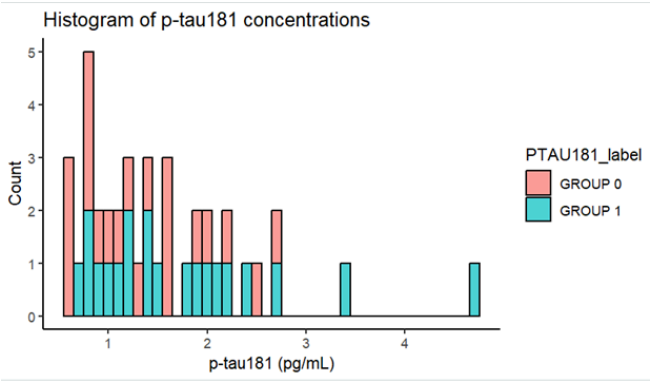
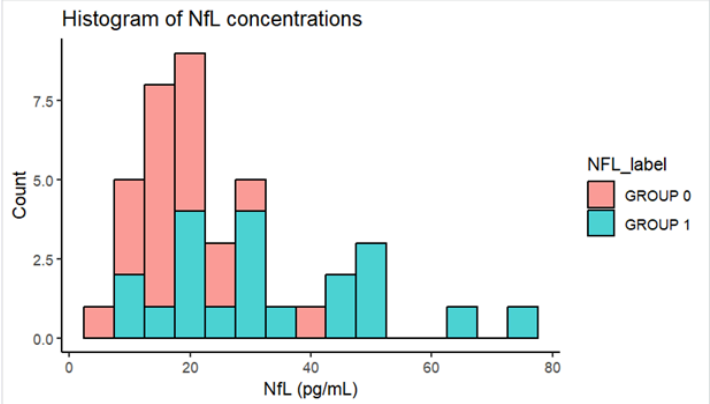
Task 1 Seminar 1 (Group A3)

Q1:

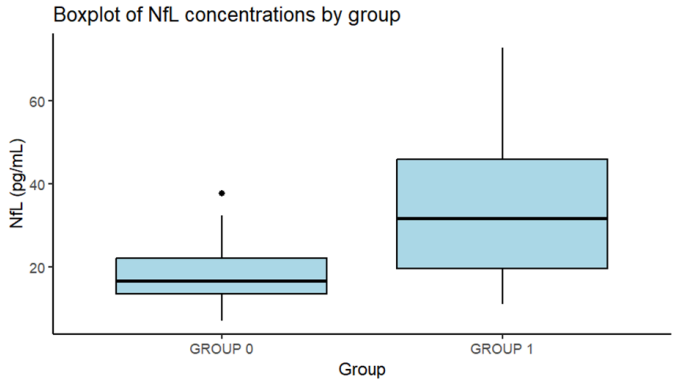
Explore the dataset visually and by carrying out appropriate statistical analyses.

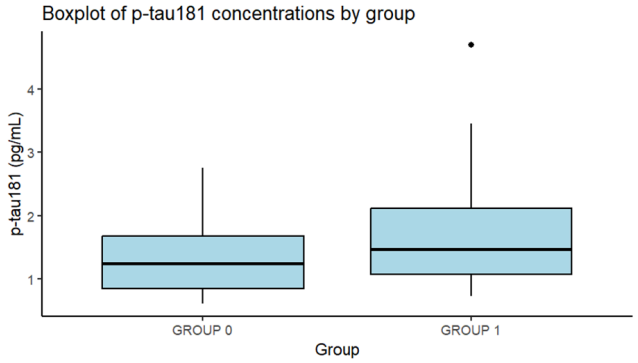
First we want to get a feel of the data and explore it visually. Since the sample size is low the visual representation is important because statistical tests lose power and p-values become less reliable.



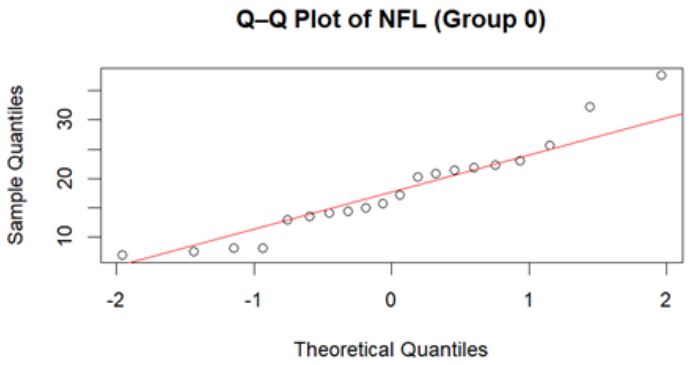


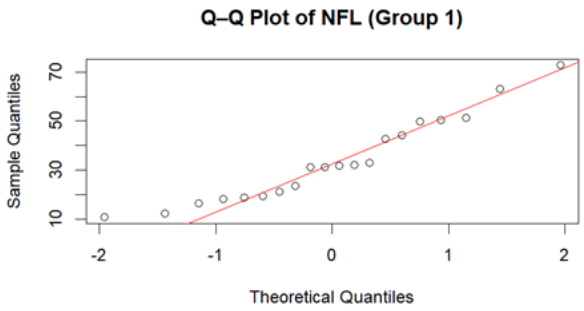
We first plot the biomarkers in a histogram. Here we see that the data looks right skewed and has some outliers.



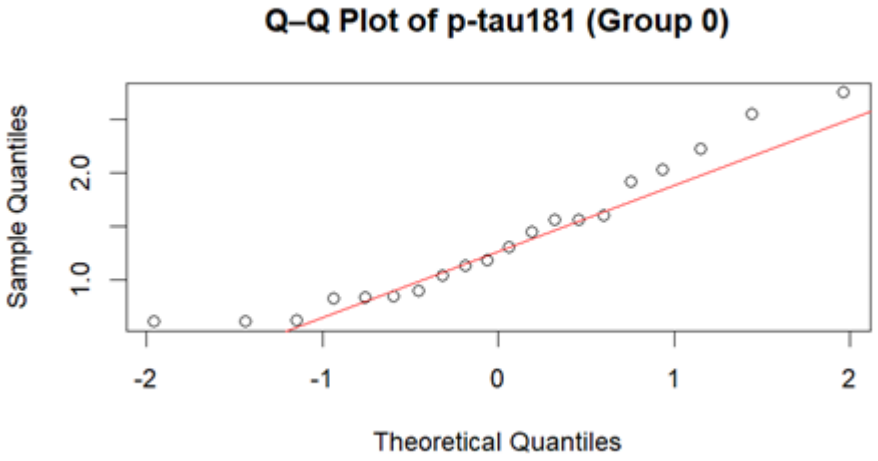


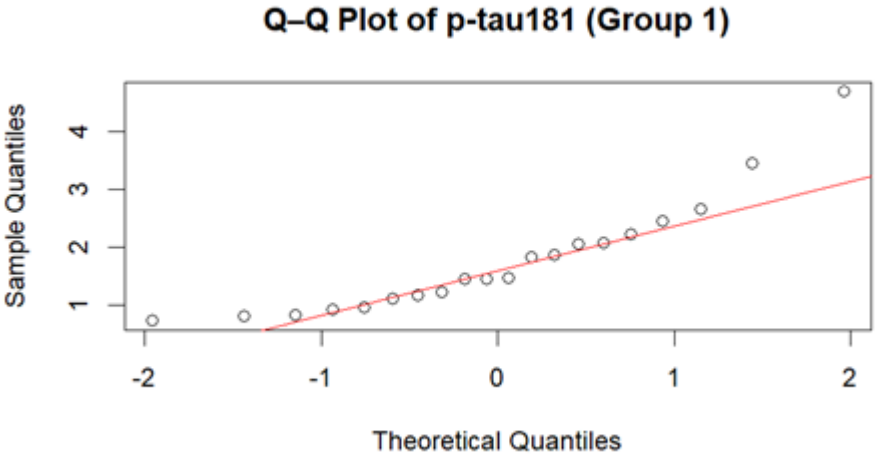
We then continue with plotting the data in boxplots. Which gives a representation of the data and how the values are distributed with the mean as a reference. This gives us an idea of the data. The p-tau181 groups look more asymmetric than NfL.





