# Reanalysis of 06-Wile

Simon Schwab*

01/06/2020

## Contents



## Reference

Wile et al. (2017). Serotonin and dopamine transporter PET changes in the premotor phase of LRRK2 parkinsonism: cross-sectional studies. *Lancet Neurology*, 16(5), 351–359. https://doi.org/10.1016/S1474-4422(17)30056-X

## Notes from reading methods section

- Dependant variable: PET radiotracer called 11C-DASB in cortex (first of many ANCOVAs reported)
- Independent variable: Group

*University of Zurich, simon.schwab@uzh.ch

  - Healty controls (n=9)
  - LRRK2 w/o PD (n=9)
  - LRRK2 w PD (n=7)
  - sporadic PD (n=13)
- Covariate: age
- age was not estimable for LRRK2 without manifest Parkinson's disease
- Design: 1-way ANCOVA with group (4 levels) as IV and age as covariate

# Reading data

Data is loaded, reshaped if necessary, and factors are specified.

```
PATH = file.path(path.expand("~"), "Data", "ancova") # ancova project folder
data = read_excel(file.path(PATH, "dataPrimaryStudies", "06-Wile", "06-Wile.xlsx"))

# group variable is dummy coded, key for AncovaVariable file
# 0 = healthy control, 1 = LRRK2 premanifest, 2 = LRRK2 affected, 3 = sporadic PD
data$Group.factor = NA
data$Group.factor[data$Group == 0] = "healthy control"
data$Group.factor[data$Group == 1] = "LRRK2 premanifest"
data$Group.factor[data$Group == 2] = "LRRK2 affected"
data$Group.factor[data$Group == 3] = "sporadic PD"
data$Group.factor = factor(data$Group.factor,
                           levels = c("healthy control", "LRRK2 premanifest",
                                      "LRRK2 affected", "sporadic PD"))
```

# Descriptives

## Dependant variable

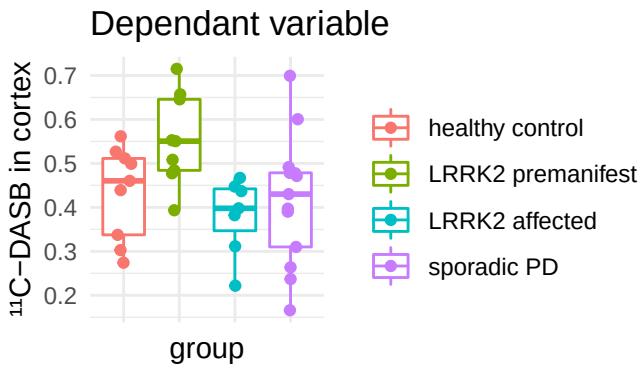
Number of samples and mean (SD) in levels of the independent variables. We reproduce Table 3 and Figure 2A of the study.

```
idx = c(1, 3, 2, 4) # sorting as in publication
tab.dv = array(NA, dim=c(4,3))
tab.dv[,1] = levels(data$Group.factor)
tab.dv[,2] = summary(data$Group.factor)
tab.dv[,3] = tapply(data$Cortex, data$Group.factor,
                    function (x) sprintf("%0.2f (%0.2f)", mean(x), sd(x)))
colnames(tab.dv) = c("group", "n", "mean (SD)")
print(tab.dv)
```

```
##      group               n    mean (SD)
## [1,] "healthy control"   "9"  "0.43 (0.10)"
## [2,] "LRRK2 premanifest" "9"  "0.55 (0.10)"
## [3,] "LRRK2 affected"    "7"  "0.38 (0.09)"
## [4,] "sporadic PD"       "13" "0.42 (0.15)"
```

```
ggplot(data, aes(y=Cortex, x=Group.factor, color=Group.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
```

```
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  xlab("group") + ylab("¹¹C-DASB in cortex") + ggtitle("Dependant variable")
```

