Pain Research - Dr. Marc O. Martel

Jiaqi Bi

2023-07-31

```r
# Two different files
#setwd("~/Desktop/Project; Daily diaries; 30-days/Pain-Research-1-Dr-Marc-O-Martel/Datasets; Project; M
#setwd("~/Desktop/Project; Daily diaries; 30-days/Pain-Research-1-Dr-Marc-O-Martel/Datasets; Project; Jc

## Load packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(ggplot2)
library(tidyr)
library(haven)## This library provides functions to read sav file into R
library(lme4)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
library(lmerTest)
```

```
##
## Attaching package: 'lmerTest'
##
## The following object is masked from 'package:lme4':
##
##     lmer
##
## The following object is masked from 'package:stats':
##
##     step
```

```r
## Baseline data
#df1 <- read_sav("~/Desktop/Project; Daily diaries; 30-days/Pain-Research-1-Dr-Marc-O-Martel/Datasets; .
df1 <- read_sav("E:/UWO/DR Marc O Martel data/Pain-Research-1-Dr-Marc-O-Martel/Datasets; Project; Jamiso
## Daily data
#df2 <- read_sav("~/Desktop/Project; Daily diaries; 30-days/Pain-Research-1-Dr-Marc-O-Martel/Datasets; .
df2 <- read_sav("E:/UWO/DR Marc O Martel data/Pain-Research-1-Dr-Marc-O-Martel/Datasets; Project; Jamiso
```

```r
## Examine the duplicated observation
df1$ID[which(duplicated(df1$ID))]
```

```
## [1] 476
```

```r
df2$StudyID[which(duplicated(df2$StudyID))]
```

```
## [1] 476
```

```r
df1 <- df1 |>
  filter(!duplicated(ID))
df2 <- df2 |>
  filter(!duplicated(StudyID)) |>
  rename(ID = StudyID)

## Time variable to numerical day (consecutive)
date_cols <- grep("^Date", names(df2), value = TRUE)

df2 <- df2 %>%
  mutate(across(all_of(date_cols), as.Date, format = "%Y-%m-%d"))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `across(all_of(date_cols), as.Date, format = "%Y-%m-%d")`.
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

```r
## Correct abberrant years
df2 <- df2 %>%
  rowwise() %>%
  mutate(across(all_of(date_cols), ~ {
    if (. != Date1 && !is.na(.)) {
      day_diff <- as.numeric(. - Date1)
      if (day_diff < 0) {
        update(., year = year(Date1))
      } else {
        .
      }
    } else {
      .
    }
```

```r
  })) %>%
  ungroup()

####### Below investigates the rest of aberrant objects #######
##### Above does not solve those entering a new year ##########
df2_investigate <- df2 |>
  pivot_longer(
  cols = starts_with("Date"),
  names_to = "Date_Number",
  values_to = "Date_Value"
) |>
  select(ID, Date_Value)

print(n = 36, df2_investigate[df2_investigate$ID == 673,])
```

```
## # A tibble: 36 x 2
##         ID Date_Value
##      <dbl> <date>
##  1     673 2023-01-04
##  2     673 2023-01-05
##  3     673 2023-01-06
##  4     673 2023-01-07
##  5     673 2023-01-08
##  6     673 2023-01-09
##  7     673 2023-01-10
##  8     673 2023-01-11
##  9     673 2023-01-12
## 10     673 2023-01-13
## 11     673 2023-01-14
## 12     673 2023-01-15
## 13     673 2023-01-16
## 14     673 2023-01-17
## 15     673 2023-01-18
## 16     673 2023-01-20
## 17     673 2023-01-21
## 18     673 2023-01-22
## 19     673 2023-01-23
## 20     673 2023-01-24
## 21     673 2023-01-25
## 22     673 2023-01-26
## 23     673 2023-01-27
## 24     673 2023-01-28
## 25     673 2023-01-29
## 26     673 2023-01-30
## 27     673 2023-01-31
## 28     673 2023-01-02
## 29     673 2023-02-08
## 30     673 2023-02-09
## 31     673 2023-02-10
## 32     673 2023-02-11
## 33     673 2023-02-12
## 34     673 2023-02-13
## 35     673 2023-02-14
## 36     673 NA
```

```r
##############################################################

## Below solved all objects date problem
df2$Date8[df2$ID == 610] <- "2023-01-01"
df2$Date8[df2$ID == 703] <- "2022-01-09"
df2$Date8[df2$ID == 741] <- "2021-01-10"
df2$Date8[df2$ID == 680] <- "2023-01-09"
df2$Date28[df2$ID == 673] <- "2023-02-02"

## Convert the date to number
df2 <- df2 |>
  mutate(across(all_of(date_cols[-1]), ~ as.numeric(. - df2$Date1 + 1)))

df2$Date1 <- 1
all_cols <- names(df2)

