STOPPD RCT Analysis Index

Erin Dickie and Navona Calarco

November 16, 2018

Contents

1 The index page

2 Verifying number of scans

2.1 Checking the TIGRLab "/archive/data"

This script pulls in and cleans up the naming of STOPPD scans as they exist in the Kimel lab file system. At earlier stages, this script helped us identify naming errors in the file system (all have since been fixed).

Purpose: The contents of the file system will, in other scripts, be checked against (1) the scans we have in XNAT, to ensure that there are no discrepancies between these databases, and also against (2) our subject inclusion list.

```
library('stringi')
library('stringr')
library('plyr')
library('tidyr')
```

```
#import spreadsheet ('ls' of file system)
terminal <- read.csv('../data/stoppd_NiiFolderContents_2018-01-25.csv', header = TRUE, stringsAsFactors
```

```
#make a new column for site component of ID
terminal$site <- str_sub(terminal$scan_id, 8, 10)

#cut out study and site component from ID (first 11 characters)
terminal$scan_id <- substring(terminal$scan_id, 12)

#make a new column for session component of ID
terminal$session <- str_sub(terminal$scan_id, -2)

#cut out session information from ID (last 3 characters)
terminal$scan_id <- stri_sub(terminal$scan_id, 1, -4)

#make a new column that captures alphabetic component of ID ('R')
terminal$contains_R <- grepl('R', terminal$scan_id, fixed=TRUE) #36 participants

#cut out the 'R' in some participant IDs (indicates repeat for controls)
terminal$scan_id <- gsub("[R]", "", terminal$scan_id)

#make a 'group' column to capture case vs. control information
terminal$group <- stri_sub(terminal$scan_id, 2, 2) #note: 1 or 2 is patient, 6 is control

#for clarity, change values in 'group' column to labels for clarity
terminal$group[terminal$group == 1] <- "patient"
terminal$group[terminal$group == 2] <- "patient"
terminal$group[terminal$group == 6] <- "control"
```

```r
#make a variable that combines unique ID and session number
terminal$id_session <- paste(terminal$scan_id, '_', terminal$session, sep='')

#write csv
write.csv(terminal, '../generated_csvs/terminal_clean_2018-01-25.csv', row.names=FALSE)

#cleanup
rm(terminal)
```

2.2 Checking XNAT

This script pulls in and cleans up the naming of STOPPD scans as they exist in XNAT. At earlier stages, this script helped us identify naming errors in XNAT (all have since been fixed).

Purpose: The contents of XNAT will, in other scripts, be checked against (1) the scans we have in our file system, to ensure that there are no discrepancies between these databases, and also against (2) our subject inclusion list.

```r
#import spreadsheets (exported from XNAT)
xnat_camh <- read.csv('../data/xnat_records/xnat_cmh_2018-01-25.csv')
xnat_nki <- read.csv('../data/xnat_records/xnat_nki_2018-01-25.csv')
xnat_pitt <- read.csv('../data/xnat_records/xnat_pmc_2018-01-25.csv')
xnat_umass <- read.csv('../data/xnat_records/xnat_umas_2018-01-25.csv')

#combine XNAT spreadsheets, take only ID and date columns
xnat <- Reduce(function(x, y) merge(x, y, all=TRUE), list(xnat_camh, xnat_nki, xnat_pitt, xnat_umass))
xnat <- xnat[c('MR.ID', 'Date') ]

#cleanup
rm (xnat_camh, xnat_nki, xnat_pitt, xnat_umass)

#import spreadsheet of data in file system (made in script 01_STOPPD_terminal)
terminal <- read.csv('../generated_csvs/terminal_clean_2018-01-25.csv')
```

```r
#remove all CAMH scans with '00' as timepoint (NOTE: '00' this is a consequence of creative naming to a
xnat$timepoint <- str_sub(xnat$MR.ID, start= -2) #make column with timepoint data
xnat <- xnat[-grep('00', xnat$timepoint),] #remove those with 00

#cut out timepoint info from subject ID string - now meaningless - and remove timepoint column
xnat$MR.ID  <- str_sub(xnat$MR.ID, 1, -4)
xnat <- xnat[, -grep('timepoint', colnames(xnat))]

#cut out study and site info from subject ID string - not needed
xnat$MR.ID <- substring(xnat$MR.ID, 12)

#make a new column for session component of ID
xnat$session <- str_sub(xnat$MR.ID, -2)
table(xnat$session)
```

```
##
##  00  01  02  03
##  17 222  77   7
```

```r
#cut out session from subject ID string - not needed
xnat$MR.ID <- str_sub(xnat$MR.ID, 1, -4)

#make a new column that captures alphabetic component of ID ('R')
xnat$contains_R <- grepl('R', xnat$MR.ID, fixed=TRUE)

#cut out the 'R' in some participant IDs (indicates repeat for controls)
xnat$MR.ID <- gsub("[R]", "", xnat$MR.ID)

#make a variable that combines unique ID and session number
xnat$id_session <- paste(xnat$MR.ID, '_', xnat$session, sep='')

#check for consistency between file system and XNAT
X <- terminal$id_session %in% xnat$id_session
  which(X == FALSE) #identical
```

```
## integer(0)
```

```r
Y <- xnat$id_session %in% terminal$id_session
  which(Y == FALSE) #identical
```

```
## integer(0)
```

```r
#write csv
write.csv(xnat, '../generated_csvs/xnat_clean_2018-01-25.csv', row.names=FALSE)

#cleanup
rm(terminal, xnat)
```

3 Decoding the Master Scan Log

This script combines information in XNAT/file system (which have already been established to be identical in script 02_STOPPD_xnat) and study logs, and randomization (recently unblinded), into a single, master spreadsheet.

Purpose: the output csv (STOPPD_participantList_2018-11-05.csv) is meant to serve as a master reference sheet for all participants that were randomized (irrespective of scan completion).

This script now also adds a column relating to whether or not the subject is ok for MR analysis (i.e. not excluded for later identified neurological condition)

Note: this script does not remove individuals who failed preprocessing, QC, or should be removed from the dataset for any other reason.

```r
library('stringi')
library('plyr')
library('tidyr')
library('stringr')
```

```r
#import spreadsheets
xnat <- read.csv('../generated_csvs/xnat_clean_2018-01-25.csv', stringsAsFactors = FALSE) #generated by
randomization <- read.csv('../data/clinical/randomization.csv', stringsAsFactors = FALSE) #from Judy (S
log <- read.csv('../data/clinical/master_log.csv', fileEncoding="latin1", na.strings=c(""," ","NA", "N/

#transform XNAT df from long to wide format
```

```
xnat <- xnat[!names(xnat) %in% c('contains_R', 'id_session')] #remove unnecessary variables
xnat <- reshape(xnat, idvar = "MR.ID", timevar = 'session', direction = "wide")
```

```
## Warning in reshapeWide(data, idvar = idvar, timevar = timevar, varying =
## varying, : multiple rows match for session=1: first taken

## Warning in reshapeWide(data, idvar = idvar, timevar = timevar, varying =
## varying, : multiple rows match for session=2: first taken
```

```
colnames(xnat) <- c('subject_id', 'first_date_xnat', 'second_date_xnat', 'third_date_xnat', 'acute_date_
```

```
#merge xnat with randomization - will get rid of controls, etc
df <- merge(randomization[c('STUDYID', 'BLINDMED')], xnat, all.x=TRUE, by.x='STUDYID', by.y = 'subject_

#rename randomization column
colnames(df)[colnames(df)=="BLINDMED"] <- "randomization"

#combine the 'notes' columns in the log file (easier to read for now)
log$Comments.1 <- paste(log$Specify.reason.if.scan.not.completed.1, log$Comments.1)
log$Comments.2 <- paste(log$Specify.reason.if.scan.not.completed.2, log$Comments.2)
log$Comments.3 <- paste(log$Specify.reason.if.scan.not.completed.3, log$Comments.3)

#make subset of log variables from log we want to merge with randomization info
log <- log[c(
  "STOPPD.clinical.Trial.ID.Imaging.ID",
  'Sex',
  'Age',
  "Date.of.randomization...Stop.PD",
  "Date.of.consent.to.imaging.study",
  "If.not.enrolled.to.imaging.study..specify.reason.",
  "Study.day.of.acute.phase.MRI",
  "Scan.completed.Y.N",
  "Date.of.MRI..1" ,
  "Study.week",
  "Scan.completed.Y.N.1",
  "Comments.1",
  "Date.of.MRI..2",
  "Study.week.1",
  "Scan.completed",
  "Comments.2",
  "Date.of.MRI..3",
  "Study.week.2",
  "Scan.completed.1",
  "Comments.3")]

#rename the columns of the variables from log we want to merge with randomization info, for clarity
colnames(log) <- c(
  'subject_id',
  'sex',
  'age',
  'randomization_date',
  'imaging_consent_date',
  'imaging_nonconsent_reason',
  'acute_date_log',
  'acute_complete_log',
  'first_date_log',
  'first_timepoint_log',
  'first_complete_log',
  'first_notes',
  'second_date_log',
  'second_timepoint_log',
  'second_complete_log',
  'second_notes',
  'third_date_log',
  'third_timepoint_log',
  'third_complete_log',
  'third_notes')

#merge the df and log data
df <- merge(df, log, all.x=TRUE, by.x = 'STUDYID', by.y='subject_id')

#reorder df columns, for clarity
df <- df[c(
  "STUDYID",
  'sex',
  'age',
  "randomization",
  'randomization_date',
  'imaging_consent_date',
  'imaging_nonconsent_reason',
  "acute_date_log",
  "acute_complete_log",
  "acute_date_xnat",
  "first_date_log",
  "first_timepoint_log",
  "first_complete_log",
  'first_notes',
  "first_date_xnat",
  "second_date_log",
  "second_timepoint_log",
  "second_complete_log",
  'second_notes',
  "second_date_xnat",
  "third_date_log",
  "third_timepoint_log",
  "third_complete_log",
  'third_notes',
  "third_date_xnat")]

#make sure dates, etc. are characters (not factors) by converting all factors to characters
i <- sapply(df, is.factor)
df[i] <- lapply(df[i], as.character)

#clean up the NA-related values (which exist in the 3 notes columsn, 'first_notes', 'second_notes', 'th
df <- data.frame(lapply(df, function(x) {
      gsub("NA NA", NA, x)
      }))
```

```
df <- data.frame(lapply(df, function(x) {
      gsub("NA", NA, x)
      }))

#alter incorrect/unclear values as required
  #acute scan
  df$acute_complete_log <- as.character(df$acute_complete_log)
  df$acute_complete_log[df$acute_complete_log == 'Y'] <- "Yes"
  df$acute_complete_log[df$acute_complete_log == "N" & df$STUDYID == '420043'] <- NA #(replace 'no' wit

  #first scan (replace 'no' with NA, to take care of inconsistent notation)
  df["first_complete_log"] <- lapply(df["first_complete_log"], function(x) {
    gsub("No", NA, x)
    })

  #second scan
    df["second_complete_log"] <- lapply(df["second_complete_log"], function(x) {
    gsub("No", NA, x)
    })

  #third scan
  df$third_timepoint_log[df$third_timepoint_log == "what would be RCT Week 36"] <- "RCT Week 36"

#remove 'day' information from 'acute_date_log' and turn into integer
df$acute_date_log <- sub('\\,.*', '', df$acute_date_log) #strip out day info
df$acute_date_log <- as.numeric(substr(df$acute_date_log, 11, 12)) #remove number, make numeric
names(df)[names(df) == 'acute_date_log'] <- 'acute_week_log' #change name of variable for clarity

#separate timepoint source and week information in 'first_timepoint_log' variable
df <- cbind(df, as.data.frame(matrix(str_split_fixed(df$first_timepoint_log, " Week ", 2), ncol = 2, by
df <- subset(df, select = -first_timepoint_log)
colnames(df)[colnames(df)=="V1"] <- "first_timepoint_log"
colnames(df)[colnames(df)=="V2"] <- "first_week_log"

#remove accidental extra space in character
df$second_timepoint_log <- as.character(df$second_timepoint_log)
df$second_timepoint_log[df$second_timepoint_log == 'Off protocol '] <- 'Off protocol'

#separate timepoint source and week information in 'second_timepoint_log' variable
df <- cbind(df, as.data.frame(matrix(str_split_fixed(df$second_timepoint_log, " Week ", 2), ncol = 2, b
df <- subset(df, select = -second_timepoint_log )
colnames(df)[colnames(df)=="V1"] <- "second_timepoint_log"
colnames(df)[colnames(df)=="V2"] <- "second_week_log"

#recode anything containing 'relapse' in 'second_timepoint_log' variable as simply 'relapse'
df$second_timepoint_log <- as.character(df$second_timepoint_log)
df$second_timepoint_log <- ifelse(grepl('Relapse', df$second_timepoint_log), "Relapse", df$second_timep

#recode anything containing 'Protocol' in 'second_timepoint_log' variable as simply 'off protocol'
df$second_timepoint_log <- ifelse(grepl('Protocol', df$second_timepoint_log), "Off protocol", df$second_

#separate timepoint source and week information in 'third_timepoint_log' variable
df <- cbind(df, as.data.frame(matrix(str_split_fixed(df$third_timepoint_log, " Week ", 2), ncol = 2, by
df <- subset(df, select = -third_timepoint_log )
colnames(df)[colnames(df)=="V1"] <- "third_timepoint_log"
colnames(df)[colnames(df)=="V2"] <- "third_week_log"

#compare dates in df that comes from log vs. XNAT (in new column)
df$first_dateDiff <- round(difftime(df$first_date_log, df$first_date_xnat, units = "days"), 2)
df$second_dateDiff <- round(difftime(df$second_date_log, df$second_date_xnat, units = "days"), 2)
df$third_dateDiff <- round(difftime(df$third_date_log, df$third_date_xnat, units = "days"), 2)

#make sure new variables are characters (not factors), and turn blank values into NA
i <- sapply(df, is.factor)
df[i] <- lapply(df[i], as.character)
df[df == ""] <- NA

#calculate the difference in weeks between scan 2 and scan 1 (i.e., calculate 'second week log' when ab
df$dateDiff_first_second <- round(difftime(df$second_date_log, df$first_date_log, units = "weeks"), 0)
df$dateDiff_first_second <- as.numeric(df$dateDiff_first_second) #turn variables into integers
df$first_week_log <- as.numeric(df$first_week_log) #turn variables into integers
df$second_week_log <- ifelse(is.na(df$second_week_log) & !is.na(df$second_timepoint_log), paste(df$date

#reorder df columns
df <- df[c(
  "STUDYID",
  'sex',
  'age',
  "randomization",
  'randomization_date',
  'imaging_consent_date',
  'imaging_nonconsent_reason',
  "acute_week_log",
  "acute_complete_log",
  "first_date_log",
  "first_timepoint_log",
  "first_week_log",
  "first_complete_log",
  'first_notes',
  "second_date_log",
  "second_timepoint_log",
  "second_week_log",
  "second_complete_log",
  'second_notes',
  "third_date_log",
  "third_timepoint_log",
  "third_week_log",
  "third_complete_log",
  'third_notes'
)]

#remove '_log' component of all variable names, for clarity
names(df) = gsub(pattern = "_log", replacement = "", x = names(df))
```

3.1 Exclusions from MR analysis and reasons

subject 320032 (PMC): incidental findings more atrophy, should excluded **subject 410012 (CMH)**: another incidental finding, case may have effected longitudinal brain morphomentry

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
df <- df %>% mutate(MR_exclusion = if_else(STUDYID %in% c("320032", "410012"), "Yes", "No"))
```

```
#make a smaller df of minimally necessary information from participants that completed 2 scans, as requ
write.csv(df, '../generated_csvs/STOPPD_masterDF_2018-11-05.csv', row.names=FALSE)

#remove participants that don't have 'Yes' in 'first_complete'
df <- df %>% filter(first_complete == "Yes") #nrow = 88, which is correct

#remove participants that don't have 'Yes' in 'second_complete'
df <-df %>% filter(second_complete == "Yes") #nrow = 74, which is correct

#remove redundant columns
df <- df %>% select(STUDYID, age, sex, randomization, MR_exclusion, first_timepoint, second_timepoint, 

df <- df %>% dplyr::rename("offlabel_timepoint" = third_timepoint)
```

```
#write.csv
write.csv(df, '../generated_csvs/STOPPD_participantList_2018-11-05.csv', row.names=FALSE)
```

4 Report Randomization numbers

This script identifies the number of participants in olanzapine vs. placebo by scan timepoint, using the logic of group inclusion that Judy and Dielle provided, and Nick and Aristotle have agreed to.

Note: this script includes data from all participants with data in Judy's master log and our file system. It has not excluded participants on any other basis (e.g., QC fail, processing fail, post-hoc clinical trial ineligibility, etc.)

4.1 Identify baseline scans

First - identify the number of baseline scans (i.e., scans completed at week 20).

```
#count the number of participants that have a 'yes' for 'completed' in "Scan.completed.1"
n_first_complete = sum(na.omit(df$first_complete == "Yes")) #88 participants completed week 20 scan

#for clarity, print the IDs of the N=88 participants that completed week 20 scans
(df %>% filter(first_complete == "Yes"))$STUDYID
```

```
##  [1] 110008 110009 110013 110016 110022 110025 110028 110030 110031 110034
## [11] 120011 120012 120015 120016 120017 120021 120026 210012 210013 210014
## [21] 210017 210020 210022 210024 210026 210030 210033 210036 210038 210042
## [31] 210048 210049 210051 220002 220003 220004 220006 220008 220009 310010
## [41] 310015 310025 310037 310051 310070 320006 320013 320021 320022 320032
## [51] 320041 320042 320043 320045 410004 410008 410009 410010 410011 410012
## [61] 410013 410015 410019 410022 410023 410029 410030 410031 410037 410039
## [71] 410040 410043 410045 410047 420005 420007 420013 420016 420018 420019
## [81] 420020 420023 420029 420032 420039 420042 420043 420044
```

The number of participants who completed their first scan is 88

RANDOMIZATION - as expected, there's no difference in first scan completion between those randomized to O vs. P group

```
#RANDOMIZATION - as expected, there's no difference in first scan completion between those randomized t
(R <- addmargins(table(df$first_complete == 'Yes', df$randomization))) #O = 45; P = 43 (total = 88)
```

```
## 
##         O  P Sum
##   TRUE 45 43  88
##   Sum  45 43  88
```

4.2 Identify week 56 scans

Second - identify the number of week 56 scans (i.e., 36 weeks after week 20).

```
#make sure that all the participants that completed week 56 scan also completed week 20
all_second_complete <- all((df$second_complete == "Yes") %in% (df$first_complete== "Yes")) #all TRUE

#count the number of participants that have a 'yes' for 'completed' in "Scan.completed" - but this incl
(n_second_complete <- sum(na.omit(df$second_complete == "Yes"))) #74 completed week 56 scan
```

```
## [1] 74
```

Subject ids of the n = 74 who completed their second scan. Note: it is TRUE that all participants who completed their second scan have baseline data.

```
#for clarity, print the IDs of the N=74 participants that completed week 56 scans
(df %>% filter(second_complete == "Yes"))$STUDYID
```

```
##  [1] 110008 110009 110013 110022 110031 110034 120011 120012 120015 120016
## [11] 120017 120021 120026 210012 210013 210014 210017 210020 210022 210026
## [21] 210030 210033 210038 210042 210049 210051 220002 220003 220004 220006
## [31] 220009 310010 310015 310025 310037 310051 320006 320013 320021 320022
## [41] 320032 320042 320043 320045 410004 410008 410009 410010 410011 410012
## [51] 410013 410015 410019 410022 410023 410029 410030 410031 410037 410039
## [61] 410040 410043 420007 420013 420016 420018 420019 420020 420023 420029
## [71] 420032 420039 420042 420043
```

Table 1: breakdown of those who where scanned at two timepoints

second_timepoint	n
Off protocol	5
RCT	41
Relapse	28

```
#count how many participants that completed week 56 scan are classified as RCT
sum(na.omit(df$second_complete == 'Yes' & df$second_timepoint == 'RCT')) #RCT = 41
```

```
## [1] 41
```

```
  (as.vector(na.omit(df$STUDYID[df$second_complete == "Yes" & df$second_timepoint == 'RCT']))) #for cla
```

```
##  [1] 110008 110009 110013 110022 110031 110034 120011 120012 120015 210012
## [11] 210013 210014 210017 210020 210030 210051 220004 310051 320006 320021
## [21] 320032 320042 320043 320045 410004 410008 410010 410013 410022 410023
## [31] 410029 410030 410037 410039 410043 420013 420020 420029 420039 420042
## [41] 420043
```

```
sum(na.omit(df$second_complete == 'Yes' & df$second_timepoint == 'Relapse')) #Relapse = 28
```

```
## [1] 28
```

```
  (as.vector(na.omit(df$STUDYID[df$second_complete == "Yes" & df$second_timepoint == 'Relapse']))) #for
```

```
##  [1] 120016 120017 120021 120026 210022 210026 210033 210038 210042 210049
## [11] 220002 220003 220006 220009 310010 310015 310025 310037 320013 410009
## [21] 410011 410012 410031 410040 420007 420016 420023 420032
```

```
sum(na.omit(df$second_complete == 'Yes' & df$second_timepoint == 'Off protocol')) #Off protocol = 5
```

```
## [1] 5
```

```
  (as.vector(na.omit(df$STUDYID[df$second_complete == "Yes" & df$second_timepoint == 'Off protocol'])))
```

```
## [1] 320022 410015 410019 420018 420019
```

```
df %>%
  filter(second_complete == "Yes") %>%
  count(second_timepoint) %>%
  kable(caption = "breakdown of those who where scanned at two timepoints")
```

```
#RANDOMIZATION- look at randomization info for those who completed a second timepoint RCT scan
(R <- addmargins(table(df$second_complete == 'Yes' & df$second_timepoint == 'RCT', df$randomization)))
```

```
##
##          O  P Sum
##   FALSE 14 24  38
##   TRUE  27 14  41
##   Sum   41 38  79
```

```
df %>%
  filter(second_complete == "Yes") %>%
  count(second_timepoint, randomization) %>%
  kable(caption = "breakdown of those who where scanned at two timepoints, by arm")
```

Table 2: breakdown of those who where scanned at two timepoints, by arm

second_timepoint	randomization	n
Off protocol	O	4
Off protocol	P	1
RCT	O	27
RCT	P	14
Relapse	O	8
Relapse	P	20

Table 3: Breakdown of thrid timepoint off-label scans 8 total

randomization	n
O	3
P	5

4.3 Identify off label scans

Third - identify the number of "off label" scans also at week 56.

