

ENDGAMES



STATISTICAL QUESTION

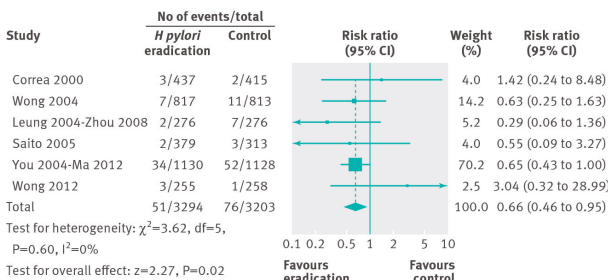
How to read a forest plot in a meta-analysis

Philip Sedgwick *reader in medical statistics and medical education*

Institute for Medical and Biomedical Education, St George's, University of London, London, UK

Researchers undertook a meta-analysis of the effect of treating *Helicobacter pylori* with eradication therapy on the subsequent occurrence of gastric cancer. Randomised controlled trials were included if the intervention consisted of seven days or more of eradication therapy, and if the control treatment was placebo or no treatment. Participants were adults who tested positive for *H pylori*. They were otherwise healthy and asymptomatic at baseline and were followed for two or more years. The primary outcome was the diagnosis of gastric cancer.¹

Six randomised controlled trials were eligible for inclusion. The results of the meta-analysis comparing the intervention with control treatment in the occurrence of gastric cancer were presented in a forest plot (figure). Fifty one (1.6%) gastric cancers occurred in 3294 participants who received eradication therapy compared with 76 (2.4%) in 3203 control subjects (relative risk 0.66, 95% confidence interval 0.46 to 0.95).



Forest plot of randomised controlled trials of *Helicobacter pylori* eradication therapy: effect on subsequent occurrence of gastric cancer

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

- a) All six trials showed a significant difference between eradication therapy and control treatment in the risk of gastric cancer
- b) The forest plot is drawn on a logarithmic scale
- c) A risk ratio greater than 1.0 indicates an increased risk of gastric cancer with the control treatment compared with eradication therapy

d) The total overall estimate of the population risk ratio indicated that eradication therapy led to a 34% lower risk of gastric cancer compared with the control treatment

e) Significant heterogeneity existed between the sample estimates of the population risk ratio of gastric cancer

Answers

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

The aim of the meta-analysis was to combine the results of trials that investigated the effectiveness of eradication therapy for *H pylori* in preventing gastric cancer among healthy asymptomatic infected people. Six trials were identified. For each trial a risk ratio, typically referred to as a relative risk, was derived that compared the risk of gastric cancer for the intervention group relative to the risk for the control group. Relative risks have been described in a previous question.² The relative risk for each trial was an estimate of the population parameter—that is, the relative risk that would be seen if the intervention was compared with the control treatment for the entire population of adults who tested positive for *H pylori* and were otherwise healthy and asymptomatic. By combining the sample estimates to form a single estimate of the population relative risk, the meta-analysis reduced the evidence to a manageable quantity.

The forest plot is a graphic representation of the results of the meta-analysis. The six trials that were included are identified by their principal author and date of publication on the left side. For each treatment group in the six trials, the number of participants who developed gastric cancer and the total number allocated to each group are shown in the column headed “No of events/total.” These data were used to calculate the sample risk ratio and 95% confidence interval for each trial, shown on the right and also represented graphically in the centre. Because the data for “No of events/total” for the intervention (eradication therapy) are presented first, followed by those for the control treatment, the presented risks ratios represent the risk of gastric cancer for eradication therapy relative to the control treatment. For each trial the sample risk ratio is represented by a square and its associated 95% confidence interval by a horizontal line in the centre of the forest plot. The size of each square is proportional to the sample size of the trial.

For all six trials the 95% confidence interval for the population risk ratio included 1.0 and therefore none of them showed a significant difference between treatment groups in the risk of gastric cancer (*a* is false). The association between the 95% confidence interval for a relative risk and 5% level of significance when hypothesis testing has been described in a previous question.³

The sample risk ratios and associated 95% confidence intervals in the centre of the forest plot are plotted on a logarithmic scale (*b* is true). Therefore, because the 95% confidence intervals were originally calculated on a logarithmic scale they appear symmetrical about the sample risk ratios 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 (risk ratio) of 1.0, representing no difference in risk of gastric cancer between the intervention and control groups. A relative risk smaller than 1.0 implies that the risk of gastric cancer was reduced with the intervention relative to the control treatment, whereas a relative risk larger than 1.0 implies that the risk of gastric cancer was increased with the intervention (*c* is false). Therefore, as indicated on the forest plot, a relative risk less than unity (1.0) “favours eradication,” whereas one greater than unity “favours control.”

