# 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.

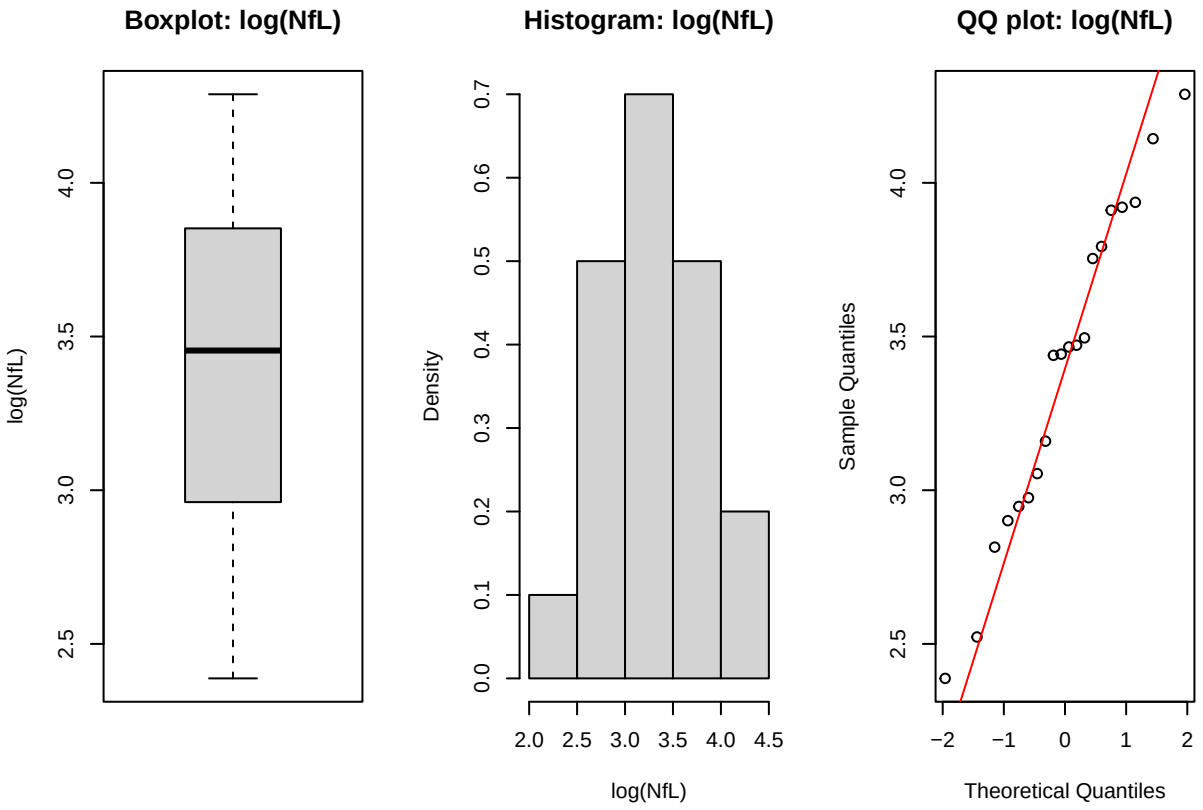
---

## 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
- $\alpha$ = 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 $1 - \beta$
- $\sigma$ = 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")
```

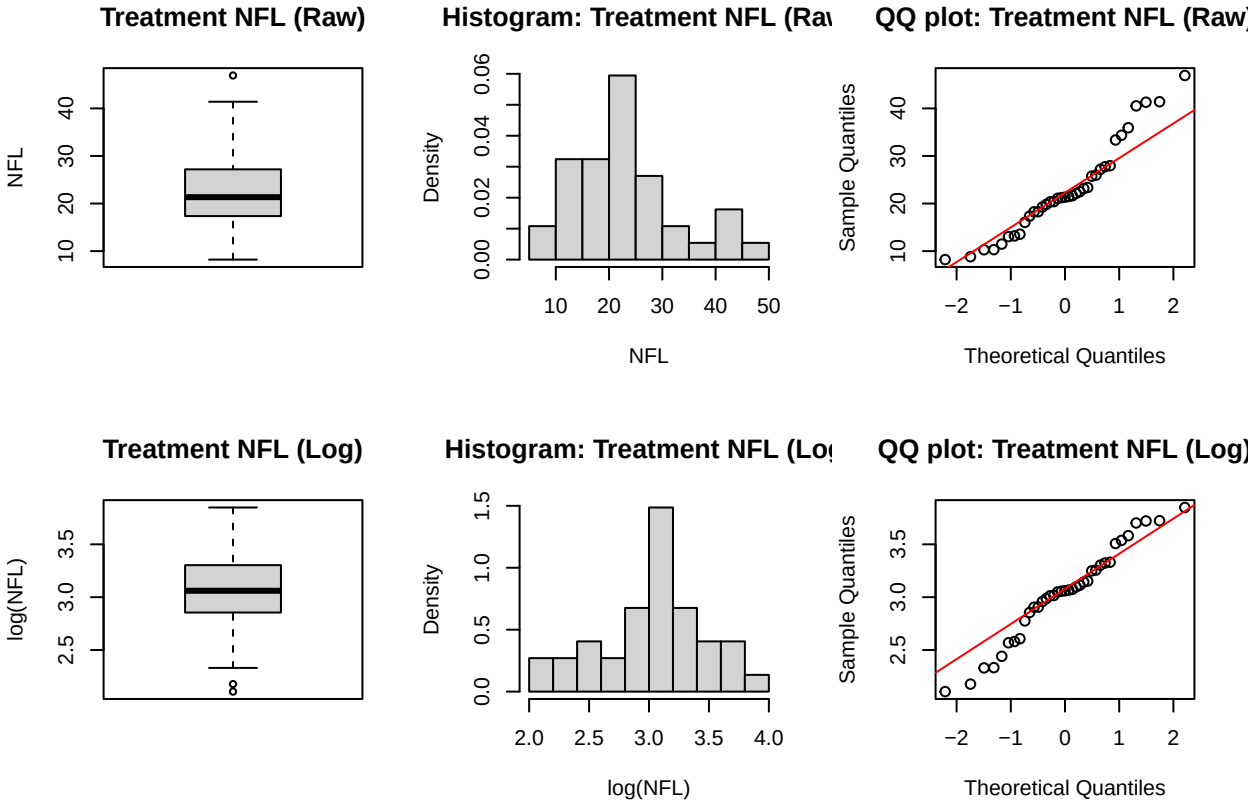
```
## Skipping install of 'CM2018rpackage' from a github remote, the SHA1 (c2bd9cd5) has not changed since
##   Use 'force = TRUE' to force installation
```

