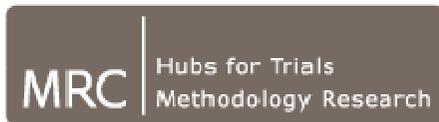
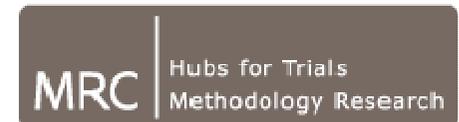


Session 12

Meta-analysis of ordinal data



Edinburgh Hub



North West Hub

Session 12

Meta-analysis of ordinal data

12.1 Meta-analysis in general

12.2 Meta-analysis of data from
proportional odds models

12.3 Practical difficulties

12.1 Meta-analysis in general

What is meta-analysis?

- A way to calculate an average
- Estimates an 'average' or 'common' effect
- Improves the precision of an estimate by using all available data

When can we do a meta-analysis?

- When more than one study has estimated an effect
- When there are no differences in the study characteristics that are likely to substantially affect outcome
- When the outcome has been measured in similar ways
- When the data are available (take care with interpretation when only some data are available)

Averaging studies

- Starting with the summary statistic for each study, how should we combine these?
- A simple average gives each study equal weight
- This seems intuitively wrong
- Some studies are more likely to give an answer closer to the 'true' effect than others

Weighting studies

- More weight to the studies which give us more information
 - More participants
 - More events
 - Lower variance
- Weight is closely related to the width of the study confidence interval: wider confidence interval = less weight

For example – Abciximab versus control in acute presumed ischaemic stroke [Sandercock 2008]

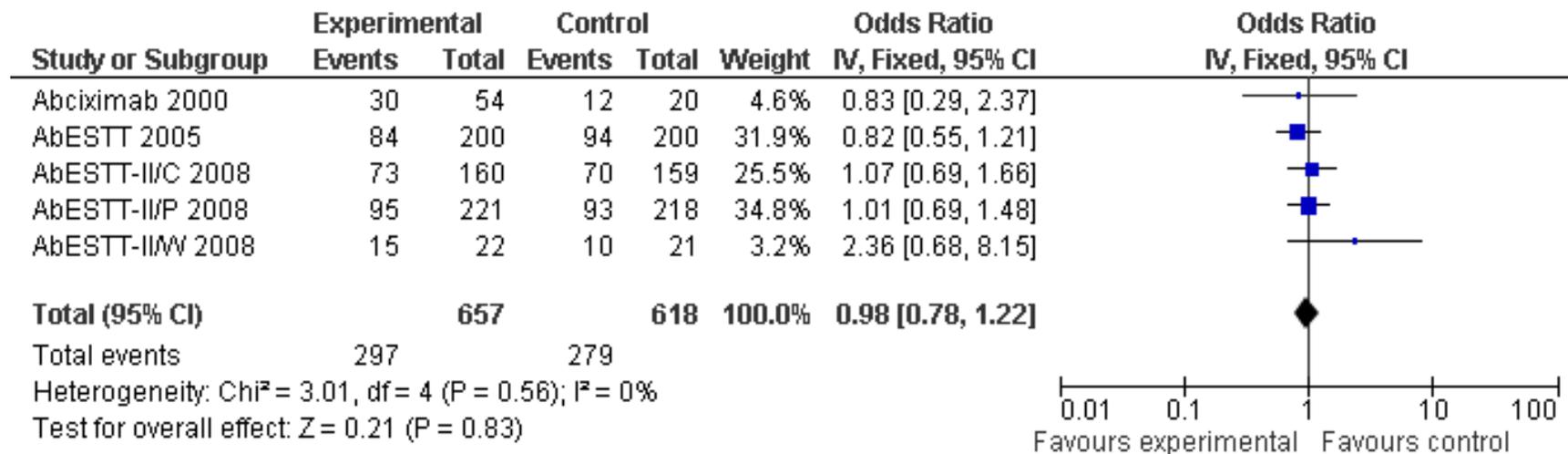
| | Dead or dependent Abciximab | Dead or dependent Control | Weight (%) |
|---------------------|--------------------------------|------------------------------|------------|
| Abciximab 2000 | 30/54 | 12/20 | 4.6 |
| AbESTT 2005 | 84/200 | 94/200 | 31.9 |
| AbESTT-II/C 2008 | 73/160 | 70/159 | 25.5 |
| AbESTT-II/P 2008 | 95/221 | 93/218 | 34.8 |
| AbESTT-II/W 2008 | 15/22 | 10/21 | 3.2 |

Displaying results graphically

- Forest plots

Abciximab vs control in acute presumed ischaemic stroke.

