

Are the results of the review valid?

What question (PICO) did the systematic review address?

What is best?

The main question being addressed should be clearly stated. The exposure, such as a therapy or diagnostic test, and the outcome(s) of interest will often be expressed in terms of a simple relationship.

Where do I find the information?

The Title, Abstract or final paragraph of the Introduction should clearly state the question. If you still cannot ascertain what the focused question is after reading these sections, search for another paper!

In this paper

Yes

No

Unclear

Comment:

F - Is it unlikely that important, relevant studies were missed?

What is best?

The starting point for a comprehensive search for all relevant studies is the major bibliographic databases (eg Medline, Cochrane, EMBASE, etc) but should also include a search of reference lists from relevant studies and contact with experts, particularly to inquire about unpublished studies. The search should not be limited to English language only. The search strategy should include both MESH terms and text words.

Where do I find the information?

The Methods section should describe the search strategy, including the terms used, in some detail. The Results section will outline the number of titles and abstracts reviewed, the number of full-text studies retrieved, and the number of studies

In this paper

Yes

No

Unclear

Comment:

A - Were the criteria used to select articles for inclusion appropriate?

What is best?

The inclusion or exclusion of studies in a systematic review should be clearly defined *a priori*. The eligibility criteria used should specify the patients, interventions or exposures and outcomes of interest. In many cases the type of study design will also be a key component of the eligibility criteria.

Where do I find the information?

The Methods section should describe in detail the inclusion and exclusion criteria. Normally, this will include the study design.

In this paper

Yes

No

Unclear

Comment:

A - Were the included studies sufficiently valid for the type of question asked?

What is best?

The article should describe how the quality of each study was assessed using predetermined quality criteria appropriate to the type of clinical question (e.g., randomization, blinding and completeness of follow-up)

Where do I find the information?

The Methods section should describe the assessment of quality and the criteria used. The Results section should provide information on the quality of the individual studies.

In this paper

Yes

No

Unclear

Comment:

T - Were the results similar from study to study?

What is best?

Ideally, the results of the different studies should be similar or homogeneous. If heterogeneity exists the authors may estimate whether the differences are significant (chi-square test). Possible reasons for the heterogeneity should be explored.

Where do I find the information?

The Results section should state whether the results are heterogeneous and discuss possible reasons. The forest plot should show the results of the chi-square test for heterogeneity and discuss reasons for heterogeneity, if present.

In this paper

Yes

No

Unclear

Comment:

What were the results?

A systematic review provides a summary of the data from the results of a number of individual studies. If the results of the individual studies are similar, a statistical method (called meta-analysis) is used to combine the results from the individual studies and an overall summary estimate is calculated. The meta-analysis gives weighted values to each of the individual studies according to their size. The individual results of the studies need to be expressed in a standard way, such as relative risk, odds ratio or mean difference between the groups. Results are traditionally displayed in a figure called a forest plot, like the one below.

Comparison: 03 Treatment versus Placebo

Outcome: 01 Effect of treatment on mortality

Study	Treatment n/N	Control n/N	OR (95%CI Fixed)	Weight %	OR (95%CI Fixed)
Brown 1998	24 / 472	35 / 499		9.6	0.71[0.42,1.21]
Geoffrey 1997	120 / 2850	182 / 2838		51.8	0.64[0.51,0.81]
Mason 1996	56 / 2051	84 / 2030		24.4	0.65[0.46,0.92]

The forest plot depicted above represents a meta-analysis of five trials that assessed the effects of a hypothetical treatment on mortality. Individual studies are represented by a black square and a horizontal line, which corresponds to the point estimate and 95% confidence interval of the odds ratio. The size of the black square reflects the weight of the study in the meta-analysis. The solid vertical line corresponds to 'no effect' of treatment - an odds ratio of 1.0. When the confidence interval includes 1 it indicates that the result is not significant at conventional levels ($P > 0.05$).

The diamond at the bottom represents the combined or pooled odds ratio of all five trials with its 95% confidence interval. In this case, it shows that the treatment reduces mortality by 34% (OR 0.66 95% CI 0.56 to 0.78). Notice that the diamond does not overlap with the 'no effect' line (the confidence interval doesn't include 1) so we can be assured that the pooled OR is statistically significant. The test for overall effect also indicates statistical significance ($p < 0.0001$).

Exploring heterogeneity

Heterogeneity can be assessed using the 'eyeball' test or more formally with statistical tests, such as the Cochran Q test. With the 'eyeball' test one looks for overlap of the confidence intervals of the trials with the summary estimate. In the example above note that the dotted line running vertically through the combined odds ratio crosses the horizontal lines of all the individual studies indicating that the studies are homogenous. Heterogeneity can also be assessed using the Cochran chi-square (Cochran Q). If Cochran Q is statistically significant there is definite heterogeneity. If Cochran Q is not statistically significant but the ratio of Cochran Q and the degrees of freedom (Q/df) is > 1 there is possible heterogeneity. If Cochran Q is not statistically significant and Q/df is < 1 then heterogeneity is very unlikely. In the example above Q/df is < 1 ($0.92/4 = 0.23$) and the p-value is not significant (0.92) indicating no heterogeneity.

Note: The level of significance for Cochran Q is often set at 0.1 due to the low power of the test to detect heterogeneity.