

Extent of resection	Control
Biopsy	2.58% (7)
Sub-total resection	38.75 (105)
Gross-total resection	58.67 (159)

MGMT	Control
Unmethylated	46.49% (126)
Weakly methylated	7.38% (20)
Methylated	46.13% (125)

Sex	Control
Female	49.45% (134)
Male	50.55% (137)

Stupp	Control
Yes	25.09% (68)
No	74.91% (203)

IDH1	Control
Wild-type	96.68% (262)
Mutant	3.32% (9)

4.2.3 Adding the background variables

The process of incorporating the independent variables to the regression followed the same approach as in the previous regression model. Dummy variables were created for all categorical variables, and reference categories were set. Examining the frequency tables for the categorical variables in Table 16, the reference categories were set the same way as in the previous regression model. Table 17 illustrates the result of what dummy variables that were included.

Table 17: The selected reference categories and dummy variables to be included in the predictive regression model for each categorical variable

Variable	Reference category	Included dummy variables
EOR	Gross-total resection	x_{biopsy}, $x_{sub\text{-}total\ resection}$
MGMT	Unmethylated	$x_{weakly\ methylated}$, $x_{methylated}$
Sex	Male	x_{female}
Stupp complete	No	x_{yes}
IDH1	Wild type	x_{mutant}

The collinearity was then assessed using the same approach as for the previous regression model: visual inspection, variance inflation factor, and correlation matrices. Similar as before, the correlation between the variables were quite low, with the most noticeable correlation emerging between age at surgery and the presence of IDH1 mutation, with a coefficient of approximately 0.304. This correlation is still not very strong, and does therefore not suggest the exclusion of either variable. The variance inflation factor for all variables can be seen in Table 18. All values are close to one, which further supports the decision not to exclude any of the variables.

4.2.4 The full predictive model

With the collinearity assessment completed, the full model was fitted as a log-linear regression with the total DXM dosage serving as the response variable.

Table 18: Variance inflation factor for all regression coefficients to be included in the predictive regression model

Variables	VIF value
Age at surgery	1.200
Tumor size	1.134
Sex, female	1.023
IDH1, mutant	1.190
Stupp complete, yes	1.222
MGMT, methylated	1.259
MGMT, weakly methylated	1.106
EOR, biopsy	1.041
EOR, sub-total resection	1.058

The estimated β-coefficients values for the full model, as well as their confidence intervals, can be seen in Table 19. Same as for the previous model, several variables' $\hat{\beta}$-coefficients were not significantly different from zero, and it was decided to further optimize the model.

Table 19: The full predictive model coefficients with confidence intervals, **significant variables**

Variables	$\hat{\beta}$-value	Lower (2.5%)	Upper (97.5%)
Intercept	**3.113**	2.753	3.473
Age at surgery	-0.004	-0.009	0.001
Tumor size	**0.046**	0.011	0.080
Sex, female	-0.023	-0.131	0.085
IDH1, mutant	-0.318	-0.644	0.008
Stupp complete, yes	-0.061	-0.197	0.076
MGMT, methylated	0.057	-0.064	0.177
MGMT, weakly methylated	0.142	-0.073	0.358
EOR, biopsy	0.096	-0.248	0.441
EOR, sub-total resection	0.081	-0.032	0.194

The model reduction was performed with a similar methodology as in the first regression. However, given the reduced sample size, it was harder to find significant variables. Therefore, the F-tests, AIC, BIC and R^2 were complemented with bootstrapping of the betas to see if some of the insignificant variables were skewed to any side of zero. As the purpose of this regression now was to predict, it was important that the model generalized well to unseen data. Cross-validation was therefore also used as a complement to previous mentioned tests.

4.2.5 Bootstrapping

The sample size for constructing the predictive regression model comprises 271 patients, only a fraction of the real population of people not on beta blockers undergoing surgery for glioblastoma. The process of drawing conclusions about an underlying population based on only a sample is called statistical inference (23). It carries significant weight, as it is the determination of whether an observed effect is genuine and can be generalized for an entire population. Similar as statistical inference, model validation has the goal of generalizing information from a limited sample (24). In model validation, the generalizing refers

to determining whether the model accurately represents the behavior of the system.

In section 4.2.4, the full predictive model was fitted, resulting in several beta coefficient that were not significantly different from zero. But how certain are actually these basic statistical estimates? The estimates of these variables are only based on a small sample of 271 patients, and it is debatable how well it than can be generalized to a whole population. To asses model validation and the uncertainty of these parameter estimates, a method called bootstrapping was used.

The bootstrap is a data-based simulation method for assigning measures of accuracy to statistical estimates, invented by Bradley Efron in 1979 (25). The basic idea with bootstrapping is that the statistical inference about a population from a sample can be modeled by resampling the sample. A form of estimated population is then created by re-performing the statistical test in question on several resamples. Figure 11 illustrates the idea.

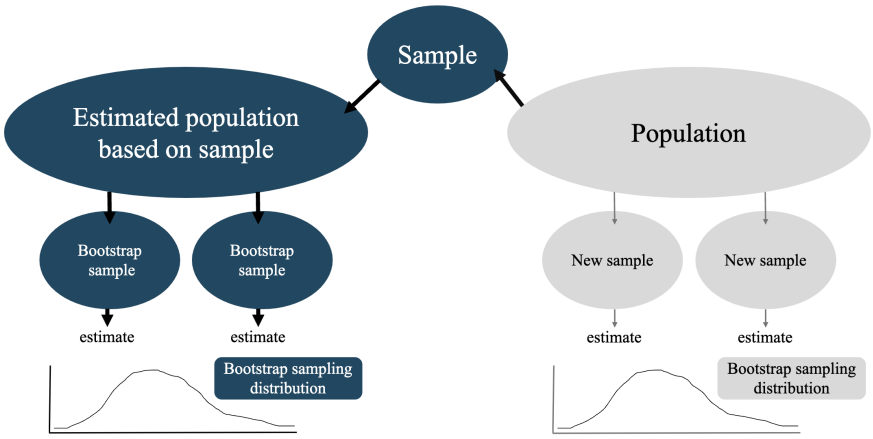

Figure 11: The idea behind bootstrapping

For assessing the parameter estimates of the model, 500 independent bootstrap samples were generated from the original sample of 271 patients. Each bootstrap sample has as many patients as the original sample (271 patients), and are generated by sampling with replacement. The regression model is then fitted to each of the bootstrapping samples, and parameter estimates are calculated for each model. This results in a total of 500 different estimates of each beta coefficient. In Figure 12 the procedure is illustrated using a sample of five patients and n number of bootstrapping samples.

The result of the bootstrapping can be seen in Figure 13, where the bootstrap distribution for each of the variables included in the predictive regression model is illustrated. These distributions gives an idea of the uncertainty associated with the parameter estimates previously presented, and will be used as a guideline in the next section where the model is to be adjusted.

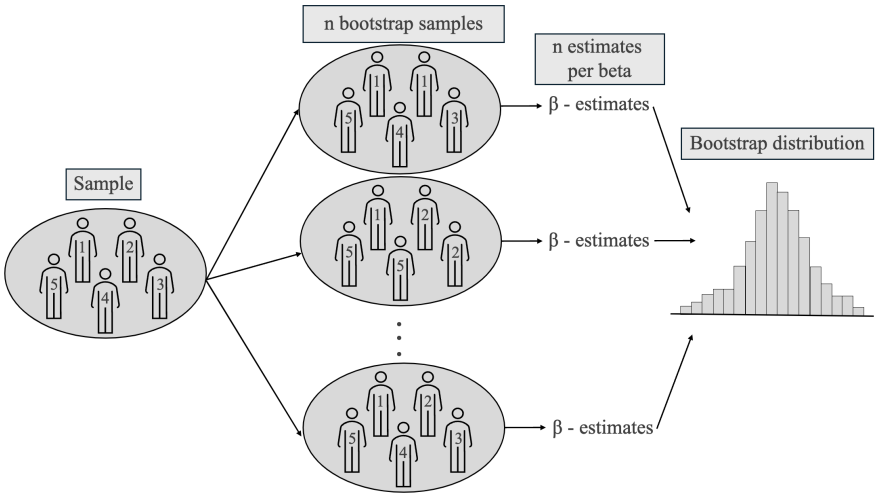

Figure 12: The process of bootstrapping

4.2.6 Cross-validation for tuning of parameters

When dividing the data into training and test set, it was decided not to include a validation set. This decision was made due to the lack of data, wanting as much as possible for the training of the model. Instead, the technique cross-validation was employed. Besides for the fact that cross-validation utilizes the available data better than a fixed validation set, it also avoids over-fitting the hyperparameters to a fixed validation set, which when the sample is small can be a distorted representation (26).

Similar as for the previous model reduction, R^2 and adjusted R^2 was again used for model selection. While these two do give an estimation of how well the model's goodness of fit is, it does not actually give any information on how the model generalizes to unseen data. Therefore, in addition to the regular R^2 for the model, an average R^2 value was calculated based on the result of 5-fold cross validation.

The k-fold cross validation technique divides the data into k subsets of approximately the same size (26). The subsets are created by random sampling units from the training set without replacement. The model is then trained on k-1 of the subsets, and then applied on the one subset not used in training the model. The measures are then calculated on how well the model generalizes to this unseen data, and the procedure is repeated k times until the model is tested on all subsets. The average of the k performance measures is then the performance measure for the cross validation. The process of acquiring the average R^2 value from the 5-fold cross validation is illustrated in Figure 14.

4.2.7 Model reduction

The model reduction was performed in the same way as in the first regression, in combination with the bootstrapping distribution and the 5-fold cross validation R^2 estimate. Bootstrapping had the effect of keeping some of the variables that looked like they could have a value different from zero, although zero was covered by the confidence intervals.

