

# MetaAnalysis Assignment

Renée Claes and Maxime Keutgen

2021-04-05

## Introduction

In this notebook, we try to reproduce some of the results of the Meta-Analysis of Hornsey & al 2016. We start by loading libraries and datasets.

```
library(haven)
library(metafor)
library(tidyverse)
library(meta)
NEP <- read_sav("NEP.sav")
ObjKnow <- read_sav("Objective Knowledge.sav")
SubjKnow <- read_sav("Subjective Knowledge.sav")
FreeMark <- read_sav("Support for Free-market Ideology.sav")
TrustScientist <- read_sav("Trust in Scientists.sav")
```

## Objective knowledge

First, we will start with one dataset, objective knowledge. A bit of data cleaning is required (shortening name of one study).

```
ObjKnow[15,1] <- "Hine, D. W., Reser et al (2013)"
```

Also, we need an estimate of the variance of the effect size. Since these are correctional studies, the correlations coefficients have been transformed to Fisher Z scores<sup>1</sup> for inferential purposes. The correctional coefficient is given by :

$$r = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

Fisher transformation is given by :

$$z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

If  $X_1 \dots X_n$  and  $Y_1, \dots, Y_n$  are i.i.d paired data, then  $z$  is asymptotically normally distributed<sup>2</sup> :

$$z \sim \mathcal{N} \left( \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right), \frac{1}{N-3} \right)$$

We can then compute the variance of the Fisher Z-scores and the variance .

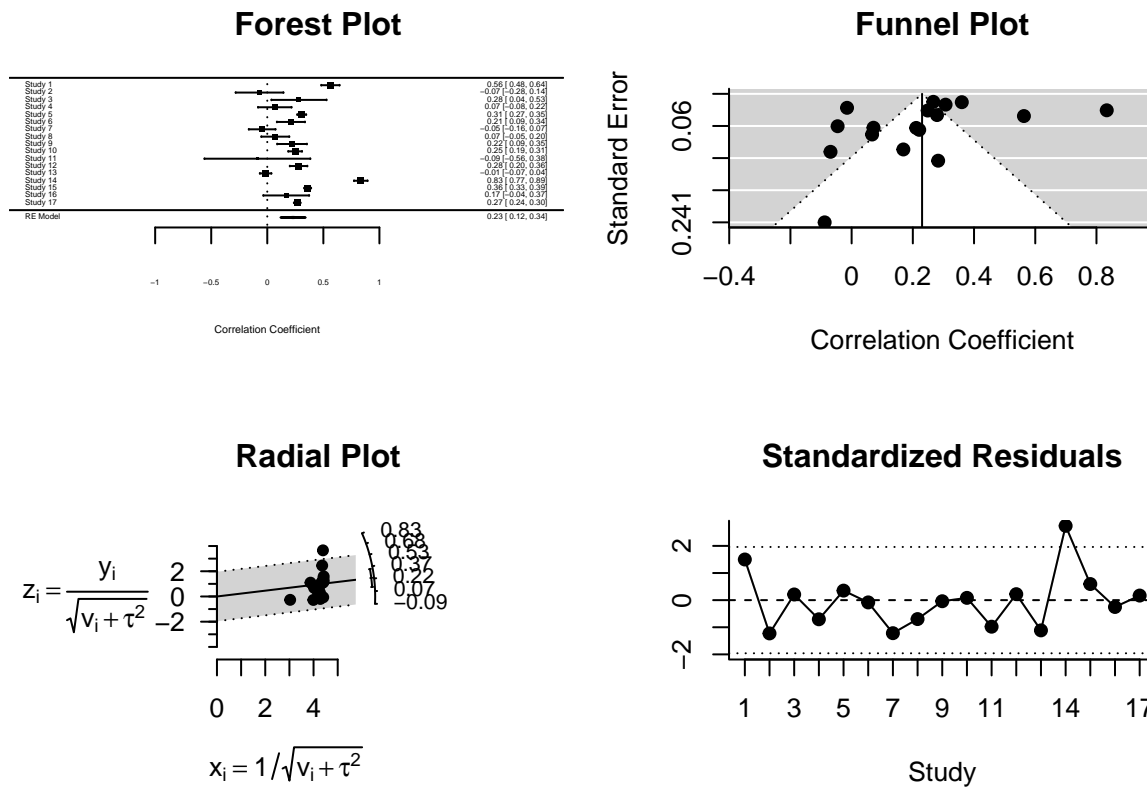
```
ObjKnow <- ObjKnow %>% mutate(Var=1/(Sample_size-3), Cor=tanh(FishersZ) )
```

We now have at our disposal point estimate of the effect size (Z scores) and of the variance of this effect size. We start by fitting a Random Effect model.

