Critical appraisal /
An introductory guide

Overview

This guide provides an introduction to critically appraising journal articles. It includes definitions of terms found in research articles, gives pointers to further reading and provides copies of CASP checklists.

In this guide

Library

June 2017

What's the purpose of critical appraisal?

There is worldwide interest in making health services more effective, to maximise the health gain they deliver. But how do we know which services are truly effective? How do purchasers know which treatments and care packages to fund? How do clinicians decide whether a particular treatment is worthwhile and affordable?

If we are to achieve the most for our population or patients, we need to base our decisions on evidence. But how do you keep up with the proliferation of literature available? One solution is to turn to review articles which summarise individual pieces of evidence. But it can still be difficult to know whether or not a review is trustworthy as evidence. Reviews themselves can draw inaccurate and confusing conclusions.

This is where critical appraisal can help. Critical appraisal skills enable you to assess systematically the trustworthiness, relevance and results of published papers. During this session, you will learn about the process and techniques used when critically appraising research papers.

Critical appraisal is a process which people can use according to their own needs. Below is a list of outcomes people have identified from knowing about and practising critical appraisal.

Being aware of evidence

- Having a greater understanding of good evidence and bad evidence (that it is possible to find and evaluate evidence to help problem-solving)
- Being aware that evidence can guide decision making
- Being aware of possible uses for research and public health findings
- Being better able to change one's mind when evidence is not supportive

Assessing and interpreting evidence: practising critical appraisal

- Being more consistent and confident in using evidence in making decisions
- Being better able to assess the reliability, results and relevance of research findings
- Understanding different functions of reviews of evidence in making decisions
- Being more discriminating in reading

Summarising evidence yourself

- Being aware of the problems of trying to summarise evidence on your own, whether on paper or in your head
- Understanding the need for appropriate and explicit methodology in reviewing literature
- Being able to use appropriate and explicit methodology in literature reviews

Critical Appraisal Learning Resources

Critical Appraisal Skills Programme (CASP) Checklists

Using ‘checklists’ can help you to evaluate the research and make a decision about whether to use that piece of research to support your practice.

The tools created by CASP from the Public Health Resource Unit are very useful checklists of 10 questions designed to help you critically appraise articles of the following types of research:

- Systematic reviews
- Randomised Controlled Trials (RCTs)
- Qualitative Research
- Economic Evaluation Studies
- Cohort Studies
- Case Control Studies
- Diagnostic Test Studies

Some of these are included in the appendices at the back of this booklet and are accessible at http://www.casp-uk.net/

You will also find links to other useful resources for practising evidence based medicine here.

Agree Instrument: Appraisal of Guidelines for Research and Evaluation

This quite comprehensive checklist, developed at St. George's, University of London, provides a framework for assessing the quality of clinical practice guidelines.

http://www.agreecollaboration.org/pdf/agreeinstrumentfinal.pdf

SIGN guideline developers' handbook

Checklists for critical appraisal

http://www.sign.ac.uk/checklists-and-notes.html

CEBMH Appraisal tools and checklists

http://www.cebm.net/index.aspx?o=1157

Glossary: Quantitative Research

Absolute Risk is the risk of a person or group developing a specific disease over a specified period of time. This can be expressed as a ratio, for example 1 in 9 (women's absolute risk of developing breast cancer over a lifetime) or a percentage, for example 11%

Absolute Risk Reduction (ARR) is the difference in the ***event rate*** between control group (CER) and treated group (EER): ARR = CER – EER.

Best point estimate is what the statisticians believe the relative risk is to the best of their ability along the ***confidence interval***. The best point estimates are the dots in the ***odds ration diagram***.

Bias is the deviation of results from the truth, due to the way(s) in which the study is conducted.

Blinding is the process of hiding the allocation of patients in a research project from the participants and/or researchers. Blinding is a key tool for eliminating bias.

Case-control Study involves identifying patients who have the outcome of interest (cases) and control patients without the same outcome, and looking back to see if they had the exposure of interest.

Case-series is a report on a series of patients with an outcome of interest. No control group is involved.

Clinical effectiveness is the extent to which an intervention (whether a treatment, procedure or service) improves the outcome for patients in practice. It is also known simply as '**effectiveness**'. Ideally, the determination of effectiveness is based on the result of one or more ***Randomised Controlled Trial***

The Cochrane Collaboration is an international endeavour in which people form many different countries systematically find, appraise and review available evidence from RCTs. These systematic reviews are then published (and kept up to date) on the Cochrane Library (http://www.thecochranelibrary.com)

Cochrane Database of Systematic Reviews (CDSR) is one of the products of the Cochrane Collaboration. It brings together all the currently available Cochrane reviews and is produced on the Internet and CD-ROM.