We go on to plot the data in QQ plots. We see that for the NfL data the data follows the line good, and we assume an approximate normality.





P-tau181 deviates substantially from the line, suggesting non-normality.

```
> shapiro_test_PTAU181_0 <- shapiro.test(df_plot[df_plot$GROUP == 0, ]$PTAU181)
> print(shapiro_test_PTAU181_0)

        Shapiro-Wilk normality test

data:  df_plot[df_plot$GROUP == 0, ]$PTAU181
W = 0.924, p-value = 0.1183


> shapiro_test_PTAU181_1 <- shapiro.test(df_plot[df_plot$GROUP == 1, ]$PTAU181)
> print(shapiro_test_PTAU181_1)

        Shapiro-Wilk normality test

data:  df_plot[df_plot$GROUP == 1, ]$PTAU181
W = 0.85896, p-value = 0.00756


> shapiro_test_NFL_0 <- shapiro.test(df_plot[df_plot$GROUP == 0, ]$NFL)
> print(shapiro_test_NFL_0)

        Shapiro-Wilk normality test

data:  df_plot[df_plot$GROUP == 0, ]$NFL
W = 0.93826, p-value = 0.2222


> shapiro_test_NFL_1 <- shapiro.test(df_plot[df_plot$GROUP == 1, ]$NFL)
> print(shapiro_test_NFL_1)

        Shapiro-Wilk normality test

data:  df_plot[df_plot$GROUP == 1, ]$NFL
W = 0.93541, p-value = 0.1961
```

To confirm, we performe the Shapiro–Wilk normality test:

- NfL (Group 0 & Group 1) P-tau181 (Group 0) : $p > 0.05$ → fail to reject normality assumption.

- P-tau181 (Group 1): $p < 0.05$ → reject normality.

Thus, NfL can reasonably be analyzed with a **t-test**, while P-tau181 requires either a non-parametric test (Mann–Whitney) or a log-transformation.

We have now looked at the data. NFL group 0, NFL group 1 are now assumed to be approximately normally distributed. Given the visualisation and the statistical test.
The p-tau181 biomarker has less of a normal distribution. Group 0 and Group 1 both follow the qqline poorly and Group 1 has a p-value lower than 0.05 in the shapiro-wilk normality test.

Moving on, we want to compare group 0 with group 1 for both biomarkers:

NFL analysis

For the NFL data we use the t-test:

```
> t_test_result <- t.test(df_plot[df_plot$GROUP == 0, ]$NFL, df_plot[df_plot$GROU
P == 1, ]$NFL)
> # Print the results
> print(t_test_result)

        Welch Two Sample t-test

data:  df_plot[df_plot$GROUP == 0, ]$NFL and df_plot[df_plot$GROUP == 1, ]$NFL
t = -3.721, df = 27.013, p-value = 0.0009209
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -24.553374  -7.099626
sample estimates:
mean of x mean of y
  17.9620   33.7885
```

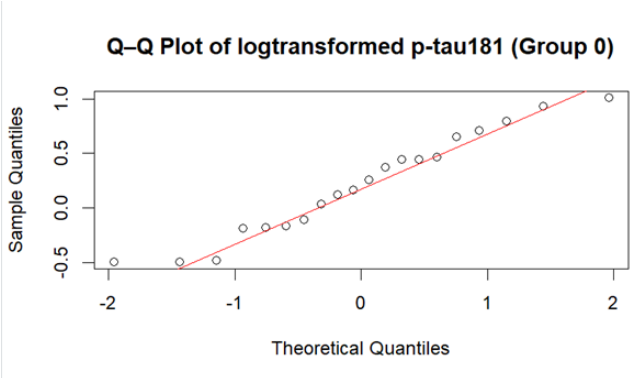
We get a p-value which is lower than 0.05 indicating that we can reject the null hypothesis. Meaning that the two groups have differing means.

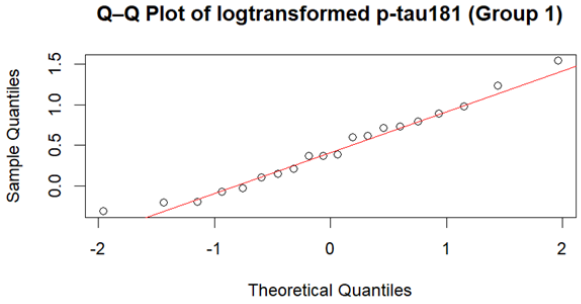
P-tau181 analysis

For the p-tau181 data we have two different things to do. Either we can use the willcoxon man whitney test, since it is non parametric meaning it does not assume a certain distribution. Or we can try to logtransform the data and see if it becomes approximately normally distributed making it possible to use the t-test.

Both tests have pros and cons. For this application we decide to use the t-test on the log data. We do this because the t-test is a more powerful test.

We start with logtransforming the data and do a qq-plot to confirm normality.





We can see that the data follows the line much better and we assume normality. For confirmation we do the shapiro test:

```
> shapiro.test(log_PTAU181_0)

        Shapiro-Wilk normality test

data:  log_PTAU181_0
W = 0.95544, p-value = 0.4573
```

```
> shapiro.test(log_PTAU181_1)

        Shapiro-Wilk normality test

data:  log_PTAU181_1
W = 0.97003, p-value = 0.7555
```

The p-values indicate that the data is normally distributed. We can now perform the t-test on the log data.

```
> t_test_result_PTAU181 <- t.test(log_data_PTAU181_0, log_PTAU181_1)
> # Print the results
> print(t_test_result_PTAU181)

        Welch Two Sample t-test

data:  log_data_PTAU181_0 and log_PTAU181_1
t = -1.4751, df = 37.808, p-value = 0.1485
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.54072530  0.08491306
sample estimates:
mean of x mean of y
0.2165314 0.4444375
```

The p-value is higher than 0.05 and we can not reject the null hypothesis. Indicating that the means of the data is similar and we can not distinguish between control and AD patients.

Q2:

What conclusions can you draw?

NfL:

- Two-sample t-test shows $p < 0.05$ → reject null hypothesis.
- Conclusion: mean NfL concentration is significantly higher in AD patients compared to controls.

P-tau181 (log-transformed):

- Log-transformation improved normality (confirmed by Q–Q plot and Shapiro–Wilk).
- Two-sample t-test on log(P-tau181): $p > 0.05$ → fail to reject null hypothesis.
- Conclusion: no significant difference in P-tau181 between groups, at least within this sample.

Q3:

How would you describe the statistical approach you are taking to solve this task? Discuss strengths and weaknesses of your approach.

Approach Taken

1. Exploratory visualization (histograms, boxplots, scatterplots, Q–Q plots).
2. Normality assessment (Shapiro–Wilk test, Q–Q plots).
3. Hypothesis testing:
    - Parametric t-test for NfL (normal).
    - Log-transform + t-test for P-tau181 (skewed).

Strengths

- Systematic workflow (EDA → normality check → hypothesis test).
- Transformation handled skewness and enabled use of a more powerful test.
- Results are straightforward to interpret for NfL.

Weaknesses

- Sample size is small (n=20/group), reducing statistical power and increasing risk of Type II error.
- Biomarkers considered separately, no combined model.
- Interpretability of log-transformed data: results need back-transformation to express differences on the original scale (e.g., as fold-change).
- A non-parametric approach (Mann–Whitney U) could have been a robustness check for P-tau181.

Conclusion

- NfL: significantly higher in AD patients, and thus shows promise as a biomarker.
- P-tau181: no significant group difference in this dataset. May require larger sample size, better measurement precision, or complementary biomarkers.

