NHS003: Introduction to critical appraisal of quantitative research

Library Services, Learning and Skills

libraryservices@kcl.ac.uk
Aim

- To increase your confidence in critically appraising quantitative research
• Your name + job

• What brings you to the session today?

• What’s your experience with reading / carrying out quantitative research?
By the end of the session…

• Identify the main concepts of critical appraisal

• Identify some areas to consider when appraising quantitative research

• Practise using a checklist to critically appraise a piece of quantitative research

• Evaluate the strengths and weakness of a piece of quantitative research as a group
Slides + further support

libguides.kcl.ac.uk/NHS

See the Training tab for slides
True or false? Critical appraisal is...

- tearing research apart
- a balanced evaluation of benefits and strengths of the research against its flaws and weaknesses
- assessment of a paper based on its results
- a process that can only be undertaken by experts and statisticians
What is critical appraisal?

• to weigh up the evidence critically to assess “its validity (closeness to the truth) and usefulness (clinical applicability)” (Sackett and Haynes, 1995)

• “Critical appraisal is concerned with the acquisition of necessary skills with which to discern clinical research papers accurately” (Ajetunmobi, 2002)
Why is it important?

• Part of the process of evidence-based practice

• Not all papers are equal: some are good, some are bad, most have strengths and weaknesses

• “[some] published articles belong in the bin and should not be used to inform practice” (Greenhalgh, 2014)
What should I consider when appraising quantitative research?

<table>
<thead>
<tr>
<th>Bias Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>selection bias</strong></td>
<td>Participants or groups chosen for study not accurately representing the intended population</td>
</tr>
<tr>
<td><strong>poor control group</strong></td>
<td>e.g. comparing a new drug with a placebo rather than the current gold standard drug</td>
</tr>
<tr>
<td><strong>performance bias</strong></td>
<td>Study groups being treated differently (beyond the intervention being studied)</td>
</tr>
<tr>
<td><strong>attrition bias</strong></td>
<td>Discounting participants who drop out of a study from results</td>
</tr>
<tr>
<td><strong>surrogate outcomes</strong></td>
<td>Focussing on results for indicators (e.g. reduction in heart rate) rather than real clinical outcomes (e.g. reducing number of heart attacks)</td>
</tr>
<tr>
<td><strong>outcome reporting bias</strong></td>
<td>Authors selectively choosing which outcomes they include, depending on the nature of the results</td>
</tr>
<tr>
<td><strong>publication bias</strong></td>
<td>Positive results and English language studies being over-represented in the literature</td>
</tr>
</tbody>
</table>
• Take 5 minutes to read the article again / skim read the article

• What do you think are the strengths and weaknesses of this piece of research?

Photo of the helmet described in this article

How do I remember what to look out for?

Use a checklist to help - https://casp-uk.net/casp-tools-checklists/
Presenting your thoughts

• Share the information from the help card
• What information is there in the article?
• Where did you find the information in the article (page number / section)?
• What do you think?
Q1. Did the trial address a clearly focused issue?

Consider PICO:

- Who is the **population** under study?
- What is the **intervention**/exposure?
- What is the **comparison**?
- What is the **outcome**?
Q2. Was the assignment of patients to treatments randomised?

- How was randomisation carried out?

- Was the allocation sequence concealed from researchers and patients?
Q2. Randomisation

• Patients should be randomly allocated to intervention / control groups

• Why? Helps ensure the groups are as comparable as possible. This means that differences in outcomes are not influenced by factors apart from the one being tested.

• How? Allocation concealment - the patients and the researchers shouldn’t be able to influence the group each patient ends up in

More info: Glasziou et al. (2007), pp.80-82
Q3. Were all of the patients who entered the trial properly accounted for at its conclusion?

• Was the trial stopped early?

• Were patients analysed in the groups to which they were randomised?
Q3. Accounting for all patients

- Why? Ignoring patients who withdraw from a clinical trial will bias the results

- Look for a flow diagram to show progress of all patients throughout the trial

- Look for use of **intention-to-treat analysis** – subjects are analysed in the group they were randomised in, even if they never receive treatment or don’t complete the trial. This type of analysis is standard for comparative studies.

