Statistics for the MRCOG

- *Sangeeta Suri*
- *Consultant Obstetrician and Gynaecologist*

Aims and Objectives

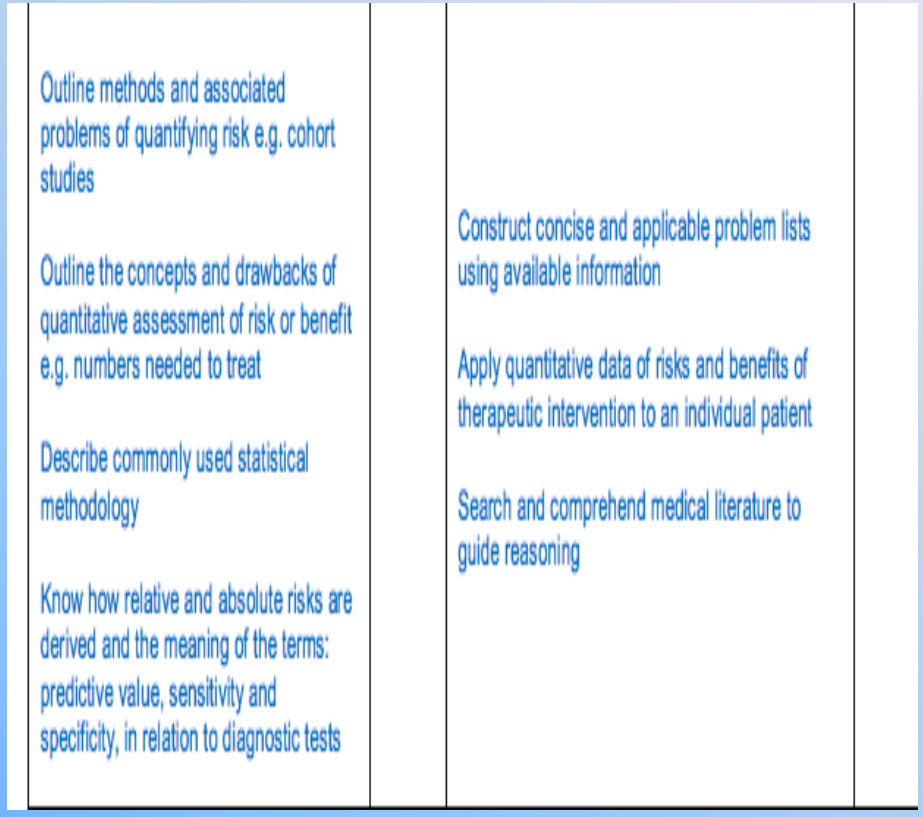
- To reintroduce the majority to statistics again
- To be able to define and explain fundamental statistical concepts
- To understand the basics of study design
- To be able to identify different types of data
- To be able to communicate statistical information to patients and colleagues using plain language
- **Most importantly – to know what you need to know to pass the exam.**

Declaration

- I am not a statistician.
- I cannot help you with statistics for a research project.
- I am an examiner for the membership exam and have written questions for MRCOG Part 2 and Part 3 (including statistics questions)

Are you expected to know statistics for MRCOG Part 2?

- Have an understanding of the statistics required for the **MRCOG Part 1** Exam
- **MRCOG Part 2** – Module 3 of curriculum
- **MRCOG Part 3** – Communicate knowledge
- Your Career – apply knowledge. EBM



Outline methods and associated problems of quantifying risk e.g. cohort studies Outline the concepts and drawbacks of quantitative assessment of risk or benefit e.g. numbers needed to treat Describe commonly used statistical methodology Know how relative and absolute risks are derived and the meaning of the terms: predictive value, sensitivity and specificity, in relation to diagnostic tests		Construct concise and applicable problem lists using available information Apply quantitative data of risks and benefits of therapeutic intervention to an individual patient Search and comprehend medical literature to guide reasoning	

Preterm Labour, Tocolytic Drugs (Green-top Guideline No. 1B)

4.1 Does tocolysis prevent preterm birth?

Use of a tocolytic drug is associated with a prolongation of pregnancy for up to 7 days but with no significant effect on preterm birth and no clear effect on perinatal or neonatal morbidity

There is no clear evidence that tocolytic drugs improve outcome and therefore it is reasonable not to use them. However, tocolysis should be considered if the few days gained would be put to good use, such as completing a course of corticosteroids or in utero transfer. **A**