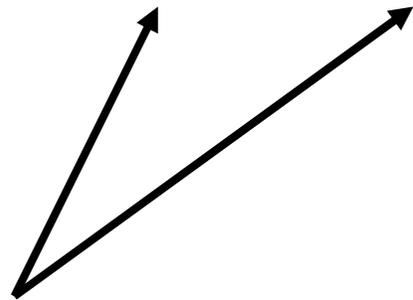
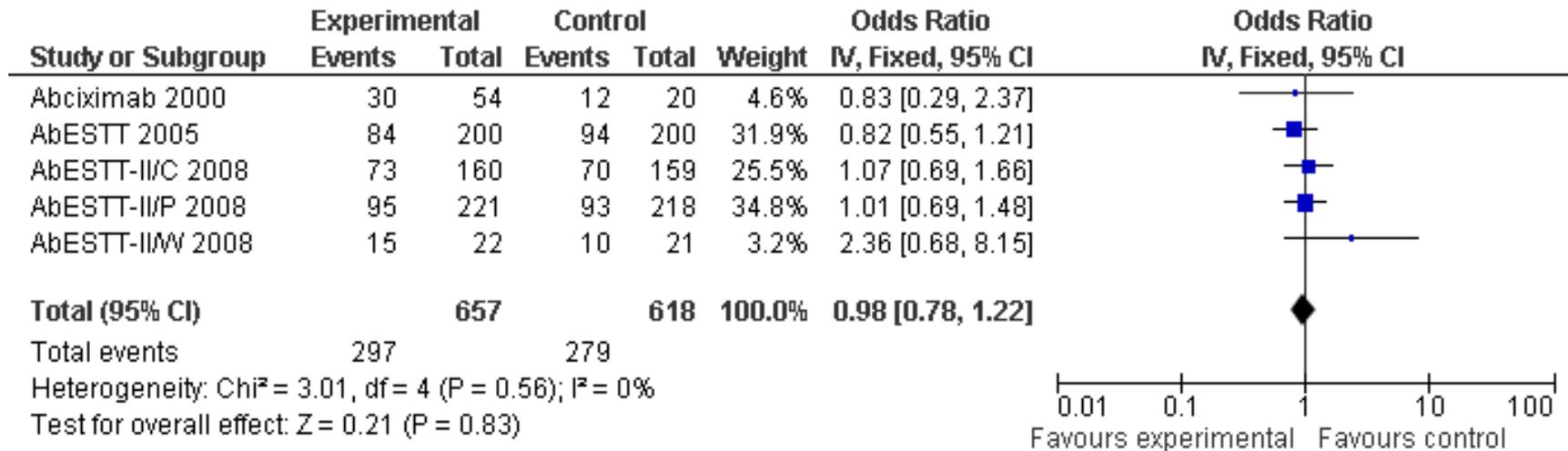
Death or Dependence at the end of follow up



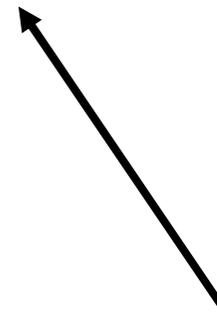
NB Here, Odds ratio <1 favours Abciximab

Could we just add the data from all the trials together?

- One approach to combining trials would be to add all the Abciximab groups together, add all the control groups together, and compare the totals
- This is wrong for several reasons, and it can give the wrong answer



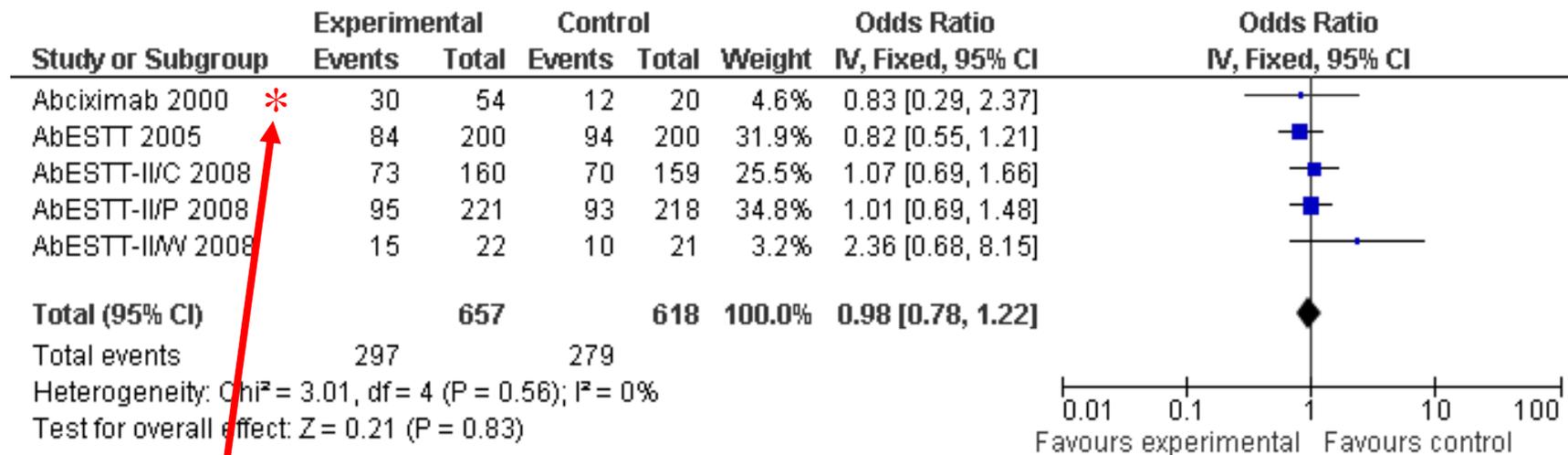
If we add up the columns we get an Odds ratio of $(297 \cdot 339) / (360 \cdot 279) = 1.002$, a higher chance of bad outcome in the experimental group



From a meta-analysis, we get Odds ratio = 0.98, a lower chance of bad outcome in the experimental group

Problems with simple addition of studies

- breaks the power of randomisation
- imbalances within trials introduce bias



The Abciximab 2000 trial contributes 8% (54/657) of all the data to the experimental column, but 3% (20/618) to the control column. Therefore it contributes more information to the average chance of death or dependency in the experimental column than it does to the control column. There is a high chance of death or dependency in this trial, so the chance of death or dependency for the experimental column is higher than the control column.

Heterogeneity

What is heterogeneity?

- Heterogeneity is variation between the studies' results

Causes of heterogeneity

Differences between studies with respect to:

- ***Patients***: diagnosis, inclusion and exclusion criteria, etc.
- ***Interventions***: type, dose, duration, etc.
- ***Outcomes***: type, scale, cut-off points, duration of follow-up, etc.
- ***Quality and methodology***: randomised or not, allocation concealment, blinding, etc.

