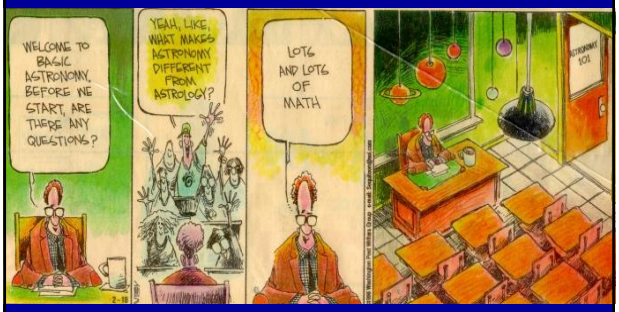
Statistics For The Terrified
Advanced level:

The Jelly Baby

Objectives for this session

- Share with you some teaching methods
- Learn from you
- Help some of you learn to make sense of results in *systematic reviews*
  - the blobbogram
  - "statistical heterogeneity"
  - the difference between fixed effects and random effects models
  - funnel plots
- Have fun!

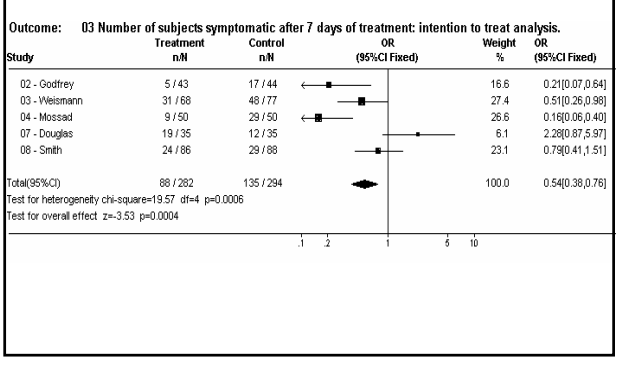
Statistics without numbers



Statistics for the terrified

- Making sense of results
- Measures of effectiveness
- P-values
- The confidence interval
- Bluff your way on the blobbogram

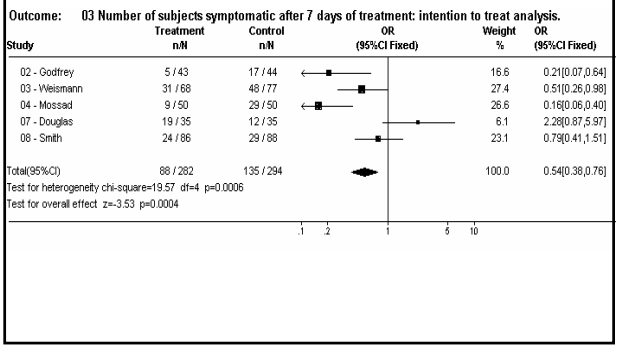
A blobbogram



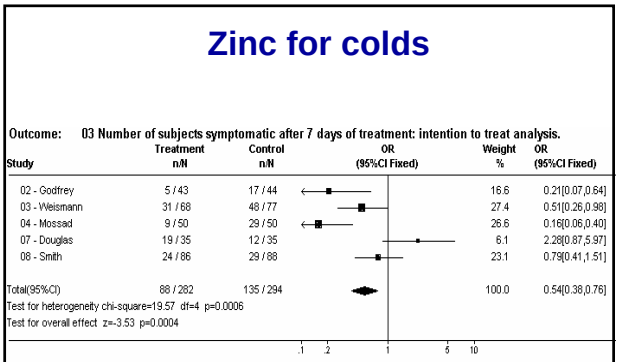
Critical appraisal of **any** study design must consider

- **Validity**
  - Can the study (results) be trusted?
- **Results**
  - What are the results and how are they (or can they be) expressed?
- **Relevance**
  - Do these results apply to the local context?

