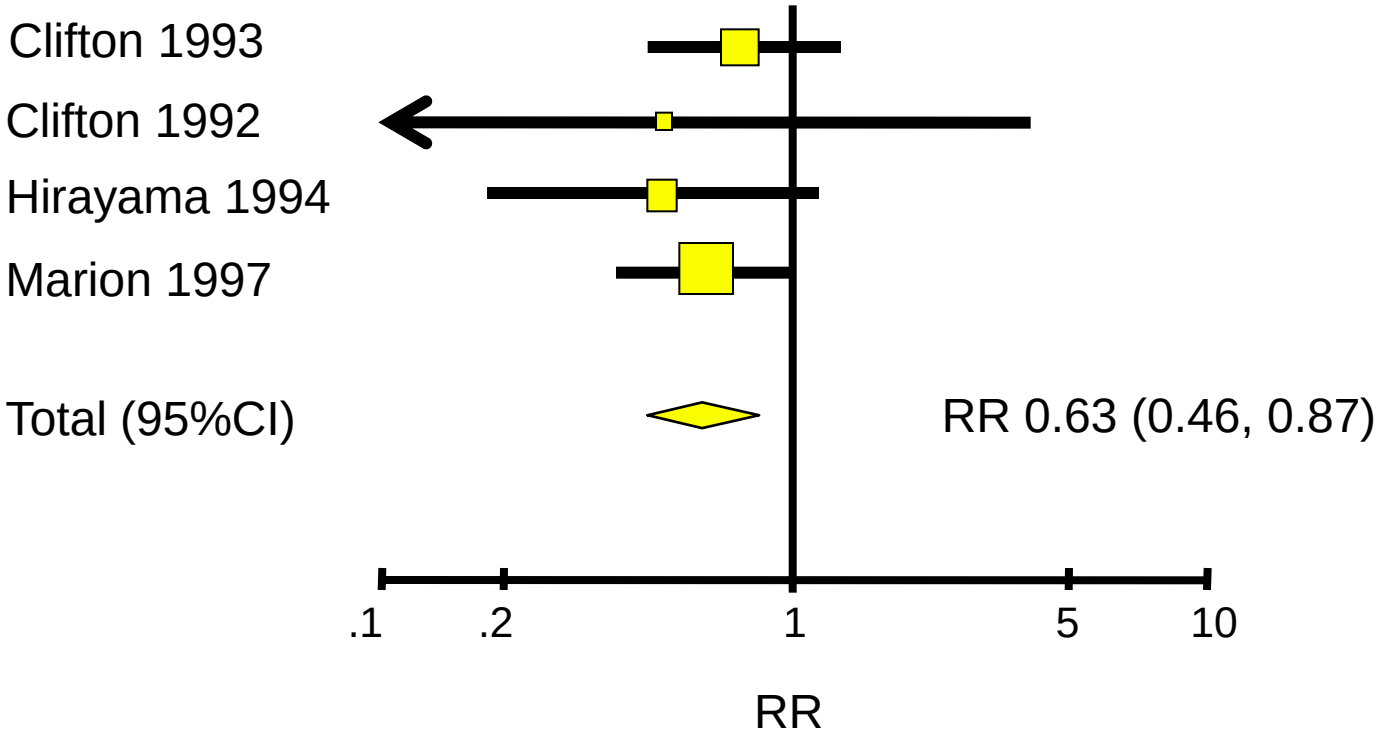
ebhc CONFERENCE 2015

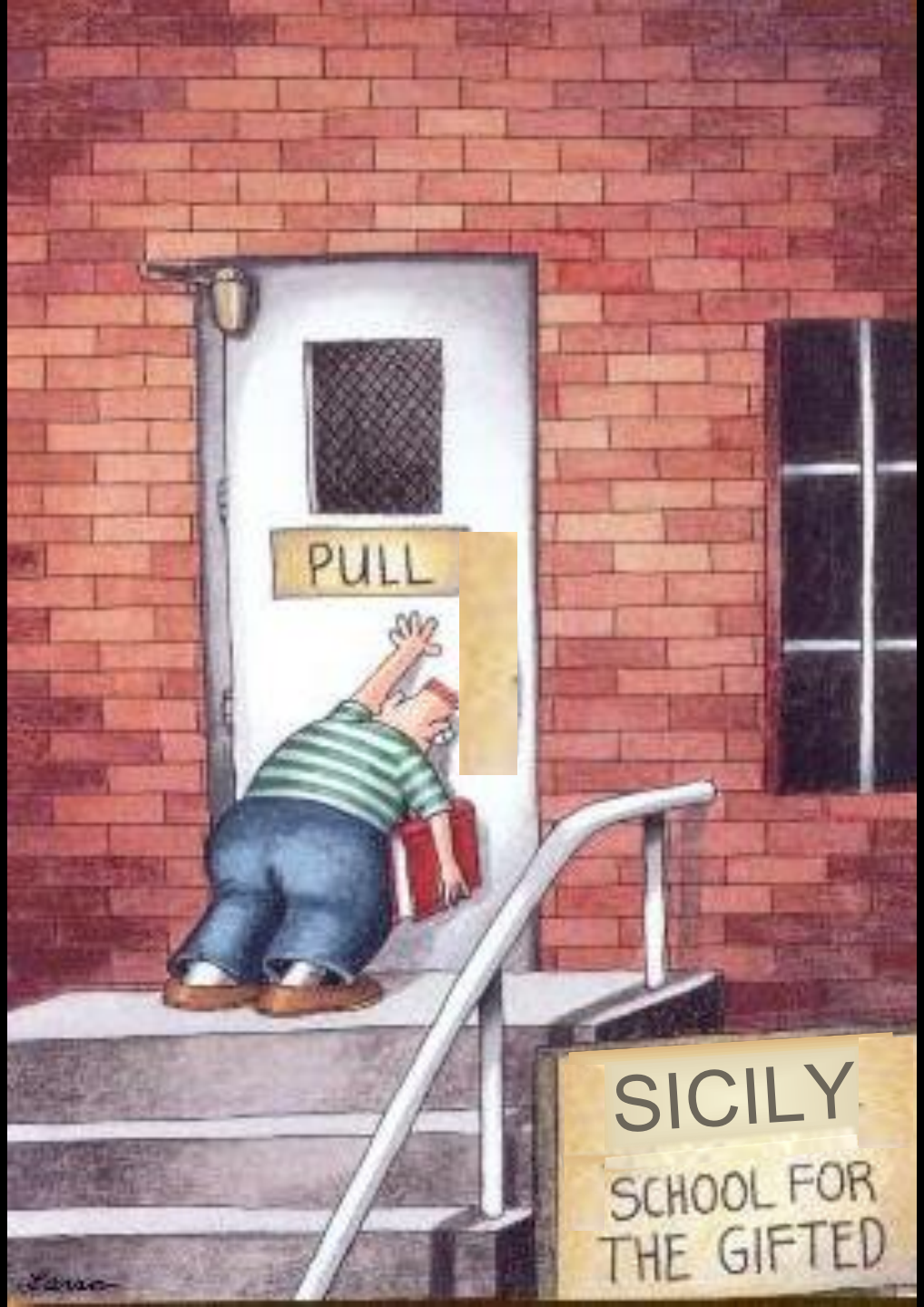
Making Sense of Results – Statistics for the Terrified

Amanda Burls
Professor Public Health, City University London
Twitter ID: #ajburls
Email: Amanda.Burls.1@city.ac.uk

