Assignment 5 - Meta-analysis of pitch in schizophrenia

Riccardo Fusaroli

3/7/2019

Building on the shoulders of giants: meta-analysis

Questions to be answered

1. What is the current evidence for distinctive vocal patterns in schizophrenia? Report how many papers report quantitative estimates, comment on what percentage of the overall studies reviewed they represent (see PRISMA chart) your method to analyze them, the estimated effect size of the difference (mean effect size and standard error) and forest plots representing it.

N.B. Only measures of pitch mean and pitch sd are required for the assignment (all the other measures have been removed from the data-set for the sake of simplicity).

2. Do the results match your own analysis from Assignment 3? If you add your results to the meta-analysis, do the estimated effect sizes change? Report the new estimates and the new forest plots.

3. Assess the quality of the literature: report and comment on heterogeneity of the studies (tau, I2), on publication bias (funnel plot), and on influential studies.

Tips on the process to follow:

- Download the data on all published articles analyzing voice in schizophrenia and the prisma chart as reference of all articles found and reviewed Data: https://www.dropbox.com/s/0l9ur0gaabr80a8/Matrix_MetaAnalysis_Diagnosis_updated290719.xlsx?dl=0 Prisma chart: https://www.dropbox.com/s/vbjf6ff73dv9tru/PRISMA%202009%20flow%20diagram_PP.pptx?dl=0
- Look through the dataset to find out which columns to use, and if there is any additional information written as comments (real world data is always messy!).
  - Hint: PITCH_F0M and PITCH_F0SD group of variables are what you need
- Following the procedure in the slides calculate effect size and standard error of the effect size per each study. N.B. we focus on pitch mean (PITCH_F0_HC_M),(PITCH_F0_SZ_M) and pitch standard deviation (PITCH_F0SD_HC_M),(PITCH_F0SD_SZ_M). . first try using lmer (to connect to what you know of mixed effects models) . then use rma() (to get some juicy additional statistics)
- Build a forest plot of the results (forest(model))
- Go back to Assignment 3, add your own study to the data table, and re-run meta-analysis. Do the results change?
- Now look at the output of rma() and check tau and I2

SETUP and LOAD-IN

```
# Setup
pacman::p_load(readxl,tidyverse,tidyr, lmerTest, lme4,metafor)

# Read in excel file
df <- read_xlsx("Matrix_MetaAnalysis_Diagnosis_updated290719.xlsx")
```

```
## New names:
## * frequency -> frequency...68
## * frequency -> frequency...73
## * frequency -> frequency...78
## * frequency -> frequency...83
## * frequency -> frequency...88
## * ...
```

```
# Counting how long the columns are without NA's
df %>% summarise(
  PitchF0HC = length(na.omit(PITCH_F0_HC_M)),
  PitchF0SZ = length(na.omit(PITCH_F0_SZ_M)),
  PitchF0SDHC = length(na.omit(PITCH_F0SD_HC_M)),
  PitchF0SDSZ = length(na.omit(PITCH_F0SD_SZ_M))
)
```

```
## # A tibble: 1 x 4
##   PitchF0HC PitchF0SZ PitchF0SDHC PitchF0SDSZ
##       <int>     <int>       <int>       <int>
## 1         6        10          15          20
```

So, we have 6 complete case-studies for Pitch-mean in healthy controls, 10 complete ones in Pitch-mean in schizophrenics, 15 complete ones in Pitch-standard-deviation-mean in healthy controls (there will be overlap with the 6) and then 20 complete ones in Pitch-standard-deviation mean in schizophrenics (again, there will be overlap here with the 10).