Zinc for colds



Zinc for colds


Outcome: 03 Number of subjects symptomatic after 7 days of treatment: intention to treat analysis.
Study
Treatment n/N
Control n/N
OR (95%CI Fixed)
Weight %
OR (95%CI Fixed)
02 - Godfrey 5 / 43 17 / 44 16.6 0.21[0.07,0.64]
03 - Weismann 31 / 68 48 / 77 27.4 0.51[0.26,0.98]
04 - Mossad 9 / 50 29 / 50 26.6 0.16[0.06,0.40]
07 - Douglas 19 / 35 12 / 35 6.1 2.28[0.87,5.97]
08 - Smith 24 / 86 29 / 88 23.1 0.79[0.41,1.51]
Total(95%CI) 88 / 282 135 / 294 100.0 0.54[0.38,0.76]
Test for heterogeneity chi-square=19.57 df=4 p=0.0006
Test for overall effect z=-3.53 p=0.0004
.1
.2
1
5
10

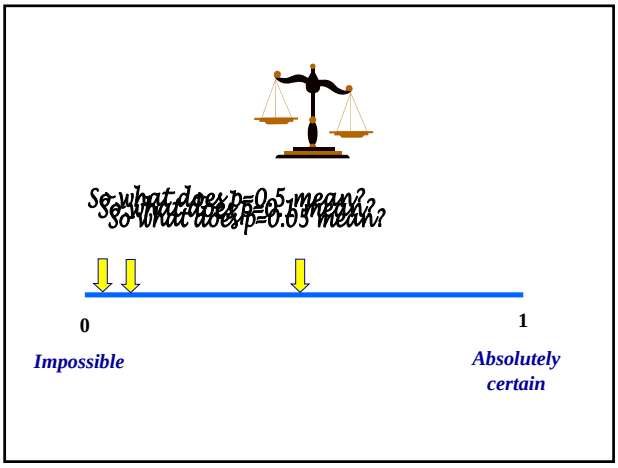
Warning!

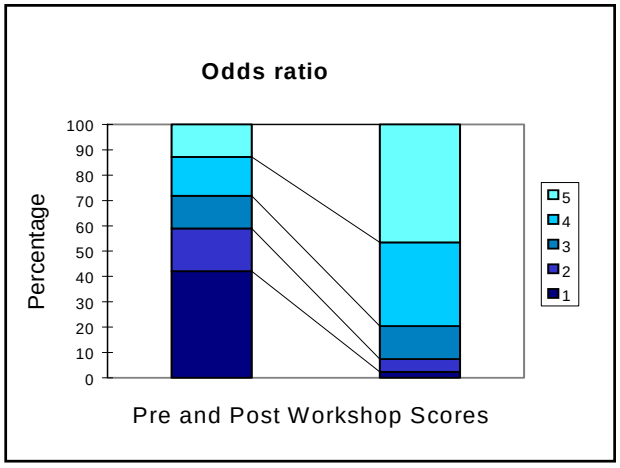
- Everything I say from now onwards assumes that the results being considered come from an unbiased study.
- It assumes that you have appraised the study and found it to be valid.

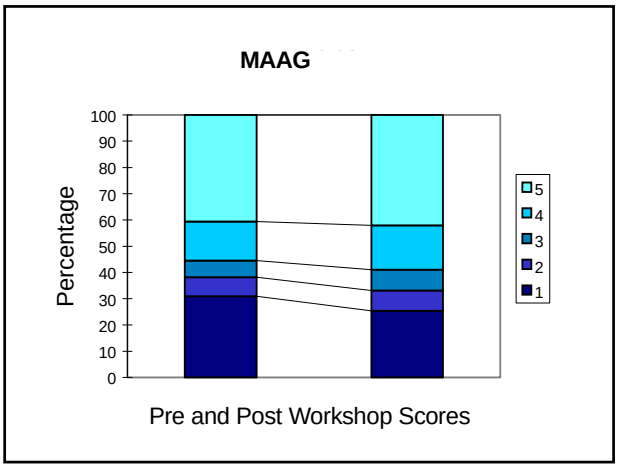
It could be due to chance!

- How can we express uncertainty due to chance?
- Null hypothesis
- How often would you get a result like this by chance if there were nothing going on?
- P-value in a nutshell

P-value in a nutshell







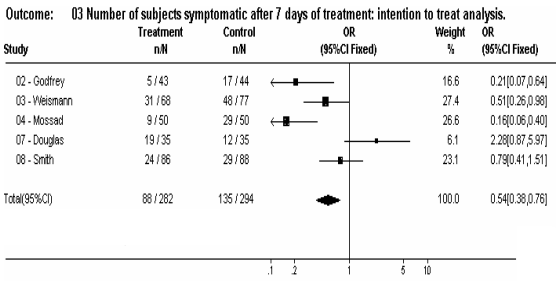
Moral:

Any observed difference between two groups, *no matter how small*, can be made to be "statistically significant" - at *any* level of significance - by taking a sufficiently large sample.

