# STOPPD mean diffusivity analysis
# (left entorhinal cortex)

## Contents

## This script analyses mean diffusivity in the left entorhinal cortex.

```r
#load libraries
library(tidyverse)
library(lme4)
library(lmerTest)
library(growthmodels)
```

```r
#bring in subject info (generated by 03_STOPPD_masterDF.Rmd)
# then take only the subjects who completed (n= 74)
df <- read_csv('../generated_csvs/STOPPD_masterDF_2018-09-06.csv') %>%
  mutate(STUDYID = as.character(STUDYID)) %>%
  filter(second_complete == "Yes")

#rename timepoint variable for clarity
colnames(df)[colnames(df)=="second_timepoint"] <- "category"

#make a datediff column for time between scans
df$dateDiff <- as.numeric(round(difftime(df$second_date, df$first_date, units = "days"), 0))
```

### Known exclusion reasons

**subject 210030 (MAS)**: serious outlier.. it appears that this participants baseline is acutally a mislabled repeat of the subject 210022's baseline data. Calculating the difference in cortical thinkness between two different people was a relatively large value..

**subject 320032 (PMC)**: incidental findings more atrophy, should exclude

### known DWI issues

**subject 410012 timepoint 02** -> scan was blacklisted "aborted" for system failure..no DWI for this participant

**subject 220009_timepoint 01** -> scan was also incomplete (this participant was only able complete the T1w)

So we will filter the data table to exclude these 4 participants (final n=70)

```r
df <- filter(df, !(STUDYID %in% c("210030", "320032", "410012", "220009")))
```

## mangling the Mean Diffusivity cata data

Erin reran the enigma DTI pipeline for only PMC using a different skull stripping parameter (-fa 0.7 to BET). We will use these numbers instead of the others in the archive here..

```r
#bring in MD data (from the filesystem)
MD_most <- read_csv('../data/enigma-DTI_archive_201809/enigmaDTI-MD-results.csv')
MD_PMC <- read_csv('../data/enigma-DTI_PMCredo_201809/enigmaDTI-MD-results.csv')

# separate id into it's parts and then drop old PMC data
MD_most <- MD_most %>%
  separate(id, into = c("study", "site", "STUDYID", "timepoint")) %>%
  filter(site != "PMC")

# separate the PMC subject id into it's parts and then bind to the data from the other sites
MD <- MD_PMC %>%
  separate(id, into = c("study", "site", "STUDYID", "timepoint")) %>%
  bind_rows(MD_most)

# drop acute ("00") and other ("03") timepoints from the analysis
MD <- MD %>%
  filter(!(timepoint %in% c("00", "03"))) %>%
  gather(tract, MD, ends_with("MD")) %>%
  unite(tract_timepoint, tract, timepoint) %>%
  spread(tract_timepoint, MD)
```

## check for missing MD data

```r
# filter the master spreadsheet for the list of completers (no output means we are ok)
df %>%
  anti_join(MD, by = "STUDYID") %>%
  summarise(`Number of missing MD values` = n()) %>%
  knitr::kable()
```

| Number of missing MD values |
| --- |
| 0 |

## merge (i.e. join) the MD data with the clinical scores

```r
all_MD <- df %>%
  select(STUDYID, sex, age, randomization, category, dateDiff) %>%
  left_join(MD, by = "STUDYID")

all_MD %>%
  filter(is.na(AverageFA_MD_01)) %>%
  summarise(`Number of missing timepoint 1 MD values` = n()) %>%
```

```
  knitr::kable()
```

| Number of missing timepoint 1 MD values |
|---|
| 0 |

```
all_MD %>%
  filter(is.na(AverageFA_MD_02)) %>%
  summarise(`Number of missing timepoint 2 MD values` = n()) %>%
  knitr::kable()
```

| Number of missing timepoint 2 MD values |
|---|
| 0 |

```
#write out clean MD speadsheet (required for subsequent MD analyses)
write.csv(all_MD, '../generated_csvs/STOPPD_MDclean.csv', row.names = FALSE)
```

