

25-VanLeijsen

Audrey Yeo & Simon Schwab*

11/13/2019

Contents

Reference	1
Notes from reading methods section	1
Reading data	2
Descriptives	2
Dependant variable	2
Covariates(s)	4
Age	4
Sex	5
Hypertension	5
Total brain volume	6
Main analysis ANCOVA	7
Comparing ANCOVA in original study with reanalysis	7
Assumptions	7
1. Homogeneity of variance	7
2. Independence between covariate and IV	8
3. Homogeneity of regression slopes	9
Notes	11

Reference

Van Leijsen et al. (2017). Plasma AB (Amyloid-B) Levels and Severity and Progression of Small Vessel Disease. Stroke, 16(5), 351–359. <https://doi.org/10.1161/STROKEAHA.117.019810>

Notes from reading methods section

- Dependant variable: AB38
- The independant variables
 - presence of microbleeds (1 or more microbleeds (n = 81) vs no microbleed (n=406))

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

- presence of lacunes (1 or more lacunes (n = 132) vs no lacunes (n = 355))
- Covariate: age, sex and hypertension, total brain volume

```
stats.orig.IV = data.frame(
  Fvalue = NA,
  df1 = NA,
  df2 = NA,
  pvalue = "<0.01") # for microbleed presence and for lacune presence groups

stats.orig.allCV = data.frame(
  Fvalue = NA,
  df1 = NA,
  df2 = NA,
  pvalue = NA)
```

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", "25-VanLeijssen", "25-VanLeijssen-1.xlsx"))
data = data.frame(data)

data$mb_presence_b.factor = NA
data$mb_presence_b.factor[data$mb_presence_b == 1] = "microbleeds"
data$mb_presence_b.factor[data$mb_presence_b == 0] = "no microbleeds"
data$mb_presence_b.factor = as.factor(data$mb_presence_b.factor)

data$lac_presence_b.factor = NA
data$lac_presence_b.factor[data$lac_presence_b == 1] = "lacunes"
data$lac_presence_b.factor[data$lac_presence_b == 0] = "no lacunes"
data$lac_presence_b.factor = as.factor(data$lac_presence_b.factor)

data$sex.factor = NA
data$sex.factor[data$sex == 1] = "male"
data$sex.factor[data$sex == 2] = "female"
data$sex.factor = as.factor(data$sex.factor)

data$hypertension.factor = data$hypertension == 1

# data$groups = as.factor(paste(data$mb_presence_b.factor, data$lac_presence_b.factor))
```

Descriptives

Dependant variable

Number of samples and mean (SD) in levels of the independant variables. We reproduce the mean and sd values of Table 2 of this study

```

tab.dv = array(NA, dim=c(4,2))
rownames(tab.dv) = c(rev(levels(data$mb_presence_b.factor)), rev(levels(data$lac_presence_b.factor)))
colnames(tab.dv) = c("n", "mean (SD)")

tab.dv[,1] = c(rev(summary(data$mb_presence_b.factor)),
               rev(summary(data$lac_presence_b.factor)))

tab.dv[1,2] = sprintf("%.1f (%.1f)", mean(data$AB38[data$mb_presence_b == 0]), sd(data$AB38[data$mb_pre
tab.dv[2,2] = sprintf("%.1f (%.1f)", mean(data$AB38[data$mb_presence_b == 1]), sd(data$AB38[data$mb_pre

tab.dv[3,2] = sprintf("%.1f (%.1f)", mean(data$AB38[data$lac_presence_b == 0]), sd(data$AB38[data$lac_p
tab.dv[4,2] = sprintf("%.1f (%.1f)", mean(data$AB38[data$lac_presence_b == 1]), sd(data$AB38[data$lac_p

print(tab.dv)