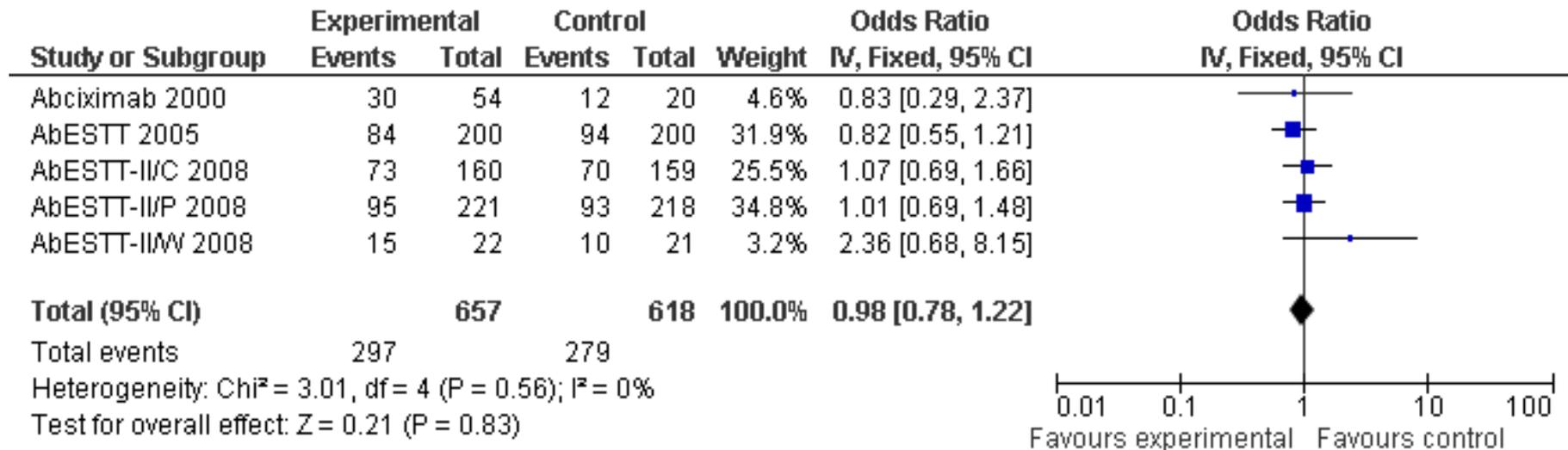
Bias versus precision

- Some features of studies might change the magnitude of the treatment effect (e.g. lack of blinding of outcome)
- Some will just affect precision (e.g. single set of high quality weighing scales vs patients' own bathroom scales).
- Bias is arguably much more important than precision.

Interpretation - Quality

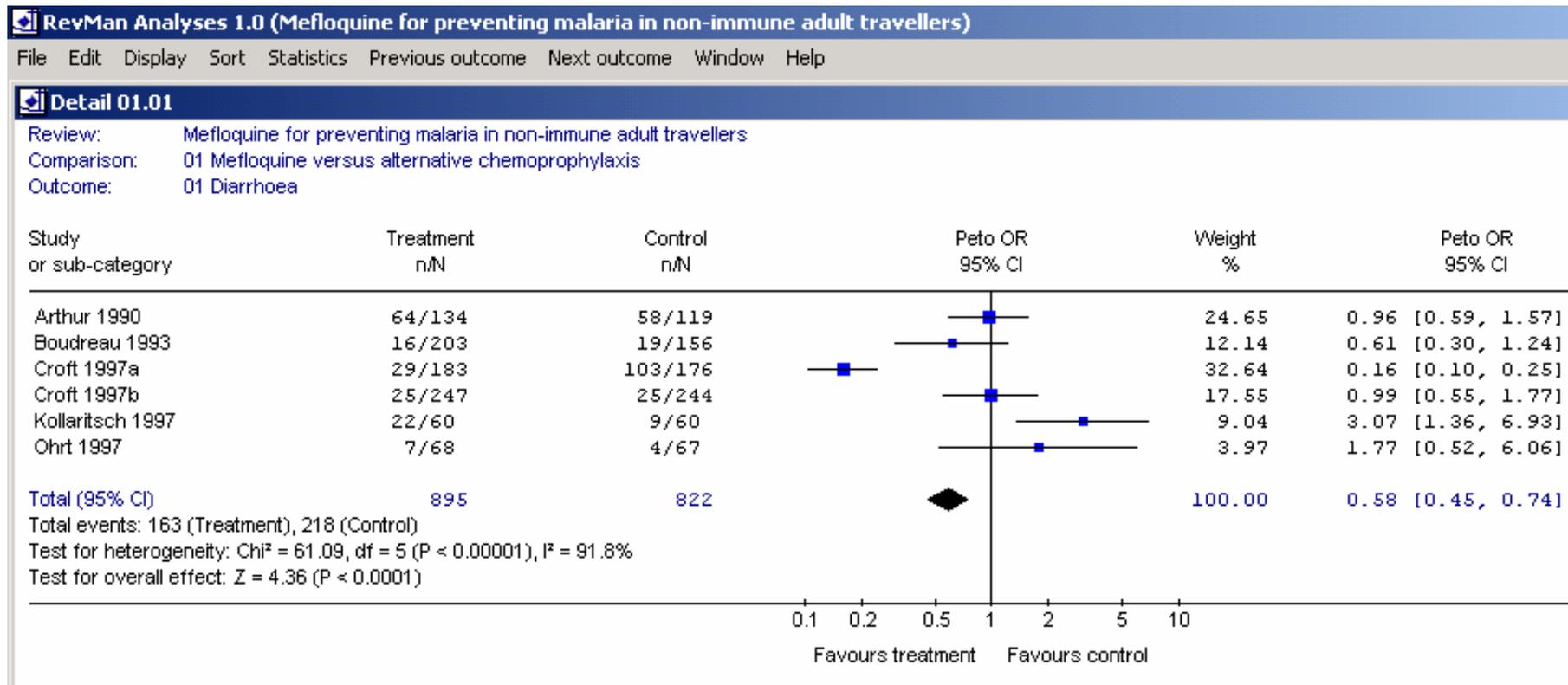
- Rubbish studies = unbelievable results
- If all the trials in a meta-analysis were of very low quality, then you should be less certain of your conclusions.
- Instead of “Treatment X cures depression”, try “There is some evidence that Treatment X cures depression, but the data should be interpreted with caution.”