Cohort Study involves identification of 2 groups (cohorts) of patients, one which did receive the exposure of interest, and one which did not, and following these cohorts forward for the outcome of interest.

Confidence Interval is the range within which the true size of effect (never exactly known) Confidence intervals usually cover the results for 95% of the subjects in a re search study (the top 2 ½ and bottom 2½ are excluded).

Continuous outcome is data from outcomes measured on a continuous scale (for example average blood pressure, range of motion of a knee joint).

Controls or control group provide the comparison in an **RCT**. They receive the usual (standard) treatment (or a placebo) while the experimental group receive the treatment being tested.

Critical Appraisal is the process of assessing and interpreting evidence, by systematically considering its validity, results and relevance to your own work.

Cross-Sectional Study is the observation of a defined population at a single point in time or time interval. Exposure and outcome are determined simultaneously.

Dichotomous outcome is data from outcomes that can be divided into two categories (e.g. dead or alive, pregnant not pregnant), where each participant must be in one or other category, and cannot be in both.

Efficacy is the extent to which an intervention improves the outcome for patients under ideal circumstances.

Embase is a large biomedical database where you can find references to thousands of pieces of research in selected journals. There is a lot of drug information on this database.

Event Rate is the proportion of patients in a group in whom the event is observed. Thus, if out of 100 patients, the event is observed in 37, the event is 0.27. Control Event Rate (CER) and Experimental Event Rate (EER) are used to refer to this in control and experimental groups of patients respectively.

Evidence-Based Health Care extends the application of the principles of Evidence-Based Medicine (see below) to all professions associated with health care, including purchasing and management.

Evidence-Based Medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.

Forest Plot is another name for an ***odds ratio diagram.***

Healthcare Databases are collections of abstracts and references to journal articles (such as Medline, CINAHL and PsycINFO) You can use databases to track down journal articles on specific topics.

Homogeneity means 'similarity'. Studies are said to be homogenous if their results vary no more than might be expected by the play of chance. The opposite of homogeneity is **heterogeneity**.

Heterogeneity occurs when there is more variation between the study results (in a systematic review) than would be expected to occur by chance alone.

Incidence is the proportion of new cases of the target disorder in the population at risk during a specified time interval.

Intention-to-treat analysis is a method of analysis for randomised trials in which all patients randomly assigned to one of the treatments are analysed together, regardless of whether or not they completed or received that treatment in order to preserve randomisation.

Line of no effect is the vertical line in the middle of an odds ratio diagram, which represents the point where treatment and control have the same effect i.e. there is no difference between the two.

Literature search is where you search the databases and other resources (depending on how thorough your search is!) to find articles on a particular topic.

Medline is an electronic database which summarises thousands of pieces of biomedical research literature, in selected journals. It is available over the Internet.

Meta-analysis is a statistical technique which summarises the results of several studies into a single estimate, giving more weight to results from larger studies.

Number needed treat (NNT) is one measure of a treatment's clinical effectiveness. It is the number of people you would need to treat with a specific intervention (e.g. aspirin for people having a heart attack) to see one occurrence of a specific outcome (e.g. prevention of death). Ideal NNT is 1. The higher the NNT, the less effective the intervention.

Odds is a ratio of the number of people who have something happen to them to the number of people that thing does not happen to.

Odds ratio (OR) diagram is a visual way of showing the results showing relative risk (RR) and confidence intervals. If the confidence interval crosses 1 then the effects of the intervention are no different from those of the control. If the result is greater (or less) than 1, then the effects of the increases (or reduces) the outcome compared to those of the control treatment. Note that the effects being measured may be adverse (e.g. death, disability) or desirable (e.g. stopping smoking).

P-value shows the likelihood of an outcome occurring by chance. A widely accepted threshold p-value is .05, i.e. if the p-value is less than .05, the observed result is said to be statistically significant, and there is 95% certainty that the outcome did not occur by chance.

Placebo is an inactive treatment often given to controls in trials. The placebo is delivered in a form which is apparently identical to the active treatment being tested in the trial, in order to eliminate psychological effects on the outcome.

Placebo effect is when patients given the placebo still show clinical improvement compared to those not given any intervention.

Pooled effect is where the results of similar ***RCT***s are brought together.

Prevalence is the number of people who have a disease or given condition at a given point in time in a defined population, or the total number of people who had a condition during a specified period.

Probability is the likelihood of something occurring by chance. Probability is expressed by ***p-values***.

Publication bias results from the fact that studies with 'positive' results are more likely to be published.

