Universität Zürich[UZH]

Master Program in Biostatistics
www.biostat.uzh.ch

STA490: Statistical Practice in Clinical Research

Prof. Ulrike Held

EBPI, Department of Biostatistics

Tissue Bridges Cut-Off Value in Spinal Cord Injury Patients

Analysis for Patrick Freund, Lynn Farner

Supervision by Torsten Hothorn

Mike Krähenbühl (*mikeeric.kraehenbuehl@uzh.ch*)

Version of December 2, 2024

1 Abstract

We investigated the predictive value of preserved tissue bridges for recovery after a cervical spinal cord injury. This analysis was based on the Nogo-A study, a randomized, placebo-controlled, multicentre, phase 2b clinical trial with 78 patients in the NG101 (intervention) group and 48 patients in the placebo group. The relationship of baseline tissue bridges with recovery, measured by upper extremity motor score (UEMS), was modelled with a generalized additive model (GAM). The analysis showed that the amount of preserved midsagittal tissue bridges had a linear positive relationship with the 6-month recovery (Δ-UEMS) in both treatment groups. The parasagittal tissue bridges showed also a positive relationship with recovery. However, in the treatment group the recovery reached a plateau for values of preserved parasagittal tissue bridges higher than (approximately) 2 mm.

(Remark: append the abstract for cut-off value if we decide to include it) A total amount of XX in preserved tissue was determined as cut-off value, separating the patients into subgroups of prognosed recovery based on the baseline tissue bridge value. Such a cut-off value can be used in clinical practice to better forecast the probability for recovery and as a selection criterion in clinical trials.

2 Introduction

Traumatic spinal cord injury (SCI) is a severe condition resulting for example from falls and road traffic injuries. Bundles of nerves and nerve fibers are damaged and transmission of nerve signals with the brain are suppressed (NINDS, 2024). SCI can lead to a loss of motor and/or sensory function below the injury site and decrease quality of life (WHO, 2024). The prognosis of recovery is crucial for the patient and the treating physician.

Biomarkers that serve for prognosis are needed. Studies have shown that the amount of preserved tissue bridges around the lesion are an important factor to characterizing and predicting the recovery after a spinal

cord injury. Tissue bridges function as preserved neuronal pathways that restore connection between the brain and the body. It has been observed that the amount of tissue bridges is positively correlated with the recovery (Huber et al., 2017), (Pfyffer et al., 2019) and that certain cut-off values are prognostic for an improved recovery after 3 months and 12 months (Pfyffer et al., 2024).

Another biomarker that showed some prognostic power is the lesion area. It can be seen as the counterpart to the tissue bridges since it represents the damaged tissue. The lesion area is negatively correlated with the recovery (Huber et al., 2017).

The aim of this analysis was to further investigate the relationship between clinical biomarkers such as preserved midsagittal/parasagittal tissue bridges and lesion volume with the recovery after a spinal cord injury. Depending on the form of this relationship, cut-off values could be determined that help to separate patients into likely and unlikely recovery. Assigning patients to subgroups depending on their predicted outcomes is crucial. It helps improving the planning, individualisation and evaluation of therapies. By segmenting patients in clinical studies into more homegeneous subgroups according to their distinct predicted recovery profile, allows to separately assess the treatment sucess in the respective subgroups. Finally, a cut-off value could also be used as inclusion criterion in a clinical trial.

3 Research Questions

1. What is the relationship between preserved tissue bridges after a spinal cord injury with the recovery over time measured by the upper extremity motor score?
2. Can a cut-off value in tissue bridges be determined that helps to separate patients into poor and good expected recovery?
3. Which biomarker (midsagittal tissue bridges, parasagittal tissue bridges or lesion volume) can be used best to determine a cut-off value?

4 Methods

Study Design

Type of study

This is an exploratory post-hoc analysis based on the study "NoGo-A antibody treatment in acute cervical spinal cord injury". The Nogo-A study was a randomized, placebo-controlled, multicentre phase 2b clinical trial. Therefore, we will only briefly explain the methods used in the Nogo-A study, but emphasise those that are relevant to answering the research questions stated in Section 3. The methods of the Nogo-A study are described in detail in the main study report. (Remark: cite as soon as published)

Anti Nogo-A Study

The study "NoGo-A antibody treatment in acute cervical spinal cord injury" was a randomised, placebo-controlled, multicentre, phase 2b, clinical trial where the effects of the treatment with an antibody was investigated. The antibody is called NG101 and it binds to the protein Nogo-A. Nogo-A is a protein that inhibits the growth of nerve fibers. The idea is that by targeting Nogo-A, the nerve fibers can regrow better and the recovery is improved.

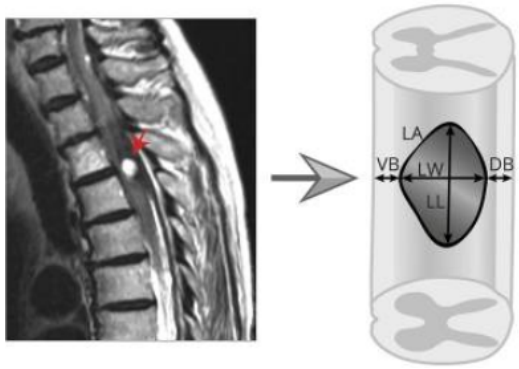

Figure 1: Typical T2W midsagittal slice (left) and schematic lesion segmentation (right) with the quantitative MRI measures analyzed (lesion area [LA], lesion length [LL], lesion width [LW], ventral midsagittal tissue bridges [VB], and dorsal midsagittal tissue bridges [DB]) (Pfyffer et al., 2019).

The study was conducted in specialized SCI centres in Germany, Switzerland, Spain and the Czech Republic. Between 2019 and 2022, 126 patients with acute traumatic cervical spinal cord injury (neurological level of injury C1-C8) were recruited and randomized to treatment (NG101) and placebo in a ratio 2:1.

The primary outcome was the change in upper extremity motor score (UEMS, according to ISNCSCI) in 6 months. The UEMS is a measure that assesses the motor function in multiple muscle groups in the arms and it can take values between 0 and 50. It was measured at baseline and after approximately 1, 3 and 6 months. The statistical analysis was mainly conducted with linear mixed-effect models. The overall primary outcome was not significant, but some subgroups showed promising results.

Tissue Bridges and Lesion Volume

Patients underwent MRI scanning at screening. The protocol included T1-weighted, sagittal T2-weighted, and axial T2-weighted anatomic scans of the cervical spinal cord, centered at lesion level. From the scans, the midsagittal tissue bridges, parasagittal tissue bridges and the lesion volume were determined.

The amount of preserved midsagittal tissue bridges (measured in mm) were calculated as the sum of the ventral and dorsal bridges that were visible on the midsagittal MRI slice. Figure 1 illustrates the schematic lesion segmentation.

Parasagittal tissue bridges were calculated similarly to midsagittal bridges. In this case, however, tissue bridges were also measured in the parasagittal MRI slices. The measurements from all slices where the lesion was visible were added together. To adjust for differences in cord size, this total was divided by the number of slices showing tissue bridges, giving the average amount of preserved tissue bridges in millimeters.

The lesion volume was also calculated by MRI screening. (Remark: add more details?)

Data collection

The primary outcome upper extremity motor score was measured at screening, baseline and 30, 84 and 168 days post baseline. The time variable in the analysis indicates the measurement timepoint of UEMS in days after start of medication. For example if the UEMS was measured at time=1, this means that it was measured 1 day after first dose. The UEMS, assessed using ISNCSCI, ranges from 0 to 50 in increments of 1. A higher UEMS indicates better motor functions.

Statistical Analysis

Data Preparation

Most data preparation was already done for the Nogo-A study. The primary dataset contained all variables except the midsagittal tissue bridges and the lesion volume. Those measures were provided in two separate excel files. We joined the baseline measures of the biomarkers with the patient id and added them to the primary dataset. The dataset with the lesion volume contained 0 where the slice thickness was missing. In this case, the volume should not be 0 but NA instead. This was corrected. (Remark: according to Lynn Farner)

Imputation Methods

There are two types of missing values that are relevant in the analysis of the relationship between baseline biomarkers and the recovery. First, the UEMS was not measured 4 times (baseline and 3 follow up) for all patients. Second, baseline biomarker values are missing for some patients. The reasons for the missing values will be provided in Section 5. In both cases missing values are not imputed. The primary aim of this exploratory analysis is to investigate the relationship between tissue bridges and recovery. No significance testing will be performed. We do not want to introduce additional uncertainty by imputing. The missing biomarker values are handled by excluding the respective patients from the analysis, hence a complete case analysis is performed.

Descriptive Statistics and Simple Methods

The following analysis is conducted for each biomarker individually. Patients are divided into four subgroups according to the quartile of the biomarker. Then, the UEMS over time is plotted by treatment group and quartile of the biomarker. This should give a first impression of the recovery profiles conditional on the treatment group and different ranges of the biomarker value.

The distributions of the biomarkers against the baseline UEMS are visualized by scatter plots. This should uncover potential surprising baseline distributions in the treatment groups.

Next, a generalized additive model (GAM), as will be described in the next Section 4, is fitted on both treatmet groups separately. The resulting smooth functions of the predictors are be presented, providing a visual guide on how the relationships are modelled. By using different values of the range of the respective biomarker, predictions for the recovery over time are made with the GAM and added to a plot with the raw patient data. Each line in the plot represents a hypothetical patient with a different baseline value for the biomarker. This should give an impression of how the recovery changes over time depending on the value of the biomarker. For each predicted recovery trajectory, the difference between the UEMS score at 6-month follow-up and baseline is calculated (Δ-UEMS). This difference is then plotted against the baseline value of the biomarker and should give an idea of how the predicted recovery changes depending on the value of the biomarker. If the relationship between the biomarker value at baseline and the 6-month recovery is non-linear, the Δ-UEMS should be represented as a straight line.

The research question of finding a cut-off value depends on this relationship of Δ-UEMS and biomarker. Only if there was a non-linear relationship where the Δ-UEMS is higher or lower over a certain range of the biomarker, a cut-off value would be meaningful.

Description of advanced statistical methods

Before we consider whether a cut-off value can be determined, we need to investigate the relationship between the tissue bridges and the recovery. Since we want to make as few assumptions as possible, a model is needed that allows for great flexibility. A generalized additive model (GAM) can model complex, non-linear

relationships between predictors and response. It extends the generalized linear model (GLM) by replacing the linear combination of the predictors $\sum \beta_j X_j$ by the sum of smooth functions of the predictors $\sum f_j(X_j)$ (Hastie and Tibshirani, 1986).

A simpler model such as a linear mixed-effects model would require to make the assumption that the relationship between each predictor and the response variable is linear. For illustration; assuming we model the problem by a linear mixed-effects model, a single coefficient for the tissue bridges would be estimated. This coefficient would represent the average effect of the tissue bridges on the recovery. But this is not what we want to achieve with this analysis. Instead, we want to investigate whether the relationship of the tissue bridges on the response (UEMS) changes depending on the value of tissue bridges. A GAM models the problem by fitting a smooth function of the tissue bridges. This smooth function can have any shape.

The research question of finding a cut-off value is dependent on the shape of the function that models the relationship between tissue bridges and recovery. If the function is linear, a cut-off value is not meaningful. But if the function suggests, for example, that there is no effect on recovery for tissue bridge values below 2 mm, but for tissue bridge values above 2 mm the recovery increases, then a cut-off value at 2 mm could be determined.