How to assess heterogeneity



Does it look like the studies agree with each other?

How to assess heterogeneity



Statistical measures of heterogeneity

- The Chi² test measures the amount of variation in a set of trials, and tells us if it is more than would be expected by chance
- Small p values suggest that heterogeneity is present
- This test is not very good at detecting heterogeneity. Often a cut-off of $p < 0.10$ is used, but lack of statistical significance does not mean there is no heterogeneity
- The more studies you have, the more likely the test is to be significant.

Statistical measures of heterogeneity (2)

- I^2 is the proportion of variation that is due to heterogeneity rather than chance
- Large values of I^2 suggest heterogeneity
- For more info see: Higgins 2003

Fixed vs random effects

- See Borenstein 2010.

Fixed effect

Philosophy behind ***fixed effect model***:

- there is one real value for the treatment effect
- all trials estimate this one value

Problems with ignoring heterogeneity:

- confidence intervals too narrow

Random effects

Philosophy behind *random effects model*:

- there are many possible real values for the treatment effect (depending on dose, duration, etc.)
- each trial estimates its own real value
- This is generally more plausible than a fixed effect, but if you have too few studies, you can't estimate the between study variation

12.2 Meta-analysis of data from proportional odds models

For each trial, data of the form...

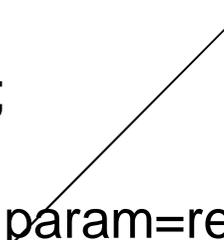
| | Rankin | | | |
|-----------|----------|----------|----------|----------|
| | 0 to 1 | 2 | 3 to 5 | 6 |
| Control | <i>n</i> | <i>n</i> | <i>n</i> | <i>n</i> |
| Abciximab | <i>n</i> | <i>n</i> | <i>n</i> | <i>n</i> |

4 of 5 trials had full data, and 1 of 5 had data in this form. All trials converted to this form.

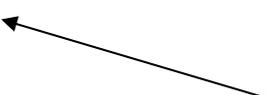
SAS code

```
data a;  
input trial $ treatment ordscale n;  
cards;  
Abciximab 0 1 4  
etc.  
;  
proc sort;  
  by trial;  
proc logistic order=internal;  
  weight n;  
  class treatment (ref='0') / param=ref order=internal;  
  model ordscale (descending) = treatment;  
  by trial;  
run;
```

Means Odds ratio <1
favours Abciximab



Analyse each trial
separately



SAS output (from each trial)

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Wald | | |
|-------------|----|----------|---------------|------------|------------|
| | | | Error | Chi-Square | Pr > ChiSq |
| Intercept 4 | 1 | -1.4652 | 0.4628 | 10.0230 | 0.0015 |
| Intercept 3 | 1 | 0.4497 | 0.4238 | 1.1258 | 0.2887 |
| Intercept 2 | 1 | 0.9766 | 0.4359 | 5.0204 | 0.0251 |
| treatment 1 | 1 | -0.2530 | 0.4796 | 0.2782 | 0.5979 |

Using Revman

1. Download Revman:
<http://ims.cochrane.org/revman/download>
2. Open Revman and create a new review
3. Choose intervention review
4. Insert title “Antiplatelet therapy for acute ischaemic stroke”
5. Choose full review
6. Go to Studies and references, included studies. Add study (just type acronym, and finish). Keep adding studies until all are entered.

Using Revman (2)

7. Go to data and analyses. Add comparison “Abciximab vs control”, and add outcome under comparison.
8. Choose Generic Inverse Variance, name it “Modified Rankin at end of follow up”, and add study data for the new outcome.
9. Shift and cursor down to select all studies.
10. Cut and paste data out of Excel sheet (or enter individual values)