Randomised controlled trial (RCT) is a trial in which the subjects are randomly assigned to two groups: one (the experimental group) receiving the intervention that is being tested, and the other (the comparison group or controls) receiving an alternative treatment. The two groups are then

followed up to see if any differences between them result. This helps people assess the effectiveness of the intervention.

Relative risk is the incidence of an event in the intervention group divided by the incidence of the same event in the control group. If the result is greater than 1 it means that the event is more likely to occur in the intervention group and if it is less than 1 the event is more likely to occur in the control group.

Review is any summary of the literature.

Systematic review is a *review* in which evidence on a topic has been systematically identified, appraised and summarised according to predetermined criteria. (Some people call this an 'overview').

Validity refers to the soundness or rigour of a study. A study is valid if the way it is designed and carried out means that the results are unbiased – that is, it gives you a 'true' estimate of clinical effectiveness.

Glossary :Qualitative research

Action research - occurs when researchers design a field experiment, collect the data, and feed it back to the activists (i.e. participants) both as feedback and as a way of modelling the next stage of the experiment

Analytic induction: use of constant comparison specifically in developing hypotheses, which are then tested in further data collection and analysis.

Case study: an in depth study of a case or cases (a 'case' can be a programme, an event, an activity, an individual), studied over time using multiple sources of information (e.g. observations, documents, archival data, interviews). Can be exploratory, explanatory, or descriptive, or a combination of these.

Consensus methods: include Delphi and nominal group techniques and consensus development conferences. They provide a way of synthesising information and dealing with conflicting evidence, with the aim of determining extent of agreement within a selected group.

Constant comparison: an iterative method of content analysis where each category is searched for in the entire data set and all instances are compared until no new categories can be identified.

Content analysis - a form of analysis which usually counts and reports the frequency of concepts/words/behaviours held within the data. The researcher develops brief descriptions of the themes or meanings, called codes. Similar codes may at a later stage in the analysis be grouped together to form categories.

Credibility - refers to the quality or trustworthiness of a piece of qualitative research. It can refer specifically to the extent to which the findings and explanations within a qualitative report are recognised and understood by the participants, but can also be extended to include considerations of all aspects of the study.

Data analysis - a systematic process of working with the data to provide an understanding of the research participant's experiences. While there are several methods of qualitative analysis that can be used, the aim is always to provide an understanding through the researcher's interpretation of the data.

Discourse analysis - the linguistic analysis of naturally occurring connected speech or written discourse. It is also concerned with language use in social contexts, and in particular with interaction or dialogue between speakers

Emerging themes - concepts (explanatory ideas) are identified from the data in the first stages of analysis and given a label or code that describes them. Concepts which are closely linked in meaning can be formed into categories: categories which have similar meanings can be brought together into a theme. The term 'emerging themes' refers to the development or 'emergence' of themes from the data and this overall method of analysis is referred to as 'thematic analysis'

Ethics - research ethics relate to the standards that should be upheld to guard participants from harm or risk. Ethical considerations should be made at each stage of the research design and include informed consent, voluntary participation and respect for confidentiality.

Ethnography - a qualitative research methodology that enables a detailed description and interpretation of a cultural or social group to be generated. Data collection is primarily through participant observation or through one-to-one interviews. The importance of gathering data on context is stressed, as only in this way can an understanding of social processes and the behaviour that comes from them be developed.

Field notes: a collective term for records of observation, talk, interview transcripts, or documentary sources. Typically includes a field diary, which provides a record of chronological events and development of research as well as the researcher's own reactions to, feeling about, and opinions of the research process

Focus groups - used to elicit the views of a group (usually around 6 to 10 individuals) who have common experiences or interests. They are brought together with the purpose of discussing a particular subject, under the guidance of a facilitator.

'Framework': a method of qualitative data analysis involving five key stages: familiarisation, identifying a thematic framework, indexing, charting, and mapping and interpretation

Grounded theory - a qualitative research methodology with systematic guides for the collection and analysis of data, that aims to generate a theory that is 'grounded in' or formed from the data and is based on inductive reasoning. This contrasts with other approaches that stop at the point of describing the participants' experiences.

Hawthorne effect: the impact of the researcher on the research participants or setting, notably in changing their behaviour.

Holistic - exploration of a research question multi-dimensionally, exhaustively and in its entirety, preserving the complexity of human behaviour.

Induction (an inductive process) - logical thought process in which generalisations are developed from specific observations: reasoning moves from the particular to the general. E.g. Grounded theory uses an inductive process, i.e. explores new, unforeseen issues that emerge during the research and theories develop/hypotheses are generated from the data.