#for (i in 2:length(date_cols)) {
#  if (any(df2[[date_cols[i]]] == 1, na.rm = TRUE)) {
#    same_day_cols <- grep(paste0("_", i), all_cols, value = TRUE)
#    df2[df2[[date_cols[i]]] == 1 & !is.na(df2_test[[date_cols[i]]]), same_day_cols] <- NA
#  }
#}

## Wide to long
df2_long <- df2 |>
  pivot_longer(cols = -ID,
               names_to = c(".value", "day"),
               names_pattern = "([A-Za-z]+)(\\d+)") |>
  select(-day) |>
  rename(Day = Date) |>
  filter(Day <= 30 & !is.na(Day)) |>
  group_by(ID) |>
  distinct(Day, .keep_all = TRUE) |>
  complete(Day = seq(1,30))

## replace all -1 to NA
df2_long <- df2_long |>
  mutate_all(~if_else(. < 0, NA_real_, .))
```

```
## `mutate_all()` ignored the following grouping variables:
## * Column `ID`
## i Use `mutate_at(df, vars(-group_cols()), myoperation)` to silence the message.
```

```r
df2_long <- df2_long |>
  rename(GlobalImprovement = Changed)

## Calculate the lagged variable
df2_long <- df2_long |>
  group_by(ID) |>
  arrange(ID, Day) |>
  mutate(AvePain_Change = AvePain - lag(AvePain),
         ActivityInt_Change = ActivityInt - lag(ActivityInt),
         Mood_Change = Mood - lag(Mood))
```

```
df2_greaterthan7 <- df2_long |>
  group_by(ID) %>%
  summarise(NonMissingGI = sum(!is.na(GlobalImprovement))) %>%
  filter(NonMissingGI >= 7) %>%
  inner_join(df2_long, by = "ID")
```

```
## lmer Analysis (GlobImp vs. AvePain_Change)
model_1 <- lmer(GlobalImprovement ~ AvePain_Change + Day + (1|ID), data = df2_greaterthan7) # Random in
summary(model_1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: GlobalImprovement ~ AvePain_Change + Day + (1 | ID)
##    Data: df2_greaterthan7
##
## REML criterion at convergence: 15963
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.2035 -0.4653  0.0104  0.4963  4.3924
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 1.300    1.140
##  Residual             1.838    1.356
## Number of obs: 4418, groups:  ID, 334
##
## Fixed effects:
##                  Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)     5.056e+00  7.627e-02  5.051e+02  66.290   <2e-16 ***
## AvePain_Change  3.912e-01  1.601e-02  4.123e+03  24.437   <2e-16 ***
## Day             5.886e-03  2.577e-03  4.236e+03   2.284   0.0224 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) AvPn_C
## AvePan_Chng  0.008
## Day         -0.464 -0.010
```

```
model_2 <- lmer(GlobalImprovement ~ AvePain_Change + Day + (Day|ID), data = df2_greaterthan7) # Random
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00293479 (tol = 0.002, component 1)
```

```
summary(model_2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: GlobalImprovement ~ AvePain_Change + Day + (Day | ID)
##    Data: df2_greaterthan7
##
## REML criterion at convergence: 15862.9
##
## Scaled residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -5.1370 -0.4367 -0.0004  0.4580  4.3993
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 1.598138 1.26418
##           Day         0.002075 0.04555  -0.44
##  Residual             1.706147 1.30620
## Number of obs: 4418, groups:  ID, 334
##
## Fixed effects:
##                 Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)    5.061e+00  8.294e-02 2.907e+02  61.017   <2e-16 ***
## AvePain_Change 3.921e-01  1.558e-02 3.976e+03  25.171   <2e-16 ***
## Day            5.328e-03  3.826e-03 2.341e+02   1.393    0.165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) AvPn_C
## AvePan_Chng  0.007
## Day         -0.579 -0.008
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00293479 (tol = 0.002, component 1)
```

```r
model_3 <- lmer(GlobalImprovement ~ AvePain_Change + Mood_Change + Day + (1|ID), data = df2_greaterthan7
summary(model_3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: GlobalImprovement ~ AvePain_Change + Mood_Change + Day + (1 |      ID)
##    Data: df2_greaterthan7
##
## REML criterion at convergence: 15831.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q      Max
## -5.1162 -0.4627  0.0070  0.4953  4.4936
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 1.280    1.131
##  Residual             1.783    1.335
## Number of obs: 4416, groups:  ID, 334
##
## Fixed effects:
##                 Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)    5.055e+00  7.551e-02 5.045e+02  66.946   <2e-16 ***
## AvePain_Change 3.502e-01  1.618e-02 4.118e+03  21.649   <2e-16 ***
## Mood_Change    1.741e-01  1.530e-02 4.124e+03  11.378   <2e-16 ***
## Day            6.057e-03  2.540e-03 4.232e+03   2.385   0.0171 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##              (Intr) AvPn_C Md_Chn
## AvePan_Chng  0.007
## Mood_Change  0.001 -0.222
## Day         -0.461 -0.011  0.001
```