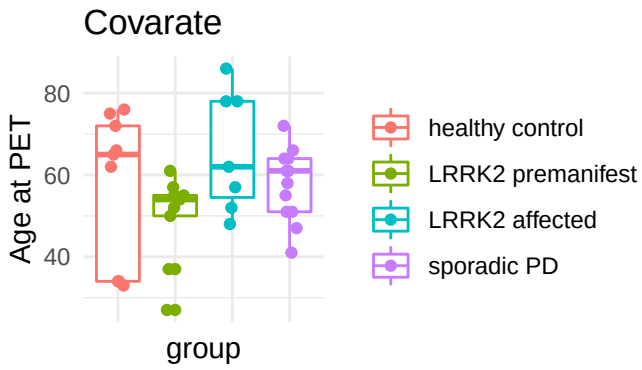



## Covariate(s)

```
idx = c(1, 3, 2, 4) # sorting as in publication
tab.cv = array(NA, dim=c(4,3))
tab.cv[,1] = levels(data$Group.factor)
tab.cv[,2] = summary(data$Group.factor)
tab.cv[,3] = tapply(data$`Age at PET`, data$Group.factor,
                    function (x) sprintf("%0.1f (%0.1f)", mean(x), sd(x)))
colnames(tab.cv) = c("group", "n", "mean (SD)")
print(tab.cv)
```

```
##      group               n    mean (SD)
## [1,] "healthy control"   "9"  "57.4 (18.4)"
## [2,] "LRRK2 premanifest" "9"  "49.8 (10.8)"
## [3,] "LRRK2 affected"    "7"  "65.9 (14.7)"
## [4,] "sporadic PD"       "13" "57.8 (8.5)"
```

```
ggplot(data, aes(y=`Age at PET`, x=Group.factor, color=Group.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  xlab("group") + ylab("Age at PET") + ggtitle("Covarate")
```



## Main analysis ANCOVA

```
# Orthogonal contrasts
contrasts(data$Group.factor) = contr.helmert(4)

fit.ancova = aov(Cortex ~ `Age at PET` + Group.factor, data = data)
# result = summary(fit) # Type I
result = Anova(fit.ancova, type=3) # Type III
print(result)
```

```
## Anova Table (Type III tests)
##
## Response: Cortex
##                Sum Sq Df F value    Pr(>F)
## (Intercept)  0.36361  1 25.1466 1.767e-05 ***
## `Age at PET` 0.00073  1  0.0502    0.8241
## Group.factor 0.12125  3  2.7951    0.0555 .
## Residuals    0.47717 33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Comparing ANCOVA in original study with reanalysis

### Independent variable

```
tab.IV = rbind(stats.orig.IV, stats.rep.IV)
rownames(tab.IV) = c("original Study", "reanalysis")
print(t(tab.IV))
```

```
##         original Study reanalysis
## Fvalue  NA             "2.80"
## df1     NA             " 3"
## df2     NA             "33"
## pvalue  "n.s."         "0.056"
## MD      NA             NA
## lowerCI NA             NA
## upperCI NA             NA
```

### Covariate

```
tab.CV = rbind(stats.orig.CV, stats.rep.CV)
rownames(tab.CV) = c("original Study", "reanalysis")
print(t(tab.CV))
```

```
##         original Study reanalysis
## Fvalue  NA             "0.05"
## df1     NA             " 1"
## df2     NA             "33"
## pvalue  NA             "0.82"
## MD      NA             NA
```

```
## lowerCI NA             NA
## upperCI NA             NA
```

# Assumptions

## 1. Homogeneity of variance

- ANOVA/ANCOVA is fairly robust in terms of the error rate when sample sizes are equal.
- When groups with larger sample sizes have larger variances than the groups with smaller sample sizes, the resulting F-ratio tends to be conservative. That is, it's more likely to produce a non-significant result when a genuine difference does exist in the population.
- Conversely, when the groups with larger sample sizes have smaller variances than the groups with smaller samples sizes, the resulting F-ratio tends to be liberal and can inflate the false positive rate.