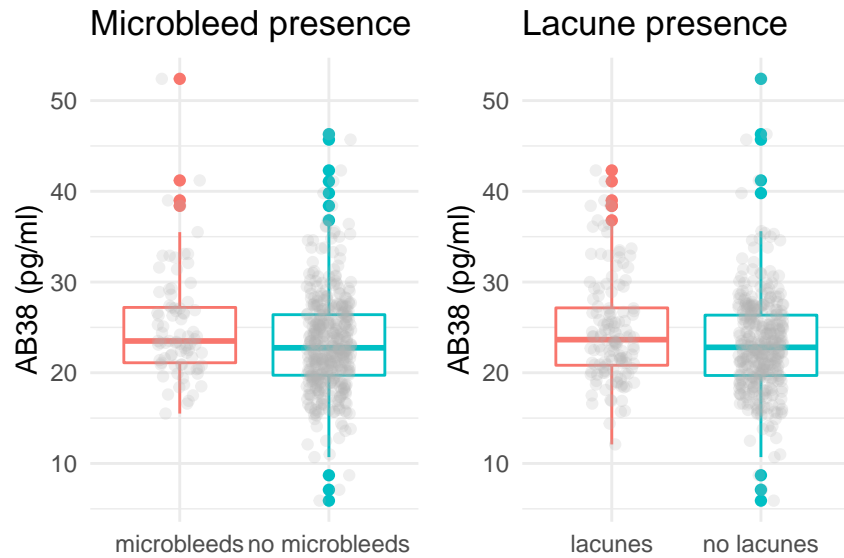
##              n      mean (SD)
## no microbleeds "406" "23.2 (5.5)"
## microbleeds   "81"  "25.2 (6.1)"
## no lacunes    "355" "23.1 (5.5)"
## lacunes       "132" "24.7 (5.9)"

p1 = ggplot(data, aes(y=AB38, x = mb_presence_b.factor, color = mb_presence_b.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0), alpha=0.2, col="gray70") +
  theme_minimal() +
  theme(legend.title = element_blank(),
        axis.title.x = element_blank(),
        legend.position = "none") +
  ggtitle("Microbleed presence") + ylab("AB38 (pg/ml)")

p2 = ggplot(data, aes(y=AB38, x = lac_presence_b.factor, color =lac_presence_b.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0), alpha=0.2, col="gray70") +
  theme_minimal() +
  theme(legend.title = element_blank(),
        axis.title.x = element_blank(),
        legend.position = "none") +
  ggtitle("Lacune presence") + ylab("AB38 (pg/ml)")

plot_grid(p1, p2, nrow = 1, ncol = 2)

```

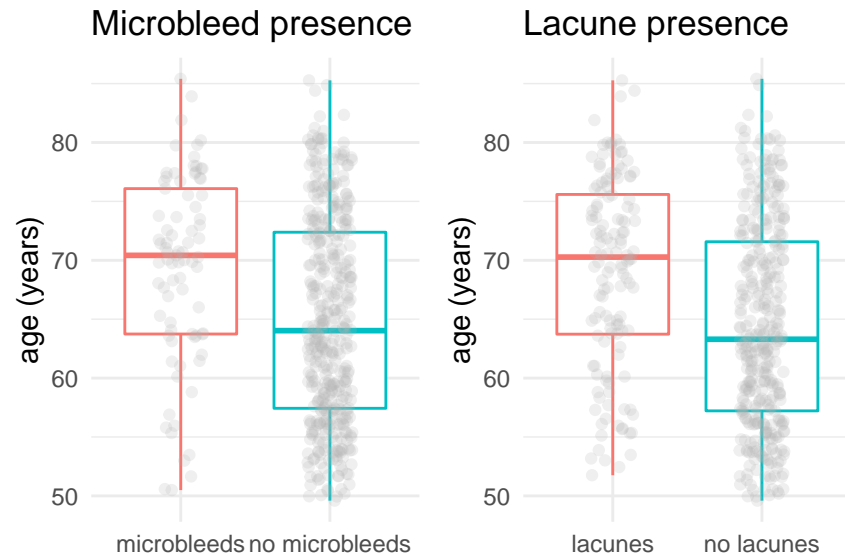


Covariates(s)

Age

```
p1 = ggplot(data, aes(y=age, x = mb_presence_b.factor, color = mb_presence_b.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0), alpha=0.2, col="gray70") +
  theme_minimal() +
  theme(legend.title = element_blank(),
        axis.title.x = element_blank(),
        legend.position = "none") +
  ggtitle("Microbleed presence") + ylab("age (years)")

p2 = ggplot(data, aes(y=age, x = lac_presence_b.factor, color = lac_presence_b.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0), alpha=0.2, col="gray70") +
  theme_minimal() +
  theme(legend.title = element_blank(),
        axis.title.x = element_blank(),
        legend.position = "none") +
  ggtitle("Lacune presence") + ylab("age (years)")
plot_grid(p1, p2, nrow = 1, ncol = 2)
```



Sex