```
#make sure timepoint is a character
df$second_timepoint <- as.character(df$second_timepoint)

#count the number of scans completed at *third* timepoint, which are by definition "off label"
n_offlable <- sum(na.omit(df$third_complete == 'Yes')) #8 off-label scans

#for clarity, print the IDs of the N=8 participants that completed off-label scans
(as.vector(na.omit(df$STUDYID[df$third_complete == "Yes"])))
```

```
## [1] 110016 210033 210049 220006 310037 320022 410019 420032
```

```
#of these, determine how many "off protocol" vs. "relapse", based on second timepoint scan
sum(na.omit(df$third_complete == 'Yes' & df$second_timepoint  == 'Off protocol')) #2 "off protocol" sca
```

```
## [1] 2
```

```
  (as.vector(na.omit(df$STUDYID[df$third_complete == "Yes" & df$second_timepoint  == 'Off protocol'])))
```

```
## [1] 320022 410019
```

```
sum(na.omit(df$third_complete == 'Yes' & df$second_timepoint == 'Relapse')) #6 relapse scans
```

```
## [1] 6
```

```
  (as.vector(na.omit(df$STUDYID[df$third_complete == "Yes" & df$second_timepoint  == 'Relapse'])))
```

```
## [1] 110016 210033 210049 220006 310037 420032
```

```
#RANDOMIZATION
df %>%
  filter(third_complete == "Yes") %>%
  count(randomization) %>%
  kable(caption = str_c("Breakdown of thrid timepoint off-label scans ", n_offlable, " total"))
```

```
df %>%
  filter(df$second_timepoint == 'Off protocol') %>%
```

Table 4: Breakdown of off-protocol scans by presence of third timepoint

randomization	third_complete	n
O	Yes	2
O	NA	2
P	NA	1

Table 5: Breakdown of thrid timepoint 'Relapse' scans by presence of third timepoint

randomization	third_complete	n
O	Yes	1
O	NA	9
P	Yes	5
P	NA	18

```
  count(randomization, third_complete) %>%
  kable(caption = str_c("Breakdown of off-protocol scans by presence of third timepoint"))
```

```
df %>%
  filter(df$second_timepoint == 'Relapse') %>%
  count(randomization, third_complete) %>%
  kable(caption = str_c("Breakdown of thrid timepoint 'Relapse' scans by presence of third timepoint"))
```

4.4 Identify "Relapse" Scans

Identify the scans completed between week 20 and week 56 which are the relapse scans (and in a small minority of cases may be a scan when somebody is moving or wants out of the study despite being well).

```
#count relapse - note: both 'relapse' and 'off protocol' is included here (everything other than 'RCT')
sum(na.omit((df$second_timepoint == 'Relapse' | df$second_timepoint == 'Off protocol') & df$second_compl
```

```
## [1] 33
```

```
#of these, count how many were "relapse" and how many were "off protocol"
sum(na.omit(df$second_timepoint == 'Relapse' & df$second_complete == 'Yes'))# 28 relapse
```

```
## [1] 28
```

```
sum(na.omit(df$second_timepoint == 'Off protocol' & df$second_complete == 'Yes'))#5 off protocol
```

```
## [1] 5
```

```
#RANDOMIZATION
(R <- addmargins(table((df$second_timepoint == 'Relapse' | df$second_timepoint == 'Off protocol') & df$s
```

```
##
##           O  P Sum
##   FALSE 28 15  43
##   TRUE  12 21  33
##   Sum   40 36  76
```

```
(R <- addmargins(table(df$second_timepoint == 'Relapse' & df$second_complete == 'Yes', df$randomization)
```

```
##
```

```
##          O  P Sum
##   FALSE 32 16  48
##   TRUE   8 20  28
##   Sum   40 36  76
```

```
(R <- addmargins(table(df$second_timepoint == 'Off protocol' & df$second_complete == 'Yes', df$randomiza
```

```
##
##          O  P Sum
##   FALSE 38 38  76
##   TRUE   4  1   5
##   Sum   42 39  81
```

```
rm(df, R)
```

5 Mangle Freesurfer Outputs

This script pulls together completion information alongside cortical thickness (CT) values and demographic information, for statistical purposes (error calculations). It is required for subsequent CT analyses. It was made in preparation for, and discussed at, the meeting with Jason Lerch.

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------------

## √ ggplot2 3.1.0     √ purrr   0.2.4
## √ tibble  1.4.1     √ dplyr   0.7.7
## √ tidyr   0.7.2     √ stringr 1.2.0
## √ readr   1.1.1     √ forcats 0.2.0

## -- Conflicts ------------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
df <- read_csv("../generated_csvs/STOPPD_masterDF_2018-11-05.csv",na = "empty") #spreadsheet created by
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   STUDYID = col_integer()
## )

## See spec(...) for full column specifications.
```

```
CT <- read_csv('../data/fs-enigma-long_201811/CorticalMeasuresENIGMA_ThickAvg.csv') #bring in CT data,
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   SubjID = col_character()
## )
## See spec(...) for full column specifications.
```

```
# remove participants that did not complete first and second scan (n=74)
# then add offlabel and dateDiff (in days columns)
# + a scan is by definition offlabel if it is the third scan
# then select the cols for analysis
```

```
df <- df %>%
  filter(first_complete == "Yes",
         second_complete == "Yes",
         MR_exclusion == "No") %>%
  mutate(offLabel  = if_else(third_complete == "Yes", "Yes", ''),
         dateDiff = round(difftime(second_date, first_date, units = "days"), 0),
         STUDYID = parse_character(STUDYID)) %>%
  rename(category = "second_timepoint") %>%
  select(STUDYID, randomization, sex, age, category, offLabel, dateDiff)
```

5.1 cleaning the CT data

```
# separating the subject id and anything afterwards to identify the longtudinal pipeline participants
# separating the subject id into site, "STUDYID" and timepoint columns
# filtering (two steps) to only include the longitudinal pipeline data
CT_long <- CT %>%
  separate(SubjID, into = c("subid", "longitudinal_pipe"), sep = '\\.', extra = "drop", fill = "right")
  separate(subid, into = c("study", "site", "STUDYID", "timepoint"), fill = "right") %>%
  filter(longitudinal_pipe == "long") %>%
  filter(timepoint != "00", timepoint != "03", timepoint != "")


# move CT from long to wide format
CT_wide <- CT_long %>%
  gather(region, thickness, ends_with('thickavg'), LThickness, RThickness, LSurfArea, RSurfArea, ICV) %
  spread(timepoint, thickness) %>%
  mutate(change = `02` - `01`) %>%
  gather(timepoint, thickness, `01`, `02`, change) %>%
  unite(newcolnames, region, timepoint) %>%
  spread(newcolnames, thickness)
```

```
# merge CT values with df
ana_df <- inner_join(df, CT_wide, by='STUDYID')

# write.csv
write_csv(ana_df, '../generated_csvs/STOPPD_participantsCT_20181111.csv')
```

5.2 report any mising values from clinical trial sample

```
anti_join(df, CT_wide, by='STUDYID') %>%
  summarise(`Number of participants missing` = n()) %>%
  knitr::kable()
```

Number of participants missing
0

```
ana_df %>%
  filter(is.na(LThickness_01)) %>%
  summarise(`Number of participants missing timepoint 01` = n()) %>%
  knitr::kable()
```

Number of participants missing timepoint 01
0

```
ana_df %>%
  filter(is.na(LThickness_02)) %>%
  summarise(`Number of participants missing timepoint 02` = n()) %>%
  knitr::kable()
```

Number of participants missing timepoint 02
0

5.3 creating an control error term calculating spreadsheet

```
## identify the repeat control in a column and mangle the STUDYID to match in a new column
CT_long1 <- CT_long %>%
  mutate(repeat_run = if_else(str_sub(STUDYID,1,1)=="R", "02", "01"),
         STUDYID = str_replace(STUDYID, 'R',""))

## extra the repeat study ids as a character vector
repeat_ids <- filter(CT_long1, repeat_run == "02")$STUDYID

## filter for only the subjects who are in the repeats list then switch to wide format
CT_wide_controls <- CT_long1 %>%
  filter(STUDYID %in% repeat_ids) %>%
  gather(region, thickness, ends_with('thickavg'), LThickness, RThickness, LSurfArea, RSurfArea, ICV) %
  unite(newcolnames, region, repeat_run) %>%
  spread(newcolnames, thickness)

#write.csv
  write.csv(CT_wide_controls, '../generated_csvs/STOPPD_errorControls_2018-11-05.csv', row.names = FALS
```

6 Cortical Thickness Analysis

This section runs the stats for average (by hemisphere) Cortical Thickness calculated with Freesurfer

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


#bring in data
df <- read_csv('../generated_csvs/STOPPD_participantsCT_20181111.csv') #generated by 05_STOPPD_error in
```

```
#make sure that STUDYID is an interger not a number
  df$STUDYID <- as.character(df$STUDYID)

#make sure that dateDiff is a number, not an interger
  df$dateDiff <- as.numeric(df$dateDiff)

# label the randomization variable
df$RandomArm <- factor(df$randomization,
                         levels = c("O", "P"),
                         labels = c("Olanzapine", "Placebo"))

#restructure data for RCT completers' only (N=40)
  RCT_CT <- df %>%
    filter(category == "RCT")



#write out clean dataframe
 # write.csv(RCT_CT, '../generated_data/df_leftCT.csv', row.names=FALSE)
```

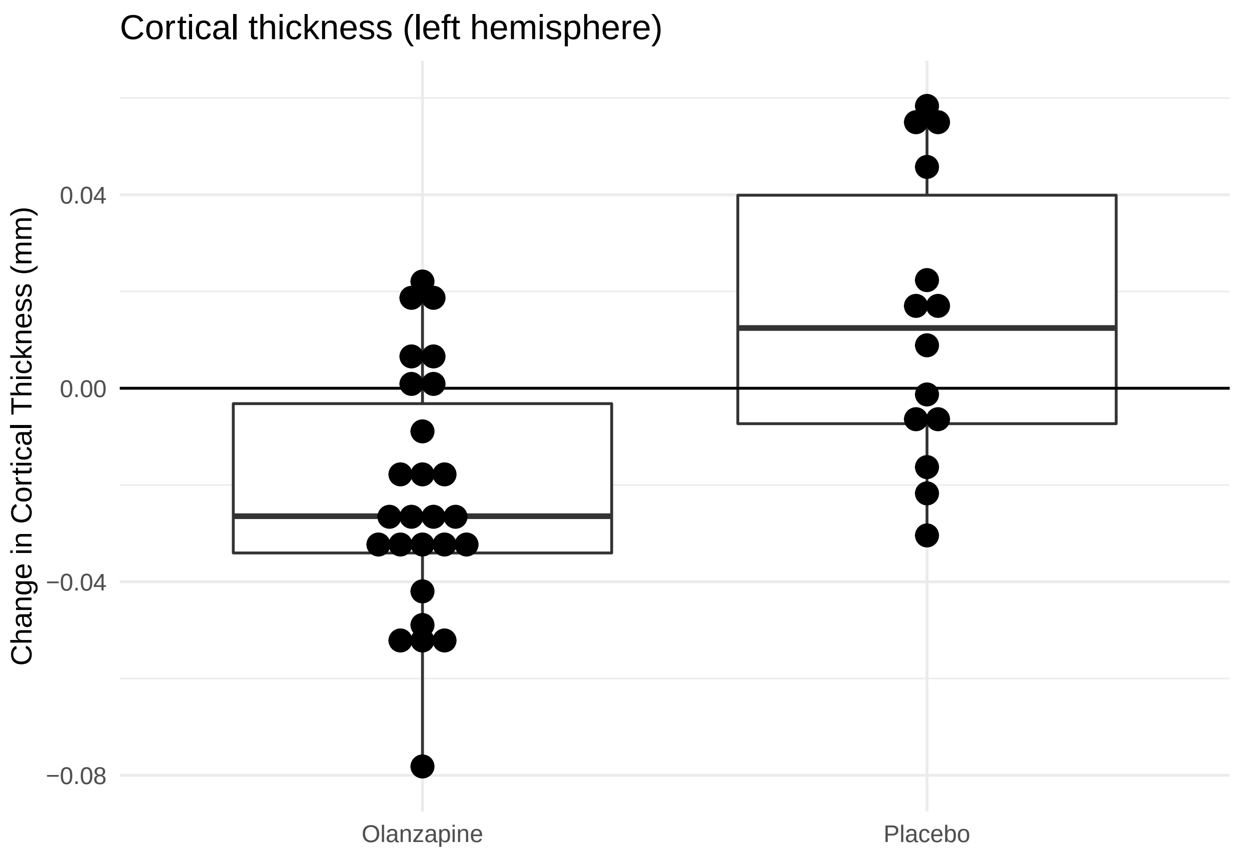
6.1 RCT only

```
RCT_CT %>% count(randomization) %>% knitr::kable()
```

randomization	n
O	26
P	14

```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
ggplot(RCT_CT, aes(x= RandomArm, y = LThickness_change)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     ggtitle("Cortical thickness (left hemisphere)") +
     xlab(NULL) +
     ylab("Change in Cortical Thickness (mm)") +
     theme_minimal()
```

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



```
#run linear model without covariates
  fit_rct <- lm(LThickness_change ~ RandomArm, data= RCT_CT)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = LThickness_change ~ RandomArm, data = RCT_CT)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo
##         -0.02253           0.03659
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = LThickness_change ~ RandomArm, data = RCT_CT)
##
## Residuals:
##       Min         1Q    Median         3Q        Max
## -0.055651 -0.018822 -0.003541  0.022344  0.044589
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.022529   0.005294  -4.255 0.000131 ***
## RandomArmPlacebo  0.036591   0.008949   4.089 0.000217 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.027 on 38 degrees of freedom
## Multiple R-squared:  0.3055, Adjusted R-squared:  0.2873
## F-statistic: 16.72 on 1 and 38 DF,  p-value: 0.0002168
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(LThickness_change ~ RandomArm + sex + age, data= RCT_CT)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = LThickness_change ~ RandomArm + sex + age, data = RCT_CT)
##
## Coefficients:
##       (Intercept)  RandomArmPlacebo              sexM               age
##        -0.0053535         0.0389342         0.0087658        -0.0004063
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = LThickness_change ~ RandomArm + sex + age, data = RCT_CT)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.054543 -0.019130 -0.001897  0.019599  0.050858
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.0053535  0.0168885  -0.317 0.753081
## RandomArmPlacebo  0.0389342  0.0090734   4.291 0.000128 ***
## sexM              0.0087658  0.0087344   1.004 0.322272
## age              -0.0004063  0.0003132  -1.297 0.202857
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02691 on 36 degrees of freedom
## Multiple R-squared:  0.3465, Adjusted R-squared:  0.292
## F-statistic: 6.363 on 3 and 36 DF,  p-value: 0.001424
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(LThickness_change ~ RandomArm + sex + age + site, data= RCT_CT)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = LThickness_change ~ RandomArm + sex + age + site,
##     data = RCT_CT)
##
## Coefficients:
##       (Intercept)  RandomArmPlacebo              sexM               age
##         0.0003982         0.0404923         0.0127657        -0.0003164
##           siteMAS           siteNKI           sitePMC
##        -0.0193116        -0.0212567        -0.0287115
```

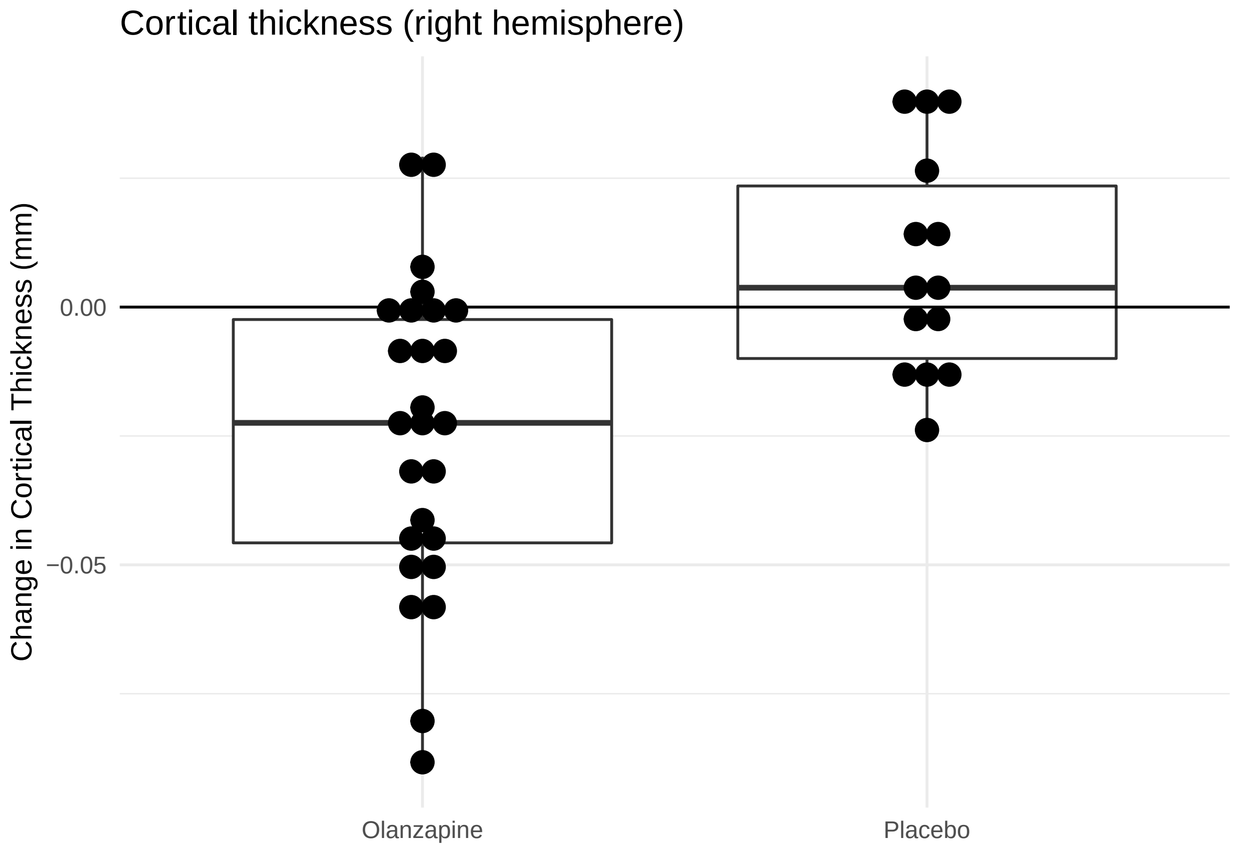
```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = LThickness_change ~ RandomArm + sex + age + site,
##     data = RCT_CT)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.045029 -0.013484 -0.001428  0.017054  0.052285
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.0003982  0.0177658   0.022   0.9823
## RandomArmPlacebo  0.0404923  0.0086868   4.661 4.99e-05 ***
## sexM              0.0127657  0.0083335   1.532   0.1351
## age              -0.0003164  0.0003244  -0.975   0.3365
## siteMAS          -0.0193116  0.0110783  -1.743   0.0906 .
## siteNKI          -0.0212567  0.0105872  -2.008   0.0529 .
## sitePMC          -0.0287115  0.0130066  -2.207   0.0343 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02527 on 33 degrees of freedom
## Multiple R-squared:  0.4717, Adjusted R-squared:  0.3757
## F-statistic: 4.911 on 6 and 33 DF,  p-value: 0.001094
```

6.1.1 looking at the same thing for Right CT

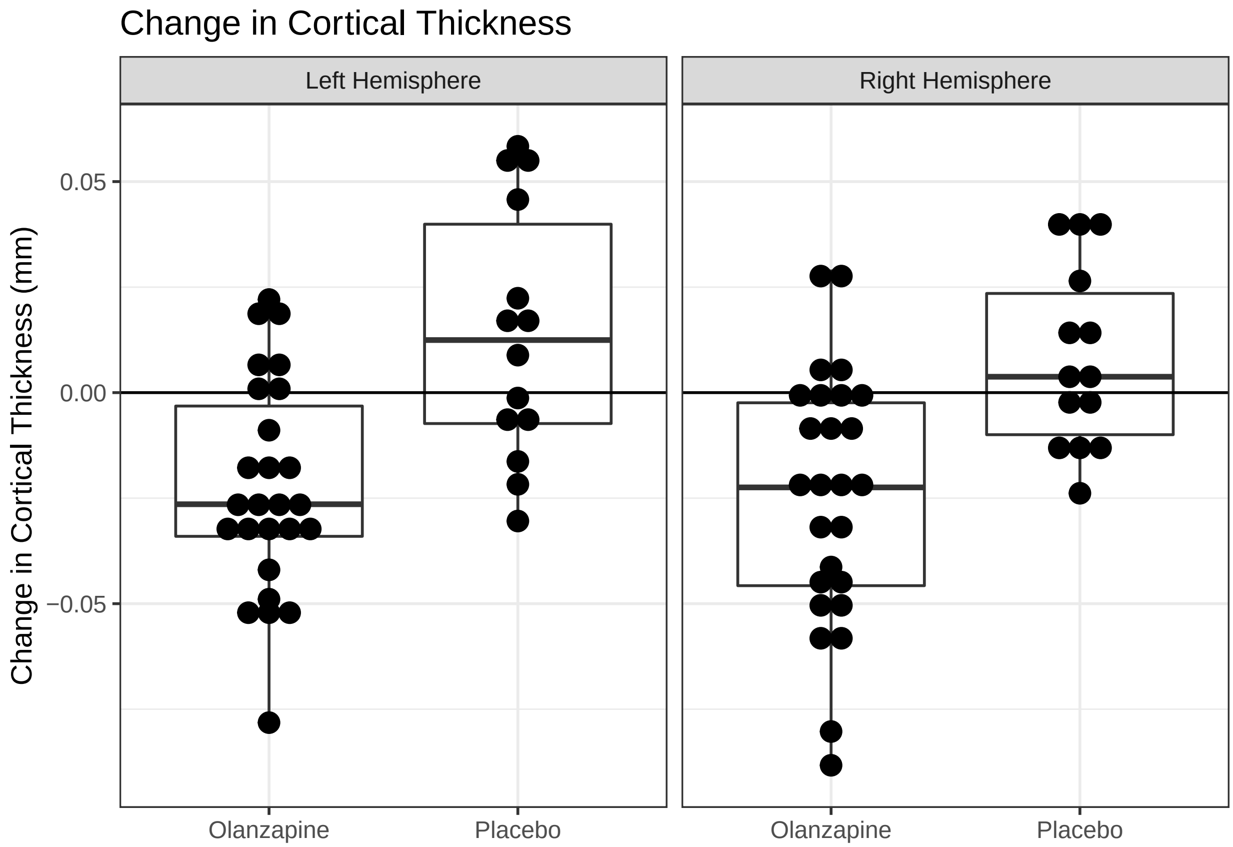
```r
#boxplot of difference in thickness (y axis) by randomization group (x axis)
ggplot(RCT_CT, aes(x= RandomArm, y = RThickness_change)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     ggtitle("Cortical thickness (right hemisphere)") +
     xlab(NULL) +
     ylab("Change in Cortical Thickness (mm)") +
     theme_minimal()
```

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