Hence, in this analysis we use a GAM to model the problem. Equation 1 shows the formula of the model that is applied. The model consists of an overall intercept, the main effects of baseline tissue bridges and time, and additionally an interaction effect for tissue bridges and time. The patient-specific random effects capture correlations between repeated measurements.

$$\text{UEMS}_{ij} = \beta_0 + f_1(\text{tissue}_i) + f_2(\text{tmstd}_{ij}) + f_3(\text{tissue}_i, \text{tmstd}_{ij}) + \alpha_{id} + \beta_{id} \times \text{tmstd} + \epsilon_{ij} \tag{1}$$

- UEMS_{ij} is the upper extremity motor score of patient i at time j. It can take values between 0 and 50.
- β_0 is the overall intercept.
- $f_1(\text{tissue}_i)$ is the smooth function of the baseline tissue bridges. It shows the effect on UEMS depending on the value of tissue bridges. It is centered at 0.
- $f_2(\text{tmstd}_{ij})$ is the smooth function of time. It shows the effect on the UEMS over time. The time variable is normalized so that time = 0 corresponds to the baseline measurement (day of first medication) and time = 1 corresponds to 6 month follow-up (day 168 after first medication). It is centered at 0.
- $f_3(\text{tissue}_i, \text{tmstd}_{ij})$ is the interaction effect of tissue bridges and time on the UEMS. It shows how the effect of tissue bridges on the UEMS changes over time. This interaction is the effect that is not already included in the main effects of tissue bridges and time. It is also centered at 0.
- $\alpha_{id} + \beta_{id} \times$ tmstd are the patient-specific random effects. They account for the fact that each patient might have an individual intercept (baseline UEMS value) or trajectory of recovery over time regardless of the value of the predictors used in the model. Without including these random effects, the model would assume that all patients with the same value of the biomarker (e.g. midsagittal tissue bridges) at a given time point would have the same baseline UEMS and identical recovery trajectory over time. This would mean that no patient-specific deviations would be allowed and therefore could lead to a biased estimation of the effect on UEMS.
- ϵ_{ij} is the error term. It is assumed to be normally distributed with mean 0 and variance σ^2. (Remark: is this true in a GAM?)

Restricted maximum likelihood (REML) is used as fitting algorithm due to its resistance to occasional severe over-fitting (Wood, 2006, p. 267). The smooth functions are approximated using penalized regression splines.

In contrast to random effects in linear mixed-effects models, where explicit parameters are estimated for each individual, the random intercepts and slopes in this GAM are specified as special cases of smooths. Same as the other terms, they are approximated using penalized regression splines. Additionally, the patient-specific slopes can be non-linear.

To avoid making assumptions about the treatment effect of NG101 in any way on the relationship between tissue bridges on the recovery, the model is fitted separately for the placebo and NG101 group.

Implementation

All analyses were performed in the R programming language (R Core Team, 2024) using base packages and the following analysis-specific packages: `mgcv` to fit generalized additive models with gam() (Wood, 2011), `lattice` to use xyplot() for plotting (Sarkar, 2008).

5 Results

Table 1: Baseline Characteristics by Treatment Group. The measures baseline measurements for UEMS and the three biomarkers are presented as mean (sd).

	Overall	Placebo	NG101
Patients	126	48	78
Time from injury to 1st injectino (days)	22.56 (5.19)	23.81 (4.62)	21.78 (5.40)
UEMS	16.39 (9.21)	17.08 (7.94)	15.96 (9.93)
Midsagittal tissue bridges	1.37 (1.58)	1.40 (1.31)	1.35 (1.75)
Parasagittal tissue bridges	1.73 (1.45)	1.99 (1.38)	1.56 (1.48)
Lesion volume	585.27 (704.80)	399.28 (234.40)	712.72 (874.94)

Anti Nogo-A Study and Neuroimaging Substudy

126 patients with acute traumatic cervical spinal cord injury (neurological level of injury C1-C8) were included in the Nogo-A study. 78 were allocated to the NG101 group and 48 received the placebo. The patients were recruited in 4 countries. The mean age was 46.2 years with a standard deviation of 16.7 years. The percentage of males was 84.9%.

103 out of 126 patients underwent MRI imaging at screening in an exploratory neuroimaging substudy. The screening was carried out on average 17.2 (sd 7.1 Remark: Nogo-A reported 6.9) days after the injury occurred. Some scans were invalid due to metal or motion artifacts and poor image quality. The midsagittal tissue bridges were calculated based on the midsagittal MRI slice and the parasagittal tissue bridges were calculated based on the parasagittal MRI slices. Table 1 shows the baseline measurements in the treatment groups.

Figure 2 shows the combinations of missing values of the three biomarkers. It is important to note that only invalid or not recorded values are considered missing values. A biomarker value of 0 is a valid measurement and therefore included in the analysis. For 90 out of 103 patients, all 3 biomarker values were recorded. 8 patients have a missing value for the parasagittal tissue bridges, but a value for midsagittal tissue bridges

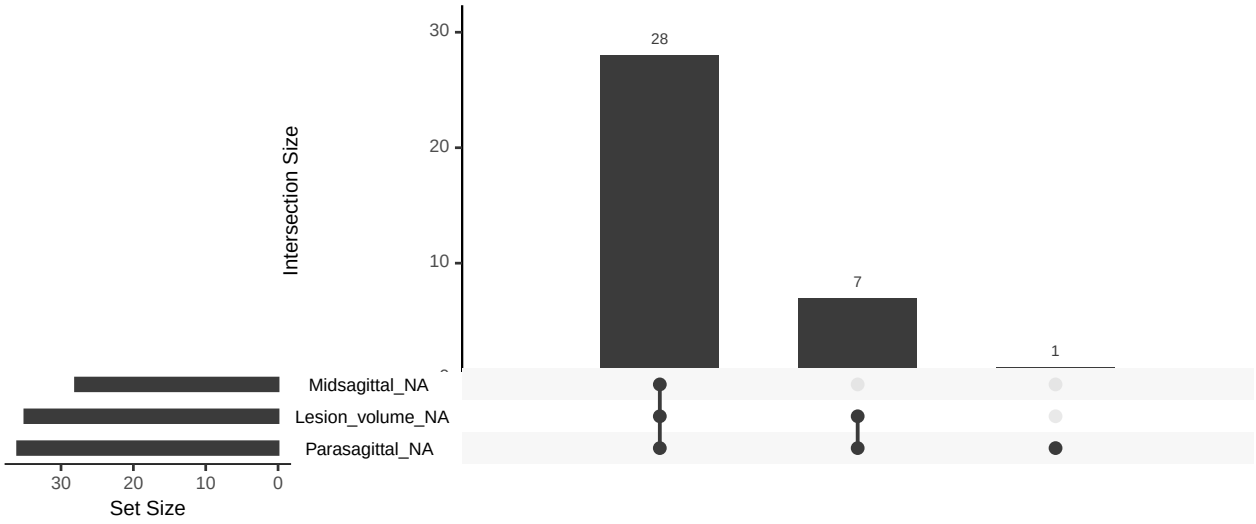

Figure 2: Visualization of the pattern of missing variables.

is recorded. Such a case can happen, for example, when the midsagittal MRI slice is valid, but one or multiple parasagittal slices are invalid because of artifacts that are not visible in the midsagittal slice.

Out of the 126 patients, 111 patients have all 4 measurements for the UEMS recorded. The other patients either discontinued after the first dose or did not receive the full dosing as per protocol. The upper extremity motor score was first measured on average 1.2 days before receiving the first dose. (Remark: Check with clinicians why some patients discontinued, or appendix of the study when published.)

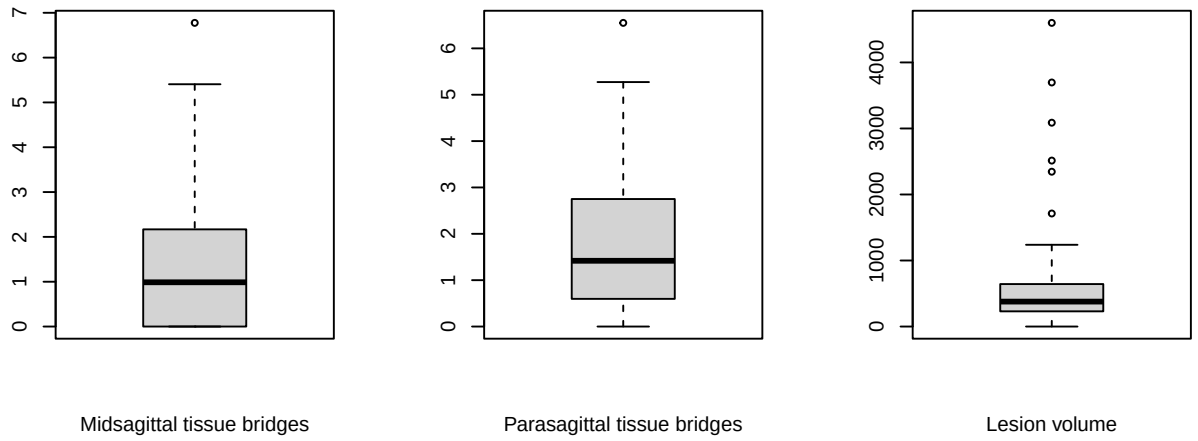

Figure 3: Boxplots of the midsagittal and parasagittal tissue bridges and the lesion volume at baseline. Concerning the lesion volume, 6 patients are considered outliers and were excluded from the analysis. They are clinically correct observations, but cause problems when fitting the model.

Figure 3 shows the distributions of the biomarkers at baseline. With regard to the lesion volume, 6 pa-

tients are considered outliers and were excluded from the further analysis. All 6 patients were part of the NG101 group. The statistical models would be strongly influenced by these outliers.

Relationship of Biomarkers with Recovery

In this section, the results are shown for each biomarker separately (midsagittal tissue bridges, parasagittal tissue bridges, lesion volume). The results are structured in the following way: First, the raw data is presented for each treatment group and corresponding quartile of the biomarker range. Second, the baseline distribution of the biomarker values against the UEMS in the treatment groups is shown. Third, the relationship of the biomarker with the recovery is analyzed by the smooth functions that have been fitted by the GAM. It is important to note that the main effects for the biomarker and the time have to be interpreted together with the interaction effect of the two variables. Only looking at the main effects individually does not explain the full relationship. Fourth, the fitted models are visualized by adding predictions to the raw data plots. Fifth, the differences in the predicted UEMS scores between baseline and 6-month follow-up (Δ-UEMS) are calculated for different values of the biomarker.

Midsagittal Tissue Bridges

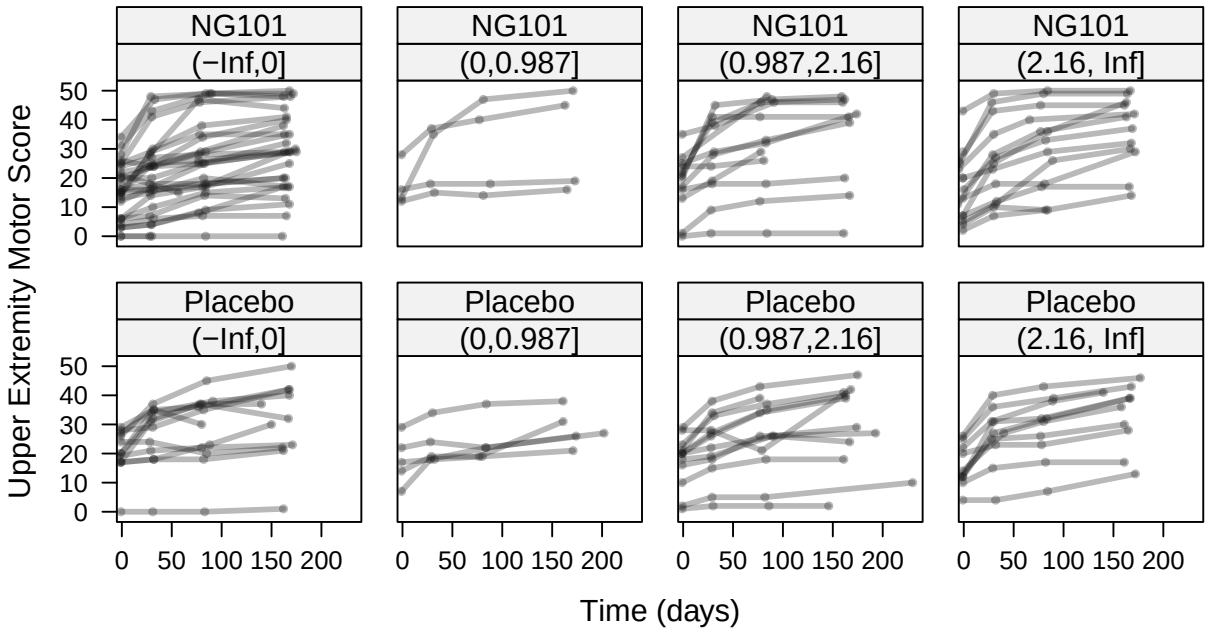

Figure 4: Patient data per treatment group separated in the four quartiles for the amount of midsagittal tissue bridges. The quartiles are based on the overall distribution of midsagittal tissue bridges.