Task 2: Clinical Evaluation of MECAS-123

Group A3

2025-09-13

Background

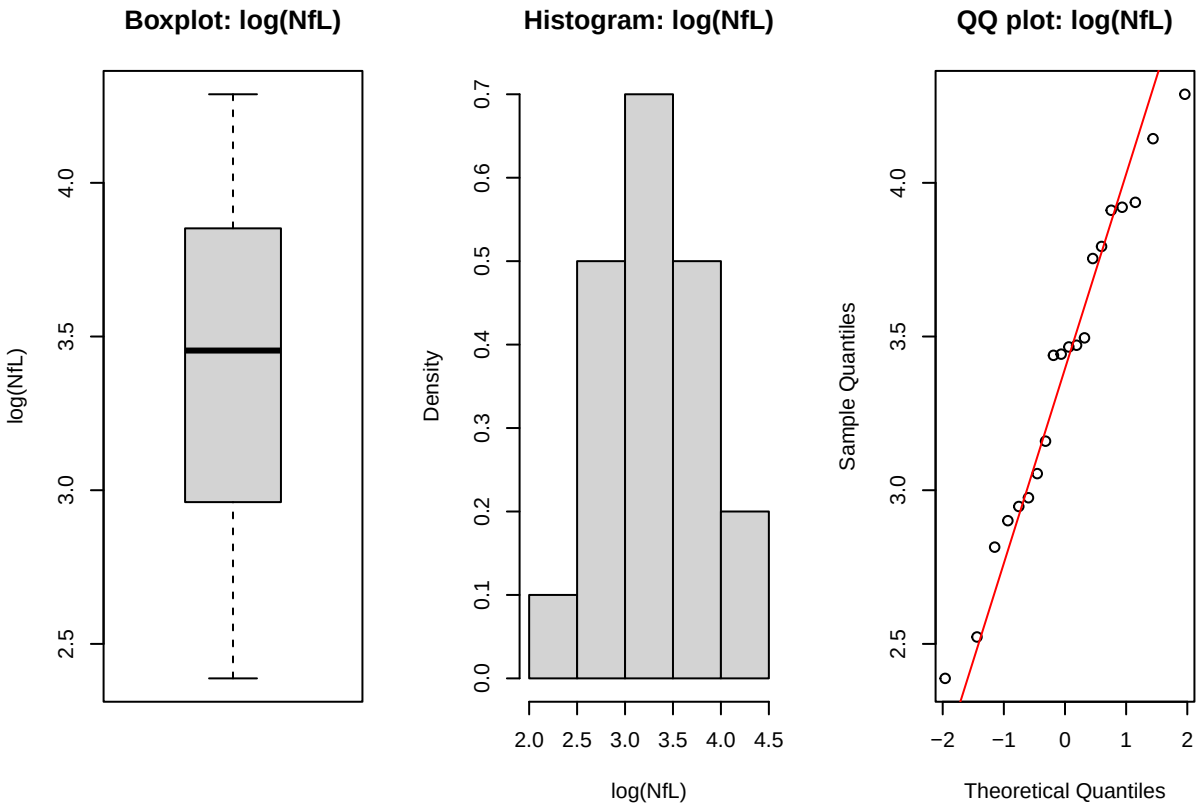
MECAS Pharma has developed a promising molecule, MECAS-123.
Preclinical studies have shown that MECAS-123 reduces the level of NfL, which could potentially slow down Alzheimer's disease progression.
A parallel-arm clinical study will be conducted in matched AD patients.
The objective is to evaluate whether MECAS-123 lowers NfL levels after 3 months compared to control.

Step 1: Load the AD dataset and power calculation

```
# Load raw dataset
Raw_1 <- read.csv("D:\\MSc\\Statistics\\Siminar1\\Data_T1.csv")

# Select AD patients only
AD_Raw <- subset(Raw_1, GROUP == 1)
NFL_AD <- AD_Raw$NFL
log_NFL_AD <- log(NFL_AD)

# Inspect first rows
head(AD_Raw)
```

```
##     X ID GROUP   NFL PTAU181
## 21 21  1     1 49.94    0.81
## 22 22  2     1 21.20    2.66
## 23 23  3     1 51.24    1.82
## 24 24  4     1 31.27    2.08
## 25 25  5     1 19.60    1.45
## 26 26  6     1 44.39    1.45
```

```
par(mfrow = c(1,3))
boxplot(log_NFL_AD, main = "Boxplot: log(NfL)",ylab = "log(NfL)")
hist(log_NFL_AD, prob = TRUE, main = "Histogram: log(NfL)", xlab = "log(NfL)")
qqnorm(log_NFL_AD, main = "QQ plot: log(NfL)"); qqline(log_NFL_AD, col="red")
```




```
shapiro_log <- shapiro.test(log_NFL_AD)
shapiro_log
```

```
##
##  Shapiro-Wilk normality test
##
## data:  log_NFL_AD
## W = 0.96817, p-value = 0.7158
```

The Shapiro test p-value is > 0.05, so log(NfL) is consistent with normal distribution. The boxplot shows approximate symmetric distribution. The histogram suggests approximate normality. The QQ plot shows points roughly along the 45-degree line.

Power Calculation

We calculate the required sample size using the following formula:

$$n = \frac{2 \cdot (z_{1-\alpha/2} + z_{1-\beta})^2 \cdot \sigma^2}{\Delta^2}$$

where:

- n = required sample size per group
- α = significance level (e.g. 0.05)

- $\beta = 1 - \text{Power}$ (e.g. 0.20 for 80% power)
- $z_{1-\alpha/2}$ = standard normal quantile at $1 - \alpha/2$
- $z_{1-\beta}$ = standard normal quantile at $1 - \beta$
- σ = standard deviation of log(NfL)
- $\Delta = \log(0.7)$ = expected treatment effect (30% reduction in geometric mean)

Thus, in our case:

$$\Delta = \log(0.7) \approx -0.357$$

```
# Define effect size on log scale
d_target <- abs(log(1 - 0.3))   # logtransfered for easier calculation of 30% reduction
d_target
```

```
## [1] 0.3566749
```

```
# Estimate standard deviation of log(NfL)
sigma_log <- sd(log_NFL_AD, na.rm = TRUE)
sigma_log
```

```
## [1] 0.5336258
```

```
# Load pwr package
if(!require(pwr)) install.packages("pwr")
```

```
## Loading required package: pwr
```

```
library(pwr)

# Parameters
alpha <- 0.05
power <- 0.80

# Compute required sample size per group
n <- pwr.t.test(d = d_target / sigma_log, sig.level = alpha, power = power, type = "two.sample")
n
```

```
##
##      Two-sample t test power calculation
##
##               n = 36.12298
##               d = 0.668399
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
# Round up to next integer
N <- ceiling(n$n)
N
```

```
## [1] 37
```

The required number of patients per arm to detect a 30% reduction in geometric mean of NfL with 80% power and alpha=0.05 is N.

Step 2: Statistical Analysis Plan (SAP) – NfL Analysis

1. Objective

To evaluate whether MECAS-123 reduces NfL levels in AD patients after 3 months compared to control.

2. Endpoint

Primary endpoint: NfL levels measured at 3 months post-treatment.

3. Study Groups

Treatment group: AD patients receiving MECAS-123 (`GROUP=1`)
Control group: AD patients receiving standard care or placebo (`GROUP=0`)

4. Statistical Analysis Strategy

Step 1: Assess Normality of Raw NfL Data

For each group, examine the distribution of raw NfL values.
Visual assessment: Histogram and Q-Q plot
Statistical test: Shapiro-Wilk test for normality

Step 2: Log Transformation

Apply natural log transformation if raw NfL is skewed: Reasons for log transformation: - Stabilizes variance and reduces skewness. - Effect is defined as a 30% reduction in geometric mean, which corresponds to arithmetic mean difference on log-scale: $\Delta = \log(0.7) = \log(0.7)$ - Simplifies interpretation and statistical testing.

Step 3: Confirm Normality of log(NfL)

Examine boxplot, histograms and Q-Q plots of log(NfL), perform Shapiro-Wilk test.