```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
RCT_CT %>%
  gather(TCT, mm, LThickness_change, RThickness_change) %>%
  mutate(ThickChange = factor(TCT, levels = c("LThickness_change", "RThickness_change"),
                              labels = c("Left Hemisphere", "Right Hemisphere"))) %>%
ggplot(aes(x= RandomArm, y = mm)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     labs(title = "Change in Cortical Thickness",x = NULL, y = "Change in Cortical Thickness (mm)") +
     facet_wrap(~ ThickChange) +
     theme_bw()
```

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



```
#run linear model without covariates
  fit_rct <- lm(RThickness_change ~ RandomArm, data= RCT_CT)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = RThickness_change ~ RandomArm, data = RCT_CT)
##
## Coefficients:
##     (Intercept)  RandomArmPlacebo
##        -0.02434           0.03260
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = RThickness_change ~ RandomArm, data = RCT_CT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06395 -0.02029  0.00013  0.02178  0.05351
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.024340   0.005360  -4.542  5.5e-05 ***
## RandomArmPlacebo  0.032596   0.009059   3.598 0.000912 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.02733 on 38 degrees of freedom
## Multiple R-squared:  0.2541, Adjusted R-squared:  0.2345
## F-statistic: 12.95 on 1 and 38 DF,  p-value: 0.0009116
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(RThickness_change ~ RandomArm + sex + age, data= RCT_CT)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = RThickness_change ~ RandomArm + sex + age, data = RCT_CT)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age
##       -0.0097657         0.0336641        -0.0044117        -0.0002401
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = RThickness_change ~ RandomArm + sex + age, data = RCT_CT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06700 -0.01799 -0.00128  0.01847  0.04950
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.0097657  0.0173882  -0.562 0.577851
## RandomArmPlacebo  0.0336641  0.0093419   3.604 0.000941 ***
## sexM             -0.0044117  0.0089928  -0.491 0.626698
## age              -0.0002401  0.0003225  -0.744 0.461500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0277 on 36 degrees of freedom
## Multiple R-squared:  0.2739, Adjusted R-squared:  0.2134
## F-statistic: 4.527 on 3 and 36 DF,  p-value: 0.008575
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(RThickness_change ~ RandomArm + sex + age + site, data= RCT_CT)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = RThickness_change ~ RandomArm + sex + age + site,
##     data = RCT_CT)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age
##       -1.111e-02         3.389e-02        -1.632e-03        -9.043e-05
##          siteMAS           siteNKI           sitePMC
##       -1.010e-02        -1.044e-02        -2.463e-02
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = RThickness_change ~ RandomArm + sex + age + site,
##     data = RCT_CT)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.072836 -0.015554  0.000948  0.019346  0.044262
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.111e-02  1.939e-02  -0.573  0.57042
## RandomArmPlacebo  3.389e-02  9.480e-03   3.575  0.00111 **
## sexM             -1.632e-03  9.095e-03  -0.179  0.85869
## age              -9.043e-05  3.540e-04  -0.255  0.79998
## siteMAS          -1.010e-02  1.209e-02  -0.835  0.40968
## siteNKI          -1.044e-02  1.155e-02  -0.903  0.37287
## sitePMC          -2.463e-02  1.419e-02  -1.735  0.09208 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02758 on 33 degrees of freedom
## Multiple R-squared:  0.3405, Adjusted R-squared:  0.2206
## F-statistic:  2.84 on 6 and 33 DF,  p-value: 0.02429
```

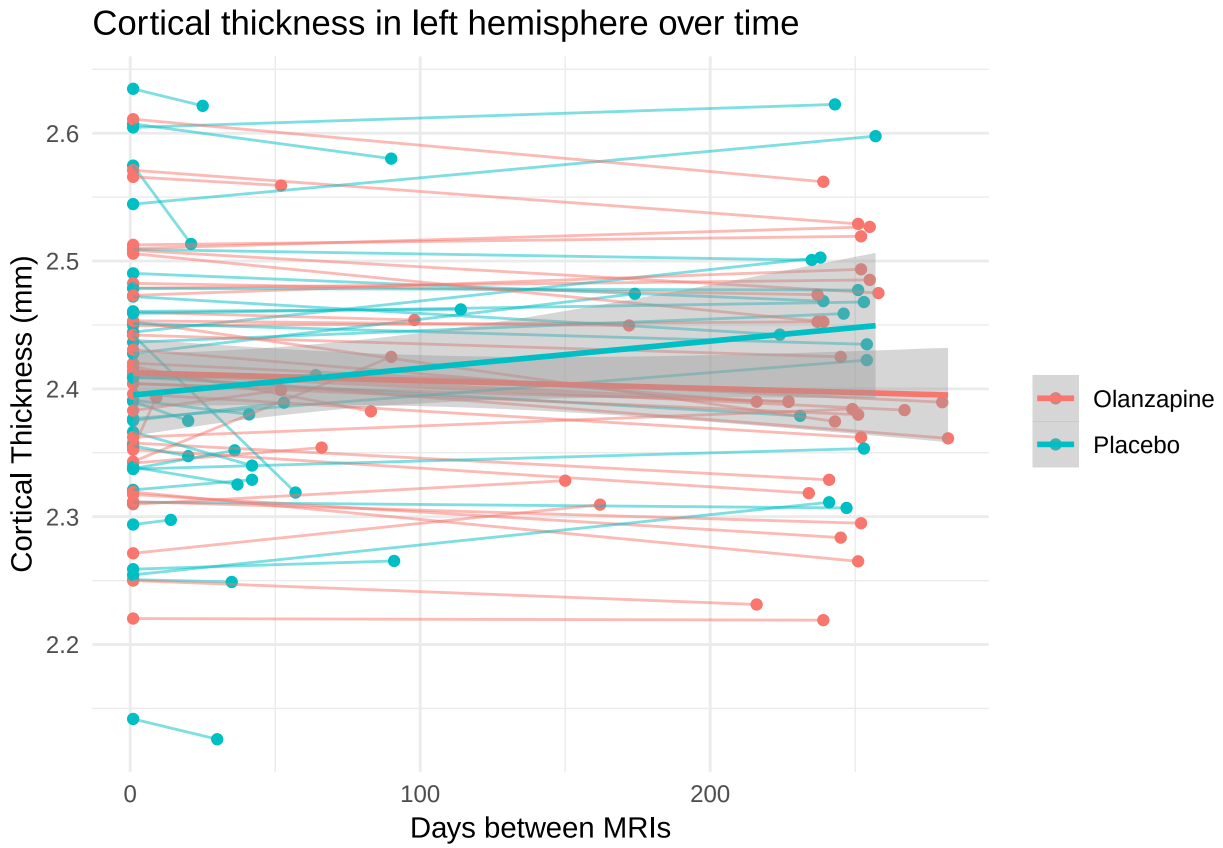
6.2 RCT & Relapse (with time as factor)

```
#restructure data for RCT & Relapse participants (N=72)
  RCTRelapse_LCT <- df %>%
    gather(thick_oldcolname, thickness, LThickness_01, LThickness_02) %>%
    mutate(model_days = if_else(thick_oldcolname == "LThickness_01", 1, dateDiff))

RCTRelapse_LCT %>% filter(model_days == 1) %>% count(randomization) %>% knitr::kable()
```

randomization	n
O	38
P	34

```
#plot all data, including outlier (participant 210030)
  RCTRelapse_LCT %>%
   ggplot(aes(x=model_days, y=thickness, colour=RandomArm)) +
   geom_point() +
   geom_line(aes(group=STUDYID), alpha = 0.5) +
   geom_smooth(method="lm", formula=y~poly(x,1)) +
   ggtitle("Cortical thickness in left hemisphere over time") +
   labs(x = "Days between MRIs", y = "Cortical Thickness (mm)", colour = NULL) +
   theme_minimal()
```



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

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: thickness ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_LCT
##
## REML criterion at convergence: -396
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.89073 -0.39603 -0.02082  0.40944  2.76834
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  STUDYID  (Intercept) 0.0054535 0.07385
##  Residual             0.0004953 0.02225
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                               Estimate Std. Error         df t value
## (Intercept)                  2.639e+00  3.484e-02  6.864e+01  75.756
## RandomArmPlacebo            -2.035e-03  1.816e-02  7.233e+01  -0.112
## model_days                  -8.012e-05  2.340e-05  7.056e+01  -3.424
## sexM                        -6.099e-03  1.792e-02  6.784e+01  -0.340
## age                           -4.053e-03  5.853e-04  6.785e+01  -6.924
## RandomArmPlacebo:model_days  1.297e-04  3.942e-05  7.148e+01   3.291
##                              Pr(>|t|)
## (Intercept)                   < 2e-16 ***
## RandomArmPlacebo              0.91106
## model_days                    0.00103 **
## sexM                          0.73470
## age                          1.96e-09 ***
## RandomArmPlacebo:model_days  0.00155 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age
## RndmArmPlcb -0.206
## model_days  -0.077  0.133
## sexM        -0.171  0.036  0.001
## age         -0.901 -0.053  0.008 -0.079
## RndmArmPl:_  0.043 -0.176 -0.594  0.004 -0.003
```

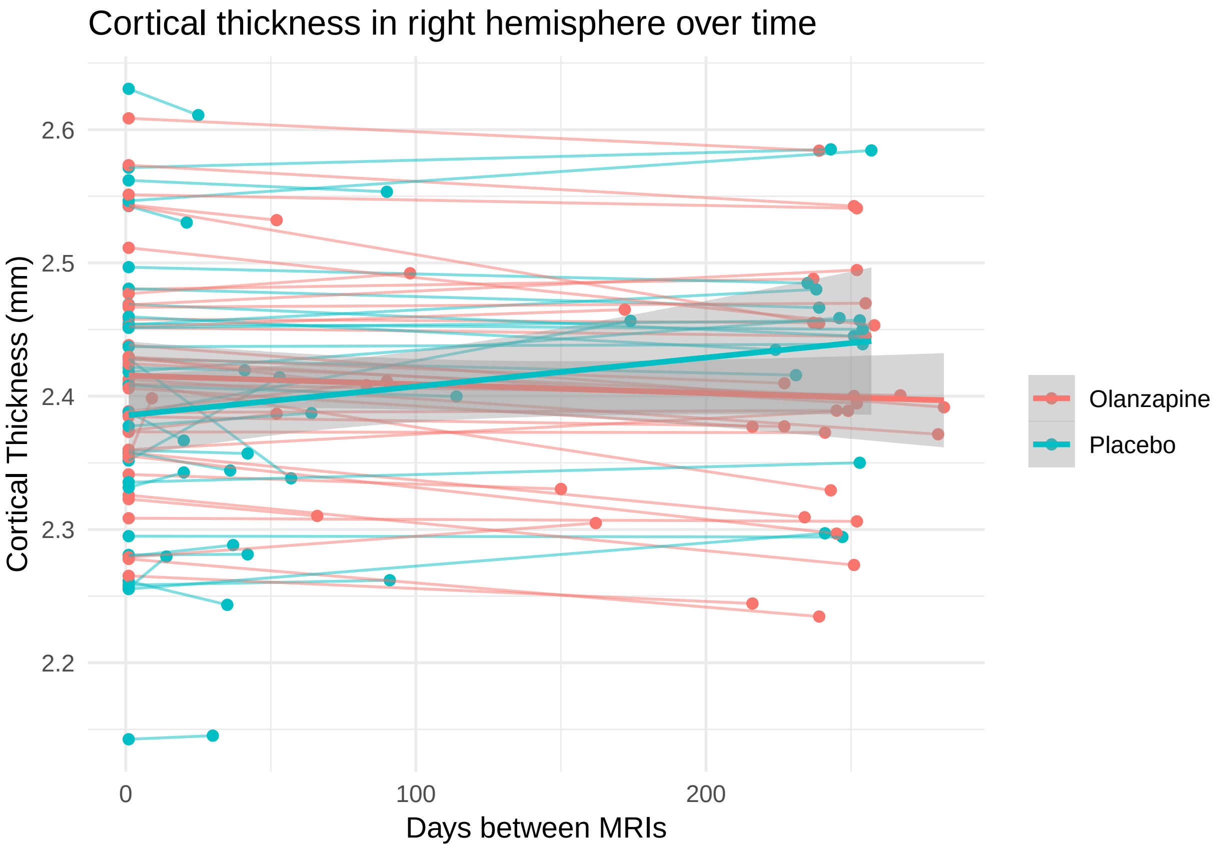
```
#run mixed linear model, with covariates
  fit_all <- lmer(thickness ~ RandomArm*model_days + sex + age + site + (1|STUDYID), data= RCTRelapse_L
  summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: thickness ~ RandomArm * model_days + sex + age + site + (1 |
##     STUDYID)
##    Data: RCTRelapse_LCT
##
## REML criterion at convergence: -380.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.93368 -0.37860 -0.00129  0.40530  2.72488
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  STUDYID  (Intercept) 0.0055987 0.07482
##  Residual             0.0004952 0.02225
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                                Estimate Std. Error         df t value
## (Intercept)                   2.640e+00  3.703e-02  6.556e+01  71.309
## RandomArmPlacebo             -1.963e-03  1.863e-02  6.894e+01  -0.105
## model_days                   -8.021e-05  2.341e-05  7.051e+01  -3.427
## sexM                         -8.192e-03  1.840e-02  6.485e+01  -0.445
## age                          -4.101e-03  5.970e-04  6.486e+01  -6.869
## siteMAS                      -6.325e-03  2.335e-02  6.486e+01  -0.271
## siteNKI                       1.359e-04  2.587e-02  6.486e+01   0.005
## sitePMC                       2.516e-02  2.662e-02  6.485e+01   0.945
## RandomArmPlacebo:model_days  1.296e-04  3.942e-05  7.143e+01   3.286
##                              Pr(>|t|)
## (Intercept)                   < 2e-16 ***
## RandomArmPlacebo             0.91640
## model_days                   0.00102 **
## sexM                         0.65759
## age                         2.98e-09 ***
## siteMAS                      0.78731
## siteNKI                      0.99582
## sitePMC                      0.34814
## RandomArmPlacebo:model_days  0.00158 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## RndmArmPlcb -0.145
## model_days  -0.074  0.129
## sexM        -0.130  0.054  0.002
## age         -0.880 -0.064  0.009 -0.076
## siteMAS     -0.291 -0.145  0.009 -0.066  0.108
## siteNKI     -0.174 -0.119 -0.010 -0.119  0.010  0.357
## sitePMC     -0.153 -0.088  0.000 -0.144 -0.009  0.343  0.319
## RndmArmPl:_  0.040 -0.172 -0.594  0.003 -0.003  0.000  0.009  0.002
```

6.2.1 Running the right hemisphere RCTRelapse

```
#restructure data for RCT & Relapse participants (N=72)
  RCTRelapse_RCT <- df %>%
    gather(thick_oldcolname, thickness, RThickness_01, RThickness_02) %>%
    mutate(model_days = if_else(thick_oldcolname == "RThickness_01", 1, dateDiff))

#plot all data, including outlier (participant 210030)
  RCTRelapse_RCT %>%
   ggplot(aes(x=model_days, y=thickness, colour=RandomArm)) +
   geom_point() +
   geom_line(aes(group=STUDYID), alpha = 0.5) +
   geom_smooth(method="lm", formula=y~poly(x,1)) +
   ggtitle("Cortical thickness in right hemisphere over time") +
   labs(x = "Days between MRIs", y = "Cortical Thickness (mm)", colour = NULL) +
   theme_minimal()
```



```
#run mixed linear model, with covariates
  fit_all <- lmer(thickness ~ RandomArm*model_days + sex + age + (1|STUDYID), data= RCTRelapse_RCT)
  summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: thickness ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_RCT
##
## REML criterion at convergence: -409
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.34720 -0.42608 -0.01215  0.43733  2.27881
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  STUDYID  (Intercept) 0.0057442 0.07579
##  Residual             0.0003947 0.01987
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                               Estimate Std. Error         df t value
## (Intercept)                  2.618e+00  3.554e-02  6.847e+01  73.658
## RandomArmPlacebo            -1.455e-02  1.847e-02  7.131e+01  -0.788
## model_days                  -8.813e-05  2.090e-05  7.041e+01  -4.216
## sexM                        -7.789e-03  1.830e-02  6.786e+01  -0.426
## age                          -3.588e-03  5.975e-04  6.786e+01  -6.004
## RandomArmPlacebo:model_days  1.281e-04  3.524e-05  7.112e+01   3.635
##                             Pr(>|t|)    
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo            0.433361    
## model_days                  7.28e-05 ***
## sexM                        0.671706    
## age                         8.40e-08 ***
## RandomArmPlacebo:model_days 0.000522 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age   
## RndmArmPlcb -0.205                            
## model_days  -0.067  0.117                     
## sexM        -0.172  0.036  0.001              
## age         -0.902 -0.053  0.007 -0.079       
## RndmArmPl:_  0.037 -0.155 -0.593  0.003 -0.002
```

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

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: thickness ~ RandomArm * model_days + sex + age + site + (1 |  
##     STUDYID)
##    Data: RCTRelapse_RCT
## 
## REML criterion at convergence: -393.2
## 
## Scaled residuals: 
##      Min       1Q   Median       3Q      Max 
## -2.37660 -0.44552 -0.00537  0.43115  2.24882 
## 
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  STUDYID  (Intercept) 0.0059298 0.07701 
##  Residual             0.0003947 0.01987 
## Number of obs: 144, groups:  STUDYID, 72
## 
## Fixed effects:
##                               Estimate Std. Error         df t value
## (Intercept)                  2.619e+00  3.788e-02  6.541e+01  69.150
## RandomArmPlacebo            -1.446e-02  1.900e-02  6.800e+01  -0.761
## model_days                  -8.825e-05  2.091e-05  7.037e+01  -4.221
## sexM                        -9.731e-03  1.883e-02  6.487e+01  -0.517
## age                         -3.638e-03  6.112e-04  6.487e+01  -5.952
## siteMAS                     -7.805e-03  2.390e-02  6.488e+01  -0.327
## siteNKI                      4.352e-03  2.649e-02  6.488e+01   0.164
## sitePMC                      2.024e-02  2.725e-02  6.487e+01   0.743
## RandomArmPlacebo:model_days  1.280e-04  3.524e-05  7.108e+01   3.631
##                             Pr(>|t|)    
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo            0.449288
## model_days                  7.14e-05 ***
## sexM                        0.607158
## age                         1.18e-07 ***
## siteMAS                     0.745036
## siteNKI                     0.870002
## sitePMC                     0.460300
## RandomArmPlacebo:model_days 0.000529 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## RndmArmPlcb -0.144
## model_days  -0.064  0.113
## sexM        -0.130  0.055  0.002
## age         -0.881 -0.065  0.008 -0.076
## siteMAS     -0.291 -0.146  0.007 -0.066  0.108
## siteNKI     -0.175 -0.119 -0.008 -0.119  0.010  0.357
## sitePMC     -0.153 -0.089  0.000 -0.144 -0.009  0.343  0.319
## RndmArmPl:_  0.035 -0.151 -0.593  0.003 -0.003  0.000  0.008  0.002
```

7 Surface Area Analysis

This script analyses hemisphere wide surface area

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


#bring in data
df <- read_csv('../generated_csvs/STOPPD_participantsCT_20181111.csv') #generated by 05_STOPPD_error in
```

```
#make sure that STUDYID is an interger not a number
  df$STUDYID <- as.character(df$STUDYID)

#make sure that dateDiff is a number, not an interger
  df$dateDiff <- as.numeric(df$dateDiff)

# label the randomization variable
df$RandomArm <- factor(df$randomization,
                       levels = c("O", "P"),
                       labels = c("Olanzapine", "Placebo"))

#restructure data for RCT completers' only (N=40)
  RCT_SA <- df %>%
    filter(category == "RCT")

#write out clean dataframe
 # write.csv(RCT_CT, '../generated_data/df_leftCT.csv', row.names=FALSE)
```