More info: Greenhalgh (2014), pp.56-57
**Figure 1. Flow of Participants Through Study**

4226 Patients admitted to the medical ICU assessed for eligibility

740 Had acute respiratory failure requiring noninvasive ventilation

657 Excluded
- 501 Did not meet entry criteria
- 456 Received NIV for <8 h
  - 40 Previously enrolled in current study
  - 10 Hypercarbic respiratory failure
- 115 Met ≥1 exclusion criteria
  - 85 Do-not-intubate order
  - 9 Upper airway obstruction
  - 1 Cardiac arrest
  - 3 No gag reflex
  - 2 Pregnant
  - 15 No research staff available
- 41 Declined to participate

83 Randomized

39 Randomized to receive NIV via face mask
  - 39 Received NIV via face mask as randomized
  - 39 Included in primary analysis

44 Randomized to receive NIV via helmet
  - 44 Received NIV via helmet as randomized
  - 44 Included in primary analysis

ICU indicates intensive care unit; NIV, noninvasive ventilation.
CONSORT 2010 Flow Diagram

Enrollment

Assessed for eligibility (n= )

Excluded (n= )
  - Not meeting inclusion criteria (n= )
  - Declined to participate (n= )
  - Other reasons (n= )

Randomized (n= )

Allocation

Allocated to intervention (n= )
  - Received allocated intervention (n= )
  - Did not receive allocated intervention (give reasons) (n= )

Follow-Up

Lost to follow-up (give reasons) (n= )
Discontinued intervention (give reasons) (n= )

Analysis

Analysed (n= )
  - Excluded from analysis (give reasons) (n= )

More info:
CONSORT Checklist
http://www.consort-statement.org/consort-statement/checklist
sections 13a ‘Participant Flow’ and 13b ‘Losses and exclusions’
4. Were patients, health workers and study personnel ‘blind’ to treatment?
Q4. Blinding

• Preventing those involved in a trial from knowing which group a patient belongs to

• Why? So that behaviour doesn’t change as a result of knowing which group a patient belongs to

• Is the description of the blinding clear? Who was blinded?

• Blinding of certain groups is not always possible

5. Were the groups similar at the start of the trial?

- What other factors might affect the outcome (e.g. age, sex, social class)?
Q5. Similar groups

• Look for a baseline table

• Check to see if the groups are similar in terms of factors that might affect the outcome

• If the groups are not balanced, is this acknowledged and what steps have been taken to overcome the problem?

Table 1. Characteristics of Patients at Baseline

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. (%) of Patients Receiving Noninvasive Ventilation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Face Mask (n = 39)</td>
</tr>
<tr>
<td>Age, median (IQR), y</td>
<td>60.9 (56.4-71.1)</td>
</tr>
<tr>
<td>Women</td>
<td>18 (46)</td>
</tr>
<tr>
<td>Black</td>
<td>22 (56)</td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>13 (33)</td>
</tr>
<tr>
<td>White, Hispanic</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Body mass index, median (IQR)</td>
<td>28 (23-35)</td>
</tr>
<tr>
<td>APACHE II(^a), median (IQR)</td>
<td>26 (23-30)</td>
</tr>
<tr>
<td>Medical History</td>
<td></td>
</tr>
<tr>
<td>Solid cancer</td>
<td>10 (26)</td>
</tr>
<tr>
<td>Hematologic cancer</td>
<td>6 (15)</td>
</tr>
<tr>
<td>Solid organ transplant</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Stem cell transplant</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Reason for Acute Respiratory Failure</td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>14 (36)</td>
</tr>
<tr>
<td>Aspiration</td>
<td>5 (13)</td>
</tr>
<tr>
<td>Extrapulmonary ARDS</td>
<td>6 (15)</td>
</tr>
<tr>
<td>Pneumonia due to immunosuppression(^b)</td>
<td>14 (36)</td>
</tr>
</tbody>
</table>
Q6. Aside from the experimental intervention were the groups treated equally?
Q6. Equal treatment

The study groups should be treated the same in every respect apart from the factor being investigated.

• Why? To keep the groups matched as closely as possible so that differences in outcomes are due to the factor being investigated.

More info: Glasziou et al. (2007), p.83
Q7. How large was the treatment effect?

• What outcomes were measured?
• Is the primary outcome clearly specified?
• What results were found for each outcome?
You have responsibility for your Trust budget and are looking at introducing a new analgesic. Which one would you fund?

A: Decreases patients’ pain rate from 80% to 30%
B: Absolute risk reduction of pain of 50%
C: Reduces the rate of pain by 62.5%
D: For every 2 patients treated with the new drug, one patient would expect a reduction in pain
Measures of treatment effect – 2 by 2 tables

*Using the trust budget example, enter the missing information in this table:*

<table>
<thead>
<tr>
<th>Outcome event</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>new analgesic</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td><strong>Control group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>placebo</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>a + c</td>
<td>b + d</td>
<td>a + b + c + d</td>
</tr>
</tbody>
</table>

From this information, different measures of treatment effect can be calculated – see handout.
A: Decreases patients’ pain rate from 80% to 30%
Control Event Rate (CER) and Experimental Event Rate (EER)

B: Absolute risk reduction of pain of 50%
Absolute Risk Reduction (ARR)

C: Reduces the rate of pain by 62.5%
Relative Risk Reduction (RRR)

D: For every 2 patients treated with the new drug, one patient would expect a reduction in pain
Number needed to treat (NNT)
Using Table 2 in the article (p.5), complete the missing information in the table.