Step 4: Hypothesis Testing

Null hypothesis (H0): No difference in mean log(NfL) between Treatment and Control. Alternative hypothesis (H1): Mean log(NfL) in Treatment < Control. Test: Two-sample t-test (Welch's t-test if variances unequal) on log(NfL) Reasons for t-test on log(NfL): - Achieves approximate normality → t-test assumptions met - Parametric test → more efficient than Wilcoxon, higher power - Mean difference can be back-transformed (exp()) → geometric mean ratio (% reduction)

Step 3: Retrieve simulated data

```
if(!require(devtools)) install.packages("devtools")
```

```
## Loading required package: devtools
```

```
## Loading required package: usethis
```

```
library(devtools)
install_github("adamdarwichkth/CM2018rpackage")
```

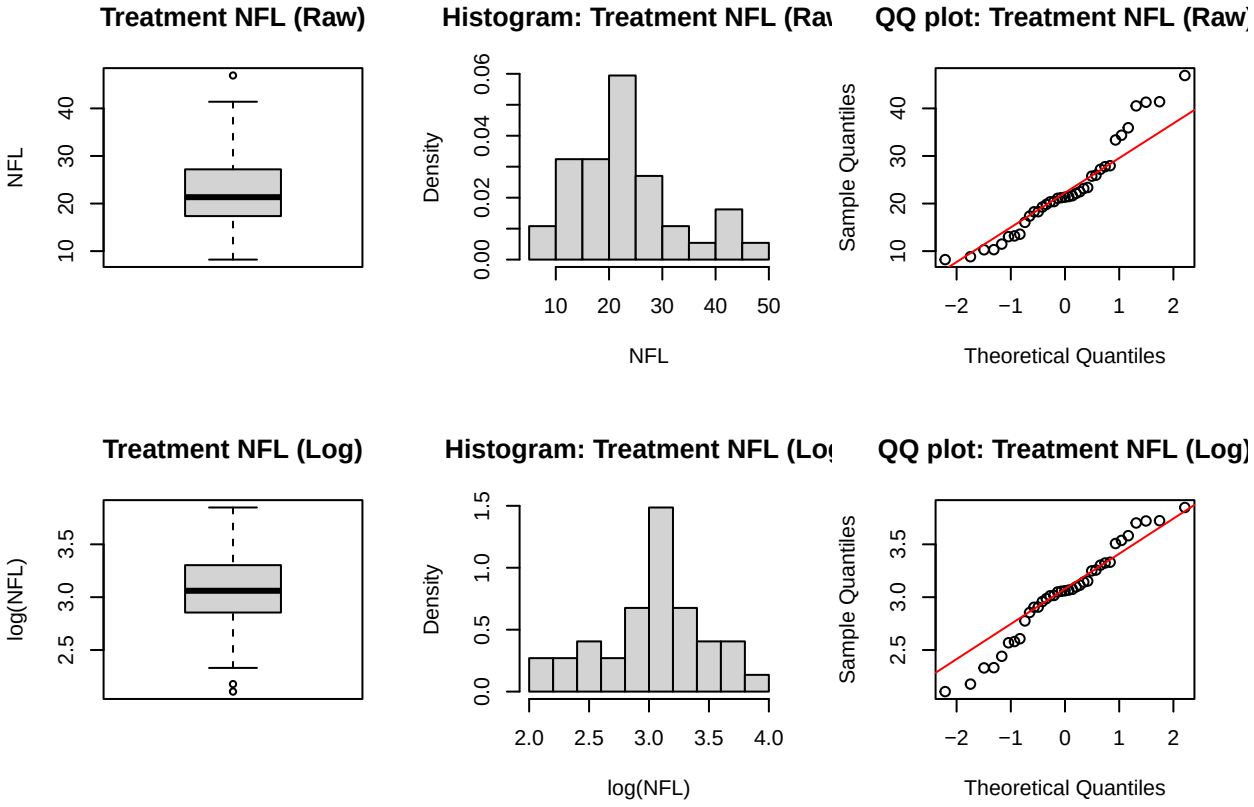
```
## Skipping install of 'CM2018rpackage' from a github remote, the SHA1 (c2bd9cd5) has not changed since
##   Use 'force = TRUE' to force installation
```

```
library(CM2018rpackage)

# Retrieve simulated AD trial data
my_dataframe <- ad_trial_data(n_per_arm = N)

# Split into treatment (GROUP==1) and control (GROUP==0)
AD_Treat <- subset(my_dataframe, GROUP == 1)
AD_Control <- subset(my_dataframe, GROUP == 0)

Treat_NFL <- AD_Treat$NFL
Control_NFL <- AD_Control$NFL

# Log-transform
log_Tr_NFL <- log(Treat_NFL)
log_Co_NFL <- log(Control_NFL)
```

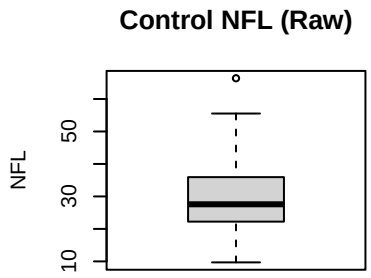
We now have two groups of simulated patients: treatment and control, and log-transformed NfL for parametric testing.

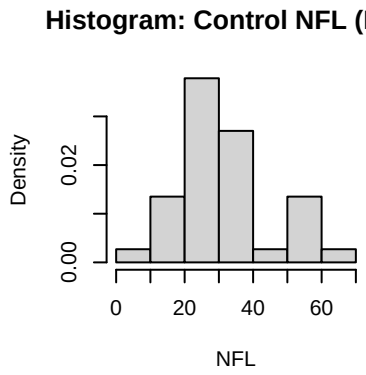
Step 4: carry out analysis

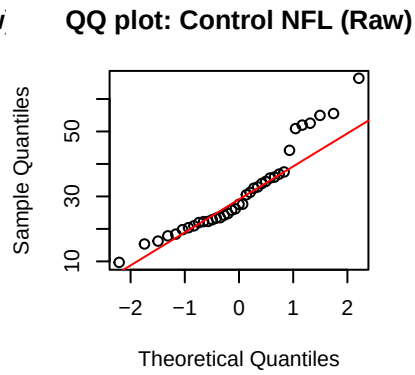
```r
# Exploratory plots and normality tests
par(mfrow=c(2,3))
# Treatment
boxplot(Treat_NFL, main = "Treatment NFL (Raw)",ylab = "NFL")
hist(Treat_NFL, prob=TRUE, main="Histogram: Treatment NFL (Raw)", xlab="NFL")
qqnorm(Treat_NFL, main="QQ plot: Treatment NFL (Raw)"); qqline(Treat_NFL, col="red")
boxplot(log_Tr_NFL, main = "Treatment NFL (Log)",ylab = "log(NFL)")
hist(log_Tr_NFL, prob=TRUE, main="Histogram: Treatment NFL (Log)", xlab="log(NFL)")
qqnorm(log_Tr_NFL, main="QQ plot: Treatment NFL (Log)"); qqline(log_Tr_NFL, col="red")
```

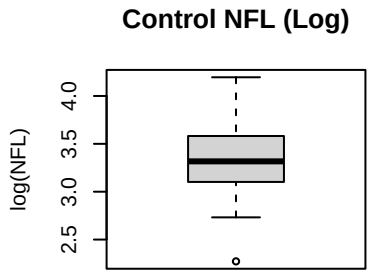


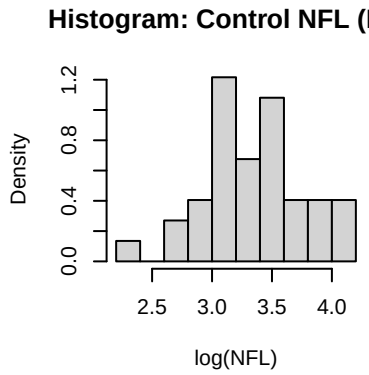

```r
# Control
boxplot(Control_NFL, main = "Control NFL (Raw)",ylab = "NFL")
hist(Control_NFL, prob=TRUE, main="Histogram: Control NFL (Raw)", xlab="NFL")
qqnorm(Control_NFL, main="QQ plot: Control NFL (Raw)"); qqline(Control_NFL, col="red")
boxplot(log_Co_NFL, main = "Control NFL (Log)",ylab = "log(NFL)")
hist(log_Co_NFL, prob=TRUE, main="Histogram: Control NFL (Log)", xlab="log(NFL)")
qqnorm(log_Co_NFL, main="QQ plot: Control NFL (Log)"); qqline(log_Co_NFL, col="red")
```

