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)")
```

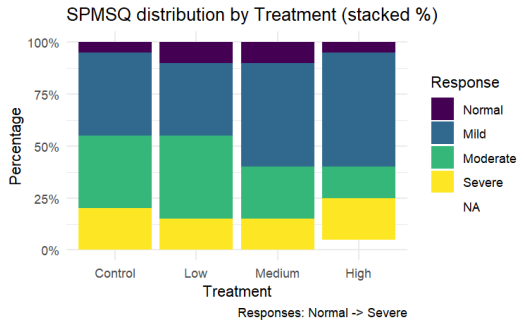

Figure 1: stacked percentage bar chart

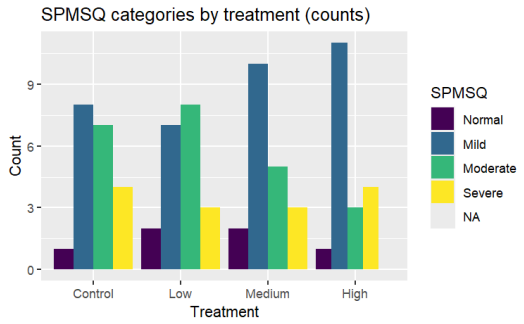

Figure 2: grouped bar plot

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()
```

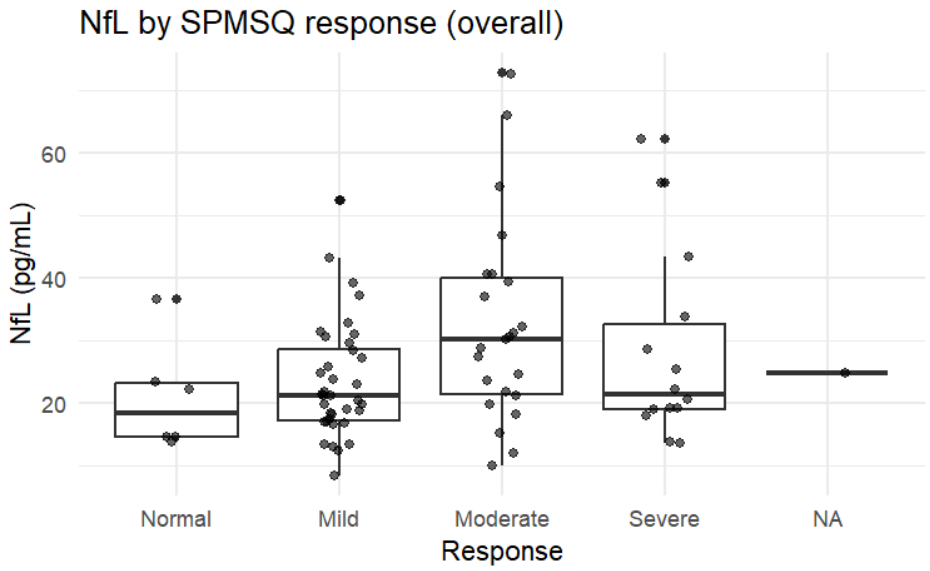

Figure 3: boxplot of NfL raw data

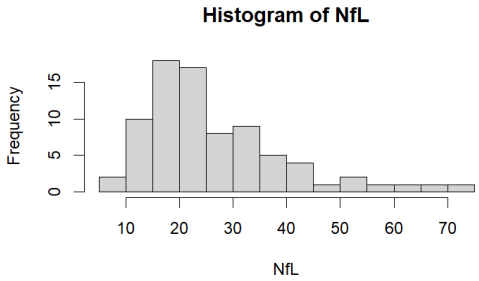

Figure 4: NfL histogram

```
        Shapiro-Wilk normality test

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

Figure 5: shapiro test of NfL

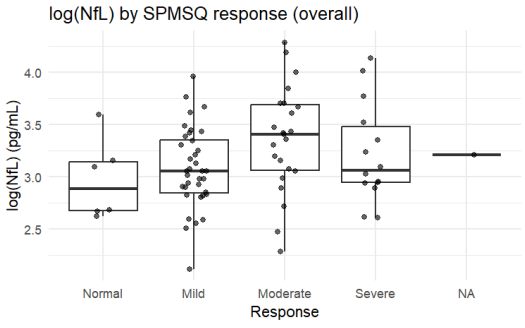

Figure 6: boxplot of log(NfL) data

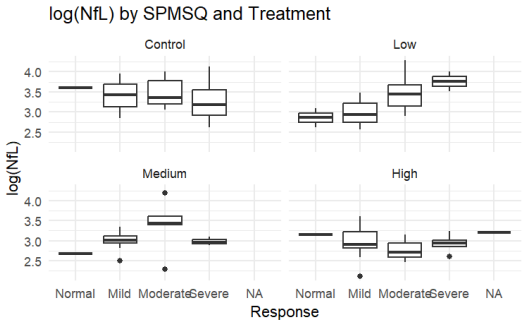

Figure 7: boxplot of log(NfL) by SPMSQ and Treatment

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

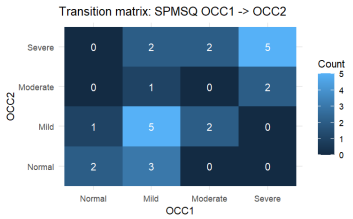

Figure 9: heatmap of transition matrix

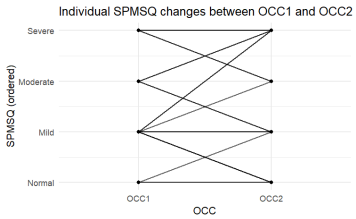

Figure 10: individual connection diagram

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.