A systematic review identified 17 trials (2800 women) comparing tocolysis with no treatment or placebo.[5] Many trials included maintenance treatment if and after contractions stopped. Some trials excluded women with ruptured membranes but they were included in others. The most frequently evaluated agent was ritodrine. Ritodrine has predominantly beta 2-receptor effects, relaxing muscles in the uterus, arterioles and bronchi. Other tocolytic drugs evaluated in these trials included isoxuprine, terbutaline, magnesium sulphate, indomethacin and atosiban. Overall, tocolytics were associated with a reduction in the odds of birth within 24 hours (odds ratio [OR] 0.47; 95% confidence interval [CI] 0.29–0.77), 48 hours (OR 0.57; 95% CI 0.38–0.83) and 7 days (OR 0.60; 95% CI 0.38–0.95). For the beta-agonists indomethacin and atosiban these effects were statistically significant, but not for magnesium sulphate. However, use of any tocolytic drug was not associated with a statistically significant reduction in births before 30 weeks of gestation (OR 1.33; 95% CI 0.53–3.33), before 32 weeks of gestation (OR 0.81; 95% CI 0.61–1.07) or before 37 weeks of gestation (OR 0.17; 95% CI 0.02–1.62).

Evidence level 1+

Preterm Labour, Tocolytic Drugs (Green-top Guideline No. 1B)

tocolytic drug was not associated with a statistically significant reduction in births before 30 weeks of gestation (OR 1.33; 95% CI 0.53–3.33) before 32 weeks of gestation (OR 0.81; 95% CI

Since this review, three further placebo-controlled trials have been reported. The largest compared atosiban with placebo (531 women).[6] Data from this study are consistent with the results of the systematic review above as, although time to delivery was not reported for all women (it was reported only for the subset of women who did not have an alternative tocolytic drug), there was no clear effect on birth before 37 weeks of gestation (relative risk [RR] 1.17; 95% CI 0.99–1.37) or before 28 weeks of gestation (RR 2.25; 95% CI 0.80–6.35).[6,7] The second study recruited 158 women and compared glyceryl trinitrate skin patches with placebo patches.[8] There was no clear difference in birth within 48 hours (RR 0.92, 95% CI 0.53–1.58) or before 37 weeks of gestation (RR 1.01; 95% CI 0.73–1.40). The third study compared glyceryl trinitrate with placebo (33 women) but was too small for any firm conclusions about the possible benefits or hazards of glyceryl trinitrate to be drawn.[9]

Evidence level 1+

This review restricted inclusion to studies in which the mean gestation at randomisation was between 28 weeks and 32 weeks of gestation but the methodology used did not allow

Evidence level 1+

Back to Basics

- **Parametric Test**

-Normal distribution of data

-Data continuous

-Independence of data (one group does not influence another)

-Homogeneity (variances between groups similar)

-Considered more powerful tests

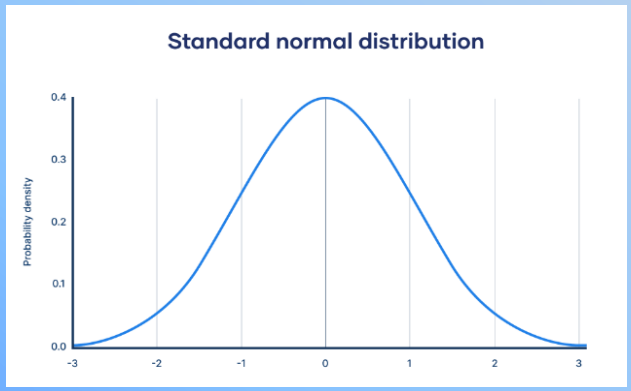
- **Non-Parametric Test**

-No assumptions wrt distribution

-Data –continuous or ordinal

-Can be tranformed to parametric data

-Independent



Examples

Parametric test	Examples of equivalent non parametric tests	Purpose of test
Two sample (unpaired) *t* test	Mann-Whitney U test	Compares two independent samples drawn from the same population
One sample (paired) *t* test	Wilcoxon matched pairs test	Compares two sets of observations on a single sample
One way analysis of variance (*F* test) using total sum of squares	Kruskall-Wallis analysis of variance by ranks	Compare three or more sets of observations on a single sample
X^2 test	Fishers's exact test	Tests the null hypothesis that the distribution of a discontinuous variable is the same in two (or more) independent samples
Product moment correlation coefficient (Pearson's *r*)	Spearman's rank correlation coefficient (rσ)	Assesses the strength of the straight line association between two continuous variables
Multiple regression by least squares method	Non parametric regression (various test)	Describes the numerical relation between a dependent variable and several predictor variables (covariates)