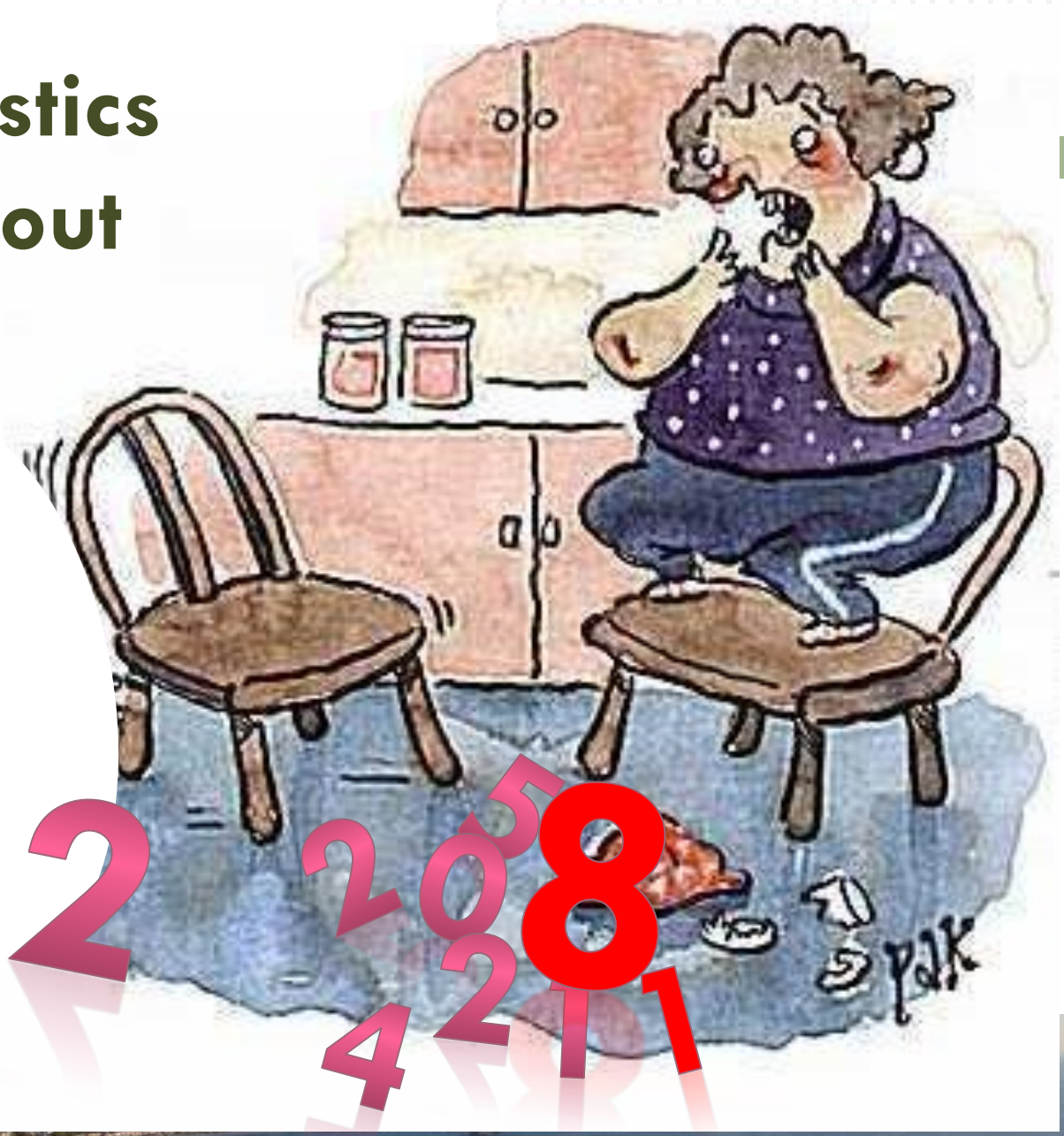
Hypothermia vs. control

In severe head injury
Mortality or incapacity (n=158)



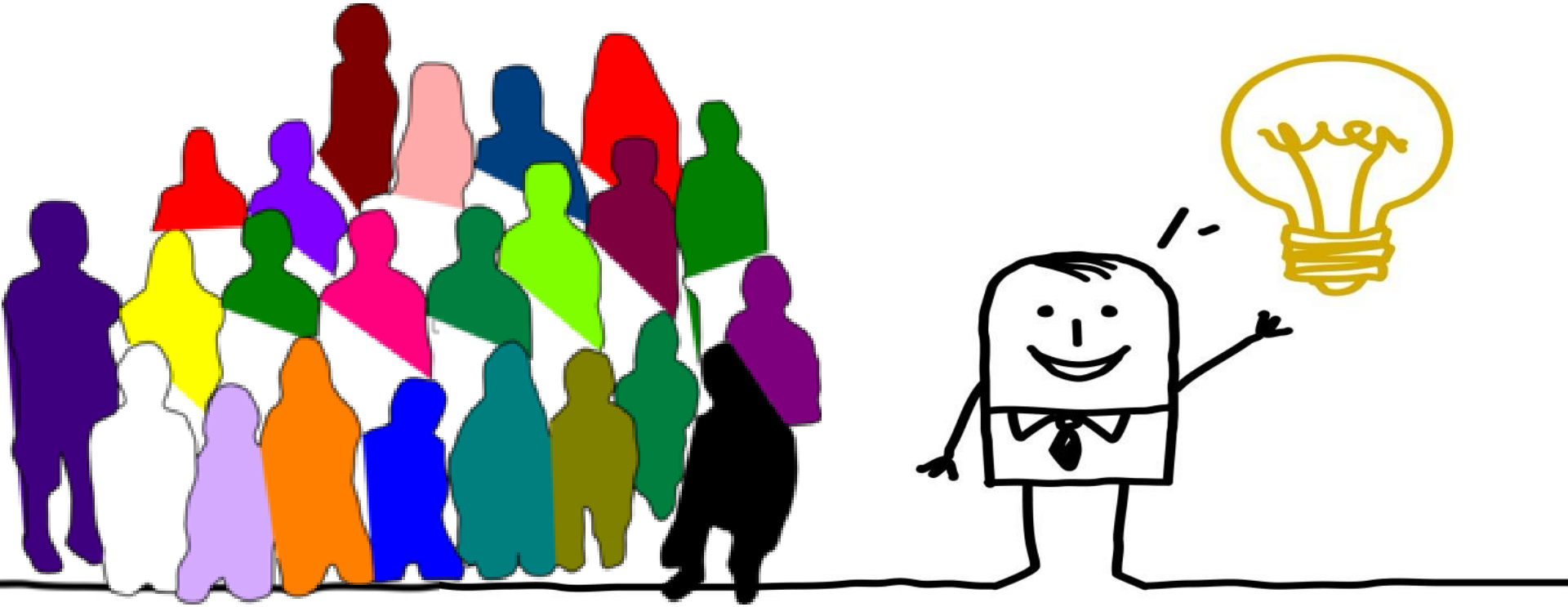

PULL
SICILY
SCHOOL FOR
THE GIFTED

Statistics without fear



Statistics without fear

What I want to know is what you want to know!



Before we start, let's limber up…

What are the important things to think about when you are using research evidence to help inform your decisions?

Critical appraisal : three things to look for in a paper

- Validity
- Results
- Relevance

Appraisal of **any** study must consider

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

Validity for an intervention study?

0:30

Validity for an intervention study?

(Randomised controlled trial)

0:29

Validity for an intervention study?

(Randomised controlled trial)

End

Validity for an RCT

- Randomised
- Concealment of allocation
- Similar baseline characteristics
- Blinding
- Treating groups the same
- Minimal losses to follow up
- Intention to treat analysis

Appraisal of **any** study must consider

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

Warning!

- Everything I say from now onwards assumes that the results being considered come from an *unbiased* study!

How are results summarised?

- Most useful studies compare at least two alternatives.
- How can the results of such comparisons be expressed?

Well conducted RCT – **no** bias!

Expressing results: What did the study show?

- Patients with backache:
  - 10 randomised to receive Potters
  - 10 randomised to receive placebo
- After 3 months:
  - 2 get better on Potters
  - 1 get better on placebo
- **Summarise this result to your neighbour in at least three different ways**