- How can we express uncertainty due to chance?
- Answer: the p-value
- But is there a better answer?

Introduction to confidence intervals

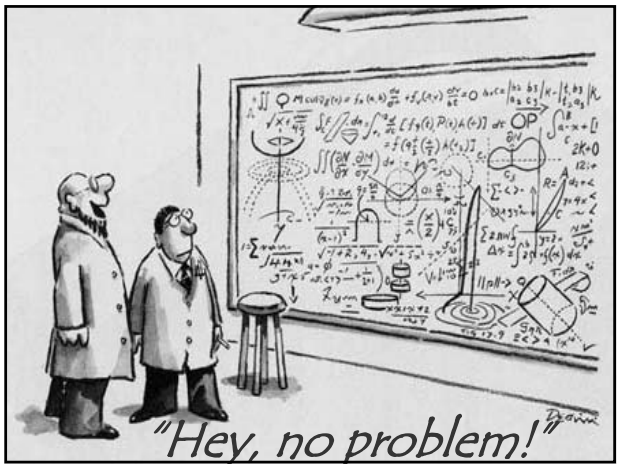
Blobbogram for zinc tablets for a cold!



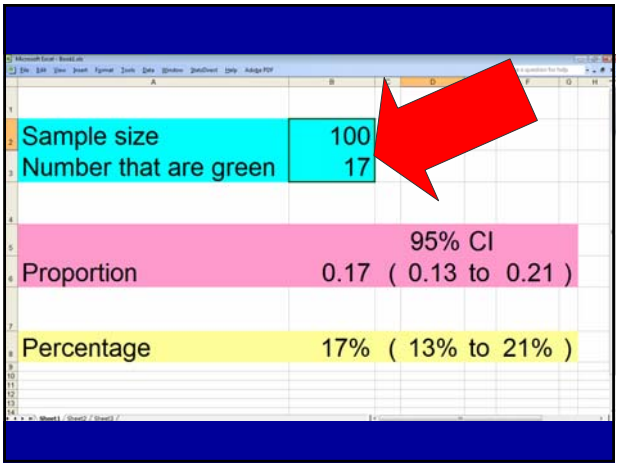
Why use a systematic review?

"The essence of good data analysis is the effective communication of clinically relevant findings"

Pocock SJ.
Clinical Trials: A Practical Approach, 1983



Combining results



5. If the results of the review have been combined, was it reasonable to do so?

- *HINT: Consider whether*
  - *The results were similar from study to study*
  - *The results of all the included studies are clearly displayed*
  - *The reasons for any variations in results are discussed*

What do we mean by "heterogeneity"?

- That things are not the same
- "Adding apples and pears"

In what way can studies be heterogeneous?

- Population
- Intervention
- Comparator
- Outocme
- Study design
- Time course
- Statistically

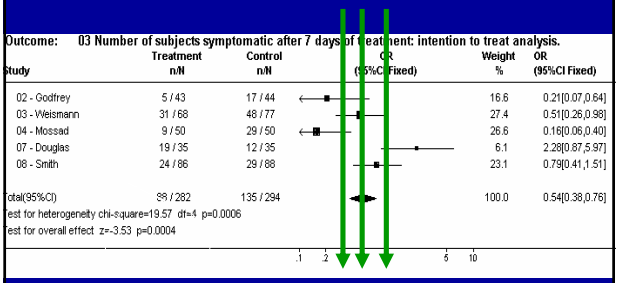
Looking for heterogeneity

- Common sense
- Clinical sense
- Statistical
  - Graphical
  - Calculation

Statistical heterogeneity

- Are the differences among the results of the studies greater than could be expected by chance?
- One way of doing this is look at blobbogram.

Zinc for colds



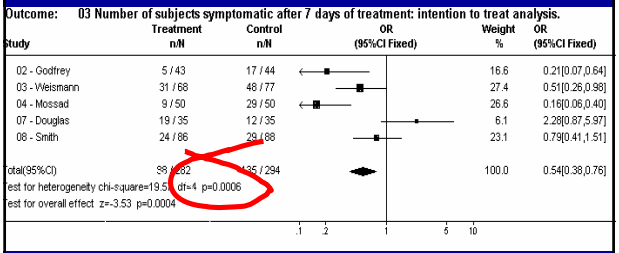
Statistical heterogeneity

- Are the differences among the results of the studies greater than could be expected by chance?
- If the CIs for the results of each study do not overlap, it means that the differences are statistically significant
- i.e. unlikely to be just due to chance – there is some underlying real difference

Statistical heterogeneity

- Tests for heterogeneity are formal statistical analyses
- They estimate how often the observed variation between study results would be expected by chance alone.
- The more significant the (the smaller the p-value), the more likely it is that the observed differences were not due to chance alone.