<table>
<thead>
<tr>
<th>Outcome event</th>
<th>endotracheal intubation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Experimental group</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>helmet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control group</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>face mask</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
</tr>
</tbody>
</table>

Using Table 2 in the article (p.5), complete the missing information in the table.

<table>
<thead>
<tr>
<th>Primary outcome, No. (%)</th>
<th>Face Mask (n = 39)</th>
<th>Helmet (n = 44)</th>
<th>Absolute Difference (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endotracheal intubation</td>
<td>i) 24 (61.5)</td>
<td>ii) 8 (18.2)</td>
<td>iii) -43.3 (-62.4 to -24.3)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

i) Control Event Rate (CER) $\frac{c}{c+d} = \frac{24}{39} = 61.5\%$

ii) Experimental Event Rate (EER) $\frac{a}{a+b} = \frac{8}{44} = 18.2\%$

iii) $18.2\% - 61.5\% = -43.3\%$
• The same data can be presented in different ways. How it is presented can depend on what the author is trying to convey.

• Relative risk values can look impressive even when the absolute risk reduction is very small - treat with caution!

Q8. How precise was the estimate of the treatment effect?

• What are the confidence limits?
Are the results statistically significant?

Is there a true difference between outcomes in the control and intervention groups? (or was it down to chance?)

Three things to look for:
• Sample size – power calculation
• P-values
• Confidence intervals
Sample size

- Is information on the power calculation included?
  - A power calculation is used to calculate the sample size necessary to detect a true difference between outcomes in the control and intervention groups
  - Power of 80-90% is standard

- What was the power in this paper?

More info: Gosall and Gosall (2015), pp.183-4
P-value

- The lower the P-value the higher the statistical significance of the finding
- P-value of 0.05 or less is frequently deemed to be “statistically significant” but this is an arbitrary figure

What are the P-values in Table 2?

Are they statistically significant?
### Table 2. Primary and Secondary Outcomes and Adverse Events

<table>
<thead>
<tr>
<th></th>
<th>Face Mask (n = 39)</th>
<th>Helmet (n = 44)</th>
<th>Absolute Difference (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary outcome, No. (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endotracheal intubation</td>
<td>24 (61.5)</td>
<td>8 (18.2)</td>
<td>-43.3 (-62.4 to -24.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Reason for intubation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory failure</td>
<td>20 (83.3)</td>
<td>3 (37.5)</td>
<td>-45.3 (-82.5 to -9.1)</td>
<td>.01</td>
</tr>
<tr>
<td>Circulatory failure</td>
<td>3 (12.5)</td>
<td>0 (0)</td>
<td>-12.5 (-25.7 to 0.7)</td>
<td>.55</td>
</tr>
<tr>
<td>Neurologic failure</td>
<td>1 (4.2)</td>
<td>5 (62.5)</td>
<td>58.3 (24.8 to 92.8)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Secondary outcomes, median (IQR), d</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ventilator-free days</td>
<td>12.5 (0.49-28)</td>
<td>28 (13.7-28)</td>
<td>8.4 (13.4 to 3.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ICU length of stay</td>
<td>7.8 (3.9-13.8)</td>
<td>4.7 (2.5-8.7)</td>
<td>-2.76 (-6.07 to 0.54)</td>
<td>.04</td>
</tr>
<tr>
<td>Hospital length of stay</td>
<td>15.2 (7.8-19.7)</td>
<td>10.1 (6.5-15.9)</td>
<td>-2.92 (-8.47 to 2.63)</td>
<td>.16</td>
</tr>
<tr>
<td><strong>Mortality, No. (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital</td>
<td>19 (48.7)</td>
<td>12 (27.3)</td>
<td>-21.4 (-41.9 to -1.0)</td>
<td>.04</td>
</tr>
<tr>
<td>90-d</td>
<td>22 (56.4)</td>
<td>15 (34.1)</td>
<td>-22.3 (-43.3 to -1.4)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Adverse events</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mask deflation</td>
<td>0 (0)</td>
<td>2 (4.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin ulceration</td>
<td>3 (7.6)</td>
<td>3 (6.8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ICU, intensive care unit; IQR, interquartile range.

*90-d Mortality includes hospital mortality.
Confidence intervals

• As studies contain samples of a population (rather than everybody in the population), papers often report the range in which we are 95% (or 99%) confident that the ‘true’ result for the population lies.

• The larger the trial = the narrower the confidence interval = the more confident we can be that the result is true for the entire population.
Confidence intervals

For an **absolute** difference, an interval that includes zero means the result is not statistically significant.

For a **ratio or relative** difference, an interval that includes one means the result is not statistically significant.