```
tab = table(data$mb_presence_b.factor, data$sex.factor)
prc = tab/rowSums(tab)
tab[,1] = sprintf("%.1f%%", prc*100)[1:2]
tab[,2] = sprintf("%.1f%%", prc*100)[3:4]
print(tab)
```

```
##
##           female male
## microbleeds  38.3% 61.7%
## no microbleeds 43.8% 56.2%
```

```
tab = table(data$lac_presence_b.factor, data$sex.factor)
prc = tab/rowSums(tab)
tab[,1] = sprintf("%.1f%%", prc*100)[1:2]
tab[,2] = sprintf("%.1f%%", prc*100)[3:4]
print(tab)
```

```
##
##           female male
## lacunes      37.9% 62.1%
## no lacunes   44.8% 55.2%
```

Hypertension

```
tab = table(data$mb_presence_b.factor, data$hypertension.factor)
prc = tab/rowSums(tab)
tab[,1] = sprintf("%.1f%%", prc*100)[1:2]
tab[,2] = sprintf("%.1f%%", prc*100)[3:4]
print(tab)
```

```
##
##           FALSE TRUE
```

```
## microbleeds 18.5% 81.5%
## no microbleeds 28.6% 71.4%

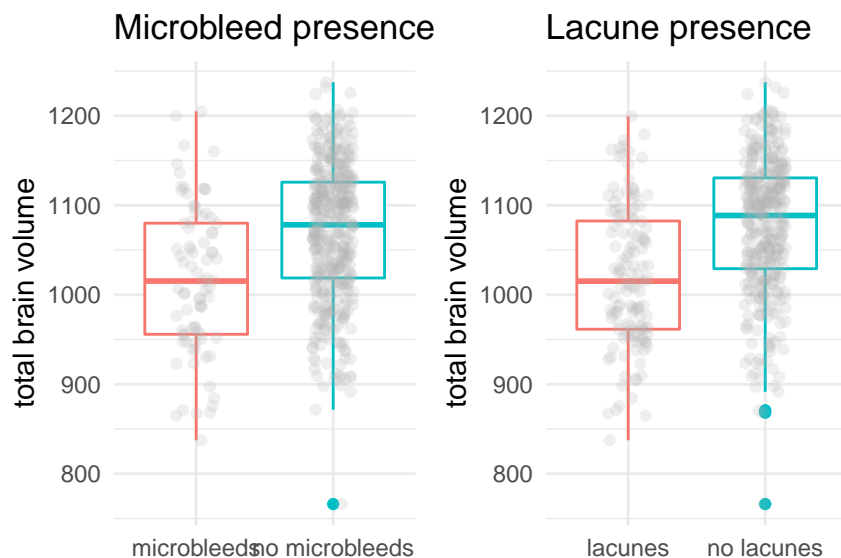
tab = table(data$lac_presence_b.factor, data$hypertension.factor)
prc = tab/rowSums(tab)
tab[,1] = sprintf("%.1f%%", prc*100)[1:2]
tab[,2] = sprintf("%.1f%%", prc*100)[3:4]
print(tab)

##
## FALSE TRUE
## lacunes 13.6% 86.4%
## no lacunes 31.8% 68.2%
```

Total brain volume

```
p1 = ggplot(data, aes(y=tbv_b, x = mb_presence_b.factor, color = mb_presence_b.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0), alpha=0.2, col="gray70") +
  theme_minimal() +
  theme(legend.title = element_blank(),
        axis.title.x = element_blank(),
        legend.position = "none") +
  ggtitle("Microbleed presence") + ylab("total brain volume")

p2 = ggplot(data, aes(y=tbv_b, x = lac_presence_b.factor, color = lac_presence_b.factor)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0), alpha=0.2, col="gray70") +
  theme_minimal() +
  theme(legend.title = element_blank(),
        axis.title.x = element_blank(),
        legend.position = "none") +
  ggtitle("Lacune presence") + ylab("total brain volume")
plot_grid(p1, p2, nrow = 1, ncol = 2)
```



Main analysis ANCOVA

```
# Orthogonal contrasts
contrasts(data$mb_presence_b.factor) = contr.helmert(2)
contrasts(data$lac_presence_b.factor) = contr.helmert(2)