ow Help

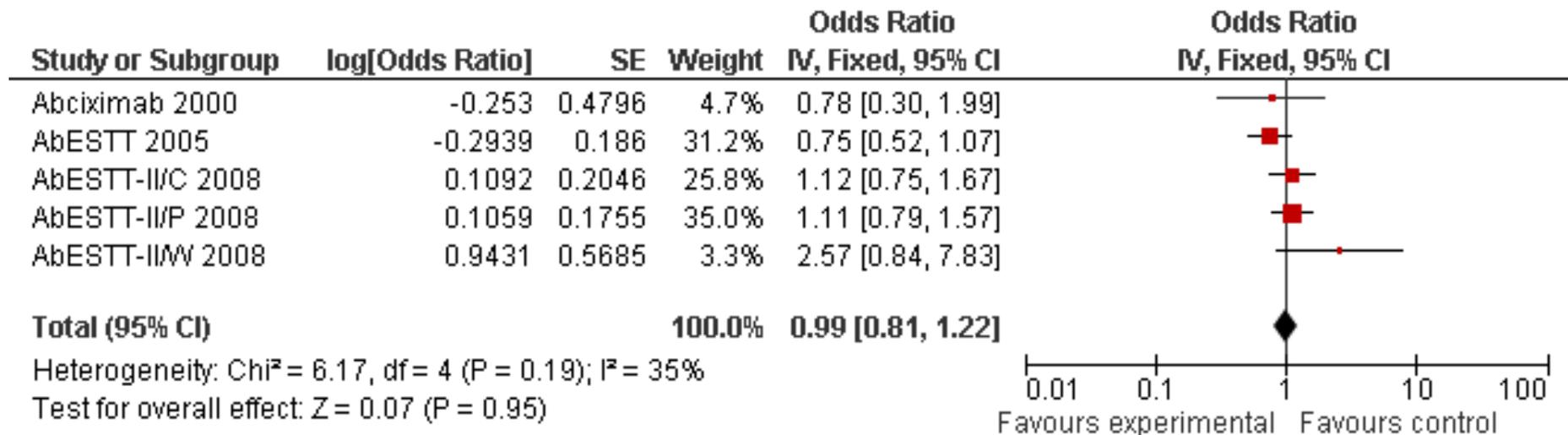
OK ABC

1.2 Modified Rankin...

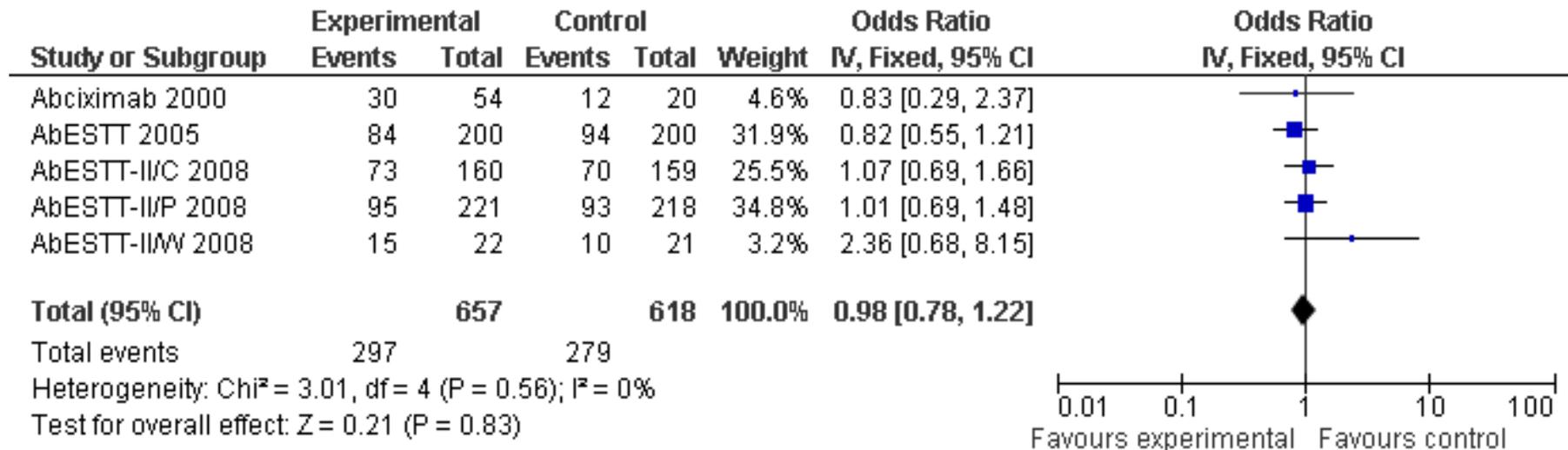
Comparison: 1 Antiplatelet versus control in acute presumed ischaemic stroke, Outcome: 1.2 Modified Rankin at the end of follow up

| Study or Subgroup | log[] | SE | Weight | IV, Fixed, 95% CI |
|---|---------|--------|---------------|--------------------------|
| <input checked="" type="checkbox"/> Abciximab 2000 | -0.253 | 0.4796 | 4.7% | 0.78 [0.30, 1.99] |
| <input checked="" type="checkbox"/> AbESTT 2005 | -0.2939 | 0.186 | 31.2% | 0.75 [0.52, 1.07] |
| <input checked="" type="checkbox"/> AbESTT-II/C 2008 | 0.1092 | 0.2046 | 25.8% | 1.12 [0.75, 1.67] |
| <input checked="" type="checkbox"/> AbESTT-II/P 2008 | 0.1059 | 0.1755 | 35.0% | 1.11 [0.79, 1.57] |
| <input checked="" type="checkbox"/> AbESTT-II/W 2008 | 0.9431 | 0.5685 | 3.3% | 2.57 [0.84, 7.83] |
| Total (95% CI) | | | 100.0% | 0.99 [0.81, 1.22] |
| Heterogeneity: Chi ² = 6.17, df = 4 (P = 0.19); I ² = 35% | | | | |
| Test for overall effect: Z = 0.07 (P = 0.95) | | | | |

