STOPPD surface area analysis (by hemisphere)

Contents

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)
```

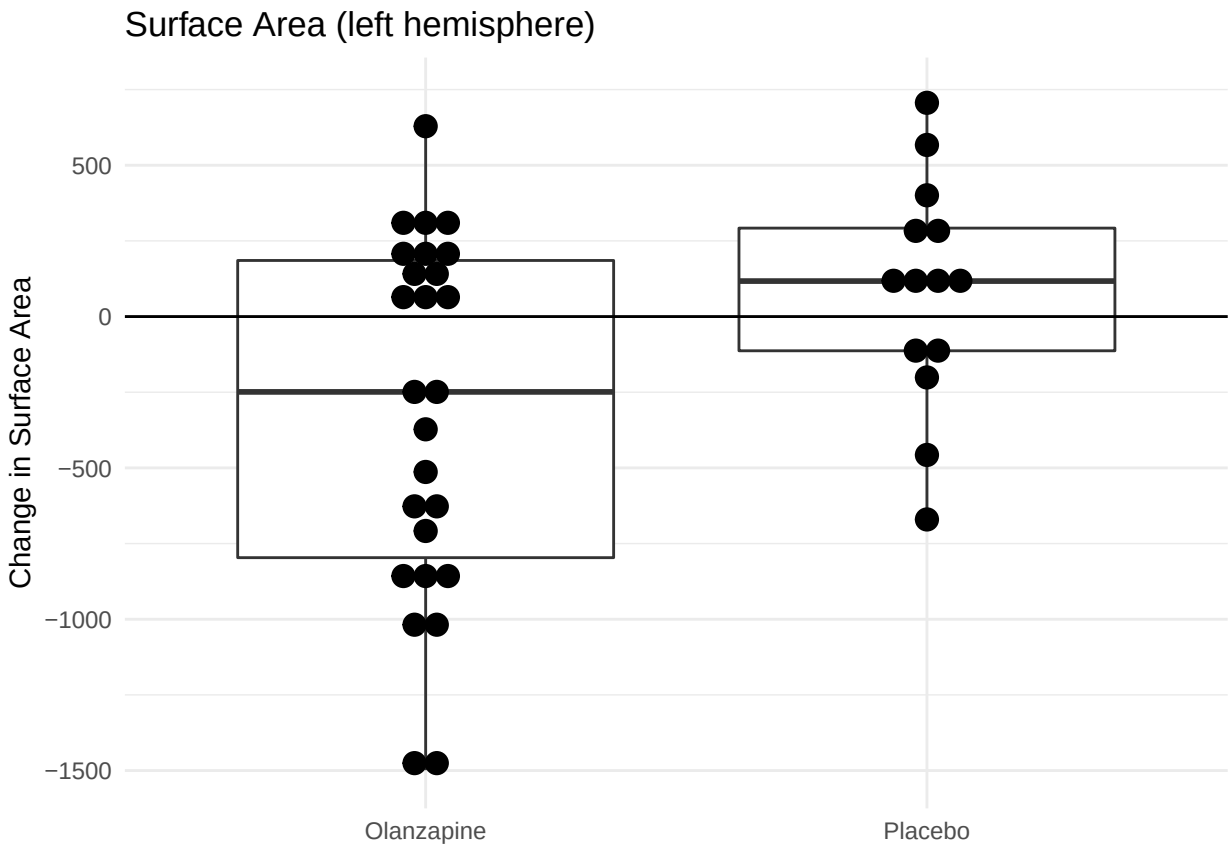
RCT only

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

randomization	n
O	26
P	14