GETTING STANDARDIZED MEAN DIFFERENCE (EFFECT SIZES)

This function "escalc" creates two new columns, where yi is the observed effect size and vi is the variance which we can use in our models. SMD = standardized mean difference, which is Cohen's d - a measure of effect size. We want it standardized because that allows us to compare across studies, and this is what we do in meta analyses.

```
# Creating effect size and sampling variance columns using escalc
WM_d <-
  escalc(
    measure = "SMD",
    n1i = SAMPLE_SIZE_HC,
    n2i = SAMPLE_SIZE_SZ,
    m1i = PITCH_F0_HC_M,
```

```
    m2i = PITCH_F0_SZ_M,
    sd1i = PITCH_F0_HC_SD,
    sd2i = PITCH_F0_SZ_SD,
    data = df
  )

# It can be seen here
test <-
  WM_d %>% select(StudyID,
                  yi,
                  vi,
                  PITCH_F0_HC_M,
                  PITCH_F0_SZ_M,
                  PITCH_F0SD_HC_M,
                  PITCH_F0SD_SZ_M)
```

MODELLING BEFORE ADDING STUDIES FROM ASSIGNMENT 3

```
# Making a mixed effects implementation
m <-
  lmer( yi ~ 1 + (1 | StudyID),
        weights = 1 / vi,
        data = WM_d,
        control = lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
        ) #Model may not have converged with 1 eigenvalue close to zero: 7.2e-10
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 7.2e-10
```

```
summary(m)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: yi ~ 1 + (1 | StudyID)
##    Data: WM_d
## Weights: 1/vi
## Control:
## lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
##
## REML criterion at convergence: 5.8
##
## Scaled residuals:
##        Min         1Q     Median         3Q        Max
## -0.0022761 -0.0011827 -0.0006473  0.0009493  0.0021579
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  StudyID  (Intercept) 1.295e-01 0.359808
```

```
##  Residual             5.472e-06 0.002339
## Number of obs: 6, groups:  StudyID, 6
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  -0.2065     0.1469  5.0000  -1.406    0.219
```

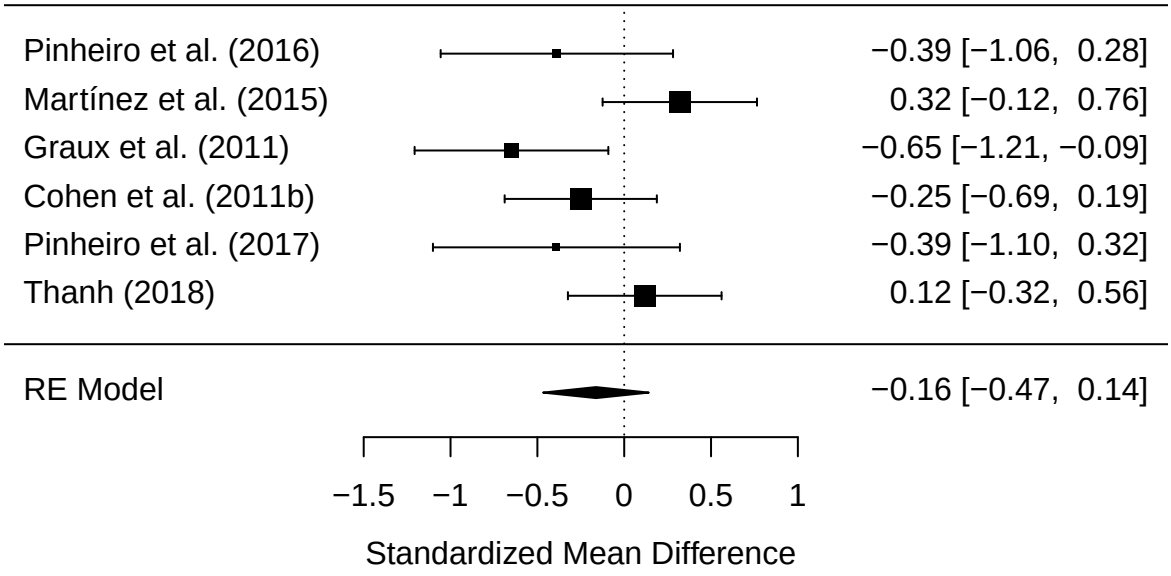
```
# A Meta-analysis optimization:
m2 <- rma(yi, vi, data = WM_d, slab = Article) # Slab = random effect. Put here as Article to get the a
```

```
## Warning in rma(yi, vi, data = WM_d, slab = Article): Studies with NAs omitted
## from model fitting.
```

