National Council for NCOR Osteopathic Research

Evidence-based practice tutorial – Critical Appraisal Skills

Earlier evidence based practice tutorials have focussed on skills to search various useful sites on the internet for evidence. Anyone who has tried searching will be familiar with the deluge of information that is available. Once the information has been found, what precisely should be done with it and how straightforward is it to try and discriminate sound and valuable research from that that is very limited in both quality and applicability? The following tutorial is intended to be a basic introduction to critical appraisal; this will be followed in Part II with a more formalised checklist system that will allow the reader to discriminate more quickly when they are familiar with the various components that form a research study. Part III will look at methodological quality and Part IV will look at statistics. It is always important to remember that many poor studies are published each year; their claims should be discounted.

What does critical appraisal mean?

This is the process by which a reader can evaluate a piece of written material and assess whether it possesses validity (i.e. is it close to the truth) and applicability (i.e. is it clinically useful). If research is being examined, critical appraisal skills are vital to decide whether the research has been well conducted and whether, ultimately, the results of the research can be implemented into our everyday practice for the benefits of our patients. Critically appraising and reviewing a paper is essentially a process to look for information that is of value.

Most research papers begin with an abstract, which summarises what the paper has attempted to investigate. The title and abstract will give a sound indication about whether the paper is likely to be relevant to your area of interest and how interesting the results are likely to be. The main body of the paper is then organised in the IMRaD format:

- Introduction
- Method(s)
- Results
- Discussion

Any research needs to be set in context and for this reason the introduction will normally include a review of previous work that has been carried out in the same subject area while trying also to highlight any gaps in the knowledge base for a particular topic. This section can also underline the clinical importance of a particular piece of research by including information about the biological, clinical, cultural, epidemiological and economic impact of the subject being investigated in addition to morbidity considerations. The introduction should draw to a close with the hypothesis that it intends to test included as a clear statement by the authors. If the hypothesis being tested is presented in a negative manner, it is known as a null hypothesis.

Method(s):

Information contained in the methods section will give a significant indication about the quality of a piece of research. This section will inform the reader how the study was carried out and how the results of the study were analysed. Information about how the study was carried out should include:

- who was involved in the research (research subjects)
- how they were recruited (e.g. by advertisements, using a particular practice etc)
- what were the inclusion criteria (e.g. age, sex, ethnicity, weight, agreed diagnostic criteria etc).

Information concerning the inclusion criteria will give the reader an indication about how generalisable the results will be to the wider population, i.e. how accurately the study group reflects the wider population intended to receive a particular type of treatment.

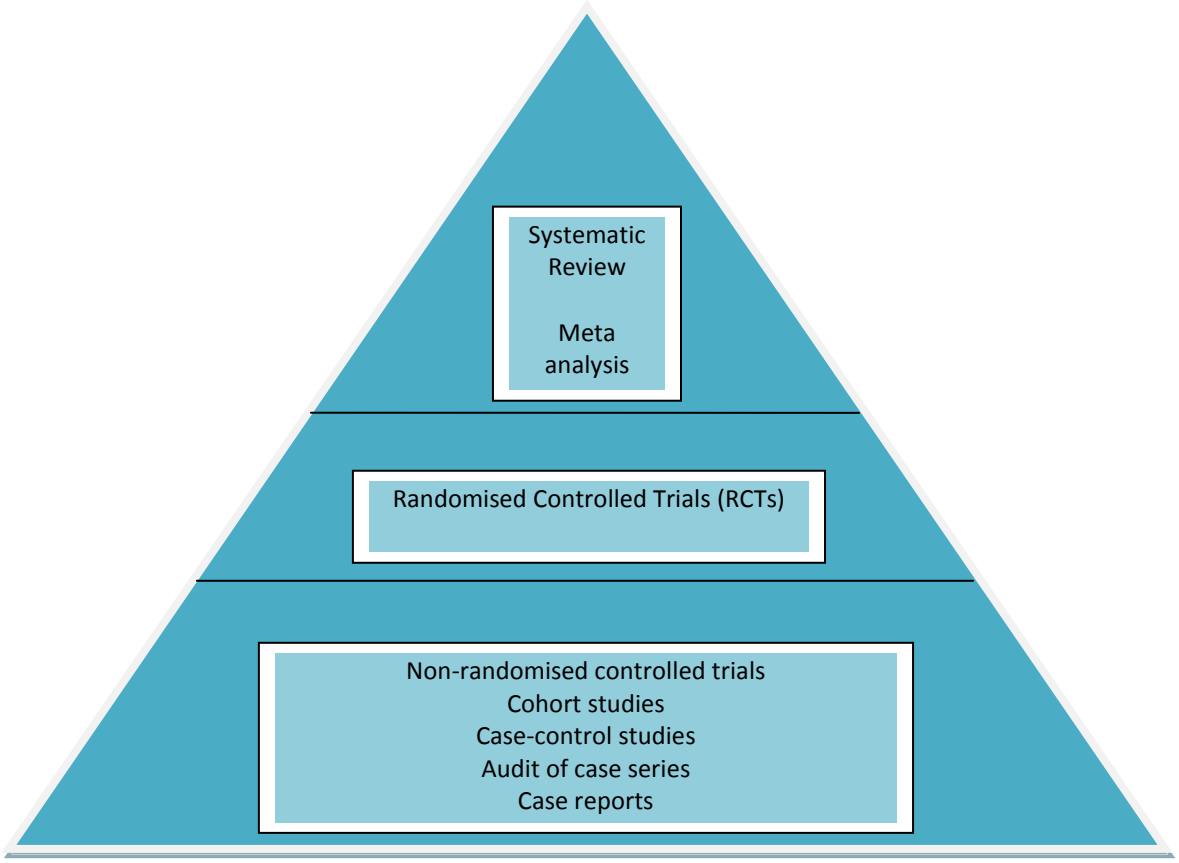
The chosen research method may be described quite briefly, but it is likely to be widely referenced for the reader to gather more extensive information should they choose to. Information should also be included in the methods section concerning how measurement procedures have been standardised, in what manner measurements have been made (e.g. particular technical instruments or measures) and the processes by which data has been recorded for later analysis. The structure of any questionnaires used to gather data should also be described. It should be made clear whether a questionnaire has been validated and the manner in which it has been tested to ensure reliability and validity.