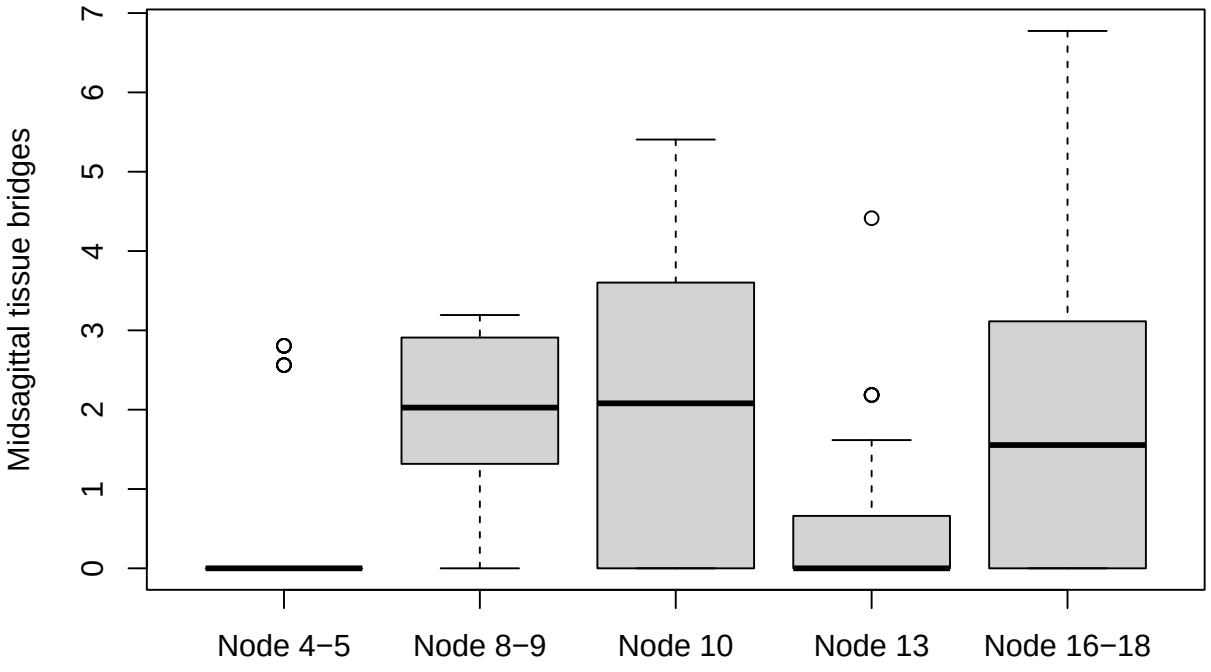

Figure 5: Boxplots of the midsagittal tissue bridges by URP-CTREE nodes.

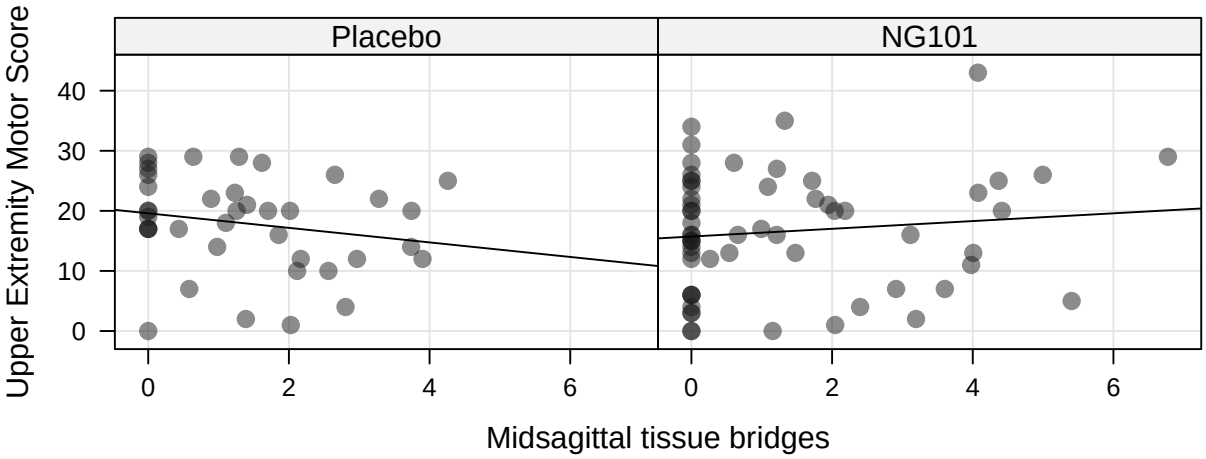

Figure 6: Scatterplot of the baseline midsagittal tissue bridges against Upper Extremity Motor Score in the two treatment groups.

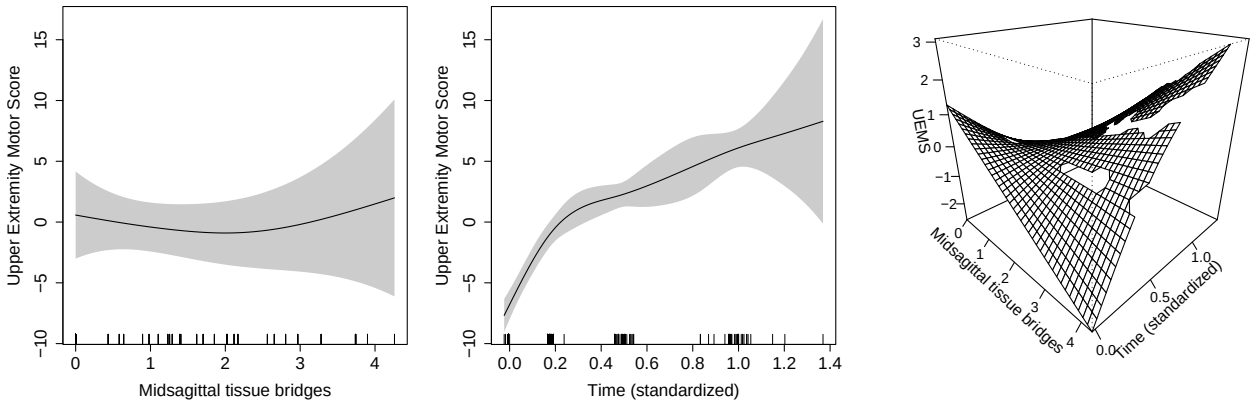

Figure 7: Smooth functions fitted by the GAM for the placebo group.

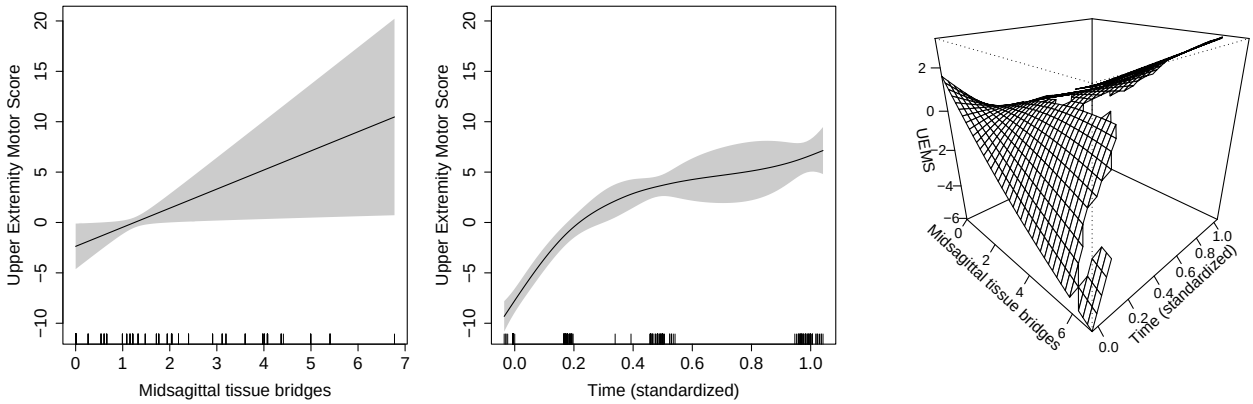

Figure 8: Smooth functions fitted by the GAM for the NG101 group.

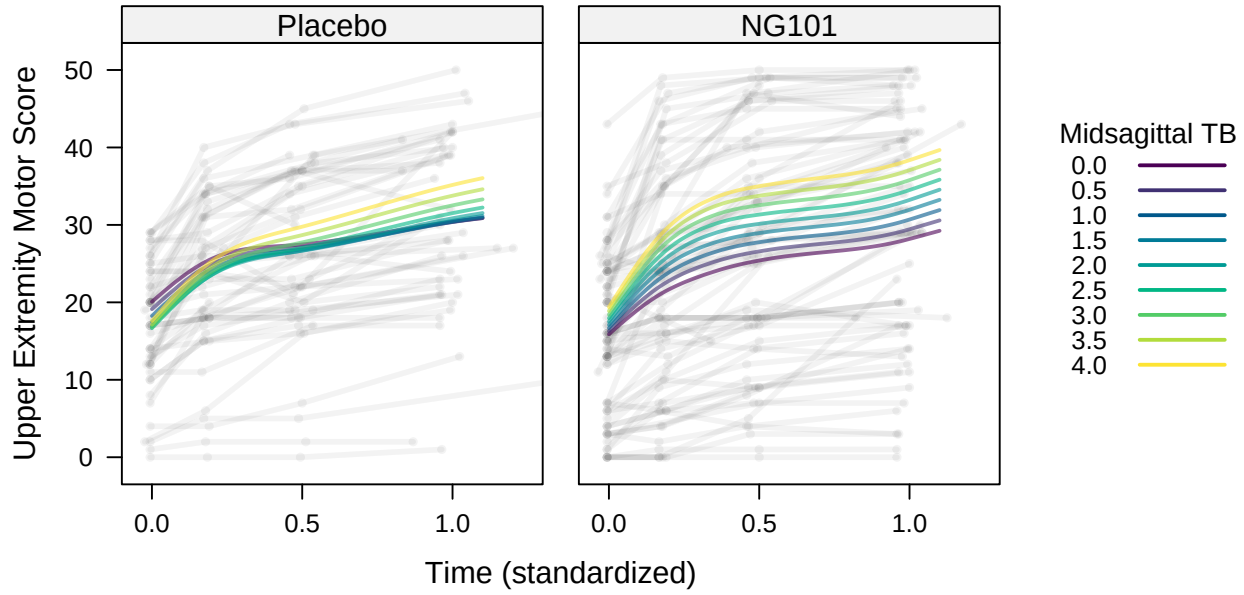

Figure 9: Predictions made with different values for midsagittal tissue bridges. For each treatment group the respective GAM was fitted separately. The only parameter that changes for the different predictions is the amount of midsagittal tissue bridges.

