Critical appraisal of the medical literature

Jane Hirst MBBS, MPH, PhD FRANZCOG
University of Oxford
What makes a study reliable?
Nine out of ten women we asked believe that “Youth Code” makes their skin firmer and younger looking

• Which design would you choose to maximise the chance of getting the result you want?

  a) Ask women buying “Youth Code” in the shops whether they agree their skin is firmer and younger looking?

  a) Ask a random sample of women to try “Youth Code” and then comment on whether they agree that their skin is firmer and younger looking?
What is critical appraisal?

- Carefully and systematically evaluate research to assess:
  - Validity (is these findings trustworthy?)
  - Value (what do the results show?)
  - Relevance (How do these results relate to my clinical practice?)
Critical appraisal: a key component of evidence based medicine

Clinical Problem → Define the question → Search for the evidence → Critical Appraisal → Decide what action to take → Evaluate your new practice
Asking the right question

- Population (P)
- Intervention (I)
- Comparator (C)
- Outcome (O)
<table>
<thead>
<tr>
<th>Patient or Problem</th>
<th>Intervention (a cause, prognostic factor, treatment, etc.)</th>
<th>Comparison Intervention (if necessary)</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tips for Building</td>
<td>Starting with your patient, ask “How would I describe a group of patients similar to mine?” Be specific.</td>
<td>Ask “What is the main alternative to the intervention?” Again, be specific.</td>
<td>Ask “What can I hope to accomplish?” or “What could this exposure really affect?” Again, be specific.</td>
</tr>
<tr>
<td>Example</td>
<td>“In patients with heart failure from dilated cardiomyopathy who are in sinus rhythm…”</td>
<td>“… would adding anticoagulation with warfarin to standard heart failure therapy…”</td>
<td>“… when compared with standard therapy alone…”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“… lead to lower mortality or morbidity from thromboembolism. Is this enough to be worth the increased risk of bleeding?”</td>
<td></td>
</tr>
</tbody>
</table>
Choosing right study design

- Some study designs are not appropriate to answer certain questions
- All study designs are prone to different biases
Pyramid of evidence

- Systematic reviews
- RCTs
- Cohorts
- Case controls
- Surveys
- Animal research
- Expert opinion
So are RCTs the gold standard for evidence?

.....depends

Slides from: K Mahtani, CEBM Oxford
Limitations of RCTs

• Excellent vs Poor RCTs – quality varies
  • Impact on interpretation of result (external validity)?
• Expensive and time consuming
  • £250k - £millions over 2-5 years+
• May not always be the right study design to answer that question
A RCT to examine if smoking causes lung cancer

- 30 healthy Oxford Students
- Randomise to 2 groups
  - Gp1 smokes 20 cigarettes per day every day
  - Gp2 no smoking
Types of research

• What is the best study design for answering this type of question?
  • Aetiology
  • Diagnosis
  • Prognosis
  • Harm
  • Effectiveness
  • Qualitative
How to critically appraise an article

• **Validity**: methods to check that the biases for which that particular study design is prone have been minimised

• Results

• Clinical relevance
Validity

Internal

External
Bias

“the systematic deviation of the results of a study from the truth because of the way it has been conducted, analysed or reported”

Burls, “What is Critical Appraisal” 2009
Internal validity

• No study is perfect and completely free from bias

• Have the researchers done all they can to minimise bias?

• Are the biases that remain unlikely to have affected the final results?
Sources of bias in clinical trials

<table>
<thead>
<tr>
<th>Source of Bias</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection bias</td>
<td>Biased allocation to comparison groups</td>
</tr>
<tr>
<td>Performance bias</td>
<td>Unequal provision of care apart from treatment under evaluation</td>
</tr>
<tr>
<td>Detection bias</td>
<td>Biased assessment of outcome</td>
</tr>
<tr>
<td>Attrition bias</td>
<td>Biased occurrence and handling of deviations from protocol and loss to follow up</td>
</tr>
</tbody>
</table>

Juni, BMJ 2001
Assessing Trials of effectiveness

Questions to ask:
1. Are the results of the trial valid?
2. What are the results?
3. Will the results help locally?

From: Critical Appraisal Skills Program, Oxford
www.casp-uk.net
Checklists for clinical trials

CONSORT: CONSolidated Standards Of Reporting Trials

Welcome to the CONSORT Website

CONSORT stands for Consolidated Standards of Reporting Trials and encompasses various initiatives developed by the CONSORT Group to alleviate the problems arising from inadequate reporting of randomized controlled trials.