End

ebhc CONFERENCE 2015

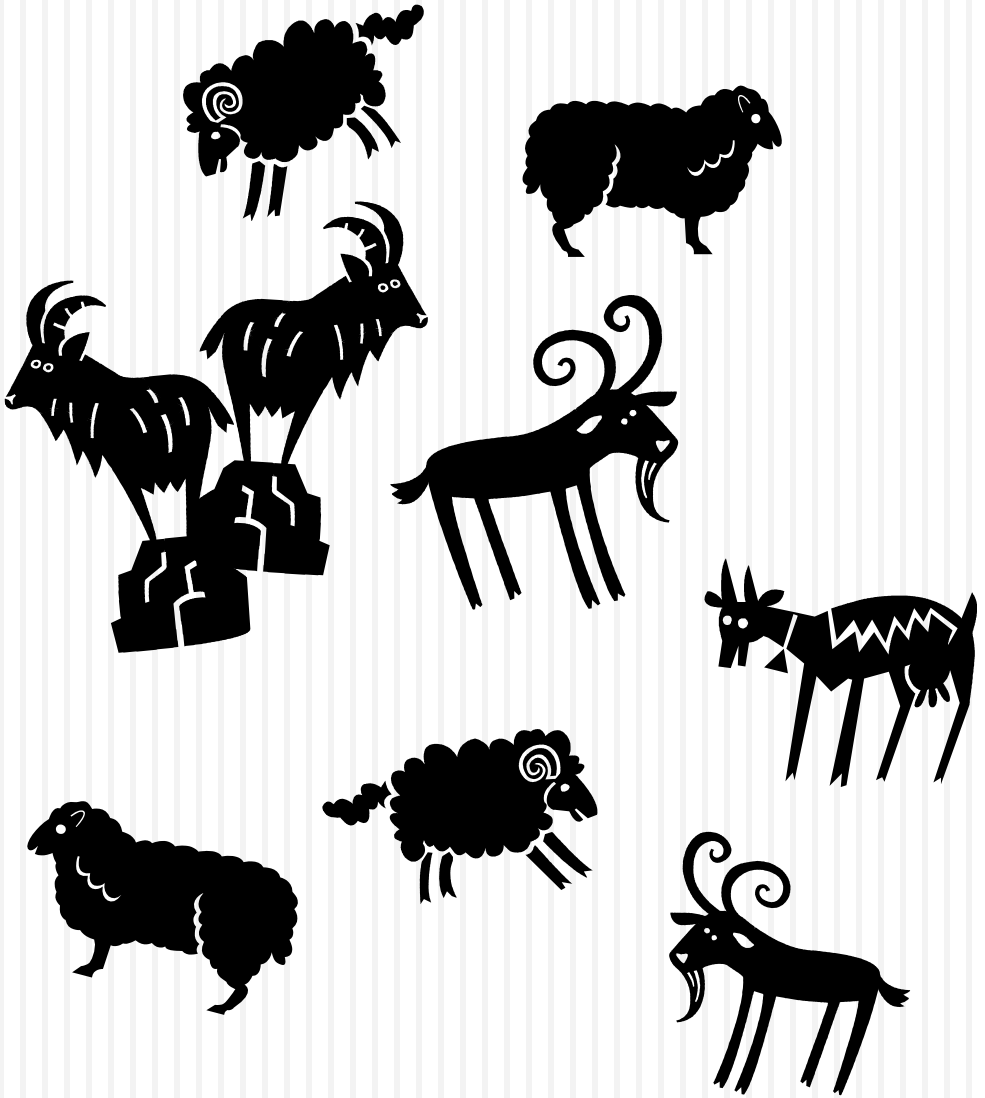
Summarise

- 2 out of 10 (20%) better on Potters
- 1 out of 10 (10%) better on placebo

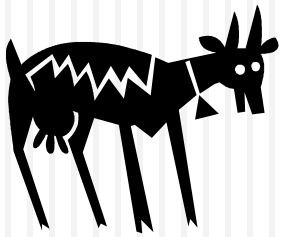
- **Twice as likely to get better on Potters**
- **An extra 10% of people get better on Potters**
- **For every 10 people with back pain given Potters, one case of back pain is improved**

Odds and Risk

- Risk (chance) is the number with the event of interest divided by the whole population
- Odds is the number with the event of interest divided by the number without the event of interest



- Chance (risk) of being a sheep?
- Odds of being a sheep?



Odds – separating the sheep from the goats

Measures of Relative Risk

How much more likely an outcome (or risk factor/exposure) is in one group compared to the other.

- Risk Ratio (RR)
  - RR = 2.0 (Twice as many recovered on Potters)
- Odds Ratio (OR)
  - OR = ?

Bottom of hill

Top

3 sheep
7 goats
Risk – 3/10 or 0.3
Odds – 3/7 nor 0.43

Bottom of hill

6 sheep
5 goats
Risk = 6/11 = 0.55
Odds = 6/5 = 1.2

Top

Measures of Relative Risk

- Risk Ratio (RR)
  - RR = 0.55
- Odds Ratio (OR)
  - OR = 0.36

Summarise

- 2 out of 10 (20%) better on Potters
- 1 out of 10 (10%) better on placebo

- **Twice as likely to get better on Potters**

Measures of Relative Risk

- Risk Ratio (RR)
  - RR = 2.0
- Odds Ratio (OR)
  - OR = 2.25

Risk difference

- The difference in the proportions recovering – the proportion of patients benefitting from treatment
- 20% improved on Potters, but 10% improved on placebo, so the risk difference is 10%

Number needed to treat (NNT)

- The number of patients to whom the new intervention needs to be given to produce one *extra* patient who is helped
- NNT = 1/risk difference
- Why?

How were the results summarised?

Two basic ways to summarise results of studies that compare groups:

1. Difference (take them away)
2. Ratio (divide)

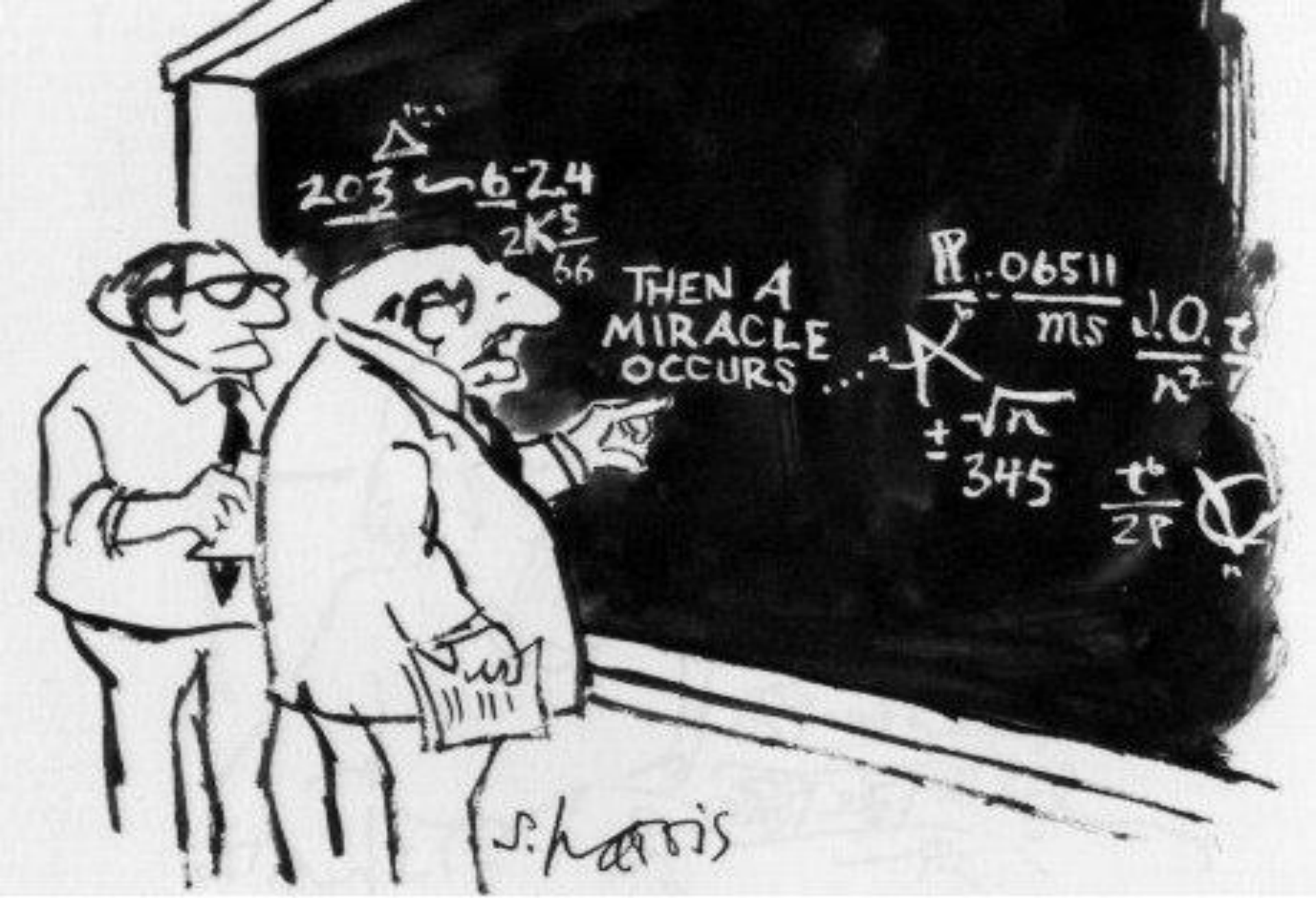
Do you think this study proves that Potters works?


I THINK I'VE
DISCOVERED SOMETHING

“It could have happened by chance!”

"It could have happened by chance!"

- What if there had been 1000 people in
  - 200 got better with Potters
  - 100 got better on placebo?
- Would you believe Potters works now?



"I think you should be more explicit here in step two."

What is the minimum number you would want in each arm to believe the trial?