```
tapply(data$Cortex, data$Group.factor, sd)
```

```
##   healthy control LRRK2 premanifest    LRRK2 affected       sporadic PD
##        0.10480960        0.10221440        0.08699411        0.14748558
```

```
leveneTest(Cortex ~ Group.factor, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3   0.821 0.4914
##       34
```

## 2. Independence between covariate and IV

When the covariate and the experimental effect (independent variable) are not independent the treatment effect is obscured, spurious treatment effects can arise and the interpretation of the ANCOVA is seriously compromised.

We test whether our groups differ on the CV. If the groups do not significantly differ then is appropriate to use the covariate.

```
fit.cv = aov(`Age at PET` ~ Group.factor, data = data)
Anova(fit.cv, type=3)
```

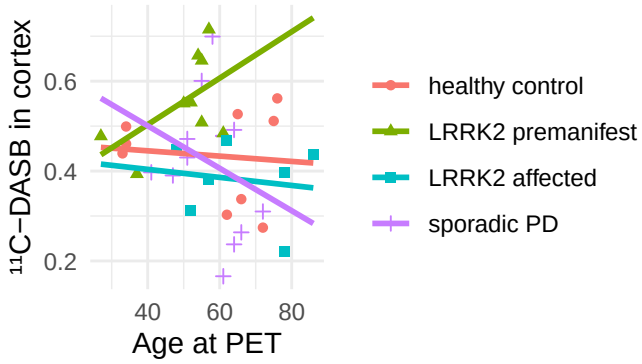
```
## Anova Table (Type III tests)
##
## Response: Age at PET
##              Sum Sq Df  F value Pr(>F)
## (Intercept)  120648  1 703.5665 <2e-16 ***
## Group.factor   1026  3   1.9942 0.1334
## Residuals      5830 34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 3. Homogeneity of regression slopes

```
fit.hrs = aov(Cortex ~ `Age at PET`*Group.factor, data = data)
Anova(fit.hrs, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: Cortex
##                              Sum Sq Df F value    Pr(>F)
## (Intercept)                 0.31357  1 21.8078 5.908e-05 ***
## `Age at PET`                0.00032  1  0.0226    0.8816
## Group.factor                0.02337  3  0.5417    0.6574
## `Age at PET`:Group.factor 0.04580  3  1.0617    0.3800
## Residuals                   0.43137 30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(data, aes(y=Cortex, x=`Age at PET`, color=Group.factor, shape=Group.factor)) +
  geom_point() +
  geom_smooth(formula = y ~ x,method=lm, se=FALSE, fullrange=TRUE) +
  theme_minimal() +
  theme(legend.title = element_blank()) +
  xlab("Age at PET") + ylab("[11]C-DASB in cortex")
```




## Additional analyses: MANCOVA

The paper performed four ANCOVAs for cortex, striatum, brainstem, and hypothalamus but did not correct for multiple testing. Therefore, we also perform a MANCOVA.

```
fit = manova(cbind(Cortex, Striatum, Brainstem, Hypothalamus) ~ Group.factor + `Age at PET`,
             data=data)
summary(fit)
```

```
##              Df  Pillai approx F num Df den Df    Pr(>F)
## Group.factor  3 0.95385   3.7293     12     96 0.0001241 ***
## `Age at PET`  1 0.36886   4.3832      4     30 0.0065618 **
## Residuals    33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Notes

- The first reported ANCOVA which was n.s. was reproduced
- p-value was not reported

- Authors also report age unadjusted analyses and report these as ANCOVAs instead of ANOVAs (p. 357) which was confusing
- Assumptions were met, except homogeneity of regression slopes showing a different relationship of outcome and covariate in the LRRK2 premanifest group
- Many analyses, they corrected for multiplicity

Data was analyzed according to recommendations by Field, Miles, & Field (2012).