Information skills: Critical appraisal – May 2024

CONTENT

Introduction to critical appraisal
Bias in studies
Study design
Randomised Controlled Trials
Systematic reviews
How to critically appraise a paper
Online resources
Glossary
Example

1. Introduction to critical appraisal

Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, and its value and relevance in a particular context.
(Amanda Burls: What is critical appraisal)

Critical appraisal is an important element of evidence-based practice (EBP). EBP is a decision-making process comprising 5 steps, focusing on developing a skill set to search, understand and apply the evidence base:

1. Asking an **answerable question**, i.e. formulating a question into a format whereby you can interrogate the medical literature and hopefully find an answer - to do this, you may use the PICO tool.
2. You then need to **search** for the evidence - if you can find a pre-appraised resource, you can miss out the next step.
3. The next step is **critical appraisal** of your results.
4. You then decide what **action** to take from your findings.
5. Finally, you **evaluate** your new or amended practice.

ASK ➡ ACQUIRE ➡ APPRAISE ➡ APPLY ➡ ASSESS/AUDIT steps 1-4

Critical appraisal is essential to:

- combat information overload by eliminating irrelevant or weak studies;

- distinguish evidence from opinion, assumptions, misreporting, and belief;
- assess the validity of the study;
- assess the usefulness and clinical applicability of the study;
- identify papers that are clinically relevant;
- recognise any potential for bias.

2. Bias in studies

When reviewing the literature published in scientific/medical journals, we should consider that papers with significant positive results are more likely to be:

- submitted and accepted for publication (**publication bias**);
- published in a major journal written in English (**Tower of Babel bias**);
- published in a journal indexed in a literature database, especially in less developed countries (**database bias**);
- cited by other authors (**citation bias**);
- published repeatedly (**multiple publication bias**);
- … and quoted by newspapers!

In **Randomised Controlled Trials,** sources of bias to look for are:

- systematic differences in the groups being compared, caused by incomplete randomisation at allocation stage (**selection bias**)
- systematic differences in the care provided apart from the intervention being evaluated at intervention stage (**performance bias**)
- systematic differences in withdrawals / exclusions of people from the trial at follow up stage (**exclusion bias**)
- systematic differences in the ways outcomes are assessed at outcomes stage (**detection bias**)

3. Study design

The following lists summarise the most common types of study design found in the medical literature.

3.1. Qualitative studies: subjective/expressed in words

Qualitative studies explore and understand people's beliefs, experiences, attitudes, behaviour and interactions. They generate non-numerical data. Examples of qualitative studies:

- **Passive observation** - systematic watching of behaviour and talk in natural occurring settings;
- **Participant observation** - observation in which the researcher also occupies a role or part in the setting, in addition to observing;

- **In depth interview** - face to face conversation with the purpose of exploring issues or topics in detail. Does not use preset questions, but is shaped by a defined set of topics;
- **Focus group** - method of group interview which explicitly includes and uses the group interaction to generate data.
- **Document** - study of documentary accounts of events, such as meetings.