CONSORT 2010 Key Documents
- CONSORT 2010 Checklist
- CONSORT 2010 Flow Diagram
- CONSORT 2010 Statement
- CONSORT 2010 Explanation and Elaboration Document

CEBM: Centre for Evidence-Based Medicine

Critical Appraisal Skills Programme (CASP)
Making sense of evidence
RAMMbo validity check

**Representative:** who did the subjects represent?

**Allocation:** randomised? Were groups similar at the start?

**Maintenance:** Were the groups treated equally? Were as many patients as possible followed-up?

**Measurements**
- blinded or objective

CEBM Oxford
Texting improves testing: a randomized trial of two-way SMS to increase postpartum prevention of mother-to-child transmission retention and infant HIV testing

Thomas A. Odeny\textsuperscript{a,b}, Elizabeth A. Bukusi\textsuperscript{a,c,d}, Craig R. Cohen\textsuperscript{e}, Krista Yuhas\textsuperscript{c}, Carol S. Camlin\textsuperscript{e} and R. Scott McClelland\textsuperscript{b,c,f,g}

Objective: Many sub-Saharan African countries report high postpartum loss to follow-up of mother–baby pairs. We aimed to determine whether interactive text messages improved rates of clinic attendance and early infant HIV testing in the Nyanza region of Kenya.

Design: Parallel-group, unblinded, randomized controlled trial.

Odeny, et al 2014 AIDS
RAMMbo validity check

**Representative:** who did the subjects represent?

**Allocation:** randomised? Were groups similar at the start?

**Maintenance:** Were the groups treated equally? Were as many patients as possible followed-up?

**Measurements** blinded or objective
Representative: Are the trial subjects representative of patients in this setting?

All people in study setting

Eligible participants

In trial
Are the trial subjects representative of HIV positive pregnant women in this setting?

Odeny, et al 2014 AIDS
RAMMbo validity check

**Representative:** who did the subjects represent?

**Allocation:** randomised? Were groups similar at the start?

**Maintenance:** Were the groups treated equally? Were as many patients as possible followed up?

**Measurements** blinded or objective
Why randomise?

• Minimises measured and unmeasured confounding
Minimising allocation bias

• Centralised computer randomisation the best

• Other methods such as sealed envelopes doubtful

• Non randomised: date of birth, alternate patients alternate days, etc
Allocation: How were participants randomised?

“A block randomization scheme with variable block sizes was used. Investigators and study staff were unaware of block numbers, sizes, or sequences. Intervention groups were assigned using sealed, opaque envelopes.”

Odeny, et al 2014 AIDS
4. Were the groups similar at the start of the trial?

☐ Yes  ☐ Can’t tell  ☐ No

Consider: Look at

- Other factors that might affect the outcome such as age, sex, social class, these may be called baseline characteristics
Allocation: were the groups similar at the start?

## Table 1. Maternal baseline characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>SMS group (n = 195)</th>
<th>Control group (n = 193)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Maternal age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–24</td>
<td>65 (33.7)</td>
<td></td>
</tr>
<tr>
<td>25–34</td>
<td>111 (57.5)</td>
<td></td>
</tr>
<tr>
<td>35+</td>
<td>17 (8.8)</td>
<td></td>
</tr>
<tr>
<td>Gestational age at enrolment – median weeks (IQR)</td>
<td>34 (22–36)</td>
<td>34 (32–36)</td>
</tr>
<tr>
<td>Shared phone</td>
<td>50 (25.9)</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>39 (20.2)</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>3 (1.6)</td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>110 (57.0)</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>55 (28.5)</td>
<td></td>
</tr>
<tr>
<td>Post-secondary</td>
<td>25 (13.0)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luo</td>
<td>177 (91.7)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>16 (8.3)</td>
<td></td>
</tr>
<tr>
<td>Married or with regular live-in partner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First pregnancy</td>
<td>110 (56.4)</td>
<td>171 (88.6)</td>
</tr>
<tr>
<td>WI/O stage (highest recorded)</td>
<td>(13.8)</td>
<td>(15.0)</td>
</tr>
<tr>
<td>1</td>
<td>110 (56.4)</td>
<td>103 (53.4)</td>
</tr>
<tr>
<td>2</td>
<td>53 (28.2)</td>
<td>57 (29.5)</td>
</tr>
<tr>
<td>3</td>
<td>23 (11.8)</td>
<td>27 (14.0)</td>
</tr>
<tr>
<td>4</td>
<td>7 (3.6)</td>
<td>6 (3.1)</td>
</tr>
<tr>
<td>Most recent CD4(^+) cell count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;200</td>
<td>22 (11.3)</td>
<td>18 (9.3)</td>
</tr>
<tr>
<td>200–349</td>
<td>40 (20.5)</td>
<td>38 (19.7)</td>
</tr>
<tr>
<td>350–500</td>
<td>54 (27.7)</td>
<td>55 (28.5)</td>
</tr>
<tr>
<td>500+</td>
<td>78 (40)</td>
<td>82 (42.5)</td>
</tr>
<tr>
<td>On ART for own health</td>
<td>101 (51.8)</td>
<td>102 (52.8)</td>
</tr>
<tr>
<td>Received ZDV prophylaxis</td>
<td>85 (43.6)</td>
<td>81 (42.0)</td>
</tr>
<tr>
<td>Received ZDV + 3TC + NVP (delivery pack)</td>
<td>60 (30.8)</td>
<td>53 (27.5)</td>
</tr>
<tr>
<td>Received ZDV + 3TC (post-delivery pack)</td>
<td>60 (30.8)</td>
<td>51 (26.4)</td>
</tr>
<tr>
<td>Nevirapine prophylaxis for baby issued</td>
<td>139 (71.3)</td>
<td>133 (68.9)</td>
</tr>
<tr>
<td>HIV diagnosed today</td>
<td>5 (2.6)</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td>HIV counselling done with partner</td>
<td>40 (20.5)</td>
<td>49 (25.4)</td>
</tr>
</tbody>
</table>