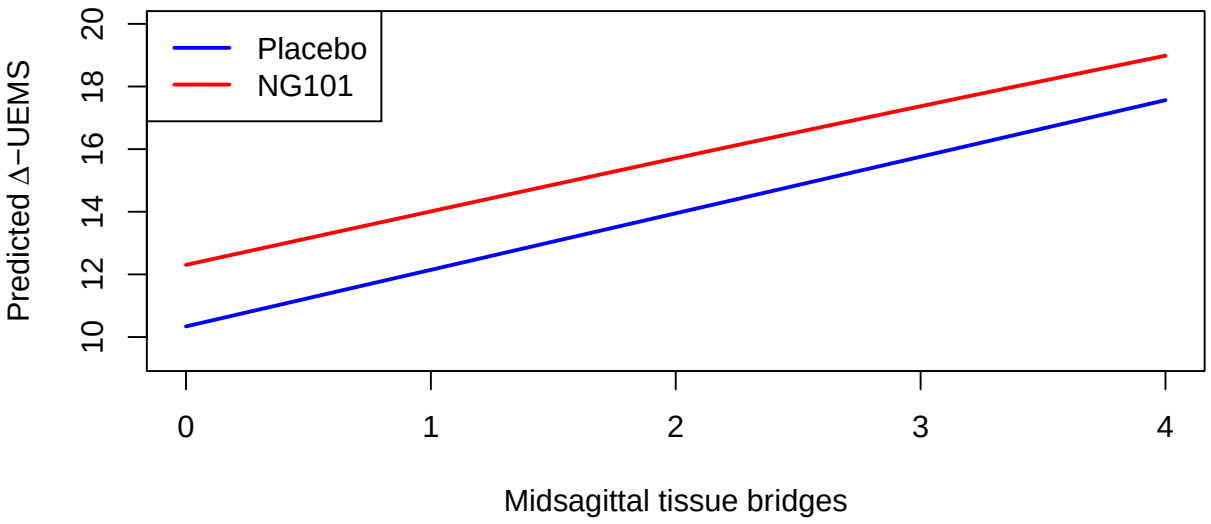

Figure 10: Differences of the predicted 6-month follow-up UEMS scores and the predicted baseline scores for different values of misagittal tissue bridges.

Parasagittal Tissue Bridges

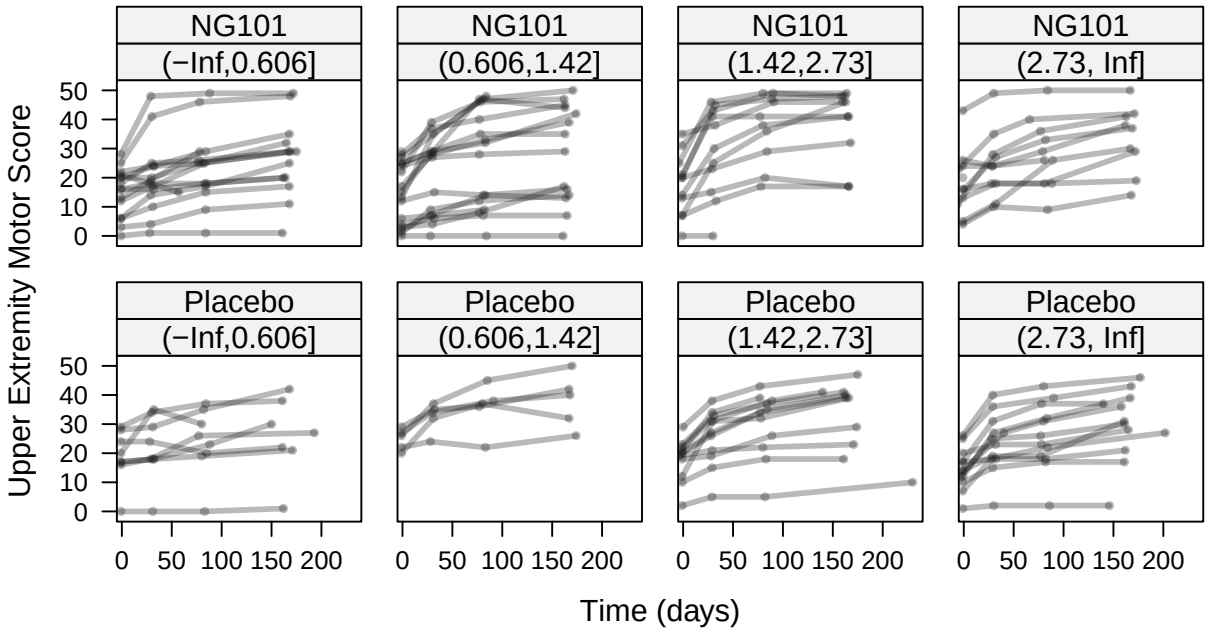

Figure 11: Patient data per treatment group separated in the four quartiles for the amount of parasagittal tissue bridges. The quartiles are based on the overall distribution of parasagittal tissue bridges.

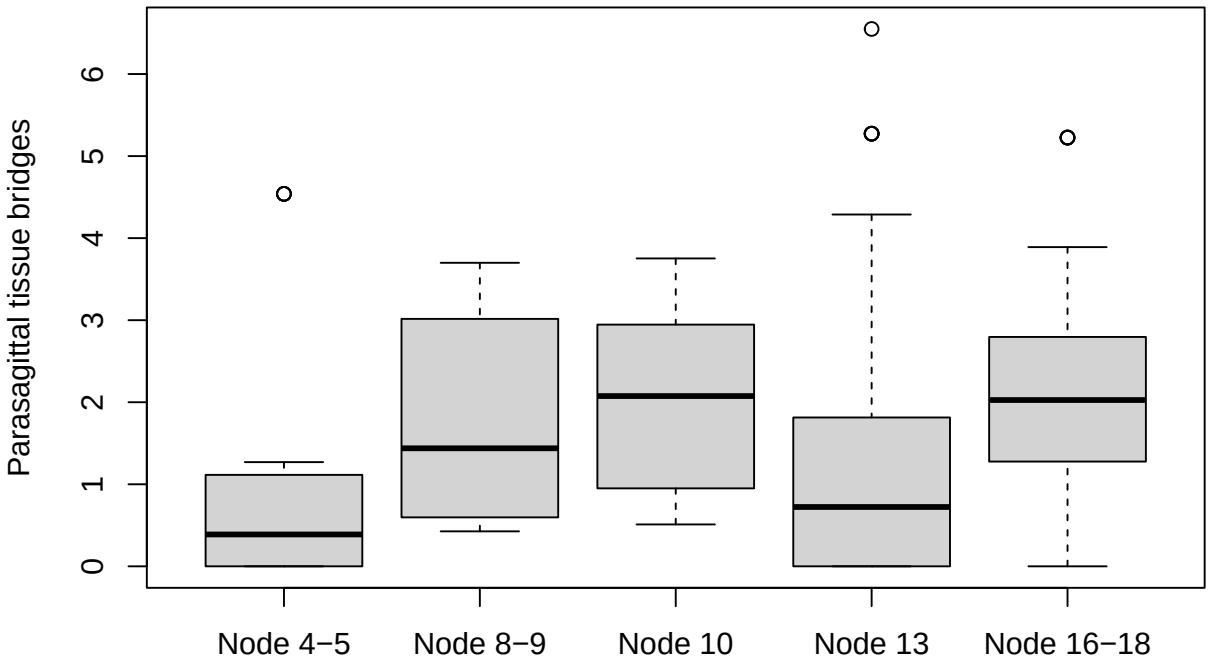

Figure 12: Boxplots of the parasagittal tissue bridges by URP-CTREE nodes.

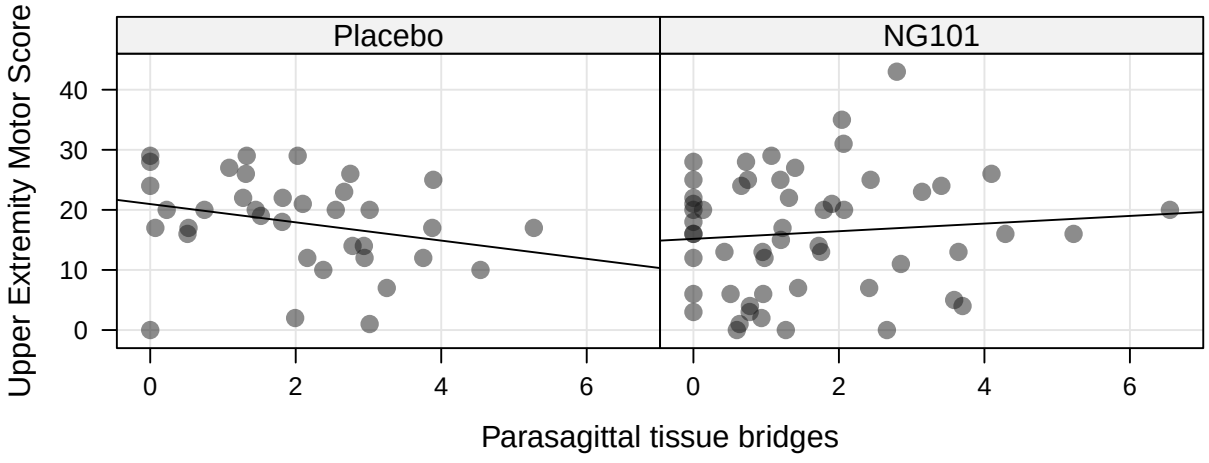

Figure 13: Scatterplot of the baseline parasagittal tissue bridges against Upper Extremity Motor Score in the two treatment groups.

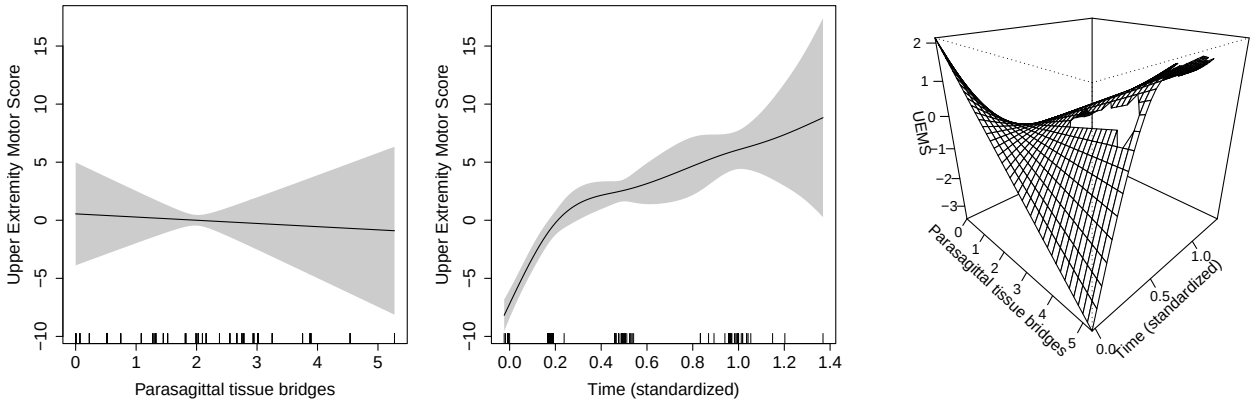

Figure 14: Smooth functions fitted by the GAM for the placebo group.

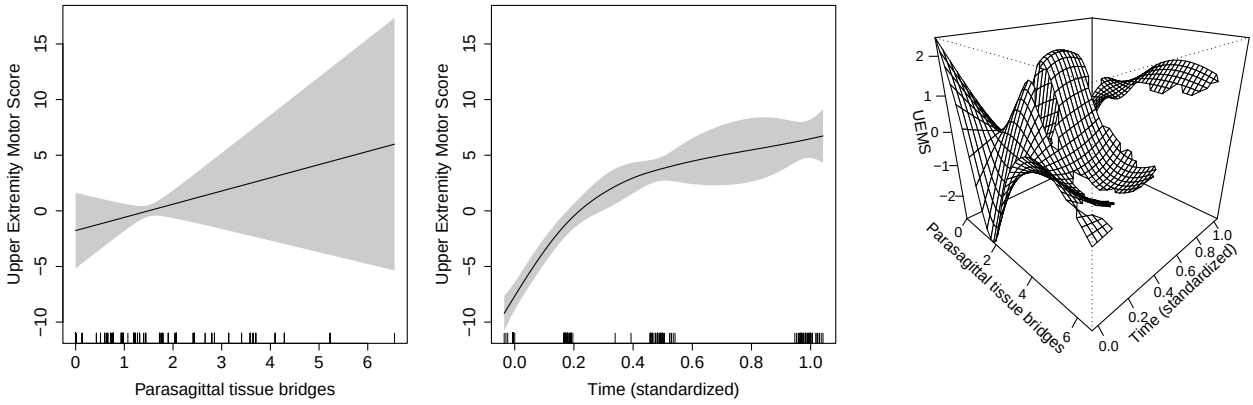

Figure 15: Smooth functions fitted by the GAM for the NG101 group.