Interpretative - exploration of the human experiential interpretation of any observed phenomena. Enables researchers to gain a better understanding of the underlying processes that may influence behaviour

Interviewing: a data collection strategy in which participants are asked to talk about the area under consideration. Interviews can be:

- Unstructured/in-depth - the researcher asks the respondent a general question regarding the area of interest and asks them to tell their own story.
- Semi-structured - the interviewer has a slightly more focused agenda than in an unstructured interview. Questions are phrased to allow the participants to tell the story in their own way and an interview guide is used to ensure information is gathered on areas of interest to the researcher.
- Structured - an interview in which the questions are pre-determined and asked to all subjects. Closed questions are used with limited response choices.

Iteration (an iterative process) - relates to the process of repeatedly returning to the source of the data to ensure that the understandings are truly coming from the data. In practice this means a constant process of collecting data, carrying out a preliminary analysis, and using that to guide the next piece of data collection and continuing this pattern until the data collection is complete.

Natural setting (naturalistic research) - the normal environment for the research participants for the issues being researched.

Observation - a strategy for data collection, involving the process of watching participants directly in the natural setting. Observation can be participative (i.e. taking part in the activity) or non-participative (the researcher watches from the outside)

Phenomenology - an approach that allows the meaning of having experienced the phenomenon under investigation to be described, as opposed to a description of what the experience was. This approach allows the reader to have a better understanding of what it was like to have experienced a particular phenomenon.

Reflexivity - the open acknowledgement by the researcher of the central role they play in the research process. A reflexive approach considers and makes explicit the effect the researcher may have had on the research findings.

Respondent validation - refers to seeking the participants' views of the initial interpretations of the data. The aim is not to ensure that the researcher are in agreement as to the meaning of the data, but that the researcher has the opportunity to incorporate the participants' responses into the analysis.

Sampling - the process of selecting participants to take part in the research on the basis that they can provide detailed information that is relevant to the enquiry.

Purposive sampling - the selection of participants who have knowledge or experience of the area being investigated. Theoretical sampling - a sampling strategy in which the selection of participants is guided by the ideas that are emerging from the data analysis.

Saturation - the point at which no further themes are generated when data from more participants are included in the analysis. The sampling process can be considered to be complete at this point.

Transferability: means that the research findings can be transferred from one context to similar situations or participants.

Triangulation - process by which the area under investigation is looked at from different (two or more) perspectives. These can include two or more methods, sample groups or investigators. Used to ensure that the understanding of an area is as complete as possible or to confirm interpretation through the comparison of different data sources.

Further Information and Reading

Websites

CASP (2014) *Critical Appraisal Skills Programme*

Available at http://www.casp-uk.net/

[Accessed: May 2016]

(Provides links to CASP critical appraisal checklists in PDF format)

BMJ Learning Understanding statistics module

Available via NHS OpenAthens at: http://learning.bmj.com/

Informed Health Online (c2006) *Evidence-based medicine*.

Available at: http://bit.ly/22gbOgs [Accessed: May 2016]

(Easy to read overview of the Evidence-Based Medicine process)

Oxford Centre for Evidence-based Medicine (2009*) Levels of Evidence and Grades of Recommendation*

Available at: http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ [Accessed: May 2016]

(Detailed breakdown of the levels of evidence)

Books and Journal Articles

Ajetunmobi, O. (2002) Making sense of critical appraisal. London: Hodder Arnold

(A clear guide to statistical analysis)

Crombie, I.K. (2002) The pocket book to critical appraisal: a handbook for health care professionals. London: BMJ Publishing Group

(A short book detailing criteria for critical appraisal)

Everitt, B.S. (2003) Medical statistics from A to Z: a guide for clinicians and medical students. Cambridge: Cambridge University Press

(A handy reference book for those statistics terms)

Gadda, S. (2006) Tips on..critically appraising a paper BMJ Careers 333 (16), pp.109 (A quick guide to critical appraisal)

Gosall, N. K. (2012) The doctor's guide to critical appraisal 3rd ed. London: PasTest

Greenhalgh, T. (2010) How to read a paper: the basics of evidence based medicine. 4th ed. London: BMJ Publishing Group

(One of the best known books on the subject with extensive coverage of critical appraisal. Also available electronically with your NHS Athens password)

Guyatt, G et al (eds) (2008) Users' guides to the medical literature: a manual for evidence-based clinical practice. 2nd ed. London : McGraw-Hill Medical.

(A classic series of papers on appraising the research literature, originally published in JAMA. Also available online from JAMA with an NHS Athens account)

Sackett, D. et al (2000) Evidence-based medicine: how to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone.

(The classic text on EBM, with a good coverage of critical appraisal)