EMQ

A: Chi squared

B: Two sample (unpaired) t test

C: Pearson's correlation test

D: One sample (paired) t test

E: Multiple regression

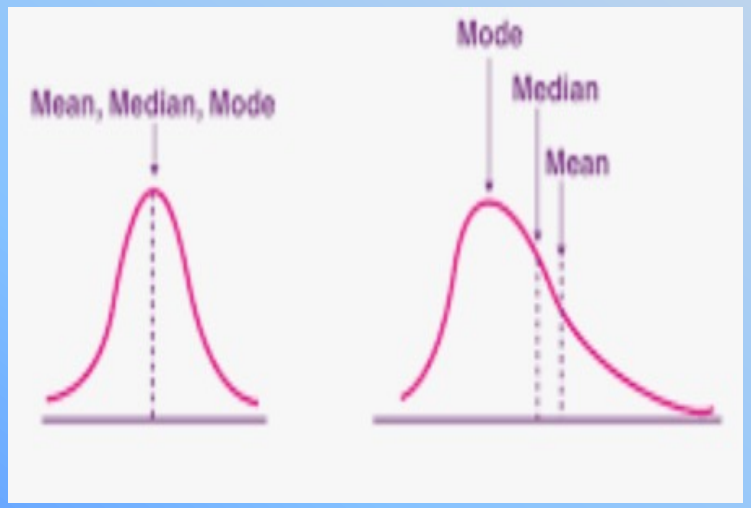
F: Analysis of variance

1. To compare the mean time of delivery between ventouse and forceps delivery.
2. To compare the weight loss in pregnant women before and after attending the healthy eating programme.
3. To compare the decision-delivery time interval in second stage for forceps, ventouse and caesarean section
4. To determine whether forceps delivery increases chances of having pelvic floor surgery later in life.
5. To assess whether HbA1c level is related to the birth weight in diabetic mothers.
6. To determine whether age, parity, smoking affects the birth weight.

B D F A C E

Mean, Mode Median

- 3, 12 ,3, 8, 9, 7, 3, 14, 9, 6, 3
- Calculate the mean mode and median
- Mean =7, Mode=3, Median =7



How to calculate Variance, Standard Deviation and Standard Error of the mean

Variance = Ave of squared differences from the mean

ie $(x_1\text{-mean})^2+(x_2\text{-mean})^2+(x_3\text{-mean})^2+(x_4\text{-mean})^2/n-1$= variance where n=4

Calculate the variance? = 14.8

Standard Deviation, represents the spread of the population

= √ Variance = 3.85

Standard Error of the mean, represents how well the sample mean approximates the pop mean. Larger the sample, the smaller the standard error, and the closer the sample mean approximates the population mean = SD/√n

Odds ratio(OR)

- OR represents the odds that a diseased group were exposed, compared to the odds of non-diseased group (controls) being exposed.

- OR=1 No difference in the odds of exposure between the two groups
- OR>1 Diseased group more likely to have been exposed compared to controls
- OR<1 Diseased group less likely to have been exposed compared to controls

Calculating OR

	OUTCOME STATUS +VE	OUTCOME STATUS -VE	
EXPOSURE STATUS +VE	A Exposed patients outcome positive	B Exposed patients outcome negative	
EXPOSURE STATUS -VE	C Unexposed patients outcome positive	D Unexposed patients outcome negative	

A = Number of exposed cases
B = Number of exposed non-cases

$$OR=(A/C)/(B/D)=AD/BC$$

C = Number of unexposed cases
D = Number of unexposed non-cases

Question?

263 women underwent a psychiatric evaluation 3 weeks into their post-natal period. Of the 186 women who did not suffer with any form of post-natal depression, 86 had previously been treated for depression pre-pregnancy. Of the 77 women who were diagnosed with some form of post-natal depression, 45 had been treated for depression pre-pregnancy.

What is the OR of post-natal depression given a pre-pregnancy history of depression?

Calculating OR

	Post-natal depression +VE	Post-natal depression -VE	
Pre-pregnancy depression +VE	A Exposed patients outcome positive 45	B Exposed patients outcome negative 86	
Pre-pregnancy depression -VE	C Unexposed patients outcome positive 32	D Unexposed patients outcome negative 100	
	77	186	

OR=(A/C)/(B/D)=AD/BC

(45/32)/(86/100)=1.63

The women with post natal depression were 1.63 times more likely to have a previous history of pre-pregnancy depression

Confidence Intervals (CI)

- A confidence interval is an indicator of your measurement's precision.
- Small/narrow CI indicates that if same question asked again for different sample population then we are reasonably sure that results would be similar. 95% CI + 95% sure of similar result
- Large/broad CI means less sure of result – ?increase no. of people sampled to increase our confidence
- CI influenced by no.of people being assessed
- YOU WILL NOT BE ASKED TO CALCULATE THIS

Relative Risk (RR)

- Risk of a certain event happening in one group vs another.