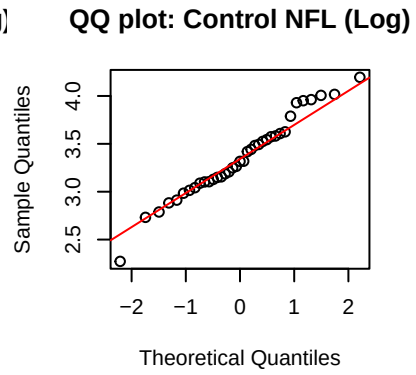













```
shapiro.test(Treat_NFL)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Treat_NFL
## W = 0.93778, p-value = 0.03923
```

```
shapiro.test(log_Tr_NFL)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  log_Tr_NFL
## W = 0.96826, p-value = 0.3632
```

```
shapiro.test(Control_NFL)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Control_NFL
## W = 0.91834, p-value = 0.009924
```

```r
shapiro.test(log_Co_NFL)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  log_Co_NFL
## W = 0.97843, p-value = 0.6771
```

Log-transformed NfL in both groups appears approximately normal: -The boxplots show that both groups become approximately symmetric. -The histograms of log-transformed NfL show a bell-shaped pattern. -The QQ plots demonstrate that the data points align much more closely with the reference line -Shapiro tests show yield non-significant results. This justifies using a t-test on log(NfL).

```r
# Two-sample t-test on log(NfL)
t_logNFL <- t.test(log_Tr_NFL, log_Co_NFL, var.equal = FALSE)
t_logNFL
```

```
##
##  Welch Two Sample t-test
##
## data:  log_Tr_NFL and log_Co_NFL
## t = -3.1586, df = 71.809, p-value = 0.00232
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5109854 -0.1155464
## sample estimates:
## mean of x mean of y
##  3.039260  3.352526
```

Conclusion

Data Distribution:

The raw NfL values showed moderate positive skewness, whereas log-transformed NfL (log(NfL)) approximated normal distribution, as confirmed by histograms, Q-Q plots, and Shapiro-Wilk tests ($p > 0.05$).

This justified the use of parametric methods on the log scale.

Power Calculation:

A 30% reduction in the geometric mean of NfL was set as the target effect.

Using the observed standard deviation of log(NfL) and a desired power of 0.80 with $\alpha = 0.05$, the required sample size per arm was calculated to be N participants.

This ensures sufficient statistical power to detect the predefined effect.

Analysis of Simulated Clinical Data:

The simulated dataset generated via the CM2018rpackage with n_per_arm = N allowed separation into Treatment and Control groups.

Log-transformed NfL was approximately normal in both groups.

A two-sample t-test on log(NfL) indicated a significant reduction in NfL levels in the Treatment group compared to Control ($p < 0.05$).

The results suggest that MECAS-123 has the potential to lower NfL levels in AD patients.

Using log-transformed data provides a clear and interpretable estimate of the geometric mean reduction, consistent with the study's effect definition.

The combination of power calculation, normality assessment, and appropriate statistical testing ensures robust and clinically meaningful conclusions.

Seminar-1-3-report

September 12, 2025

1. Propose a statistical analysis for multiple comparisons. We have data divided as 20 patients per study (total 80) as the sample size is low usage of non-parametertic techniques are to be preferred and multi group statistical tests are to be utilized.

We make the following assumption

- The data for each group are samples originate from the same distribution.
- We assume the data to be normally distributed.

2. Carry out the analysis using the attached data Carried out below

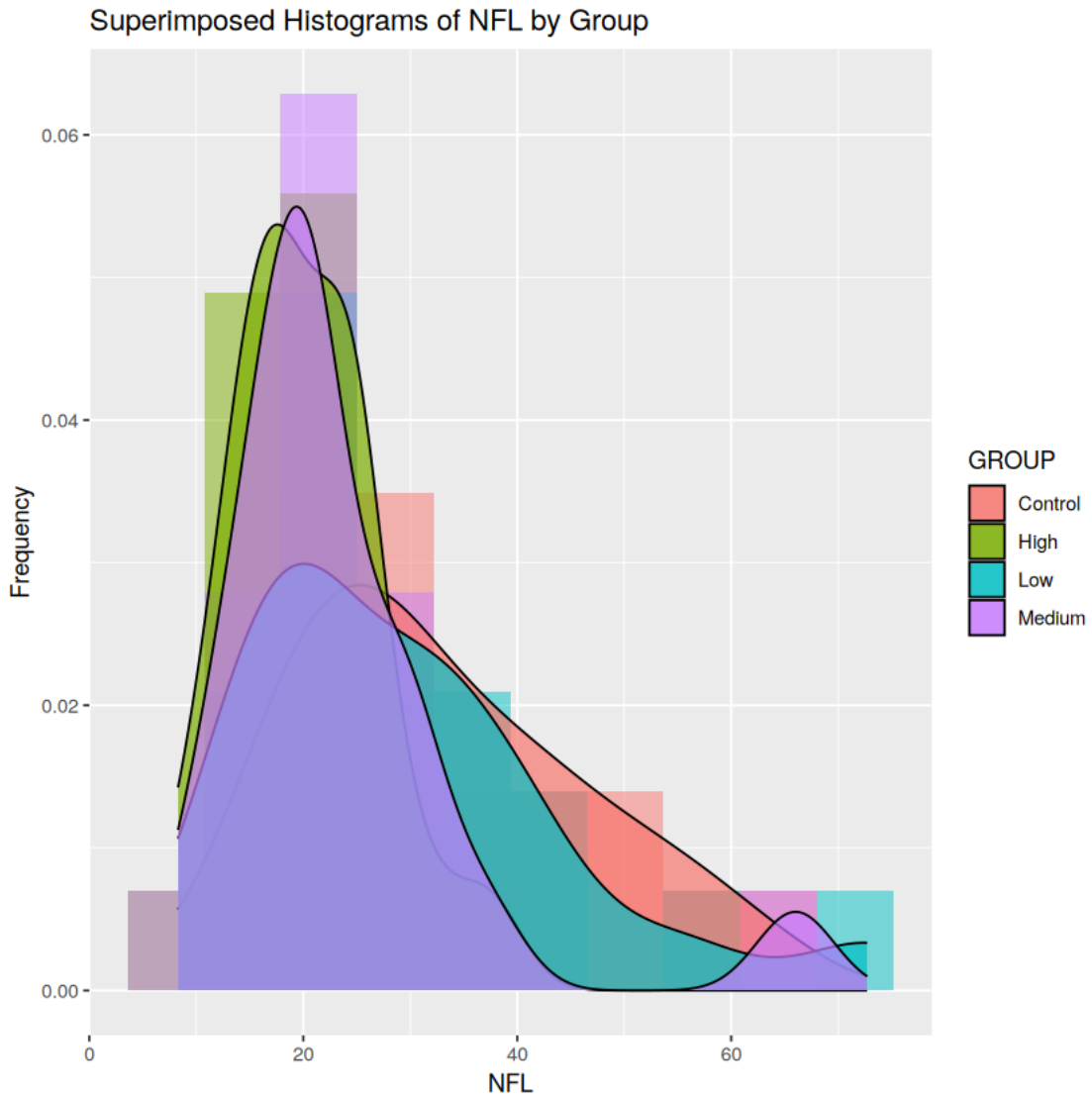
[13]:
```
# a first look at the data
head(data)
summary(data)
```