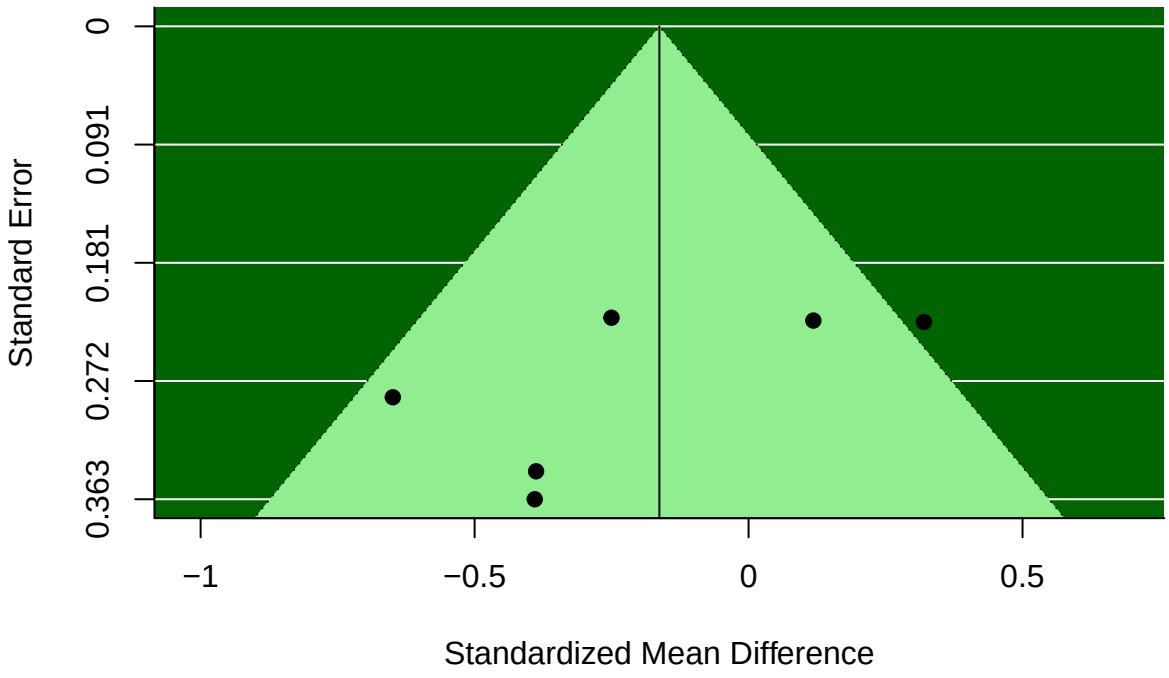
```
summary(m2)
```

```
##
## Random-Effects Model (k = 6; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
##  -2.1749    4.3497    8.3497    7.5686   14.3497
##
## tau^2 (estimated amount of total heterogeneity): 0.0712 (SE = 0.0908)
## tau (square root of estimated tau^2 value):      0.2668
## I^2 (total heterogeneity / total variability):   50.29%
## H^2 (total variability / sampling variability):  2.01
##
## Test for Heterogeneity:
## Q(df = 5) = 9.8472, p-val = 0.0797
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb   ci.ub
##  -0.1628  0.1554  -1.0476  0.2948  -0.4672  0.1417
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Making a forest and funnel plot
forest(m2)
```




```
funnel(x = m2, back = 'DarkGreen', shade = 'lightGreen') # Christmas colors
```




```
# Checking publication bias with ranktest
ranktest(m2)
```

```
##
## Rank Correlation Test for Funnel Plot Asymmetry
##
## Kendall's tau = -0.2000, p = 0.7194
```

```
# What if we have a fixed effect?
m3 <-
  lmer(yi ~ 1 + TYPE_OF_TASK + (1 | StudyID),
       weights = 1 / vi,
       data = WM_d,
       control = lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
       )
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 5.9e-10
```

```
summary(m3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: yi ~ 1 + TYPE_OF_TASK + (1 | StudyID)
##    Data: WM_d
## Weights: 1/vi
## Control:
## lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
##
## REML criterion at convergence: 4.6
##
## Scaled residuals:
##        Min         1Q     Median         3Q        Max
## -0.0014862 -0.0005657 -0.0002657  0.0000000  0.0018983
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  StudyID  (Intercept) 1.734e-01 0.416405
##  Residual             5.912e-06 0.002431
## Number of obs: 6, groups:  StudyID, 6
##
## Fixed effects:
##                     Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)        -0.27677    0.20820  3.00000  -1.329    0.276
## TYPE_OF_TASKFREE    0.02659    0.46555  3.00000   0.057    0.958
## TYPE_OF_TASKSOCIAL  0.39504    0.46555  3.00000   0.849    0.458
##
## Correlation of Fixed Effects:
##               (Intr) TYPE_OF_TASKF
## TYPE_OF_TASKF -0.447
## TYPE_OF_TASKS -0.447  0.200
## convergence code: 0
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

ADDING STUDIES FROM ASSIGNMENT 3:

```
# Making a placeholder dataframe of same dimensions
new <- data.frame(matrix(NA, nrow = 4, ncol = 147))

# Filtering study 1-4 - the Danish ones from assignment 3
olddf <- read_csv('pitch_data.csv') %>% filter(Study == 1:4) %>% mutate(Diagnosis = as.factor(Diagnosis
                                                Study = as.factor(Study),
                                                Subject = as.factor(Subject))
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
```

```
##   X1 = col_double(),
##   soundname = col_character(),
##   Filename = col_character(),
##   Study = col_double(),
##   Diagnosis = col_double(),
##   Subject = col_double(),
##   Trial = col_character(),
##   Mean = col_double(),
##   SD = col_double(),
##   IQR = col_double(),
##   Median = col_double()
## )
```

```
## Warning in Study == 1:4: longer object length is not a multiple of shorter
## object length
```

```
# Finding the information we need
obj <- olddf %>% group_by(Study, Diagnosis) %>%
  summarise(meanF0 = mean(Mean),
            SDF0 = sd(Mean),
            SampleSize = n_distinct(Subject)
            )
```

```
## `summarise()` regrouping output by 'Study' (override with `.groups` argument)
```

```
colnames(new) = colnames(df)

# Adding it - manually
new$ArticleID = c(49, 50, 51, 52)

new$StudyID = c(51, 52, 53, 54)

new$Authors = c('Unknown1', 'Unknown2', 'Unknown3', 'Unknown4')

new$PITCH_F0_HC_M = c(obj[1,3], obj[3,3], obj[5,3], obj[7,3])
new$PITCH_F0_SZ_M = c(obj[2,3], obj[4,3], obj[6,3], obj[8,3])

new$PITCH_F0_HC_SD = c(obj[1,4], obj[3,4], obj[5,4], obj[7,4])
new$PITCH_F0_SZ_SD = c(obj[2,4], obj[4,4], obj[6,4], obj[8,4])

new$SAMPLE_SIZE_HC = c(obj[1,5], obj[3,5], obj[5,5], obj[7,5])
new$SAMPLE_SIZE_SZ = c(obj[2,5], obj[4,5], obj[6,5], obj[8,5])

new$Article = c('A3P2 - Study 1', 'A3P2 - Study 2', 'A3P2 - Study 3', 'A3P2 - Study 4')

# The new dataframe is then the 6 studies plus the new 4 studies from Assignment 3:
new_df <- rbind(df, new)
```

GETTING STANDARDIZED MEAN DIFFERENCE (EFFECT SIZES) W. MORE STUDIES