```r
#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`.
```




```r
#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
```

```
#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
```

```
  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
```

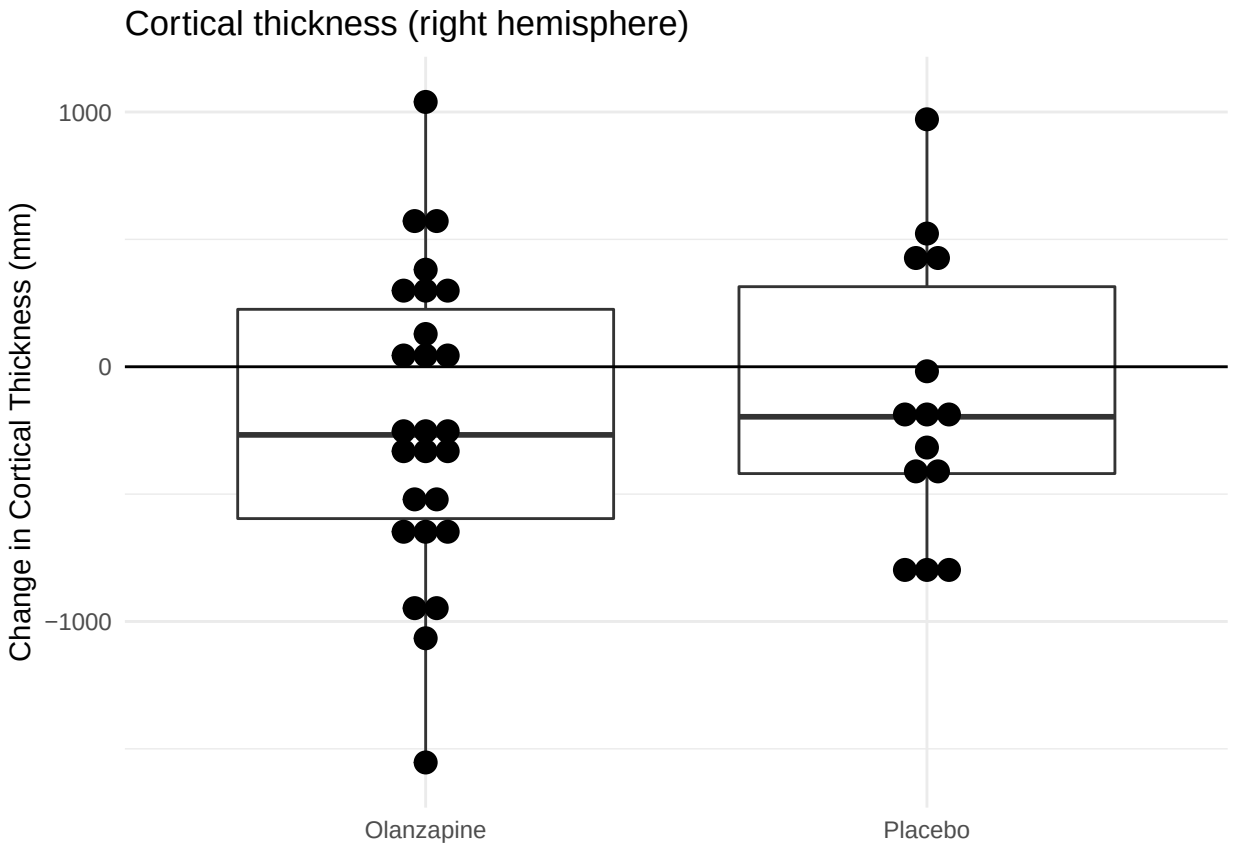
```
summary(fit_rct)
```