## RCT only

```
#boxplot of difference in MD in whole skeleton (y axis) by randomization group (x axis)
ggplot(RCT_MD, aes(x= randomization, y = diffAverageSkel_MD)) +
  geom_boxplot(outlier.shape = NA) +
  geom_dotplot(binaxis = 'y', stackdir = 'center') +
  ggtitle("Whole skeleton mean diffusivity") +
  xlab("randomization condition") +
  ylab("difference between timepoints")
```

```
## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```

## Whole skeleton mean diffusivity
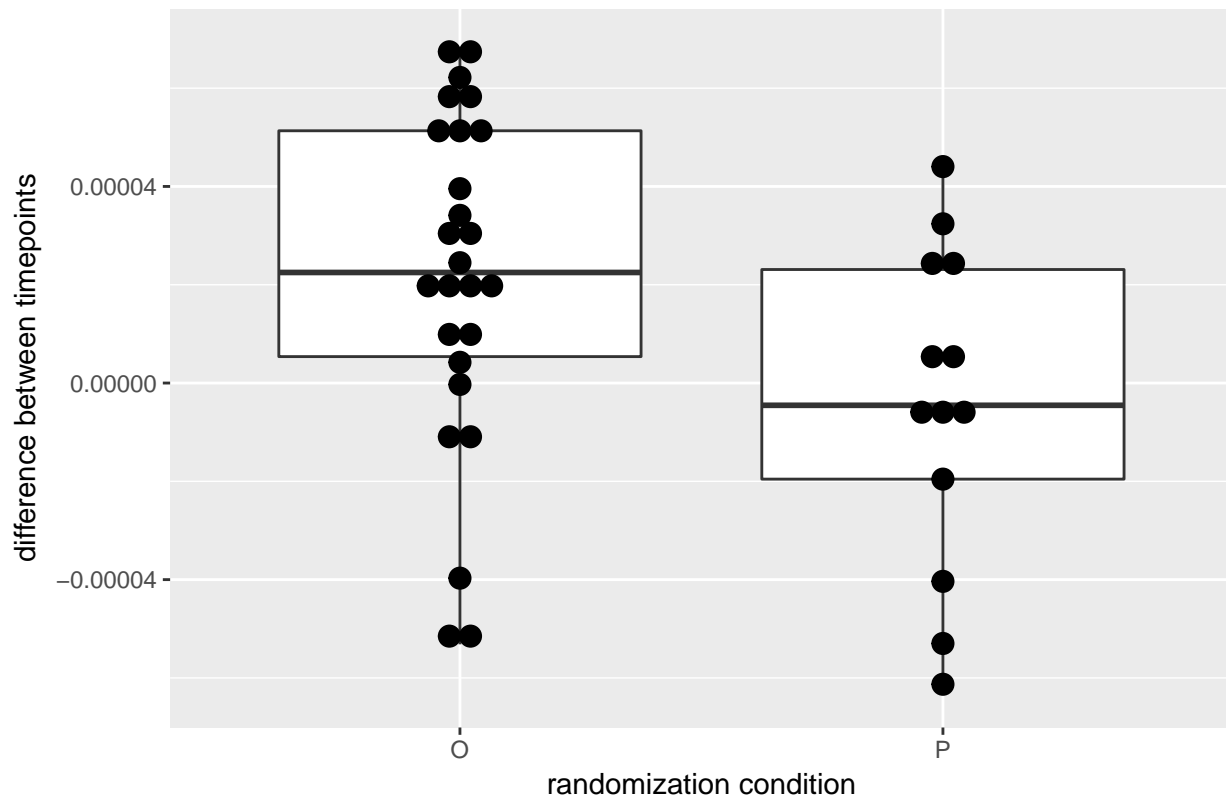
```
fit_rct <- lm(diffAverageSkel_MD ~ randomization, data= RCT_MD)
print(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization, data = RCT_MD)
##
## Coefficients:
##    (Intercept)   randomizationP
##     0.00002180      -0.00002606
```