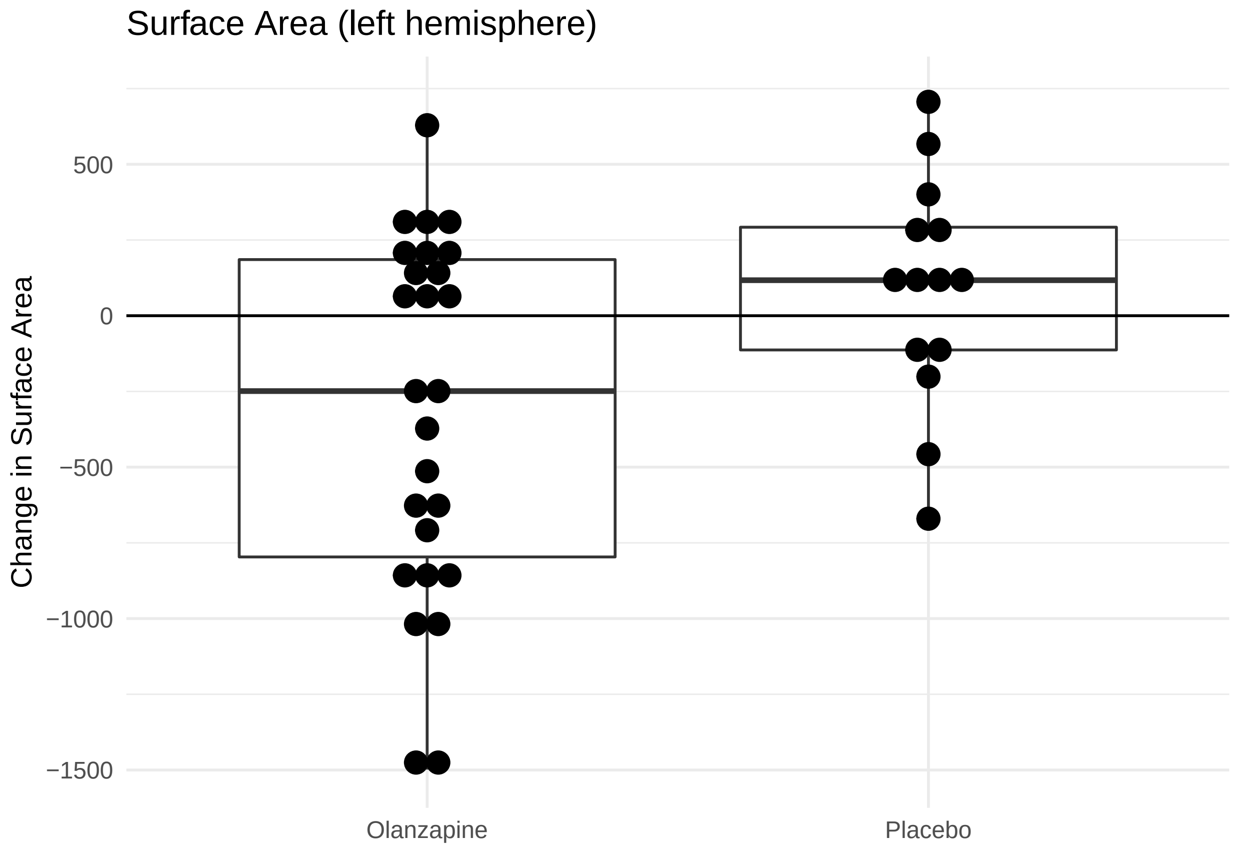
7.1 RCT only

```
RCT_SA %>% count(randomization) %>% knitr::kable()
```

randomization	n
O	26
P	14

```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
ggplot(RCT_SA, aes(x= RandomArm, y = LSurfArea_change)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     ggtitle("Surface Area (left hemisphere)") +
     xlab(NULL) +
     ylab("Change in Surface Area") +
     theme_minimal()
```

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



```
#run linear model without covariates
  fit_rct <- lm(LSurfArea_change ~ RandomArm, data= RCT_SA)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = LSurfArea_change ~ RandomArm, data = RCT_SA)
##
```

```
## Coefficients:
##      (Intercept)  RandomArmPlacebo  
##           -316.6             399.2
```

```r
  summary(fit_rct)
```

```
## 
## Call:
## lm(formula = LSurfArea_change ~ RandomArm, data = RCT_SA)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -1182.2  -333.1    43.8   441.8   945.5 
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)   
## (Intercept)        -316.6      103.1   -3.07  0.00394 **
## RandomArmPlacebo    399.2      174.3    2.29  0.02763 * 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 525.8 on 38 degrees of freedom
## Multiple R-squared:  0.1213, Adjusted R-squared:  0.09819 
## F-statistic: 5.246 on 1 and 38 DF,  p-value: 0.02763
```

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

```
## 
## Call:
## lm(formula = LSurfArea_change ~ RandomArm + sex + age, data = RCT_SA)
## 
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age  
##           525.53            477.83            -37.09            -15.79
```

```r
  summary(fit_rct)
```

```
## 
## Call:
## lm(formula = LSurfArea_change ~ RandomArm + sex + age, data = RCT_SA)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1013.00  -252.25    52.49   330.08   994.27 
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)   
## (Intercept)        525.53     305.14   1.722  0.09360 . 
## RandomArmPlacebo   477.83     163.94   2.915  0.00609 **
## sexM               -37.09     157.81  -0.235  0.81551   
## age                -15.79       5.66  -2.791  0.00836 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

```
## Residual standard error: 486.1 on 36 degrees of freedom
## Multiple R-squared:  0.2884, Adjusted R-squared:  0.229
## F-statistic: 4.862 on 3 and 36 DF,  p-value: 0.006104
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(LSurfArea_change ~ RandomArm + sex + age + site, data= RCT_SA)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = LSurfArea_change ~ RandomArm + sex + age + site,
##     data = RCT_SA)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age
##          727.113           508.771           -74.066           -20.838
##          siteMAS           siteNKI           sitePMC
##           -7.221            12.760           500.924
```

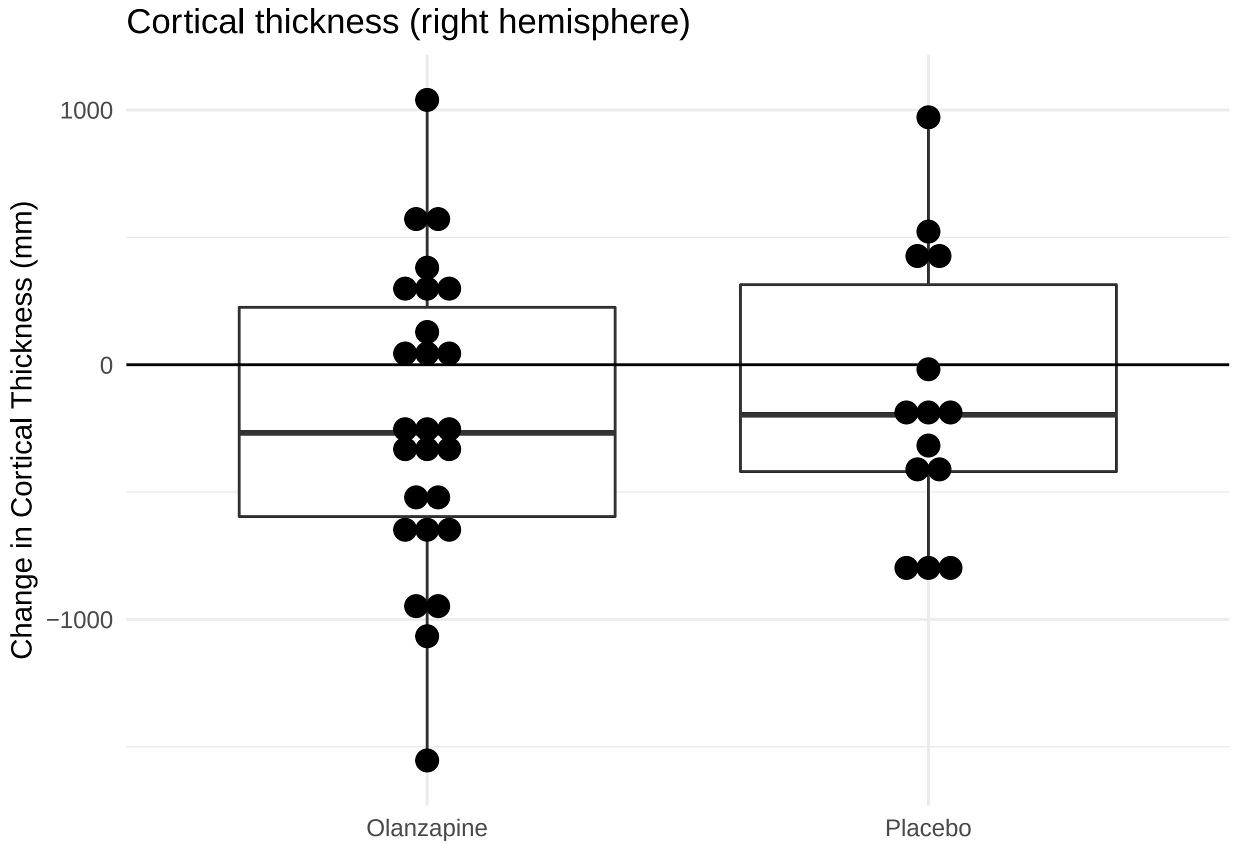
```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = LSurfArea_change ~ RandomArm + sex + age + site,
##     data = RCT_SA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1053.2  -327.3    74.4   383.7   670.0
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       727.113    334.417   2.174  0.03696 *
## RandomArmPlacebo  508.771    163.517   3.111  0.00383 **
## sexM              -74.066    156.867  -0.472  0.63992
## age               -20.838      6.106  -3.412  0.00172 **
## siteMAS            -7.221    208.534  -0.035  0.97259
## siteNKI            12.760    199.290   0.064  0.94934
## sitePMC           500.924    244.831   2.046  0.04879 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 475.6 on 33 degrees of freedom
## Multiple R-squared:  0.3756, Adjusted R-squared:  0.262
## F-statistic: 3.308 on 6 and 33 DF,  p-value: 0.01163
```

7.1.1 looking at the same thing for Right SA

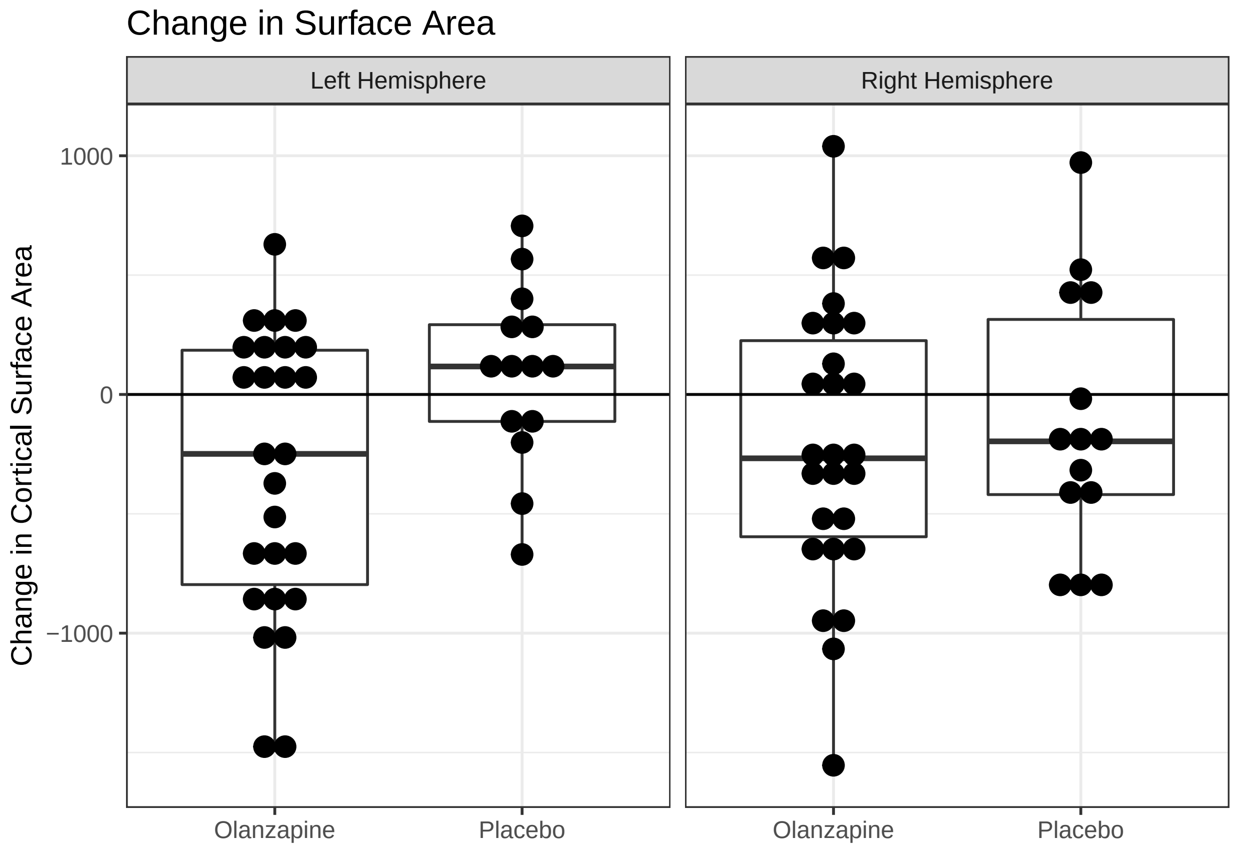
```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
ggplot(RCT_SA, aes(x= RandomArm, y = RSurfArea_change)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     ggtitle("Cortical thickness (right hemisphere)") +
     xlab(NULL) +
     ylab("Change in Cortical Thickness (mm)") +
     theme_minimal()
```

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



```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
RCT_SA %>%
  gather(TCT, mm, LSurfArea_change, RSurfArea_change) %>%
  mutate(ThickChange = factor(TCT, levels = c("LSurfArea_change", "RSurfArea_change"),
                              labels = c("Left Hemisphere", "Right Hemisphere"))) %>%
ggplot(aes(x= RandomArm, y = mm)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     labs(title = "Change in Surface Area",x = NULL, y = "Change in Cortical Surface Area") +
     facet_wrap(~ ThickChange) +
     theme_bw()
```

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



```
#run linear model without covariates
  fit_rct <- lm(RSurfArea_change ~ RandomArm, data= RCT_SA)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = RSurfArea_change ~ RandomArm, data = RCT_SA)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo
##          -215.28             90.53
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = RSurfArea_change ~ RandomArm, data = RCT_SA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1338.22  -341.50   -63.29   476.45  1254.88
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -215.28     112.33  -1.916   0.0628 .
## RandomArmPlacebo      90.53     189.87   0.477   0.6363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 572.8 on 38 degrees of freedom
## Multiple R-squared:  0.005946,   Adjusted R-squared:  -0.02021
## F-statistic: 0.2273 on 1 and 38 DF,  p-value: 0.6363
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(RSurfArea_change ~ RandomArm + sex + age, data= RCT_SA)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = RSurfArea_change ~ RandomArm + sex + age, data = RCT_SA)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age
##          205.920           143.174           153.063            -9.417
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = RSurfArea_change ~ RandomArm + sex + age, data = RCT_SA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1168.56  -430.29   -71.21   390.09  1433.95
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       205.920    358.036   0.575    0.569
## RandomArmPlacebo  143.174    192.356   0.744    0.462
## sexM              153.063    185.168   0.827    0.414
## age                -9.417      6.641  -1.418    0.165
##
## Residual standard error: 570.4 on 36 degrees of freedom
## Multiple R-squared:  0.06603,    Adjusted R-squared:  -0.0118
## F-statistic: 0.8484 on 3 and 36 DF,  p-value: 0.4766
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(RSurfArea_change ~ RandomArm + sex + age + site, data= RCT_SA)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = RSurfArea_change ~ RandomArm + sex + age + site,
##     data = RCT_SA)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age
##           457.35            181.79            118.38            -15.16
##          siteMAS           siteNKI           sitePMC
##           -42.89            -40.79            523.20
```

```
  summary(fit_rct)
```

```
##
## Call:
```

```
## lm(formula = RSurfArea_change ~ RandomArm + sex + age + site,
##     data = RCT_SA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1019.99  -393.21   -12.28   387.01  1153.36
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       457.345    396.183   1.154   0.2566
## RandomArmPlacebo  181.787    193.718   0.938   0.3549
## sexM              118.375    185.840   0.637   0.5285
## age               -15.158      7.234  -2.095   0.0439 *
## siteMAS           -42.885    247.049  -0.174   0.8632
## siteNKI           -40.794    236.098  -0.173   0.8639
## sitePMC           523.197    290.050   1.804   0.0804 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 563.5 on 33 degrees of freedom
## Multiple R-squared:  0.1646, Adjusted R-squared:  0.0127
## F-statistic: 1.084 on 6 and 33 DF,  p-value: 0.3924
```

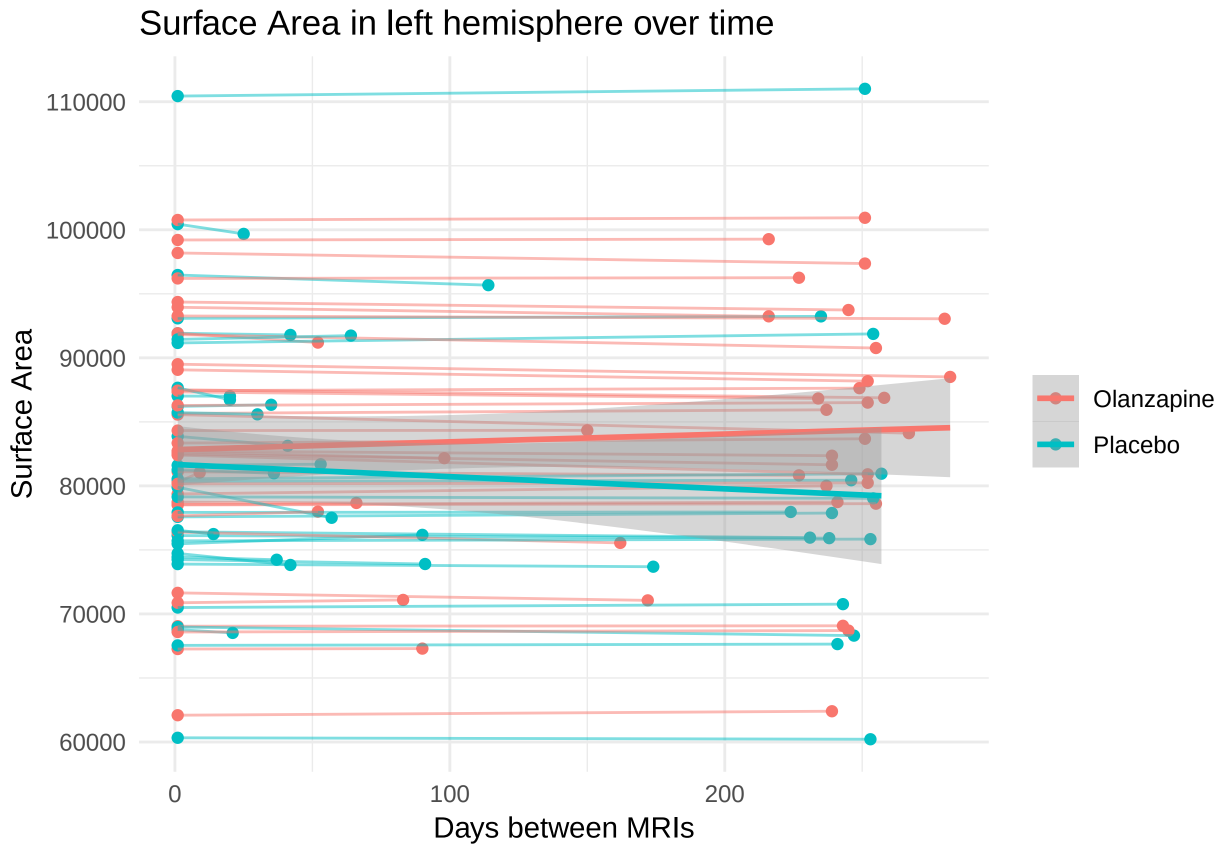
7.2 RCT & Relapse (with time as factor)

```r
#restructure data for RCT & Relapse participants (N=72)
  RCTRelapse_LSA <- df %>%
    gather(oldcolname, SurfArea, LSurfArea_01, LSurfArea_02) %>%
    mutate(model_days = if_else(oldcolname == "LSurfArea_01", 1, dateDiff))

RCTRelapse_LSA %>% filter(model_days == 1) %>% count(randomization) %>% knitr::kable()
```

randomization	n
O	38
P	34

```r
#plot all data, including outlier (participant 210030)
  RCTRelapse_LSA %>%
   ggplot(aes(x=model_days, y=SurfArea, colour=RandomArm)) +
   geom_point() +
   geom_line(aes(group=STUDYID), alpha = 0.5) +
   geom_smooth(method="lm", formula=y~poly(x,1)) +
   ggtitle("Surface Area in left hemisphere over time") +
   labs(x = "Days between MRIs", y = "Surface Area", colour = NULL) +
   theme_minimal()
```