A SECRET BALLOT

Assume similar effect size:
10% better with placebo
20% with Potters

- Write on a piece of paper your estimate
- Fold your paper in half and half again
- Swap it with your neighbour
- Swap the paper again with someone else
- Keep swapping until you don't know who's paper you have

ebhc CONFERENCE 2015

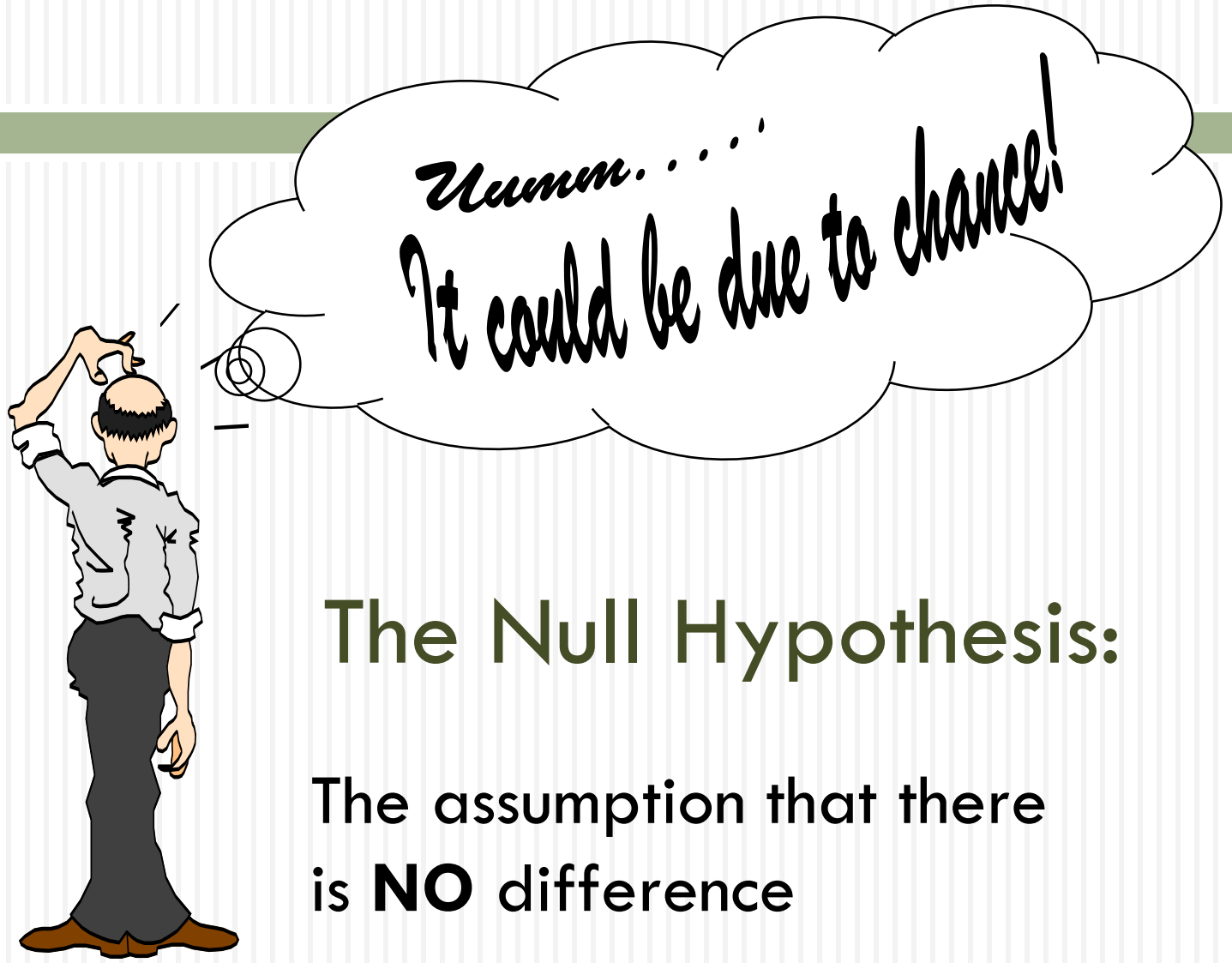
Scores

- 0-20
- 21-40
- 41-60
- 61-100
- 101-200
- >200

Quantifying uncertainty

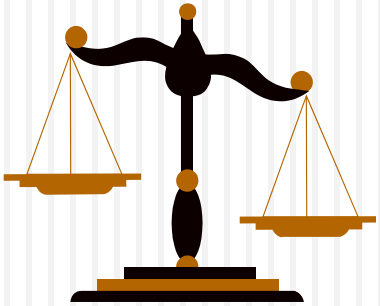
p-value

ebhc CONFERENCE 2015



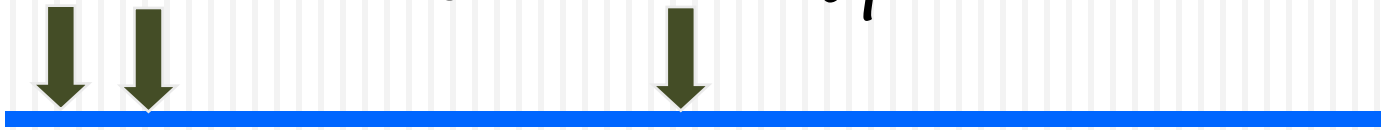
The Null Hypothesis:

The assumption that there is **NO** difference



So what does p=0.1 mean?

So what does p=0.05 mean?



0

1

Impossible

Absolutely certain

“Statistical significance”

- When the result observed is unlikely to have occurred by chance more often than 1 in 20 of the time
- $p<0.05$

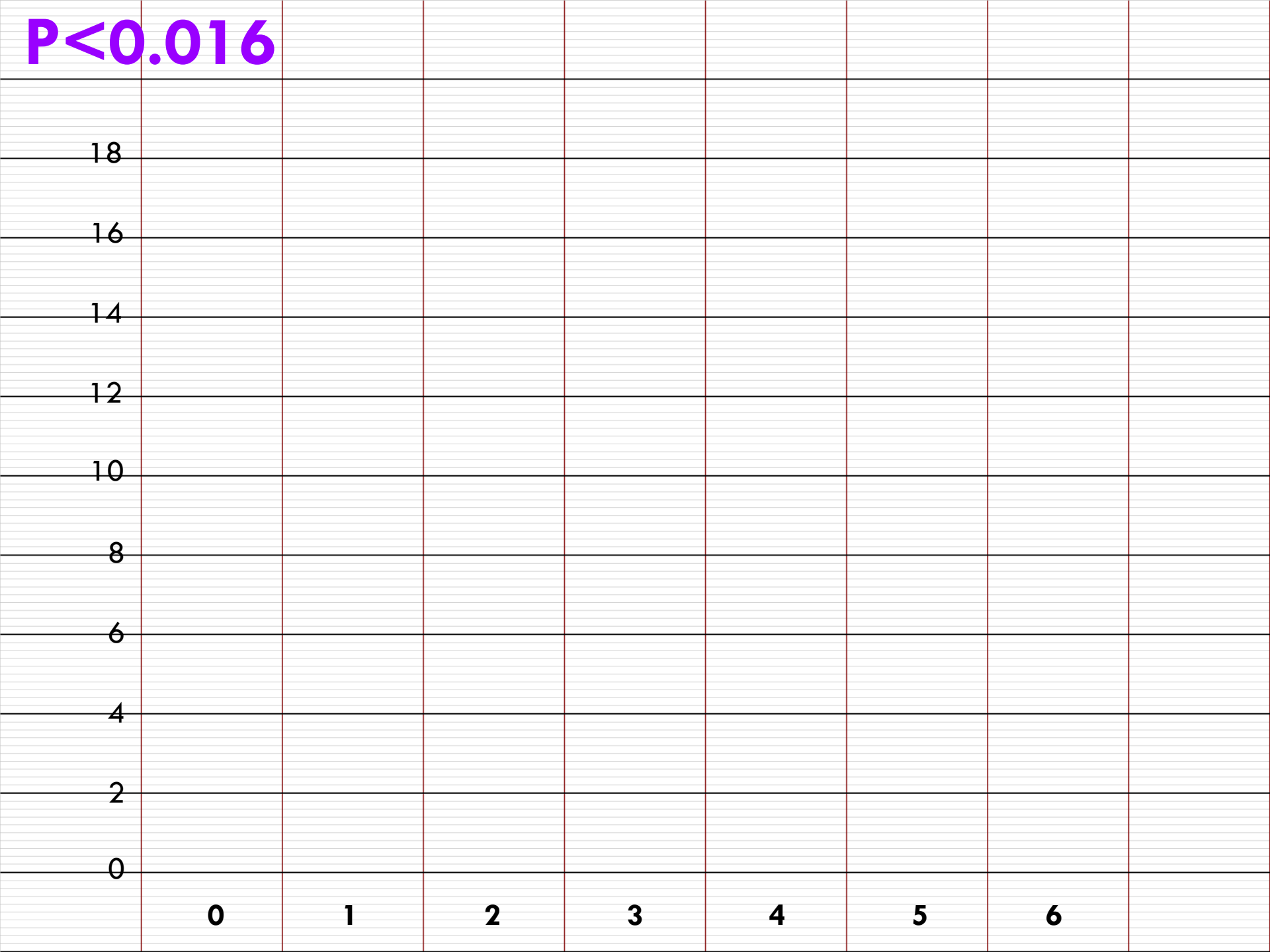
“Statistical significance”

- The p-value indicates the chance of a result, as or more extreme than the result observed, occurring **if** the null hypothesis (no difference) is true
- The p-value gives the strength of evidence **against** the null hypothesis (lower is more)
- Most studies use a “significance level” of 95% ($p<0.05$)

Potters	Placebo	P-value
2/10	1/10	P = 0.531
4/20	2/20	P = 0.376
6/30	3/30	P = 0.278
8/40	4/40	P = 0.210
10/50	5/50	P = 0.161
12/60	6/60	P = 0.125
14/70	7/70	P = 0.097
16/80	8/80	P = 0.076
18/90	9/90	P = 0.060
20/100	10/100	P = 0.048
100/500	50/500	P < 0.0001
200/1000	100/1000	P < 0.0001

Why p<0.05 as the cut-off?