```
##
## 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
```

looking at the same thing for Right SA

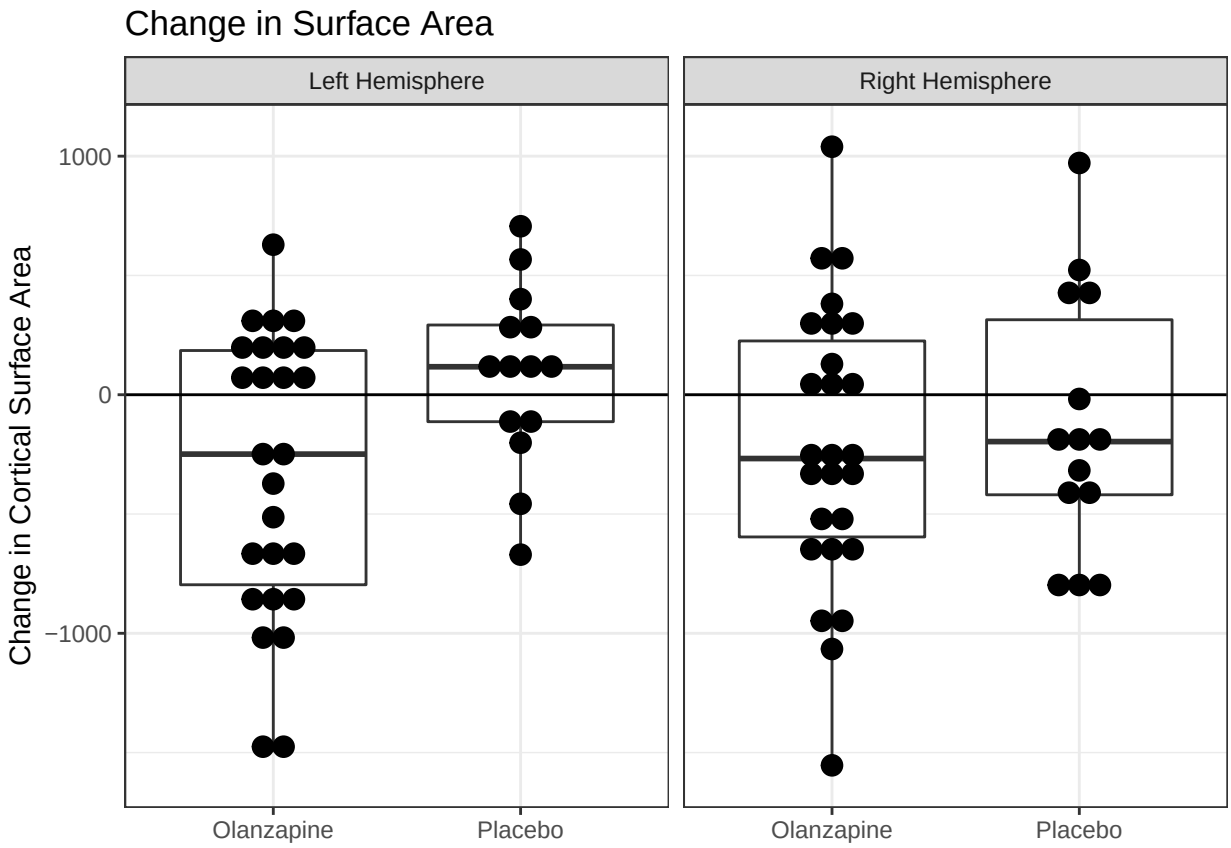
```
#boxplot of difference in thickness (y axis) by randomization group (x axis)
ggplot(RCT_SA, aes(x= RandomArm, y = RSurfArea_change)) +
     geom_boxplot(outlier.shape = NA) +
     geom_dotplot(binaxis = 'y', stackdir = 'center') +
     geom_hline(yintercept = 0) +
     ggtitle("Cortical thickness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

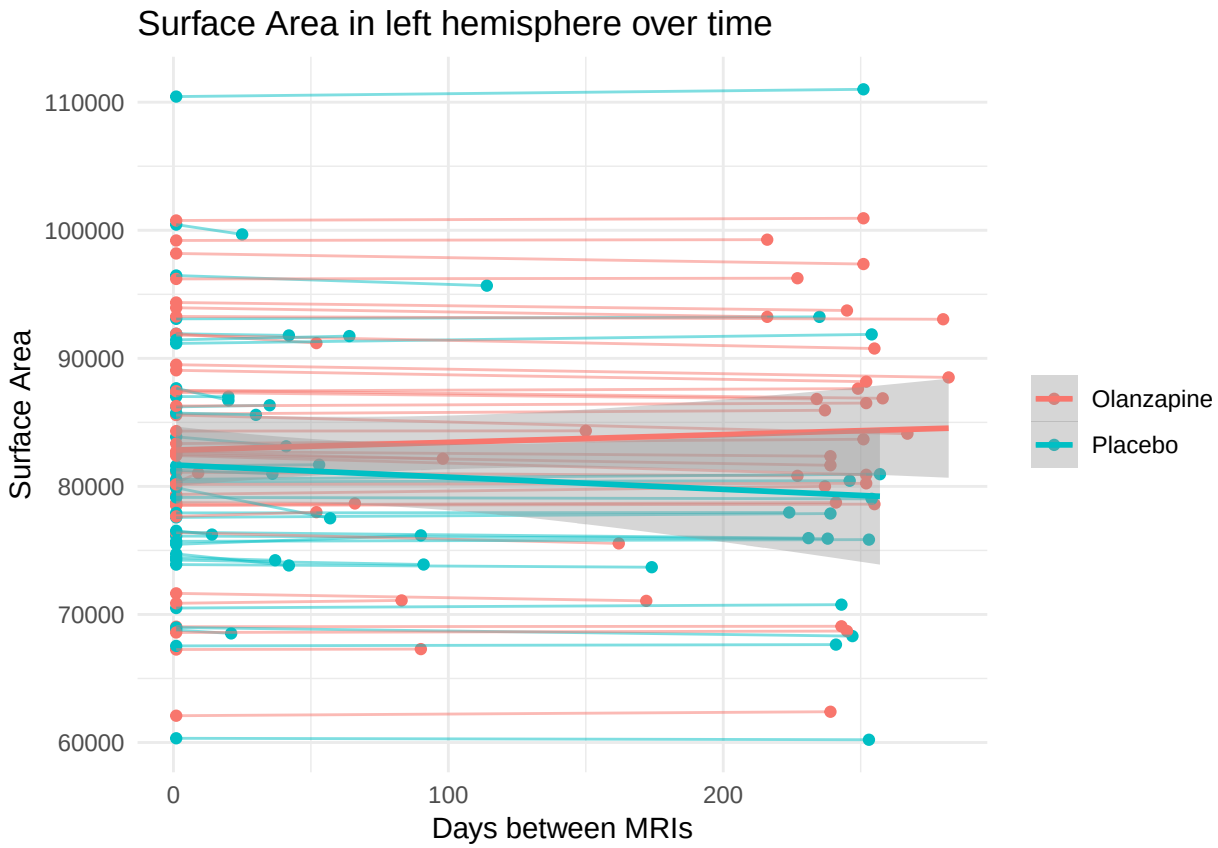
RCT & Relapse (with time as factor)

```
#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