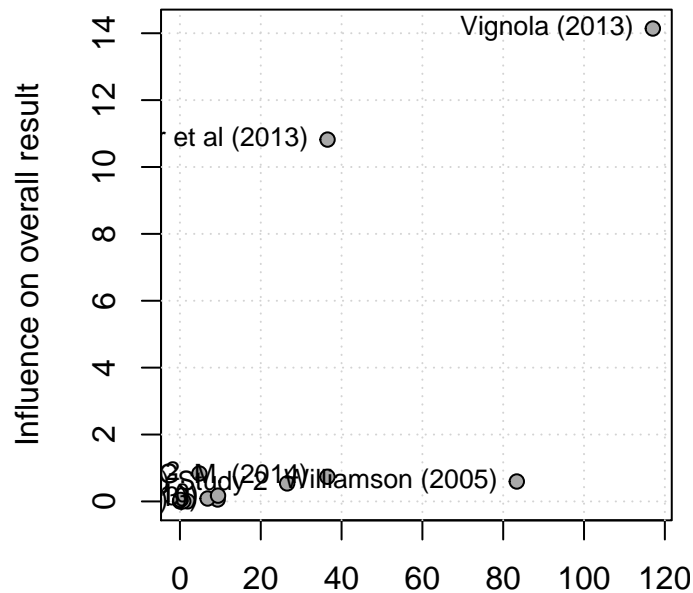
```
REM.metafor <- rma(vi = ObjKnow$var, method = "REML", ri = ObjKnow$Cor, measure="COR", ni=ObjKnow$Sample_size)
```

Random effect models with RMA function has an estimate of 0.2302, ci 0.11-0.34

```
plot.rma.uni(REM.metafor)
```



```
REM.meta <- metacor(cor = Cor,n=Sample_size,studlab = Study_name,comb.random = T,data = ObjKnow,method.
baujat.meta(REM.meta)
```



**Contribution to overall heterogeneity** Comment : The funnel plot and residual plot show that study 14 (Williamson 2015) should be investigated. The outlier map shows that Vignola has a high influence on the overall result but sample size is satisfactory (N=1462). We could not find the estimate

of correlation between belief in climate change and objective knowledge ( $r = 0.015$  according to the authors of the meta-analysis) <https://link.springer.com/article/10.1007%2Fs11027-012-9364-8>

```
REM.meta <- metacor(cor = Cor,n=Sample_size,studlab = Study_name,comb.random = T,data = ObjKnow,method.
```

Estimate : 0.2464, 95% CI :[0.1097; 0.3740]. Estimator of  $\tau^2$  is REML. Should we use different estimators ? The results agreed with article figure 2 from article. Test for heterogeneity is significant. Tau is an estimate of between study variance,  $I^2$  heterogeneity.

We can now also do funnel plots and forestplots. Should be nice to do them in ggplot.