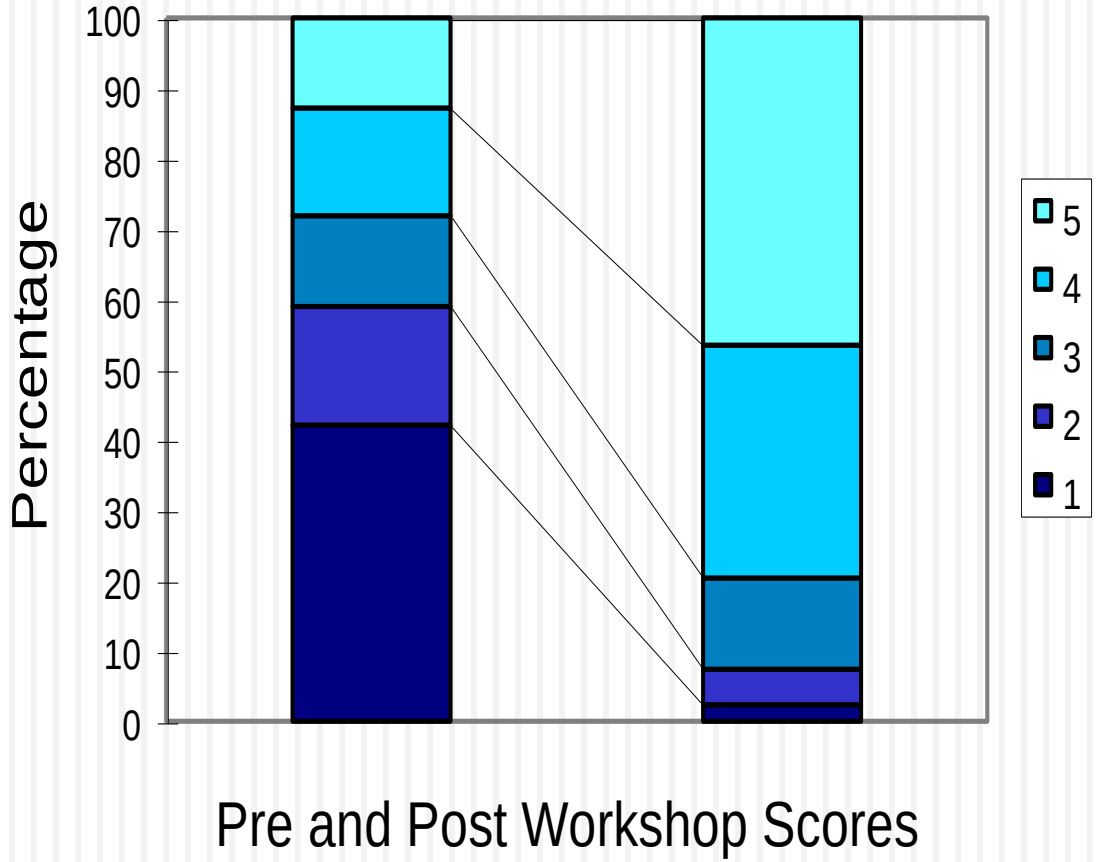
- Convention!
- The p-value is a measure of the strength of the evidence against the null hypothesis (assuming an unbiased trial)
- No magic cut-off between "statistically significant" and not (although many behave as if there were)


P<0.016
18
16
14
12
10
8
6
4
2
0
0
1
2
3
4
5
6

Self-assessed understanding - score

- 5 - I understand the term and could explain it
- 4 - I understand the term but could not define it
- 3 - I know have a vague idea what it means
- 2 - I have heard it but don't know what it means
- 1 - I have never heard of the term

“Odds ratio”


100
90
80
70
60
50
40
30
20
10
0
Percentage
5
4
3
2
1
Pre and Post Workshop Scores

Do you think this is likely to have happened by chance?

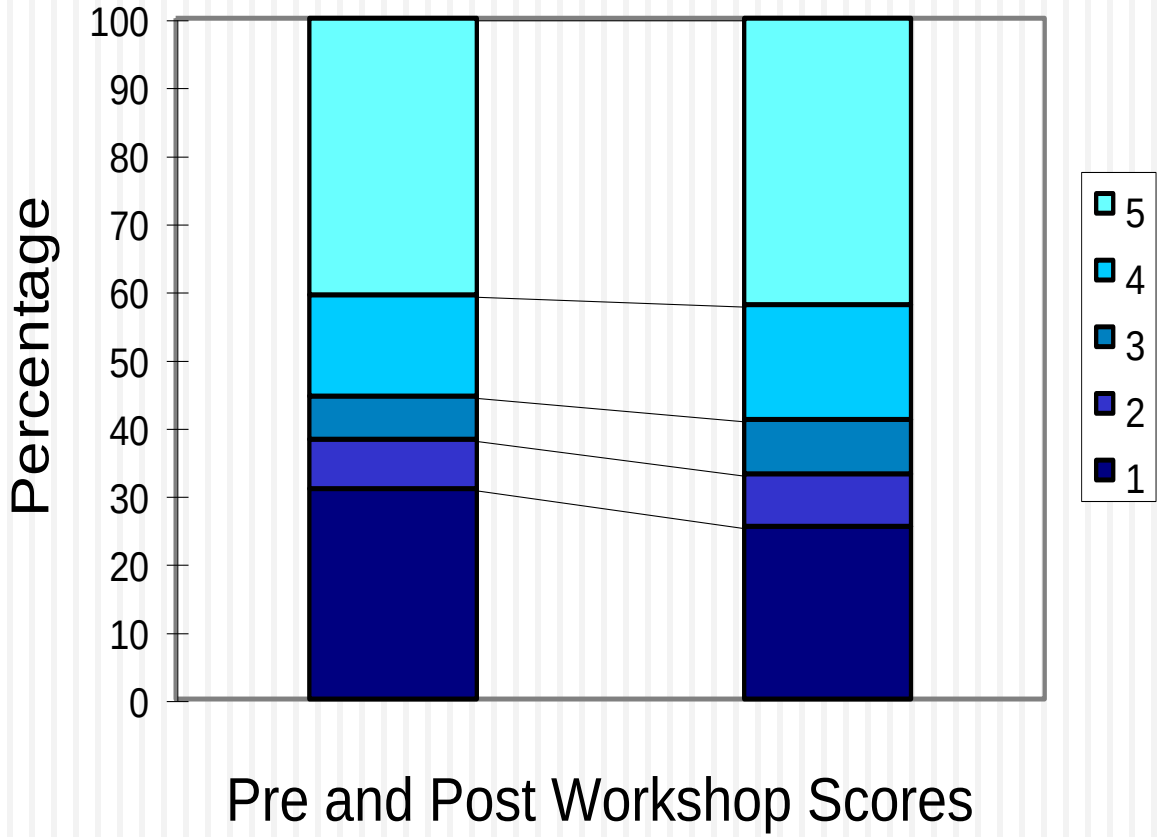
1. Yes
2. Don't know
3. No

Do you think this is likely to have happened by chance?

1. Yes
2. Don't know (~1000)
3. No

# P<0.00001

“MAAG”


100
90
80
70
60
50
40
30
20
10
0
Percentage
5
4
3
2
1
Pre and Post Workshop Scores

Do you think this is likely to have happened by chance?

1. Yes
2. Don't know
3. No

# P<0.00001

Limitation of the p-value

Any 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.