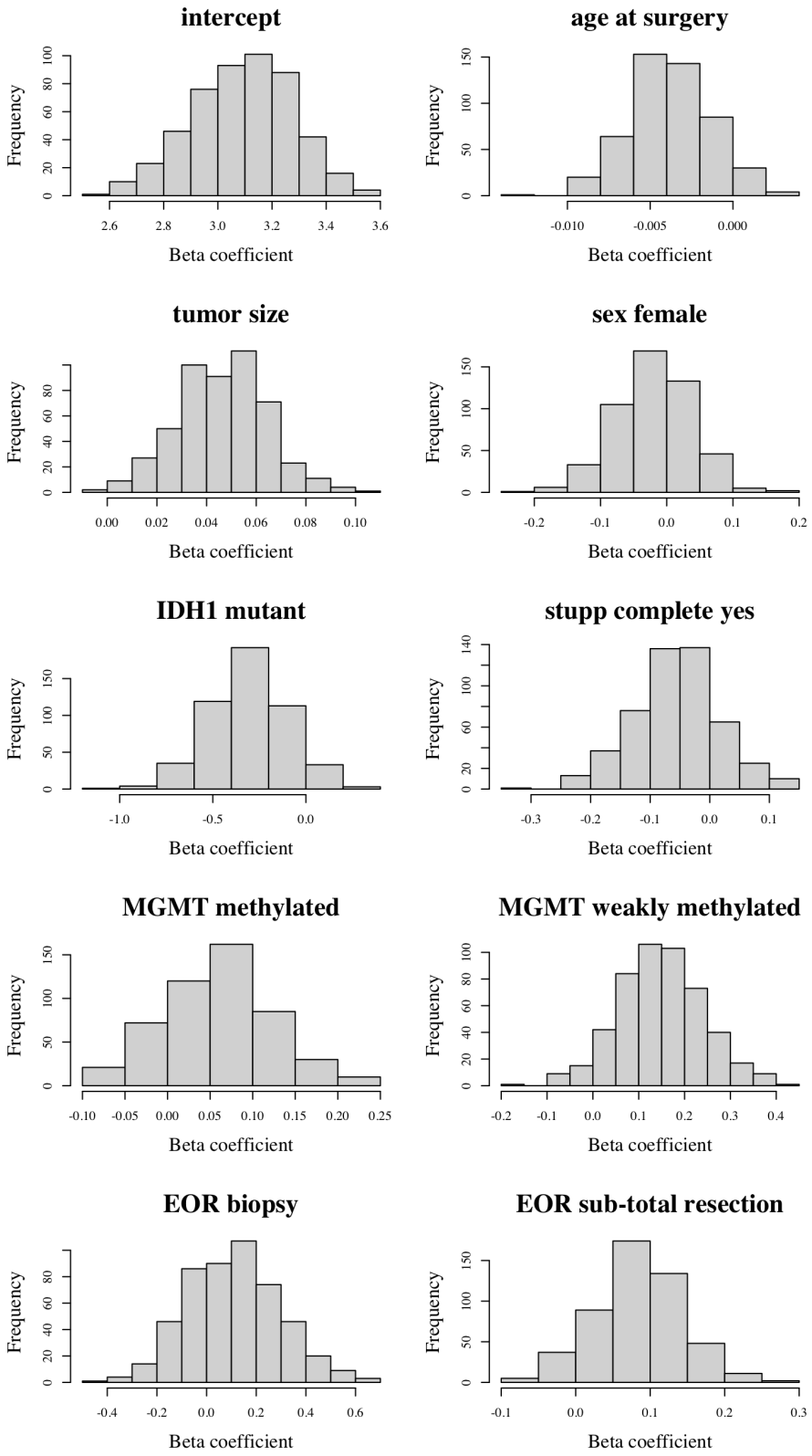

Figure 13: The result from bootstrapping on parameter estimates

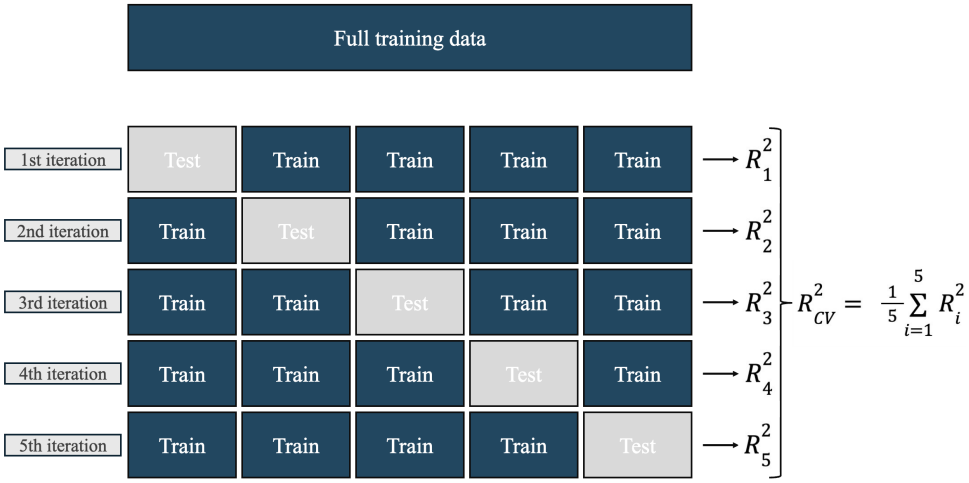

Figure 14: The process of acquiring the average R^2 from 5-fold cross validation

The test results were more ambiguous this time around, as seen in Table 20. AIC and BIC were best for the model excluding sex, stupp complete, MGMT and EOR, and the partial F-test also deemed it as good as the full model. R^2 showed the highest value for the full model, adjusted R^2 for the one excluding Sex, stupp complete and MGMT, and R^2 from cross validation showed the highest value when excluding sex, stupp complete, MGMT and age at surgery. It was decided to stop the reduction at the model favored by adjusted R^2. This model had overall good values for all measurement, but the main reason being that there are medical knowledge supporting the inclusion of age at surgery, IDH1, and EOR. Looking back at the bootstrapping distribution of these three covariates, Figure 13, they were also tilted slightly away from zero, suggesting they might have a significance should more data be available.

Table 20: Reduced non-beta blocker models and test results, p-value for partial F-test. Stupp refers to stupp complete

Excluded variables	R^2	R^2_{adj}	R^2_{CV}	BIC	AIC	F_{part}
Full model	**0.061**	0.028	0.020	385	345	N/A
Sex	0.060	0.032	0.021	379	343	0.649
Sex, Stupp	0.058	0.033	0.035	374	342	0.648
Sex, Stupp, MGMT	0.052	**0.034**	0.031	365	340	0.660
Sex, Stupp, MGMT, Age	0.044	0.030	**0.044**	361	340	0.486
Sex, Stupp, MGMT, IDH1	0.040	0.025	0.026	363	341	0.323
Sex, Stupp, MGMT, EOR	0.043	0.033	0.042	**356**	**338**	0.561

The final model for non-beta blocker patients thus became the one including age at surgery, tumor size, IDH1, and EOR. The coefficient values and their confidence intervals for the model is presented in Table 21.

4.2.8 Testing on test set and beta blocker patients

To finally test if beta blocker patients got a lower prescribed total DXM dosage compared to non-beta blocker patients, the final model was tested on the test

Table 21: Final non-beta blocker model

Variables	β-value	Lower (2.5%)	Upper (97.5%)
Intercept	**3.066**	2.727	3.401
Age at surgery	-0.003	-0.008	0.001
Tumor size	**0.050**	0.014	0.082
IDH1, mutant	-0.300	-0.618	0.017
EOR, biopsy	0.071	-0.270	0.412
EOR, sub-total resection	0.088	-0.023	0.198

set as well as a random sample of beta blocker patients. Only 70 of the 104 beta blocker patients were randomly chosen and tested for comparability in sample size.

The procedure was:

1. Predict total (log) dose of DXM based on the independent variables in the test and beta blocker data
2. Calculate the residuals as the difference between true (log) values and predicted values
3. Analyze these to see if there is a difference in performance on the test set and the beta blocker patients

The residual analysis conducted involved computing mean squared error (MSE) or mean absolute error (MAE), as well as visual inspection of the residuals to see if residuals would be skewed in any direction. For n samples with predicted values $\hat{y}_i^2$ and true values y_i they are computed as (9):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

The MSE is more sensitive towards outliers, as the residuals are squared. A larger difference becomes even more influential.

At first glance, the model seemed to perform slightly better on the non-beta blocker patients in the test set, with a lower mean absolute error (MAE) and mean square error (MSE) as seen in Table 22. However, when the residuals were plotted against the predicted values, a couple of outliers were spotted for the beta blocker patients. The fact that the mean absolute error was more equal for the two sets compared to the mean squared error indicates that these outliers had a contribution.

Studying the residual plots in Figure 15 and 16, the residuals looks nice and randomly scattered. There were five outliers for the beta blocker patients that could have contributed to the slightly "worse" mean errors, as seen in Figure 16.

The five samples with residuals below -0.5 were removed, yielding a new plot seen in Figure 17. With these samples removed, there was no longer an indication

Table 22: Residual analysis for predictions on test set and beta blocker patients

Evaluation	Test set	Beta blocker set
MAE	0.306	0.314
MSE	0.135	0.167

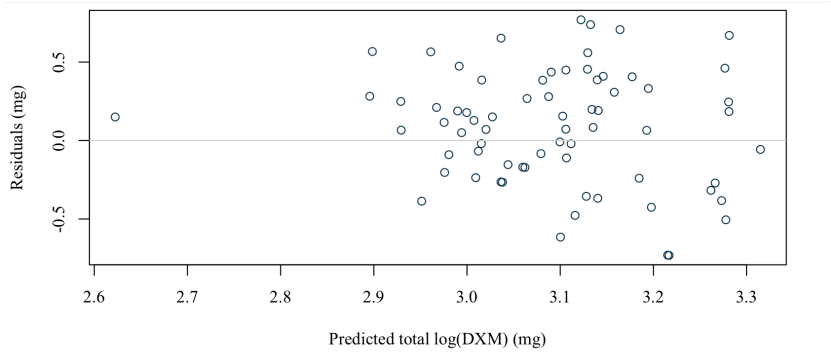

Figure 15: Residuals for (n = 70) test data prediction against log predicted values

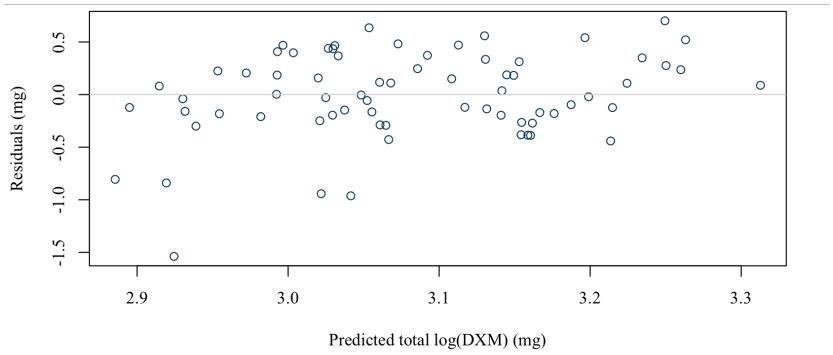

Figure 16: Residuals for (n = 70) beta blocker patient prediction against log predicted values

that the predictive model performed worse on beta blocker patients. The new error measurements, seen in Table 23, rather showed that it gave slightly better predictions for beta blocker patients compared to the test data.