```
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization, data = RCT_MD)
##
## Residuals:
##          Min          1Q       Median           3Q          Max
## -0.000074955 -0.000016430 -0.000000265  0.000028775  0.000048305
##
## Coefficients:
##                  Estimate   Std. Error t value Pr(>|t|)
## (Intercept)    0.000021795  0.000006622   3.291   0.0022 **
## randomizationP -0.000026061  0.000011470  -2.272   0.0290 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00003377 on 37 degrees of freedom
## Multiple R-squared:  0.1224, Adjusted R-squared:  0.09872
## F-statistic: 5.162 on 1 and 37 DF,  p-value: 0.02899
```

```r
#run linear model with covariates of sex and age
fit_rct <- lm(diffAverageSkel_MD ~ randomization + sex + age, data= RCT_MD)
print(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization + sex + age,
##     data = RCT_MD)
##
## Coefficients:
##    (Intercept)  randomizationP            sexM              age
##   0.00001578270  -0.00002642399   0.00000240912    0.00000009383
```

```r
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization + sex + age,
##     data = RCT_MD)
##
## Residuals:
##          Min          1Q       Median           3Q          Max
## -0.000078858 -0.000015710  0.000001514  0.000027177  0.000051303
##
## Coefficients:
##                    Estimate     Std. Error t value Pr(>|t|)
## (Intercept)     0.00001578270  0.00002188374   0.721   0.4756
## randomizationP -0.00002642399  0.00001208614  -2.186   0.0356 *
## sexM            0.00000240912  0.00001153089   0.209   0.8357
## age             0.00000009383  0.00000040916   0.229   0.8200
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00003466 on 35 degrees of freedom
## Multiple R-squared:  0.1256, Adjusted R-squared:  0.0506
## F-statistic: 1.675 on 3 and 35 DF,  p-value: 0.1901
```

```r
#run linear model with covariates of sex, age and site
fit_rct <- lm(diffAverageSkel_MD ~ randomization + sex + age + site, data= RCT_MD)
print(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization + sex + age +
##     site, data = RCT_MD)
##
## Coefficients:
##    (Intercept)  randomizationP            sexM              age
##   0.0000015618   -0.0000283856    0.0000026839     0.0000003249
##        siteMAS          siteNKI          sitePMC
```

```
##    0.0000082173    0.0000118892   -0.0000128727
```

```r
summary(fit_rct)
```

```
## 
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization + sex + age +
##     site, data = RCT_MD)
## 
## Residuals:
##          Min          1Q      Median          3Q         Max
## -0.000072052 -0.000012354  0.000002321  0.000020826  0.000049127
## 
## Coefficients:
##                   Estimate    Std. Error t value Pr(>|t|)
## (Intercept)     0.0000015618  0.0000248965   0.063   0.9504
## randomizationP -0.0000283856  0.0000124375  -2.282   0.0293 *
## sexM            0.0000026839  0.0000119040   0.225   0.8231
## age             0.0000003249  0.0000004581   0.709   0.4834
## siteMAS         0.0000082173  0.0000159846   0.514   0.6107
## siteNKI         0.0000118892  0.0000148117   0.803   0.4281
## sitePMC        -0.0000128727  0.0000181798  -0.708   0.4840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.00003532 on 32 degrees of freedom
## Multiple R-squared:  0.1698, Adjusted R-squared:  0.01415
## F-statistic: 1.091 on 6 and 32 DF,  p-value: 0.3889
```

## RCT & Relapse (with time as factor)