# fit.ancova= aov(AB38 ~ mb_presence_b.factor*lac_presence_b.factor , data = data)
fit.ancova = aov(AB40 ~ age + sex.factor + hypertension.factor + tbv_b + mb_presence_b.factor*lac_presence_b.factor, data = data)
result = Anova(fit.ancova, type = 3) # Type III

print(result)

## Anova Table (Type III tests)
##
## Response: AB40
##
##              Sum Sq Df F value    Pr(>F)
## (Intercept)    17807   1  13.6054 0.0002514 ***
## age            16434   1  12.5570 0.0004334 ***
## sex.factor         195   1   0.1493 0.6993628
## hypertension.factor  1963   1   1.4997 0.2213140
## tbv_b            2023   1   1.5454 0.2144299
## mb_presence_b.factor  7304   1   5.5811 0.0185545 *
## lac_presence_b.factor  1730   1   1.3216 0.2508822
## mb_presence_b.factor:lac_presence_b.factor    129   1   0.0983 0.7540404
## Residuals      626906 479
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing ANCOVA in original study with reanalysis

We skip this section as no p-values or test statistics are reported.

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 sample sizes, the resulting F-ratio tends to be liberal and can inflate the false positive rate.
- In this study, statistical descriptives show that the highest and lowest variances seem close. Also, there is homogeneity of variance as p values exceed 0.05 for the Levene's test

```
tapply(data$AB38, data$mb_presence_b.factor, sd)
```

```
##      microbleeds no microbleeds
##      6.069102      5.471873
```

```

leveneTest(AB38 ~ mb_presence_b.factor, data = data)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 1  0.0912 0.7628
##      485

tapply(data$AB38, data$lac_presence_b.factor, sd)

##      lacunes no lacunes
## 5.870429 5.459909

leveneTest(AB38 ~ lac_presence_b.factor, data = data)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 1  1.3772 0.2412
##      485

```

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.

```

# Age
fit.cv.age = aov(age ~ lac_presence_b + mb_presence_b, data = data)
Anova(fit.cv.age, type=3)

## Anova Table (Type III tests)
##
## Response: age
##      Sum Sq Df    F value    Pr(>F)
## (Intercept) 1372627 1 19282.9584 < 2.2e-16 ***
## lac_presence_b 1334 1 18.7464 1.816e-05 ***
## mb_presence_b 638 1 8.9616 0.002898 **
## Residuals 34453 484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Sex
fit.cv.sex = glm(sex.factor ~ lac_presence_b + mb_presence_b, family = binomial, data = data)
Anova(fit.cv.sex, type=3)

## Analysis of Deviance Table (Type III tests)
##
## Response: sex.factor
##      LR Chisq Df Pr(>Chisq)
## lac_presence_b 1.36028 1 0.2435
## mb_presence_b 0.33341 1 0.5637

# Hypertension
fit.cv.hypertension = glm(hypertension.factor ~ lac_presence_b + mb_presence_b, family = binomial, data = data)
Anova(fit.cv.hypertension, type=3)

```



```
## Analysis of Deviance Table (Type III tests)
##
## Response: hypertension.factor
##           LR Chisq Df Pr(>Chisq)
## lac_presence_b 14.8425 1 0.0001169 ***
## mb_presence_b   0.7439 1 0.3884022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Total brain volume
fit.cv.tbw = aov(tbv_b ~ lac_presence_b + mb_presence_b, data = data)
Anova(fit.cv.age, type=3)

## Anova Table (Type III tests)
##
## Response: age
##           Sum Sq Df    F value    Pr(>F)
## (Intercept) 1372627 1 19282.9584 < 2.2e-16 ***
## lac_presence_b   1334 1   18.7464 1.816e-05 ***
## mb_presence_b    638 1    8.9616 0.002898 **
## Residuals     34453 484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Homogeneity of regression slopes

- We test the interaction between the IV and the CV

```
# Age
Anova(aov(AB38 ~ age*mb_presence_b.factor, data = data), type=3)

## Anova Table (Type III tests)
##
## Response: AB38
##           Sum Sq Df F value    Pr(>F)
## (Intercept)    581.2 1 19.6293 1.164e-05 ***
## age            528.9 1 17.8617 2.841e-05 ***
## mb_presence_b.factor    16.5 1 0.5560 0.4562
## age:mb_presence_b.factor    28.4 1 0.9603 0.3276
## Residuals     14301.0 483
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova(aov(AB38 ~ age*lac_presence_b.factor, data = data), type=3)