```
#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_LS
  summary(fit_all)
```

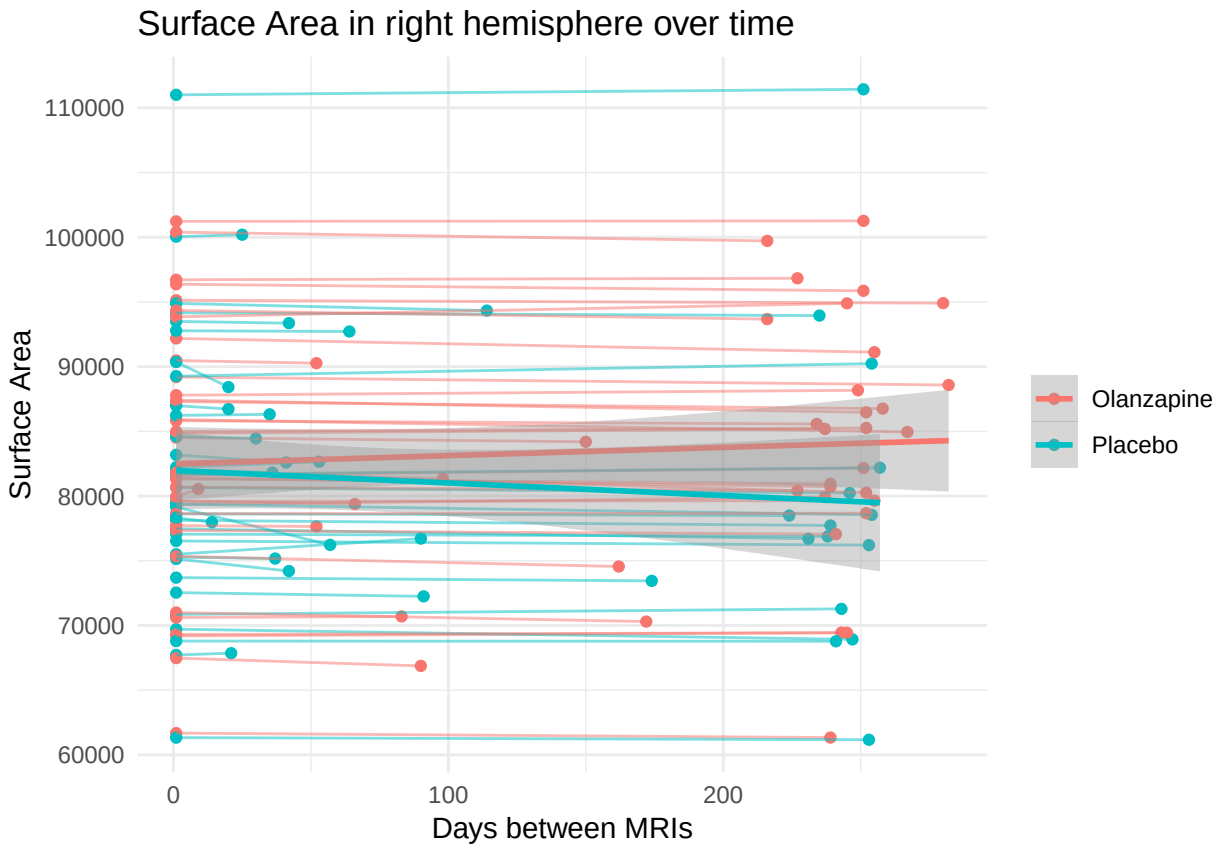
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## SurfArea ~ RandomArm * model_days + sex + age + site + (1 | STUDYID)
##    Data: RCTRelapse_LSA
##
## REML criterion at convergence: 2482.9
##
## Scaled residuals:
##      Min       1Q   Median       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
```