```
#run mixed linear model, with covariates
  fit_all <- lmer(SurfArea ~ RandomArm*model_days + sex + age + (1|STUDYID), data= RCTRelapse_LSA)
  summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: SurfArea ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_LSA
##
## REML criterion at convergence: 2546.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.93057 -0.46073 -0.01316  0.45586  2.94909
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 60349627 7768.5
##  Residual               163892  404.8
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                           Estimate Std. Error        df t value
## (Intercept)             83706.3319  3576.7100   68.0230  23.403
## RandomArmPlacebo        -2037.5203  1839.8200   68.1405  -1.107
## model_days                 -1.2902     0.4267   70.0200  -3.024
## sexM                    11595.6870  1845.3747   67.9977   6.284
## age                          -102.3534    60.2578    67.9978  -1.699
## RandomArmPlacebo:model_days     1.2418     0.7210    70.0491   1.722
##                             Pr(>|t|)    
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo             0.27199    
## model_days                   0.00349 ** 
## sexM                         2.7e-08 ***
## age                          0.09397 .  
## RandomArmPlacebo:model_days  0.08943 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age   
## RndmArmPlcb -0.201                            
## model_days  -0.014  0.024                     
## sexM        -0.172  0.037  0.000              
## age         -0.903 -0.055  0.001 -0.079       
## RndmArmPl:_  0.008 -0.032 -0.592  0.001  0.000
```

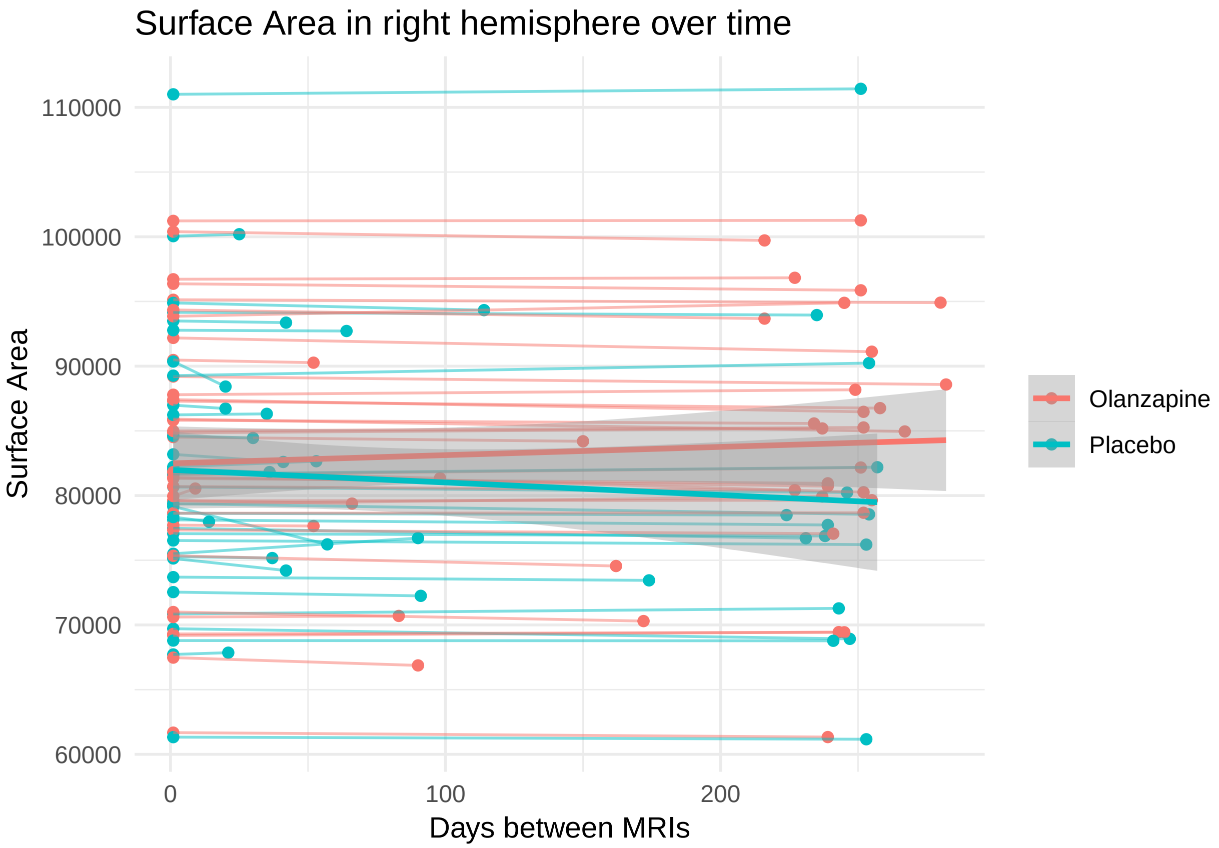
```
#run mixed linear model, with covariates
  fit_all <- lmer(SurfArea ~ RandomArm*model_days + sex + age + site + (1|STUDYID), data= RCTRelapse_LSA
  summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: 
## SurfArea ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_LSA
## 
## REML criterion at convergence: 2482.9
## 
## Scaled residuals: 
##      Min       1Q   Median       3Q      Max 
## -2.95470 -0.46259 -0.00638  0.45970  2.92462 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 52724012 7261.1  
##  Residual               163892  404.8  
## Number of obs: 144, groups:  STUDYID, 72
## 
## Fixed effects:
##                               Estimate Std. Error         df t value
## (Intercept)                 84101.9412  3509.6531    65.0236  23.963
## RandomArmPlacebo            -2019.4088  1744.0368    65.1500  -1.158
## model_days                     -1.2904     0.4267    70.0215  -3.024
## sexM                        10935.7319  1748.4462    64.9973   6.255
## age                          -118.5383    56.7389    64.9976  -2.089
## siteMAS                     -2376.9627  2218.6137    64.9977  -1.071
## siteNKI                      1074.8893  2458.6459    64.9977   0.437
## sitePMC                      7182.3728  2529.8337    64.9972   2.839
## RandomArmPlacebo:model_days     1.2397     0.7210    70.0557   1.719
##                             Pr(>|t|)    
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo             0.25113
## model_days                   0.00348 **
## sexM                        3.52e-08 ***
## age                          0.04061 *
## siteMAS                      0.28796
## siteNKI                      0.66342
## sitePMC                      0.00603 **
## RandomArmPlacebo:model_days  0.08997 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## RndmArmPlcb -0.139
## model_days  -0.014  0.025
## sexM        -0.130  0.055  0.000
## age         -0.882 -0.066  0.002 -0.076
## siteMAS     -0.292 -0.147  0.002 -0.066  0.108
## siteNKI     -0.175 -0.119 -0.002 -0.119  0.010  0.357
## sitePMC     -0.153 -0.089  0.000 -0.144 -0.009  0.343  0.319
## RndmArmPl:_  0.008 -0.034 -0.592  0.001 -0.001  0.000  0.002  0.000
```

7.2.1 Running the right hemisphere RCTRelapse

```
#restructure data for RCT & Relapse participants (N=72)
  RCTRelapse_RSA <- df %>%
    gather(oldcolname, SurfArea, RSurfArea_01, RSurfArea_02) %>%
    mutate(model_days = if_else(oldcolname == "RSurfArea_01", 1, dateDiff))

#plot all data, including outlier (participant 210030)
  RCTRelapse_RSA %>%
   ggplot(aes(x=model_days, y=SurfArea, colour=RandomArm)) +
   geom_point() +
   geom_line(aes(group=STUDYID), alpha = 0.5) +
   geom_smooth(method="lm", formula=y~poly(x,1)) +
   ggtitle("Surface Area in right hemisphere over time") +
   labs(x = "Days between MRIs", y = "Surface Area", colour = NULL) +
   theme_minimal()
```



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

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: SurfArea ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_RSA
##
## REML criterion at convergence: 2567.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1080 -0.3529 -0.0027  0.3822  3.1187
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 61430078 7837.7
##  Residual               214731  463.4
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                             Estimate Std. Error        df t value
## (Intercept)               83306.0885  3609.3846   68.0297  23.080
## RandomArmPlacebo          -1357.7075  1856.8569   68.1809  -0.731
## model_days                   -1.0748     0.4884   70.0260  -2.201
## sexM                      11573.0016  1862.1832   67.9972   6.215
## age                          -101.0096    60.8067    67.9974  -1.661
## RandomArmPlacebo:model_days     0.3065     0.8253    70.0634   0.371
##                             Pr(>|t|)
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo              0.4672
## model_days                    0.0311 *
## sexM                        3.57e-08 ***
## age                           0.1013
## RandomArmPlacebo:model_days   0.7114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age
## RndmArmPlcb -0.201
## model_days  -0.016  0.027
## sexM        -0.172  0.037  0.000
## age         -0.903 -0.055  0.002 -0.079
## RndmArmPl:_  0.009 -0.036 -0.592  0.001 -0.001
```

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

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## SurfArea ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_RSA
##
## REML criterion at convergence: 2503.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.13482 -0.35071  0.00312  0.37950  3.09151
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 54283287 7367.7
##  Residual               214731  463.4
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                               Estimate Std. Error         df t value
## (Intercept)                 83705.9315  3562.0240    65.0303  23.500
## RandomArmPlacebo            -1340.1774  1770.2952    65.1912  -0.757
## model_days                     -1.0753     0.4884    70.0277  -2.202
## sexM                        10923.1943  1774.4877    64.9969   6.156
## age                          -117.0597    57.5840    64.9973  -2.033
## siteMAS                     -2383.7886  2251.6589    64.9974  -1.059
## siteNKI                      1177.5286  2495.2662    64.9974   0.472
## sitePMC                      6985.0155  2567.5131    64.9968   2.721
## RandomArmPlacebo:model_days     0.3044     0.8253    70.0711   0.369
##                             Pr(>|t|)
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo             0.45176
## model_days                   0.03098 *
## sexM                        5.22e-08 ***
## age                          0.04616 *
## siteMAS                      0.29366
## siteNKI                      0.63858
## sitePMC                      0.00835 **
## RandomArmPlacebo:model_days  0.71338
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## RndmArmPlcb -0.139
## model_days  -0.016  0.028
## sexM        -0.130  0.055  0.000
## age         -0.882 -0.066  0.002 -0.076
## siteMAS     -0.292 -0.147  0.002 -0.066  0.108
## siteNKI     -0.175 -0.119 -0.002 -0.119  0.010  0.357
## sitePMC     -0.153 -0.089  0.000 -0.144 -0.009  0.343  0.319
## RndmArmPl:_  0.009 -0.038 -0.592  0.001 -0.001  0.000  0.002  0.000
```

7.3 Dealing with the confusion..

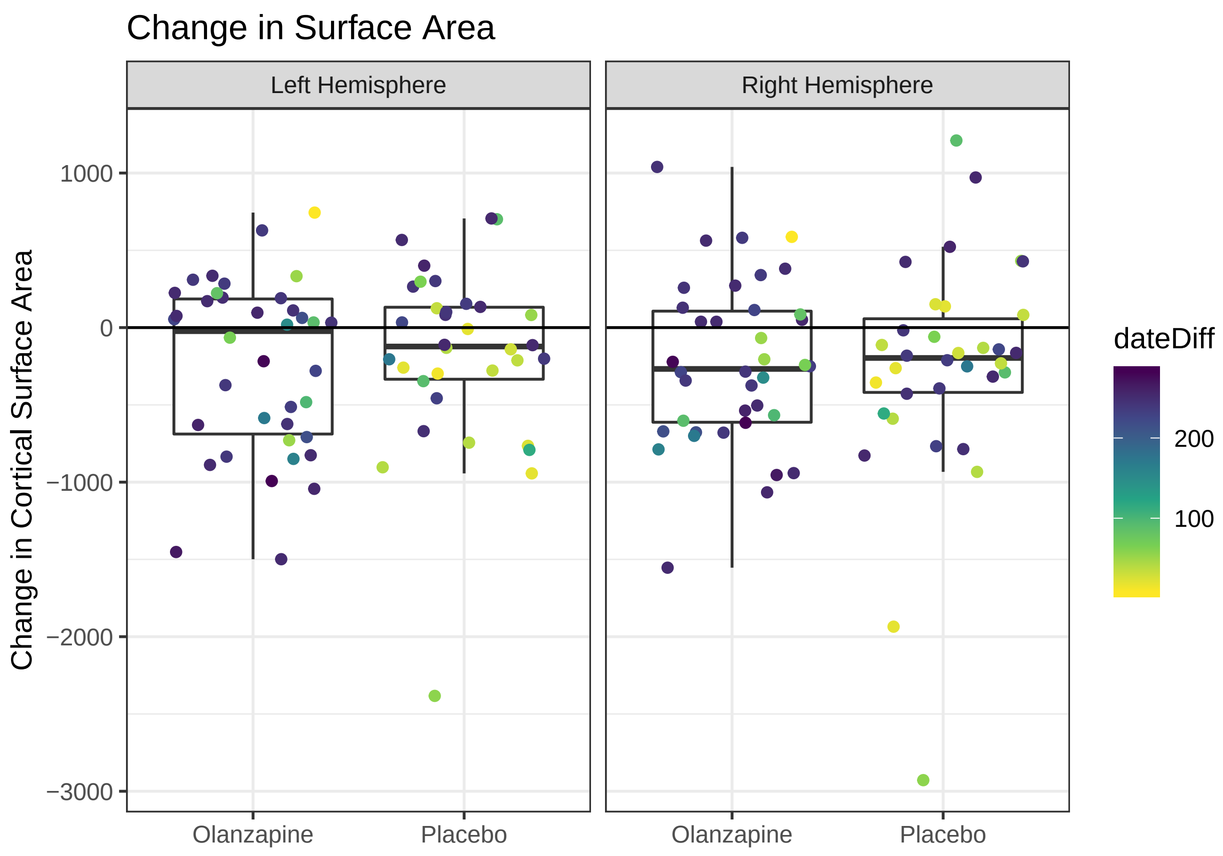
So I (maybe for one) Was Confused by the way that the two findings above seems to go in opposite directions.

I.e. More the RCT analysis shows a decrease in surface area with Olanzapine, while the longitudinal fit is trending upward.

I thought it might be useful to rebuild the first plot, but with the whole sample, with point color representing the time between scans

Note that the dark blue dots would be the one's included in the RCT analysis

```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
df %>%
  gather(TCT, mm, LSurfArea_change, RSurfArea_change) %>%
  mutate(ThickChange = factor(TCT, levels = c("LSurfArea_change", "RSurfArea_change"),
                              labels = c("Left Hemisphere", "Right Hemisphere"))) %>%
ggplot(aes(x= RandomArm, y = mm)) +
     geom_boxplot(outlier.shape = NA) +
     geom_jitter(aes(color = dateDiff)) +
     geom_hline(yintercept = 0) +
     labs(title = "Change in Surface Area",x = NULL, y = "Change in Cortical Surface Area") +
     facet_wrap(~ ThickChange) +
     scale_color_viridis_c(direction = -1) +
     theme_bw()
```



8 Whole Skeleton Fractional Anisotropy

```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= 72 - note two were excluded for IF)
df <- read_csv('../generated_csvs/STOPPD_masterDF_2018-11-05.csv') %>%
  mutate(STUDYID = as.character(STUDYID)) %>%
  filter(second_complete == "Yes", MR_exclusion == "No")

#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))
```

8.1 Known exclusion reasons

8.1.0.1 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 2 participants (final n=71)

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

8.2 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..

```
#bring in FA data (from the filesystem)
FA_most <- read_csv('../data/enigma-DTI_archive_201811/enigmaDTI-FA-results.csv')
FA_PMC <- read_csv('../data/enigma-DTI_PMCredo_201809/enigmaDTI-FA-results.csv')

# separate id into it's parts and then drop old PMC data
FA_most <- FA_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
FA <- FA_PMC %>%
  separate(id, into = c("study", "site", "STUDYID", "timepoint")) %>%
  bind_rows(FA_most)

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

8.3 check for missing FA data

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

Number of missing FA values
0

8.4 merge (i.e. join) the FA data with the clinical scores

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

```r
all_FA %>%
  filter(is.na(AverageFA_FA_01)) %>%
  summarise(`Number of missing timepoint 1 FA values` = n()) %>%
  knitr::kable()
```

Number of missing timepoint 1 FA values
0

```r
all_FA %>%
  filter(is.na(AverageFA_FA_02)) %>%
  summarise(`Number of missing timepoint 2 FA values` = n()) %>%
  knitr::kable()
```

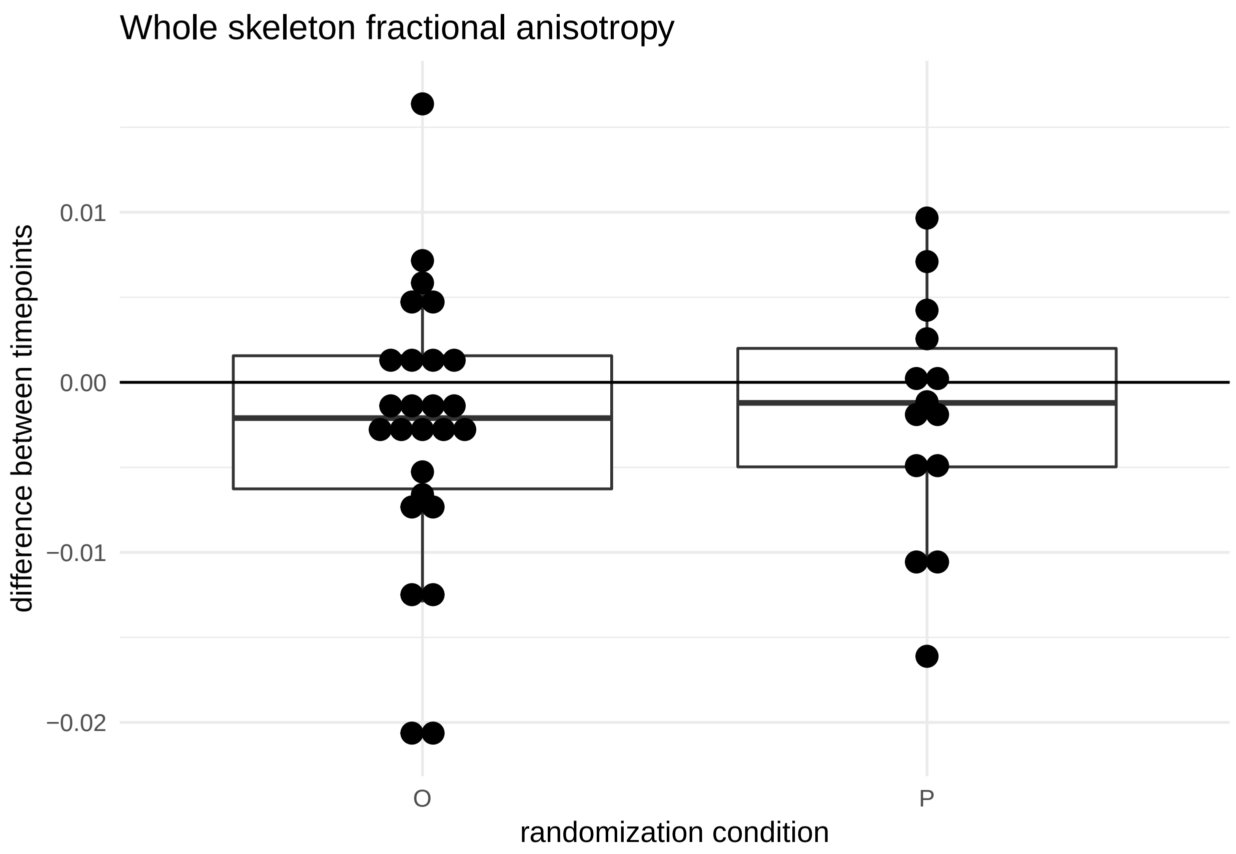
Number of missing timepoint 2 FA values
0

```r
#write out clean FA speadsheet (required for subsequent FA analyses)
write.csv(all_FA, '../generated_csvs/STOPPD_FAclean.csv', row.names = FALSE)
```

8.5 RCT only

```r
#boxplot of difference in FA in whole skeleton (y axis) by randomization group (x axis)
ggplot(RCT_FA, aes(x= randomization, y = diffAverageSkel_FA)) +
   geom_boxplot(outlier.shape = NA) +
   geom_dotplot(binaxis = 'y', stackdir = 'center') +
   geom_hline(yintercept = 0) +
   ggtitle("Whole skeleton fractional anisotropy") +
   xlab("randomization condition") +
   ylab("difference between timepoints") +
   theme_minimal()
```

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



```
fit_rct <- lm(diffAverageSkel_FA ~ randomization, data= RCT_FA)
print(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_FA ~ randomization, data = RCT_FA)
##
## Coefficients:
##   (Intercept)  randomizationP
##    -0.0026337       0.0006364
```

```
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_FA ~ randomization, data = RCT_FA)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0182273 -0.0032754  0.0007225  0.0043244  0.0190097
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.0026337  0.0015172  -1.736   0.0907 .
## randomizationP  0.0006364  0.0025645   0.248   0.8054
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.007736 on 38 degrees of freedom
## Multiple R-squared:  0.001618,   Adjusted R-squared:  -0.02466
## F-statistic: 0.06158 on 1 and 38 DF,  p-value: 0.8054
```

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

```
##
## Call:
## lm(formula = diffAverageSkel_FA ~ randomization + sex + age,
##     data = RCT_FA)
##
## Coefficients:
##    (Intercept)  randomizationP            sexM             age
##      0.0066641       0.0017724       0.0030420      -0.0002049
```

```
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_FA ~ randomization + sex + age,
##     data = RCT_FA)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.0176902 -0.0037512  0.0003699  0.0044837  0.0169146
##
## Coefficients:
##                   Estimate  Std. Error t value Pr(>|t|)
## (Intercept)     0.00666412  0.00459470   1.450   0.1556
## randomizationP  0.00177238  0.00246851   0.718   0.4774
## sexM            0.00304198  0.00237628   1.280   0.2087
## age            -0.00020489  0.00008522  -2.404   0.0215 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00732 on 36 degrees of freedom
## Multiple R-squared:  0.1531, Adjusted R-squared:  0.08256
## F-statistic:  2.17 on 3 and 36 DF,  p-value: 0.1085
```