3.2. Quantitative studies: objective/expressed in numbers

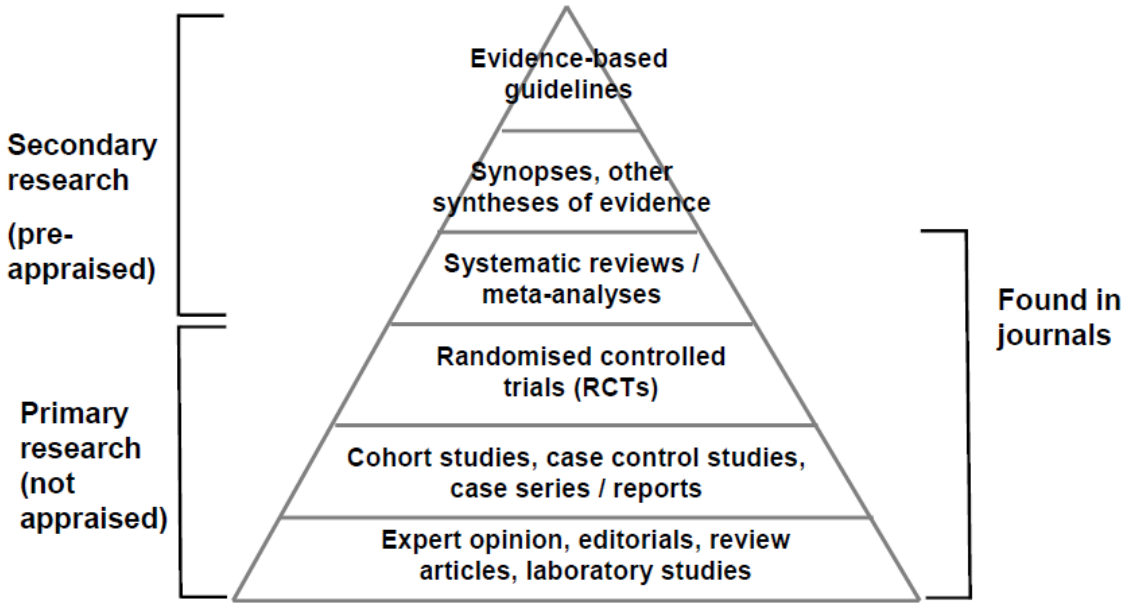
Quantitative studies generate numerical data or data that can be converted into numbers. Examples of quantitative studies:

- **Case report** - report on a single patient;
- **Case series** - report on a series of patients (no control group);
- **Case control study** (retrospective study) - identifies patients with a particular outcome (cases) and control patients without the outcome. Looks back and explores exposures and possible links to outcome. These types of studies are often less reliable than randomised controlled trials and cohort studies because showing a statistical relationship does not mean than one factor necessarily caused the other.
- **Cohort study** (prospective studies) - follows groups of people who are selected on the basis of their exposure to a particular agent (e.g.: vaccine, drug) over a period of time, which can be years. They are compared with another group who are not affected by the condition or treatment. These studies are not as reliable as randomised controlled clinical trials as the two groups may differ in a way other than the variable being studied.

Key quantitative studies:

- **Randomised Controlled Trial (RCT)** - a clinical trial in which participants are randomly allocated to a test treatment and a control; involves concurrent enrolment and follow-up of both groups; gold standard in testing the efficacy of an intervention (therapy/prevention);
- **Systematic review** - identifies and critically appraises all research on a specific topic, and combines valid studies; increasingly important in evidence-based medicine; different from **review article** (which is a summary of more than one paper on a specific topic, and which may or may not be comprehensive);
- **Meta-analysis** - a systematic review that uses quantitative methods to summarise the results.

The following diagram shows a model for the organisation of some quantitative studies. Different types of studies are located at different levels of the **hierarchy of evidence**:



There are also **other types** of quantitative studies, such as:

- **Cross-sectional survey** - the observation of a defined population at a single point in time or time interval. Exposure and outcome are determined simultaneously. Gold standard in diagnosis and screening research;
- **Decision analysis** - uses the results of primary studies to generate probability trees to be used in making choices about clinical management or resource allocation;
- **Economic analysis** - uses the results of primary studies to say whether a particular course of action is a good use of resources.

3.3. Critical appraisal of different study designs

To critically appraise a journal article, you would have to start by assessing the research methods used in the study. This is done using **checklists** which are specific to the study design:

CASP Checklists
SIGN Checklists
CEBM Critical Appraisal Tools
JBI Critical Appraisal Tools
AMSTAR Checklist (for Systematic Reviews)

Critical appraisal for antiracism

Traditional appraisal tools do not prompt appraisers to consider issues such as minoritised ethnic groups' under-representation, and the effect of racial bias on medical research and its application. To address these issues, a short checklist of five questions is provided to be used alongside existing tools: **Critically Appraising for Antiracism**