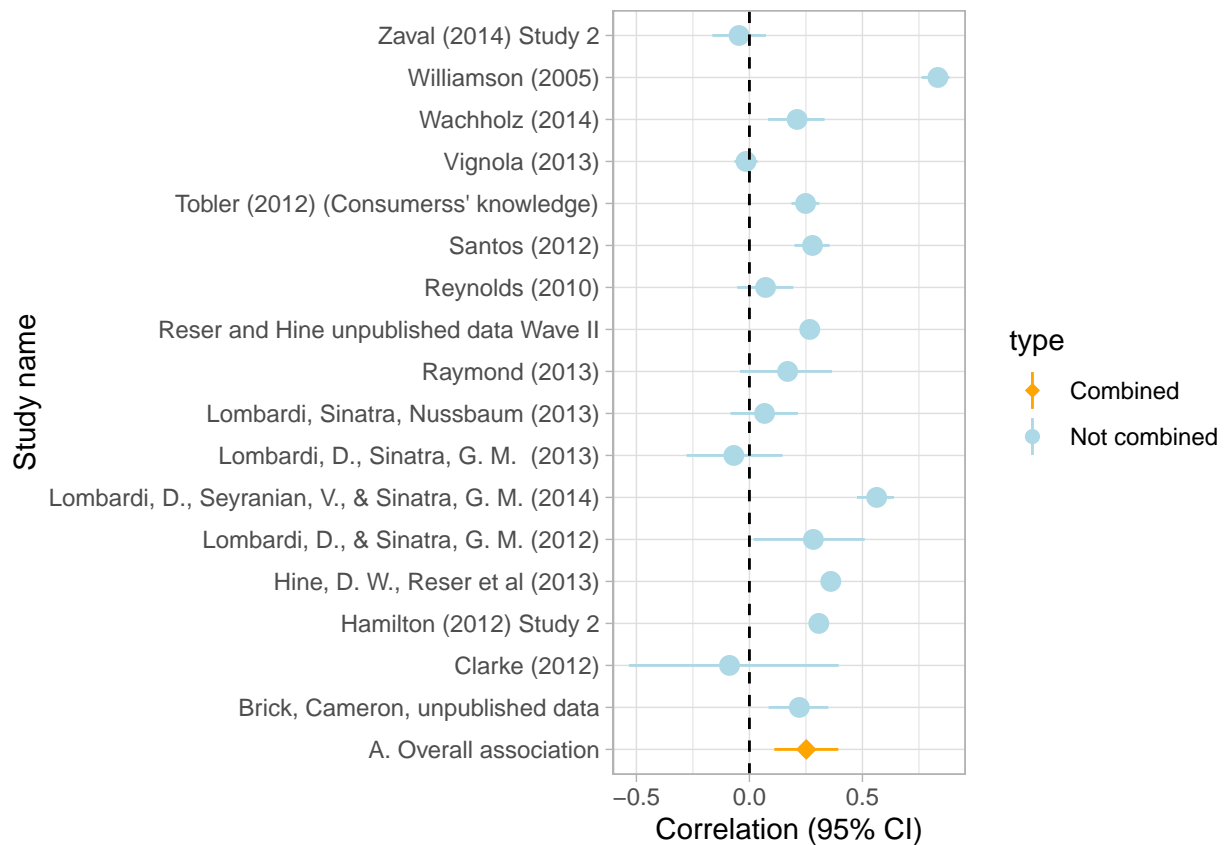
```
t <- tibble(ObjKnow$Study_name,tanh(REM.meta$lower),tanh(REM.meta$upper),
            REM.meta$cor,floor(ObjKnow$Sample_size))
names(t) <- c("name","lower","upper","estimate","sample.size")
t <- t %>% add_row(name="A. Overall association",lower=REM.meta$lower.random,upper=REM.meta$upper.random,
                  estimate=REM.meta$TE.random,sample.size=floor(sum(ObjKnow$Sample_size)))
t %>% filter(name=="Williamson (2005)")
```

```
## # A tibble: 1 x 5
##   name          lower upper estimate sample.size
##   <chr>          <dbl> <dbl>    <dbl>         <dbl>
## 1 Williamson (2005) 0.762 0.885    0.834           99
```

```
t$type <- ifelse(t$name=="A. Overall association","Combined","Not combined")
```

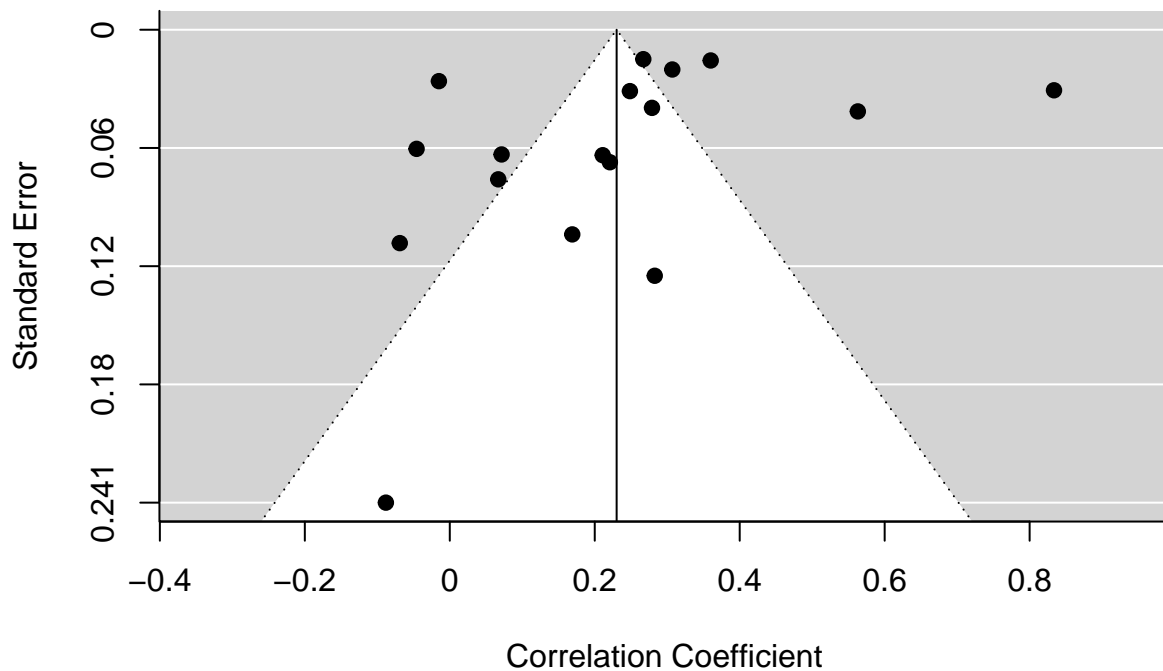
Forest plot

```
fp <- ggplot(data=t, aes(x=name, y=estimate, ymin=lower, ymax=upper,shape=type)) +
  geom_pointrange(aes(color=type),fatten = 6) +
  geom_hline(yintercept=0, lty=2) + # add a dotted line at x=1 after flip
  coord_flip() + # flip coordinates (puts labels on y axis)
  xlab("Study name") + ylab("Correlation (95% CI)") +
  scale_shape_manual(values=c(18,16)) + scale_color_manual(values = c("orange","lightblue"))+theme
print(fp)
```



