

ENDGAMES

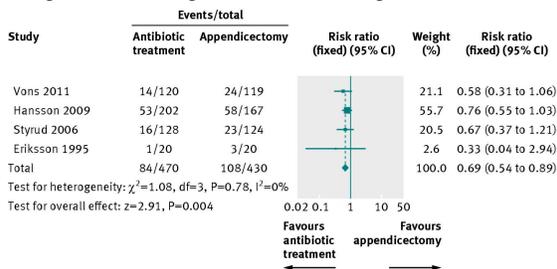
STATISTICAL QUESTION

How to read a forest plot

Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, Tooting, London, UK

Researchers undertook a meta-analysis of the safety and efficacy of antibiotic treatment compared with appendicectomy for the primary treatment of uncomplicated acute appendicitis. Randomised controlled trials were included if they investigated adult patients presenting with uncomplicated acute appendicitis, diagnosed by haematological and radiological investigations. The primary outcome measure was the presence of complications, including wound infection, perforated appendicitis, or peritonitis. Four randomised controlled trials were identified.¹ The results of the meta-analysis for complications were presented in a 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.
- e) No significant heterogeneity existed between the sample estimates of the population relative risk.

Answers

Statements *a*, *c*, *d*, and *e* are all true, whereas *b* is false.

Four randomised controlled trials were identified that compared antibiotic treatment with appendicectomy for the treatment of uncomplicated acute appendicitis. The primary outcome in each trial was complications. For each trial a relative risk was derived that compared the risk of complications for antibiotic treatment relative to appendicectomy. Relative risk has been described in previous questions.^{2,3} Each sample relative risk was an estimate of the population parameter—that is, the relative risk that would be observed if antibiotic treatment was compared with appendicectomy for the entire population of all adults with uncomplicated acute appendicitis. The purpose of the meta-analysis was to combine the results of the four trials and achieve a single estimate of the population relative risk of complications for antibiotic treatment compared with appendicectomy.

The forest plot presents the results of the meta-analysis graphically. The trials are identified by their principal author on the left side. For each treatment group in the four trials, the number of participants who experienced a complication and the total number in each group are shown in the column headed “Events/total.” These data were used to calculate the estimated relative risk and 95% confidence interval, shown on the right for each trial and represented graphically in the centre. Because the data for “Events/total” for antibiotic treatment are presented first, followed by those for appendicectomy, the presented relative risks therefore represent the risk of complications for antibiotic treatment relative to appendicectomy. The sample relative risk is represented by a square and its associated 95% confidence interval by the horizontal line. The size of each square is proportional to the sample size of the trial.

For all four trials the 95% confidence interval for the population relative risk included 1.0, and therefore not one of them demonstrated a significant difference between treatment groups in the risk of complications (*a* is true). The relation between the 95% confidence interval for a relative risk and 5% level of significance when hypothesis testing has been described in a previous question.⁴

The graphical representation of the sample relative risks and associated 95% confidence intervals are plotted on a logarithmic scale (b is false). As the 95% confidence intervals were originally calculated on a logarithmic scale, they therefore appear symmetrical about the sample relative risk in the forest plot. The solid vertical line in the centre of the graph is the “line of no effect”—that is, a relative risk of 1.0, which represents no difference in risk between antibiotic treatment and appendicectomy. A relative risk smaller than 1.0 would imply that the risk of complications was reduced for antibiotic treatment, relative to appendicectomy (c is true), whereas the risk would be increased if the relative risk was larger than 1.0. Therefore, as indicated on the forest plot, a relative risk less than unity “favours antibiotic treatment,” whereas one greater than unity “favours appendicectomy.”

A total overall estimate of the population relative risk was obtained by pooling the relative risks from the four trials. However, the trials did not contribute equally to the pooled result, as the total estimate was not an average of the individual estimates. The contribution of each trial is indicated under the heading “Weight (%)” The percentage weight contributed by a trial is determined by the precision of its sample estimate of the population parameter, and trials with more precise estimates—those with narrower confidence intervals—contributed more.

