---

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

9. Were the methods used to combine the findings of studies appropriate?
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

---

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ Yes
☐ No
☐ Can't answer
☐ Not applicable

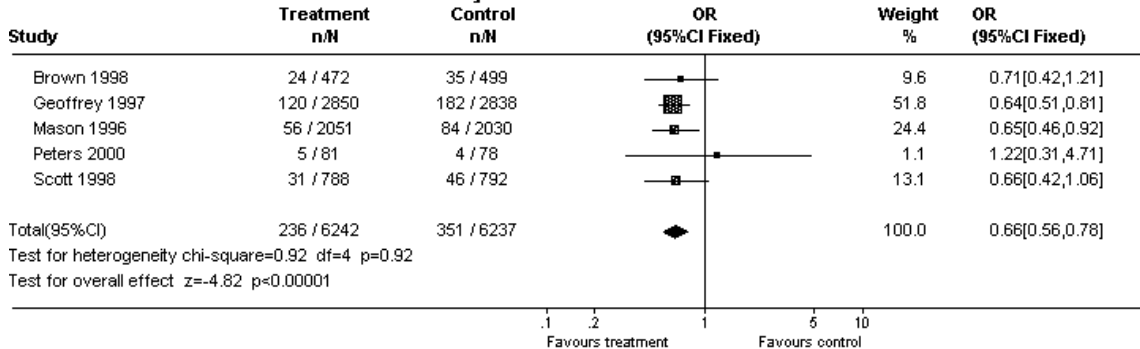| Clinical Importance	
A. What were the results of the review? *(Are the results of all included studies clearly displayed? Are the results similar from study to study? Is there a clinical bottom line? If the study results combined, was it appropriate to do so?)* B. How precise are the results? *(What is the confidence interval? p-value?)* C. Did the interpretation of the review's results accurately reflect the results themselves? Are the results generalizable?	

How are the results presented?

The 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 calculation. 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, like the one below, called a **forest plot**.

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]
Peters 2000	5 / 81	4 / 78		1.1	1.22[0.31,4.71]
Scott 1998	31 / 788	46 / 792		13.1	0.66[0.42,1.06]
Total(95%CI)	236 / 6242	351 / 6237		100.0	0.66[0.56,0.78]

Test for heterogeneity chi-square=0.92 df=4 p=0.92
Test for overall effect z=-4.82 p<0.00001

.1 .2   1   5   10
Favours treatment   Favours control

The forest plot depicted above represents a meta-analysis of 5 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 5 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 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.