Quantitative Critical Appraisal: Summary

The aim of Critical Appraisal:
To assess whether a reported piece of research is good enough to be used in decision making

Where does Critical Appraisal fit in?

Step: 1. An issue arises, e.g. is this the best treatment I can offer?
2. Formulate a focused question
3. Search for the best evidence available
4. Critically appraise the evidence
5. Apply the evidence to your patient / practice

Hierarchy of Evidence

Systematic Review
- see Critical Appraisal Toolkit 7

RCT
- see Critical Appraisal Toolkit 6

Cohort Study
- see Critical Appraisal Toolkit 2

Case-Control Study
- see Critical Appraisal Toolkit 1

Case Report

Searching the Literature - Did the reviewers try to identify all relevant studies?

Which databases were used?
Look for a good range: Medline, Cochrane Library, Embase, Cinahl etc.

What search terms were used?
Have subject headings as well as keywords been used? Are the subject headings and keywords used appropriate for the search?

Did any hand searching take place?
Were relevant articles in print copies of journal issues searched?

Is any foreign language research included?
Bias

Bias can take several forms, both when a trial or study is in progress and afterwards at publication:

**Sampling Bias:**
- Publication bias
- Indexing bias
- Search bias
- Reference bias
- Multiple publication bias
- Multiply-used subjects bias

**Selection Bias:**
- Inclusion criteria bias
- Selector bias
- Extractor bias
- Quality score bias
- Reporting bias
- Recording error bias

**Within Study Bias:**

**Other Biases:**
- Geographical Bias
- Follow-up time bias

The researcher only usually considers publication bias, although all should be avoided. See: Meta-analysis in Medical Research - Leandro for more details.

Unpublished studies
Has some attempt been made to find unpublished research by contacting researchers in the field?

Grey literature
Have conference proceedings, government documents, scientific papers been searched for relevant studies?

References
Have references and references of references in primary sources been checked for relevant research papers?

Qualitative Scales / Instruments

When reading a paper, it is important to evaluate the procedural and methodological aspects of the paper, as well as learning from the findings. To do this, a number of quality scores have been devised. In terms of the evaluating the studies to be included in the systematic review, the following criteria may be used:

1. Was the assignment to the treatment groups really random?
2. Was the treatment allocation concealed?
3. Were the groups similar at baseline in terms of prognostic factors?
4. Were the eligibility criteria specified?
5. Were outcome assessors blinded to the treatment allocation?
6. Was the care provider blinded?
7. Was the patient blinded?
8. Were the point estimates and measure of variability presented for the primary outcome measure?
9. Did the analyses include an intention to treat analysis?

In terms of the overall systematic review or meta-analysis, the following criteria may be used. The toolkit you’re using in this workshop asks similar questions:
To perform a meta-analysis we need to group only similar studies together so that we can be confident of the overall results. Due to randomisation, there will be some variation in results due to chance. However we need to know whether there is more variation than we’d expect by chance alone. This excess variation is called heterogeneity.

To detect it, look at the:

- Odds-ratio Diagram. If the confidence intervals do not widely overlap, suspect some heterogeneity.
- \( \chi^2 \) Chi-squared calculation: Test for heterogeneity \( \chi^2 = 12.62 \) df=4 \( p=0.01 \). Suspect heterogeneity if \( \chi^2 \) a lot greater than df (degrees of freedom) or p-value is less than 0.05. P value is more useful.
- \( I^2 \) = less than 25%

Heterogeneity may be caused by patients being at different stages of a disease or condition; dosage changes; length of follow up (treatment effects may vary over time) and co-interventions (may be taking something else that affects the intervention being tested). Subgroup analyses, meta-regression and use of random effects meta-analysis can help to limit heterogeneity. Are the reasons for heterogeneity explained? Has one or more of the methods used to limit heterogeneity been used?

**Sub-group analyses** are meta-analyses on subgroups of the studies.

**Meta-regression** is a test to see if there is evidence of different effects in different subgroups of trials. For example, you can use meta-regression to test whether treatment effects are bigger in low quality studies than in high quality studies.

**Random effects meta-analysis** makes the assumption that individual studies are estimating different treatment effects.