Having conducted both the "statistical" and "predictive" regression approach, none yielding any significant results when determining the beta blocker significance, it was clear that, based on this data, no contribution from beta blockers on prescribed DXM dosage could be found.

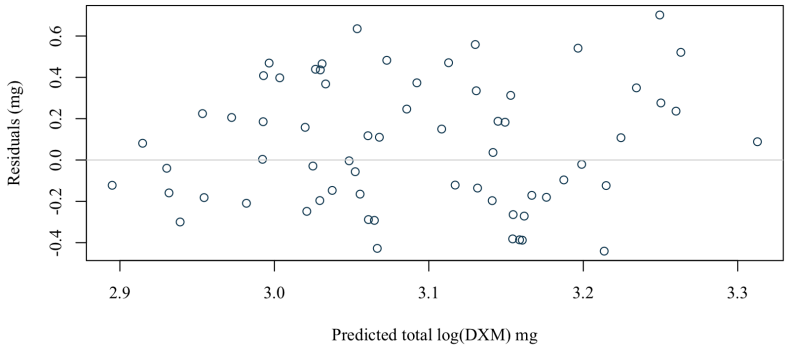

Figure 17: Residuals for (n = 65) beta blocker patient prediction against log predicted values, outliers removed

Table 23: Residual analysis for predictions on test set and beta blocker patients with five removed outliers

Evaluation	Test set	Beta blocker set without outliers
MAE	0.306	0.260
MSE	0.134	0.094

5 Sensitivity analysis

Before summarizing the results, it is found necessary to discuss the potential sources of error of the study and how they are handled. To design a completely flawless study is probably never possible, and in this case the observational data collection brings with it several issues that could potentially be avoided in an experimental study. These, and other issues with this study will be presented here.

5.1 Sample size

5.1.1 Low frequency categories

One source of error is derived from the six categorical variables included in the research. This is a multi-dimensional problem and therefore not an easy task to visualize. As seen in the initial frequency Table 3, there are in total two covariates with three categories, and four with two. In an ideal case, one would have patients belonging to all different combinations of categories (9). In total there is 144 unique combinations for the case and control group respectively. This criterion is not fulfilled in the data, where for example only one individual in the case group went through biopsy. There will be at least one man or female (or any other category for that matter) that has not had a biopsy in the case group.

A solution to this problem would be to simply remove the patients belonging to low frequency categories. However, that would reduce the sample size and

the scope of conclusions. Instead, it is noted that the results cannot be fully reliable without more data across categories being gathered.

5.1.2 Required overall sample size

The need for more data extend beyond the categorical frequencies. Given the number of covariates included, an increase in overall sample size would help to gain certainty in the estimated coefficients. Figure 18 displays the change in confidence interval width for the beta blocker coefficient with regards to sample size given the full model presented in the "statistical approach" section 4.1.5.

By experimenting with the sample, and drawing random samples from the pool of 441 patients, the certainty in the beta estimate for the beta blocker coefficient can be visualized. The curve is expected to flatten out with a sample size sufficient enough, however the derivative is still negative as the sample size reaches 100%. It is noted that the decrease in variance is heavier between 20% and 60% compared to the end, indicating that an acceptable size is approaching. Nevertheless, it is not reached within the current sample size.

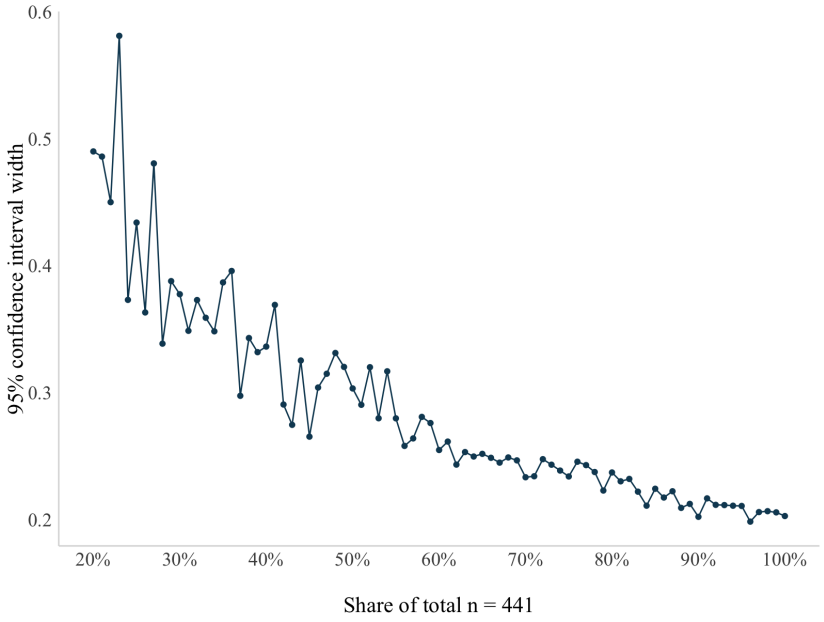

Figure 18: Beta blocker beta variance against share of sample size, initial regression

5.2 Outliers

There are several ways to identify outliers. In the previous section, 4.2.8, residual analysis was computed to identify possible outliers by plotting residuals against the predicted values. This was not done for creating the model, but rather for the interpretation of the test data and beta blocker data. This resulted in the removal of five outliers in the beta blocker data when interpreting the result. In this section, additional outlier analysis will be presented.

The term outlier refers to an observation that, in some sense, is inconsistent

with the rest of the observations in the dataset. In the second edition of the book "Applied Regression Analysis: A research Tool" (9), the authors opted to redefine the concept of outliers, introducing distinct terminology for various types of outliers. This resulted in the following three terms:

1. Outlier: The value of the dependent variable is in some sense inconsistent with the rest of the sample.
2. Potentially influential observation: An outlier in the span of the independent variables - outliers in the **X**-space.
3. Outlier in residual: The observed residual is larger than might reasonably be expected from random variation alone.

This sensitivity analysis will take use of these three terms, and as the data for the two regression models are different, this outlier analysis will be divided into two separate parts: one for the "statistical approach" and one for the "predictive approach". As the reduced model in the "statistical approach" no longer include the beta blocker parameter, the full model before reduction will instead be used for this outlier sensitivity analysis.

5.2.1 Outlier analysis for the "statistical approach"

The first step conducted to identify possible outliers was by visual inspection of the data. Both scatterplots and boxplots against the dependent variable total DXM were studied. Figures 29 and 30 in the Appendix show the named scatterplots and boxplots, as well as highlights the outliers. In total six outliers were identified: patients with study IDs 250, 331, 373, 443, 526, and 537. Four of which where non-beta blocker patients. Previously, study ID 331 was found as an potentially influential observation when studying the scatterplot of age at surgery vs tumor size, see Figure 25 in the Appendix. Now it has also been identified as an outlier.

Potentially influential observations are far from the gravity of the **X**-space (9). This distance can be measured through leverage, v_{ii}, which is calculated as

$$v_{ii} = \mathbf{x}_i\text{'}(\mathbf{X'X})^{-1}\mathbf{x}_i$$

where $\mathbf{x}_i$' is the i-th row of $\mathbf{X}$ (9). The limits on v_{ii} are $1/n \leq v_{ii} \leq 1/c$, where c is the number of rows in $\mathbf{X}$ that have the same values as the i-th row. A rule of thumb is that a leverage above $2(p+1)/n$ is considered high, where p is the number of independent variables and n is the sample size. When calculating the leverages, 33 observations had a leverage higher than the threshold of $2(p+1)/n$. Only two of the previous mentioned outliers (study ID 331 and 373) had a leverage considered high. An observation having a high leverage does however not imply that they are actually influential. To further research the influence of these potentially influential observations, as well as the six previous mentioned outliers, visual inspection of Cook's distance was performed.

Cook's distance measures the effect of an observation i on the estimate $\hat{\beta}$, and can be calculated as (9):

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' \cdot (\hat{Var}(\hat{\beta}))^{-1} \cdot (\hat{\beta}_{(i)} - \hat{\beta})}{(p+1)}.$$

There is no unanimous consensus on what cut-off values to use to highlight a too large Cook's distance. One common cut-off is to say that observation i may be influential if $D_i > F_{(0.5,p+1,n-(p+1))}$, or that it is not a problem as long as $D_i < 4/\text{n}$ (9). The latter one of the two will be used here. Figures 19 and 20 show the Cook's distance for all the observations, where in the first figure the previous six identified outliers are highlighted, and in the second figure, the previously 33 observations with high leverage are highlighted. Out of the six outliers, four had a Cook's distance over the limit, and out of the 33 observations with high leverage, 15 had a Cook's distance over the limit. Two of these are the same patients, study ID 331 and 373. It remained 12 observations with high Cook's distance, not previously detected as outliers or high leverage observations.

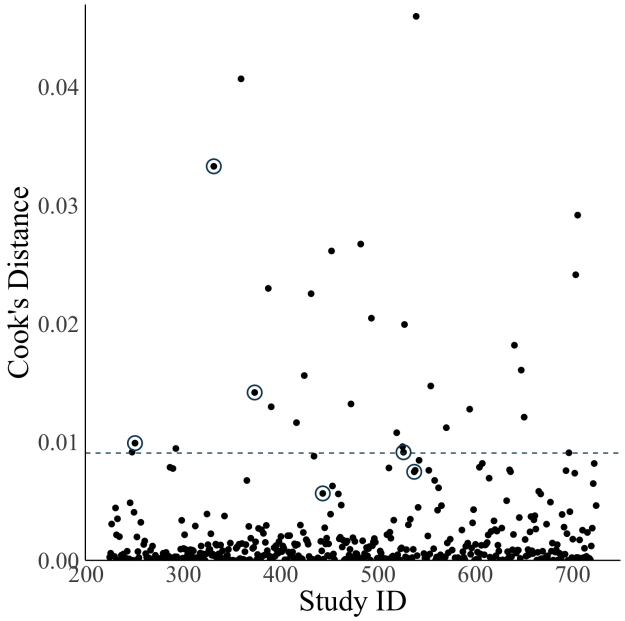

Figure 19: Plot of Cook's distance where previous six detected outliers are highlighted

The study IDs 331 and 337 frequently appeared as outliers across various outlier measurements. Although there is no indication of any measurements errors, it might have been beneficial to temporarily exclude these observations prior to model fitting. To assess the impact of these, the regression model was refitted using data that excluded them. It was particularly important to examine if the exclusion of the two outliers influenced the significance of the $\hat{\beta}_{\text{beta blocker}}$ estimate. However, refitting the model with the new data did not change the insignificance of the $\hat{\beta}_{\text{beta blocker}}$ estimate.