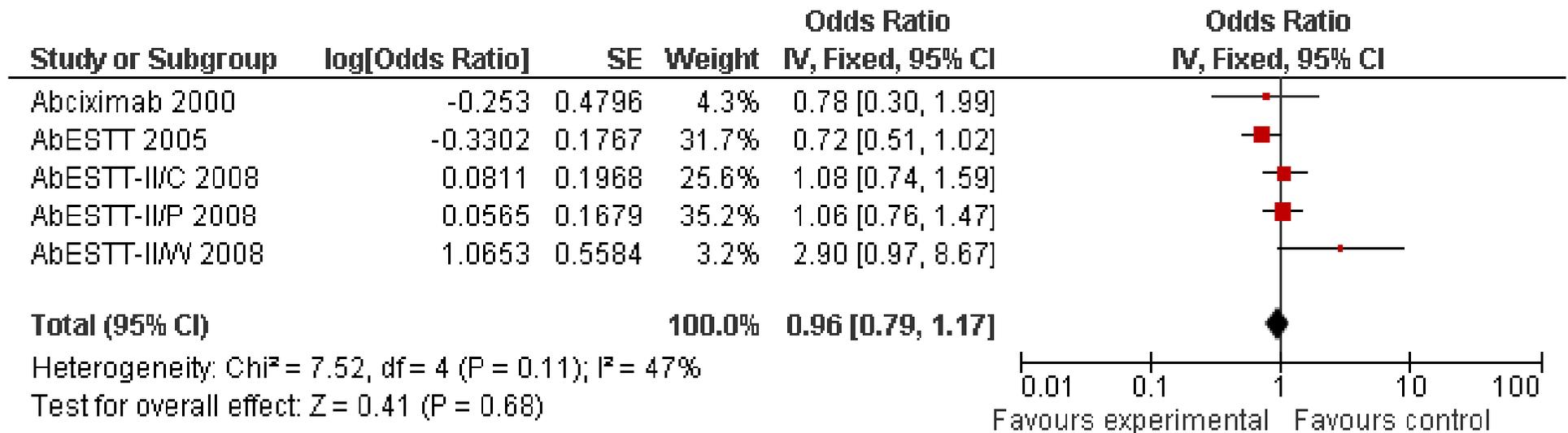
Meta-analysis of ordinal outcomes



Meta-analysis of binary outcomes



Meta-analysis of full available data



Abciximab had fewer categories than other trials – weight decreases

Meta-analysis in SAS

- It's possible, but tricky – Anne Whitehead has some code

Meta-analysis in R (1)

```
library(RODBC)
```

```
channel <-
```

```
  odbcConnectExcel("C:/abcixi.xls")
```

```
abcixi <- sqlFetch(channel, "Sheet2")
```

```
odbcClose(channel)
```

Excel Sheet →

| trial | est | se |
|-----------|---------|--------|
| Abciximab | -0.253 | 0.4796 |
| AbESTT | -0.2939 | 0.186 |
| AbESTT2C | 0.1092 | 0.2046 |
| AbESTT2P | 0.1059 | 0.1755 |
| AbESTT2W | 0.9431 | 0.5685 |

Meta-analysis in R (2)

```
library(grid)
```

```
library(rmeta)
```

```
abmeta <- meta.summaries(abcixi$est,  
  abcixi$se, names=abcixi$trial,  
  method="fixed", logscale=TRUE)
```

↑
Or "random"

Meta-analysis in R (3)

```
lowCI <- exp(abmeta$effects - (1.96 * abmeta$stderrs))
upCI <- exp(abmeta$effects + (1.96 * abmeta$stderrs))
lowCIsumm <- exp(abmeta$summary - (1.96 * abmeta$se.summary))
upCIsumm <- exp(abmeta$summary + (1.96 * abmeta$se.summary))

tabletext<-cbind(c("", "Study", abmeta$names, NA, "Summary"),
  c("Odds", "Ratio", round(exp(abmeta$effects), digits=2), NA,
    round(exp(abmeta$summary), digits=2)),
  c("Lower 95%", "CI Limit", round(lowCI, digits=2), NA,
    round(lowCIsumm, digits=2)),
  c("Upper 95%", "CI Limit", round(upCI, digits=2), NA,
    round(upCIsumm, digits=2)))
```

Meta-analysis in R (4)

```
m<- c(NA,NA,abmeta$effects,NA,abmeta$summary)
l<- m-c(NA,NA,abmeta$stderrs,NA,abmeta$se.summary)*1.96
u<- m+c(NA,NA,abmeta$stderrs,NA,abmeta$se.summary)*1.96
```

```
xt <- c(0.1, 1, 10)
```

```
forestplot(tabletext,m,l,u,zero=0,
  is.summary=c(TRUE,TRUE,rep(FALSE,6),TRUE),
  clip=c(log(0.01),log(100)), xlog=TRUE,
  col=meta.colors(box="royalblue",line="darkblue",
  summary="royalblue"), xticks=xt)
```