4. Randomised Controlled Trials (RCTs)

4.1. Mechanisms to control bias in RCTs

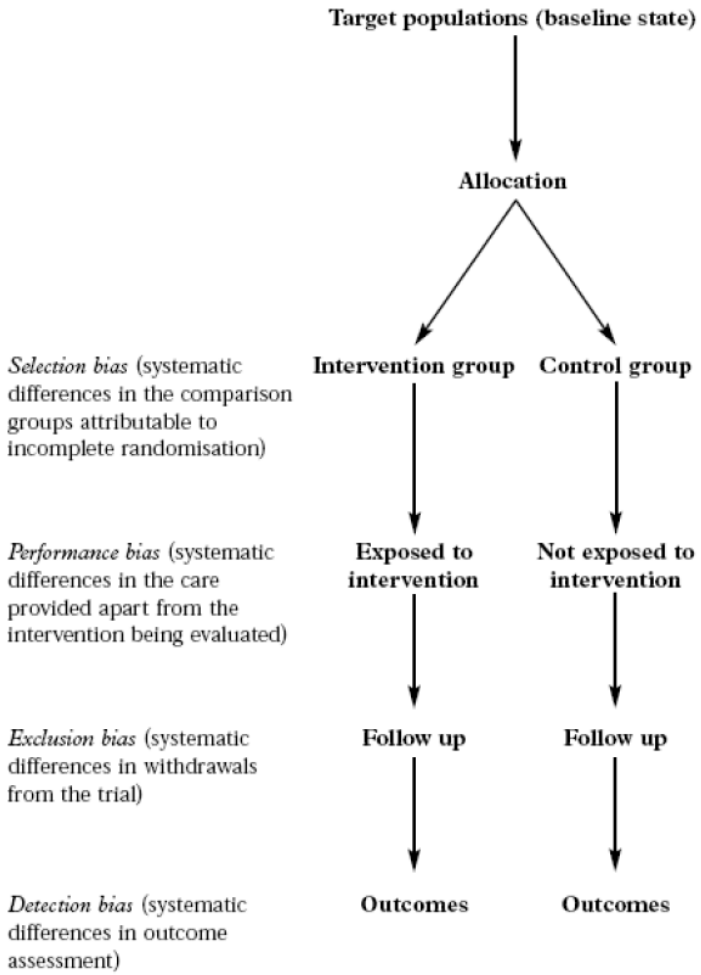
RCTs control bias by randomisation and blinding. Randomisation indicates that participants are randomly allocated to treatment or control group.

- Acceptable methods of randomisation include random numbers, either from tables or computer-generated.
- Unacceptable methods include last digit of date of birth, date seen in clinic etc.
- Stratified randomisation is often used to avoid confounding factors, i.e. to ensure equal distribution of participants with a characteristic thought to affect prognosis or response.

Blinding means masking who is getting treatment and control.

- Single blinding: participants do not know.
- Double blinding: neither the participants nor those giving the intervention know.
- Triple blinding: statisticians doing the analysis also do not know.

The following diagram illustrates the sources of bias in RCTs:



4.2. Advantages and disadvantages of RCTs

Advantages:
- allow for rigorous evaluation of a single variable;
- potentially eradicate bias;
- allow for meta-analysis.

Disadvantages:
- expensive;
- time consuming;
- ethically problematic at times - a trial is sometimes stopped early if dramatic effects are seen.

