Assessing study quality: critical appraisal

Assessing study quality: why?

Why critical appraise?

- Evidence informed decision-making assumes people make better decisions if informed by **best available evidence**

- Systematic reviews should distinguish between studies that are **more susceptible or less susceptible to bias**
  - Prioritise best available evidence

- Conclusions of a systematic review should interpret the evidence in light of the quality of evidence (i.e. least biased) studies on a given subject

The Cochrane Handbook states that the evaluation of the validity of the included studies is "an essential component of a Cochrane review, and should influence the analysis, interpretation and conclusions of the review."

Critical appraisal allows you to

1. Exclude studies that fail to meet your methodological criteria

OR

2. Stratify included studies by those with relatively high or relatively low risks of bias

OR

3. A mixture of the two

For transparency it is important to state clearly the criteria by which you appraise your studies (ideally *a priori)*.

To help there are a number of CA tools / checklists.

Assessing study quality: what do you look for?

Determine study design

- Key criteria for inclusion/exclusion of studies

- Needs to be pre-specified

- Deciding study design can be very confusion

- Key elements of study design...

Determine study design: key questions

- Does the study report outcome data on the effects of an intervention?
  - What is the intervention?

- How was the intervention assigned? Randomised?

- Was there a control group?

- Was outcome data collected before & after the intervention?

- Was the outcome data collected from the same people? Was it a cohort study? Or cross-sectional data?

- Was study retrospective?

Retrospective studies

- Used in natural experiments- look back for data

- Were the features of the study described carried out AFTER the study was designed:

  - Identification of participants?

  - Assessment before intervention?

  - Actions/choices leading to an individual becoming a member of a group?

  - Assessment of outcomes?

Determine study design

- Randomised: cluster/individual
  - Experimental & quasi-experimental

- Non-randomised designs
  - Agree consistent labels for designs
  - Do not assume authors use correct or same labelling system
  - *(see earlier best available evidence session)*
    - Retrospective/Prospective/Before & After...
    - Cohort/cross-sectional

Study design: Inclusion/exclusion criteria

- In addition to issues of: randomisation & use of control group, think about whether or not to include studies if:

  - Reports outcomes at follow-up only rather than **changes in outcome** before and after the intervention

  - Unclear if a **cohort** has been used, i.e. outcome data are reported before & after the intervention but the study has used **cross-sectional data** at both time points

  - If it is not clear how participants were **selected** or **assigned** to the intervention

Example questions about experimental designs

1. How were comparison/control groups formed?
   - Random assignment
   - Other (specify)

2. If random assignment, specify design:
   - Simple/systematic (individuals/families)
   - Stratified/blocked (identify stratifying variables)
   - Yoked pairs (created by timing of enrollment into the study)
   - Matched pairs (identify matching variables)
   - Cluster (group) randomized
   - Other (specify)
   - Can't tell

More on experimental research design

3. Who performed group assignment?
- Research staff
- Program staff
- Can't tell
- Other (specify)

4. How was random assignment performed?
- computer generated
- random numbers table
- coins or dice
- other (describe)
- can't tell

Important aspects of study quality (interventions)

- **Assignment to intervention**
  - Changes in assignment/exposure to intervention during the study
  - Crossover designs
  - Multiple intervention and/or control groups
  - Natural experiment: how was intervention allocated-pragmatic to area etc?

- **Contamination**
  - Were any of the control group exposed to the intervention during the study?

- **Selection of participants**
  - Do they represent the target population?
    - e.g. wealthy people receiving food aid
    - What was baseline response for study?

Important aspects of study quality (interventions)

- **Missing data**
  - How many people completed the study?
    - Attrition, dropouts, withdrawals at follow-up
    - Characteristics of dropouts

- **Selective outcome reporting**
  - Did the authors only report the positive findings
  - Needs to be confirmed in a protocol

- **participants, assessors, analysts…**
- **Blinding:**
  - Can you blind people to the intervention?
  - But may be useful to blind the assessors and/or analysts

Important aspects of study quality (interventions)

- **Confounding**
  - Randomisation should deal with this but this needs to be confirmed

  - Comparability of control & intervention groups at baseline (prior to intervention)
    - Key characteristics (these may vary from one topic to another, e.g. housing condition)
    - Key outcome, e.g. health status

  - May be able to control for confounding statistically

Assessing quality of other study designs

- Interrupted time series
  - Example: Is the intervention independent of other changes?
  - Are there sufficient data points to allow inference?

- Case control study
  - E.g. Similarity of groups, valid measurements?

- Cross-sectional survey
  - E.g. Size: statistical power? Outcome measure?

- Qualitative studies…

External validity

Things to consider in a study:

- **Reach & representativeness of study sample**
  - Define/compare target study sample/setting & wider target population/setting for intervention
  - Eligibility of population/setting

- **Program or policy implementation & adaptation**
  - Details of intervention and heterogeneity of intervention across sample/settings, specialist training
  - Mediating factors on intervention effect, e.g. context

- **Outcomes for decision making**
  - Generalisable outcomes, consideration of range of outcomes, dose-response, costs

- **Maintenance & institutionalisation**
  - Long term effects, sustainability of intervention (including cost)

From: Green LW, Glasgow RE. Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology. *Eval Health Prof* 2006;29(**1**):126-153.