```
# Creating effect size and sampling variance columns using escalc
WM_d2 <-
  escalc(
    measure = "SMD",
    n1i = as.numeric(SAMPLE_SIZE_HC),
    n2i = as.numeric(SAMPLE_SIZE_SZ),
    m1i = as.numeric(PITCH_F0_HC_M),
    m2i = as.numeric(PITCH_F0_SZ_M),
    sd1i = as.numeric(PITCH_F0_HC_SD),
    sd2i = as.numeric(PITCH_F0_SZ_SD),
    data = new_df
  )

# The new dataframe and effect sizes can be seen here
test2 <-
  WM_d2 %>% select(StudyID,
                   yi,
                   vi,
                   PITCH_F0_HC_M,
                   PITCH_F0_SZ_M,
                   PITCH_F0_HC_SD,
                   PITCH_F0_SZ_SD)
```

MODELLING AFTER ADDING STUDIES FROM ASSIGNMENT 3

```
# Making a mixed effects implementation
m22 <-
  lmer( yi ~ 1 + (1 | StudyID),
        weights = 1 / vi,
        data = WM_d2,
        control = lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
        ) #Model may not have converged with 1 eigenvalue close to zero: 7.2e-10
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(m22)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: yi ~ 1 + (1 | StudyID)
##    Data: WM_d2
## Weights: 1/vi
## Control:
## lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
```

```
##
## REML criterion at convergence: 8.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.73162 -0.79363 -0.09357  0.66473  1.33850
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  StudyID  (Intercept) 0.0      0.000
##  Residual             1.6      1.265
## Number of obs: 10, groups:  StudyID, 10
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) -0.02595    0.10608  9.00000  -0.245    0.812
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

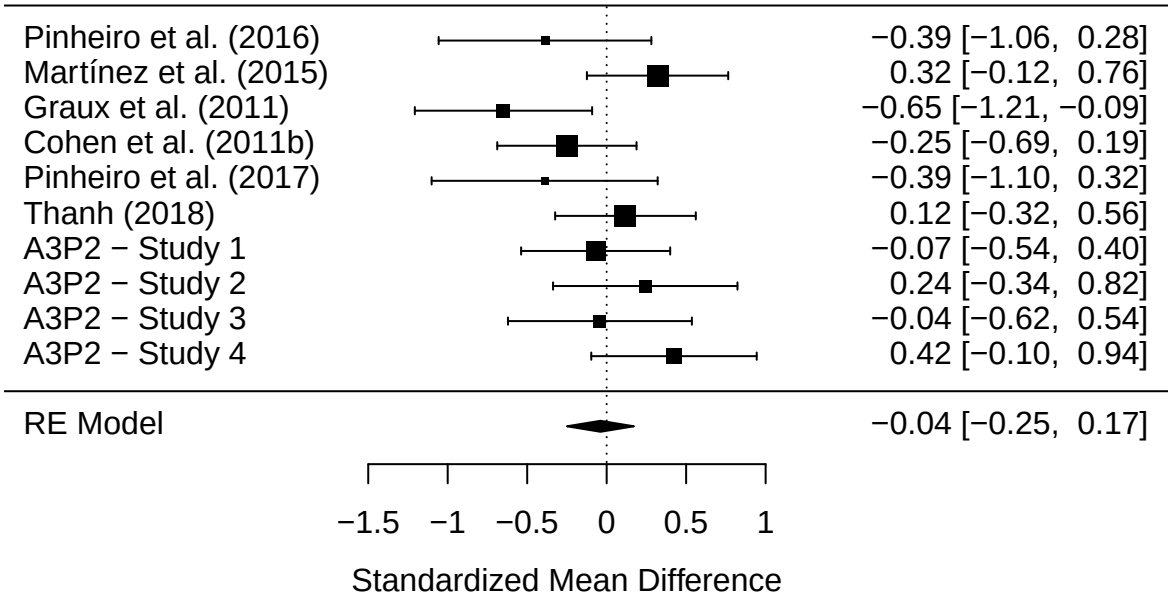
```
# A Meta-analysis optimization:
m22 <- rma(yi, vi, data = WM_d2, slab = Article)
```

```
## Warning in rma(yi, vi, data = WM_d2, slab = Article): Studies with NAs omitted
## from model fitting.
```