Regarding all of the observations exhibiting high leverage, it was probably the right decision to include them in the regression analysis. Before removal of observations with high leverage it is necessary to check that they do in fact have an influence on the model. Removing observations with high leverage without assessing their impact on the model can result in misidentifying outliers. Additionally, high leverage observations may also be legitimate data points that

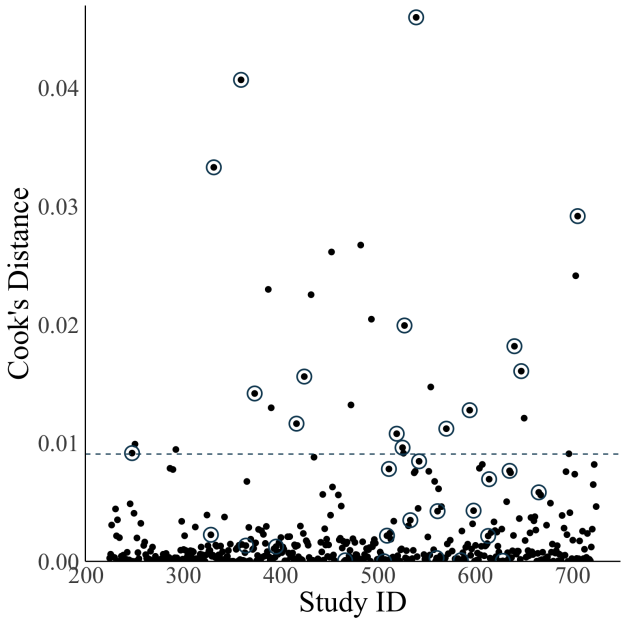

Figure 20: Plot of Cook's distance where previous 33 detected observations with high leverage are highlighted

represent extreme cases within the sample. Analysis of Figure 31 and 32 in the Appendix indicates that this is often the case, and the removal of these observations could therefore lead to incorrect conclusions and reduced generalizability.

There are in total 29 observations with Cook's distance higher than the set threshold. The removal of observations always comes at the cost of statistical power, and to remove these observations without further investigation could be detrimental. It can imply a loss of information, biases in the analysis, an overreaction to influential observations, and so on. When dealing with outliers, it's crucial to consider the context of the analysis. What may be considered influential in one study, may not be influential in another. As the goal for this regression model is to answer if patients that are on beta blocker are prescribed a smaller dose of DXM compared to patients that are not on beta blocker, the estimation of $\hat{\beta}_{\text{beta blocker}}$ is what is of most importance. To investigate the impact of the mentioned observations, and potential other observations, on the estimate $\hat{\beta}_{\text{beta blocker}}$ the measurement DFBETAS was used.

Each $DFBETAS_{j(i)}$ is the standardized change in $\hat{\beta}_j$ when the ith observation is deleted from the analysis (9). It is calculated as

$$DFBETAS_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_j(i)}{s_i\sqrt{c_{jj}}},$$

where s_i^2 is the variance estimate from a regression where observation i is excluded, and c_{jj} is the *(j+1)*-th diagonal element of $(\mathbf{X'X})^{-1}$. In Figure 21, the DFBETAS for the estimated $\hat{\beta}_{\text{beta blocker}}$ are plotted, and the observations with previously high Cook's distance are highlighted. It should be mentioned that neither study ID 331 or 373 had a high DFBETA value. Analyzing observations with high DFBETA values revealed no definitive reasons for their exclusion.

However, given their significant influence on the parameter, it was decided to reevaluate the model without these observations. This was to assess any potential impact on the significance of the $\hat{\beta}_{\text{beta blocker}}$ estimate. Not too surprisingly, the beta blocker parameter estimate was still not significantly different from zero.

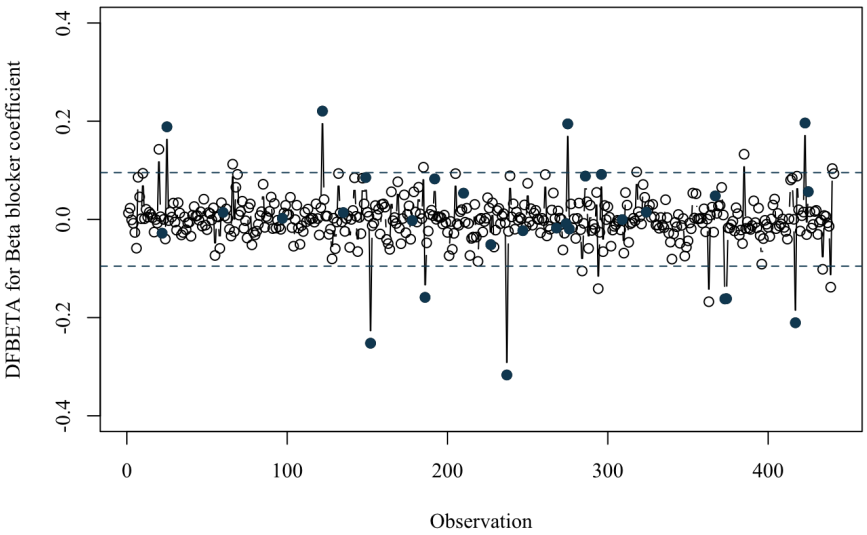

Figure 21: DFBETAs for beta blocker coefficient in statistical regression, observations with Cook's distance above threshold highlighted

5.2.2 Outlier analysis for the "predictive approach"

As a subset of the full data was used in the predictive approach, potential outliers had to be re-identified. It turned out that all of the outliers for non-beta blocker patients presented for the previous regression model were included in the training data set and remained outliers. This meant that study ID 331, 373, 443 and 537 remained outliers for this part as well.

When measuring the Cook's distance, only one of the identified outliers showed to be an influential observation in the model, as seen in Figure 22. Namely, study ID 331. This is a 24 year old male, with a large tumor of 12 cm preoperatively that was given a fairly low total DXM dosage of 16 mg. As seen in Figure 22, there are several other observations with a high Cook's distance, however these were not previously identified as outliers.

The leverage analysis showed 21 observations with leverage above the $2(p+1)/n$ threshold, a subset of the 33 previously mentioned. This represents almost 10% of the training data with $n = 271$. It is not too surprising given that several unbalanced categorical covariates were used for this analysis. The Cook's distance was above the threshold for 13 of these, as seen in Figure 23. Their removal can be discussed, nevertheless they were kept. It is particularly important when building predictive models that they not only fit the existing data, but can do well for new data as well. To remove too many outliers from the training data

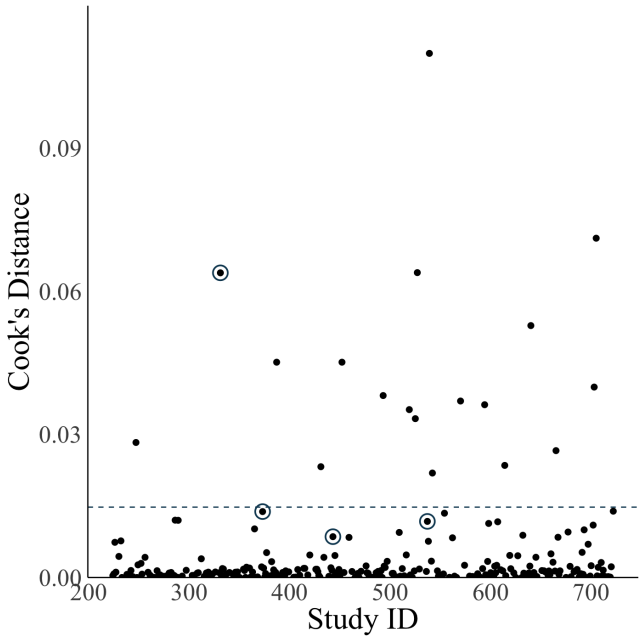

Figure 22: Cook's distance for observations in the train data set in the predictive model, the four outliers highlighted

could make the model perform worse on unseen data where extreme values also exist.

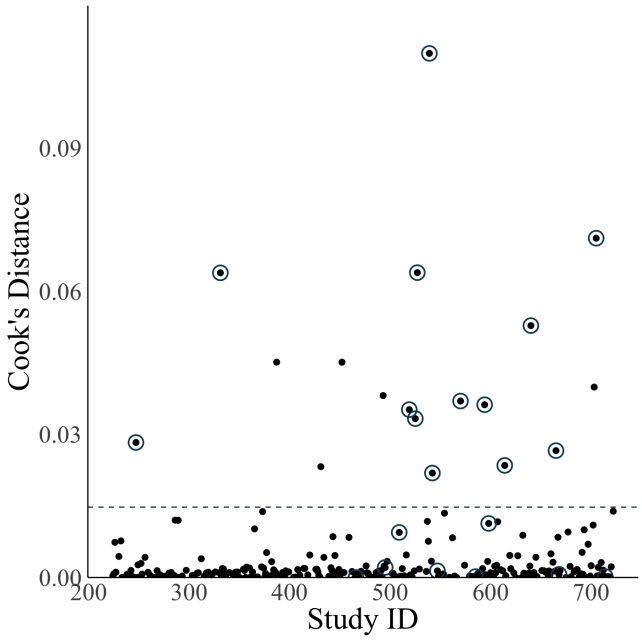

Figure 23: Cook's distance for observations in the train data set in the predictive model, 21 high leverage points highlighted

5.3 Discrepancies in data

There are also sources of error in the data collection that needs to be addressed.

