

Pain Research - Dr. Marc O. Martel

Jiaqi Bi

2023-07-31

```
# 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; J

## 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

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

library(lmerTest)

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

```

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

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

## [1] 476
df2$StudyID[which(duplicated(df2$StudyID))]

## [1] 476
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))

## Correct aberrant 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 {
      .
    }
  })

```

```

})) %>%
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

```

```
#####

## 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(Date = Date) |>
  filter(Date <= 30 & !is.na(Date)) |>
  group_by(ID) |>
  distinct(Date, .keep_all = TRUE) |>
  complete(Date = 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.

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

## Calculate the lagged variable
df2_long <- df2_long |>
  group_by(ID) |>
  arrange(ID, Day) |>
  mutate(AvePain_Change = abs(AvePain - lag(AvePain)),
         ActivityInt_Change = abs(ActivityInt - lag(ActivityInt)),
         Mood_Change = abs(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: 16518.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.2103 -0.4505 -0.0073  0.4567  4.3841
##
## Random effects:
## Groups Name Variance Std.Dev.
## ID      (Intercept) 1.323  1.150
## Residual          2.100  1.449
## Number of obs: 4418, groups: ID, 334
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  5.007e+00  8.208e-02 6.162e+02  60.994  <2e-16 ***
## AvePain_Change 3.712e-02  2.517e-02 4.337e+03   1.475  0.1403
## Day           6.694e-03  2.755e-03 4.247e+03   2.429  0.0152 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) AvPn_C
## AvePan_Chng -0.295
## Day          -0.474  0.049

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

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0451398 (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: 16447.9
##
## Scaled residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -4.3174 -0.4243 -0.0074  0.4422  4.7548
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   ID       (Intercept) 1.569232 1.25269
##           Day          0.001871 0.04325  -0.40
##   Residual                1.979504 1.40695
## Number of obs: 4418, groups: ID, 334
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  5.005e+00  8.732e-02 3.316e+02  57.322  <2e-16 ***
## AvePain_Change 4.061e-02  2.488e-02 4.291e+03   1.632  0.1027
## Day          6.406e-03  3.863e-03 2.338e+02   1.658  0.0986 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) AvPn_C
## AvePan_Chng -0.275
## Day          -0.561  0.035
## optimizer (nlptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0451398 (tol = 0.002, component 1)
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: 16507.1
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -4.2123 -0.4462 -0.0071  0.4511  4.3904
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   ID       (Intercept) 1.304      1.142
##   Residual                2.097      1.448
## Number of obs: 4416, groups: ID, 334
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  4.959e+00  8.365e-02 6.663e+02  59.280  < 2e-16 ***
## AvePain_Change 2.954e-02  2.536e-02 4.325e+03   1.165  0.24416
## Mood_Change   5.852e-02  2.276e-02 4.374e+03   2.571  0.01017 *
## Day           7.145e-03  2.756e-03 4.243e+03   2.592  0.00958 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:

```

```

##          (Intr) AvPn_C Md_Chn
## AvePan_Chng -0.261
## Mood_Change -0.214 -0.123
## Day         -0.474  0.044  0.040
model_4 <- lmer(GlobalImprovement ~ AvePain_Change + Mood_Change + Day + (Day|ID), data = df2_greaterthan7)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0475696 (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: 16435.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.3210 -0.4188 -0.0087  0.4389  4.4953
##
## Random effects:
##   Groups    Name      Variance Std.Dev. Corr
##   ID        (Intercept) 1.570335 1.2531
##   Day                0.001893 0.0435  -0.42
##   Residual                1.975620 1.4056
## Number of obs: 4416, groups: ID, 334
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  4.955e+00  8.925e-02 3.553e+02  55.518 < 2e-16 ***
## AvePain_Change 3.300e-02  2.505e-02 4.277e+03   1.317  0.18785
## Mood_Change   6.012e-02  2.274e-02 4.366e+03   2.644  0.00823 **
## Day           6.960e-03  3.874e-03 2.353e+02   1.797  0.07366 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) AvPn_C Md_Chn
## AvePan_Chng -0.242
## Mood_Change -0.206 -0.122
## Day         -0.563  0.030  0.040
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0475696 (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 16510 16542 -8249.8    16500
## model_2     7 16444 16488 -8214.8    16430 70.05  2  6.148e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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 16494 16533 -8241.2    16482
## model_4     8 16427 16478 -8205.5    16411 71.341  2  3.224e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## 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: 16517.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.2044 -0.4503 -0.0094  0.4557  4.3869
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## ID      (Intercept)    1.304      1.142
## Residual                    2.101      1.450
## Number of obs: 4418, groups: ID, 334
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  4.986e+00  8.077e-02 5.909e+02  61.733  < 2e-16 ***
## Mood_Change  6.261e-02  2.260e-02 4.385e+03   2.770  0.00562 **
## Day          6.734e-03  2.755e-03 4.247e+03   2.445  0.01454 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) Md_Chn
## Mood_Change -0.257
## Day         -0.480  0.046

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| = 0.340787 (tol = 0.002, component 1)
```

```
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: 16447.3
```

```
##
```

```
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -4.2921 -0.4196 -0.0094  0.4351  4.4971
```

```
##
```

```
## Random effects:
```

```
## Groups   Name                Variance Std.Dev. Corr
```

```
## ID      (Intercept) 1.406269 1.18586
```

```
##        Day          0.001771 0.04208 -0.36
```

```
## Residual                1.991874 1.41134
```

```
## Number of obs: 4418, groups: ID, 334
```

```
##
```

```
## Fixed effects:
```

```
##              Estimate Std. Error      df t value Pr(>|t|)
```

```
## (Intercept) 4.984e+00  8.359e-02 3.457e+02 59.632 < 2e-16 ***
```

```
## Mood_Change 6.506e-02  2.260e-02 4.376e+03  2.879  0.00401 **
```

```
## Day         6.626e-03  3.820e-03 2.421e+02  1.735  0.08410 .
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Correlation of Fixed Effects:
```

```
##              (Intr) Md_Chn
```

```
## Mood_Change -0.252
```

```
## Day         -0.556  0.043
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
```

```
## Model failed to converge with max|grad| = 0.340787 (tol = 0.002, component 1)
```

```
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 16508 16540  -8249    16498
```

```
## model_6     7 16442 16487  -8214    16428 69.978  2 6.373e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```