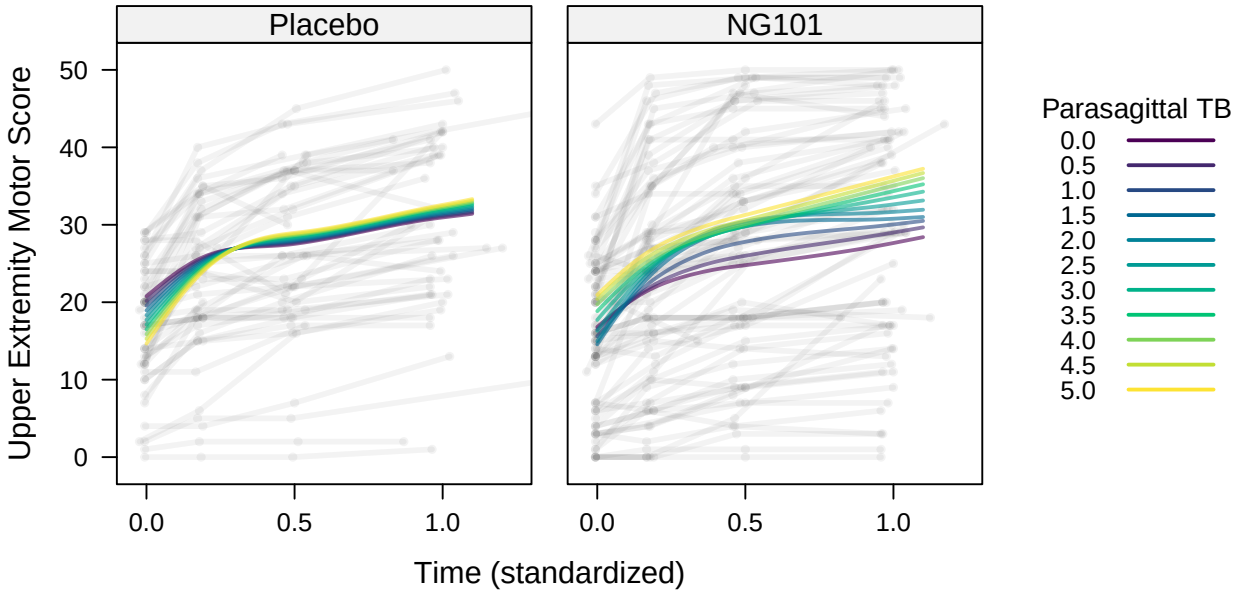

Figure 16: Predictions made with different values for parasagittal tissue bridges. For each treatment group the respective GAM was fitted separately. The only parameter that changes for the different predictions is the amount of parasagittal tissue bridges.

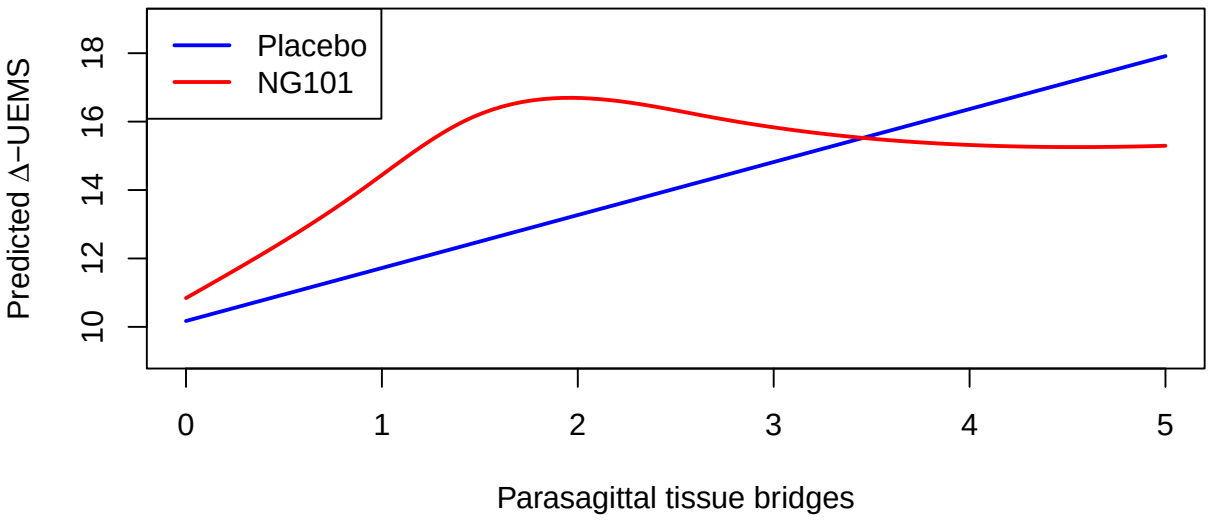

Figure 17: Differences of the predicted 6-month follow-up UEMS scores and the predicted baseline scores for different values of parasagittal tissue bridges.

Lesion Volume

Important: 6 patients with lesion volume > 1200 were excluded and set to NA. They are considered as outliers. Check with the study team!

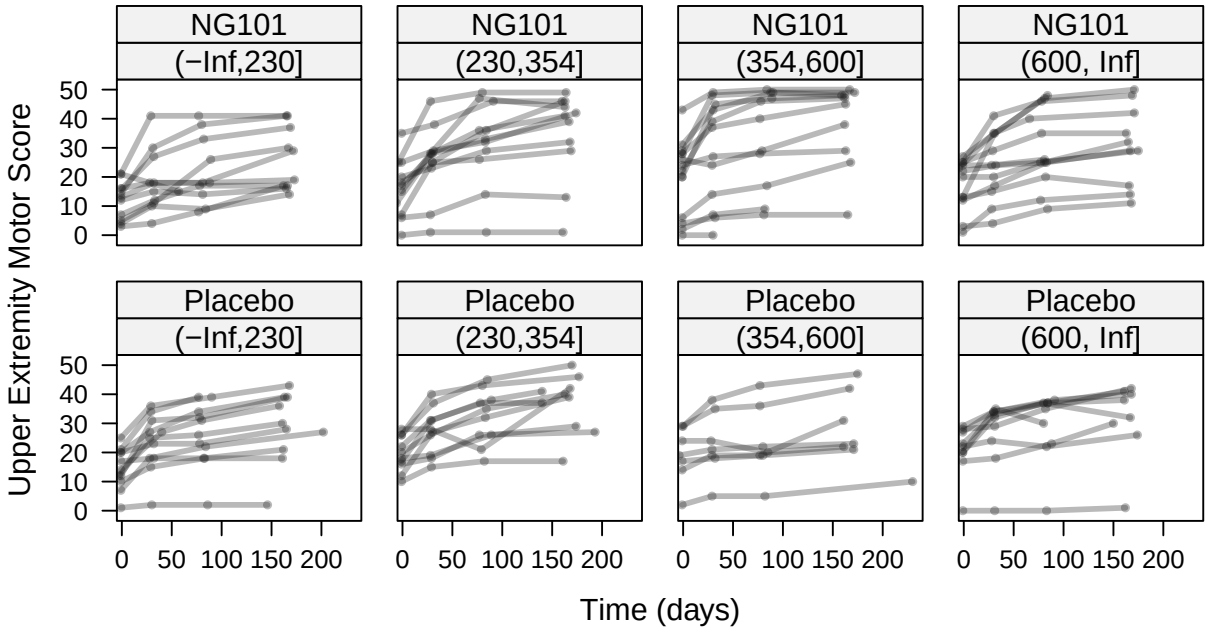

Figure 18: Patient data per treatment group separated in the four quartiles for the lesion volume. The quartiles are based on the overall distribution of lesion volume.

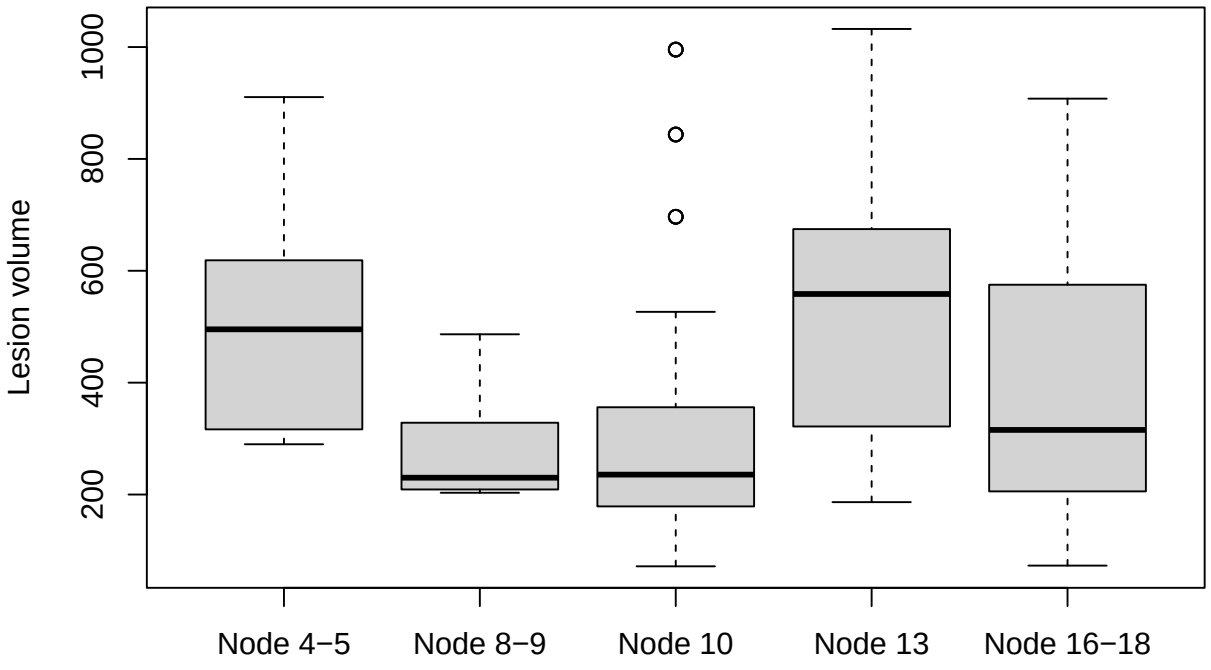

Figure 19: Boxplots of the lesion volume by URP-CTREE nodes.

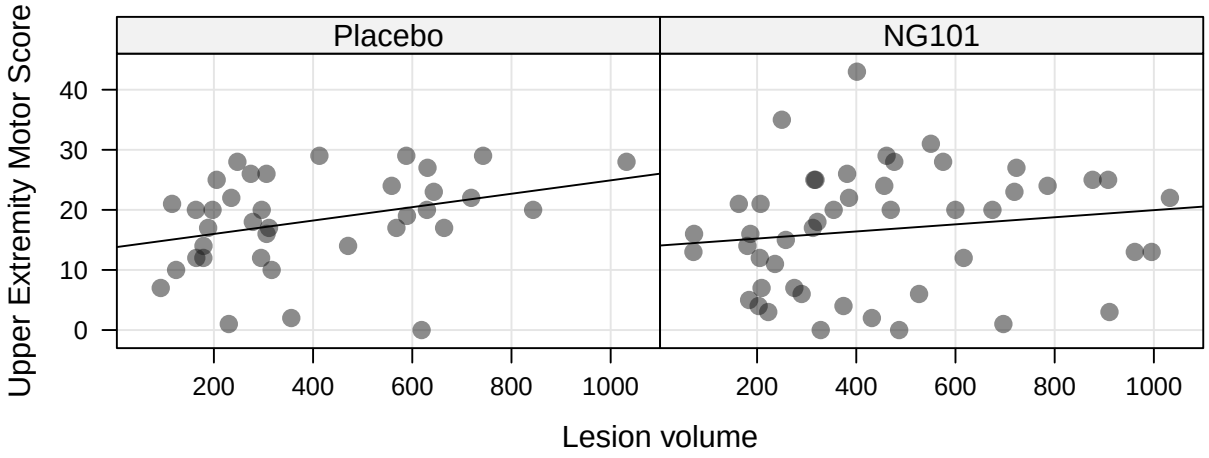

Figure 20: Scatterplot of the baseline lesion volume against Upper Extremity Motor Score in the two treatment groups. 6 patients were considered outliers and are not shown in this plot.