1. The ordination of DXM
2. The relationship between prescribed DXM and edema
3. Patients that passed away during time of treatment
4. The meaning of beta blocker consumption
5. The selection of beta blocker patients

To begin with, the prescription of DXM is complicated. Doctors often prescribe set dosages (3) such as 20 mg or 40 mg, meaning that even in an ideal case the prescription would not follow a perfect function of the independent variables. It is, for example, unlikely that a patient would be prescribed 23.4 mg of DXM, although that may be the optimal dose. However, the dosages are not set enough to be considered categorical data. In this study, they were treated as continous (to avoid having over 30 categories), however in reality it is a mix. Additionally, a prescription is not equal to intake. A patient may be prescribed a dose and choose not to consume it.

Furthermore, the prescribed DXM dose does not correlate perfectly with the severity of the edema. It is the best measurement to be found in the data, but depending on how the patient is experiencing the edema different dosages can be prescribed for the same severity. If one patient experience heavier symptoms than another for the same level of swollenness, it is likely that it would be prescribed a larger dose and vice versa. A solution to this problem is to observe the edema directly through x-rays, a data collection that is currently ongoing at this point of time.

In addition, some patients may have died during the course of treatment, causing their DXM dose to become zero at a certain point in time. For the patients that had a zero dosage at the end of the treatment cycle, it was not fully clear in the if they had a true zero DXM dosage or if they had passed away. The patients that passed away at some point during the treatment but still received a first dose of DXM therefore remained in the study. This is of course problematic as it skews the total DXM value.

As mentioned in the previous section 4.1.1, seven patients were never prescribed any DXM (had a total DXM dose of zero) and were therefore removed. The reason these patients were never prescribed any DXM was unclear. It could be that they had no need for DXM, or it could be that they passed away, or the data input was faulty. So, although the exclusion of these patients did entail some loss of generalization, the decision to exclude them was still made. It was supported by the fact that the data seemed unreliable, as well as the difficulty in including them for the log-linear regression.

Moreover, the beta blocker status of the patients was somewhat vague. That they are classified as beta blocker patients does not necessarily mean that they were on beta blockers at the time of surgery, but rather that they had taken them throughout their lives. For future studies, this needs to be more clearly

defined as the half-life of beta blockers is not more than days (3) and thus timing does matter.

Finally, the observational nature of the study made it impossible to control who were given the treatment. Beta blocker patients are a certain type of patients, and it is likely that other related diseases such as obesity and diabetes could be over-represented in this group (3). This could in itself affect the prescribed DXM dosage. More data regarding this has not been collected, however it is something to keep in mind when observing the results.

5.4 Breach of assumptions

To be able to fit a linear regression, the underlying assumption is that the residuals are normally distributed (9). These assumptions are slightly breached already for the full model in the "statistic approach", as seen in Figure 24 where a straight line would be expected in the QQ-plot. This is a fairly strict constraint when working with real data. Given the small sample size and the fact that DXM in reality is not perfectly continuous, it was decided that the residuals followed a good enough distribution.

When reducing the model in the regression part, F-tests and t-tests are used. The reliability of these also builds upon the assumption that the residuals are normally distributed throughout the model reduction. As the full model residuals were not normally distributed to begin with, the tests could not be trusted blindly. They were thus used as guidelines rather than a source of truth.

5.5 Generalization of results

To whom do the results of this study apply? As seen in Table 3 and Figures 2 and 3, the general population is not fully represented in this study. The mean age at surgery for the patients is over 60 years, and females are slightly underrepresented. However, the aim of this study is not that the sample is representable for the entire human population, but rather for glioblastoma patients. These patients are often older and match the sample distribution quite well (3). Conclusions regarding infratentorial tumors and tumors that are IDH1 mutant are harder to draw given their low representation in the data, however they are also more rare.

There are also few cases of biopsy, so those cases need to be studied more carefully. Biopsies are often part of initial diagnostics that are followed up with a sub- or gross-total resection (3). In afterthought, these patients could probably have been removed from the study.

5.6 P-hacking

Finally, p-value hacking should also be addressed. When using multiple significance tests to prove the same result from one data set, there is a risk to get false positives through the phenomenon referred to as p-hacking (27). In this study, this is done twice:

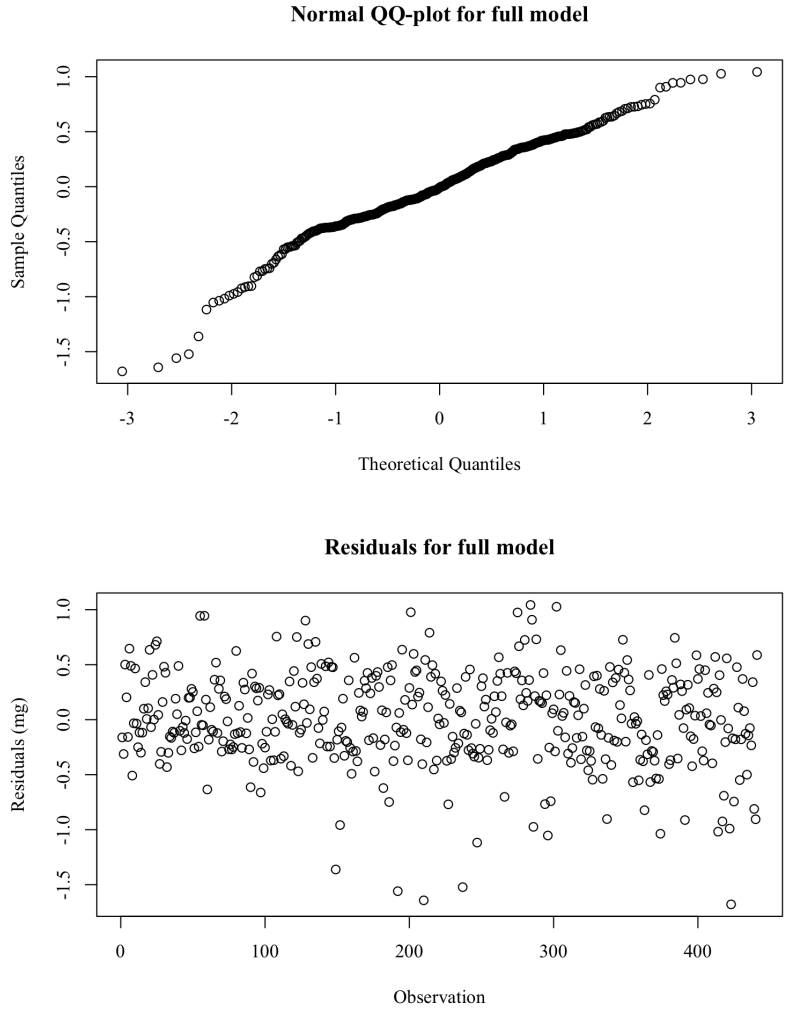

Figure 24: Normal QQ-plot and residuals for the full model in the "statistical approach"

1. In the initial analysis when using propensity score matching and inverse probability weighting

2. In the regression part, when two regression models are fit to the same data

One way to reduce the risk of p-hacking is to lower the significance level when using multiple tests (27). The relevant variable in this study was the beta blocker status which showed to be insignificant across all tests. Consequently, there were no positive or false positive found, and no wrong conclusions were drawn through p-hacking. Should the same study method be used for a new data set in the future, one should be aware that the significance level should be reduced from 0.05 when using multiple tests. Multiple methods were used here to do a more robust analysis, as well as experimenting with different mathematical methods.

6 Topics of future research

Summarizing the takeaways from the sensitivity analysis for future studies, this was a study based on observational data. Should someone want to proceed with studies within the subject, despite the lack of results observed in this paper, there are some suggested improvements.

To begin with, observe x-rays of the edema rather than prescribed DXM dosage. This is a more direct measure and would reduce bias associated with the medical prescription and intake. Secondly, design an experimental study instead. If ethical aspects allow, a random selection of beta blocker patients would be preferred as this could reduce the bias of underlying diseases related to beta blocker medication. Thirdly, investigating other relations than linear. Exploring beyond linear regression allows for techniques better suited to capturing the complexity of real-world data relationships. Additionally, only proceed with patients that are alive through all stages of treatment to avoid the risk of dead patients affecting the average dosage. Finally, try to gain a sample size larger than around 100 beta blocker patients to increase the statistical power of the analysis.

7 Conclusion and final remarks

This thesis aims to answer the question if glioblastoma patients on beta blockers were associated with a lower severity of edema at the time around surgery. Edema was measured through prescribed dosages of dexamethasone. To answer the question, the following analysis was done:

1. Checked if prescribed DXM differed between the case and control group without covariate control
2. Controlled for confounders in the case and control group through propensity score matching and inverse probability weighting, and then checked for significant difference
3. Fitted a linear model to the full data set to see if the beta blocker covariate had significant contribution to the model
4. Fitted a linear model to parts of the non-beta blocker patient data set and see if it fitted as well to beta blocker patients as to the test data

The conclusion was that beta blocker status did not show to be significantly associated with prescribed dexamethasone at the time around surgery for glioblastoma patients. It is understandable that the clinical experience indicated such a relationship, nevertheless it seems to derive from the characteristics of beta blocker patients such as their age and extent of resection rather than the effect of the medicine.

This study aimed to explore the relationship between beta blocker intake and severity of edema, and builds upon the assumption that edema can be measured through the prescription of DXM. The hope is to find an alternative treatment for edema, removing the need for steroids that currently inhibits the effect of the new oncolytic virotherapy treatment. However, the beta blockers' direct

effect on patient survival has not been explored in this study, and is a highly relevant subject to investigate before incorporating beta blockers as a part of glioblastoma treatment.

As a final remark, the authors would like to express their gratitude towards the people who have helped bring this thesis to life. A special thank you to our supervisor at Lund University, Andreas Jakobsson, as well as our supervisor at the hospital. Both of you have inspired us with your great knowledge and skills in the mathematical as well as medical field. You have challenged us along the way and ensured that we a study that we can feel proud to present. To the reader of this thesis, thank you for taking part of our work. Please do not hesitate to ask questions and suggest improvements.

8 Reference list

(1) American Brain Tumor Association/Yu J, Ghiaseddin A, Ahluwalia M. Glioblastoma [Internet]. Chicago: American Brain Tumor Association; 2023 [2022-12; 2024-05-01]. Available from: *https://www.abta.org/tumor_ types/glioblastoma-gbm/*

(2) National Brain Tumor Society. About Glioblastoma [Internet]. Newton: National Brain Tumor Society; 2024 [Not identified; 2024-03-16]. Available from: *https://braintumor.org/events/glioblastoma-awareness-day/about-glioblastoma/*

(3) Personal communication, J Gerstl 2024-03-11

(4) Stummer W. Mechanisms of tumor-related brain edema. Journal of Neurosurgery. 2007; 22(5): 1-7.

