Critical appraisal of research

Alan Fricker
Library Liaison Manager
Learning objectives

• Understand the principles of critical appraisal and its role in evidence based practice
• Understand the different levels of quantitative evidence
• Be able to appraise research and judge its validity
• Be able to assess the relevance of research to their own work
• Know about resources available to help them to critically appraise research
• Be able to critically appraise quantitative research as a group
What is evidence based practice?

Evidence-based practice is the integration of

• **individual clinical expertise**
  with the

• **best available external clinical evidence** from
  systematic research
  and

• **patient’s values and expectations**
The evidence-based practice (EBP) process.

- Decision or question arising from a patient’s care.
- Formulate a focused question.
- Search for the best evidence.
- Appraise the evidence.
- Apply the evidence.
Why does evidence from research fail to get into practice?

- 75% cannot understand the statistics
- 70% cannot critically appraise a research paper

Using research for Practice: a UK experience of the barriers scale. Dunn, V. et al.
What is critical appraisal?

- Weighing up evidence to see how useful it is in decision making
- Balanced assessment of benefits and strengths of research against its flaws and weaknesses
- Assess research process and results
- Skill that needs to be practiced by all health professionals as part of their work
Why do we need to critically appraise?

• “It usually comes as a surprise to students to learn that some (the purists would say 99% of) published articles belong in the bin and should not be used to inform practice”

Greenhalgh 2001
Why do we need to critically appraise?

• studies which don't report their methods fully overstate the benefits of treatments by around 25%

• studies funded by a pharmaceutical company were found to be 4 times as likely to give results that were favourable to the company than independent studies
  Lexchin et al, BMJ, 2003
Sources of bias

- poor control group / control dosage
- surrogate outcomes
- ignore drop outs
- modify trial length
- misuse baseline statistics
- statistics overload
How do I appraise?

• Mostly common sense.
• You don’t have to be a statistical expert!
• **Checklists** help you focus on the most important aspects of the article.
• Different checklists for different types of research.
• Will help you decide if research is valid and relevant.
Research methods

Quantitative
• Uses numbers to describe and analyse
• Useful for finding precise answers to defined questions

Qualitative
• Uses words to describe and analyse
• Useful for finding detailed information about people’s perceptions and attitudes
Levels of quantitative evidence.

- Systematic reviews
- Randomized controlled trials
- Prospective studies (cohort studies)
- Retrospective studies (case control)
- Case series and reports

- NB quality assessment!
Systematic reviews

• Thorough search of literature carried out.
• All RCTs (or other studies) on a similar subject synthesised and summarised.
• Meta-analysis to combine statistical findings of similar studies.
Randomised controlled trials (RCTs)

- Participants are randomised.
- If possible should be blinded.
- Intention to treat analysis
Cohort studies

- prospective
- groups (cohorts)
- exposure to a risk factor
- followed over a period of time
- compare rates of development of an outcome of interest
- Confounding factors and bias
Case-control studies

• Retrospective
• Subjects confirmed with a disease (cases) are compared with non-diseased subjects (controls) in relation to possible past exposure to a risk factor.
• Confounding factors and bias
Qualitative research

- Complements quantitative
- Natural settings
- No gold standard – grounded theory common
- Inductive and iterative
- Purposive – not statistical – sampling
- Triangulation – varied research methods
Appraising primary research

Are the results valid?
• Is the research question focused (and original)?
• Was the method appropriate?
• How was it conducted?

What are the results?
• How was data collected and analysed?
• Are they significant?

Will the results help my work with patients?
Appraising RCTs

• Recruitment and sample size
• Randomisation method and controls
• Confounding factors
• Blinding
• Follow-up (flow diagram)
• Intention to treat analysis
• Adjusting for multiple analyses
“As a non-statistician I tend only to look for 3 numbers in the methods:

• Size of sample
• Duration of follow-up
• Completeness of follow-up”

Greenhalgh 2010
Sample size calculation

Based on primary outcome measure. Depends upon various factors:

- Significance level (usually 5%)
- Power (minimum 80%)
- Variability of the measure (e.g. Standard Deviation)
- Level of clinically significant difference
Appraising cohort/case control studies

- Recruitment – selection bias
- Exposure - measurement, recall or classification bias
- Confounding factors & adjustment
- Time-frames
- Plausibility
Appraising qualitative research

• Sampling method and theoretical saturation
• Triangulation of methods
• Independent analysis
• Reflexivity
• Respondent validation
• Plausible interpretation
• Analysis should be done using explicit, systematic, justified and reproducible methods
Appraising systematic reviews

• Was a thorough literature search carried out?
• How was the quality of the studies assessed?
• If results were combined, was this appropriate?
Publication bias

papers with more ‘interesting’ results are more likely to be:

– Submitted for publication
– Accepted for publication
– Published in a major journal
– Published in the English language
Publication bias

• All SSRI trials registered with FDA
• 37 studies were assessed by FDA as positive - 36 of these were published.
• 22 studies with negative or inconclusive results were not published and 11 were written up as positive.