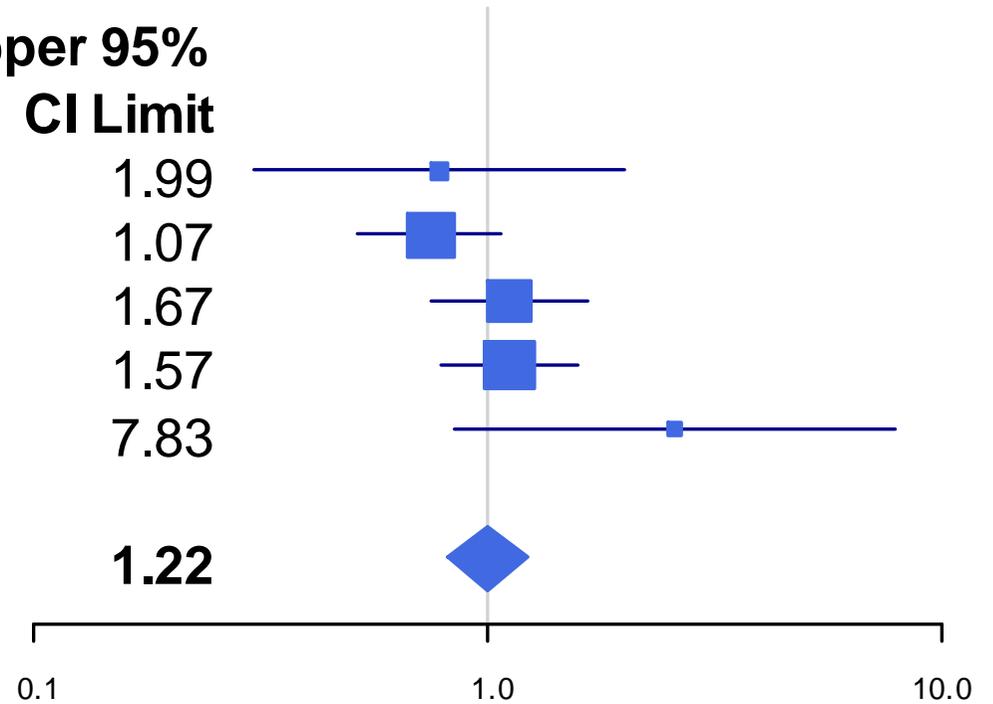
Number of
studies +1

Log scale

Clips long CIs

Manual tick
marks

| Study | Odds Ratio | Lower 95% CI Limit | Upper 95% CI Limit |
|----------------|-------------------|---------------------------|---------------------------|
| Abciximab | 0.78 | 0.3 | 1.99 |
| AbESTT | 0.75 | 0.52 | 1.07 |
| AbESTT2C | 1.12 | 0.75 | 1.67 |
| AbESTT2P | 1.11 | 0.79 | 1.57 |
| AbESTT2W | 2.57 | 0.84 | 7.83 |
| Summary | 0.99 | 0.81 | 1.22 |



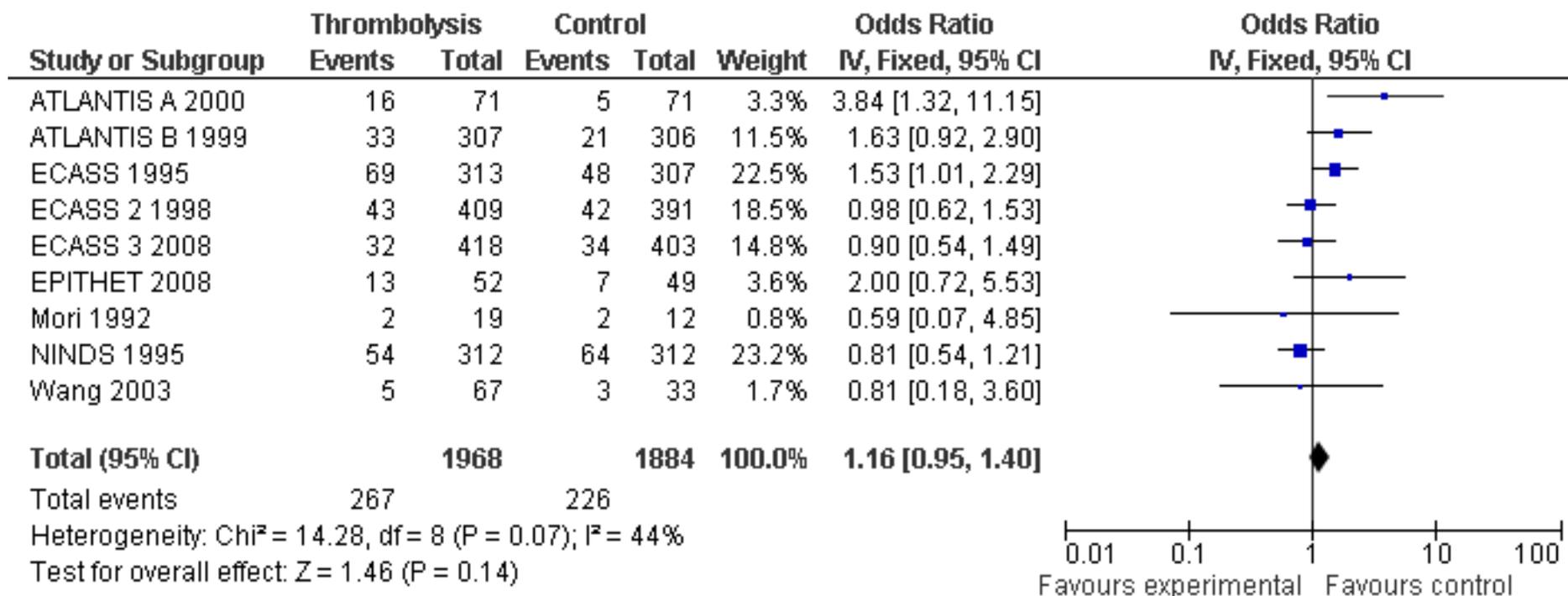
12.3 Practical difficulties

Pitfalls, etc.

- IMPACT head injury investigators have found that the proportional odds assumption mostly holds in their trial data.
- They say even if the data deviate considerably from proportional odds, it still gives a useful summary.
- However, it will hide 'kill or cure' effects if used without any other summary measures.