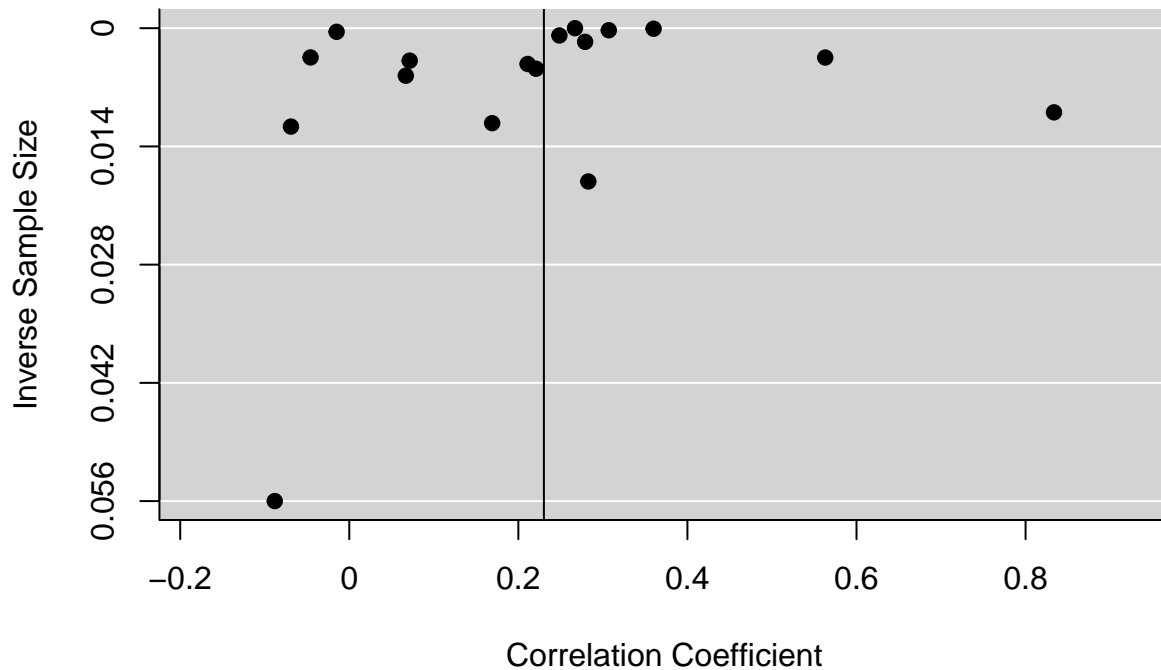
What about Funnel Plot ?

`funnel(REM.metafor)`



Commentary of the funnel plot : There seems to be considerable publication bias. In particular, the study of Williamson, with a reported correlation of 0.8 looks astonishing. One should investigate further.

```
funnel(REM.metafor,yaxis="ninv")
```



When reading article of Williamson et al, one see that there is probably *no publication bias*, the study has been well carried out, sample size is acceptable. However there seem to be a sampling bias (people studied were highly educated researchers). As authors of the study phrase it :

"In our study (with a sample consisting mainly of researchers and government employees) there was no indication that experts would be resistant to policy initiatives to prepare for, adapt to, or mitigate climate change. These experts recognized that climate change is caused by humans, they felt that some immediate action may be necessary, they recognized that climate change is having a significant impact on both forested ecosystems and forest-based communities (and that these impacts will continue in the future), and they felt that the impacts on ecosystems and communities tended toward being unacceptable. A recent study by Stedman et al. (2004) reinforces these findings<sup>3</sup>.