4.3. Preliminary statistical concepts in RCTs

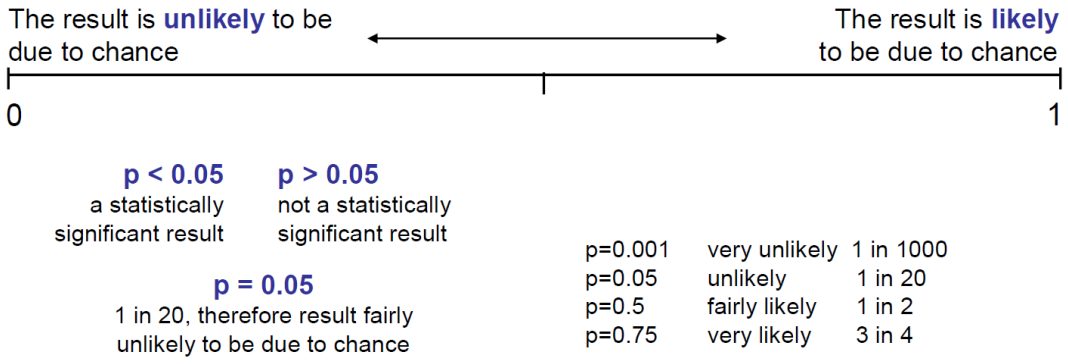
Baseline characteristics - both the control and the intervention group should be broadly similar in factors like age, sex distribution and level of illness.

Sample size calculation (**Power calculation**) - a trial should be big enough to have a high chance of detecting a worthwhile effect if it exists. Statisticians can work out before the trial begins how large the sample size should be in order to have a good chance of detecting a true difference between the intervention and control groups.

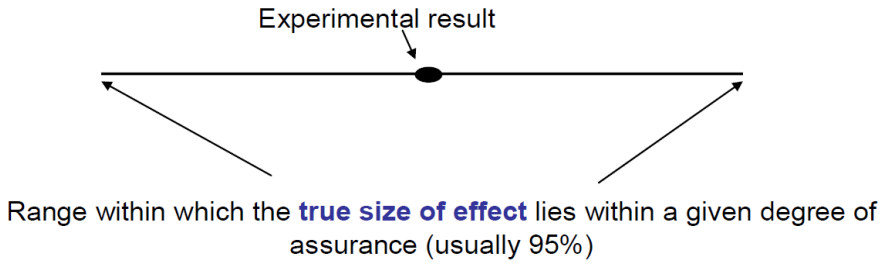
Intention to treat - all data on participants including those who withdraw from the trial should be analysed. Failure to do so may lead to underestimation/ overestimation of results.

4.4. Presenting the results of RCTs

p-value - the probability value refers to the probability that any particular outcome would have arisen by chance. p-values can range from 0 (impossible for the event to happen by chance) to 1 (the event will certainly happen).
If p=0.001 the likelihood of a result happening by chance is extremely low: 1 in 1000
If p=0.05 it is fairly unlikely that the result happened by chance: 1 in 20
A p-value of less than 1 in 20 (p<0.05) is statistically significant.



Confidence interval – it is used in the same way as p-values in assessing the effects of chance but can give you more information. Any result obtained in a sample of patients can only give an estimate of the result which would be obtained in the whole population. The real value will not be known, but the confidence interval can show the size of the likely variation from the true figure. A 95% confidence interval means that there is a 95% chance that the true size of effect will lie within this range (this is equivalent to a p-value of 0.05).



Short video: https://youtu.be/lRMihDrXZtY?feature=shared

The larger the trial the narrower the confidence interval, and therefore the more likely the result is to be definitive. In an odds ratio diagram if the confidence interval crosses the line of zero difference (no effect) it can mean either that there is no significant difference between the treatments and/or that the sample size was too small to allow us to be confident where the true result lies.

Intention-to-treat analysis – it is one in which all the participants in a trial are analysed according to the intervention to which they were allocated, whether they received it or not (for whatever reason). Intention-to-treat analyses are favoured in assessments of effectiveness as they reflect the non-compliance and treatment changes that are likely to occur when the intervention is used in practice and because of the risk of bias when participants are excluded from the analysis.

4.5. Quantifying the risk of benefit/harm in RCTs

When talking about the chance of something happening, e.g. death, hip fracture, we can talk about:
• risk and relative risk
or
• odds and odds ratio

- **Experimental Event Rate** (**EER**) - in the treatment group, number of patients with outcome divided by total number of patients.