Thrombolysis (tPA) for acute ischaemic stroke

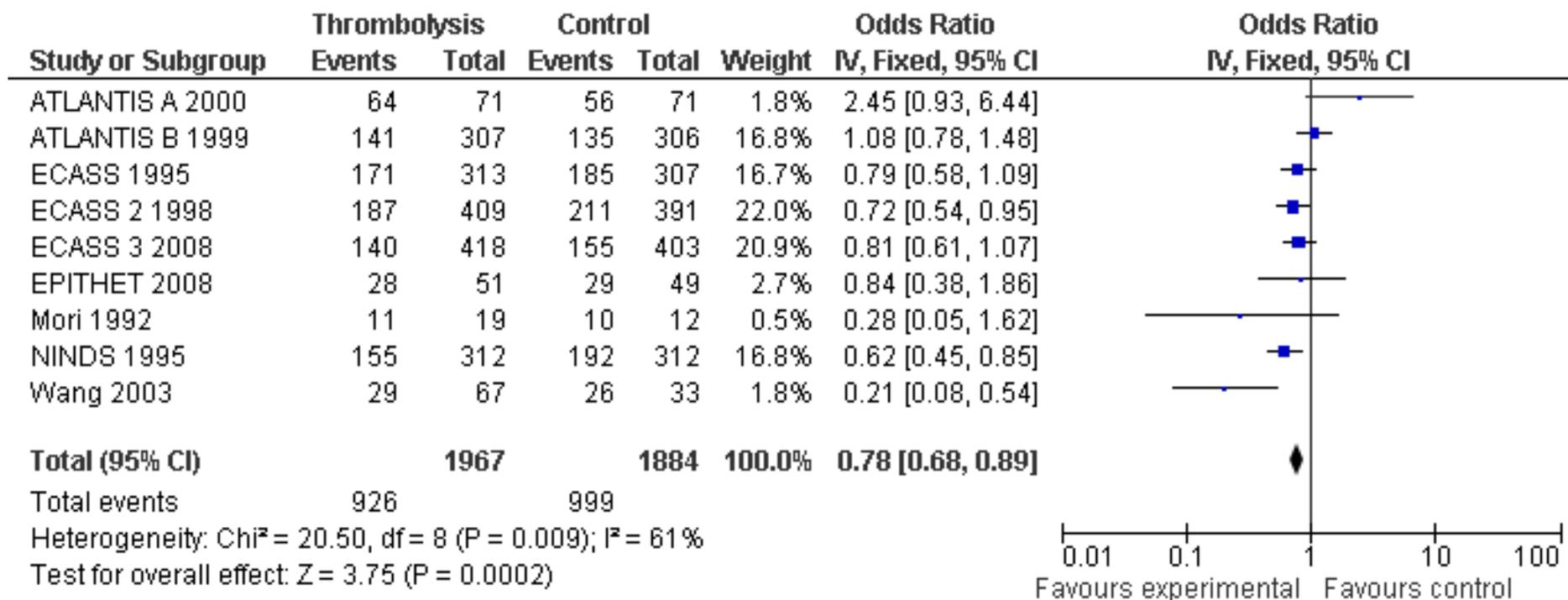
– Death during follow up



From Wardlaw 2009 (Only studies that report both death, and death and dependency included)

Thrombolysis (tPA) for acute ischaemic stroke

– Death or dependency during follow up



From Wardlaw 2009 (Only studies that report both death, and death and dependency included)

Collecting data

- You need the numbers of patients in each category of the ordinal scale for each intervention group if the proportional odds ratio method will be used.
- Full data probably more likely for shorter scales and more recent papers??

Gøtzsche paper – arthritis trials

Gøtzsche PC (2001). Reporting of outcomes in arthritis trials measured on ordinal and interval scales is inadequate in relation to meta-analysis. *Annals of the Rheumatic Diseases* 60, 349-352.

This paper examines whether ordinal scales were optimally reported (which includes reporting of original ordered categories but also allows mean and SD for pain scales). For the non-pain scales, there is no evidence of better reporting over time. Only one third to one half of ordinal scales were optimally reported.

ECASS1 (1995)

Hacke W, Kaste M, Fleschi C, Toni D, Lesaffre E, von Kummer R, et al (1995). Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke. The European Cooperative Acute Stroke Study (ECASS). *JAMA* 274, 1017-1025.

This paper presents the modified Rankin scale results as median score plus p-value in a table.

In the text it states “In the ITT analysis 29.3% of patients in the placebo arm and 35.7% of the rt-PA treated patients had RS scores better than 2 at 90 days (Table 3)”

NINDS (1995)

The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group (1995). Tissue plasminogen activator for acute ischaemic stroke. *New England Journal of Medicine* 333(24), 1581-1587.

This paper has a diagram with modified Rankin scale split into 4 categories. For each treatment group only percentages are given, not actual numbers.

ECASS3 [Hacke 2008]

Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, et al (2008). Thrombolysis with alteplase 3 to 4.5 hours after ischemic stroke. *New England Journal of Medicine* 359, 1317-1329.

This more recent paper has a diagram showing the full modified Rankin scale, split by treatment group. Only percentages are reported, but these are reported to one decimal place, which should be sufficient to work out the exact values.

You could mix binary and ordinal data...

- Reminder: The odds ratio calculated from the proportional odds model can be interpreted as the odds of success on the experimental intervention relative to control, irrespective of how the ordered categories might be divided into success or failure.
- If proportional odds holds, you could combine:
 - The original Rankin scale in 7 categories
 - A summarised Rankin scale in 4 categories
 - Binary data where the scale has been split at 0-2 vs 3-6
 - Dead vs Alive (category 6 on the scale vs 0-5).
- However, more categories = more weight in meta-analysis
- If proportional odds does not fit, mixing binary and ordinal data will increase heterogeneity.

Mixing different scales

- Methods are available for combining data from scales that are related but have different definitions for their categories (discussed in Anne Whitehead's book – Meta-analysis of controlled clinical trials, section 9.3).

Where next?

- An MRC project.
 - Practical methods for ordinal data meta-analysis in stroke
 - 1 June 2010 to 31 May 2012
- a. Review the methods available for meta-analysis of ordinal outcomes.
- b. Investigate using each of these methods in real data:
 - how often sufficient data are presented (or can be obtained),
 - how often the available data fulfil any distributional assumptions (and whether there are sufficient data to check assumptions),
 - how easy to understand the results are, and how much detail they show of the way the treatment effect operates.
 - assess the added statistical power gained by using ordinal and continuous data methods over binary methods.
- c. Develop a Cochrane workshop on ordinal methods.

Other free sources of help and advice

- The Cochrane Reviewer's handbook
 - <http://www.cochrane-handbook.org/>
- The Cochrane distance learning material
 - <http://www.cochrane-net.org/openlearning/>
- The Cochrane RevMan user guide.
 - <http://ims.cochrane.org/revman/documentation/rm5userguide.pdf>
(user guide)
 - <http://ims.cochrane.org/revman/documentation/Statistical-methods-in-RevMan-5.pdf> (all the equations)
- R meta analysis documentation.
 - <http://rss.acs.unt.edu/Rdoc/library/rmeta/html/00Index.html>