How sensible is the meta-analysis to study 14 ? Let's construct a hypothesis test. We know that the Fisher transform of correlations coefficients is normally distributed. We also know that the sum of normally distributed random variables is also normally distributed. We can thus construct a z-statistic :

$$z = \frac{\text{arctanh}(r'_1) - \text{arctanh}(r'_2)}{S} \sim N(0, 1)$$

With S a pooled estimate of the two estimates of variances of the Fisher transform of the regression estimates :

$$S = \sqrt{S_{r'_1}^2 + S_{r'_2}^2}$$

```
set.seed(1)
random.sel <- sample(17,16)
with.study14 <- ObjKnow[random.sel,]
rma.study14 <- metacor(cor = Cor,n=Sample_size,studlab = Study_name,comb.random = T,data = with.study14)
wo.study14 <- ObjKnow[-14,]
rma.wo.study14 <- metacor(cor = Cor,n=Sample_size,studlab = Study_name,comb.random = T,data = wo.study14)
```

We get an estimate of 0.2598 for a random selection of 17 studies of the meta-analysis (including study 14) and we get an estimate of 0.2015 when we exclude study 14. We construct the z statistic as follows :

$$z = \frac{\text{arctanh}(0.2598) - \text{arctanh}(0.2015)}{0.01727241} = 3.57$$

With S an estimator of variance of which we take an upper bound estimate, given by Cauchy-Schwartz inequality :

$$S \leq \sqrt{\frac{1}{13485-3}} + \sqrt{\frac{1}{13337-3}} = 0.01727241$$

13485 is sample size without study 14. 13337.23 is sample size without study 8 (random selection). Looks like we have evidence against null hypothesis. Study 14 is peculiar and significantly influence the correlation coefficient. We believe it should not have been included.

$$pvalue = 18E - 05$$

## Study of the moderators

In this section, we try to replicate the use of moderators.

```
REM.group.country <- metacor(cor = Cor,n=Sample_size,studlab = Study_name,comb.random = T,data = ObjKnow)
summary(REM.group.country)
```

```
## Number of studies combined: k = 17
##
##              COR              95%-CI      z  p-value
## Random effects model 0.2485 [0.1641; 0.3293] 5.64 < 0.0001
##
## Quantifying heterogeneity:
## tau^2 = 0.0285 [0.0241; 0.1409]; tau = 0.1688 [0.1553; 0.3753]
## I^2 = 95.2% [93.6%; 96.4%]; H = 4.57 [3.95; 5.30]
##
## Test of heterogeneity:
##      Q d.f.  p-value
## 334.59   16 < 0.0001
##
## Results for subgroups (random effects model):
##      k      COR      95%-CI  tau^2    tau    Q    I^2
## Country = USA      9  0.1912 [ 0.0563; 0.3192] 0.0381 0.1952 94.40 91.5%
## Country = Switzerland 1  0.2485 [ 0.1863; 0.3088] --    --    0.00  --
## Country = South Africa 2  0.1834 [-0.1397; 0.4712] 0.0360 0.1898  2.05 51.3%
## Country = Costa Rica 1 -0.0150 [-0.0662; 0.0363] --    --    0.00  --
## Country = Canda     1  0.8337 [ 0.7616; 0.8854] --    --    0.00  --
## Country = Australia 3  0.2947 [ 0.2062; 0.3784] 0.0051 0.0714 20.14 90.1%
##
## Test for subgroup differences (random effects model):
##      Q d.f.  p-value
## Between groups 165.59   5 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - DerSimonian-Laird estimator for tau^2
## - Jackson method for confidence interval of tau^2 and tau
## - Fisher's z transformation of correlations
```

## References

- [1] Fisher, R. A. (1915). “Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population”. *Biometrika*. 10 (4): 507–521. doi:10.2307/2331838. hdl:2440/15166. JSTOR 2331838
- [2] Fisher, R. A. (1921). “On the ‘probable error’ of a coefficient of correlation deduced from a small sample” (PDF). *Metron*. 1: 3–32.
- [3] Williamson, T., Parkins, J., & McFarlane, B. (2005). Perceptions of climate change risk to forest ecosystems and forest-based communities. *The Forestry Chronicle*, 81, 710-716