- **Control Event Rate** (**CER**) - in the control group, number of patients with outcome divided by total number of patients.

- **Relative Risk** or **Risk Ratio (RR**) - is a ratio of proportions, the risk of the outcome occurring in the intervention group compared with the control group.

RR= EER/CER

RR <1 if group represented in the numerator is at lower “risk” of the event.
Want this if the event is a bad outcome e.g. death.

RR >1 if group represented in numerator is at greater “risk” of the event.
Want this if the event is a good outcome e.g. smoking cessation.

- **Absolute Risk Reduction (ARR)** or **Absolute Benefit Increase** (**ABI**) - absolute amount by which the intervention reduces (or increases) the risk of outcome.
  If the experimental intervention makes a bad outcome more likely we talk about an ARI. ARR= CER - EER
  Use this term if the event is bad e.g. death.

- **Relative Risk Reduction (RRR**) – the difference in the risk of the event between the control and experimental groups, relative to the control group.
  RRR = (CER - EER)/CER.
  An alternative way of calculating the relative risk reduction is to use the relative risk: RRR = (1 - RR).
  Use this term if the event is bad e.g. death.

- **Relative Benefit Increase (RBI)** - **t**he difference in the risk of the event between the control and experimental groups, relative to the control group.
  RBI = (CER - EER)/CER.
  An alternative way of calculating the relative benefit increase is to use the relative risk: RBI = (1 - RR).
  Use this term if the event is good e.g. smoking cessation.

- **Odds of outcome** - in each patient group, the number of patients with an outcome divided by the number of patients without the outcome.

- **Odds ratio** - odds of outcome in treatment group divided by odds of outcome in control group.
  If the outcome is negative, an effective treatment will have an odds ratio <1;
  If the outcome is positive, an effective treatment will have an odds ratio >1.
  (In case control studies, the odds ratio refers to the odds in favour of exposure to a particular factor in cases divided by the odds in favour of exposure in controls).

- **Number needed to treat** (**NNT**) - the number of patients who needed to be treated to prevent the occurrence of one adverse event (e.g. complication, death) or promote the occurrence of one beneficial event (e.g. cessation of smoking).
  NNT=1/ARR
  Ideal NNT=1
  The higher the NNT, the less effective the treatment.
  For every *n* patients treated with the intervention one *extra* patient would be expected to benefit.

4.6. Critical appraisal of RCTs

Factors to look for:
- allocation (randomisation, stratification, confounders);
- blinding;
- follow up of participants (intention to treat);
- data collection (bias);
- sample size (power calculation);
- presentation of results (clear, precise);
- applicability to local population.

5. Systematic reviews

5.1. Mechanisms to control bias in systematic reviews

Systematic reviews provide an overview of all primary studies on a topic and try to obtain an overall picture of the results.

To avoid bias, systematic reviews must:
- contain a statement of objectives, materials and methods;
- follow an explicit and reproducible methodology.

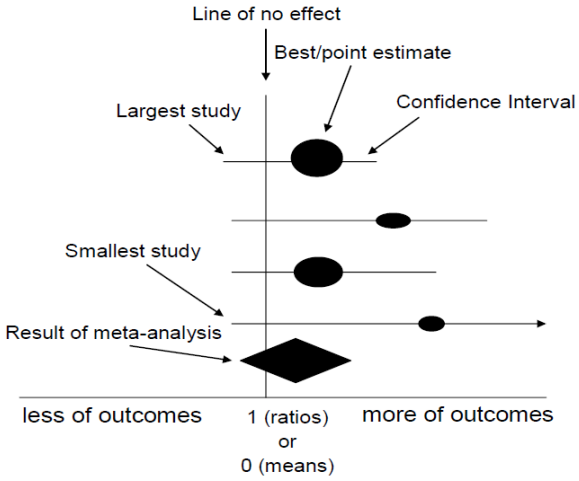
In a systematic review, all the primary studies identified are critically appraised and only the best ones are selected. A **meta-analysis** (i.e. a statistical analysis) of the results from selected studies may be included.