(5) Shoaf M, Desjardins A. Oncolytic Viral Therapy for Malignant Glioma and Their Application in Clinical Practice. Neurotherapeutics. 2022; 19(6): 1818–1831.

(6) NHS. Beta blockers [Internet]. England: Department of Health and Social Care; 2024 [2022-12-02; 2024-03-15]. Available from: *https://www.nhs.uk/conditions/beta-blockers/*

(7) Tewarie I, Senders J, Hulsbergen A, Kremer S, Broekman M. Beta-blockers and glioma: a systematic review of preclinical studies and clinical results. Neurosurg Rev. 2021;44(2):669-677.

(8) Kim H, Park Y, Lee H, Kwon DD, Song J, Chang I. Propranolol Inhibits the Proliferation of Human Glioblastoma Cell Lines through Notch1 and Hes1 Signaling System. PMC. 2021;64(5):716-725.

(9) Rawlings J, Pantula S, Dickey D. Applied Regression Analysis: A Research Tool. Second Edition. New York: Springer-Verlag; 1998.

(10) Brain Tumour Research. Glioblastoma Multiforme [Internet]. Milton Keynes: Brain Tumour Research; 2024 [Not identified; 2024-04-01]. Available from: *https://braintumourresearch.org/pages/types-of-brain-tumours-glioblastoma-multiforme-gbm*

(11) Mass General Brigham Incorporated. IDH Mutation Brain Tumors [Internet]. Boston: Mass General Brigham Incorporated; 2024 [Not identified; 2024-04-01]. Available from: *https://www.massgeneralbrigham.org/en/patient-care/international/treatments-and-specialties/cancer-care/neuro-oncology/idh-mutant-tumors*

(12) National Cancer Institute. Brain and Spine Tumor Anatomy and Functions [Internet]. Bethesda: National Cancer Institute; 2024 [2023-10-23; 2024-04-03]. Available from: *https://www.cancer.gov/rare-brain-spine-tumor/tumors/anatomy*

(13) Pfund Z, Szapary L, Jaszberenyi O, Nagy F, J Czopf. Headache in intracranial tumors. Cephalalgia 1999; 19(99): 787–790.

(14) Loghin M and Levin VA. Headache related to brain tumors. Curr Treat Options Neurol 2006; 8(1): 21–32.

(15) Rosenbaum, Paul R.; Rubin, Donald B. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 1983; 70(1): 41–55.

(16) Chesnaye N, Stel V, Tripepi G, Dekker F, Fu E, Zoccali C, Jager K. An introduction to inverse probability of treatment weighting in observational research. Clinical Kidney Journal. 2022; 15(1): 14-20.

(17) Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine. 1900; Series 5:157-174.

(18) Fisher R. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. Journal of the Royal Statistical Society. 1922; 85(1): 87-94.

(19) Mann H, Whitney D. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics. 1947; 18: 50-60.

(20) Spearman C. The proof and measurement of association between two things. American Journal of Psychology. 1904; 15 (1): 72–101.

(21) Pan W, Bai H. Propensity score interval matching: using bootstrap confidence intervals for accommodating estimation errors of propensity scores. BMC Med Res Methodol. 2015; 15: 53.

(22) Chowdhury1 M, Turin T. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health. 2020; 8(1): e000262.

(23) Bandyopadhyay P, Forster M, editors. Philosophy of Statistics. North Holland; 2011.

(24) Cadrin S, Karr L, Mariani S, editors. Stock identification Methods. Second Edition. Academic Press; 2014.

(25) Efron B, Tibshirani R. An Introduction to the Bootstrap. Boca Raton: Chapman and Hall/CRC; 2011.

(26) Roitberg B. Reference Module in Life Sciences. Upplaga. [Not identified]: Elsevier; 2016.

(27) Head M, Holman L, Kahn A, Jennions M. The Extent and Consequences of P-Hacking in Science. PLoS Biol. 2015; 13(3): e1002106.

9 Appendix

Table 24: Beta blocker correlation coefficient (95% lower limit, upper limit) for Pearson's product moment correlation ρ

DXM	Age	EOR	MGMT	Tumor size
Pre-op	**-0.23** (-0.4, -0.04)	0.01 (-0.18, 0.20)	-0.2 (-0.21, 0.18)	**0.32** (0.13, 0.48)
Post-op	-0.15 (-0.33, 0.04)	**-0.31** (-0.48, -0.13)	-0.02 (-0.21, 0.17)	0.17 (-0.03, 0.35)
2w post-op	-0.14 (-0.32, 0.05)	-0.13 (-0.32, 0.06)	0.01 (-0.18, 0.20)	-0.03 (-0.22, 0.17)
6-8w post-op	-0.05 (-0.24, 0.14)	-0.08 (-0.27,0.11)	0.06 (-0.13, 0.25)	0.16 (-0.04, 0.34)
Total	**-0.28** (-0.45, -0.09)	**-0.20** (-0.38, -0.01)	-0.00 (-0.19, 0.19)	**0.33** (0.15, 0.49)

Table 25: Non-beta blocker correlation coefficient (95% lower limit, upper limit) for Pearson's product moment correlation ρ

DXM	Age	EOR	MGMT	Tumor size
Pre-op	-0.08 (-0.18, 0.02)	-0.02 (-0.13, 0.08)	-0.03 (-0.13, 0.08)	**0.24** (0.14, 0.34)
Post-op	-0.04 (-0.14, 0.06)	0.02 (-0.08, 0.13)	-0.00 (-0.11, 0.10)	0.03 (-0.07, 0.14)
2w post-op	-0.04 (-0.15, 0.06)	**-0.10** (-0.21, -0.01)	-0.02 (-0.13 0.08)	0.04 (-0.07, 0.14)
6-8w post-op	-0.05 (-0.15, 0.06)	**-0.12** (-0.22, -0.01)	-0.03 (-0.13, 0.07)	-0.01 (-0.11, 0.09)
Total	-0.09 (-0.20, 0.01)	-0.07 (-0.17, 0.03)	-0.03 (-0.14, 0.07)	**0.17** (0.07, 0.27)

Table 26: Frequency tables for data with patients removed with any missing data (n=452)

Extent of resection	Case	Control
Biopsy	0.97% (1)	3.44% (12)
Sub-total resection	47.57% (49)	38.68% (135)
Gross-total resection	51.46% (53)	57.88% (202)

MGMT	Case	Control
Unmethylated	58.25% (60)	45.27% (158)
Weakly methylated	1.94% (2)	8.02% (28)
Methylated	39.81% (41)	46.70% (163)

Sex	Case	Control
Female	37.86% (39)	47.56% (166)
Male	62.14% (64)	52.44% (183)

Stupp	Case	Control
Yes	22.33% (23)	25.21% (88)
No	77.67% (80)	74.79% (261)

IDH1	Case	Control
Wild type	98.06% (101)	96.56% (337)
Mutant	1.94% (2)	3.44% (12)

Tumor location	Case	Control
Supratentorial	99.03% (102)	99.14% (346)
Infratentorial	0.97% (1)	0.86% (3)

Table 27: Frequency tables for data used in log-linear regression (n=441)

Extent of resection	Case	Control
Biopsy	1.00% (1)	3.20% (11)
Sub-total resection	49.00% (49)	38.42% (131)
Gross-total resection	50.00% (50)	58.36% (199)

MGMT	Case	Control
Unmethylated	58.00% (58)	46.04% (157)
Weakly methylated	2.00% (2)	8.21% (28)
Methylated	40.00% (40)	45.75% (156)

Sex	Case	Control
female	37.00% (37)	47.51% (162)
male	63.00% (63)	52.49% (179)

Stupp	Case	Control
Yes	22.00% (22)	24.93% (85)
No	78.00% (78)	75.07% (256)

IDH1	Case	Control
Wild-type	98.00% (98)	96.48% (329)
Mutant	2.00% (2)	3.52% (12)

Tumor location	Case	Control
Supratentorial	100% (100)	100% (341)

Table 28: Correlation between variables (continuous and dummy) in regression "statistical approach"

Variables	Methylated	Weakly meth.	Biposy
Methylated	1	-0.241	0.047
Weakly methylated	-0.241	1	0.010
Biposy	0.047	0.010	1
Sub-total resection	0.009	-0.023	-0.139
Stupp complete yes	0.260	0.036	0.035
IDH1 mutant	0.046	0.054	-0.030
Female	-0.004	-0.010	0.044
Beta blocker yes	-0.048	-0.103	-0.057
Age at surgery	0.045	-0.041	0.072
Tumor size	-0.030	0.067	-0.055
Variables	**Sub-total res.**	**Stupp, yes**	**IDH1, mutant**
Methylated	0.009	0.260	0.046
Weakly methylated	-0.023	0.036	0.054
Biopsy	-0.139	0.035	-0.030
Sub-total resection	1	-0.126	0.008
Stupp complete yes	-0.126	1	0.169
IDH1 mutant	0.008	0.169	1
Female	0.035	-0.067	-0.086
Beta blocker yes	0.090	-0.029	-0.036
Age at surgery	-0.028	-0.196	-0.304
Tumor size	0.079	-0.028	0.187
Variables	**Female**	**Beta blocker, yes**	**Age**
Methylated	-0.004	-0.048	0.045
Weakly methylated	-0.010	-0.103	-0.041
Biopsy	0.044	-0.057	0.072
Sub-total resection	0.035	0.090	-0.028
Stupp complete yes	-0.067	-0.029	-0.196
IDH1 mutant	-0.086	-0.036	-0.304
Female	1	-0.088	0.025
Beta blocker yes	-0.088	1	0.252
Age at surgery	0.025	0.252	1
Tumor size	-0.059	-0.017	-0.191

Variables	Tumor size
Methylated	-0.030
Weakly methylated	0.067
Biopsy	-0.055
Sub-total resection	0.079
Stupp complete yes	-0.028
IDH1 mutant	0.187
Female	-0.059
Beta blocker yes	-0.017
Age at surgery	-0.191
Tumor size	1