The total overall estimate of the population relative risk is presented in the row labelled “Total” and is given as 0.69 (95% confidence interval 0.54 to 0.89). It is graphically represented by a diamond; the centre of the diamond equals the total overall relative risk, whereas the extreme points indicate the limits of the 95% confidence interval. The vertical dotted line through the centre of the diamond and graph represents the overall estimated relative risk. Therefore, the meta-analysis of complications showed a relative risk reduction of 31% for antibiotic treatment compared with appendicectomy (d is true). As the 95% confidence interval did not include 1.0, the total overall estimate was therefore significant at the 5% level of significance—that is, there was a significant difference between treatments in the risk of complications. The P value for the test of significance of the total overall estimate was 0.004, as shown in the text “Test for overall effect: $z=2.91$, $P=0.004$,” corroborating the inference from the 95% confidence interval. The value for z is the test statistic resulting from the statistical test used to derive the P value.

A meta-analysis must incorporate a statistical test of heterogeneity to assess the extent of variation between the sample estimates. Statistical tests of heterogeneity have been described in a previous question.⁵ Statistical homogeneity would have existed in the above example if the sample relative risks were similar in magnitude and the variation between them was no more than expected when taking samples from the same population—that is, any variation between them was minimal. If statistical homogeneity did not exist, then statistical heterogeneity would be present, and the sample estimates would differ substantially. The result of the statistical test of heterogeneity influences how the total overall estimate would have been obtained. Furthermore, the presence of heterogeneity might suggest that the relative risk of complications differed between subgroups in the population.

The results of the statistical test of heterogeneity are shown in the text “Test for heterogeneity: $\chi^2=1.08$, $df=3$, $P=0.78$, $I^2=0\%$.” The test is performed in a similar way to traditional statistical hypothesis testing, there being a null and alternative hypothesis. Simply, the null hypothesis indicates that homogeneity exists, while the alternative hypothesis states that heterogeneity is present. The P value for the test of heterogeneity was 0.78, indicating that there was no evidence to reject the null hypothesis in favour of the alternative. Therefore, the conclusion was that homogeneity existed between the sample estimates (e is true). The value for χ^2 is the test statistic resulting from the statistical test used to derive the P value. The value for degrees of freedom (“df”) equals the number of trials minus one and is used along with the test statistic to calculate the P value.

Higgins I^2 statistic is often also used to test for heterogeneity. This statistic represents the percentage of variation between the sample estimates that is due to heterogeneity. It can take values from 0% to 100%, with 0% indicating that statistical homogeneity exists between the sample estimates. Significant statistical heterogeneity is often considered to be present if I^2 is $\geq 50\%$. The value of I^2 in the above example is 0%, corroborating the inference of the statistical test that statistical homogeneity existed (e is true).

Because of the presence of statistical homogeneity between the sample estimates, a so called fixed effects meta-analysis was performed; this is indicated in the headings of the forest plot. If statistical heterogeneity had existed (that is, if statistical homogeneity had not existed), a random effects meta-analysis would have been undertaken. The difference between these approaches is the method used to calculate the total overall effect. A random effects meta-analysis produces a wider confidence interval for the total overall effect than a fixed effects meta-analysis, resulting in a less accurate total overall effect size.

Not one of the four trials demonstrated a significant difference between treatments in the risk of complications for patients with uncomplicated acute appendicitis. However, the total overall estimate from the meta-analysis confirmed a significantly reduced risk of complications for antibiotic treatment compared with appendicectomy. This is one of the advantages of a meta-analysis. By combining results from across studies, it aggregates the numbers of study participants, thereby providing a total overall estimate that has increased power and precision. This means that a significant effect may be seen overall that was not seen in any of the individual studies.

The researchers concluded that antibiotics were effective and safe and merited consideration as primary treatment for patients with uncomplicated acute appendicitis.

Competing interests: None declared.

- 1 Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. *BMJ* 2012;344:e2156.
- 2 Sedgwick P, Marston L. Relative risks. *BMJ* 2010;341:c3983.
- 3 Sedgwick P. Absolute and relative risks. *BMJ* 2012;345:e5613.
- 4 Sedgwick P. Confidence intervals and statistical significance: rules of thumb. *BMJ* 2012;345:e4960.
- 5 Sedgwick P. Meta-analyses: tests of heterogeneity. *BMJ* 2012;344:e3971.

Cite this as: *BMJ* 2012;345:e8335

© BMJ Publishing Group Ltd 2012