Practical meta-analysis in R Boerhaave course

# Introduction

Download from the Boerhaave website <http://www.boerhaave.nu/dta/> the 3 dataset, and save them on the desktop.

Install and load the libraries

# load libraries

install.packages(c("foreign", "metafor", "meta"))

library(foreign)

library(metafor)

library(meta)

# set your working directory using setwd

setwd("D:/Meta-analysis course LUMC")

**Exercise I**

We will use the STATA dataset diuretic.dta to perform a meta-analysis. This dataset contains 9 studies with data on the use of diuretics in pregnancy and the risk of pre-eclampsia.

Open this dataset in R by typing

diuretic = read.dta("diuretic.dta")

and take a look at the created dataframe by typing:

**View**(diuretic)

a. Use the print command to get information about the number of studies, the number of participants per trial and arm, and the number of patients with pre-eclampsia in each arm:

print(diuretic)

In the first analysis we perform the meta-analysis based on odds ratios from all studies. Therefore we calculate the log odds ratio and its standard error for each study.

b. First calculate the number of mothers without pre-eclampsia (abbreviated as h for healthy) in the treated and control arm. You can do this with the following comments

hc = with(diuretic,nc-pec)

ht = with(diuretic,nt-pet)

Note that the command with is used to indicate in which dataframe (in our example diuretic) you want to perform the calculations

View the new variables

c. Now calculate the odds ratio in each study.

d. Calculate the log odds ratio in each study

e. The standard error of the log odds ratio in each study can be obtained by

selogor = with(diuretic, sqrt(1/pet + 1/pec + 1/ht + 1/hc))

f. List the newly calculated measures. Which trials have the smallest standard errors? What do you expect from the weighing of the trials?

g. Now we will perform a standard meta-analysis of these 9 trials. Therefore we use the rma command of the package metafor. Type

res1 = rma(yi =logor, sei =selogor, data=diuretic, method="FE", measure="OR")

res1

What do you think, is this a fixed or random effects analysis? Why?

A forestplot can be obtained by

forest(res1)

h. This output was on the logarithmic scale. To obtain the effect estimate with 95% CI on the odds ratio scale type

predict(res1, transf=exp, digits=4)

j. To obtain a more sophisticated graph you may type:

forest(res1, xlim=c(-16, 6), at=log(c(0.05, 0.25, 1, 4)), order = "obs", atransf=exp, slab = diuretic$trialid,

ilab = cbind(diuretic$pet, ht, diuretic$pec, hc),

ilab.xpos=c(-9.5,-8,-6,-4.5), cex=0.75, ylim=c(-1,12),

refline = res1$b)

There are many options in R to improve the graph output such as

## Some "standard" modifications

# order: sort by "obs", "fit", "prec", etc

# slab: change study labels

# ilab: study information

# transf: apply function to effects

# psize: symbol sizes

h. Now change the fixed effect option into random, by choosing the method REML:

res2 = rma(yi =logor, sei =selogor, data=diuretic, method="REML", measure="OR")

res2

Look at the output. What do you observe? How large is the between study variance?

Performing a meta analysis is more flexible when data are in a 2x2 form each study. If such 2x2 data are available, a range of effect estimates can be obtained (odds ratios, risk ratios, risk differences). We therefore use the metabin command from the package meta.

The data must be supplied as follows:

i. number of events in the treatment group, ii. Number of no events in the treatment group, iii. Number of events in the control group, iv. Number of no events in the control group.

k. Type:

res3 = metabin(pet, ht, pec, hc, data = diuretic, studlab = trialid, sm="OR", method="MH", comb.random=TRUE)

res3

Compare the results to the results obtained earlier. Are they identical? Explain.

Note: the metabin command can also be used to perform the exact analyses for dichotomous outcomes.

**Exercise II.**

The streptok .dta dataset contains results from trials on the effect of streptokinase on mortality after myocardial infarction.

1. Import the dataset and explore the variables.
2. Generate for the treated and untreated group a variable which contains the number of persons without the outcome (h1 and h0).
3. Perform a meta-analysis using odds ratios as outcome. Use the command:

# Odds ratio as outcome

res1 = metabin(cases1, h1, cases0, h0, data = strep, studlab = trialnam, sm="OR", method="MH", comb.random=TRUE)

res1

Is this a fixed effect or a random effect meta analysis? Look at the results.