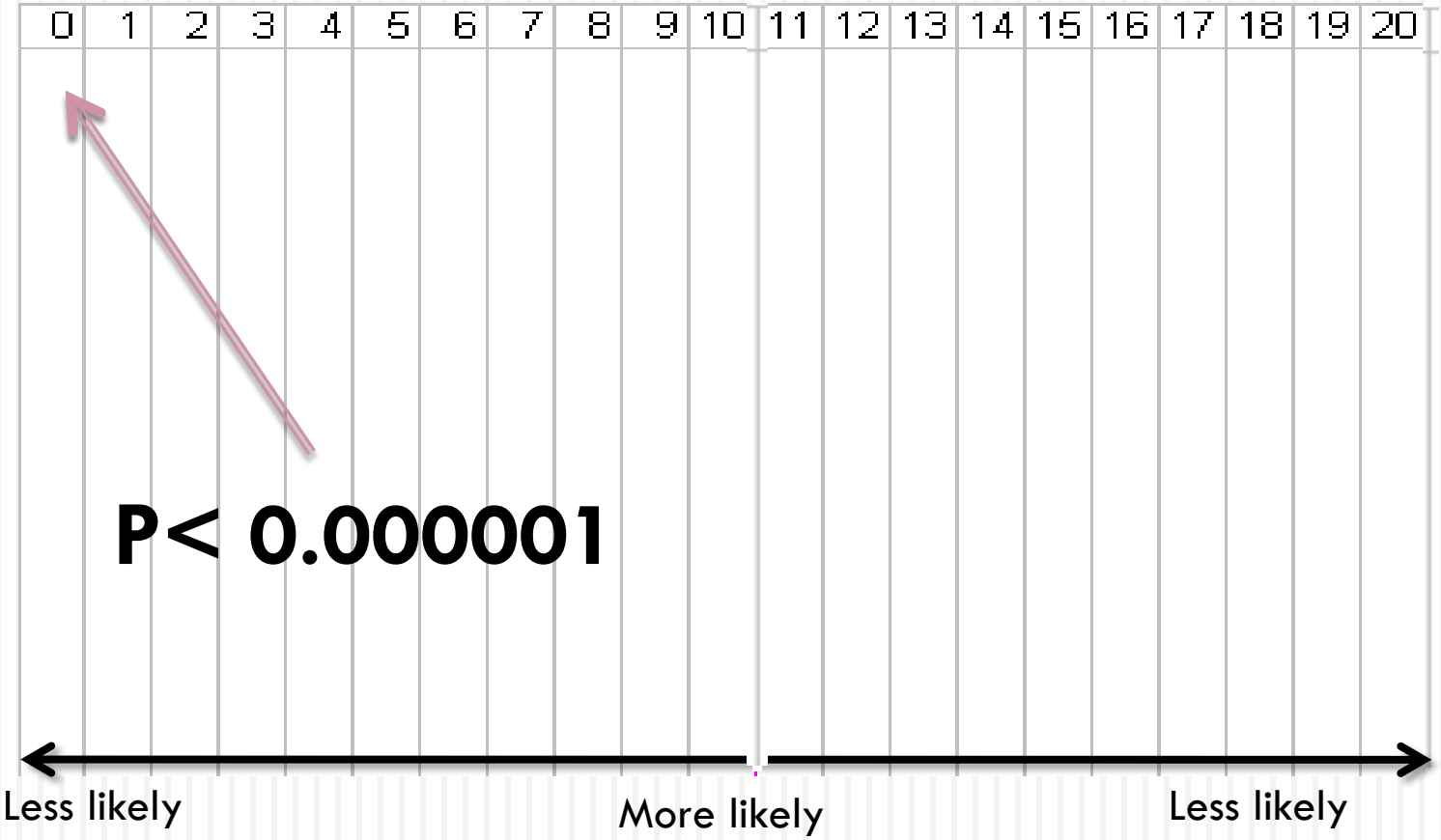
Is there a better way of expressing uncertainty?

- Yes - the confidence interval

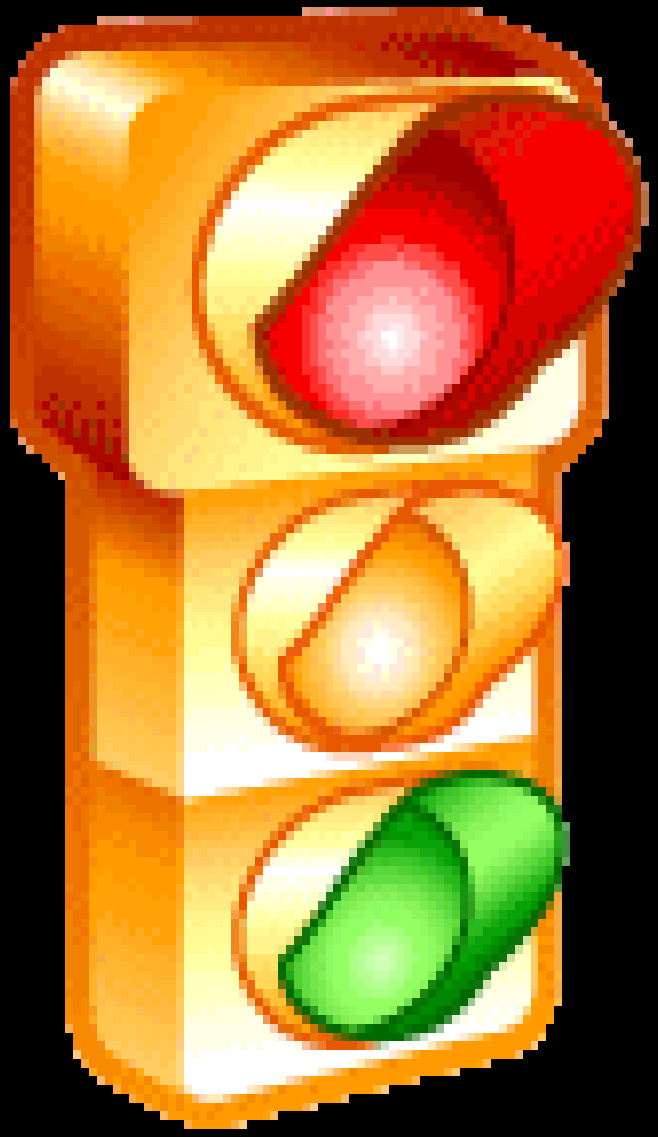
Introduction to confidence intervals

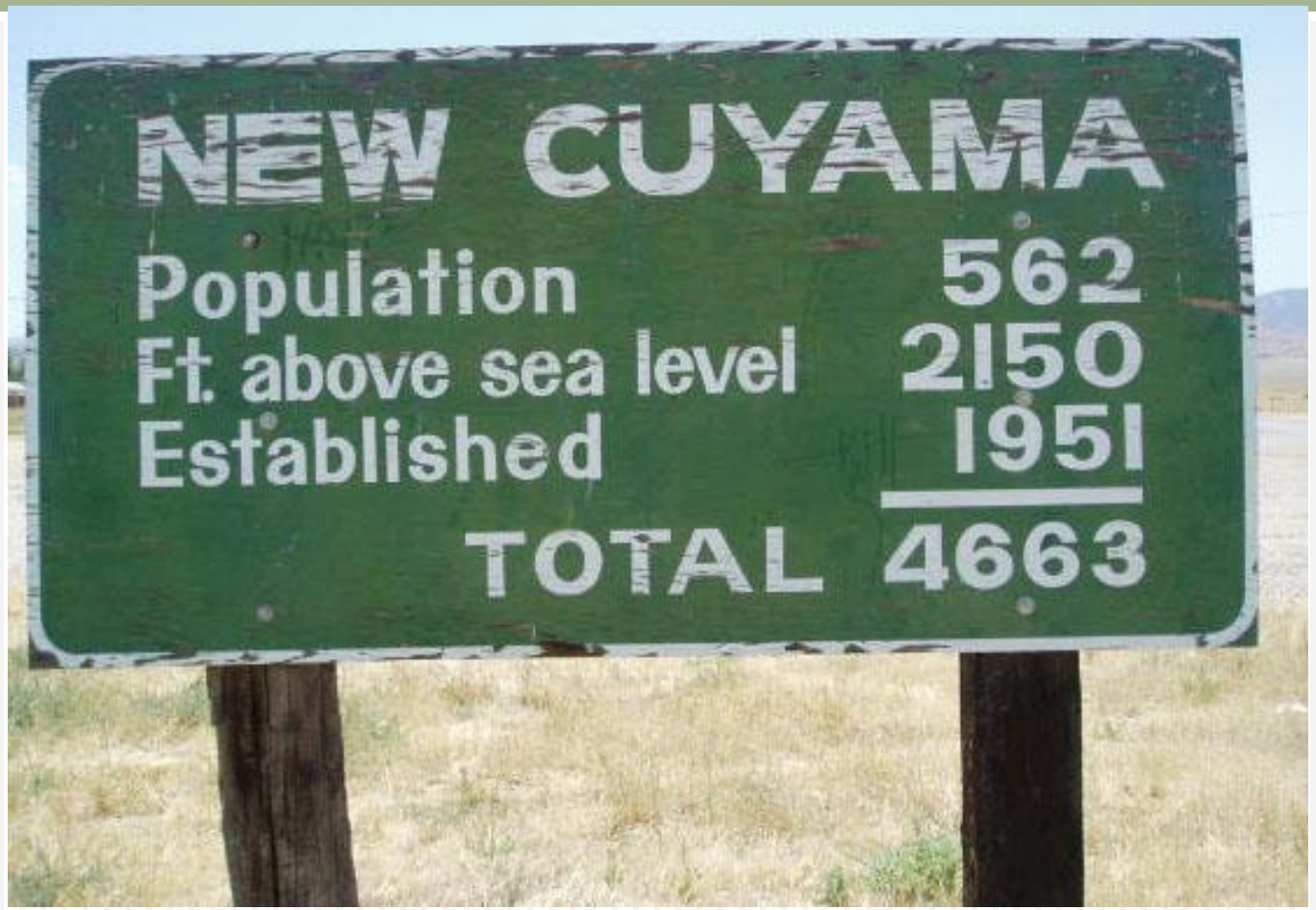
- CIs are a way of showing the uncertainty surrounding a point estimate.