```r
RCTRelapse_wholeskelMD <- RCTRelapse_MD %>%
  filter(Tract == "AverageFA")
#plot
RCTRelapse_wholeskelMD %>%
  ggplot(aes(x=model_days, y=MD, colour=randomization)) +
  geom_point() +
  geom_line(aes(group=STUDYID), alpha = 0.5) +
  geom_smooth(method="lm", formula=y~poly(x,1)) +
  ggtitle("Whole skeleton mean diffusivity over time") +
  xlab("days between MRIs") +
  ylab("Mean Diffusivity") +
  theme_minimal()
```

## Whole skeleton mean diffusivity over time
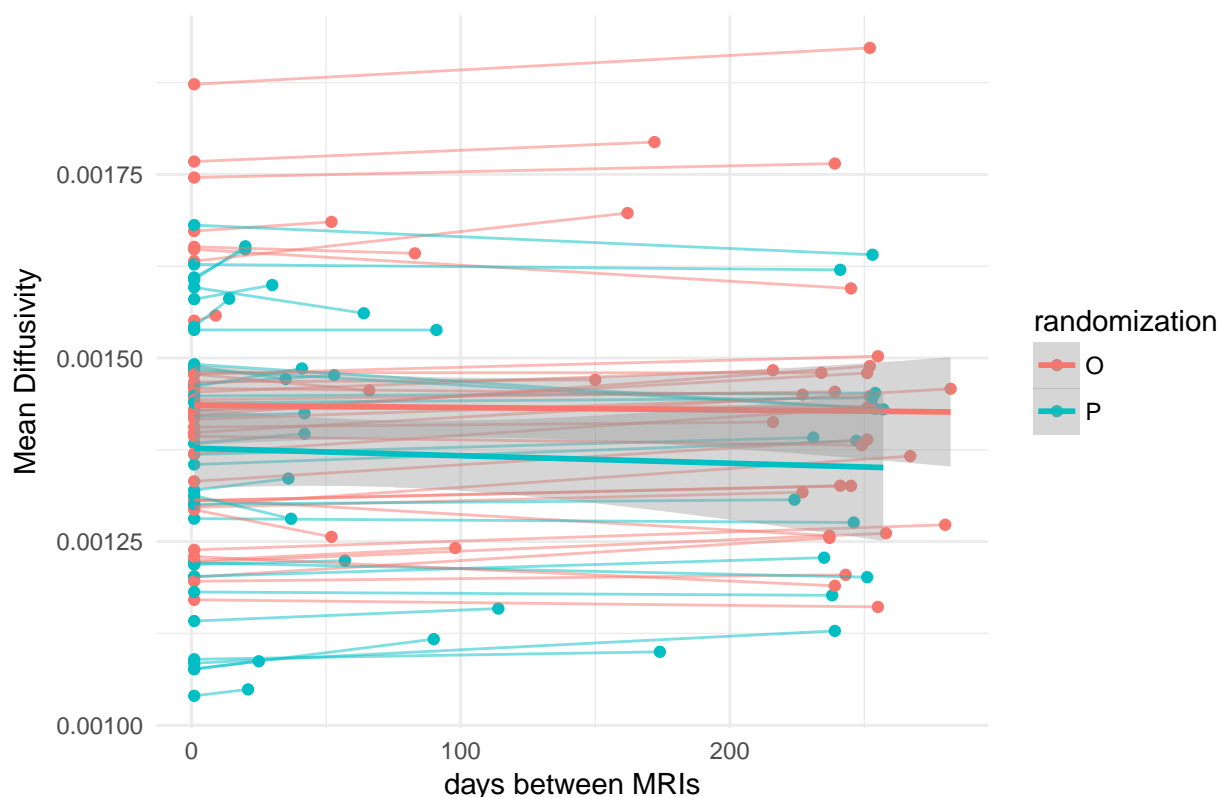


```r
#run mixed linear model, with covariates
fit_all <- lmer(MD ~ randomization*model_days + sex + age + (1|STUDYID), data= RCTRelapse_wholeskelMD)
summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: MD ~ randomization * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_wholeskelMD
##
## REML criterion at convergence: -2172.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.79473 -0.45375 -0.02211  0.38740  1.75336
##
## Random effects:
##  Groups   Name        Variance      Std.Dev.
##  STUDYID  (Intercept) 0.0000000162720 0.00012756
##  Residual             0.0000000004398 0.00002097
## Number of obs: 140, groups:  STUDYID, 70
##
## Fixed effects:
##                            Estimate     Std. Error            df
## (Intercept)              0.00097123189 0.00005922449 66.23783533131
## randomizationP          -0.00006682099 0.00003104642 67.33483764239
## model_days               0.00000008709 0.00000002214 68.19357217598
## sexM                     0.00003295832 0.00003101516 65.99532783955
```