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

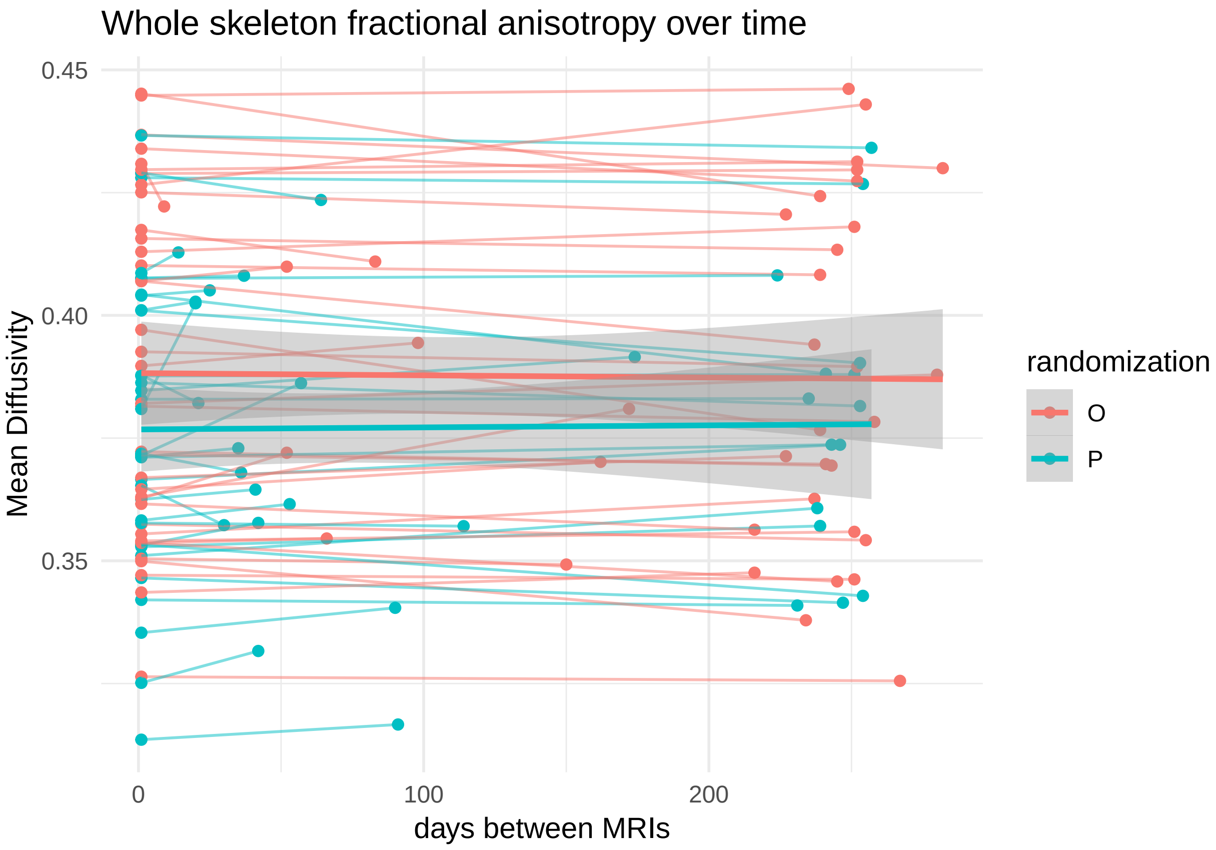
```
##
## Call:
## lm(formula = diffAverageSkel_FA ~ randomization + sex + age +
##     site, data = RCT_FA)
##
## Coefficients:
##    (Intercept)  randomizationP            sexM             age
##      0.0055312       0.0011892       0.0024924      -0.0002075
##        siteMAS         siteNKI         sitePMC
##      0.0048105       0.0015693       0.0027403
```

```r
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_FA ~ randomization + sex + age +
##     site, data = RCT_FA)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0164312 -0.0039600  0.0001338  0.0050820  0.0187285
##
## Coefficients:
##                   Estimate  Std. Error t value Pr(>|t|)
## (Intercept)     0.00553120  0.00519594   1.065   0.2948
## randomizationP  0.00118917  0.00254062   0.468   0.6428
## sexM            0.00249237  0.00243729   1.023   0.3139
## age            -0.00020752  0.00009488  -2.187   0.0359 *
## siteMAS         0.00481051  0.00324006   1.485   0.1471
## siteNKI         0.00156932  0.00309643   0.507   0.6157
## sitePMC         0.00274025  0.00380401   0.720   0.4764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00739 on 33 degrees of freedom
## Multiple R-squared:  0.2089, Adjusted R-squared:  0.06501
## F-statistic: 1.452 on 6 and 33 DF,  p-value: 0.2251
```

8.6 RCT & Relapse (with time as factor)

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



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

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: FA ~ randomization * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_wholeskelFA
##
## REML criterion at convergence: -705.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.94972 -0.36140 -0.01493  0.37639  2.03848
##
## Random effects:
##  Groups   Name        Variance   Std.Dev.
##  STUDYID  (Intercept) 0.00091790 0.030297
##  Residual             0.00002934 0.005417
## Number of obs: 142, groups:  STUDYID, 71
##
## Fixed effects:
##                             Estimate   Std. Error           df t value
## (Intercept)              0.420038504  0.014072666 67.294984696  29.848
## randomizationP          -0.010187596  0.007320306 68.631328224  -1.392
## model_days              -0.000006710  0.000005717 69.240746206  -1.174
## sexM                    -0.000889927  0.007300269 67.001340480  -0.122
## age                       -0.000570356  0.000237119 67.001860376  -2.405
## randomizationP:model_days  0.000002899  0.000009637 69.582543938   0.301
##                                       Pr(>|t|)    
## (Intercept)               <0.0000000000000002 ***
## randomizationP                         0.1685    
## model_days                             0.2446    
## sexM                                   0.9033    
## age                                    0.0189 *  
## randomizationP:model_days              0.7644    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) rndmzP mdl_dy sexM   age   
## randomiztnP -0.205                            
## model_days  -0.047  0.082                     
## sexM        -0.174  0.050  0.002              
## age         -0.899 -0.061  0.004 -0.085       
## rndmztnP:m_  0.026 -0.108 -0.593  0.002 -0.001
```

8.7 just wanna through in one site model..

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

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: 
## FA ~ randomization * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_wholeskelFA
## 
## REML criterion at convergence: -763.4
## 
## Scaled residuals: 
##      Min       1Q   Median       3Q      Max 
## -2.22673 -0.36093  0.00773  0.40476  1.97403 
## 
## Random effects:
##  Groups   Name        Variance   Std.Dev.
##  STUDYID  (Intercept) 0.00026606 0.016311
##  Residual             0.00002933 0.005416
## Number of obs: 142, groups:  STUDYID, 71
## 
## Fixed effects:
##                               Estimate   Std. Error           df t value
## (Intercept)                0.446664113  0.008121600 64.854939100  54.997
## randomizationP            -0.001694241  0.004143105 68.953740506  -0.409
## model_days                -0.000006481  0.000005707 69.765390246  -1.136
## sexM                       0.007528548  0.004063003 64.002130344   1.853
## age                       -0.000644155  0.000131327 64.007125598  -4.905
## siteMAS                   -0.045128486  0.005245359 64.005598123  -8.604
## siteNKI                   -0.054431216  0.005671919 64.015089285  -9.597
## sitePMC                   -0.057667867  0.005835298 63.997753494  -9.883
## randomizationP:model_days  0.000002411  0.000009591 70.909497124   0.251
##                                        Pr(>|t|)    
## (Intercept)               < 0.0000000000000002 ***
## randomizationP                          0.6839    
## model_days                              0.2599    
## sexM                                    0.0685 .  
## age                         0.0000067266336686 ***
## siteMAS                     0.0000000000027939 ***
## siteNKI                     0.0000000000000519 ***
## sitePMC                     0.0000000000000168 ***
## randomizationP:model_days               0.8023    
## ---
## 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 -0.141                                                 
## model_days  -0.082  0.144                                          
## sexM        -0.126  0.072  0.004                                   
## age         -0.878 -0.076  0.009 -0.086                            
## siteMAS     -0.290 -0.172  0.005 -0.092  0.124                     
## siteNKI     -0.175 -0.122 -0.011 -0.122  0.013  0.355              
## sitePMC     -0.153 -0.091 -0.001 -0.146 -0.007  0.341  0.320       
## rndmztnP:m_  0.044 -0.190 -0.595  0.002 -0.002  0.002  0.010  0.003
```

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

```
## Warning in rm("df", "fit_all", "fit_rct", "FA", "plot", "RCT_FA",
## "RCTRelapse_FA"): object 'plot' not found
```

9 Whole Skeleton Mean Diffusivity

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

```
#bring in subject info (generated by 03_STOPPD_masterDF.Rmd)
# then take only the subjects who completed (n= 72 - note two were excluded for IF)
df <- read_csv('../generated_csvs/STOPPD_masterDF_2018-11-05.csv') %>%
  mutate(STUDYID = as.character(STUDYID)) %>%
  filter(second_complete == "Yes", MR_exclusion == "No")

#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))
```

9.1 Known exclusion reasons

9.1.0.1 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 2 participants (final n=71)

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

9.2 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_201811/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)
```

9.3 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

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

```
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()
```

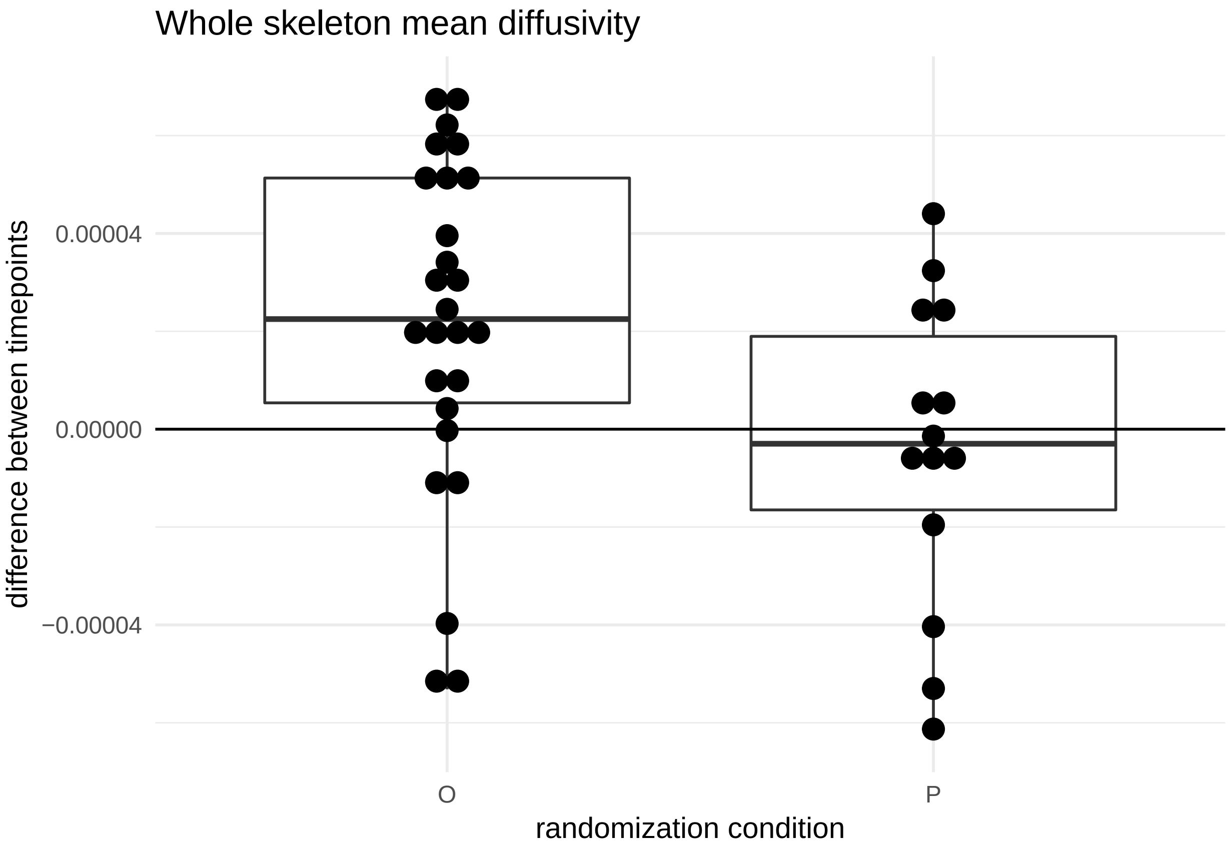
Number of missing timepoint 2 MD values
0

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

9.5 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') +
   geom_hline(yintercept = 0) +
   ggtitle("Whole skeleton mean diffusivity") +
   xlab("randomization condition") +
   ylab("difference between timepoints") +
   theme_minimal()
```

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



```
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     -0.00002586
```

```
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization, data = RCT_MD)
##
## Residuals:
##          Min           1Q       Median           3Q          Max
## -0.000074955 -0.000016010  0.000001087  0.000028189  0.000048102
##
## Coefficients:
##                    Estimate   Std. Error t value Pr(>|t|)
## (Intercept)     0.000021795  0.000006535   3.335  0.00191 **
## randomizationP -0.000025858  0.000011047  -2.341  0.02459 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.00003332 on 38 degrees of freedom
## Multiple R-squared:  0.126,  Adjusted R-squared:  0.103
## F-statistic: 5.479 on 1 and 38 DF,  p-value: 0.02459
```

```
#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
##  0.00001593673  -0.00002622481   0.00000257258   0.00000008944
```

```
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization + sex + age,
##     data = RCT_MD)
##
## Residuals:
##          Min           1Q       Median           3Q          Max
## -0.000078824 -0.000015346  0.000001442  0.000027004  0.000051108
##
## Coefficients:
##                       Estimate     Std. Error t value Pr(>|t|)
## (Intercept)     0.00001593673  0.00002145081   0.743   0.4623
## randomizationP -0.00002622481  0.00001152451  -2.276   0.0289 *
## sexM            0.00000257258  0.00001109389   0.232   0.8179
## age             0.00000008944  0.00000039787   0.225   0.8234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00003417 on 36 degrees of freedom
## Multiple R-squared:  0.1292, Adjusted R-squared:  0.05661
## F-statistic:  1.78 on 3 and 36 DF,  p-value: 0.1684
```

```
#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)
```

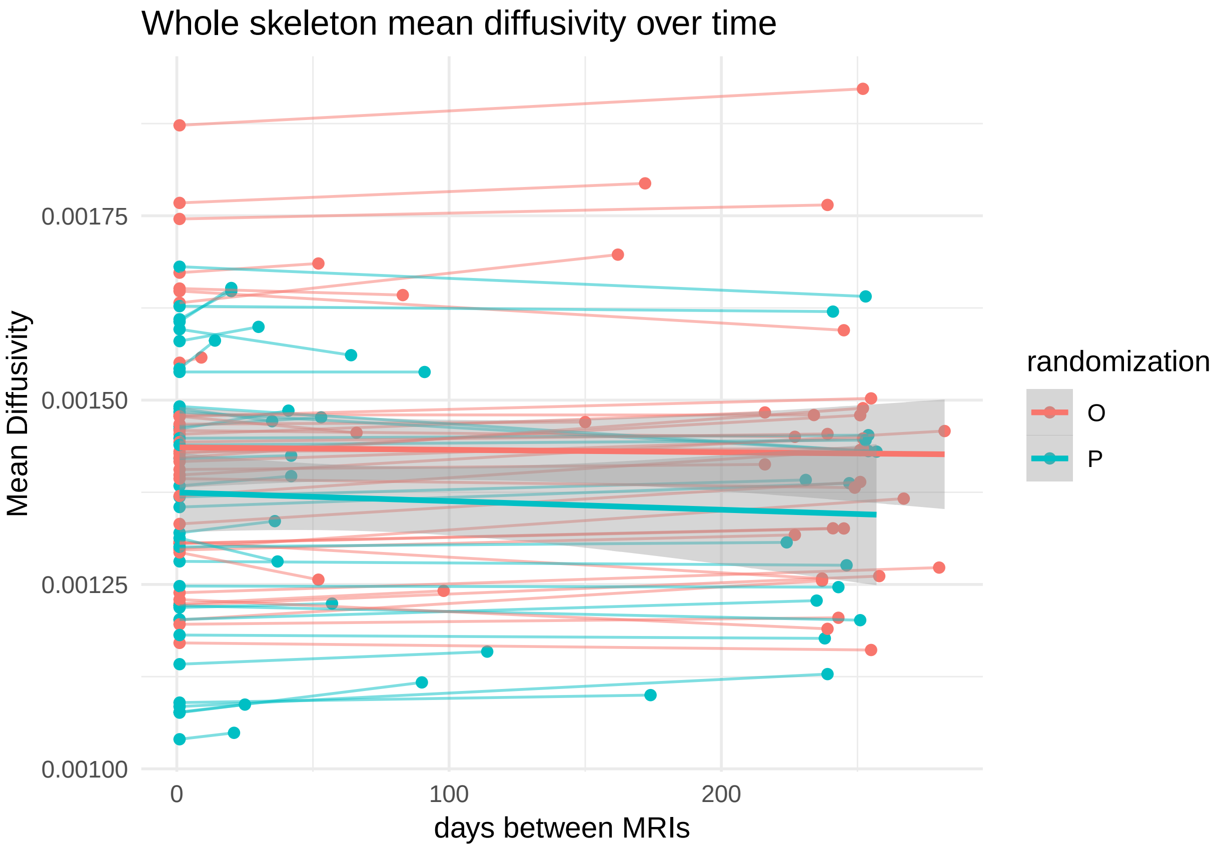
```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization + sex + age +
##     site, data = RCT_MD)
##
## Coefficients:
##    (Intercept)  randomizationP            sexM             age
##   0.0000015270   -0.0000284379    0.0000026363    0.0000003261
##        siteMAS         siteNKI         sitePMC
##   0.0000081404    0.0000119016   -0.0000128742
```

```
summary(fit_rct)
```

```
##
## Call:
## lm(formula = diffAverageSkel_MD ~ randomization + sex + age +
##     site, data = RCT_MD)
##
## Residuals:
##          Min           1Q       Median           3Q          Max
## -0.000072064 -0.000010679  0.000000839  0.000020198  0.000049068
##
## Coefficients:
##                      Estimate    Std. Error t value Pr(>|t|)
## (Intercept)     0.0000015270  0.0000244528   0.062   0.9506
## randomizationP -0.0000284379  0.0000119565  -2.378   0.0233 *
## sexM            0.0000026363  0.0000114702   0.230   0.8196
## age             0.0000003261  0.0000004465   0.730   0.4703
## siteMAS         0.0000081404  0.0000152481   0.534   0.5970
## siteNKI         0.0000119016  0.0000145722   0.817   0.4199
## sitePMC        -0.0000128742  0.0000179022  -0.719   0.4771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00003478 on 33 degrees of freedom
## Multiple R-squared:  0.1733, Adjusted R-squared:  0.02303
## F-statistic: 1.153 on 6 and 33 DF,  p-value: 0.3545
```

9.6 RCT & Relapse (with time as factor)

```
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()
```



```
#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: -2207.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.80780 -0.44415 -0.02353  0.38354  1.76631
##
## Random effects:
##  Groups   Name        Variance        Std.Dev.
##  STUDYID  (Intercept) 0.0000000160882 0.00012684
##  Residual             0.0000000004334 0.00002082
## Number of obs: 142, groups:  STUDYID, 71
##
## Fixed effects:
##                               Estimate     Std. Error             df
## (Intercept)                0.00096997852  0.00005883295 67.24042295176
## randomizationP            -0.00006889045  0.00003058032 68.37097299833
## model_days                 0.00000008709  0.00000002198 69.19446523698
## sexM                       0.00003079208  0.00003052508 66.99217346591
## age                            0.00000802914  0.00000099148 66.99260766274
## randomizationP:model_days -0.0000000009395  0.00000003705 69.48340161813
##                           t value             Pr(>|t|)    
## (Intercept)                16.487 < 0.0000000000000002 ***
## randomizationP             -2.253             0.027486 *  
## model_days                  3.963             0.000178 ***
## sexM                        1.009             0.316726    
## age                         8.098       0.000000000016 ***
## randomizationP:model_days  -2.536             0.013471 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) rndmzP mdl_dy sexM   age   
## randomiztnP -0.205                            
## model_days  -0.043  0.075                     
## sexM        -0.174  0.050  0.002              
## age         -0.899 -0.061  0.004 -0.085       
## rndmztnP:m_  0.024 -0.099 -0.593  0.002 -0.001
```

9.7 just wanna through in one site model..

```
#run mixed linear model, with covariates
fit_all <- lmer(MD ~ randomization*model_days + sex + age + site + (1|STUDYID), data= RCTRelapse_wholes
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: -2201.2
## 
## Scaled residuals: 
##      Min       1Q   Median       3Q      Max 
## -1.76367 -0.44166 -0.00876  0.39208  1.91515 
## 
## Random effects:
##  Groups   Name        Variance        Std.Dev.  
##  STUDYID  (Intercept) 0.0000000077233 0.00008788
##  Residual             0.0000000004336 0.00002082
## Number of obs: 142, groups:  STUDYID, 71
## 
## Fixed effects:
##                                 Estimate     Std. Error             df
## (Intercept)                0.00107616983  0.00004312397 64.39556122315
## randomizationP            -0.00003551473  0.00002183608 66.56430055822
## model_days                 0.00000008624  0.00000002197 69.34666962025
## sexM                       0.00005950979  0.00002160794 63.94765683051
## age                        0.00000761020  0.00000069842 63.95027065884
## siteMAS                   -0.00019539464  0.00002789578 63.94948973534
## siteNKI                   -0.00018424524  0.00003016377 63.95445709082
## sitePMC                    -0.00018719886  0.00003103365 63.94534349730
## randomizationP:model_days -0.00000009587  0.00000003699 69.94739066144
##                            t value             Pr(>|t|)
## (Intercept)                 24.955 < 0.0000000000000002 ***
## randomizationP              -1.626             0.108585
## model_days                   3.926             0.000201 ***
## sexM                         2.754             0.007655 **
## age                         10.896 0.000000000000000325 ***
## siteMAS                     -7.004 0.000000001833884447 ***
## siteNKI                     -6.108 0.000000066268061845 ***
## sitePMC                     -6.032 0.000000089508213317 ***
## randomizationP:model_days  -2.592             0.011617 *
## ---
## 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 -0.137
## model_days  -0.059  0.105
## sexM        -0.126  0.072  0.003
## age         -0.879 -0.077  0.006 -0.086
## siteMAS     -0.290 -0.174  0.004 -0.092  0.124
## siteNKI     -0.175 -0.122 -0.008 -0.122  0.013  0.355
## sitePMC     -0.153 -0.092  0.000 -0.146 -0.007  0.341  0.320
## rndmztnP:m_  0.032 -0.139 -0.594  0.001 -0.002  0.001  0.007  0.002
```

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

10 Freesurfer Derived Subcortical Volumes

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------------------------
```

```
## √ ggplot2 3.1.0     √ purrr   0.2.4
## √ tibble  1.4.1     √ dplyr   0.7.7
## √ tidyr   0.7.2     √ stringr 1.2.0
## √ readr   1.1.1     √ forcats 0.2.0
```

```
## -- Conflicts ---------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:tidyr':
##
##     expand
```