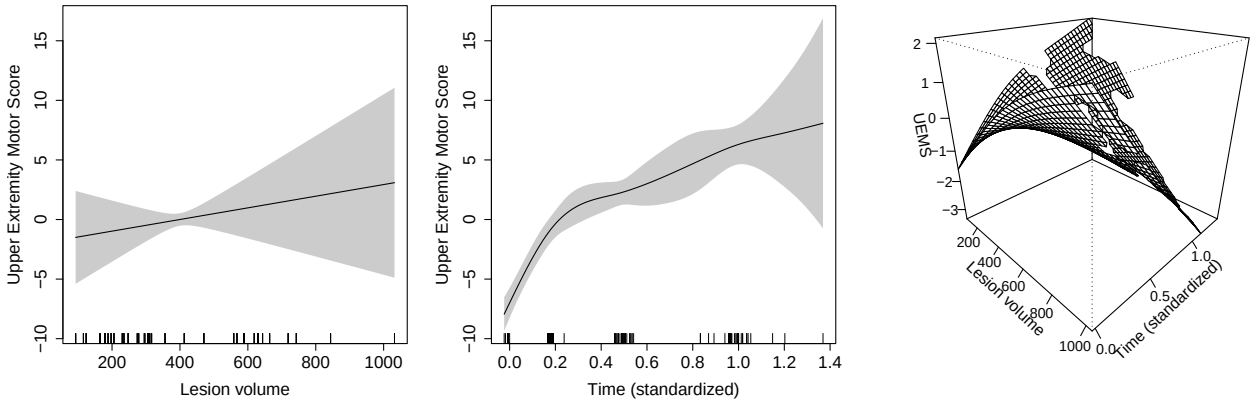

Figure 21: Smooth functions fitted by the GAM for the placebo group.

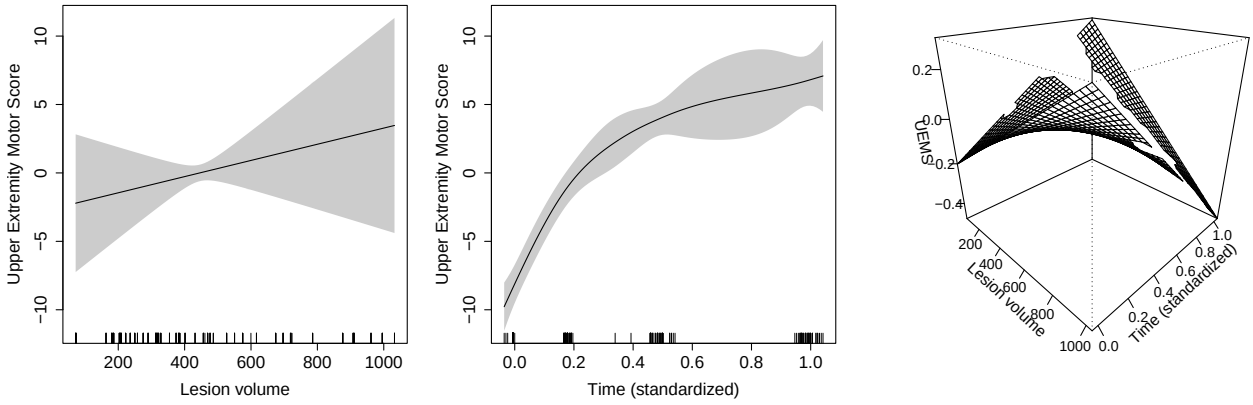

Figure 22: Smooth functions fitted by the GAM for the NG101 group.

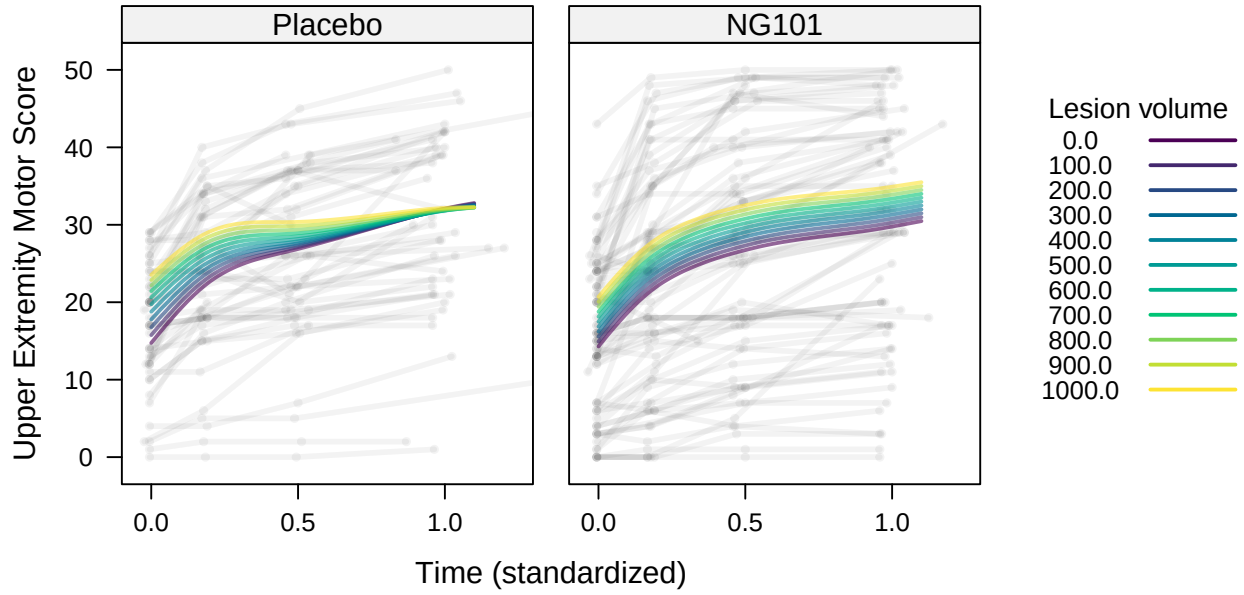

Figure 23: Predictions made with different values for lesion volume. For each treatment group the respective GAM was fitted separately. The only parameter that changes for the different predictions is the amount of lesion volume.

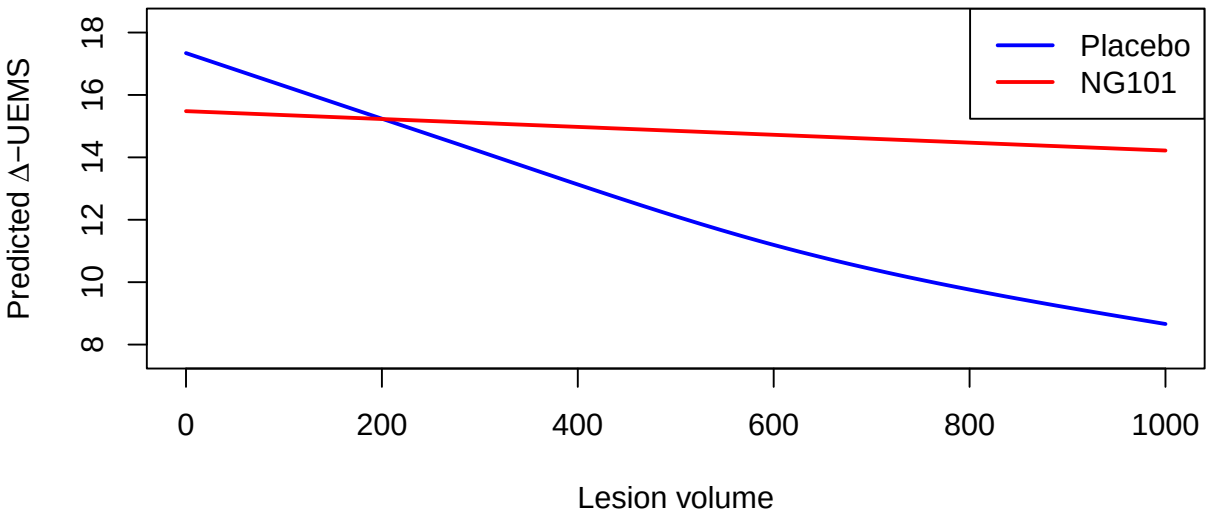

Figure 24: Differences of the predicted 6-month follow-up UEMS scores and the predicted baseline scores for different values of lesion volume.

5.1 Cut-Off Values

If we decide to find cut-off values, write it here.

6 Conclusion

Our findings regarding the association of the midsagittal tissue bridges with the recovery support the results of previous studies. More preserved midsagittal tissue bridges correspond to an improved recovery as seen in Figure 9. Figure 10 suggests that the relationship between the amont of midsagittal tissue bridges and the 6-month recovery is linear in both treatment groups. The baseline distribution of the midsagittal tissue bridges against the UEMS in the treatment groups is shown in Figure 6. It slightly differs among the treatment groups. In the placebo group, smaller tissue bridge values seem to be associated with higher UEMS values, while in the NG101 group, the opposite is the case.

The baseline distribution of the parasagittal tissue bridges against the UEMS in the treatment groups, as shown in Figure 13, is similar to the midsagittal tissue bridges. Also the recovery profile in the placebo group is comparable, as can be seen in Figure 17. However, in the NG101 group the recovery reaches a plateau after a certain value of the parasagittal tissue bridges (Remark: approx. 2mm, check where turning point). Why this is the case remains unclear.

The lesion volume in the placebo group shows a rather surprising baseline distribution as visualized in Figure 20. Patients with higher lesion volumes seem to have higher baseline UEMS values. This is contradictory to what one would expect. In the NG101 group, this effect is less extreme. The recovery profile seems to be not directly affected by the lesion volume. Figure 23 shows that independent of the lesion volume, all trajecotries converge to the same UEMS level. In the NG101 group, the Δ-UEMS is almost constant over the range of lesion volumes (Figure 24). In the placebo group, the Δ-UEMS decreases with increasing lesion volume. However, we assume this is due to the questionable baseline distribution.

To summarize, the midsagittal and parasagittal tissue bridges seem to be good predictors for the recovery. However, the association between spared tissue and 6-month recovery appears to be linear (with exception of the parasagittal tissue bridges in the NG101 group). Based on this data, the lesion volume does not seem to be a good predictor for the recovery.

Since we could not detect an obvious non-linear relationship of tissue bridges with recovery, a cut-off value for the tissue bridges is not meaningful.

(Remark: Discuss if we agree on this conclusion or nevertheless decide to find a cut-off value)

7 Generative AI Declaration

During the preparation of this report, I used Github Copilot and ChatGPT in order to be more efficient with coding. After using this tools/services, I reviewed and edited the content as needed and I take full responsibility for the content of the report.

References

Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science* **1** 297 – 310.
URL `https://doi.org/10.1214/ss/1177013604`

Huber, E., Lachappelle, P., Sutter, R., Curt, A. and Freund, P. (2017). Are midsagittal tissue bridges predictive of outcome after cervical spinal cord injury? *Annals of Neurology* **81** 740–748.
URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.24932`

NINDS (2024). Spinal cord injury. Accessed: 2024-11-18.
URL https://www.ninds.nih.gov/health-information/disorders/spinal-cord-injury

Pfyffer, D., Huber, E., Sutter, R., Curt, A. and Freund, P. (2019). Tissue Bridges Predict Recovery after Traumatic and Ischemic Thoracic Spinal Cord Injury. *Neurology* **93** 10.1212/WNL.0000000000008318.