3TC, lamivudine; IQR, interquartile range; NVP, nevirapine; ZDV, zidovudine.

No major differences between groups.
RAMMbo validity check

**Representative:** who did the subjects represent?

**Allocation:** randomised? Were groups similar at the start?

**Maintenance:** Were the groups treated equally? Were as many patients as possible followed-up?

**Measurements** blinded or objective
Maintenance: Were the groups treated equally?

5. Aside from the experimental intervention, were the groups treated equally?

Study staff called participants in the SMS arm weekly beginning at 38 weeks gestation to ascertain whether delivery had occurred. Delivery dates for participants in the control arm were abstracted from clinic records. If control women did not return, they were contacted either in person or by phone. Women’s return visits and infant HIV testing data were extracted from clinic records.

Odeny, et al 2014 AIDS
6. Were all of the patients who entered the trial properly accounted for at its conclusion?

Consider:
- Was the trial stopped early?
- Were patients analysed in the groups to which they were randomised?
Maintenance: were as many patients as possible followed-up?

1. Trial profile showing flow of study participants.
Intention to treat

• Once a participant is randomised, they should be analysed to the group they were assigned to

• Pros
  • Reflects “real life” e.g non compliance
  • Unbiased estimate of true effect
  • Maintains sample size thus maintaining statistical power

• Cons
  • Noncompliance provides little data on efficacy
  • Treatment effect may be conservative
  • Dropouts/non-compliant/compliant subjects are different

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159210/
RAMMbo validity check

Representative: who did the subjects represent?

Allocation: randomised? Were groups similar at the start?

Maintenance: Were the groups treated equally? Were as many patients as possible followed-up?

Measurements blinded or objective
3. Were patients, health workers and study personnel blinded?

Consider:
- Health workers could be; clinicians, nurses etc
- Study personnel – especially outcome assessors
Measurements blinded or objective

Women’s return visits and infant HIV testing data were extracted from clinic records.

Odeny, et al 2014 AIDS
(B) What are the results?

7. How large was the treatment effect?

Consider:
- What outcomes were measured?
- Is the primary outcome clearly specified?
- What results were found for each outcome?
- Is there evidence of selective reporting of outcomes?

8. How precise was the estimate of the treatment effect?

Consider:
- What are the confidence limits?
- Were they statistically significant?
What does this study tell us?

• **P values** (hypothesis testing):
  • Tests to exclude the null hypothesis

• **Confidence intervals** (estimation of effect)
  • Range of values within which the true effect is likely to lie
  • Wider the confidence interval, less precision in result

• **Relative Risk**

• **Absolute Risk**

• **Odds Ratios**

• **Number needed to treat**
In the SMS group, 38 of 194 (19.6%) women attended a postpartum clinic visit compared to 22 of 187 (11.8%) in the control group [relative risk (RR) 1.66, 95% confidence interval (CI) 1.02–2.70, \( P = 0.04 \)].

In the per-protocol analysis, women in the SMS arm had a significantly higher probability of attending clinic within 8 weeks compared to those in the control arm (RR 1.83, 95% CI 1.11–3.01).
(C) Will the results help locally?

9. Can the results be applied in your context? (or to the local population?)

Consider:

- Do you have reason to believe that your population of interest is different to that in the trial
- If so, in what way?
10. Were all clinically important outcomes considered?

Consider:

- Is there other information you would like to have seen?
- Was the need for this trial clearly described?

11. Are the benefits worth the harms and costs?

Consider:

- Even if this is not addressed by the trial, what do you think?
Conclusion

• Critical appraisal helps us decide whether evidence is valid, what the results tell us and whether the study is relevant to our setting

• Checklists are available to help

• Don’t believe everything you read in journals!