A data.frame: 6 × 4

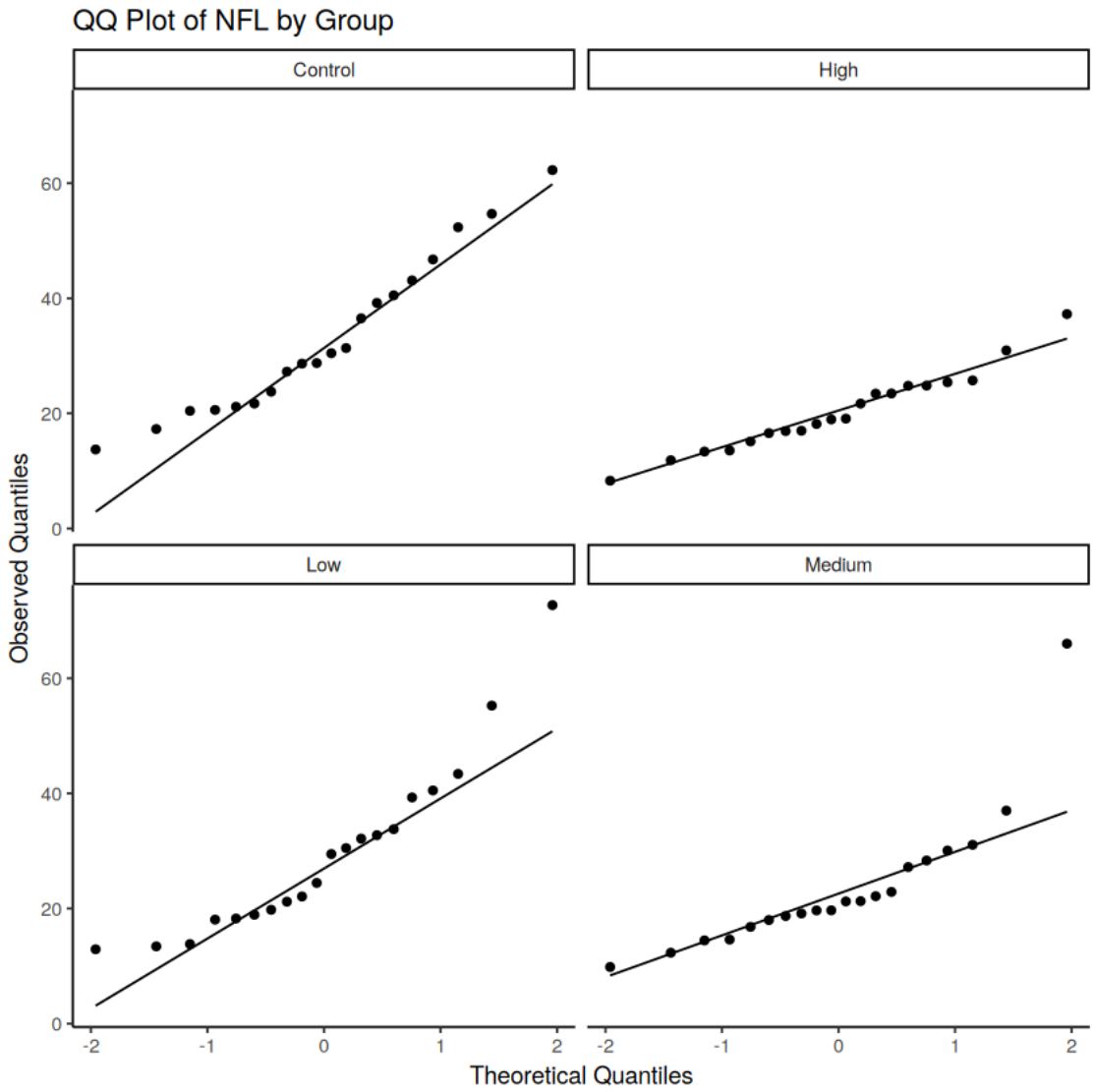
	X <int>	ID <int>	GROUP <chr>	NFL <dbl>
1	1	1	Control	39.195
2	2	2	Control	30.453
3	3	3	Control	46.740
4	4	4	Control	20.433
5	5	5	Control	21.704
6	6	6	Control	40.507

```
       X                ID           GROUP                NFL        
 Min.   : 1.00   Min.   : 1.00   Length:80          Min.   : 8.295  
 1st Qu.:20.75   1st Qu.:20.75   Class :character   1st Qu.:18.137  
 Median :40.50   Median :40.50   Mode  :character   Median :22.509  
 Mean   :40.50   Mean   :40.50                      Mean   :26.622  
 3rd Qu.:60.25   3rd Qu.:60.25                      3rd Qu.:31.131  
 Max.   :80.00   Max.   :80.00                      Max.   :72.712  
```

[ ]:
```
# Plot histograms for each group with overlayed density curves
ggplot(data, aes(x = NFL, fill = GROUP)) +
  geom_histogram(aes(y = ..density..), alpha = 0.5, position = "identity", bins␣
 ↪= 10) +
  geom_density(alpha = 0.7) +
  labs(title = "Superimposed Histograms of NFL by Group", x = "NFL", y =␣
 ↪"Frequency")
```



[9]:
```
ggplot(data, aes(sample = NFL)) +
  stat_qq() +
  stat_qq_line() +
  facet_wrap(~ GROUP) +
  labs(title = "QQ Plot of NFL by Group", x = "Theoretical Quantiles", y =␣
 ↪"Observed Quantiles") +
  theme_classic()
```



3. Plot the data, interpret and comment on the results. Here in the QQ Plot we can observe Control & High to follow the normal line quite well vice versa Low and Medium are diverging away from the normal line. These observations question our assumptions made about the data.

- 2/4 groups seem to be following normal distribution and other 2/4 are not.
- We can visually observe only 2/4 groups to be following the normal distribution.

To verify our observations we use Kruskal-Wallis rank sum test to see if they are sampled from the same distribution.

[10]:
```
kruskal.test(NFL ~ GROUP, data)
```

```
        Kruskal-Wallis rank sum test
```

```
data:  NFL by GROUP
Kruskal-Wallis chi-squared = 12.64, df = 3, p-value = 0.005483
```

Kruskal-Wallis rank sum test gives us the p-value of 0.005483 which is very less than our tolerance of 0.05, that would entail that the groups are sampled from different distributions i.e. both our assumptions are proven wrong.

- Data for each group have been sampled from different distributions.
- As our previous assumption is untrue that would also mean that not all the groups are normal. If they were p-value would be higher as they would have been sampled from normal distribution.

Now we will use Conover–Iman test to do pairwise comparision. It is based on Kruskal-Wallis test.

[11]:
```
library(conover.test)


conover.test(data$NFL, data$GROUP)
```

```
  Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 12.6402, df = 3, p-value = 0.01


                           Comparison of x by group
                                (No adjustment)
Col Mean-|
Row Mean |    Control       High        Low
---------+----------------------------------
    High |   3.429637
         |    0.0005*
         |
     Low |   1.099522  -2.330114
         |     0.1375    0.0112*
         |
  Medium |   2.723321  -0.706316   1.623798
         |    0.0040*     0.2411     0.0543

alpha = 0.05
Reject Ho if p <= alpha/2
```

The Conover-Iman test likewise preserves the ranks that the Kruskal-Wallis uses, and uses a pooled variance estimate to construct post hoc t test statistics.

From the p-values we can observe that there is significant p values for Control vs High group = 0.0005 & Medium vs Control = 0.004. There is also a very significant correlation between Medium vs Hight but is irrelevant to our testing.

Results 1. The groups are not distributed normally or sampled from same distribution. 2. The dosage *MECAS-123* is only significant for High and Medium dosage and shows an improvement in reduced NfL levels. + For High dosage 12.7037 pg/mL mean reduction. (Significant) + For Medium dosage 9.49985 pg/mL mean reduction. (Significant) + For Low dosage 3.38745 pg/mL mean reduction. (Insignificant)

Task 4: Clinical outcome scales of cognitive impairment

Group A3

September 13, 2025

Data Loading

Data4A includes 80 participants with 4 treatments and 4 responses (SPMSQ). Data3 includes the SPMSQ data of these 80 participants using different treatments. Data4B includes SPMSQ data of 25 participants from two different occasions.

```
'data.frame':        80 obs. of  4 variables:
 $ X        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ ID       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Treatment: int  0 0 0 0 0 0 0 0 0 0 ...
 $ Response : chr  "Mild" "Mild" "Moderate" "Mild" ...

    Mild Moderate   Normal   Severe     <NA>
      36       23        6       14        1
'data.frame':        50 obs. of  4 variables:
 $ X    : int  1 2 3 4 5 6 7 8 9 10 ...
 $ ID   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ OCC  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ SPMSQ: chr  "Mild" "Moderate" "Normal" "Mild" ...

    Mild Moderate   Normal   Severe
      19        7        8       16
```

```
spmsq_levels <- c("Normal", "Mild", "Moderate", "Severe")

data4A <- data4A %>%
  mutate(
    Treatment = factor(Treatment, levels = c(0,1,2,3),
                       labels = c("Control","Low","Medium","High")),
    Response = factor(Response, levels = spmsq_levels, ordered = TRUE)
  )

data4B <- data4B %>%
  mutate(
    SPMSQ = factor(SPMSQ, levels = spmsq_levels, ordered = TRUE)
  )
```

We first changed the numeric representation of the Treatment column to labels "Control", "Low", "Medium" and "High" for a clearer demonstration.