The forest plot can be obtained by

forest(res1, order = "obs", atransf=exp,

slab = strep$trialnam,

refline = res1$b)

What do you observe?

d. The subset option can be used to perform restrictions on the dataset. Now restrict the analysis to studies up to and including the year 1977

res2 = metabin(cases1, h1, cases0, h0, data = strep, subset = strep$year <= 1977, studlab = trialnam, sm="OR", method="Inverse", comb.random=TRUE)

res2

What do you see?

1. We see that the ISIS-2 trial has a very large influence on the results. Start by identifying the trial number of ISIS-2 by searching in the data-browser. Then perform a meta analysis excluding this trial (hint: use the subset command. Hint 2. Not equal is noted in R as != )
2. A nice command is metacum, which performs a cumulative meta-analysis. You first have to sort the dataset by year:

# Cumulative meta-analysis

res4 = metacum(res1, sortvar=year, pooled="random")

res4

forest(res4)

Can you think of reasons why this cumulative meta-analysis was performed?

1. If you have time left, you could ask for the help file:

help(metabin)

Look through all the options that are available and try out some of them.

**Exercise III**

The data set cholesterol.dta contains the data of 28 trials on the effect of cholesterol reduction on the risk of ischaemic heart disease (IHD) events. In this exercise we concentrate on describing the between studies variability in the effect and how the effect is possibly modified by the average cholesterol reduction achieved in each trial (the variable cholreduc), type of intervention (the variable intervention: drug or diet) or type of population (variable ihdentry: with IHD, without known IHD, with or without IHD).

Use the commands view and summary, respectively, to get informed on the data set.

1. The log odds ratio and its standard error are already calculated for us. Use them to first fit a fixed effect meta-analysis

res1 = rma(yi=logor, vi=varlogor, data=chol , method="FE", measure="OR")

res1

An estimate of the pooled odds ratio and the forest plot with weights can be obtained by

predict(res1, transf=exp, digits=4)

forest(res1, atransf=exp, slab = chol$trialname, showweights=TRUE)

Look at the forest plot. Do you think that there is between studies heterogeneity?  
What is the result of the formal test on heterogeneity?

1. Now perform a random effects analysis.

Compare the weights of the different trials with the fixed effect analysis. For instance, what has happened with the weight of the largest trial?

Also compare the overall effect, confidence interval and significance.

1. To describe the heterogeneity between trials R gives two quantitative measures.   
   The first one is I-squared. What is its value here? What is its interpretation?   
   The second one is tau-squared. What is its interpretation? Calculate exp(1.96\*tau) and use it to make a prediction interval. What is the interpretation of it?
2. With the command predict(res2, transf=exp, digits=4) , you will get both a confidence interval and a prediction interval (credible interval). The prediction interval is somewhat larger than you have calculated by yourself. Why?
3. In the rest of the exercise we try to explain heterogeneity by trial characteristics.  
   First we let the effect depends on the patient population (variable ihdentry: with IHD, without known IHD, with or without IHD). This can be done by adding the mods option, in which you specify the variables for the regression analysis

res3 = rma(y=logor, vi=varlogor, mods = ~ factor(ihdentry) - 1, data=chol)

res3

1. To investigate whether the effect depends on a continuous trial characteristic, we perform a second meta-regression. We look at the average achieved cholesterol level reduction (the variable cholreduc).  
   res4 = rma(y=logor, vi=varlogor, mods = ~ cholreduc , data=chol)

res4

Also calculate the effect on the OR scale.

Study the output. What is increase in treatment effect per unit achieved cholesterol reduction? (estimate, 95% CI, P-value) What is your conclusion?

1. Compare model res2 (the model without cholrec as covariate) to model res4.

What percentage of the between studies variance is explained by differences in cholesterol reduction?

1. To investigate whether the effect is different for different types interventions (drug or diet), perform a meta-regression on the variable intervention. (Leave out the one trial with surgical intervention). What is your conclusion?
2. To see an example of a multivariable meta-regression, perform a regression on both the cholesterol reduction and the type of intervention.
3. Finally we look at the funnel plot. Ask R to draw a funnel plot for the fitted ranom effect model:

funnel(res2, main="Standard Error")

Judge the plot. What do you think? What does it mean when a point is outside the dotted lines? You can also specify the option refline =0.

funnel(res2, main="Standard Error", refline=0)

What does it mean now when a point is outside the dotted lines?

1. By specifying the option egger you get a trend line reflecting the association between effect and standard error.

metafunnel logor selogor, nullforce egger

1. You can get Egger’s test by:

regtest(res2)