## Anova Table (Type III tests)
##
## Response: AB38
##           Sum Sq Df F value    Pr(>F)
## (Intercept)   1056.2 1 35.5735 4.750e-09 ***
## age           616.0 1 20.7488 6.639e-06 ***
## lac_presence_b.factor    4.1 1 0.1398 0.7087
## age:lac_presence_b.factor   10.4 1 0.3490 0.5550
## Residuals     14340.2 483
## ---
```

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

```
# Sex
```

```
Anova(aov(AB38 ~ sex.factor*mb_presence_b.factor, data = data), type=3)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: AB38
```

	Sum Sq	Df	F value	Pr(>F)
## (Intercept)	59156	1	1906.9045	< 2e-16 ***
## sex.factor	45	1	1.4411	0.23055
## mb_presence_b.factor	6	1	0.2052	0.65078
## sex.factor:mb_presence_b.factor	89	1	2.8782	0.09043 .
## Residuals	14984	483		

```
## ---
```

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

```
Anova(aov(AB38 ~ sex.factor*lac_presence_b.factor, data = data), type=3)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: AB38
```

	Sum Sq	Df	F value	Pr(>F)
## (Intercept)	85711	1	2753.4007	<2e-16 ***
## sex.factor	8	1	0.2410	0.6237
## lac_presence_b.factor	32	1	1.0324	0.3101
## sex.factor:lac_presence_b.factor	32	1	1.0291	0.3109
## Residuals	15035	483		

```
## ---
```

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

```
# Hypertension
```

```
Anova(aov(AB38 ~ hypertension.factor*mb_presence_b.factor, data = data), type=3)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: AB38
```

	Sum Sq	Df	F value	Pr(>F)
## (Intercept)	26794.2	1	880.3884	< 2.2e-16 ***
## hypertension.factor	214.6	1	7.0523	0.008178 **
## mb_presence_b.factor	16.8	1	0.5534	0.457306
## hypertension.factor:mb_presence_b.factor	6.0	1	0.1983	0.656270
## Residuals	14699.9	483		

```
## ---
```

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

```
Anova(aov(AB38 ~ hypertension.factor*lac_presence_b.factor, data = data), type=3)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: AB38
```

	Sum Sq	Df	F value	Pr(>F)
## (Intercept)	30776.7	1	1008.2994	< 2.2e-16 ***
## hypertension.factor	229.4	1	7.5146	0.006347 **
## lac_presence_b.factor	6.7	1	0.2190	0.640013
## hypertension.factor:lac_presence_b.factor	7.6	1	0.2495	0.617620
## Residuals	14742.8	483		

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

# Total brain volume
Anova(aov(AB38 ~ tbv_b*mb_presence_b.factor, data = data), type=3)

## Anova Table (Type III tests)
##
## Response: AB38
##
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	2705.7	1	90.1924	< 2.2e-16 ***
tbv_b	412.9	1	13.7621	0.0002316 ***
mb_presence_b.factor	10.3	1	0.3448	0.5573285
tbv_b:mb_presence_b.factor	6.0	1	0.2013	0.6538747
Residuals	14489.6	483		

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

Anova(aov(AB38 ~ tbv_b*lac_presence_b.factor, data = data), type=3)

## Anova Table (Type III tests)
##
## Response: AB38
##
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	3154.9	1	104.9384	< 2.2e-16 ***
tbv_b	440.8	1	14.6613	0.0001456 ***
lac_presence_b.factor	0.1	1	0.0035	0.9525567
tbv_b:lac_presence_b.factor	0.1	1	0.0029	0.9570343
Residuals	14520.9	483		

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

Notes

- We reproduced the non-significant result, i.e. the lack of evidence for microbleeds and lacunes to have an effect on Abeta38 using ANCOVA and the 4 covariates.
- However, we could not reproduce an effect of lacunes reported when not taking in account the covariates, and we could not reproduce the Ab40 result.
- No exact p-values and no test statistics are reported for the analysis with Abeta38 and Abeta40.
- Assumptions were not met: three covariates age, hypertension and total brain volume were not independent from the group variable (microbleed or lacunes).

Reference : Field, Miles & Miles (2012), Discovering Statistics, Published by Sage Pub, UK.