Q1. Visualization of the four study arms

We first used a stacked percentage bar plot to visualize the distribution of SPMSQ categories (Normal, Mild, Moderate, Severe) within each treatment arm. This type of plot displays the relative composition of each category as a proportion of the total in that arm, which is useful for directly comparing the percentage distribution of responses across groups regardless of group size. We additionally created a

grouped (side-by-side) bar plot showing the absolute counts of participants in each SPMSQ category per treatment arm. This complements the percentage plot by revealing the actual numbers of participants behind each percentage, which helps interpret differences when sample sizes vary.

```
# calculate the frequencies and percentage
spmsq_tab <- data4A %>%
  group_by(Treatment, Response) %>%
  summarise(n = n()) %>%
  ungroup() %>%
  group_by(Treatment) %>%
  mutate(freq = n / sum(n))

# stacked percentage bar chart
ggplot(spmsq_tab, aes(x = Treatment, y = freq, fill = Response)) +
  geom_bar(stat = "identity", position = "fill") +
  scale_y_continuous(labels = scales::percent) +
  labs(y = "Percentage", title = "SPMSQ distribution by Treatment (stacked %)",
       caption = "Responses: Normal -> Severe") +
  theme_minimal()

# grouped bar plot
ggplot(data4A, aes(x = factor(Treatment),
                   fill = Response)) +
  geom_bar(position = "dodge") +
  labs(x = "Treatment", y = "Count",
       fill = "SPMSQ",
       title = "SPMSQ categories by treatment (counts)")
```

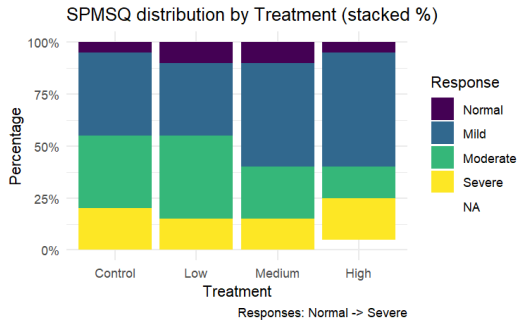


Figure 1: stacked percentage bar chart

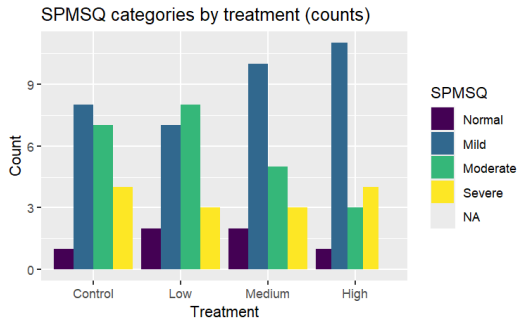


Figure 2: grouped bar plot

There is a missing value in the SPMSQ data of the High-dose group. Both plots show that the distribution of SPMSQ responses differs across treatment arms. For example, the High-dose arm appears to have a larger proportion of Mild impairment compared with the Control and Low arms, which show relatively more Moderate responses. However, because some categories have small counts, these visual differences should be interpreted cautiously and confirmed by statistical testing.

Q2. Visualizations of SPMSQ given the biomarker in Task 3

To explore the relationship between cognitive impairment and the biomarker NfL, we merged the SPMSQ data from Data4A with the NfL values from Data3 by participant ID. The boxplot was made with raw NfL data at first to show the original distribution. Since NfL concentrations showed a pronounced right-skewed distribution in histograms and failed the Shapiro–Wilk normality test, we applied a natural logarithm transformation before modelling. This transformation compresses very

high values and spreads out low values, making the distribution more symmetric and reducing the influence of outliers. It also allows the effect of NfL to be interpreted on a multiplicative ("relative change") scale, which is more meaningful for biological concentrations. We then plotted boxplots of log-transformed NfL across the four SPMSQ categories and across treatment arms. Boxplots display the median and interquartile range and highlight potential outliers.

```r
# merge data3 and data4A
merged_34A <- left_join(data4A,data3[,c("ID","NFL")],by="ID")

# boxplot of raw data
ggplot(merged_34A, aes(x = Response, y = NFL)) +
  geom_boxplot() +
  geom_jitter(width = 0.15, alpha = 0.6) +
  labs(title = "NfL by SPMSQ response (overall)",
       y = "NfL (pg/mL)") +
  theme_minimal()

# NFL normal or not
hist(data3$NFL, breaks=20, main="Histogram of NfL", xlab="NfL")
shapiro.test(data3$NFL)

# NFL isn't normal -> log-transform
merged_34A$logNFL <- log(merged_34A$NFL)

# boxplot with log(NFL)
ggplot(merged_34A, aes(x = Response, y = logNFL)) +
  geom_boxplot() +
  geom_jitter(width = 0.15, alpha = 0.6) +
  labs(title = "log(NfL) by SPMSQ response (overall)",
       y = "log(NfL) (pg/mL)") +
  theme_minimal()

# boxplot by different treatment groups
ggplot(merged_34A, aes(x = Response, y = logNFL)) +
  geom_boxplot() + facet_wrap(~Treatment) +
  labs(title = "log(NfL) by SPMSQ and Treatment", y = "log(NfL)") +
  theme_minimal()
```

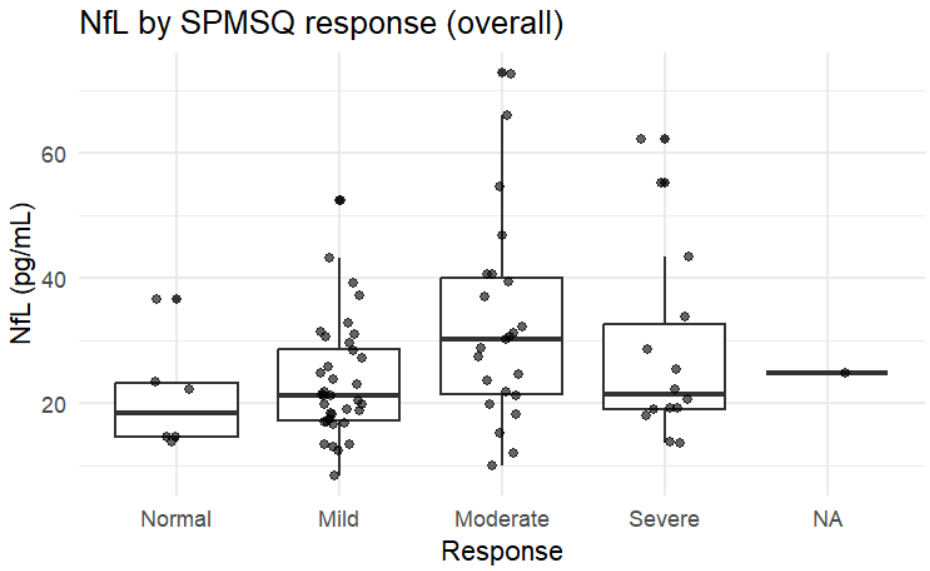


Figure 3: boxplot of NfL raw data

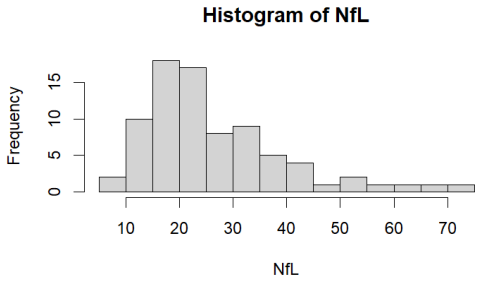


Figure 4: NfL histogram

```
        Shapiro-Wilk normality test

data:  data3$NFL
W = 0.87543, p-value = 1.287e-06
```