- Commonly used in epidemiology and EBM, where RR helps identify the risk of developing a disease after an exposure (i.e. a drug/treatment or an environmental exposure) vs the risk of developing a disease in absence of the exposure

Relative Risk

- A 2x2 table is the basis for many epidemiological calculations.
- A = No of people who both had the exposure and developed the disease
- B = No of people who had the exposure but did not develop the disease
- C = No of people who did not have the exposure but did develop the disease
- D = No of people who neither had the exposure nor developed the disease

Relative Risk

	DISEASE +VE	DISEASE -VE	TOTAL NO
EXPOSURE +VE	A	B	-
EXPOSURE -VE	C	D	-
TOTAL NO	-	-	

Relative Risk

A study looks at 300 Women all of whom had mild dyskaryosis on scan 150 of whom were also positive for high risk HPV.

They are followed for the next 20 years to assess the risk of developing cervical cancer.

At the end of the study they found that 50 women who were HPV positive developed Cervical Ca as did 10 women who were HPV –ve.

What is the relative risk of developing Ca cervix if you are HPV+ve?

Relative Risk

	DISEASE +VE	DISEASE -VE	TOTAL NO
EXPOSURE +VE	50	100	150
EXPOSURE -VE	10	140	150
TOTAL NO	60	240	

RR=(A/A+B)/(C/C+D)= (50/150)/10/150)=
0.33333/0.06666667

RR=4.99999997=5

Sensitivity and Specificity

- *Sensitivity is the ability of a test to detect disease*
- *Specificity is the ability of a test to detect health*

- Tests with low sensitivity –waste of time/money
- With sensitivity FP may occur but a highly sensitive test won't miss the disease.
- Sensitivity=(TP/TP+FN) x100%

- Specificity highly specific test have low FP test ie healthy people won't be identified as sick
- Specificity=(TN/TN+FP) x100%

Sensitivity and Specificity

	DISEASE POSITIVE	DISEASE NEGATIVE	TOTAL
TEST POSITIVE	TP	FP	-
TEST NEGATIVE	FN	TN	-
TOTAL	-	-	

Sensitivity and Specificity

- Test is 90% sensitive in 100 people with disease
- Test is 80% specific in 100 people without disease
- *FILL IN THE BOX*

	DISEASE +VE	DISEASE -VE	
TEST +VE			
TEST -VE			

Sensitivity and Specificity

- *Question*
- If a test has 75% sensitivity and 80% specificity where 200 have the disease and 400 people are without the disease, what is the False negative rate?

Sensitivity and Specificity

	DISEASE POSITIVE	DISEASE NEGATIVE	TOTAL
TEST POSITIVE			-
TEST NEGATIVE			-
TOTAL	200	400	

Sensitivity = TP/TP+FNx100% → 0.75

=TP/200 =0.75x200=TP=150

200=TP+FN→200=150+FN→ FN=50

Positive and Negative Predictive Values

- ***Positive predictive value*** - The chance that if the test is postive the patient has the disease

(TP/TP+FP) x100%

When prevalence increases so does the PPV

- ***Negative Predictive value*** - when tested negative for the disease then they don't have the disease
- (TN/TN+FN) x100%
- When prevalence increases NPV decreases