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()
```




```
#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_RS
  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
```

Dealing with the confusion..

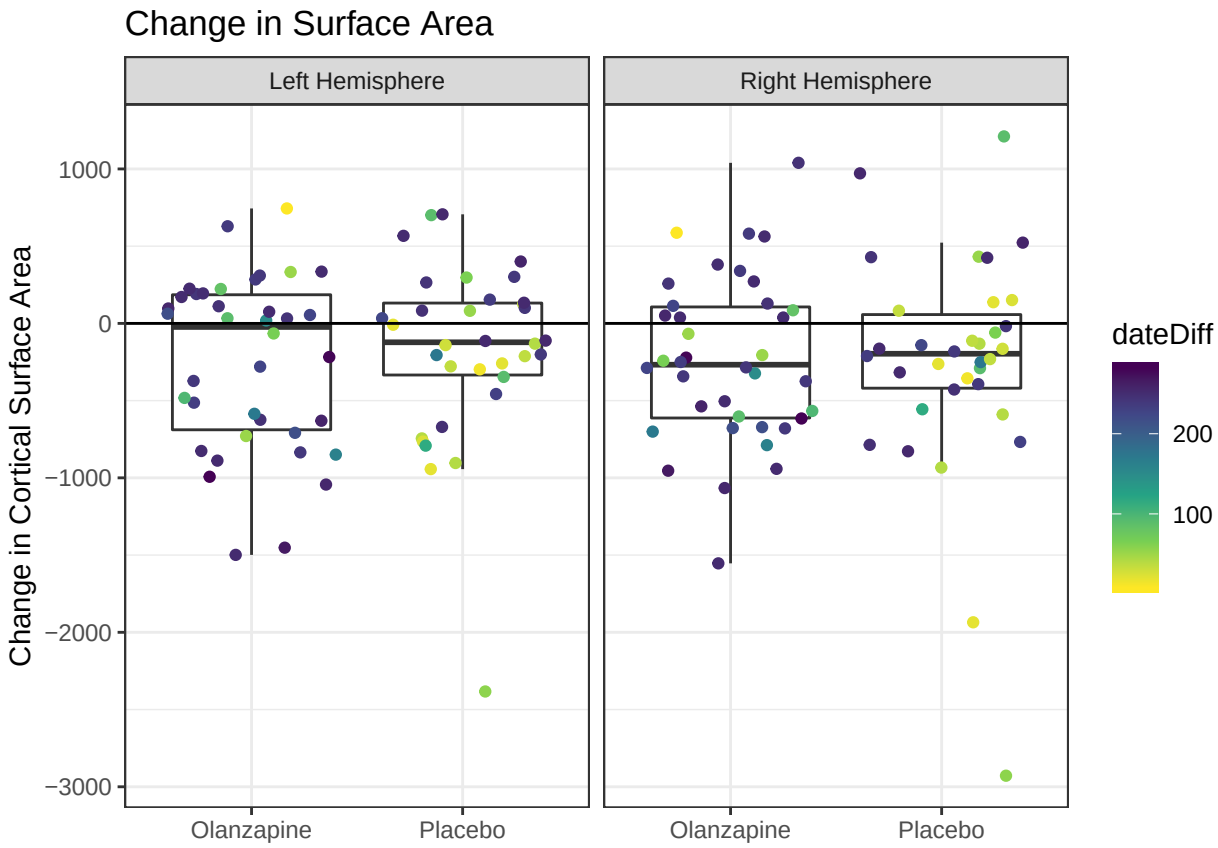
So I (maybe for one) Was Confused by the way that the two findings above seems to go in opposite directions.

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()
```


Change in Surface Area
Left Hemisphere
Right Hemisphere
Change in Cortical Surface Area
1000
0
−1000
−2000
−3000
Olanzapine
Placebo
Olanzapine
Placebo
dateDiff
200
100