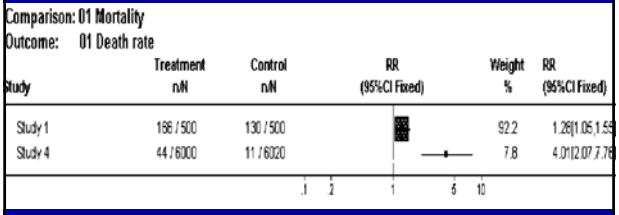
Zinc for colds



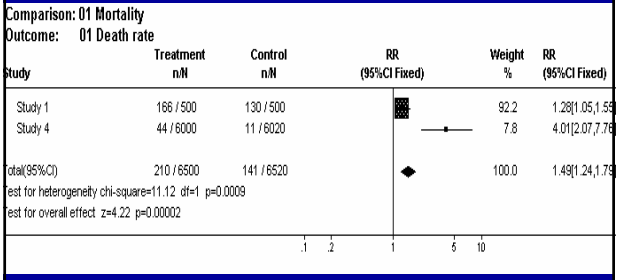
Combining studies when there is heterogeneity

- What can we do?

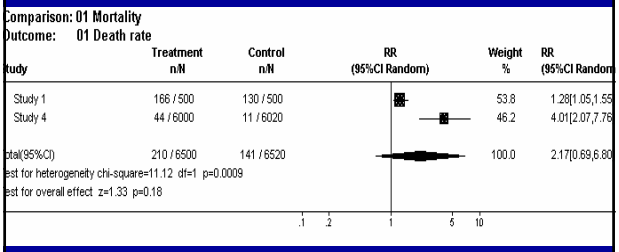
Random effects vs fixed effects



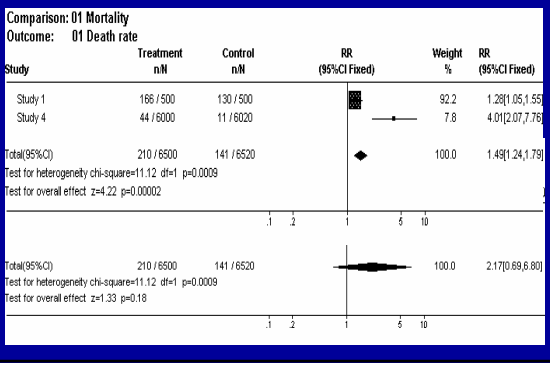
Random effects vs fixed effects



Random effects vs fixed effects



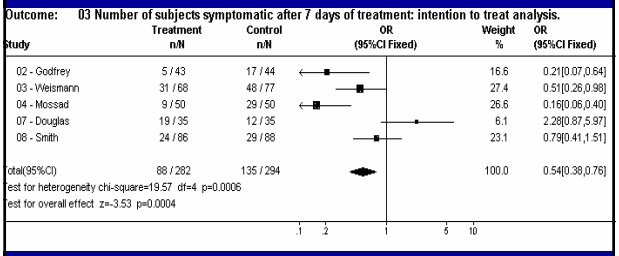
Random effects vs fixed effects



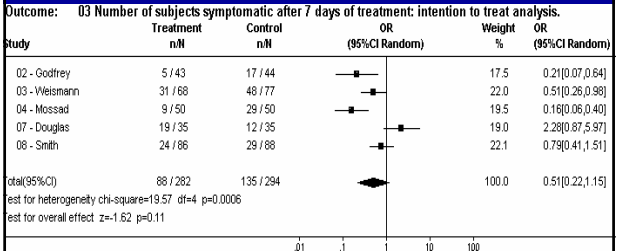
Random or fixed effects – Which is right?