*Turner et al. NEJM, 2008.*
Results

• How was data collected?
• Which statistical analyses were used?
• How precise are the results?
• How are the results presented?
Statistical tests

Type of test used depends upon:
• Type of data – categorical, continuous etc
• One- or two-tailed (or sided) significance
• Independence and number of samples
• Number of observations and variables
• Distribution of data, e.g. normal
Estimates of effect

The observed relationship between an intervention and an outcome

Express results in terms of likely harms or benefits

Relative and absolute measures, e.g.:

- Odds ratios, absolute and relative risks/benefits,
- mean Difference,
- Number needed to treat (NNT)
Confidence intervals (CI)

• The range of values within which the “true” value in the population is found
• 95% CI – 95% confident the population lies within those limits
• Wide CI = less precise estimates of effect
CIs and statistical significance

• When quoted alongside an absolute difference a CI that includes zero is statistically non-significant.
• When quoted alongside a ratio or relative difference a CI that includes one is statistically non-significant.

• IF statistically significant:
• Less than zero (or one) = less of the outcome in the treatment group.
• More than zero (or one) = more of the outcome.
• BUT - Is the outcome good or bad?
Cardiac deaths - less = good
Smoking cessation – more = good

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment n/N</th>
<th>Control n/N</th>
<th>Peto Odds Ratio 95% CI</th>
<th>Weight (%)</th>
<th>Peto Odds Ratio 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottraux 1983</td>
<td>10 / 138</td>
<td>9 / 140</td>
<td></td>
<td>21.5</td>
<td>1.14 [0.45, 2.88]</td>
</tr>
<tr>
<td>Hill 1993</td>
<td>7 / 22</td>
<td>2 / 20</td>
<td></td>
<td>8.8</td>
<td>3.54 [0.82, 15.23]</td>
</tr>
<tr>
<td>McDowell 1985</td>
<td>29 / 183</td>
<td>11 / 93</td>
<td></td>
<td>37.2</td>
<td>1.38 [0.68, 2.80]</td>
</tr>
<tr>
<td>Minthorn-Biggs 2000</td>
<td>19 / 50</td>
<td>2 / 25</td>
<td></td>
<td>16.5</td>
<td>4.34 [1.50, 12.66]</td>
</tr>
<tr>
<td>Pederson 1981</td>
<td>8 / 31</td>
<td>0 / 9</td>
<td></td>
<td>5.6</td>
<td>4.82 [0.77, 30.10]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>82 / 456</td>
<td>25 / 319</td>
<td></td>
<td>100.0</td>
<td>2.19 [1.42, 3.37]</td>
</tr>
</tbody>
</table>

- Test for heterogeneity chi-square = 8.78 df = 6 p = 0.1183
- Test for overall effect = 3.55 p = 0.0004
P values

• P stands for probability - how likely is the result to have occurred by chance?
• P value of less than 0.05 means likelihood of results being due to chance is less than 1 in 20 = “statistically significant”.

• P values and confidence intervals should be consistent
• Confidence intervals provide a range of values
Statistical significance

Remember:

• statistical significance does not necessarily equal clinical significance
• statistical non-significance may be due to small sample size - check CI upper/lower limits – might a bigger sample result in a statistically significant difference?
Are results relevant?

- Can I apply these results to my own practice?
- Is my local setting significantly different?
- Are these findings applicable to my patients?
- Are findings specific/detailed enough to be applied?
- Were all outcomes considered?
The good news!

- Some resources have already been critically appraised for you.
- An increasing number of guidelines and summaries of appraised evidence are available on the internet.
Summary

- Search for resources that have already been appraised first, e.g. Guidelines, Cochrane systematic reviews.
- Search down through levels of evidence, e.g. systematic reviews, RCTs.
- Use checklists to appraise research.
- How can these results be put into practice?