**Odds Ratio Diagrams**

Review: Group behaviour therapy programmes for smoking

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment n/N</th>
<th>Control n/N</th>
<th>Odds Ratio 95% CI</th>
<th>Weight %</th>
<th>Odds Ratio 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottraux 1983</td>
<td>10/138</td>
<td>9/140</td>
<td>31.4</td>
<td>1.14</td>
<td>[1.43, 102.61]</td>
</tr>
<tr>
<td>Grant 2003</td>
<td>0/20</td>
<td>1/20</td>
<td>5.5</td>
<td>0.32</td>
<td>[0.01, 8.26]</td>
</tr>
<tr>
<td>Hill 1993</td>
<td>7/22</td>
<td>2/20</td>
<td>5.4</td>
<td>4.20</td>
<td>[0.76, 23.32]</td>
</tr>
<tr>
<td>Leung 1991</td>
<td>9/32</td>
<td>1/32</td>
<td>2.7</td>
<td>12.13</td>
<td>[1.43, 102.61]</td>
</tr>
<tr>
<td>McDowell 1985</td>
<td>29/183</td>
<td>11/93</td>
<td>46.5</td>
<td>1.40</td>
<td>[0.67, 2.95]</td>
</tr>
<tr>
<td>Minthorn-Briggs 2000</td>
<td>19/50</td>
<td>2/25</td>
<td>6.3</td>
<td>7.05</td>
<td>[1.49, 33.33]</td>
</tr>
<tr>
<td>Pederson 1981</td>
<td>8/31</td>
<td>0/9</td>
<td>2.1</td>
<td>6.87</td>
<td>[0.36, 131.31]</td>
</tr>
<tr>
<td>Total 95% CI</td>
<td>476</td>
<td>339</td>
<td>100.0</td>
<td>2.17</td>
<td>[1.37, 3.45]</td>
</tr>
</tbody>
</table>

Total events 82 (Treatment), 26 (Control)

Test for heterogeneity \( \chi^2 = 10.36 \) df=6 \( p=0.11 \)

Test for overall effect \( z=3.28 \) \( p=0.001 \)

A meta-analysis diamond touching the line of no effect shows an inconclusive result.
Measures of Risk

Efficacy of hip protectors in reducing the incidence of hip fractures

<table>
<thead>
<tr>
<th>Incidence of hip fracture</th>
<th>Intervention Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>66/1872</td>
<td>90/3263</td>
</tr>
</tbody>
</table>

Absolute Risk (AR)  
\[ \text{AR} = \frac{66}{1872} = 0.035 \text{ or } 3.5\% \]

Relative Risk (AR)  
\[ \text{RR} = \frac{0.035}{0.027} = 1.29 \]

Odds Ratio (OR)  
\[ \text{OR} = \frac{66/(1872-66)}{90/(3263-90)} = 1.29 \]

If outcome is positive (e.g. pain relief), an RR/OR >1 means treatment better than control

Absolute Risk Reduction (ARR)  
\[ \text{ARR} = \frac{66}{1872} - \frac{90}{3263} = 0.08 \]

Relative Risk Reduction (RRR)  
\[ \text{RRR} = \left(1 - \frac{0.035}{0.027}\right) \times 100 = -29\% \]

Numbers Needed to Treat (NNT)  
\[ \text{NNT} = \frac{66}{1872} = 3.5\%, \frac{90}{3263} = 2.7\%, 3.5-2.7 = 0.8 \]

Number Needed to Treat (NNT)

NNT (Number Needed to Treat): the number of patients who need to be treated to prevent 1 adverse outcome.

Specifies the treatment, its duration, and the adverse outcome being prevented. Reported as a whole number, calculated as 1/ARR, rounded to the next highest whole number, and accompanied by its 95% confidence interval (CI):

\[ \frac{1}{\text{ARR}} = \frac{1}{6.8\%} = 14.7, \text{ rounded to } 15, \text{ with a } 95\% \text{ CI from } 9 \text{ to } 35. \]

Readers can convert this to an NNT for a specific patient by estimating that patient's susceptibility relative to the average control patient in the trial report, expressing it as a decimal fraction, F, and then dividing the reported NNT by F. If a reader's patient is judged to be half as susceptible as the average control patient in the example, F = 0.5 and NNT/F = 15/0.5 = 30.

Definition from EBM Glossary
(Taken from EBM Glossary - Evidence Based Medicine Volume 125 Number 1)

Bandolier: Number needed to Treat:
http://www.jr2.ox.ac.uk/bandolier/band59/NNT1.html#Heading11

Clinical vs. Statistical Significance

Statistical Significance is the degree to which a value is greater or smaller than would be expected by chance. Typically, a relationship is considered statistically significant when the probability of obtaining that result by chance is less than 5% if there were, in fact, no relationship in the population.

Clinical Significance, by contrast, is a measure of whether a research result "matters" in the real world. Sometimes changes observed in clinical studies may prove to be statistically significant, but they may not be clinically significant. For example, if one drug is slightly more effective than an older drug, the difference may be statistically significant yet trivial - and not worth the greater expense.

Clinical Significance Calculator:
http://www.healthcare.ubc.ca/calc/clinsig.html