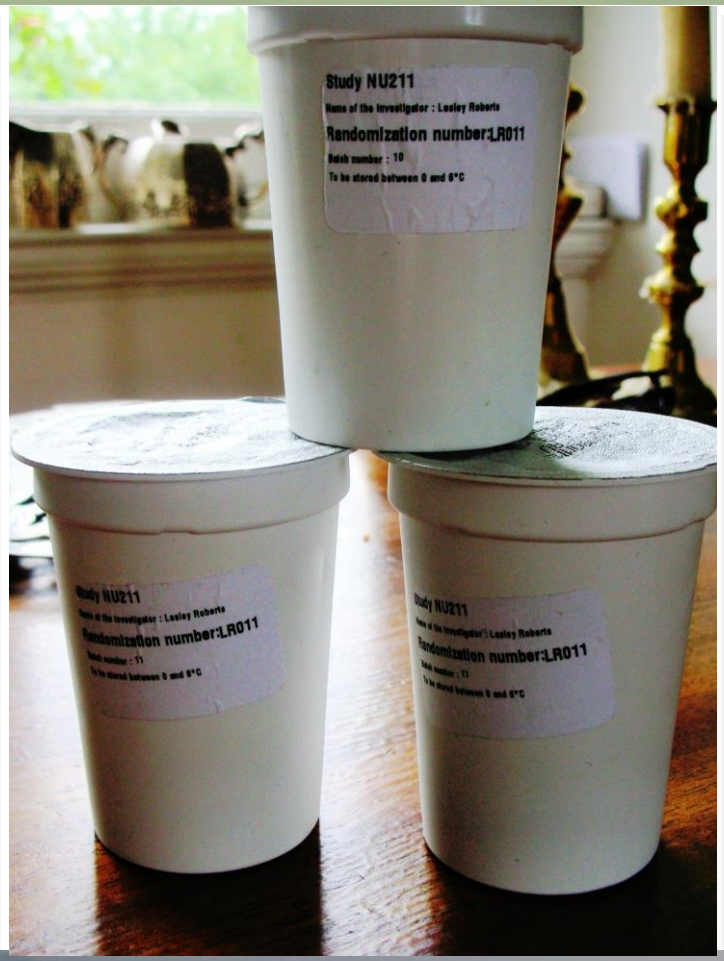
How many **Red** sweets did I pick?



Statistical significance does not imply clinical significance!




NEW CUYAMA
Population 562
Ft. above sea level 2150
Established 1951
TOTAL 4663



Probiotic yoghurt trial

ebhc CONFERENCE 2015

Sent: 19 January 2010 15:45

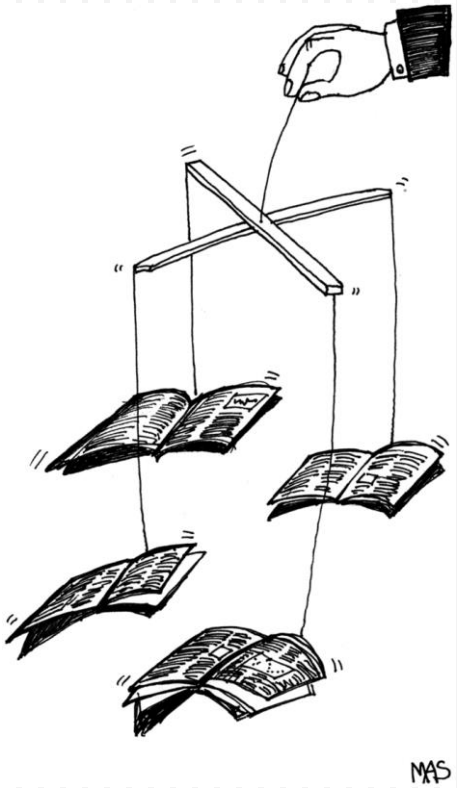
To: Dr Amanda Burls

Subject: RE: Yoghurt trial

Dear Amanda,

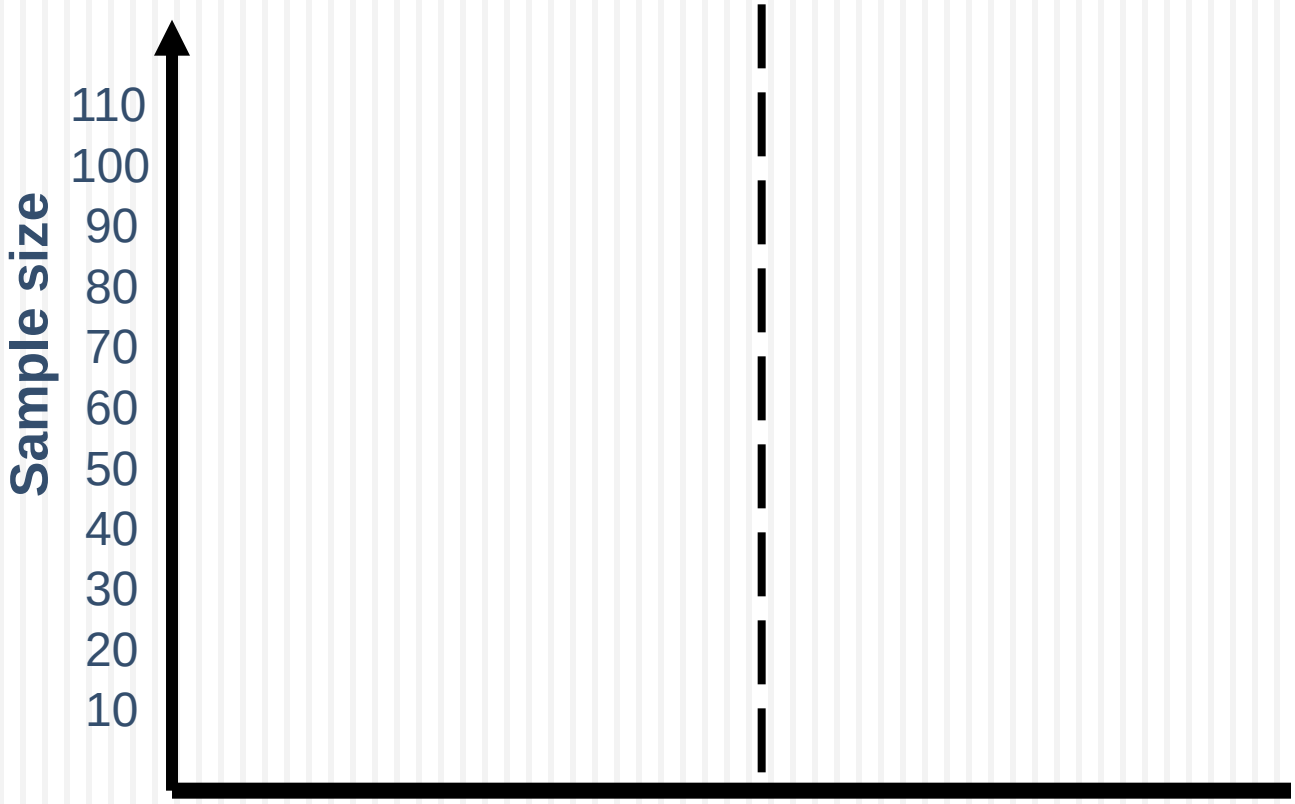
The trial is not yet in press - this is in part due to the much longer than anticipated further analysis of the data at the funders request. In summary this was a negative trial - although both groups demonstrated benefit, those in the active product group did not show greater benefit and at times the difference actually favoured the control product….