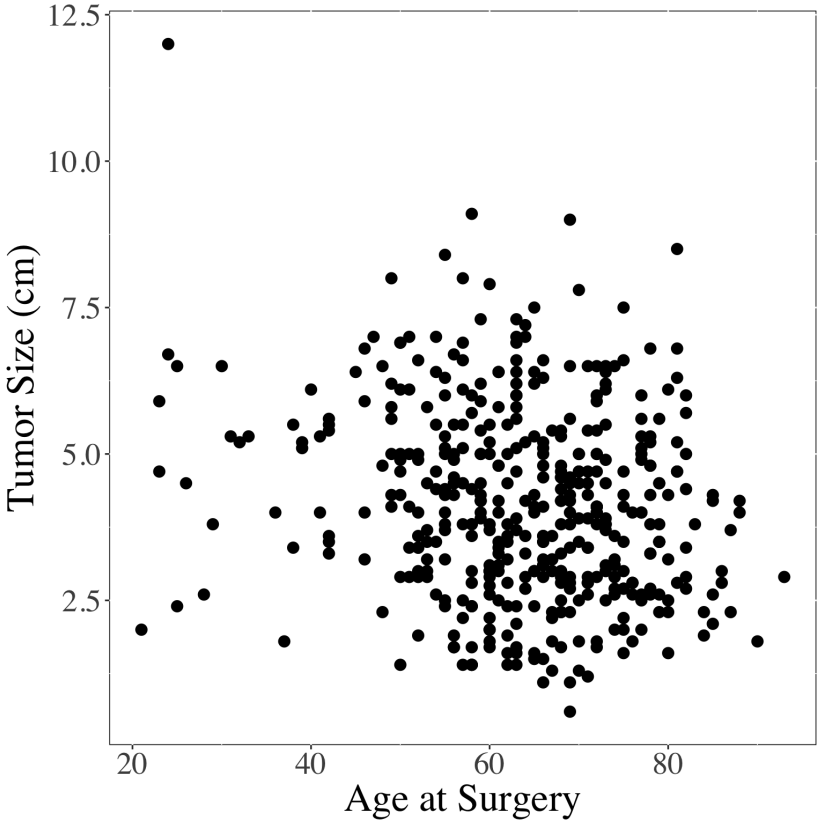

Figure 25: Scatter plot of age at surgery vs tumor size in "statistical approach"

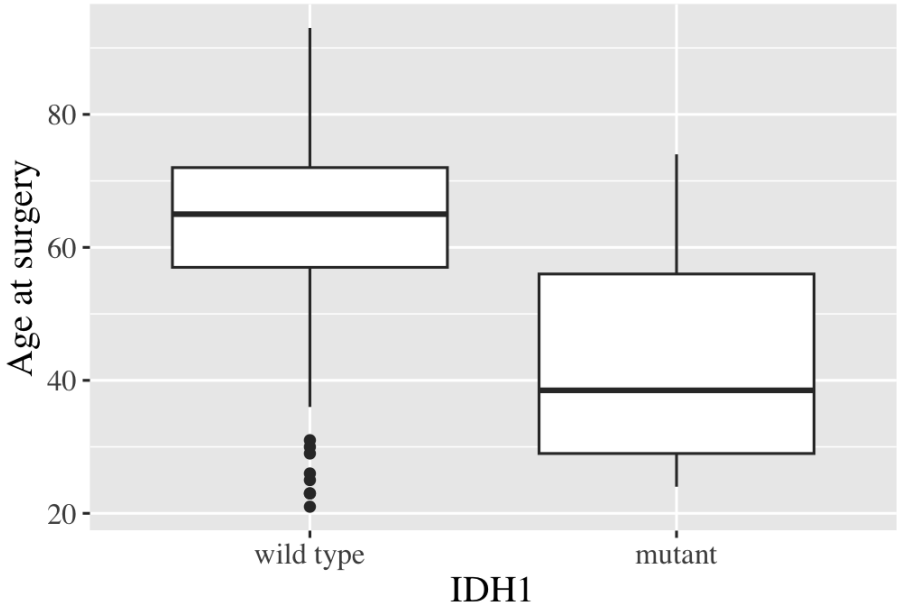

Figure 26: Boxplot of IDH1 (wild type or mutant) vs age at surgery

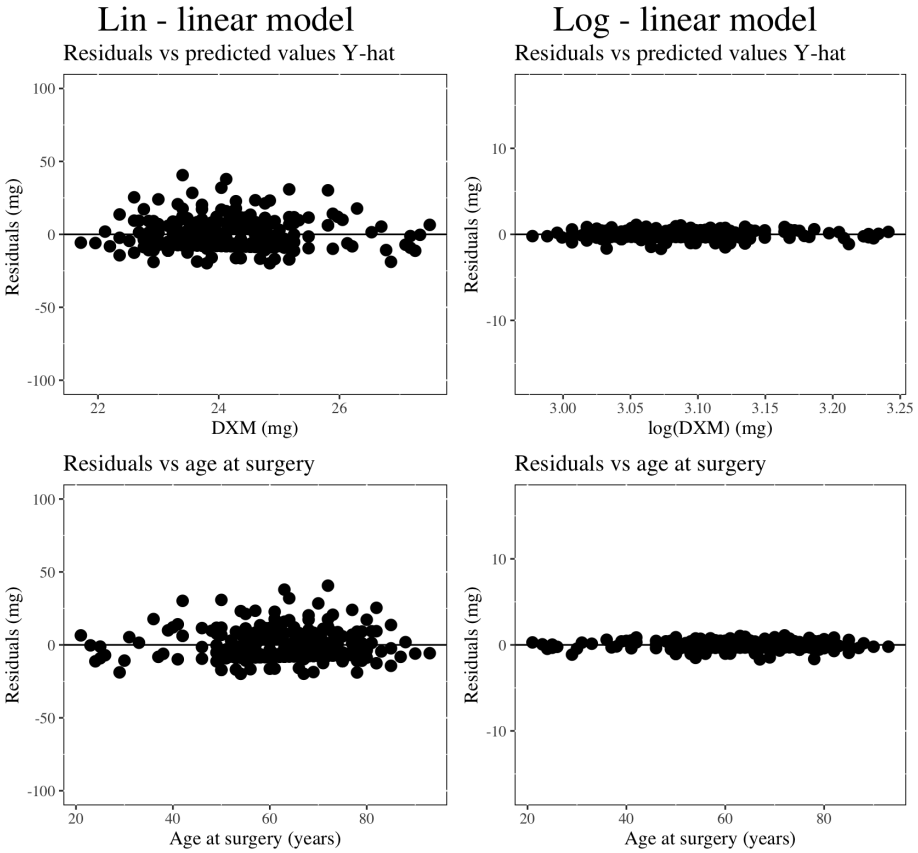

Figure 27: Residuals plotted against age at surgery and predicted DXM for model selection in "predictive approach"

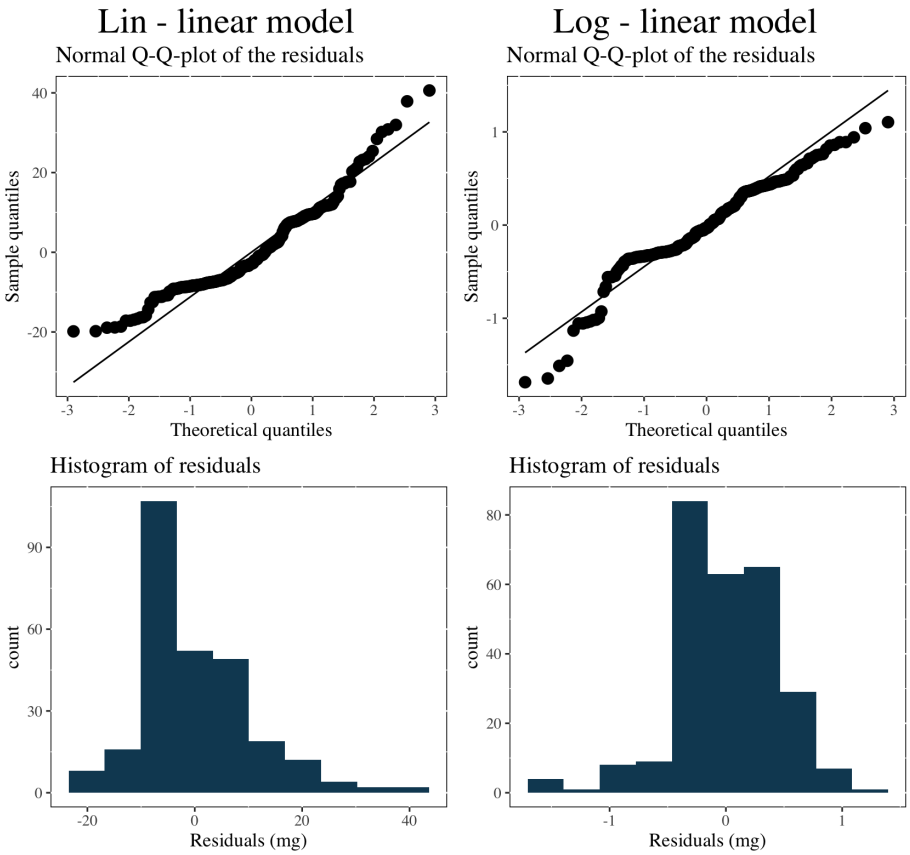

Figure 28: QQ-plot and histogram of residuals for for model selection in "predictive approach"

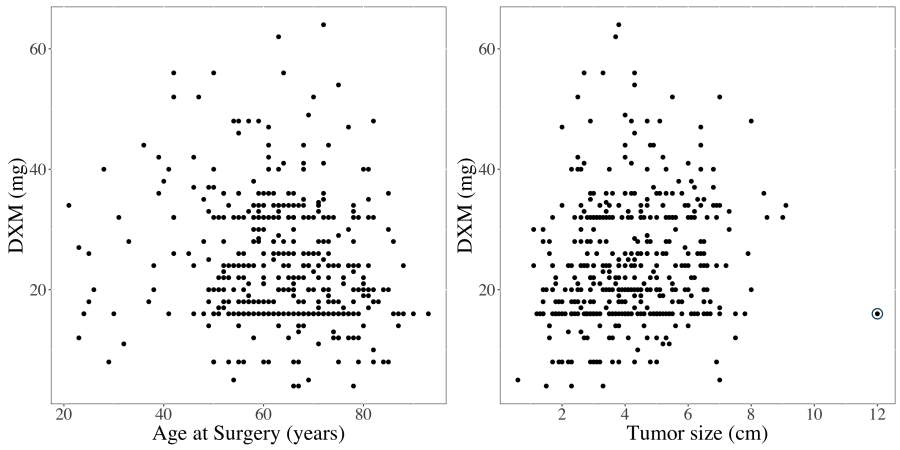

Figure 29: Scatterplots of the independent continuous variables against total DXM. Based on data used for "statistical approach". Study ID 331 identified as outlier