```
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)
```

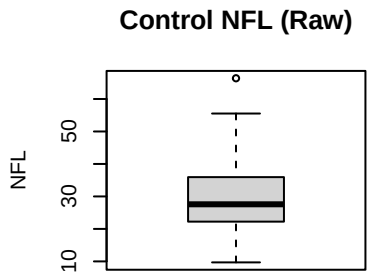
We now have two groups of simulated patients: treatment and control, and log-transformed NfL for parametric testing.

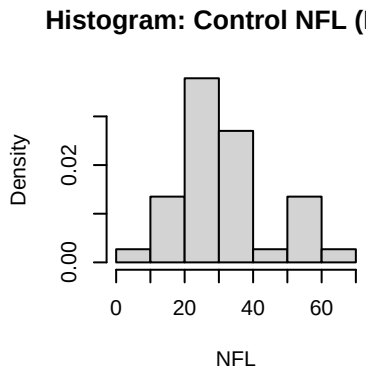
# Step 4: carry out analysis

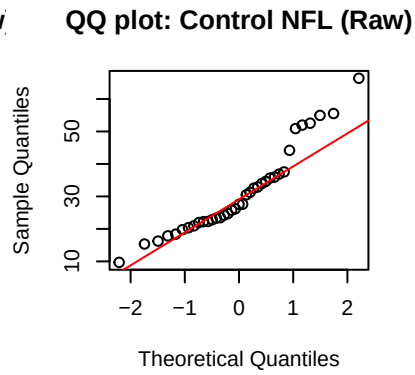
```r
# Exploratory plots and normality tests
par(mfrow=c(2,3))
# Treatment
boxplot(Treat_NFL, main = "Treatment NFL (Raw)",ylab = "NFL")
hist(Treat_NFL, prob=TRUE, main="Histogram: Treatment NFL (Raw)", xlab="NFL")
qqnorm(Treat_NFL, main="QQ plot: Treatment NFL (Raw)"); qqline(Treat_NFL, col="red")
boxplot(log_Tr_NFL,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

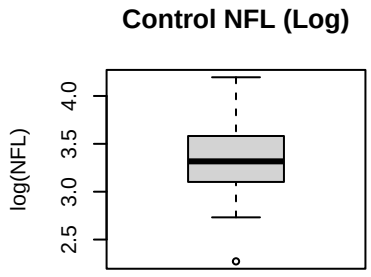


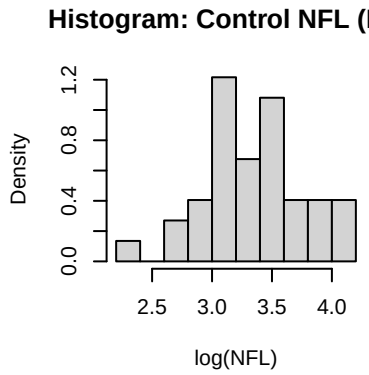

```r
# Control
boxplot(Control_NFL, main = "Control NFL (Raw)",ylab = "NFL")
hist(Control_NFL, prob=TRUE, main="Histogram: Control NFL (Raw)", xlab="NFL")
qqnorm(Control_NFL, main="QQ plot: Control NFL (Raw)"); qqline(Control_NFL, col="red")
boxplot(log_Co_NFL, main = "Control NFL (Log)",ylab = "log(NFL)")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

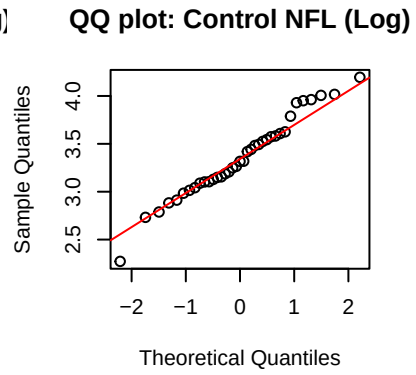



```
shapiro.test(Treat_NFL)
```

```
##
##  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.