Looking for bias in systematic reviews

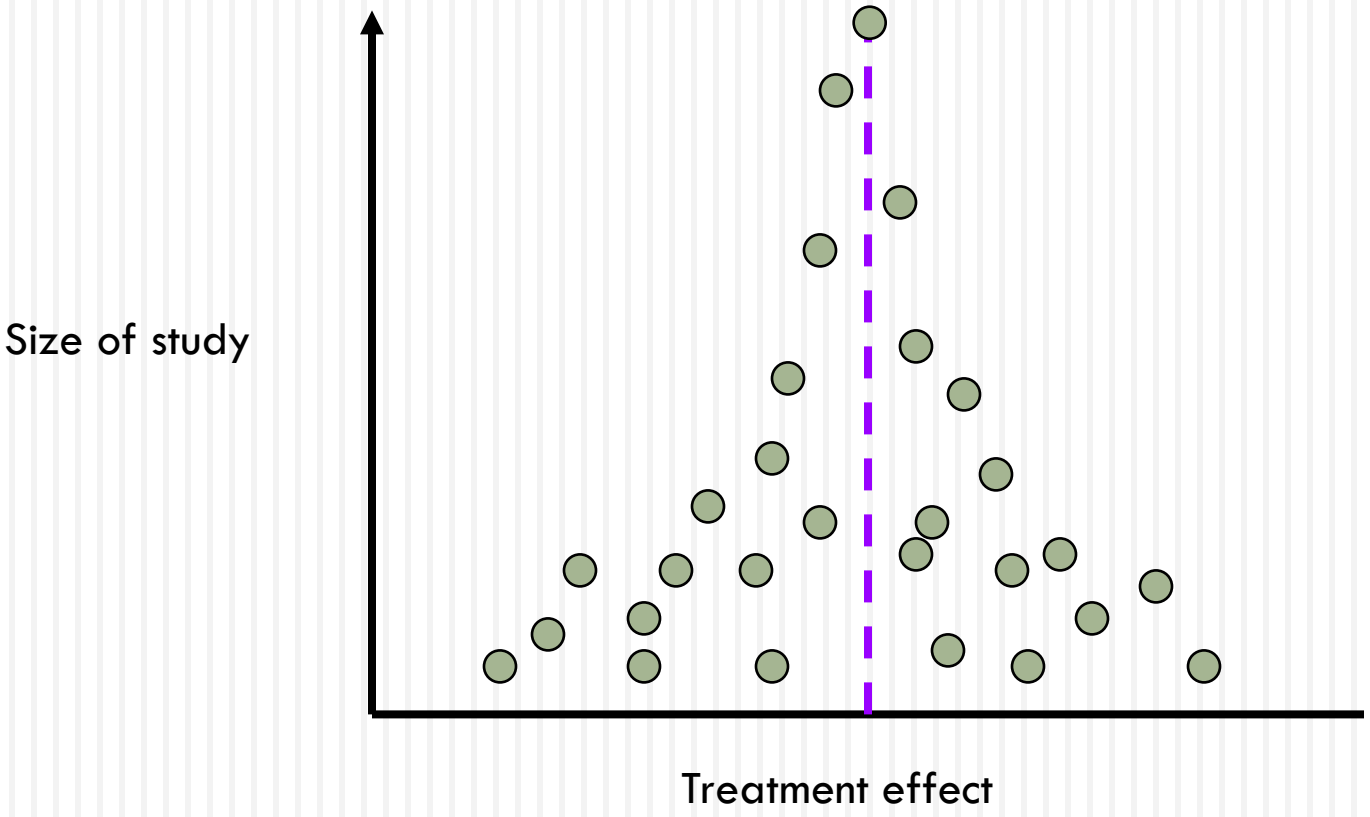



IN 1765 ON THIS SPOT
NOTHING HAPPENED

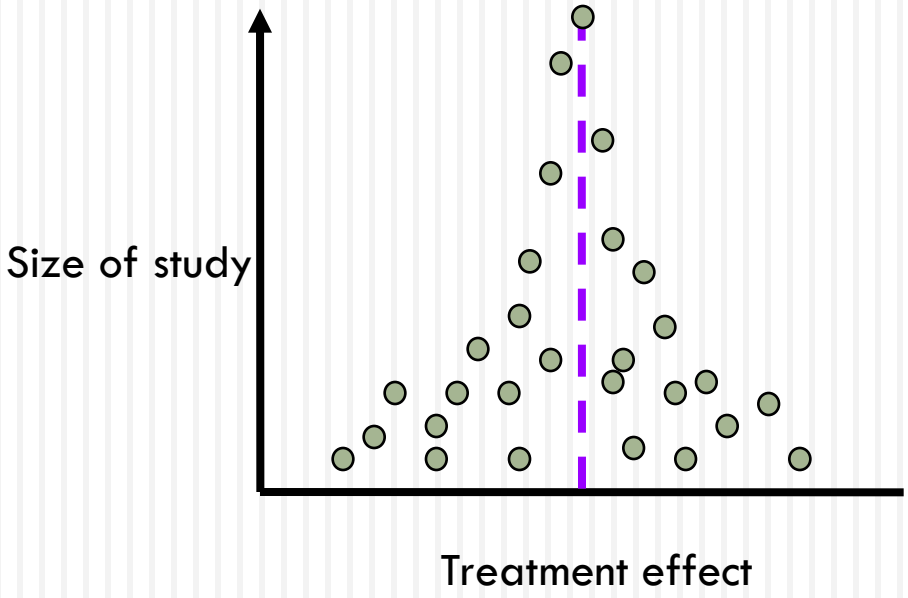
Uncertainty due to chance



A funnel plot

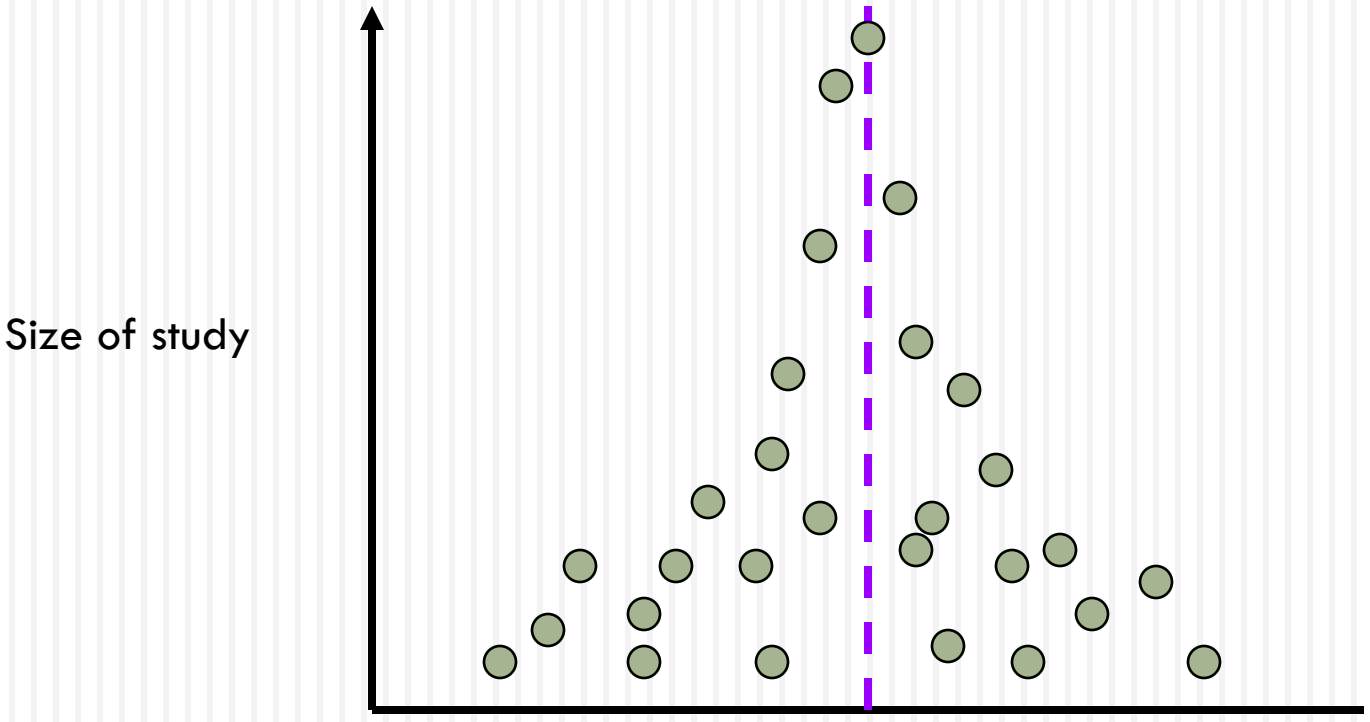


Funnel plots



- 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



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


ebhc
CONFERENCE
2015