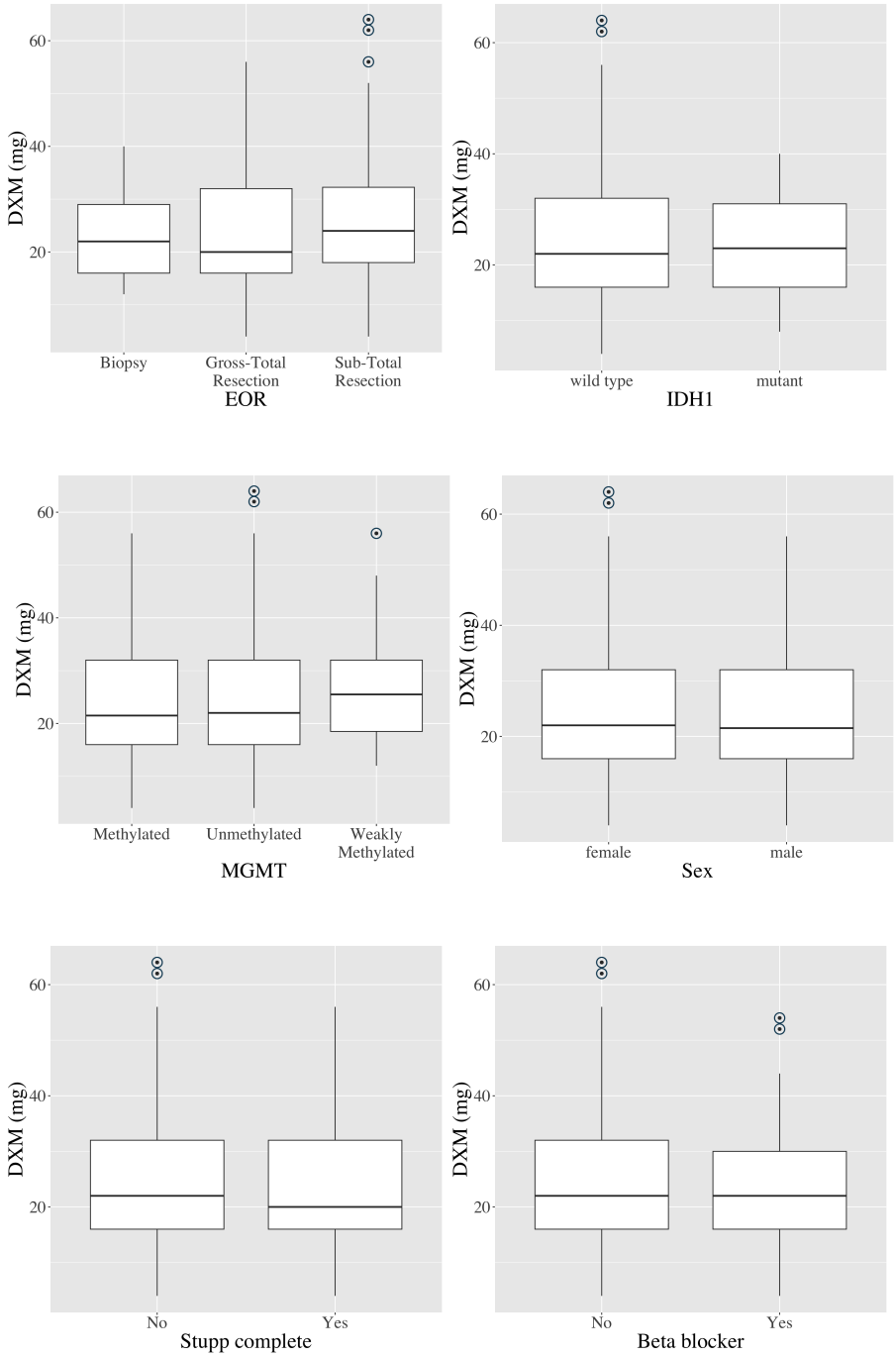

Figure 30: Boxplots of the independent categorical variables against total DXM. Based on data used for "statistical approach". Study IDs 250, 443, 537, 373, and 536 identified as outliers

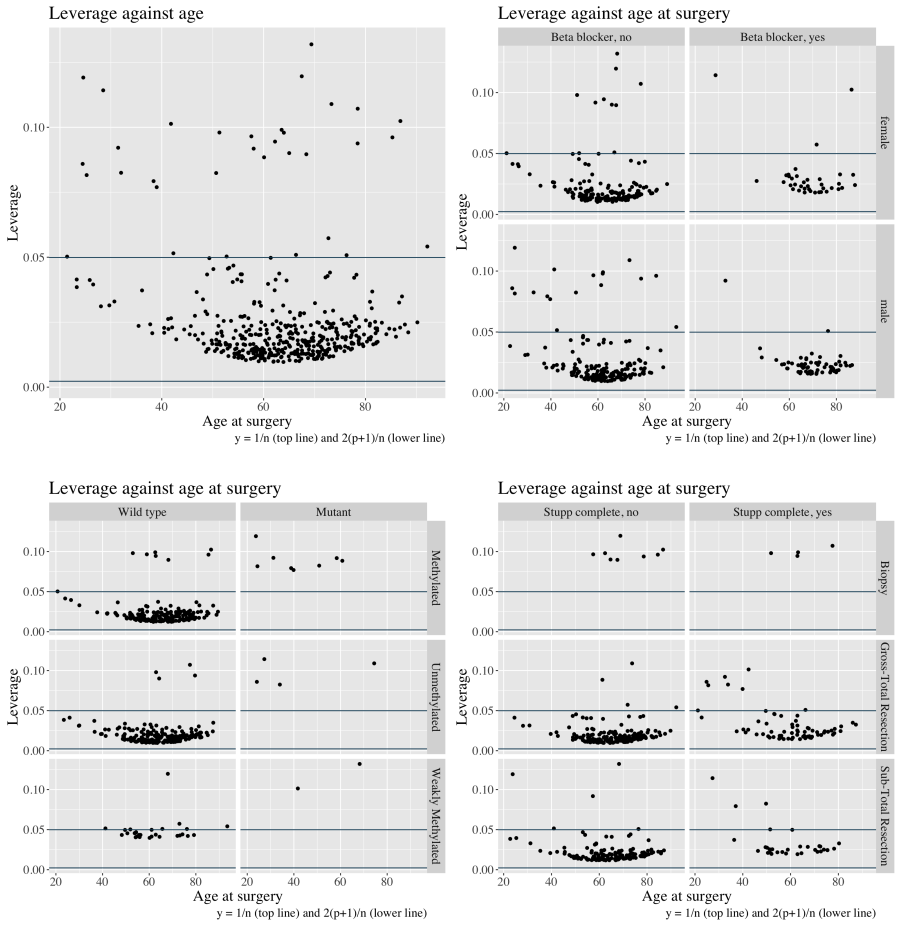

Figure 31: Leverage against age based on full patient group, beta blocker status, IDH1 status and Stupp complete for the full statistical regression model

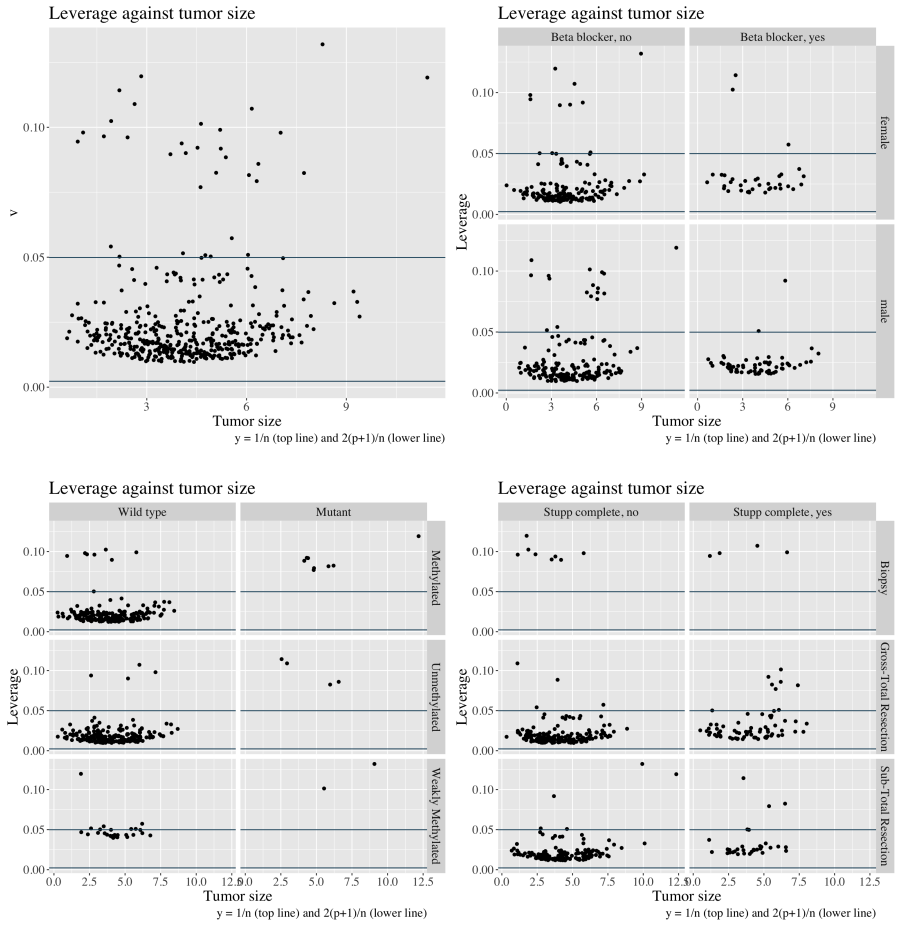

Figure 32: Leverage against tumor size based on full patient group, beta blocker status, IDH1 status and Stupp complete for the full statistical regression model