```
## Loading required package: methods
```

```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
df <- read_csv("../generated_csvs/STOPPD_masterDF_2018-11-05.csv",na = "empty") #spreadsheet created by
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   STUDYID = col_integer()
## )

## See spec(...) for full column specifications.
```

```r
FS <- read_csv('../data/fs-enigma-long_201811/LandRvolumes.csv') #bring in subcortical data, from pipel
```

```
## Parsed with column specification:
## cols(
##   SubjID = col_character(),
##   LLatVent = col_double(),
##   RLatVent = col_double(),
##   Lthal = col_double(),
##   Rthal = col_double(),
##   Lcaud = col_double(),
##   Rcaud = col_double(),
##   Lput = col_double(),
##   Rput = col_double(),
##   Lpal = col_double(),
##   Rpal = col_double(),
##   Lhippo = col_double(),
##   Rhippo = col_double(),
##   Lamyg = col_double(),
##   Ramyg = col_double(),
##   Laccumb = col_double(),
##   Raccumb = col_double(),
##   ICV = col_double()
## )
```

```r
# remove participants that did not complete first and second scan (n=74)
# then add offlabel and dateDiff (in days columns)
# + a scan is by definition offlabel if it is the third scan
# then select the cols for analysis
df <- df %>%
  filter(first_complete == "Yes",
         second_complete == "Yes",
         MR_exclusion == "No") %>%
  mutate(offLabel  = if_else(third_complete == "Yes", "Yes", ''),
         dateDiff = round(difftime(second_date, first_date, units = "days"), 0),
         STUDYID = parse_character(STUDYID),
         age = parse_number(age)) %>%
  rename(category = "second_timepoint") %>%
  select(STUDYID, randomization, sex, age, category, offLabel, dateDiff)
```

10.1 cleaning the CT data

```r
# separating the subject id and anything afterwards to identify the longtudinal pipeline participants
# separating the subject id into site, "STUDYID" and timepoint columns
# filtering (two steps) to only include the longitudinal pipeline data
FS_long <- FS %>%
  separate(SubjID, into = c("subid", "longitudinal_pipe"), sep = '\\.', extra = "drop", fill = "right")
  separate(subid, into = c("study", "site", "STUDYID", "timepoint"), fill = "right") %>%
  filter(longitudinal_pipe == "long") %>%
  filter(timepoint != "00", timepoint != "03", timepoint != "")

# adding columns that combine L and R
FS_long_plus <- FS_long %>%
  mutate(Thalamus = Lthal + Rthal,
         Hippocampus = Lhippo + Rhippo,
         Striatum = Lcaud + Rcaud + Lput + Rput)


# move CT from long to wide format
FS_wide <- FS_long_plus %>%
  gather(region, volume, -study, -site, -timepoint, -STUDYID, -longitudinal_pipe) %>%
  spread(timepoint, volume) %>%
  mutate(change = `02` - `01`) %>%
  gather(timepoint, volume, `01`, `02`, change) %>%
  unite(newcolnames, region, timepoint) %>%
  spread(newcolnames, volume)
```

```r
# merge CT values with df
ana_df <- inner_join(df, FS_wide, by='STUDYID') %>%
    mutate(STUDYID = as.character(STUDYID),
         dateDiff = as.numeric(dateDiff),
         RandomArm = factor(randomization,
                            levels = c("O", "P"),
                            labels = c("Olanzapine", "Placebo")))

# write.csv
write_csv(ana_df, '../generated_csvs/STOPPD_participants_LandRVolumes_20181116.csv')
```

10.2 report any mising values from clinical trial sample

```r
anti_join(df, FS_wide, by='STUDYID') %>%
  summarise(`Number of participants missing` = n()) %>%
  knitr::kable()
```

Number of participants missing
0

```
ana_df %>%
  filter(is.na(ICV_01)) %>%
  summarise(`Number of participants missing timepoint 01` = n()) %>%
  knitr::kable()
```

Number of participants missing timepoint 01
0

```
ana_df %>%
  filter(is.na(ICV_02)) %>%
  summarise(`Number of participants missing timepoint 02` = n()) %>%
  knitr::kable()
```

Number of participants missing timepoint 02
0

10.3 creating an control error term calculating data frame

```
## identify the repeat control in a column and mangle the STUDYID to match in a new column
FS_long1 <- FS_long_plus %>%
  mutate(repeat_run = if_else(str_sub(STUDYID,1,1)=="R", "02", "01"),
         STUDYID = str_replace(STUDYID, 'R',""))

## extra the repeat study ids as a character vector
repeat_ids <- filter(FS_long1, repeat_run == "02")$STUDYID

## filter for only the subjects who are in the repeats list then switch to wide format
FS_wide_controls <- FS_long1 %>%
  filter(STUDYID %in% repeat_ids) %>%
  gather(region, volume, -study, -site, -timepoint, -STUDYID, -longitudinal_pipe, -repeat_run) %>%
  unite(newcolnames, region, repeat_run) %>%
  spread(newcolnames, volume)

#write.csv
  write.csv(FS_wide_controls, '../generated_csvs/STOPPD_errorControls_LandRVolumes_2018-11-05.csv', row

rm(FS_long1, repeat_ids)
```

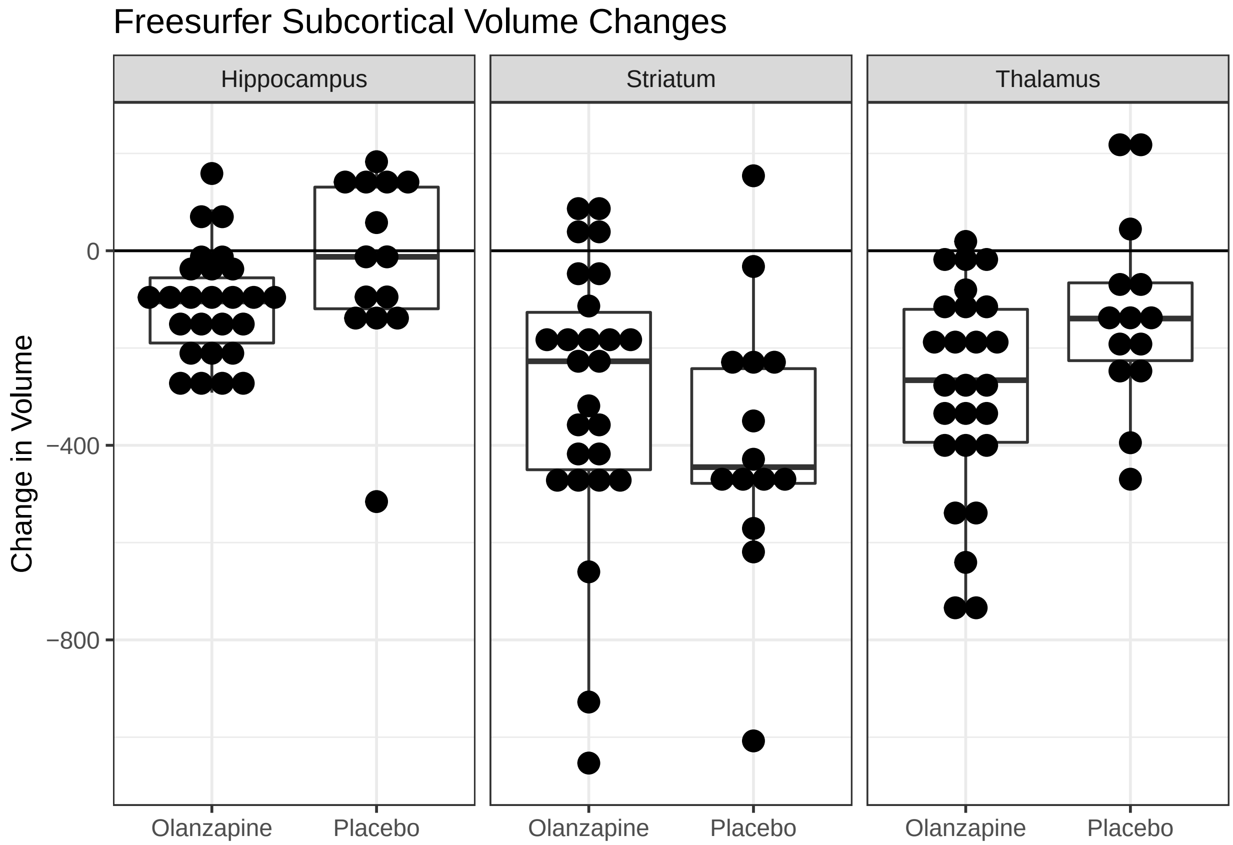
10.4 run RCT analysis (because it's simpler across volumes)

```
# make sure that STUDYID is an character not a number
# make sure that dateDiff is a number, not an interger
# label the randomization variable
RCT_SubCort <- ana_df %>%

  filter(category == "RCT")
```

```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
RCT_SubCort  %>%
  gather(region, volume_change, Thalamus_change, Hippocampus_change, Striatum_change) %>%
  mutate(Region = str_replace(region, '_change','')) %>%
ggplot(aes(x= RandomArm, y = volume_change)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     ggtitle("Freesurfer Subcortical Volume Changes") +
     xlab(NULL) +
     ylab("Change in Volume") +
     facet_wrap(~Region) +
     theme_bw()
```

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



### Running RCT Linear Models

10.4.0.1 Thalamus

```
#run linear model without covariates
  fit_rct <- lm(Thalamus_change ~ RandomArm, data= RCT_SubCort)
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = Thalamus_change ~ RandomArm, data = RCT_SubCort)
##
## Residuals:
##      Min       1Q  Median      3Q     Max
## -454.67 -105.09    3.09  155.71  366.04
##
## Coefficients:
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -286.13      41.25  -6.937    3e-08 ***
## RandomArmPlacebo   157.29      69.73   2.256   0.0299 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 210.3 on 38 degrees of freedom
## Multiple R-squared:  0.1181, Adjusted R-squared:  0.09489
## F-statistic: 5.089 on 1 and 38 DF,  p-value: 0.02992
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(Thalamus_change ~ RandomArm + sex + age, data= RCT_SubCort)
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = Thalamus_change ~ RandomArm + sex + age, data = RCT_SubCort)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -484.06 -102.67   17.14  139.25  326.59
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -188.053    134.129  -1.402   0.1695
## RandomArmPlacebo  169.144     72.061   2.347   0.0245 *
## sexM               30.398     69.369   0.438   0.6639
## age                -2.146      2.488  -0.863   0.3940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 213.7 on 36 degrees of freedom
## Multiple R-squared:  0.1377, Adjusted R-squared:  0.06582
## F-statistic: 1.916 on 3 and 36 DF,  p-value: 0.1444
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(Thalamus_change ~ RandomArm + sex + age + site, data= RCT_SubCort)
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = Thalamus_change ~ RandomArm + sex + age + site,
##     data = RCT_SubCort)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -396.26 -147.24   65.82  121.84  294.44
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -302.140    144.384  -2.093   0.0442 *
## RandomArmPlacebo  145.518     70.598   2.061   0.0472 *
## sexM                7.308     67.727   0.108   0.9147
## age                -1.325      2.636  -0.503   0.6186
## siteMAS           171.904     90.034   1.909   0.0649 .
## siteNKI           179.824     86.043   2.090   0.0444 *
## sitePMC            90.282    105.706   0.854   0.3992  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 205.3 on 33 degrees of freedom
## Multiple R-squared:   0.27,  Adjusted R-squared:  0.1373
## F-statistic: 2.035 on 6 and 33 DF,  p-value: 0.08872
```

10.4.0.2 Striatum

```
#run linear model without covariates
  fit_rct <- lm(Striatum_change ~ RandomArm, data= RCT_SubCort)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = Striatum_change ~ RandomArm, data = RCT_SubCort)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo  
##          -296.60            -91.83  
```

```
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = Striatum_change ~ RandomArm, data = RCT_SubCort)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -756.70 -141.60    8.11  153.80  542.53 
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)    
## (Intercept)       -296.60      55.23  -5.370 4.15e-06 ***
## RandomArmPlacebo   -91.83      93.35  -0.984    0.331    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 281.6 on 38 degrees of freedom
## Multiple R-squared:  0.02483,    Adjusted R-squared:  -0.0008284 
## F-statistic: 0.9677 on 1 and 38 DF,  p-value: 0.3315
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(Striatum_change ~ RandomArm + sex + age, data= RCT_SubCort)
  print(fit_rct)
```

```
##
## Call:
## lm(formula = Striatum_change ~ RandomArm + sex + age, data = RCT_SubCort)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age  
##         -169.875           -80.518           -12.233            -2.318  
```

```r
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = Striatum_change ~ RandomArm + sex + age, data = RCT_SubCort)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -772.16 -141.82    9.74  154.93  552.85
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -169.875    180.214  -0.943    0.352
## RandomArmPlacebo  -80.518     96.821  -0.832    0.411
## sexM              -12.233     93.203  -0.131    0.896
## age                -2.318      3.343  -0.693    0.492
##
## Residual standard error: 287.1 on 36 degrees of freedom
## Multiple R-squared:  0.0397, Adjusted R-squared:  -0.04032
## F-statistic: 0.4961 on 3 and 36 DF,  p-value: 0.6873
```

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

```
##
## Call:
## lm(formula = Striatum_change ~ RandomArm + sex + age + site,
##     data = RCT_SubCort)
##
## Coefficients:
##      (Intercept)  RandomArmPlacebo              sexM               age
##         -147.226           -69.518           -25.309            -3.412
##          siteMAS           siteNKI           sitePMC
##          -21.034            84.805           157.298
```

```r
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = Striatum_change ~ RandomArm + sex + age + site,
##     data = RCT_SubCort)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -742.29 -116.85   26.69  199.56  431.93
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -147.226    205.644  -0.716    0.479
## RandomArmPlacebo  -69.518    100.552  -0.691    0.494
## sexM              -25.309     96.463  -0.262    0.795
## age                -3.412      3.755  -0.909    0.370
## siteMAS           -21.034    128.235  -0.164    0.871
## siteNKI            84.805    122.550   0.692    0.494
## sitePMC           157.298    150.555   1.045    0.304
## 
## Residual standard error: 292.5 on 33 degrees of freedom
## Multiple R-squared:  0.08654,    Adjusted R-squared:  -0.07955 
## F-statistic: 0.521 on 6 and 33 DF,  p-value: 0.7881
```

10.4.0.3 Hippocampus

```r
#run linear model without covariates
  fit_rct <- lm(Hippocampus_change ~ RandomArm, data= RCT_SubCort)
  summary(fit_rct)
```

```
## 
## Call:
## lm(formula = Hippocampus_change ~ RandomArm, data = RCT_SubCort)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -490.91  -89.61    7.16   94.86  270.16 
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)    
## (Intercept)       -111.36      27.95  -3.984 0.000296 ***
## RandomArmPlacebo    86.58      47.25   1.832 0.074752 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 142.5 on 38 degrees of freedom
## Multiple R-squared:  0.08118,    Adjusted R-squared:  0.057 
## F-statistic: 3.357 on 1 and 38 DF,  p-value: 0.07475
```

```r
#run linear model with covariates of sex and age
  fit_rct <- lm(Hippocampus_change ~ RandomArm + sex + age, data= RCT_SubCort)
  summary(fit_rct)
```

```
## 
## Call:
## lm(formula = Hippocampus_change ~ RandomArm + sex + age, data = RCT_SubCort)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -498.44  -90.15   -0.46   84.67  257.79 
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)  
## (Intercept)         5.198     89.606   0.058   0.9541  
## RandomArmPlacebo   97.320     48.141   2.022   0.0507 .
## sexM               -6.910     46.342  -0.149   0.8823  
## age                -2.171      1.662  -1.306   0.1998  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 142.8 on 36 degrees of freedom
## Multiple R-squared:  0.1268, Adjusted R-squared:  0.05408 
## F-statistic: 1.743 on 3 and 36 DF,  p-value: 0.1756
```

```
#run linear model with covariates of sex and age
  fit_rct <- lm(Hippocampus_change ~ RandomArm + sex + age + site, data= RCT_SubCort)
  summary(fit_rct)
```

```
##
## Call:
## lm(formula = Hippocampus_change ~ RandomArm + sex + age + site,
##     data = RCT_SubCort)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -491.31  -98.13    3.17   97.16  258.53
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        32.179    103.540   0.311   0.7579
## RandomArmPlacebo   97.302     50.627   1.922   0.0633 .
## sexM              -10.971     48.568  -0.226   0.8227
## age                -2.748      1.891  -1.454   0.1555
## siteMAS            22.156     64.565   0.343   0.7337
## siteNKI           -24.291     61.703  -0.394   0.6964
## sitePMC            47.155     75.803   0.622   0.5382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 147.3 on 33 degrees of freedom
## Multiple R-squared:  0.1483, Adjusted R-squared:  -0.006514
## F-statistic: 0.9579 on 6 and 33 DF,  p-value: 0.4684
```

10.5 RCT & Relapse (with time as factor)

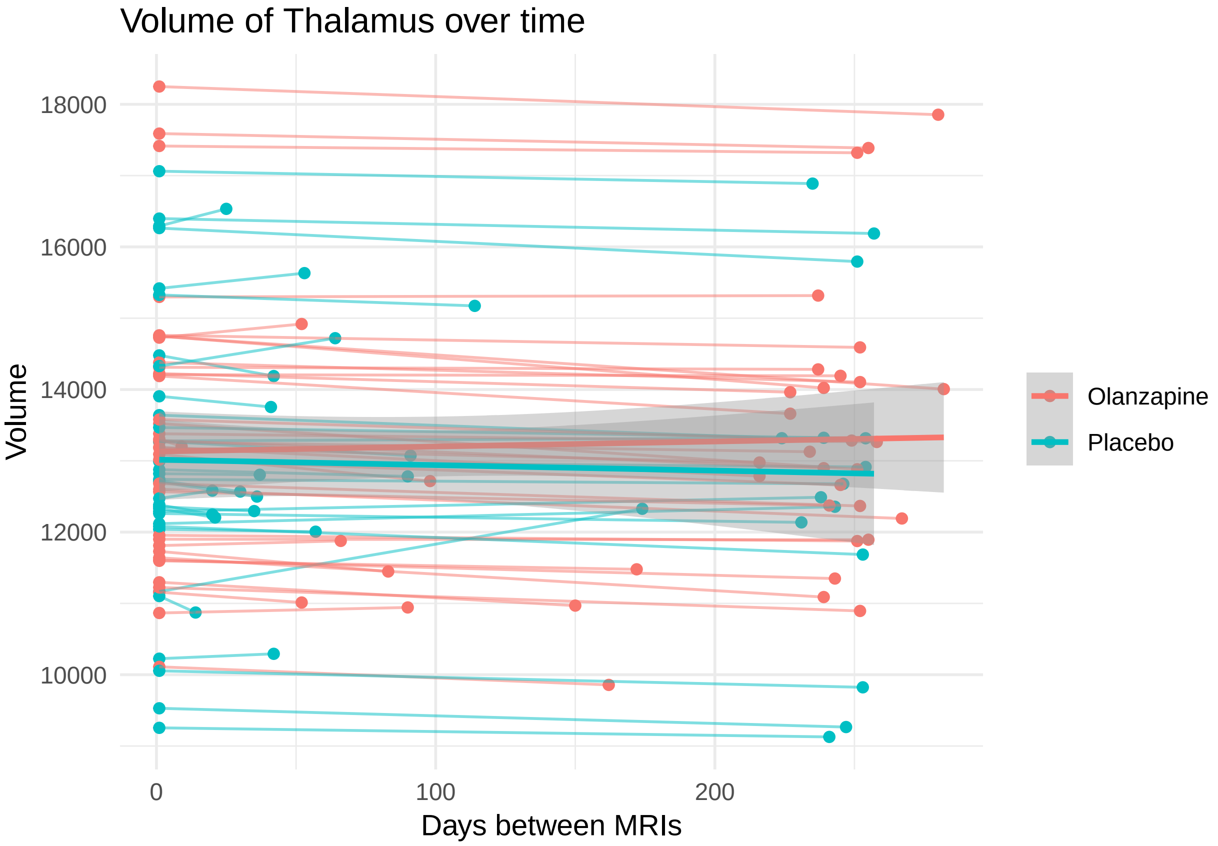
10.5.1 Thalamus

```
#restructure data for RCT & Relapse participants (N=72)
  RCTRelapse_Thalamus <- ana_df %>%
    gather(oldcolname, volume, Thalamus_01, Thalamus_02) %>%
    mutate(model_days = if_else(oldcolname == "Thalamus_01", 1, dateDiff))

RCTRelapse_Thalamus %>% filter(model_days == 1) %>% count(randomization) %>% knitr::kable()
```

randomization	n
O	38
P	34

```
#plot all data, including outlier (participant 210030)
  RCTRelapse_Thalamus %>%
   ggplot(aes(x=model_days, y=volume, colour=RandomArm)) +
   geom_point() +
   geom_line(aes(group=STUDYID), alpha = 0.5) +
   geom_smooth(method="lm", formula=y~poly(x,1)) +
   ggtitle("Volume of Thalamus over time") +
   labs(x = "Days between MRIs", y = "Volume", colour = NULL) +
   theme_minimal()
```



```
#run mixed linear model, with covariates
  fit_all <- lmer(volume ~ RandomArm*model_days + sex + age + (1|STUDYID), data= RCTRelapse_Thalamus)
  print(fit_all)
```

```
## Linear mixed model fit by REML ['lmerModLmerTest']
## Formula: volume ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_Thalamus
## REML criterion at convergence: 2193.44
## Random effects:
##  Groups   Name        Std.Dev.
##  STUDYID  (Intercept) 1396.1
##  Residual              171.2
## Number of obs: 144, groups:  STUDYID, 72
## Fixed Effects:
##                (Intercept)             RandomArmPlacebo
##                 15637.4008                    -181.4066
##                 model_days                         sexM
##                    -1.1851                    1898.4711
##                        age  RandomArmPlacebo:model_days
##                   -58.6352                       0.8093
```