PPV and NPV

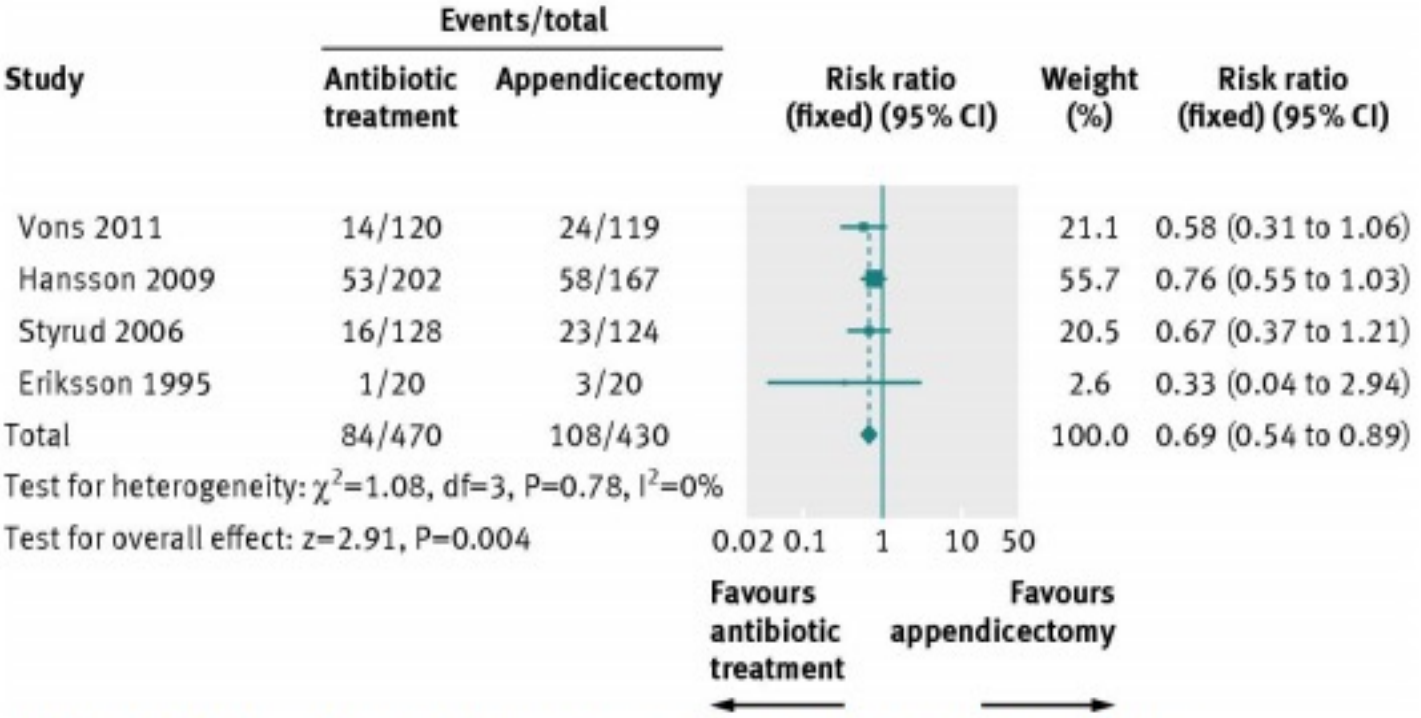
	DISEASE POSITIVE	DISEASE NEGATIVE	TOTAL
TEST POSITIVE	150	80	-
TEST NEGATIVE	50	320	-
TOTAL	200	400	

What are the PPV and NPV?

PPV =(TP/TP+FP)x100% =(150/150+80) x100% = 65.2%

NPV= (TN/TN+FN) x100% = (320/320+50)x100% = 86.5%

Forest Plot



Antibiotic treatment versus appendicectomy for uncomplicated acute appendicitis: forest plot for complications

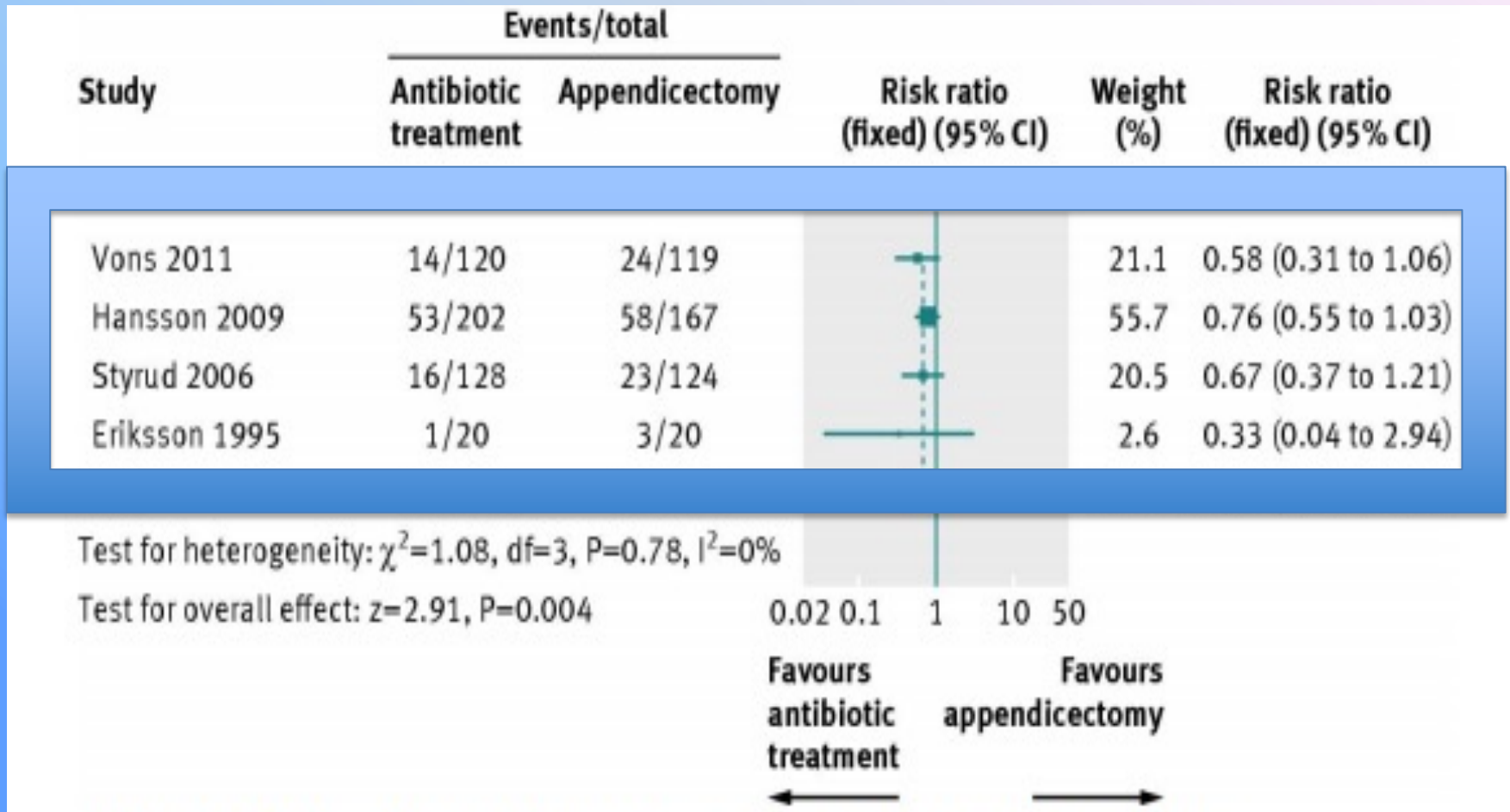
Which of the following statements, if any, are true?