Pfyffer, D., Smith, A. C., Weber, I., Kenneth A, Grillhoesl, A., Mach, O., Draganich, C., Berliner, J. C., Tefertiller, C., Leister, I., Maier, D., Schwab, J. M., Thompson, A., Curt, A. and Freund, P. (2024). Prognostic value of tissue bridges in cervical spinal cord injury: a longitudinal, multicentre, retrospective cohort study. *The Lancet Neurology* **23** 816–825.
URL https://doi.org/10.1016/S1474-4422(24)00173-X

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York.
URL http://lmdvr.r-forge.r-project.org

WHO (2024). Spinal cord injury: Fact sheet. Accessed: 2024-11-18.
URL https://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury

Wood, S. (2006). *Generalized Additive Models: An Introduction With R*, vol. 66.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Journal of the Royal Statistical Society (B)* **73** 3–36.

8 Appendix

8.1 Computational Details

This document was generated on Dezember 02, 2024 at 20:24. R version and packages used to generate this report:

- R version 4.4.1 (2024-06-14 ucrt), `x86_64-w64-mingw32`
- Locale: `LC_COLLATE=German_Germany.utf8`, `LC_CTYPE=German_Germany.utf8`, `LC_MONETARY=German_Germany.utf8`, `LC_NUMERIC=C`, `LC_TIME=German_Germany.utf8`
- Time zone: `Europe/Zurich`
- TZcode source: `internal`
- Running under: `Windows 11 x64 (build 22631)`
- Matrix products: default
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: biostatUZH 2.2.7, colorspace 2.1-1, dplyr 1.1.4, ggplot2 3.5.1, knitr 1.48, lattice 0.22-6, MASS 7.3-60.2, mgcv 1.9-1, naniar 1.1.0, nlme 3.1-164, RColorBrewer 1.1-3, readxl 1.4.3, stringr 1.5.1, survival 3.6-4, tableone 0.13.2, xtable 1.8-4
- Loaded via a namespace (and not attached): boot 1.3-30, cellranger 1.1.0, class 7.3-22, cli 3.6.3, cmprsk 2.2-12, codetools 0.2-20, compiler 4.4.1, data.table 1.16.2, DBI 1.2.3, digest 0.6.37, e1071 1.7-16, evaluate 1.0.0, fansi 1.0.6, farver 2.1.2, forcats 1.0.0, future 1.34.0, future.apply 1.11.2, generics 0.1.3, globals 0.16.3, glue 1.7.0, grid 4.4.1, gridExtra 2.3, gtable 0.3.5, haven 2.5.4, highr 0.11, hms 1.1.3, labeling 0.4.3, labelled 2.13.0, lava 1.8.0, lifecycle 1.0.4, listenv 0.9.1, lme4 1.1-35.5, magrittr 2.0.3, Matrix 1.7-0, minqa 1.2.8, mitools 2.4, munsell 0.5.1, nloptr 2.1.1, parallel 4.4.1, parallelly 1.38.0, pillar 1.9.0, pkgconfig 2.0.3, plyr 1.8.9, prodlim 2024.06.25, proxy 0.4-27, psy 1.2, R6 2.5.1, Rcpp 1.0.13, ReplicationSuccess 1.3.3, rlang 1.1.4, rstudioapi 0.17.1, scales 1.3.0, splines 4.4.1, stringi 1.8.4, survey 4.4-2, tibble 3.2.1, tidyselect 1.2.1, tools 4.4.1, UpSetR 1.4.0, utf8 1.2.4, vctrs 0.6.5, visdat 0.6.0, withr 3.0.1, xfun 0.47