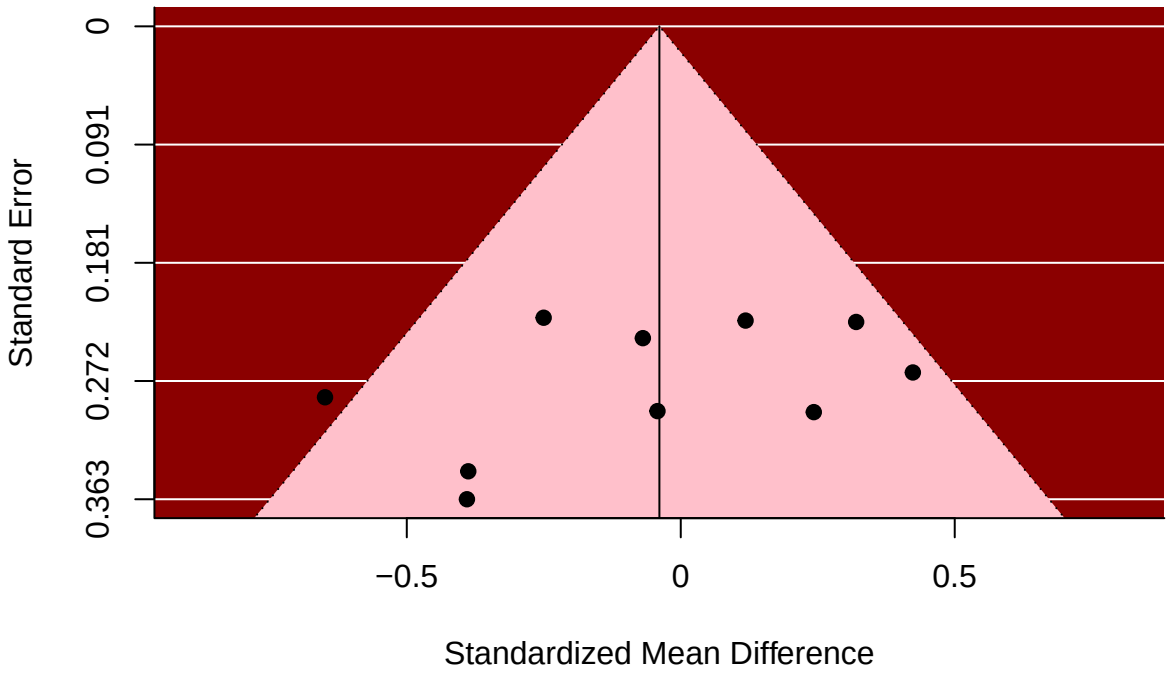
```
summary(m22)
```

```
##
## Random-Effects Model (k = 10; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
##  -3.2293    6.4586   10.4586   10.8531   12.4586
##
## tau^2 (estimated amount of total heterogeneity): 0.0424 (SE = 0.0537)
## tau (square root of estimated tau^2 value):      0.2060
## I^2 (total heterogeneity / total variability):   37.37%
## H^2 (total variability / sampling variability):  1.60
##
## Test for Heterogeneity:
## Q(df = 9) = 14.3982, p-val = 0.1088
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb   ci.ub
##  -0.0390  0.1074  -0.3628  0.7167  -0.2496  0.1716
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Making a forest and funnel plot
forest(m22)
```




```
funnel(x = m22, back = 'DarkRed', shade = 'pink', col = 'Black') # Christmas colors
```




```
# Checking publication bias with ranktest
ranktest(m22)
```

```
##
## Rank Correlation Test for Funnel Plot Asymmetry
##
## Kendall's tau = -0.1111, p = 0.7275
```

```
# What if we have a fixed effect?
m32 <-
  lmer(yi ~ 1 + TYPE_OF_TASK + (1 | StudyID),
       weights = 1 / vi,
       data = WM_d2,
       control = lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
