

Martel Paper - Daily OA Pain

2024-03-16

Data Wrangling

```
## Load packages
library(tidyverse)
library(ggplot2)
library(tidyr)
library(haven) ## This library provides functions to read sav file into R
library(lme4)
library(lmerTest)
```

```
## Read data
data_paper <- read_sav("Dataset; 2024.1.sav")
checkdf3 <- data_paper |>
  subset(ID == 2072) |>
  select(c(ID,
           Level1_Even_DateIn,
           Level1_Even_TimeIn))
```

- DT_i combines D_i and T_i : 'DateTime' variable
- DT_0 is the first response 'DateTime' for each patient
- W_i is the adjusted 'Wave_Day' variable
- Add a grace period G for calculating the adjusted W_i , in our case $G = 6$ hours
- Calculate the datetime difference H_i in **hours** from the first response, incorporating the grace period:

$$H_i = DT_i - DT_{i-1}$$

- Then apply the grace period indicator I_i :

$$I_i = \begin{cases} 1 & \text{if } H_i \leq 24 + G \\ \left\lceil \frac{H_i - G}{24} \right\rceil & \text{otherwise} \end{cases}$$

- The initial response for 'Wave_Day' is 1, i.e., $W_0 = 1$, then the adjusted 'Wave_Day' W_i is

$$W_i = \sum_{i=0}^{i-1} I_i$$

```
## Consecutive Days - Grace Period 6 hours
data_paper <- data_paper |>
  mutate(Lev1_DateTimeIn = as.POSIXct(strptime(paste(Level1_Even_DateIn,
                                                    Level1_Even_TimeIn),
                                                    format="%Y-%m-%d %H:%M:")) |>

  arrange(ID, Lev1_DateTimeIn) |>
  group_by(ID) |>
  mutate(
    TimeDiffHours = as.numeric(difftime(Lev1_DateTimeIn,
                                         lag(Lev1_DateTimeIn,
                                              default = first(Lev1_DateTimeIn)),
                                         units = "hours")), # T diff
    WithinGracePeriod = if_else(TimeDiffHours <= 30,
                                1,
                                ceiling((TimeDiffHours - 6) / 24)), # Check grace perio
    Wave_Day_Adjusted = cumsum(WithinGracePeriod) # Adjusted Wave_Day
  ) |>
  ungroup()

## Check if the above approach is correct
checkdf <- data_paper |> select(c(ID,
                                 Lev1_DateTimeIn,
                                 TimeDiffHours,
                                 WithinGracePeriod,
                                 Wave_Day_Adjusted))
checkdf2 <- checkdf |> subset(ID == 2072) # Weird ID 2072
max(checkdf$Wave_Day_Adjusted)

## [1] NA

## Fill in the gap of Wave_Day
data_paper2 <- data_paper |>
  group_by(ID) |>
  complete(Wave_Day = 1:14) |>
  ungroup()

## Check the aberrant values
summary(data_paper2$IndexLev1_NegativeAffect_Total) # Lev 1 Negative Affect

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    0.000   8.667  20.500  25.154  36.667  97.333   1220

summary(data_paper2$IndexLev1_Catastrophizing_Total) # Lev 1 Catas

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    0.000   2.792  12.667  23.018  37.750 100.000   1220
```

```
summary(data_paper2$IndexLev2_QST_BaselinePPTH) # Lev 2 PPTHs?
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      67.0   235.5   369.0   395.3   473.5   1200.0   1313
```

```
summary(data_paper2$IndexLev2_QST_TSPAve) # Lev 2 TSP?
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##     -2.50    1.50    7.50   13.93   15.00   94.00   1276
```

```
summary(data_paper2$IndexLev2_QST_CpmTrialAve) # Lev 2 CPM?
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##     37.56  103.21  119.69  122.83  135.13  251.76   1321
```

$$APE(t_i) = I(PAIN(t_i) - PAIN(t_i - 1) \geq 20)$$

```
## APE index
```