"Using a random effects model substitutes the unrealistic assumption of the fixed effects model with another equally unrealistic assumption"

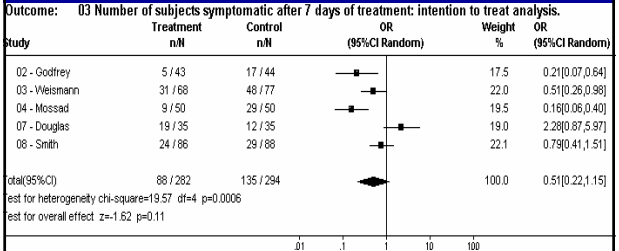
Zinc for colds – fixed effects



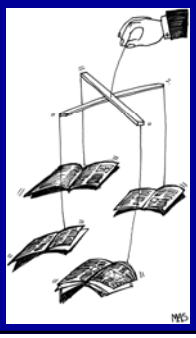
Zinc for colds – random effects



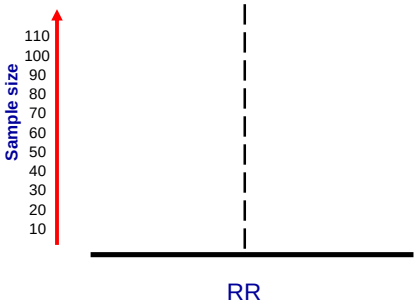
Zinc for colds – random effects



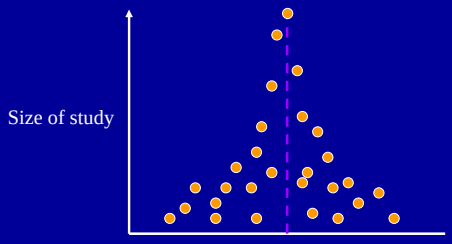
Looking for bias in systematic reviews



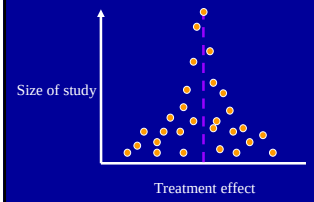
Publication bias



A funnel plot

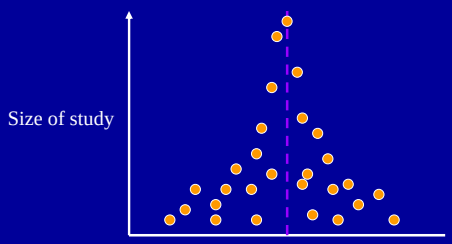


Funnel plots

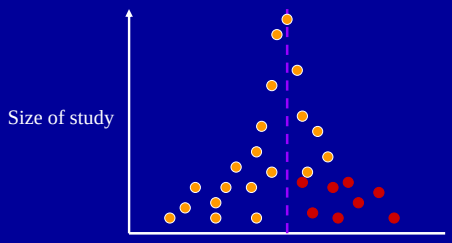


- Are scatter plots of treatment effect estimated from individual studies (x axis) against a measure of each study's sample size (y axis).
- The precision in the estimation of the treatment effect increases as sample size increases.
- Effect estimates from small studies scatter more widely at the bottom of the graph, with the spread narrowing among larger studies.
- In the absence of bias the plot should resemble a symmetrical inverted funnel.

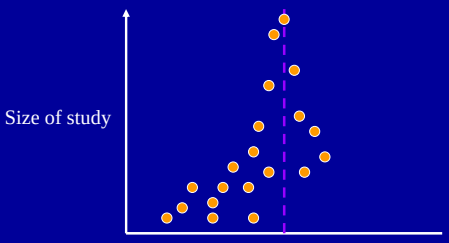
A funnel plot



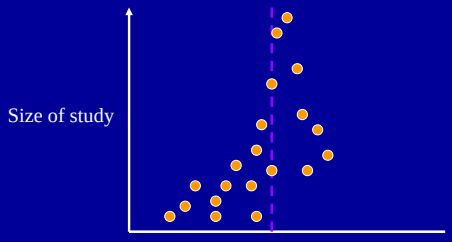
A funnel plot



A funnel plot



A funnel plot



Sources of asymmetry

- Publication bias
- Poor methodological quality of smaller studies
- Poor methodological design
- True heterogeneity i.e. Size of effect differs according to study size
  - for example, due to differences in the intensity of interventions or differences in underlying risk between studies of different sizes
- Chance

What should we do if there is heterogeneity?