Assessing quality of study **OR** data?

- Study design
- Contamination
- Selection of participants
- Missing data
- Selective outcome reporting
- Blinding of assessors
- Confounding

- Some aspects may vary by outcome
- **Recommended that certain aspects are assessed by outcome**

Assessing quality by outcome

- Likely to be the same across all study data
  - Study design
  - Contamination
  - Selection of participants
  - Selective outcome reporting

- Likely to vary by outcome
  - Missing data
  - Blinding of assessors
  - Confounding

Assessing other aspects of study quality: rating tools

- More than 200 scales and checklists available, few if any appropriate for systematic reviews (Deeks et al., 2003)

- Overall study quality scores have questionable reliability/validity (Jüni et al., 2001)
    - Mix up different methods issues which may have different impacts on reliability/validity of data

    - Preferable to examine potential influence of key components of methodological quality individually

- Avoid numerical scales of overall study quality
    - Especially in meta-analysis

What Cochrane says *(see Cochrane Handbook)*

"*The Collaboration's recommended tool for assessing risk of bias is neither a scale nor a checklist. It is a **domain-based evaluation**, in which critical assessments are made separately for different domains…. The most realistic assessment of the validity of a study may involve subjectivity…*

What are the current recommendations?

The Cochrane handbook suggests that reviewers:

- Critically appraise each outcome of interest

- Do not give an overall summary score

- Instead, group your criteria into a number of domains

- Assess the risk of bias for each domain

- Identify which domains are most important for your particular review (i.e. carry the greatest risk of bias)

- Assign a 'low', 'high' or 'unclear' risk of bias label to each relevant outcome in each included study.

Cochrane critical appraisal tool
Domains and Criteria

- ***Selection bias***
  - Random sequence generation.
  - Allocation concealment.
- ***Performance bias.***
  - Blinding of participants and personnel *(for each outcome)*
- ***Detection bias.***
  - *Blinding of outcome assessment (for each outcome)*
- ***Attrition bias.***
  - *Incomplete outcome data (for each outcome)*
- ***Reporting bias.***
  - *Selective reporting.*
- ***Other bias.***
  - *Other sources of bias*

(Cochrane handbook Table 8.5.a)

What are the current recommendations?

The Cochrane handbook suggests that reviewers:

- Critically appraise each outcome of interest

- Do not give an overall summary score

- Instead, group your criteria into a number of domains

- Assess the risk of bias for each domain

- Identify which domains are most important for your particular review (i.e. carry the greatest risk of bias)

- Assign a 'low', 'high' or 'unclear' risk of bias label to each relevant outcome in each included study.

Some examples of CA tools…

- **The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses**
  http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
  - Case control
  - Cohort

- **EPHPP "Quality Assessment Tool for Quantitative Studies"**

  (sometimes known as the 'Hamilton tool')

  http://www.ephpp.ca/tools.html

Some examples of CA tools…

- **The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses**
  http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
  - Case control
  - Cohort

- **EPHPP "Quality Assessment Tool for Quantitative Studies"**

  (sometimes known as the 'Hamilton tool')

  http://www.ephpp.ca/tools.html

  See handout

EPHPP "Quality Assessment Tool for Quantitative Studies"

1. Selection bias (representative sample, response rate)
2. Study design (with optional questions on randomisation)
3. Confounders (control group comparability)
4. Blinding (assessor and participants)
5. Data collection methods (validity, reliability)
6. Withdrawals/drop outs (Reported? Explained? Attrition rates?)
7. Intervention integrity (Exposure, consistency, contamination)
8. Analysis (appropriate statistics, intention to treat?)

An accompanying 'Dictionary' explains each of the questions in more detail and provides guidance on scoring.

EPHPP 'SCORING'

ASSUMES YOU HAVE AT LEAST TWO INDEPENDENT REVIEWERS WHO MUST REACH A FINAL AGREEMENT

<u>SIMPLE GLOBAL RATING FOR THIS PAPER</u>

- STRONG (no WEAK ratings)
- MODERATE (one WEAK rating)
- WEAK (two or more WEAK ratings)

Example from housing review

- Critical appraisal adapted from EPHPP "Quality Assessment Tool for Quantitative Studies"

- But some changes were made
  - **Blinding** (can't blind participants against housing improvement)
  - **Intervention integrity** (miss out question on consistency of intervention)
  - **Analysis** (miss out questions on intention to treat)

- Describe & justify changes in tool in final report

Critical appraisal: summary

- Essential to assess bias & interpret data appropriately
  - Use in decisions about what is included in synthesis

- MUST be done by two independent reviewers to minimise bias in reviewers assessment

- Poor reporting of studies makes it difficult to work out what was done- especially study design
  - Especially difficult for non-randomised studies

- Much debate about how to assess quality of non-randomised studies
  - Many tools available
  - Use a recommended tool

Critical appraisal: summary

- Interpretation of what and how much bias is introduced by study shortcomings is subject of much debate

- Appraise at an outcome-level if you have >1 outcome

- If it is not reported or is not clear- consider it as not done

- Critical appraisal is difficult!
  - Don't spend all day scratching your head. Take a deep breath and make a decision with co-reviewer
  - Keep a record to maintain transparency and consistency