Figure 5: shapiro test of NfL

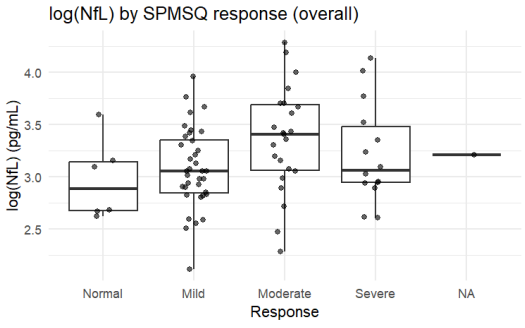


Figure 6: boxplot of log(NfL) data

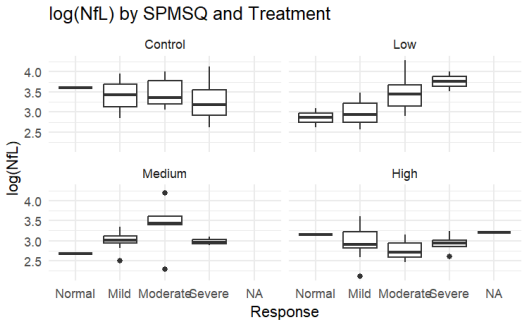


Figure 7: boxplot of log(NfL) by SPMSQ and Treatment

These plots show that in the more severe SPMSQ category, the log(NfL) values are relatively high, suggesting that increased NfL may be associated with worse cognitive status. To quantify this association, we fitted an ordinal logistic regression model(proportional-odds model) using the polr function. This model is appropriate because SPMSQ is an ordered categorical outcome (Normal ¡ Mild ¡ Moderate ¡ Severe). The model estimates the log-odds of being at or below each SPMSQ category as a linear function of predictors. A positive coefficient for log(NfL) means that higher NfL

increases the probability of belonging to a more severe category. We converted coefficients to odds ratios (OR = e^{β}) to aid interpretation.

```
merged_34A$Response <- factor(merged_34A$Response, levels = spmsq_levels, ordered = TRUE)

# logistic regression model
ord_model <- polr(Response ~ logNFL + Treatment, data = merged_34A, Hess = TRUE)
summary(ord_model)

# p-value of the coefficient
ctable <- coef(summary(ord_model))
pvals <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
cbind("p value" = pvals)
```

```
Call:
polr(formula = Response ~ logNFL + Treatment, data = merged_34A,
    Hess = TRUE)

Coefficients:
                   Value Std. Error  t value
logNFL           0.94216     0.5130  1.83639
TreatmentLow    -0.04143     0.5796 -0.07149
TreatmentMedium -0.28917     0.6087 -0.47504
TreatmentHigh    0.01323     0.6601  0.02004

Intercepts:
                Value   Std. Error t value
Normal|Mild      0.3397  1.7968     0.1891
Mild|Moderate    3.0748  1.8133     1.6957
Moderate|Severe  4.5436  1.8532     2.4517

Residual Deviance: 188.1605
AIC: 202.1605
(1)
                   p value
logNFL          0.06629935
TreatmentLow    0.94300917
TreatmentMedium 0.63475602
TreatmentHigh   0.98401055
Normal|Mild     0.85002854
Mild|Moderate   0.08994091
Moderate|Severe 0.01421740
```

In our analysis, the coefficient for log(NfL) was positive (0.942), corresponding to an OR $\approx$ 2.57 (exp(0.942) $\approx$ 2.57), meaning that for each unit increase in log(NfL) the odds of being in a more severe SPMSQ category roughly doubled. However, the p-value was 0.06, just above 0.05, indicating that while the direction is consistent with the hypothesis, the evidence is only suggestive and not conventionally statistically significant. This could reflect the relatively small sample size or the limited sensitivity of SPMSQ.

Q3. Visualise the SPMSQ for the same individuals at two occasions

Data4B allows us to examine intra-individual variability and test the stability of SPMSQ over time. We first created a transition matrix showing how many participants moved from each SPMSQ category at occasion 1 to each category at occasion 2. Such a matrix highlights patterns of improvement, worsening, or stability between time points. We visualised the matrix as a heatmap, where the diagonal cells

represent no change and off-diagonal cells indicate transitions. We also drew an individual connection diagram connecting each person's category at occasion 1 and occasion 2 to show the direction of change on an individual level.

```
# ID OCC1 OCC2
data4B_wide <- data4B %>%
  dplyr::select(ID, OCC, SPMSQ) %>%
  pivot_wider(names_from = OCC, values_from = SPMSQ, names_prefix = "OCC")

# transfer table
trans_table <- table(data4B_wide$OCC1, data4B_wide$OCC2)
trans_table

# transfer heat map
trans_df <- as.data.frame(as.table(trans_table))
colnames(trans_df) <- c("OCC1","OCC2","Count")
ggplot(trans_df, aes(x = OCC1, y = OCC2, fill = Count)) +
  geom_tile() +
  geom_text(aes(label = Count), color = "white") +
  labs(title = "Transition matrix: SPMSQ OCC1 -> OCC2") +
  theme_minimal()

# Individual connection diagram
level_map <- setNames(1:4, spmsq_levels)
data4B_wide <- data4B_wide %>%
  mutate(occ1_num = level_map[as.character(OCC1)],
         occ2_num = level_map[as.character(OCC2)])

data4B_long <- data4B_wide %>%
  dplyr::select(ID, occ1_num, occ2_num) %>%
  pivot_longer(cols = starts_with("occ"), names_to = "OCC", values_to = "Score") %>%
  mutate(OCC = ifelse(OCC=="occ1_num", "OCC1", "OCC2"))

ggplot(data4B_long, aes(x = OCC, y = Score, group = ID)) +
  geom_line(alpha = 0.6) +
  geom_point() +
  scale_y_continuous(breaks = 1:4, labels = spmsq_levels) +
  labs(title = "Individual SPMSQ changes between OCC1 and OCC2",
       y = "SPMSQ (ordered)") +
  theme_minimal()
```

	Normal	Mild	Moderate	Severe
Normal	2	1	0	0
Mild	3	5	1	2
Moderate	0	2	0	2
Severe	0	0	2	5

Figure 8: transition matrix

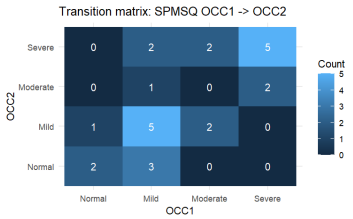


Figure 9: heatmap of transition matrix

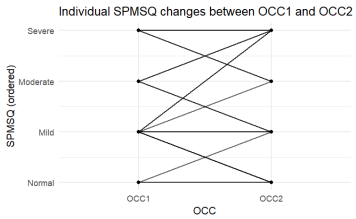


Figure 10: individual connection diagram

To formally assess agreement, we calculated a weighted Cohen's kappa coefficient. Kappa measures how well the two sets of ratings agree beyond chance; weighting accounts for the ordered nature of SPMSQ categories by giving partial credit to near-agreements. A kappa around 0.3–0.4 indicates fair agreement, while values above 0.6 indicate good agreement.

```r
# Prepare data: two columns of factor levels
kappa_data <- data4B_wide %>% dplyr::select(OCC1, OCC2)
# kappa2
kappa2_res <- kappa2(as.data.frame(kappa_data), weight = "squared")
kappa2_res
```

```
Cohen's Kappa for 2 Raters (Weights: squared)

 Subjects = 25
   Raters = 2
    Kappa = 0.368

        z = 1.91
  p-value = 0.0559
```

In our data, the weighted kappa was about 0.368, indicating fair agreement and suggesting that SPMSQ scores are moderately stable but not perfectly consistent across occasions.

Finally, to test for a systematic shift at group level we applied a Wilcoxon signed-rank test, the non-parametric equivalent of a paired t-test. This test ranks the paired differences and checks whether the median difference is zero.

```r
# Wilcoxon signed-rank
wilcox_res <- wilcox.test(data4B_wide$occ1_num, data4B_wide$occ2_num, paired = TRUE)
wilcox_res
```

```
        Wilcoxon signed-rank test with continuity correction

data:  data4B_wide$occ1_num and data4B_wide$occ2_num
V = 42, p-value = 0.8217
alternative hypothesis: true location shift is not equal to 0
```

Our p-value was 0.8217, larger than 0.05, indicating no significant overall change in SPMSQ between the two occasions—although some individuals improved and some worsened, there was no consistent group-level trend.

Together, these analyses show that the SPMSQ exhibits moderate intra-individual variability but no clear systematic change over time in this sample, which supports its use as a quick screening instrument but also highlights the importance of considering repeated measures and individual variability in future studies.