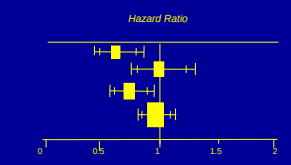
Chemotherapy for ovarian cancer: paclitaxel + platinum vs platinum

Josie Sandercock, Mahesh Parmar, Valter Torri, Wendi Qian. *BJC*, 2002; **87**: 815-824
MRC Clinical Trials Unit, London and Mario Negri Institute, Milan

Survival

	Paclitaxel + platinum	Platinum
GOG111	98/184	137/202
GOG132	154/201	158/ 200
OV10	183/342	220/338
ICON3	345/710	674/1364

$\chi^2_{het} = 13.75$ (3df) $p = 0.003$



paclitaxel/platinum better platinum based better

Making sense of the data

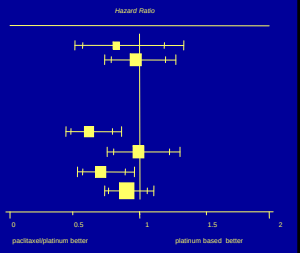
A number of explanations for heterogeneity have been proposed:

- **patients**
  different types of patients included in the trials
- **crossover**
  different rates of crossover to taxanes on the control arms
- **research arms**
  research arms differ in effectiveness
- **control arms**
  control arms differ in effectiveness

Type of patient

Survival

	Paclitaxel /platinum	Platinum based
Optimal resi dual disease		
OV10	54/132	57/116
ICON3	130/380	254/737
χ^2het = 0.595 (1df) p = 0.44		
Suboptimal residual disease		
GOG111	98/184	137/202
GOG -132	154/201	158/200
OV10	128/209	162/221
ICON3	215/330	420/627
χ^2het = 10.80 (3df) p = 0.013		

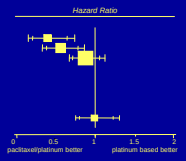


Crossover before progression

Survival

	Paclitaxel /platinum	Platinum
Little or no crossover before progression		
GOG111	98/184	137/202
OV10	183/342	220/338
ICON3	345/710	674/1364
χ^2het = 11.80 (2df) p = 0.003		
Substantial crossover before progression		
GOG132	154/201	158/200

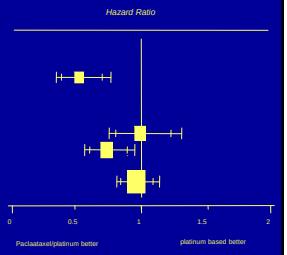
0.83 83.05



Crossover on progression

Survival

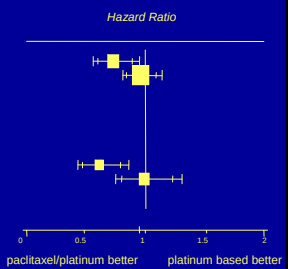
	Paclitaxel /platinum	Platinum based
Little or no crossover on progression		
GOG111	98/184	137/202
Substantial crossover on progression		
GOG132	154/201	158/200
OV10	183/342	220/338
ICON3	345/710	674/1364
χ^2het = 6.00 (2df) p = 0.05		



Paclitaxel schedule

Survival

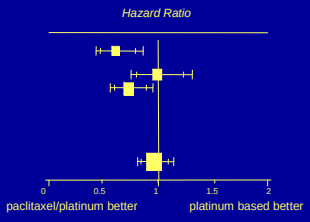
	Paclitaxel /platinum	Platinum
3 hour infusion of paclitaxel		
OV10	183/342	220/338
ICON3	345/710	674/1364
χ^2het = 5.14 (1df) p = 0.023		
24 hour infusion of paclitaxel		
GOG111	98/184	137/202
GOG-132	154/201	158/200
χ^2het = 7.93 (1df) p = 0.0054		



Platinum agent used

Survival

	Paclitaxel /platinum	Platinum
Cisplatin used in research arm		
GOG111	98/184	137/202
GOG132	154/201	158/200
OV10	183/342	220/338
χ^2het = 8.6 (2df) p = 0.014		
Carboplatin used in research arm		
ICON3	345/710	674/1364



Differences in control arms

Overall survival

	Paclitaxel /platinum	Platinum
cyclophosphamide/cisplatin		
GOG111	98/184	137/202
OV10	183/342	220/338
χ^2het = 1.17 (1df) p = 0.28		
single agent platinum or CAP		
GOG132	154/201	158/200
ICON3 carbo	230/478	472/943
ICON3 (CAP)	115/232	202/421
χ^2het = 0.33 (2df) p = 0.85