```
model_4 <- lmer(GlobalImprovement ~ AvePain_Change + Mood_Change + Day + (Day|ID), data = df2_greatertha
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0080542 (tol = 0.002, component 1)
```

```
summary(model_4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: GlobalImprovement ~ AvePain_Change + Mood_Change + Day + (Day |
##     ID)
##    Data: df2_greaterthan7
##
## REML criterion at convergence: 15723.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.0140 -0.4439 -0.0015  0.4708  4.5065
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 1.602434 1.26587
##           Day         0.002127 0.04611  -0.45
##  Residual             1.648244 1.28384
## Number of obs: 4416, groups:  ID, 334
##
## Fixed effects:
##                 Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)    5.060e+00  8.269e-02 2.914e+02  61.192   <2e-16 ***
## AvePain_Change 3.512e-01  1.571e-02 3.969e+03  22.350   <2e-16 ***
## Mood_Change    1.736e-01  1.484e-02 3.962e+03  11.695   <2e-16 ***
## Day            5.407e-03  3.821e-03 2.355e+02   1.415    0.158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) AvPn_C Md_Chn
## AvePan_Chng  0.007
## Mood_Change  0.000 -0.222
## Day         -0.585 -0.008 -0.001
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0080542 (tol = 0.002, component 1)
```

```
## Model comparison for above analysis
anova(model_1, model_2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: df2_greaterthan7
## Models:
## model_1: GlobalImprovement ~ AvePain_Change + Day + (1 | ID)
## model_2: GlobalImprovement ~ AvePain_Change + Day + (Day | ID)
```

```
##           npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model_1      5 15953 15985 -7971.5    15943
## model_2      7 15858 15902 -7921.8    15844 99.331  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(model_3, model_4)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: df2_greaterthan7
## Models:
## model_3: GlobalImprovement ~ AvePain_Change + Mood_Change + Day + (1 | ID)
## model_4: GlobalImprovement ~ AvePain_Change + Mood_Change + Day + (Day | ID)
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model_3      6 15817 15855 -7902.4    15805
## model_4      8 15713 15764 -7848.6    15697 107.57  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
## lmer Analysis (GlobImp vs. Mood_Change)
model_5 <- lmer(GlobalImprovement ~ Mood_Change + Day + (1|ID), data = df2_greaterthan7) # random inter
summary(model_5)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: GlobalImprovement ~ Mood_Change + Day + (1 | ID)
##    Data: df2_greaterthan7
##
## REML criterion at convergence: 16284.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1992 -0.4668 -0.0036  0.4831  4.5103
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 1.292    1.137
##  Residual             1.987    1.410
## Number of obs: 4418, groups:  ID, 334
##
## Fixed effects:
##               Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 5.046e+00  7.709e-02 5.213e+02  65.462   <2e-16 ***
## Mood_Change 2.482e-01  1.575e-02 4.133e+03  15.759   <2e-16 ***
## Day         6.339e-03  2.677e-03 4.246e+03   2.368   0.0179 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Md_Chn
## Mood_Change  0.002
## Day         -0.477 -0.001
```

```r
model_6 <- lmer(GlobalImprovement ~ Mood_Change + Day + (Day|ID), data = df2_greaterthan7) # Add random
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
```

```
## Model failed to converge with max|grad| = 0.00397988 (tol = 0.002, component 1)
```

```r
summary(model_6)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: GlobalImprovement ~ Mood_Change + Day + (Day | ID)
##    Data: df2_greaterthan7
##
## REML criterion at convergence: 16203.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3118 -0.4244 -0.0109  0.4523  4.5002
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 1.552117 1.24584
##           Day         0.001928 0.04391  -0.42
##  Residual             1.863542 1.36512
## Number of obs: 4418, groups:  ID, 334
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 5.049e+00  8.297e-02 2.889e+02  60.850   <2e-16 ***
## Mood_Change 2.477e-01  1.537e-02 3.982e+03  16.119   <2e-16 ***
## Day         5.917e-03  3.831e-03 2.347e+02   1.545    0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Md_Chn
## Mood_Change  0.002
## Day         -0.577 -0.002
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00397988 (tol = 0.002, component 1)
```

```r
anova(model_5, model_6) # prefer random slope + intercept
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: df2_greaterthan7
## Models:
## model_5: GlobalImprovement ~ Mood_Change + Day + (1 | ID)
## model_6: GlobalImprovement ~ Mood_Change + Day + (Day | ID)
##         npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## model_5    5 16274 16306  -8132    16264
## model_6    7 16198 16243  -8092    16184 79.939  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```