a) Not one of the four trials showed a significant difference between antibiotic treatment and appendicectomy in the risk of complications.

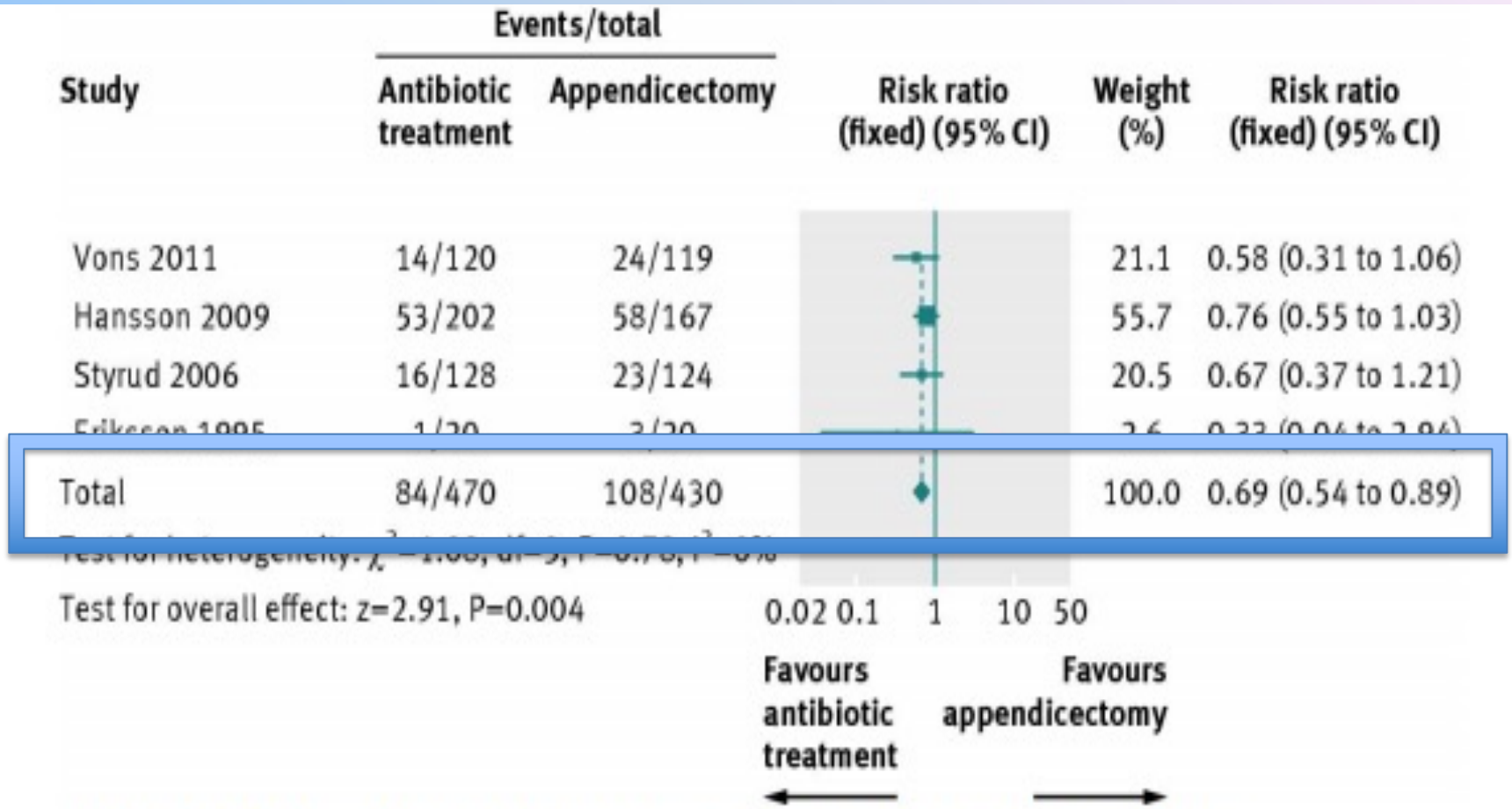
b) The forest plot is drawn on a linear scale.