```
  summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: volume ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_Thalamus
## 
## REML criterion at convergence: 2193.4
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -3.6119 -0.4282  0.0145  0.3924  3.5470 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 1949229  1396.1  
##  Residual               29305   171.2  
## Number of obs: 144, groups:  STUDYID, 72
## 
## Fixed effects:
##                              Estimate Std. Error         df t value
## (Intercept)                15637.4008   645.0482    68.1290  24.242
## RandomArmPlacebo            -181.4066   332.4480    68.7735  -0.546
## model_days                    -1.1851     0.1804    70.1127  -6.570
## sexM                        1898.4711   332.6682    67.9899   5.707
## age                          -58.6352    10.8628    67.9907  -5.398
## RandomArmPlacebo:model_days    0.8093     0.3046    70.2726   2.657
##                             Pr(>|t|)    
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo             0.58706    
## model_days                  7.46e-09 ***
## sexM                        2.75e-07 ***
## age                         9.26e-07 ***
## RandomArmPlacebo:model_days  0.00976 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age   
## RndmArmPlcb -0.201                            
## model_days  -0.032  0.056                     
## sexM        -0.172  0.037  0.000              
## age         -0.903 -0.054  0.003 -0.079       
## RndmArmPl:_  0.018 -0.074 -0.592  0.002 -0.001
```

```r
#run mixed linear model, with covariates
  fit_all <- lmer(volume ~ RandomArm*model_days + sex + age + site + (1|STUDYID), data= RCTRelapse_Thala
  print(fit_all)
```

```
## Linear mixed model fit by REML ['lmerModLmerTest']
## Formula: 
## volume ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_Thalamus
## REML criterion at convergence: 2140.5
## Random effects:
##  Groups   Name        Std.Dev.
##  STUDYID  (Intercept) 1313.0  
##  Residual              171.2  
## Number of obs: 144, groups:  STUDYID, 72
## Fixed Effects:
##                 (Intercept)             RandomArmPlacebo
##                  15877.4970                    -149.6319
##                  model_days                         sexM
##                     -1.1878                    1844.5154
##                         age                      siteMAS
##                    -61.8649                    -822.5404
##                     siteNKI                      sitePMC
##                    821.9333                      37.9774
## RandomArmPlacebo:model_days
##                      0.8111
```

```
summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## volume ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_Thalamus
##
## REML criterion at convergence: 2140.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5816 -0.4404  0.0083  0.3873  3.5784
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 1724006  1313.0
##  Residual               29304   171.2
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                               Estimate Std. Error         df t value
## (Intercept)                 15877.4970   637.1250    65.1359  24.921
## RandomArmPlacebo             -149.6319   317.2828    65.8206  -0.472
## model_days                     -1.1878     0.1804    70.1251  -6.586
## sexM                         1844.5154   317.2625    64.9936   5.814
## age                           -61.8649    10.2955    64.9952  -6.009
## siteMAS                      -822.5404   402.5790    64.9957  -2.043
## siteNKI                       821.9333   446.1341    64.9957   1.842
## sitePMC                        37.9774   459.0477    64.9931   0.083
## RandomArmPlacebo:model_days     0.8111     0.3046    70.3103   2.663
##                             Pr(>|t|)
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo              0.6388
## model_days                  7.00e-09 ***
## sexM                        2.02e-07 ***
## age                         9.35e-08 ***
## siteMAS                       0.0451 *
## siteNKI                       0.0700 .
## sitePMC                       0.9343
## RandomArmPlacebo:model_days   0.0096 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## RndmArmPlcb -0.140
## model_days  -0.033  0.059
## sexM        -0.130  0.055  0.001
## age         -0.882 -0.066  0.004 -0.076
## siteMAS     -0.292 -0.147  0.004 -0.066  0.108
## siteNKI     -0.175 -0.119 -0.004 -0.119  0.010  0.357
## sitePMC     -0.153 -0.089  0.000 -0.144 -0.009  0.343  0.319
## RndmArmPl:_  0.018 -0.078 -0.592  0.001 -0.001  0.000  0.004  0.001
```

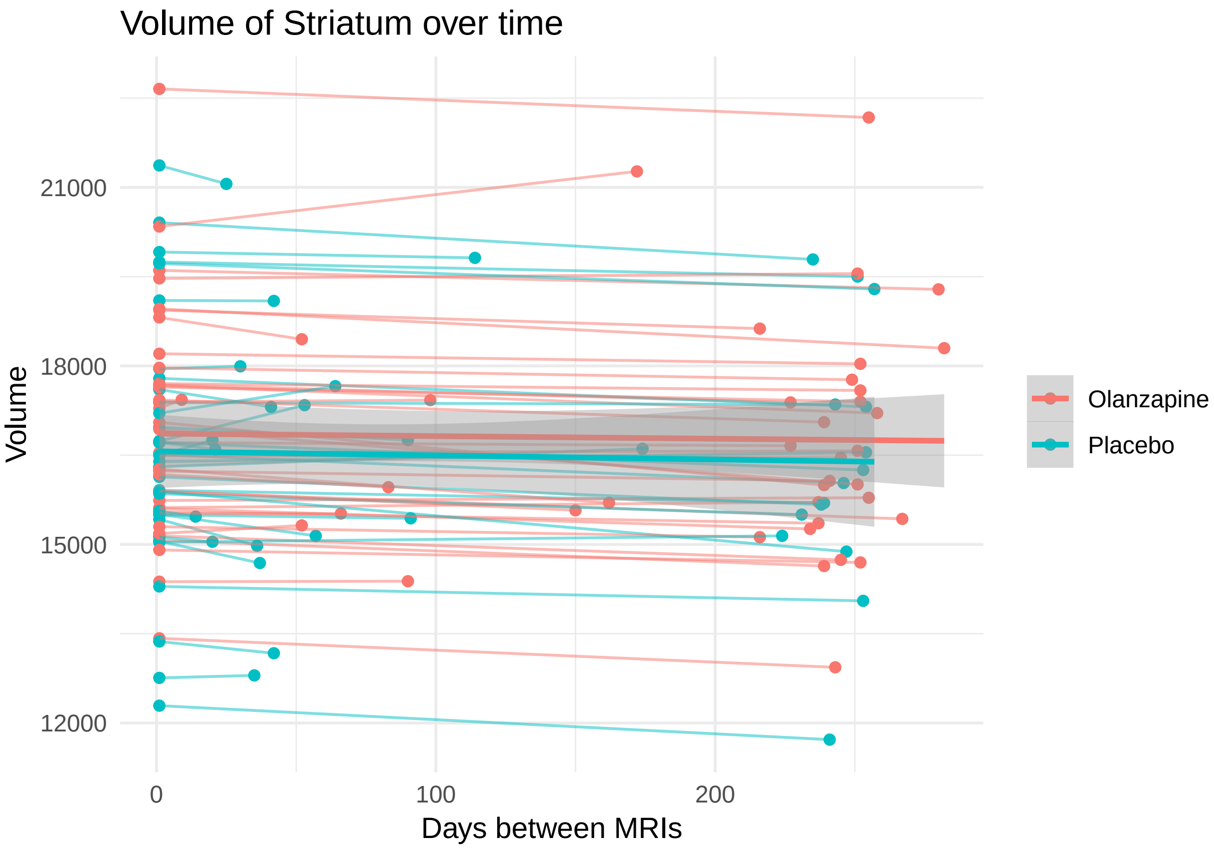
10.5.2 Striatum

```
#restructure data for RCT & Relapse participants (N=72)
  RCTRelapse_Striatum <- ana_df %>%
    gather(oldcolname, volume, Striatum_01, Striatum_02) %>%
    mutate(model_days = if_else(oldcolname == "Striatum_01", 1, dateDiff))

RCTRelapse_Striatum %>% filter(model_days == 1) %>% count(randomization) %>% knitr::kable()
```

randomization	n
O	38
P	34

```
#plot all data, including outlier (participant 210030)
  RCTRelapse_Striatum %>%
   ggplot(aes(x=model_days, y=volume, colour=RandomArm)) +
   geom_point() +
   geom_line(aes(group=STUDYID), alpha = 0.5) +
   geom_smooth(method="lm", formula=y~poly(x,1)) +
   ggtitle("Volume of Striatum over time") +
   labs(x = "Days between MRIs", y = "Volume", colour = NULL) +
   theme_minimal()
```



```
#run mixed linear model, with covariates
  fit_all <- lmer(volume ~ RandomArm*model_days + sex + age + (1|STUDYID), data= RCTRelapse_Striatum)
  print(fit_all)
```

```
## Linear mixed model fit by REML ['lmerModLmerTest']
## Formula: volume ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_Striatum
## REML criterion at convergence: 2250.495
## Random effects:
##  Groups   Name        Std.Dev.
##  STUDYID  (Intercept) 1674.5
##  Residual              215.6
## Number of obs: 144, groups:  STUDYID, 72
## Fixed Effects:
##               (Intercept)              RandomArmPlacebo
##                18786.7886                     -199.5106
##                model_days                          sexM
##                   -1.1427                     1603.1890
##                       age  RandomArmPlacebo:model_days
##                  -47.6384                       -0.1886
```

```
  summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: volume ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_Striatum
##
```

```
## REML criterion at convergence: 2250.5
## 
## Scaled residuals: 
##      Min       1Q   Median       3Q      Max 
## -2.37595 -0.36241  0.03762  0.33458  2.82797 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 2804106  1674.5  
##  Residual               46470   215.6  
## Number of obs: 144, groups:  STUDYID, 72
## 
## Fixed effects:
##                               Estimate Std. Error         df t value
## (Intercept)                 18786.7886   774.0095    68.1517  24.272
## RandomArmPlacebo             -199.5106   399.0078    68.8613  -0.500
## model_days                     -1.1427     0.2271    70.1337  -5.031
## sexM                         1603.1890   399.1561    67.9984   4.016
## age                           -47.6384    13.0339    67.9994  -3.655
## RandomArmPlacebo:model_days    -0.1886     0.3836    70.3099  -0.492
##                             Pr(>|t|)    
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo            0.618657    
## model_days                  3.61e-06 ***
## sexM                        0.000150 ***
## age                         0.000502 ***
## RandomArmPlacebo:model_days 0.624440    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age   
## RndmArmPlcb -0.202                            
## model_days  -0.034  0.059                     
## sexM        -0.172  0.036  0.000              
## age         -0.903 -0.054  0.003 -0.079       
## RndmArmPl:_  0.019 -0.078 -0.592  0.002 -0.001
```

```
#run mixed linear model, with covariates
  fit_all <- lmer(volume ~ RandomArm*model_days + sex + age + site + (1|STUDYID), data= RCTRelapse_Stria
  print(fit_all)
```

```
## Linear mixed model fit by REML ['lmerModLmerTest']
## Formula: 
## volume ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_Striatum
## REML criterion at convergence: 2204.864
## Random effects:
##  Groups   Name        Std.Dev.
##  STUDYID  (Intercept) 1680.8  
##  Residual              215.6  
## Number of obs: 144, groups:  STUDYID, 72
## Fixed Effects:
##                 (Intercept)             RandomArmPlacebo  
##                  19124.4744                    -100.8250  
##                 model_days                          sexM
##                    -1.1423                     1673.9180
##                        age                       siteMAS
##                   -48.9143                     -595.3613
##                    siteNKI                       sitePMC
##                  -807.5124                     -307.2624
## RandomArmPlacebo:model_days
##                    -0.1903
```

```
summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## volume ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_Striatum
##
## REML criterion at convergence: 2204.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.38638 -0.35547  0.03791  0.32934  2.81721
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 2825222  1680.8
##  Residual               46470   215.6
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                              Estimate Std. Error         df t value
## (Intercept)                19124.4744   815.4829    65.1363  23.452
## RandomArmPlacebo            -100.8250   406.0694    65.7992  -0.248
## model_days                    -1.1423     0.2271    70.1259  -5.029
## sexM                        1673.9180   406.0846    64.9986   4.122
## age                          -48.9143    13.1779    65.0002  -3.712
## siteMAS                     -595.3613   515.2865    65.0006  -1.155
## siteNKI                     -807.5124   571.0354    65.0006  -1.414
## sitePMC                     -307.2624   587.5646    64.9981  -0.523
## RandomArmPlacebo:model_days   -0.1903     0.3836    70.3052  -0.496
##                            Pr(>|t|)
## (Intercept)                 < 2e-16 ***
## RandomArmPlacebo           0.804680
## model_days                 3.63e-06 ***
## sexM                       0.000109 ***
## age                        0.000429 ***
## siteMAS                    0.252157
## siteNKI                    0.162099
## sitePMC                    0.602793
## RandomArmPlacebo:model_days 0.621329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## RndmArmPlcb -0.140
## model_days  -0.033  0.058
## sexM        -0.130  0.055  0.001
## age         -0.882 -0.066  0.004 -0.076
## siteMAS     -0.292 -0.147  0.004 -0.066  0.108
## siteNKI     -0.175 -0.119 -0.004 -0.119  0.010  0.357
## sitePMC     -0.153 -0.089  0.000 -0.144 -0.009  0.343  0.319
## RndmArmPl:_  0.017 -0.077 -0.592  0.001 -0.001  0.000  0.004  0.001
```

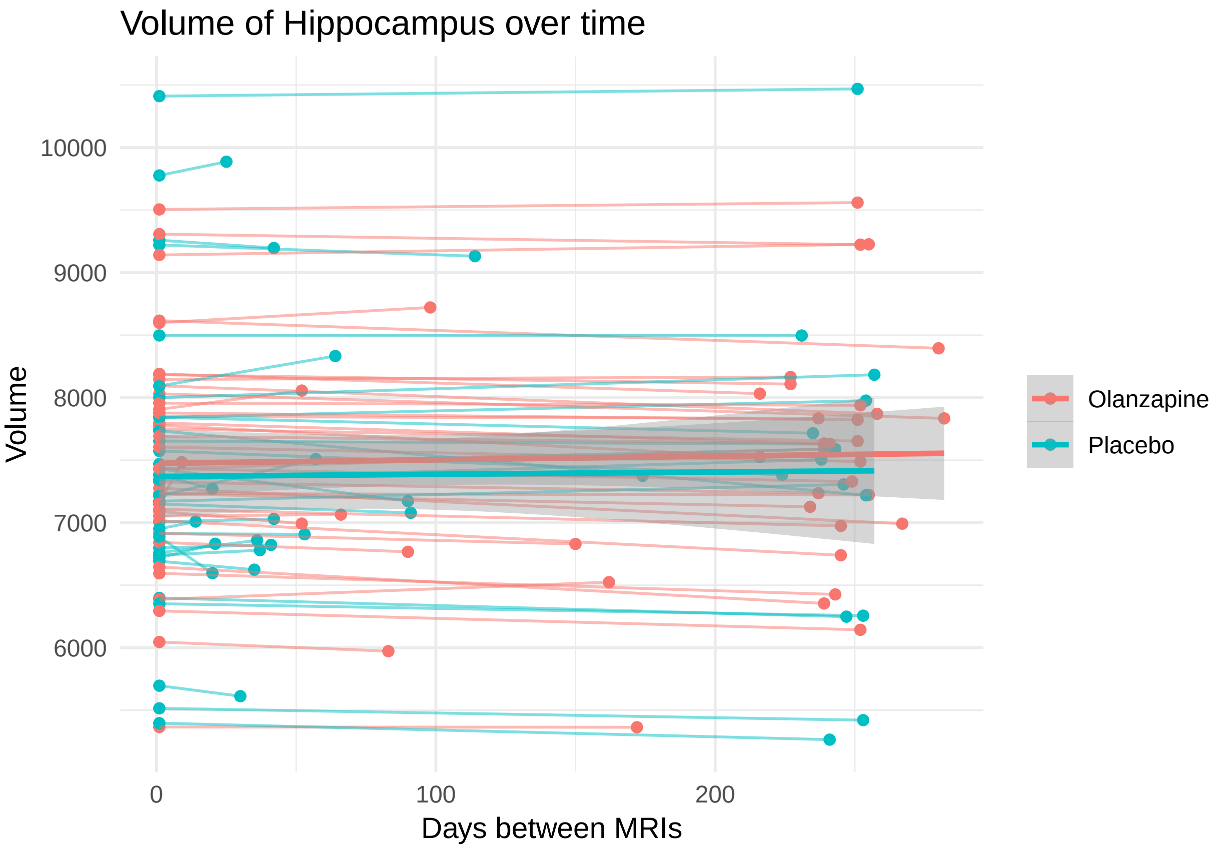
10.5.3 Hippocampus

```
#restructure data for RCT & Relapse participants (N=72)
  RCTRelapse_Hippocampus <- ana_df %>%
    gather(oldcolname, volume, Hippocampus_01, Hippocampus_02) %>%
    mutate(model_days = if_else(oldcolname == "Hippocampus_01", 1, dateDiff))

RCTRelapse_Hippocampus %>% filter(model_days == 1) %>% count(randomization) %>% knitr::kable()
```

randomization	n
O	38
P	34

```
#plot all data, including outlier (participant 210030)
  RCTRelapse_Hippocampus %>%
   ggplot(aes(x=model_days, y=volume, colour=RandomArm)) +
   geom_point() +
   geom_line(aes(group=STUDYID), alpha = 0.5) +
   geom_smooth(method="lm", formula=y~poly(x,1)) +
   ggtitle("Volume of Hippocampus over time") +
   labs(x = "Days between MRIs", y = "Volume", colour = NULL) +
   theme_minimal()
```



```
#run mixed linear model, with covariates
  fit_all <- lmer(volume ~ RandomArm*model_days + sex + age + (1|STUDYID), data= RCTRelapse_Hippocampus
  print(fit_all)
```

```
## Linear mixed model fit by REML ['lmerModLmerTest']
## Formula: volume ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_Hippocampus
## REML criterion at convergence: 2049.905
## Random effects:
##  Groups   Name        Std.Dev.
##  STUDYID  (Intercept) 840.5
##  Residual             100.6
## Number of obs: 144, groups:  STUDYID, 72
## Fixed Effects:
##               (Intercept)             RandomArmPlacebo
##                 8989.3972                     -84.5737
##                model_days                         sexM
##                   -0.4047                     607.4115
##                       age  RandomArmPlacebo:model_days
##                  -31.6143                       0.2634
```

```
  summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: volume ~ RandomArm * model_days + sex + age + (1 | STUDYID)
##    Data: RCTRelapse_Hippocampus
##
```

```
## REML criterion at convergence: 2049.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.41294 -0.41410 -0.00566  0.40252  2.36012
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 706389   840.5
##  Residual              10111   100.6
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                              Estimate Std. Error        df t value
## (Intercept)                 8989.3972   388.2347   68.1293  23.155
## RandomArmPlacebo             -84.5737   200.0678   68.7432  -0.423
## model_days                    -0.4047     0.1060   70.1138  -3.820
## sexM                         607.4115   200.2276   67.9968   3.034
## age                          -31.6143     6.5381   67.9975  -4.835
## RandomArmPlacebo:model_days    0.2634     0.1790   70.2661   1.472
##                             Pr(>|t|)
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo            0.673815
## model_days                  0.000285 ***
## sexM                        0.003420 **
## age                         7.94e-06 ***
## RandomArmPlacebo:model_days 0.145501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age
## RndmArmPlcb -0.201
## model_days  -0.031  0.055
## sexM        -0.172  0.037  0.000
## age         -0.903 -0.054  0.003 -0.079
## RndmArmPl:_  0.017 -0.073 -0.592  0.002 -0.001
```

```
#run mixed linear model, with covariates
  fit_all <- lmer(volume ~ RandomArm*model_days + sex + age + site + (1|STUDYID), data= RCTRelapse_Hipp
  print(fit_all)
```

```
## Linear mixed model fit by REML ['lmerModLmerTest']
## Formula:
## volume ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_Hippocampus
## REML criterion at convergence: 2005.898
## Random effects:
##  Groups   Name        Std.Dev.
##  STUDYID  (Intercept) 827.3
##  Residual             100.6
## Number of obs: 144, groups:  STUDYID, 72
## Fixed Effects:
##                 (Intercept)             RandomArmPlacebo
##                   8920.7798                    -106.7365
##                  model_days                        sexM
##                     -0.4048                    548.9133
##                         age                     siteMAS
##                    -32.1879                     46.0261
##                     siteNKI                     sitePMC
##                    116.0375                    631.6705
## RandomArmPlacebo:model_days
##                      0.2637
```

```
summary(fit_all)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## volume ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_Hippocampus
##
## REML criterion at convergence: 2005.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.40104 -0.40874 -0.00025  0.39616  2.37266
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  STUDYID  (Intercept) 684488   827.3
##  Residual              10111   100.6
## Number of obs: 144, groups:  STUDYID, 72
##
## Fixed effects:
##                              Estimate Std. Error        df t value
## (Intercept)                 8920.7798   401.2060   65.1204  22.235
## RandomArmPlacebo            -106.7365   199.7292   65.7164  -0.534
## model_days                    -0.4048     0.1060   70.1110  -3.820
## sexM                         548.9133   199.7986   64.9966   2.747
## age                          -32.1879     6.4837   64.9980  -4.964
## siteMAS                       46.0261   253.5270   64.9984   0.182
## siteNKI                      116.0375   280.9562   64.9984   0.413
## sitePMC                      631.6705   289.0890   64.9961   2.185
## RandomArmPlacebo:model_days    0.2637     0.1790   70.2722   1.474
##                             Pr(>|t|)
## (Intercept)                  < 2e-16 ***
## RandomArmPlacebo            0.594864
## model_days                  0.000285 ***
## sexM                        0.007766 **
## age                         5.26e-06 ***
## siteMAS                     0.856506
## siteNKI                     0.680959
## sitePMC                     0.032497 *
## RandomArmPlacebo:model_days 0.145021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) RndmAP mdl_dy sexM   age    sitMAS sitNKI sitPMC
## RndmArmPlcb -0.140
## model_days  -0.031  0.055
## sexM        -0.130  0.055  0.001
## age         -0.882 -0.066  0.004 -0.076
## siteMAS     -0.292 -0.147  0.004 -0.066  0.108
## siteNKI     -0.175 -0.119 -0.004 -0.119  0.010  0.357
## sitePMC     -0.153 -0.089  0.000 -0.144 -0.009  0.343  0.319
## RndmArmPl:_  0.017 -0.073 -0.592  0.001 -0.001  0.000  0.004  0.001
```