

## Critical Appraisal Worksheet: Systematic Review/Meta-analysis

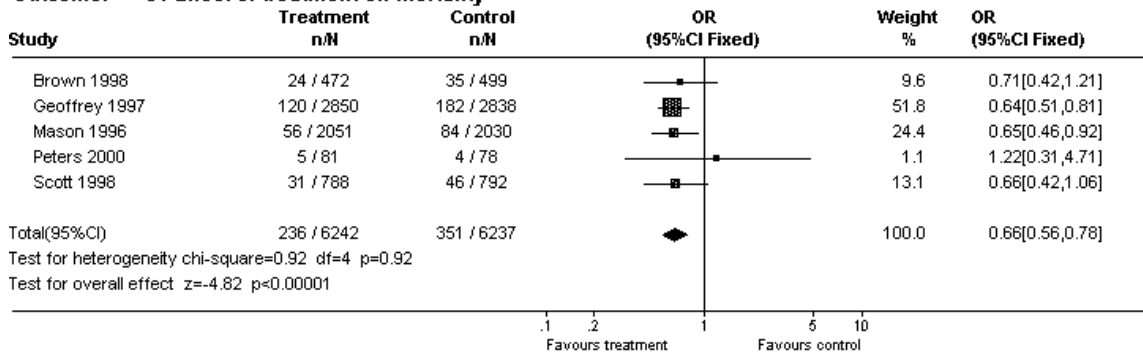
<b>1. What question did the systematic review address?</b>	
The main question being addressed should be clearly stated. The exposure, such as therapy or diagnostic test, and the outcome(s) of interest will often be expressed in terms of a simple relationship.	
<b>2. Is it likely that important, relevant studies were identified?</b>	
The starting point for comprehensive search for all relevant studies is the major bibliographic databases (e.g. MEDLINE, Cochrane, EMBASE, etc). It 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also include a search of reference lists from relevant studies, and contact with experts, particularly to inquire about unpublished studies.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 controlled vocabulary terms and text words.	
<b>3. Were the criteria used to select articles for inclusion predetermined, clearly stated and appropriate?</b>	
The inclusion or exclusion of studies in systematic reviews should be a 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.	
<b>4. Were the included studies sufficiently valid for the type of question asked?</b>	
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)	
<b>5. Were the results similar from study to study?</b>	
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.	
<b>6. Clinical Importance</b>	
<p>6a. What were the results of the review?  <i>(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?)</i></p> <p>6b. How precise are the results?  <i>(What is the confidence interval? p-value?)</i></p> <p>6c. Did the interpretation of the review's results accurately reflect the results themselves? Are the results generalizable?</p>	

## 7. 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**



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.

Adapted from:

- Clegg, Andrew, Hewitson, Paul, and Milne, Ruairidh. Explicit and reproducible: how to assess the quality of the evidence in a systematic review. **StudentBMJ**. 8:24. Feb. 2000
- Heneghan, Carl and Badenoch, Douglas. **Evidence-based Medicine Toolkit, 2<sup>nd</sup> edition**. Oxford: Blackwell Publishing; BMJ Books. 2006. p. 27.
- Systematic Review Appraisal Sheet. Centre for Evidence Based Medicine. University of Oxford [\[http://www.cebm.net/index.aspx?o=1567\]](http://www.cebm.net/index.aspx?o=1567)