```
## age                          0.00000798688   0.00000100086 65.99338155713
## randomizationP:model_days -0.00000009381   0.00000003809 68.49572304325
##                                   t value            Pr(>|t|)
## (Intercept)                        16.399 < 0.0000000000000002 ***
## randomizationP                     -2.152            0.034962 *
## model_days                          3.934            0.000198 ***
## sexM                                1.063            0.291813
## age                                 7.980     0.0000000000288 ***
## randomizationP:model_days          -2.463            0.016298 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) rndmzP mdl_dy sexM   age
## randomiztnP -0.197
## model_days  -0.043  0.075
## sexM        -0.166  0.068  0.002
## age         -0.898 -0.072  0.004 -0.096
## rndmztnP:m_  0.024 -0.098 -0.581  0.003 -0.002
```

```r
#cleanup
rm('df', 'fit_all', 'fit_rct', 'MD', 'plot', 'RCT_FA', 'RCTRelapse_FA')
```

**just wanna through in one site model..**

```r
#run mixed linear model, with covariates
fit_all <- lmer(MD ~ randomization*model_days + sex + age + site + (1|STUDYID), data= RCTRelapse_wholeskel
summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## MD ~ randomization * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_wholeskelMD
##
## REML criterion at convergence: -2165.5
##
## Scaled residuals:
##      Min      1Q   Median      3Q      Max
## -1.75027 -0.45651 -0.00872  0.40133  1.90149
##
## Random effects:
##  Groups   Name        Variance       Std.Dev.
##  STUDYID  (Intercept) 0.0000000078425 0.00008856
##  Residual             0.0000000004399 0.00002097
## Number of obs: 140, groups:  STUDYID, 70
##
## Fixed effects:
##                               Estimate     Std. Error              df
## (Intercept)                0.00107620814  0.00004345570 63.38845538623
## randomizationP            -0.00003508047  0.00002215435 65.50652097021
## model_days                 0.00000008625  0.00000002213 68.34133906756
## sexM                       0.00006003452  0.00002200440 62.95205381331
```

```
## age                          0.00000760310  0.00000070508 62.94985076260
## siteMAS                     -0.00019464558  0.00002847356 62.95516476445
## siteNKI                     -0.00018439503  0.00003040878 62.95436840632
## sitePMC                     -0.00018734952  0.00003128515 62.94544058940
## randomizationP:model_days  -0.00000009601  0.00000003802 68.96745578641
##                             t value          Pr(>|t|)
## (Intercept)                  24.766 < 0.0000000000000002 ***
## randomizationP               -1.583           0.118134
## model_days                    3.898           0.000224 ***
## sexM                          2.728           0.008241 **
## age                          10.783 0.000000000000000614 ***
## siteMAS                      -6.836 0.000000003867958716 ***
## siteNKI                      -6.064 0.000000082815682000 ***
## sitePMC                      -5.988 0.000000111344241240 ***
## randomizationP:model_days    -2.525           0.013880 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) rndmzP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## randomiztnP -0.136
## model_days  -0.059  0.105
## sexM        -0.124  0.087  0.003
## age         -0.878 -0.083  0.006 -0.094
## siteMAS     -0.286 -0.152  0.004 -0.066  0.113
## siteNKI     -0.175 -0.124 -0.008 -0.125  0.015  0.346
## sitePMC     -0.153 -0.094 -0.001 -0.149 -0.005  0.332  0.320
## rndmztnP:m_  0.031 -0.138 -0.582  0.004 -0.003  0.004  0.007  0.002
```