5.2. Forrest plot (Blobbogram)

A blobbogram or forest plot is a graphical display used to present the result of a meta-analysis. Selected studies must be tested for homogeneity, which should be >50%. A quick way to check for homogeneity is to look at the confidence intervals for each study - if they don't overlap, the studies are likely to be heterogeneous. More rigorous tests of homogeneity include $\chi 2$ (chi-squared).

If studies are homogeneous, a fixed-effect model is normally used in the meta-analysis. This means that results are only interpreted within the populations/samples in the included studies.

If studies are heterogeneous, a random-effects model is used. This means that results are interpreted across the wider population. A different underlying effect is assumed for each study and an additional source of variation is added to the model.



5.2. Advantages and disadvantages of systematic reviews

Advantages:

- allow for rigorous pooling of results;
- may increase overall confidence from small studies;
- potentially eradicate bias;
- may be updated if new evidence becomes available;
- may have the final say on a clinical query;
- may identify areas where more research is needed.

Disadvantages:

- expensive;
- time consuming;
- may be affected by publication bias - a test called **Funnel Plot** can be used to test for publication bias;
- normally summarise evidence up to two years before (due to the time required for the execution of the systematic review).

5.3. Critical appraisal of systematic reviews

Factors to look for:

- literature search (did it include published and unpublished materials as well as non-English language studies? Was personal contact with experts sought?);
- quality-control of studies included (type of study; scoring system used to rate studies; analysis performed by at least two experts);
- homogeneity of studies;
- presentation of results (clear, precise);
- applicability to local population.

6. How to critically appraise a paper

Some key questions to consider when critically appraising a paper:

- Is the study question relevant to my field?
- Does the study add anything new to the evidence in my field?
- What type of research question is being asked? A well-developed research question usually identifies three components: the group or population of patients, the studied parameter (e.g. a therapy or clinical intervention) and outcomes of interest.
- Was the study design appropriate for the research question?
- Did the methodology address important potential sources of bias?
- Does the study population represent the patient population (e.g. with regards to ethnicity)?
- Was the study performed according to the original protocol? Deviations from the planned protocol can affect the validity or relevance of a study.
- Does the study test a stated hypothesis? Is there a clear statement of what the investigators expect the study to find which can be tested, and confirmed or refuted?
- Were the statistical analyses performed correctly? The approach to dealing with missing data, and the statistical techniques that have been applied should be specified. Original data should be presented clearly so that readers can check the statistical accuracy of the paper.
- Do the data justify the conclusions? Watch out for definite conclusions based on statistically insignificant results, generalised findings from a small sample size, and statistically significant associations being misinterpreted to imply a cause and effect.
- Are there any conflicts of interest? Who has funded the study and can we trust their objectivity? Do the authors have any potential conflicts of interest, and have these been declared?
- Will the results help me manage my patients? At the end of the appraisal process you should have a better appreciation of how strong the evidence is, and ultimately whether or not you should apply it to your patients.

7. Online resources

CASP

The Critical Appraisal Skills Programme (CASP) is a programme within Workforce Development by Solutions for Public Health (SPH), a not-for-profit NHS public health organisation. CASP aims to enable individuals to develop the skills to find and make sense of research evidence. The website gives access to critical appraisal checklists which guide the appraisal of different types of study.

Centre for Evidence Based Medicine

The Centre for Evidence Based Medicine is the first of several centres around the country whose broad aim is to promote evidence-based health care and provide

support and resources to anyone who wants to make use of them. It includes a wide range of EBM resources including critical appraisal tools.

How to read a paper
Links to the series of articles that make up the book 'How to read a paper'. The articles are available online free of charge from the BMJ website.

8. Glossary

9. Example

Critical appraisal of an article using the CASP checklist for systematic reviews:
https://prezi.com/view/wzAzPZlVl0UJJtNy4Gs5/