The total overall estimate of the population relative risk was obtained by pooling the sample estimates of the relative risk from the six trials. However, the six trials did not contribute equally to the pooled result—that is, the total estimate was not an average of the estimates across all trials. The amount that each trial contributed is indicated under the heading “Weight (%).” The percentage weight contributed by a trial was determined by the precision of its sample estimate for the population parameter. Trials with more precise estimates, as indicated by those with narrower confidence intervals, had a greater weight.

The total overall estimate of the population relative risk is presented in the row labelled “Total” (towards the bottom of the plot) and is equal to 0.66 (95% confidence interval 0.46 to 0.95). It is represented graphically by the diamond; the centre of the diamond equals the total overall estimated relative risk and the ends of the diamond indicate the limits of the 95% confidence interval. The vertical dotted line through the centre of the diamond represents the total overall estimated relative risk. The meta-analysis therefore showed an overall 34% reduction in the risk of gastric cancer for eradication therapy relative to control treatment (*d* is true). Because the 95% confidence interval for the total overall estimate did not include 1.0, the reduction in the risk of gastric cancer was significant at the 5% level of significance. The P value for the test of significance of the total overall estimate was 0.02, as shown in the text “Test for overall effect: $z=2.27$, $P=0.02$,” 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.

It was essential that the meta-analysis incorporated a statistical test of heterogeneity before the total overall estimate of the population relative risk was obtained. A test of heterogeneity, described in a previous question,⁴ assesses the extent of variation between the sample estimates in a meta-analysis. The test is performed in a similar way to traditional statistical hypothesis testing, there being a null and alternative hypothesis. The null hypothesis states that statistical homogeneity exists—that is, the sample relative risks are of a similar magnitude and the variation between them is no more than would be expected when taking samples from the same population—that is, any variation between them is minimal. The alternative hypothesis states that statistical heterogeneity is present, and the sample estimates

differ substantially. If statistical heterogeneity is present it influences how the total overall estimate is obtained. Furthermore, the presence of statistical heterogeneity might suggest that the effects of treatment differ between subgroups (such as ethnic groups) in the population. An example is described in a previous question.⁵

The results of the statistical test of heterogeneity for the above meta-analysis are shown in the text “Test for heterogeneity: $\chi^2=3.62$, $df=5$, $P=0.60$, $I^2=0\%$.” The P value for the test of heterogeneity was 0.60, indicating that there was no evidence to reject the null hypothesis in favour of the alternative. Therefore, homogeneity existed between the sample estimates (*e* is false). 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, simply referred to as I^2 , is often also used as an alternative test for heterogeneity.⁴ 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. Significant statistical heterogeneity is often considered to be present if I^2 is 50% or more. The value of I^2 in the above example is 0%, which corroborates the inference from the statistical test of the hypotheses that statistical homogeneity existed (*e* is false). The presence of statistical homogeneity between the sample estimates meant that a so called fixed effects meta-analysis was performed. 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 estimate. A random effects meta-analysis produces a wider confidence interval for the total overall estimate than a fixed effects meta-analysis, resulting in a less accurate total overall estimate. However, if a random effects meta-analysis had been performed despite there being statistical homogeneity, it would have produced similar results to a fixed effects meta-analysis.

None of the six trials showed a significant difference between eradication therapy and control treatment in the risk of gastric cancer in adults who tested positive for *H pylori* and were otherwise healthy and asymptomatic. However, the total overall estimate from the meta-analysis confirmed a significantly reduced risk of gastric cancer for eradication therapy compared with control treatment. 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. In the study above, the researchers concluded that eradicating *H pylori* reduced the incidence of gastric cancer in healthy asymptomatic infected adults.

Competing interests: None declared.

- 1 Ford AC, Forman D, Hunt RH, et al. Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. *BMJ* 2014;348:g3174.
- 2 Sedgwick P, Joekes K. What are the risks? *BMJ* 2015;350:h2931.
- 3 Sedgwick P. Confidence intervals and statistical significance: rules of thumb. *BMJ* 2012;345:e4960.
- 4 Sedgwick P. Meta-analyses: what is heterogeneity? *BMJ* 2015;350:h1435.
- 5 Sedgwick P. Meta-analyses: heterogeneity and subgroup analysis. *BMJ* 2013;346:f4040.

Cite this as: *BMJ* 2015;351:h4028

© BMJ Publishing Group Ltd 2015

ENDGAMES