Practical meta-analysis in STATA Boerhaave course

# Introduction

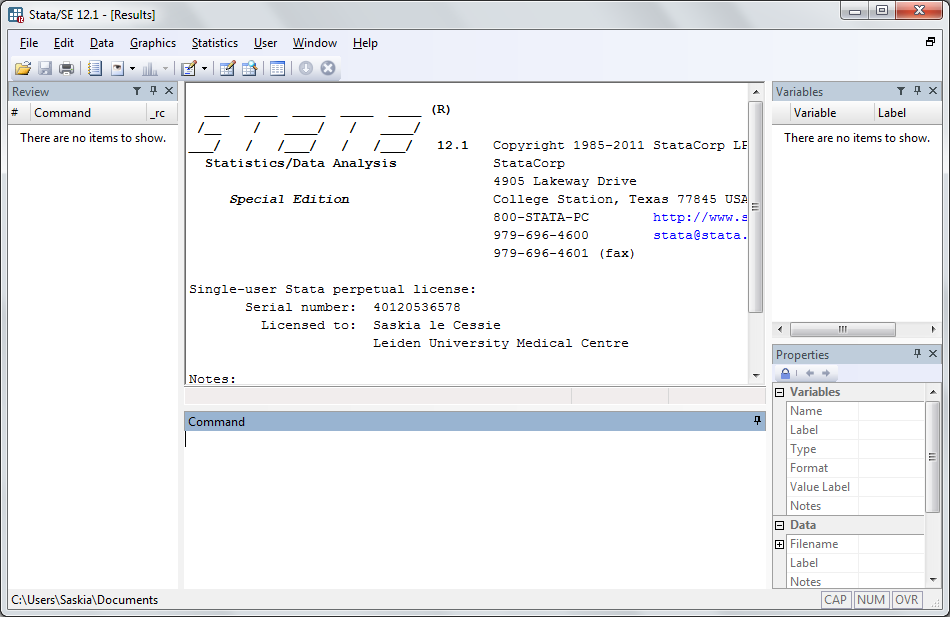
Download from the Boerhaave website <http://www.boerhaave.nu/dta/> the 3 dataset, and save them on the desktop.

In the exercises we will use Stata. Working with Stata is comparable to SPSS. After opening Stata you will see the following screen:

Variables and labels

Edit data

Browse data



The screen consists of 5 sub-screens. The output of all analyses appears in the middle top screen. In the Command screen, you can type commands.

Some user written STATA software needs to be installed. Type the following commands in the command panel on your screen:

net from <http://www.stata-press.com/data/mais>

net install mais

After installing this file, type

spinst\_mais

to obtain additional user written commands (mais=meta-analysis in stata).

**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 STATA. Look in the right upper panel on your screen to see which variables are in the dataset. You can also see the data by clicking on the Data Editor (Browse) icon at the top of the screen.

a. Use the list 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:

list trialid trial nt nc pet pec

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

generate hc=nc-pec

generate ht=nt-pet

Make a new listing

c. Now use generate to 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

gen selogor=sqrt(1/pet+1/pec+1/ht+1/hc)

(The command gen is the abbreviation of generate. STATA will recognize unambiguous abbreviations)

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. The metan command is a basic STATA command for fixed and random effect meta-analysis. Type

metan logor selogor

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

h. This output was on the logarithmic scale. To obtain output on the odds ratio scale type

metan logor selogor, eform

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

metan logor selogor, eform lcols(trialid) effect(odds ratio) xlab (.5,1, 2) xtitle(odds ratio) favours(favours diuretic # favours control)

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

Xlabel (.5, 1, 2, 4) defines x-axis. Any number of points can be defined

xtick (#, ….) adds tick marks to the x-axis

textsize (#) specifies the font size for text in the graph

effect (odds ratio) labels the effect size of the summary statistic

summaryonly shows only summary estimates in the graph

lcols (varlist) defines columns of additional data

h. Now add the option random:

metan logor selogor, eform random

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

j. To get obtain both analyses within one output, type:

metan logor selogor, eform second(random)

The metan command 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). The default is the Mantel-Haenszel risk ratio.

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. When the metan command is performed on 2x2 data, the eform option is no longer needed. Type:

metan pet ht pec hc

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

l. Now perform the analysis but with the odds ratio as outcome

metan pet ht pec hc, or

metan pet ht pec hc, or fixedi lcols(trialid)

What is the difference between these two analyses?

m. Look at the list of variables at the right part of the screen. List these variables and try to identify what they represent

**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 with persons without the outcome (h1 and h0).
3. Perform a meta-analysis using odds ratios as outcome. Use the command:

metan cases1 h1 cases0 h0, or lcols(trialnam)

Is this a fixed effect or a random effect meta analysis? Look at the results and the forrest plot. What do you observe?

d. The if command in STATA can be used to perform restrictions on the dataset. Now restrict the analysis to studies up to and including the year 1977

metan cases1 h1 cases0 h0 if year <=1977, or lcols(trialnam)

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 if command. Hint 2. Not equal is noted in Stata as != )
2. A nice command is metacum, which performs a cumulative meta-analysis. You first have to sort the dataset by year:

sort year

metacum cases1 h1 cases0 h0, or lcols(trialnam year) xlab(.1, 1, 10) xtick (.5, 1, 2)

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(metan)

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).

Type in the commands describe and notes, respectively, to get informed on the data set. Have a look at the data set using the ‘data-editor (browse)’ button.

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

metan logor selogor, fixed eform xlab(.1,.2, .5, 1, 2, 5,10) force nobox effect(odds ratio)

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 STATA 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. Now we ask STATA to calculate a prediction interval for us and to show it in the Forest plot. (The option to use is rfdist)

metan logor selogor, random eform xlab(.1,.2, .5, 1, 2, 5,10) force nobox effect(odds ratio) rfdist

The prediction interval is somewhat larger than you have calculated by yourself. Why? (Two reasons)

1. In the rest of the exercise we try to explain heterogeneity by trial characteristics.  
   First look at whether the effect depends on the patient population (variable ihdentry: with IHD, without known IHD, with or without IHD). This can be done by the option by(ihdentry)
2. To investigate whether the effect depends on a continuous trial characteristic, we perform a meta-regression. As an example we look at the average achieved cholesterol level reduction (the variable cholreduc).

metareg logor cholreduc, wsse(selogor) graph

Also repeat this command with the extra option eform.

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

1. To investigate how much of the between trials heterogeneity is explained by achieved cholesterol reduction, fit a regression leaving out the explanatory variable cholreduc:

metareg logor, wsse(selogor)

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)

metareg logor intervention if intervention < 3, wsse(selogor) graph eform

What is your conclusion?

1. To see an example of a multivariable meta-regression, perform a regression on both the cholesterol reduction and the type of intervention.
2. Finally we look at the funnel plot. Ask STATA to draw a funnel plot:

metafunnel logor selogor

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 forcenull.

metafunnel logor selogor, forcenull

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:

metabias logor selogor, egger