       ) #Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 5.9e-10
```

```r
summary(m32)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: yi ~ 1 + TYPE_OF_TASK + (1 | StudyID)
##    Data: WM_d2
## Weights: 1/vi
## Control:
## lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.nRE = "ignore")
##
## REML criterion at convergence: 4.6
##
## Scaled residuals:
##        Min         1Q     Median         3Q        Max
## -0.0014862 -0.0005657 -0.0002657  0.0000000  0.0018983
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  StudyID  (Intercept) 1.734e-01 0.416405
##  Residual             5.912e-06 0.002431
## Number of obs: 6, groups:  StudyID, 6
##
## Fixed effects:
##                     Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)        -0.27677    0.20820  3.00000  -1.329    0.276
## TYPE_OF_TASKFREE    0.02659    0.46555  3.00000   0.057    0.958
## TYPE_OF_TASKSOCIAL  0.39504    0.46555  3.00000   0.849    0.458
##
## Correlation of Fixed Effects:
##               (Intr) TYPE_OF_TASKF
## TYPE_OF_TASKF -0.447
## TYPE_OF_TASKS -0.447  0.200
## convergence code: 0
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

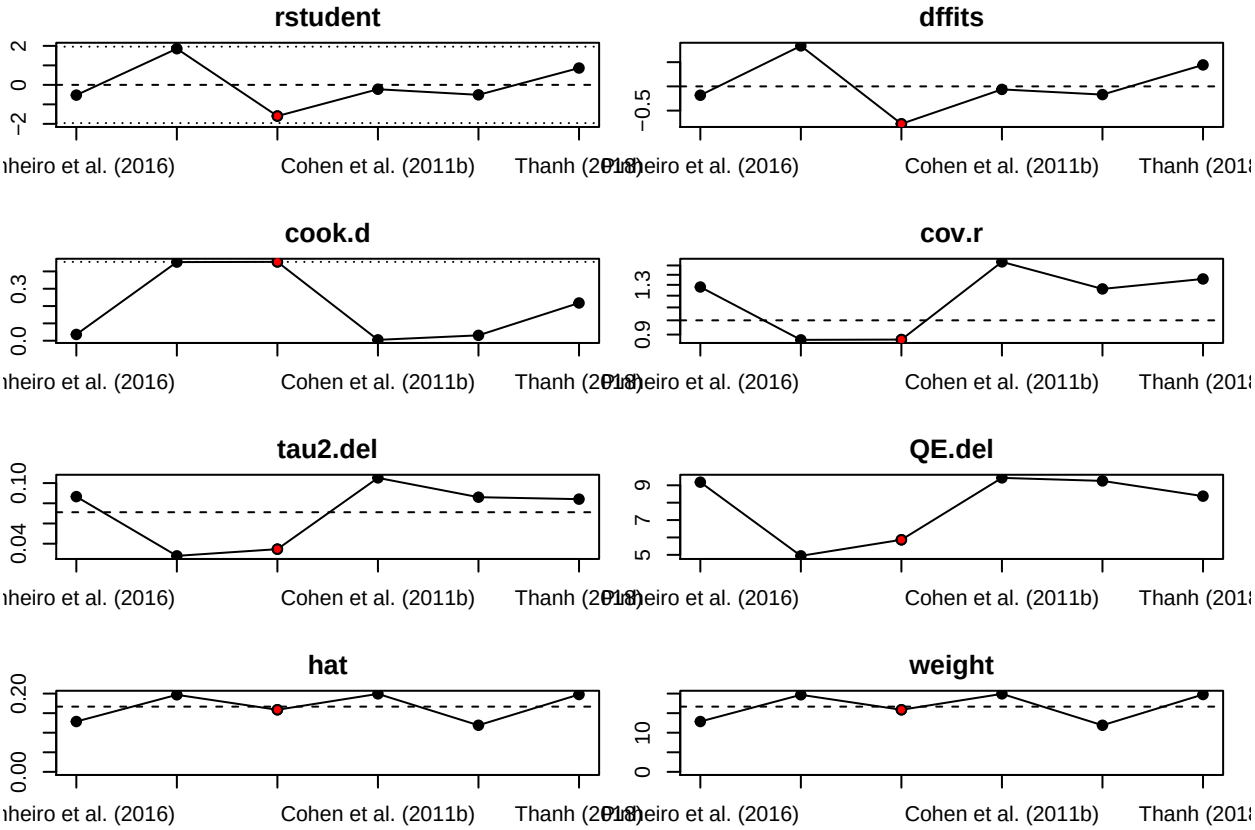
#FINDING INFLUENTIAL STUDIES