c) A relative risk less than 1.0 represents a reduced risk of complications for antibiotic treatment compared with Appendicectomy.

d) The meta-analysis of complications showed a relative risk reduction of 31% for antibiotic treatment compared with appendicectomy.



Antibiotic treatment versus appendicectomy for uncomplicated acute appendicitis: forest plot for complications

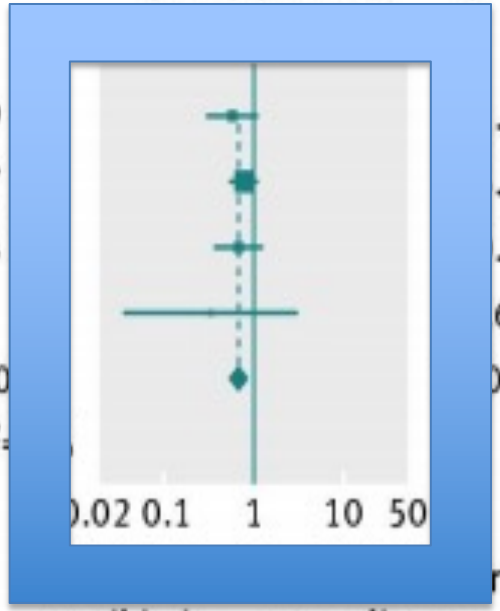


Antibiotic treatment versus appendicectomy for uncomplicated acute appendicitis: forest plot for complications

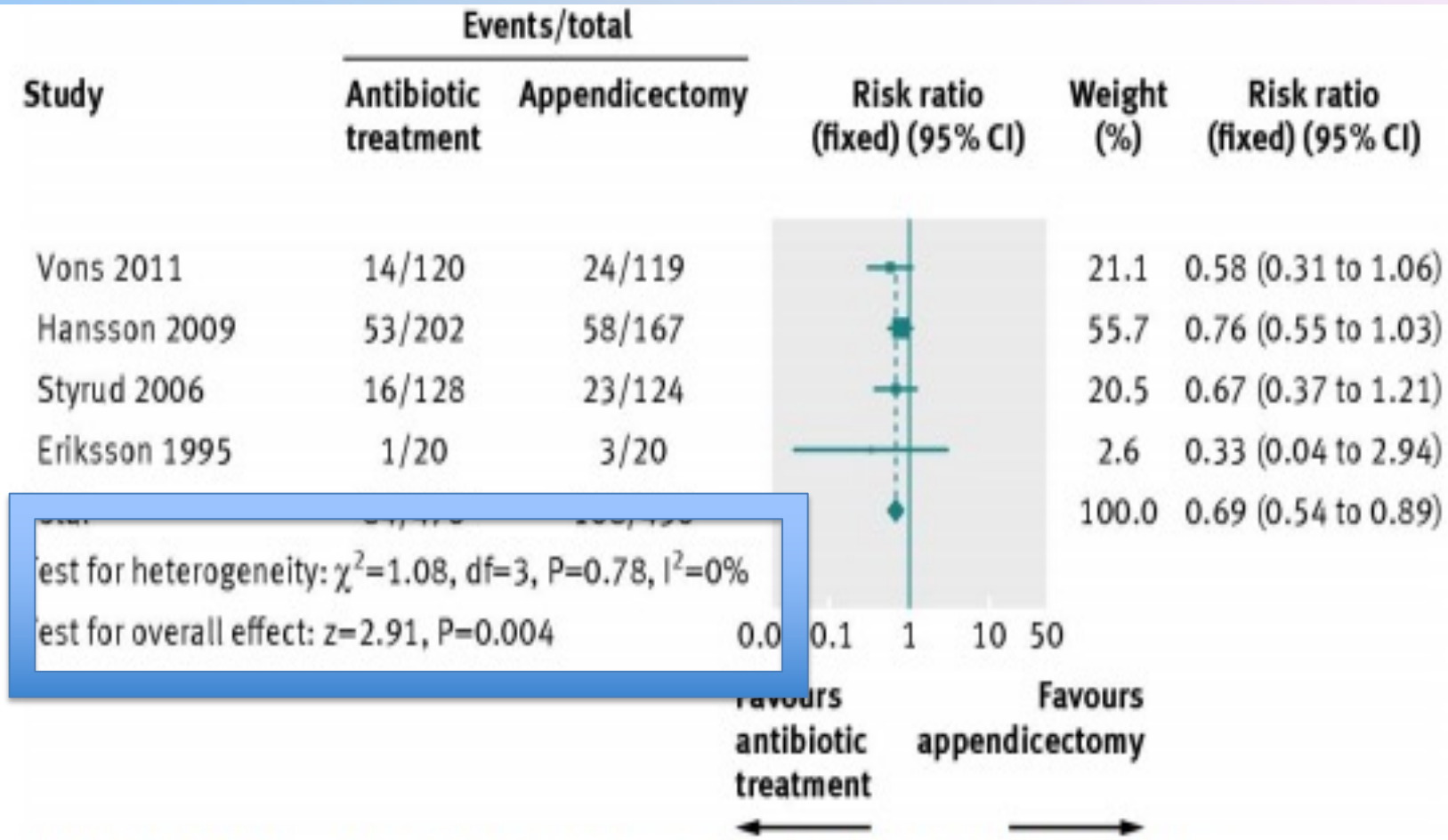
Study	Events/total: Antibiotic treatment	Events/total: Appendicectomy	Risk ratio (fixed) (95% CI)	Weight (%)	Risk ratio (fixed) (95% CI)
Vons 2011	14/120	24/119		.1	0.58 (0.31 to 1.06)
Hansson 2009	53/202	58/167		.7	0.76 (0.55 to 1.03)
Styrud 2006	16/128	23/124		.5	0.67 (0.37 to 1.21)
Eriksson 1995	1/20	3/20		6	0.33 (0.04 to 2.94)
Total	84/470	108/430		.0	0.69 (0.54 to 0.89)