8.2 Code

Please provide ALL code with which you produced this report by code chunk reuse, i.e. name all your chunks and display them here by typing «`chunkname`». You can also display code from external scripts, see the example for data preparation below. Try to format your code such that it fits in the lines. include comments to indicate which section the chunks were used in.

```
##############################################################################
# code for packages, settings
##############################################################################
## Import external functions
## ----------------------------------------------------------------------------

## Packages
## ----------------------------------------------------------------------------
library(RColorBrewer) # colors for plots
library(tableone) # for Table 1 functions
library(xtable) # formatting tables and generating the tex code
library(biostatUZH) # EBPI-written package, if not installed, uncomment code below
#devtools::install\_github(repo = "felix-hof/biostatUZH")
library(ggplot2) # customizable plots
library(stringr) # to prettify tables
library(readxl) # to import data
library(lattice) # for xyplot
library(mgcv) # for GAMs
library(colorspace) # sequential colors
library(dplyr)
library(naniar) # for missing data visualization (gg_miss_upset)
# library(gtsummary) # for table creation
### include project-specific packages here as well (e.g., lme4 for linear mixed effects models)
### if possible do not load libraries in chunks further below or in scripts that you source
```

```
################################################
### only load libraries that you really use!####
################################################

## Additional settings
## ---------------------------------------------------------------------------
cols <- brewer.pal(3, "Set1")
options(width = 85, digits = 4, show.signif.stars = FALSE)
```

```
#############################################################################
# code for preparation: data import and joining
#############################################################################
## ---------------------------------------------------------------------------
## Import Data
## ---------------------------------------------------------------------------

# Import primary dataset
# (it already includes parasagittal tissue bridges)
source("../data/NISCI_data.R")


# Import midsagittal tissue bridges
tissue_mid <- read_excel("../data/lesion_parameters_LF.xlsx")
# Subset only first visit
tissue_mid <- subset(tissue_mid, ExamStage == "NISCI01")
# add total_bridges from tissue_mid to datc (where tissue_mid(NISCI_ID) == datc(id))
datc$tissue_mid <- tissue_mid$total_bridges[match(datc$id, tissue_mid$NISCI_ID)]


# Import lesion volume
lesion_volume <- read_excel("../data/Lesion_volume_LF.xlsx")
# rename Slice Thickness
names(lesion_volume)[4] <- "Slice_Thickness"
# Subset only first visit
lesion_volume <- subset(lesion_volume, scan == 1)
# According to Lynn Farner (Zoom 31.10.24), if Slice_Thickness is NA, the lesion volume
# should also be NA and not 0. Hence, these values are set to NA here:
lesion_volume$Volume2[is.na(lesion_volume$Slice_Thickness)] <- NA
# add Volume2 to datc (where lesion_volume(id) == datc(id))
datc$lesion_volume <- lesion_volume$Volume2[match(datc$id, lesion_volume$id)]

# amount of lesion_volume > 1200 (7 patients)
# sum(lesion_volume$Volume2 > 1200, na.rm = TRUE)

# amount of visit_id == 2 with lesion_volume > 1200 (joined 6 patients,
# one patient was not in the main dataset)
# sum(datc$lesion_volume > 1200 & datc$visit_id == 2, na.rm = TRUE)

# Set 2 patients to NA where lesion volume not correct in dataset:

# Email by Lynn Farner: "Ich haben die Daten Ã¼berprÃ¼ft und es stimmt leider nicht ganz. BSL-001 und BYH-025 soll
datc[datc$id =="BSL-001",]$lesion_volume <- NA
datc[datc$id =="BYH-025",]$lesion_volume <- NA

# store dataset including outliers
datc_with_outliers <- datc
```

```r
# remove outliers for lesion volume
# set lesion_volume to NA where lesion_volume > 1200
datc$lesion_volume[datc$lesion_volume > 1200] <- NA


# subset data for treatment and placebo
datc_placebo <- subset(datc, trtplot=="Placebo")
datc_treatment <- subset(datc, trtplot=="NG101")


################################################################################
# code for results: data collection
################################################################################

# mean time and sd for MRI timepoint (days after injury)
MRI_timepoint_mean <- mean(tissue_mid$time, na.rm = TRUE)
MRI_timepoint_sd <- sd(tissue_mid$time, na.rm = TRUE)


# to get the baseline values, we analyze the dataset including outliers

# subset only baseline
visit_2 <- subset(datc_with_outliers, visit_id == 2)

# Number of total patients (126)
total_patients <- length(unique(visit_2$id))

# total countries
total_countries <- length(unique(visit_2$country))

# total patients in treatment and placebo
total_treatment <- sum(visit_2$trtplot == "NG101")
total_placebo <- sum(visit_2$trtplot == "Placebo")

# age mean and sd
mean_age <- mean(visit_2$bl_age, na.rm = TRUE)
sd_age <- sd(visit_2$bl_age, na.rm = TRUE)

# male percentage
percentage_male <- table(visit_2$bl_sex)[1]/total_patients

# total patients with tissue_mid
total_tissue_mid <- sum(!is.na(visit_2$tissue_mid))

# table of number of records per patient (0, 1, 2, 3, 4)
number_UEMS_records <- table(table(datc$id))



# table 1

# subset only baseline measures
visit_2 <- subset(datc_with_outliers, visit_id == 2)

## Get variables names
#dput(names(visit_2))
```

```r
## Vector of variables to summarize
myVars <- c("medistart", "uems","tissue_mid", "SR", "lesion_volume")

# Create a TableOne object with an overall column
tab_one <- CreateTableOne(vars = myVars, strata = "trt", data = visit_2,
                          addOverall = TRUE, test = FALSE)

# Convert TableOne to data frame for LaTeX formatting
tab_one_df <- print(tab_one, printToggle = FALSE, noSpaces = TRUE)

rownames(tab_one_df) <- c("Patients", "Time from injury to 1st injectino (days)", "UEMS",
                          "Midsagittal tissue bridges", "Parasagittal tissue bridges", "Lesion volume")

colnames(tab_one_df) <- c("Overall", "Placebo", "NG101")

tab_one_xt <- xtable(tab_one_df, caption = "Baseline Characteristics by Treatment Group.
                     The measures baseline measurements for UEMS and the three biomarkers
                     are presented as mean (sd).",
                     label = "tab:table_one_results")

# Print xtable
print(tab_one_xt, include.rownames = TRUE, caption.placement = "top", type = "latex")



# Check for missingness across the three biomarker variables, then group and summarize

missing_summary <- visit_2 %>%
  mutate(
    tissue_mid_missing = is.na(tissue_mid),
    SR_missing = is.na(SR),
    lesion_volume_missing = is.na(lesion_volume)
  ) %>%
  group_by(
    tissue_mid_missing,
    SR_missing,
    lesion_volume_missing
  ) %>%
  summarize(
    patient_count = n(),
    .groups = "drop"
  )

tab3_xt <- xtable(missing_summary, caption = "Combination of missing values in the three
                  biomarkers under investigation.", label = "tab:table_missing")


# Print xtable
print(tab3_xt, caption.placement = "top", type = "latex", include.rownames=FALSE)



# create a plot of the missing value combinations (same as table above)

missing_set <- visit_2[, c("tissue_mid", "SR", "lesion_volume")]
colnames(missing_set) <- c("Midsagittal", "Parasagittal", "Lesion_volume")
```

```r
# Generate a missing data heatmap
gg_miss_upset <- gg_miss_upset(missing_set,
                               nsets = 5)

# Display the plot
gg_miss_upset



# Select individuals with missing parasagittal but recorded midsagittal tissue bridges
# show them in a table
# -> according to Lynn Farner, Patient BHY-002 indeed has a missing value for parasagittal tissue but a valid lesi

missing_SR <- visit_2 %>%
  filter(is.na(SR) & !is.na(tissue_mid)) %>%
  select(id, tissue_mid, SR, lesion_volume)

# Replace NA values with "Missing"
missing_SR[is.na(missing_SR)] <- "Missing"

colnames(missing_SR) <- c("Patient-ID", "Tissue midsagittal", "Tissue parasagittal", "Lesion volume")

tab_SR_missing <- xtable(missing_SR, caption = "8 Patients that have recorded midsagittal tissue bridge
                         values but no parasagittal tissue bridges. This should not be possible.
                         Check with clinicians.",
                         label = "tab:table_SR_missing")

# Print xtable
print(tab_SR_missing, include.rownames = FALSE,
      caption.placement = "top", type = "latex")



# mean time of UEMS measurements in the respective visit_id (in days after first dose)
mean_time <- datc %>%
  group_by(visit_id) %>%
  summarize(mean_time = mean(tm, na.rm = TRUE))


# Boxplots of the biomarkers at baseline

par(mfrow=c(1,3))
boxplot(visit_2$tissue_mid, xlab = "Midsagittal tissue bridges")
boxplot(visit_2$SR, xlab = "Parasagittal tissue bridges")
boxplot(lesion_volume$Volume2, xlab = "Lesion volume")

# amount of patients with lesion volume > 1200 (they are excluded from the analysis)
patients_lesion_1200 <- sum(datc_with_outliers[datc_with_outliers$visit_id ==2,]$lesion_volume >1200,
                            na.rm = TRUE)


##############################################################################
# code for results: relationship of biomarkers with recovery (code-chunk-reuse)
##############################################################################
# Define quartiles for the selected biomarker, independent of treatment groups and based on visit 2
quantiles <- quantile(subset(datc, visit_id == 2)[[biomarker]], prob = 1:3 / 4, na.rm = TRUE)
```

```r
# Categorize the biomarker based on quartiles
datc[[paste0(biomarker, "_c")]] <- cut(datc[[biomarker]], breaks = c(-Inf, quantiles, Inf),
                                       as.ordered = TRUE)

# Plot the raw data of recovery separately per treatment group and quartile of biomarker
xyplot(
    uems ~ tm | get(paste0(biomarker, "_c")) + trtplot, data = datc,
    group = id, xlim = c(-5, 240), type = "b", between = list(x = 1, y = 1),
    pch = 20, cex = 0.7, lwd = 3, col = rgb(0.1, 0.1, 0.1, 0.3),
    xlab = "Time (days)", ylab = "Upper Extremity Motor Score",
    scales = list(
         x = list(alternating = 1), # Show x-axis tick labels only at the bottom,
         y = list(alternating = 1), # Show y-axis tick labels only at the left
         tck = c(1, 0) # Remove top ticks
       )
    #,main = paste("UEMS over time by", biomarker_name, "quartiles and treatment group")
)

# Plot the baseline UEMS against the baseline biomarker value

xyplot(as.formula(paste("uems ~", biomarker, "|  trtplot")),
       data = subset(datc, visit_id == 2),
       group = id,
       type = "b",
       pch = 16,
       cex = 1.2,
       lwd = 2,
       col = rgb(.1, .1, .1, .5),
       xlab = biomarker_name_upper,
       ylab = "Upper Extremity Motor Score",
       #main = paste("Baseline", biomarker_name, "vs UEMS"),
       grid = TRUE,
                         scales = list(
         x = list(alternating = 1), # Show x-axis tick labels only at the bottom
         tck = c(1, 0) # Remove top ticks
       ),
                         panel = function(x, y, ...) {
         panel.xyplot(x, y, ...) # Plot the points
         panel.lmline(x, y, col = "black", lwd = 1)  # Add a linear regression line
       }
       )


# fit GAM for each treatment group separately

# Define the formula
gam_formula <- as.formula(paste("uems ~ s(", biomarker, ") + s(tmstd) + ti(", biomarker, ", tmstd) +
                                s(id, bs = 're') + s(id, tmstd, bs = 're')"))

# Fit the GAM
gam_placebo <- gam(gam_formula,
                   data = datc_placebo, method = "REML")

gam_treatment <- gam(gam_formula,
                     data = datc_treatment, method = "REML")
```

```r
# <<summary_table>>


# plot the smooth functions fitte dby the GAM for the placebo group

par(mfrow = c(1, 3))


plot(gam_placebo, scheme = 1, select = 1,
     seWithMean = TRUE, rug = TRUE,
     xlab = biomarker_name_upper, ylab = "Upper Extremity Motor Score",
     cex.lab = 1.5, cex.axis = 1.5)

plot(gam_placebo, scheme = 1, select = 2,
     seWithMean = TRUE, rug = TRUE,
     xlab = "Time (standardized)", ylab = "Upper Extremity Motor Score",
     cex.lab = 1.5, cex.axis = 1.5)

plot(gam_placebo, scheme = 1, select = 3,
     seWithMean = TRUE, rug = TRUE, main = "UEMS",
     xlab = biomarker_name_upper, ylab = "Time (standardized)", ticktype = "detailed",
     cex.lab = 1.45, cex.axis = 1.3, cex.main = 1.45, theta = 45,  # Rotate around the z-axis
     phi = 25)    # Adjust elevation angle (tilt the plot vertically)



# plot the smooth functions fitted by the GAM for the treatment group

par(mfrow = c(1, 3))

plot(gam_treatment, scheme = 1, select = 1,
     seWithMean = TRUE, rug = TRUE,
     xlab = biomarker_name_upper, ylab = "Upper Extremity Motor Score",
     cex.lab = 1.5, cex.axis = 1.5)

plot(gam_treatment, scheme = 1, select = 2,
     seWithMean = TRUE, rug = TRUE,
     xlab = "Time (standardized)", ylab = "Upper Extremity Motor Score",
     cex.lab = 1.5, cex.axis = 1.5,)

plot(gam_treatment, scheme = 1, select = 3,
     seWithMean = TRUE, rug = TRUE, main = "UEMS",
     xlab = biomarker_name_upper, ylab = "Time (standardized)", ticktype = "detailed",
     cex.lab = 1.45, cex.axis = 1.3,
     cex.main = 1.45, theta = 45,  # Rotate around the z-axis
     phi = 25)    # Adjust elevation angle (tilt the plot vertically)



# predict recovery trajectories based on different values of the biomarker

models <- list(gam_placebo, gam_treatment)

# Prediction function using biomarker variable
make_prediction <- function(biomarker_value, model) {
  # new_data includes time at baseline up to 6 months
  # (used standardized time = 1.1 just for visualization)
```

```r
  new_data <- expand.grid(id = 1, tmstd = seq(0, 1.1, length = 100))
  # curent biomarker value is added to the new data
  # current "biomarker" is defined before the code chunk
  new_data[[biomarker]] <- biomarker_value

  # Make predictions excluding random effects
  prediction <- data.frame(
    tmstd = new_data$tmstd,
    pred = predict(model, newdata = new_data, exclude = "s(id)")
  )
  return(prediction)
}

# Set up colors for plotting
colors_plot <- sequential_hcl(length(biomarker_values), palette = "Viridis")

# Legend key setup
# legend_key <- list(
#   text = list(labels = paste0(format(biomarker_values, nsmall = 1))),  # Round to 1 decimal place
#   lines = list(col = colors_plot, lwd = 2),  # Color corresponding to each biomarker value
#   columns = 4
# )

legend_key <- list(
  text = list(labels = paste0(format(biomarker_values, nsmall = 1))), # Add descriptive labels
  lines = list(col = colors_plot, lwd = 2),  # Corresponding colors and line widths
  space = "right",  # Place the legend to the right of the plot
  columns = 1,  # Single column layout for vertical alignment
  title = biomarker_name_legend,  # Clear and concise title
  cex.title = 0.9,  # Slightly smaller title for compactness
  cex = 0.8  # Slightly smaller text for the labels
)

# Plot the raw recovery data per treatment group and
# add the predictions for different values of the biomarker
xyplot(uems ~ tmstd | trtplot,
       data = datc,
       group = id,
       xlim = c(-0.1, 1.3),
       type = "b",
       between = list(x = 1, y = 1),
       pch = 20,
       cex = .7,
       lwd = 3,
       col = rgb(.1, .1, .1, alpha = 0.05),  # Fade primary data lines
       xlab = "Time (standardized)",
       ylab = "Upper Extremity Motor Score",
       #main =  paste("Predictions for different", biomarker_name, "values"),
       key = legend_key,
       scales = list(
         x = list(alternating = 1), # Show x-axis tick labels only at the bottom
         tck = c(1, 0) # Remove top ticks
       ),
       # add predictions
       panel = function(x, y, subscripts, ...) {
         panel.xyplot(x, y, subscripts = subscripts, ...)
         # panel_index indicating treatment group (1 = placebo, 2 = NG101)
```

```
        panel_index <- panel.number()

        # for each biomarker value, make predictions and add to plot
        for (i in seq_along(biomarker_values)) {
            biomarker_value <- biomarker_values[i]
            # Generate predictions for each treatment group and biomarker value
            preds <- make_prediction(biomarker_value = biomarker_value, model = models[[panel_index]])
            panel.lines(preds$tmstd, preds$pred, col = colors_plot[i], alpha = 0.6, lwd = 2)
        }
    })



# this chunk generates predictions for baseline UEMS and 6-month UEMS for different values
# of the biomarker and plots the delta-UEMS over the range of the biomarker

# number of predicted delta-UEMS
N <- 100

# predictions made along the sequence of min(biomarker_value) to max(biomarker_value) of the biomarker
prediction_sequence <- seq(from= min(biomarker_values), to= max(biomarker_values), length.out = N)

# Create new data for predictions using biomarker variable dynamically
#new_data <- expand.grid(id = 1, tmstd = c(0, 1), biomarker_value = biomarker_values)

#new_data <- data.frame(id = rep(1, 2 * length(biomarker_values)),
# tmstd = rep(c(0, 1), each = length(biomarker_values)))

# new_data for tmstd=0 and tmstd=1
new_data <- data.frame(id = rep(1, 2 * N), tmstd = rep(c(0,1), each = N))

# add biomarker value sequence to new data
new_data[[biomarker]] <- rep(prediction_sequence, 2)


# Generate predicted UEMS for placebo and treatment groups at tmstd=0 and tmstd=1
predictions <- data.frame(
    tmstd = new_data$tmstd,
    biomarker = new_data[[biomarker]],
    pred_placebo = predict(gam_placebo, newdata = new_data, exclude = "s(id)"),
    pred_treatment = predict(gam_treatment, newdata = new_data, exclude = "s(id)")
)


# Calculate the predicted delta change between tmstd = 1 and tmstd = 0 for each biomarker value
predicted_delta <- data.frame(
    biomarker_value = prediction_sequence,
    delta_placebo = predictions[predictions$tmstd == 1,]$pred_placebo
    - predictions[predictions$tmstd == 0,]$pred_placebo,
    delta_treatment = predictions[predictions$tmstd == 1,]$pred_treatment
    - predictions[predictions$tmstd == 0,]$pred_treatment
)


# Plot the predicted delta-UEMS for each biomarker value
plot(predicted_delta$biomarker_value, predicted_delta$delta_placebo, type = "l", col = "blue", lwd = 2,
     ylim = c(min(predicted_delta[,2:3])-1, max(predicted_delta[,2:3])+1),
```

```
     xlab = biomarker_name_upper,
     ylab = expression("Predicted " * Delta * "-UEMS")
     #, main = bquote("Predicted " * Delta * "-UEMS for different" ~ .(biomarker_name) ~ "values")
     )


# Add line for the treatment group
lines(predicted_delta$biomarker_value, predicted_delta$delta_treatment, col = "red", lwd = 2)

# Add legend
legend(legend_location, legend = c("Placebo", "NG101"), col = c("blue", "red"), lwd = 2)




################################################################################
# code for results: relationship of biomarkers with recovery (call of the code chunks)
################################################################################

# not explicitly showed here because the prepared code chunks would be displayed each time again.

# midsagittal
# <<quartile_plot_other_biomarker>>
# <<baseline_uems_plot_tissue_mid>>
# <<plot_gam_fit>>
# <<plot_gam_fit_NG101>>
# <<plot_gam_predictions>>
# <<plot_gam_delta>>

# parasagittal
# <<quartile_plot_para>>
# <<baseline_uems_plot_tissue_para>>
# <<plot_gam_fit_para>>
# <<plot_gam_fit_NG101_para>>
# <<plot_gam_predictions_para>>
# <<plot_gam_delta_para>>

# lesion volume
# <<quartile_plot_lesion>>
# <<baseline_uems_plot_tissue_lesion>>
# <<plot_gam_fit_lesion>>
# <<plot_gam_fit_NG101_lesion>>
# <<plot_gam_predictions_lesion>>
# <<plot_gam_delta_lesion>>
```