```r
# Influential points on the meta analysis without our studies
inf1<- influence(m2)
inf1$ids <- rownames(print(inf1))
```

```
##
##                       rstudent  dffits cook.d  cov.r tau2.del QE.del    hat
## Pinheiro et al. (2016)  -0.5262 -0.1826 0.0361 1.2793   0.0866 9.1799 0.1286
## Martínez et al. (2015)   1.8551  0.8370 0.4538 0.8656   0.0280 4.9417 0.1969
## Graux et al. (2011)     -1.6002 -0.7739 0.4550 0.8674   0.0346 5.8661 0.1586
## Cohen et al. (2011b)    -0.2243 -0.0623 0.0050 1.5403   0.1052 9.4274 0.1992
## Pinheiro et al. (2017)  -0.5115 -0.1697 0.0309 1.2612   0.0860 9.2530 0.1190
## Thanh (2018)             0.8629  0.4438 0.2178 1.3578   0.0841 8.3757 0.1976
##                        weight    dfbs inf
## Pinheiro et al. (2016) 12.8591 -0.1803
```

```
## Martínez et al. (2015) 19.6867  0.8137
## Graux et al. (2011)    15.8640 -0.7871   *
## Cohen et al. (2011b)   19.9247 -0.0632
## Pinheiro et al. (2017) 11.9019 -0.1672
## Thanh (2018)           19.7636  0.4464
```

```
plot(inf1)
```




```
# Influential points on the meta analysis with our studies
inf2<- influence(m22)
inf2$ids <- rownames(print(inf2))
```

```
##
##                       rstudent  dffits cook.d  cov.r tau2.del  QE.del    hat
## Pinheiro et al. (2016)  -0.9058 -0.2525 0.0641 1.0864   0.0432 13.2024 0.0726
## Martínez et al. (2015)   1.3088  0.4789 0.2079 1.0520   0.0337 11.7008 0.1230
## Graux et al. (2011)     -2.1651 -0.7627 0.4238 0.7611   0.0089  9.1448 0.0935
## Cohen et al. (2011b)    -0.7141 -0.2554 0.0696 1.2051   0.0487 13.2271 0.1249
## Pinheiro et al. (2017)  -0.8671 -0.2294 0.0531 1.0854   0.0439 13.3332 0.0663
## Thanh (2018)             0.5212  0.2142 0.0526 1.2774   0.0561 13.9247 0.1236
## A3P2 - Study 1          -0.0843 -0.0030 0.0000 1.2978   0.0591 14.3610 0.1158
## A3P2 - Study 2           0.8037  0.2643 0.0732 1.1596   0.0486 13.5016 0.0888
## A3P2 - Study 3          -0.0025  0.0221 0.0005 1.2327   0.0559 14.3948 0.0891
## A3P2 - Study 4           1.5324  0.4861 0.2095 0.9846   0.0297 11.2144 0.1022
##                        weight    dfbs inf
```

```
## Pinheiro et al. (2016)  7.2626 -0.2522
## Martínez et al. (2015) 12.2991  0.4753
## Graux et al. (2011)     9.3549 -0.7818
## Cohen et al. (2011b)   12.4941 -0.2568
## Pinheiro et al. (2017)  6.6327 -0.2289
## Thanh (2018)           12.3619  0.2165
## A3P2 - Study 1         11.5846 -0.0030
## A3P2 - Study 2          8.8784  0.2633
## A3P2 - Study 3          8.9103  0.0220
## A3P2 - Study 4         10.2215  0.4868
```

```
plot(inf2)
```