Test for heterogeneity: χ^2=1.08, df=3, P=0.78, I^2=

Test for overall effect: z=2.91, P=0.004



Antibiotic treatment versus appendicectomy for uncomplicated acute appendicitis: forest plot for complications



Antibiotic treatment versus appendicectomy for uncomplicated acute appendicitis: forest plot for complications

Glossary

Meta-analysis: multiple independent studies (preferably RCT's) – same topic. Provides a more comprehensive and robust estimate of overall effect.

Randomised Control Trial: Subjects randomly allocated to design/control group (double blind) - to minimize bias. **Gold standard**

Case-Controlled study: Retrospective. Compares subjects with condition (cases) to those without (controls) to identify contributing factors which caused development of condition.

Cohort Study: Longitudinal research design. Follows individuals over time to investigate relationships between certain exposures/characteristics and development of a specific outcome

Bias: Errors/deviations from true results due to flaws in study design, data collection or analysis – leads to incorrect or misleading conclusions eg not randomly allocating to groups can lead to a characteristic being over represented in one group. Can affect validity of results

Classification of evidence levels

1++ High-quality meta-analyses, systematic reviews of randomised controlled trials or randomised controlled trials with a very low risk of bias

1+ Well-conducted meta-analyses, systematic reviews of randomised controlled trials or randomised controlled trials with a low risk of bias

1– Meta-analyses, systematic reviews of randomised controlled trials or randomised controlled trials with a high risk of bias

2++ High-quality systematic reviews of case–control or cohort studies or high-quality case–control or cohort studies with a very low risk of confounding, bias or chance and a high probability that the relationship is causal

2+ Well-conducted case–control or cohort studies with a low risk of confounding, bias or chance and a moderate probability that the relationship is causal

2– Case–control or cohort studies with a high risk of confounding, bias or chance and a significant risk that the relationship is not causal

3 Non-analytical studies; e.g. case reports, case series

4 Expert opinion

Grades of recommendations

A At least one meta-analysis, systematic reviews or randomised controlled trial rated as 1++ and directly applicable to the target population; or
A systematic review of randomised controlled trials or a body of evidence consisting principally of studies rated as 1+, directly applicable to the target population and demonstrating overall consistency of results

B A body of evidence including studies rated as 2++ directly applicable to the target population and demonstrating overall consistency of results; or
Extrapolated evidence from studies rated as 1++ or 1+

C A body of evidence including studies rated as 2+ directly applicable to the target population and demonstrating overall consistency of results; or
Extrapolated evidence from studies rated as 2++

D Evidence level 3 or 4; or
Extrapolated evidence from studies rated as 2+

Good practice point

✓ Recommended best practice based on the clinical experience of the guideline development group

Any Questions?