- Where are the confidence intervals in Table 2?
- Do these confidence intervals relate to absolute or relative differences?
- Are the results statistically significant?
### Table 2. Primary and Secondary Outcomes and Adverse Events

<table>
<thead>
<tr>
<th>Primary outcome, No. (%)</th>
<th>Face Mask (n = 39)</th>
<th>Helmet (n = 44)</th>
<th>Absolute Difference (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endotracheal intubation</td>
<td>24 (61.5)</td>
<td>8 (18.2)</td>
<td>−43.3 (−62.4 to −24.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reason for intubation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory failure</td>
<td>20 (83.3)</td>
<td>3 (37.5)</td>
<td>−45.3 (−82.5 to −9.1)</td>
<td>.01</td>
</tr>
<tr>
<td>Circulatory failure</td>
<td>3 (12.5)</td>
<td>0 (0)</td>
<td>−12.5 (−25.7 to 0.7)</td>
<td>.55</td>
</tr>
<tr>
<td>Neurologic failure</td>
<td>1 (4.2)</td>
<td>5 (62.5)</td>
<td>58.3 (24.8 to 92.8)</td>
<td>.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary outcomes, median (IQR), d</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventilator-free days</td>
<td>12.5 (0.49-28)</td>
<td>28 (13.7-28)</td>
<td>8.4 (13.4 to 3.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ICU length of stay</td>
<td>7.8 (3.9-13.8)</td>
<td>4.7 (2.5-8.7)</td>
<td>−2.76 (−6.07 to 0.54)</td>
<td>.04</td>
</tr>
<tr>
<td>Hospital length of stay</td>
<td>15.2 (7.8-19.7)</td>
<td>10.1 (6.5-15.9)</td>
<td>−2.92 (−8.47 to 2.63)</td>
<td>.16</td>
</tr>
<tr>
<td>Mortality, No. (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital</td>
<td>19 (48.7)</td>
<td>12 (27.3)</td>
<td>−21.4 (−41.9 to −1.0)</td>
<td>.04</td>
</tr>
<tr>
<td>90 d&lt;sup&gt;a&lt;/sup&gt;</td>
<td>22 (56.4)</td>
<td>15 (34.1)</td>
<td>−22.3 (−43.3 to −1.4)</td>
<td>.02</td>
</tr>
</tbody>
</table>

### Adverse events

<table>
<thead>
<tr>
<th>Adverse events</th>
<th>Face Mask (n = 39)</th>
<th>Helmet (n = 44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mask deflation</td>
<td>0 (0)</td>
<td>2 (4.5)</td>
</tr>
<tr>
<td>Skin ulceration</td>
<td>3 (7.6)</td>
<td>3 (6.8)</td>
</tr>
</tbody>
</table>

Abbreviations: ICU, intensive care unit; IQR, interquartile range.

<sup>a</sup> 90-d Mortality includes hospital mortality.
Help with stats

The resources on the Critical Appraisal tab of Library Guide include:

• Guide to p-values and confidence intervals
• Biomedical Research Centre at Guy’s runs a research methods programme every Spring and Autumn, including courses on stats
• Statistical advice consultancy for staff across King’s Health Partners (KHP) - 1hr booked appointments with a professional statistician
• Books available from Library Services
• Don’t get bogged down with stats – they are only one part of critical appraisal!

• Statistical significance doesn’t necessarily equal clinical significance…
Q9. Can the results be applied in your context?

Q10. Were all clinically important outcomes considered?

Q11. Are the benefits worth the harms and costs?
Transferability

• Can I apply these results to my own practice?
• Group under study: are they the same as your patients?
  - socio-cultural origin, gender, age etc.
  - location of research: country, setting
  - are the exclusion criteria reasonable?
• Is there other information you would like to have seen?
• Does the paper consider the results from the perspective of difference stakeholders?
Tips for success

• Group work
• Read all the paper
• Keep calm and carry on!
• Review and feedback
• Consider your context
• Resources on Critical Appraisal tab of Library Guide

“Undertaking a critical appraisal is really using your everyday skills, and applying them in a more structured and systematic way” - Dawes (2005)
References

• Ajetunmobi, O. (2001) Making sense of critical appraisal


• Glasziou, P. et al. (2007) Evidence-based practice workbook (2nd ed)


• Greenhalgh, T (2014) How to read a paper: the basics of evidence-based medicine (5th ed)


Other support

NHS libguide - libguides.kcl.ac.uk/nhs

• Using the libraries
• Info on using databases + accessing full-text journal articles
• Other sessions (finding evidence quickly, using databases, introduction to critically appraising qualitative research)
Reflection

Next month you’ll be emailed a link to a survey asking what impact today’s session has had for you. This helps us ensure that sessions are as useful as possible for NHS staff.

As part of the survey you’ll be asked to leave your details if you would like a certificate.