A wide variety of research methods and styles exist. Research can be described as qualitative or quantitative; it can also be described as either primary or secondary. Primary studies report research first hand, whereas secondary studies consist of summaries or analyses of primary studies. Examples of primary research are experiments (e.g. involving either animals or human volunteers), clinical trials (e.g. investigating the effect of a treatment intervention on a group of patients and then following them closely over a period of time) and surveys (e.g. a particular area of

interest is measured in a group of professionals, patients or other targeted group of individuals).

Secondary research, by comparison, consists of overviews (e.g. non-systematic reviews, systematic reviews and meta-analyses), clinical guidelines, decision analyses and economic analyses.

A hierarchy of research evidence exists where the relevance of each type of evidence is evaluated by the wider research community:



This can be expanded further to show how various levels of evidence are assessed by external agencies e.g. the National Institute for Clinical Excellence (NICE).

Level	Evidence
Ia	Evidence from systematic review and meta-analysis of randomised controlled trials
Ib	evidence from at least one randomised controlled trial
IIa	evidence from at least one controlled study without randomisation
IIb	evidence from at least one other type of quasi-experimental study
III	evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case control studies
IV	evidence from expert committee reports or opinions and/or clinical experience of expected authorities

It is important for the reader to decide whether the research method chosen is the most appropriate to answer the hypothesis being investigated.

Results:

The results section describes what the researchers found; these findings are normally presented in a table. The data should be presented in a logical manner with fuller explanations present in the accompanying text. The text should highlight the key findings in the results, but will tend to give the researchers' interpretation of the findings. When looking at the results section, it can be valuable to refer back to the original research question/hypothesis to assess whether the results truly address this. If the original hypothesis has not been addressed, the question must arise whether this is because the researchers have failed to gather appropriate data or the findings have not been what were anticipated. Any inconsistencies in the data should be apparent in the results section.

Discussion:

This section focuses on considering the implications of the study's findings. The extent to which research is generalisable bestows a value on the research. A common criticism of some qualitative research is that it is not widely generalisable and focuses too exclusively on the environment in which it has been carried out. This is a less common criticism of quantitative research.

It can also be important to examine negative results. It is unfortunate that many journals persist in refusing to publish negative findings. This prevents wide dissemination of all research that has been conducted and can result in unnecessary research being repeated. This presents a considerable ethical problem; it can result in volunteer subjects repeatedly participating in studies that have been shown to fail and wastes goodwill and other valuable resources in the process.

Tools to help critical appraisal

Critiquing research reports can vary slightly depending on the type of research e.g. qualitative or quantitative and on the research design e.g. a case report or a randomised controlled trial (RCT). However, common features exist to help to critique all research studies. The table given below will attempt to highlight some of the main areas which need to be considered for a randomised controlled trial design.

QUESTION	CONSIDERATION
Does the introduction and literature review adequately place the research question in context?	Is the material included in the literature review relevant to the research question?

Has the research hypothesis been clearly stated and is it appropriate to the research question and supporting literature?	Are the key terms in the study well defined?
Has the research study stated a clear and focused question?	Is the population that has been studied clear to the reader? Is the intervention administered clear? Are the outcomes of the study clear?
Is the research design chosen appropriate to answer the research question?	What alternatives, if any, could have been chosen?
Are the methods and procedures clearly described in sufficient detail?	Could the study be easily replicated from this information?
Consider the research study participants.	What were the inclusion and exclusion criteria? Are the selected participants representative and appropriate to the study? Are the participants properly orientated and well motivated? What is their understanding of the task involved in being part of the study? Are their instructions clear and precise? Have sufficient numbers of participants been selected i.e. is the sample size (N) appropriate to give the research study statistical power? Was a power calculation performed to determine the sample size and minimise the results occurring being due to chance occurrence(s)? How have the participants been allocated to intervention and control groups (in RCTs)? Has the selection process been truly random? What method of randomisation was used e.g. computer/telephone/envelopes? Was a method used to balance the randomisation e.g. stratification? Are there any differences between the groups at the beginning of the trial? Could any of these differences have affected the outcomes (i.e. acted as confounding factors? Has participant attrition occurred? (i.e. have patients dropped out of the study). If yes, does this bias the sample?
Consider the blinding processes that have been used.	Were all the personnel involved with the trial e.g. researchers, support staff, participants blinded? Was blinding possible for the trial? Can observer bias be identified? Was blinding necessary for the trial? Has every effort been made to achieve blinding?

How was the data collected?	Is the independent variable being assessed appropriate to the research question? Are the levels of independent variable appropriate? Is the dependent variable appropriate to the study? Was data collected in all groups in the same manner and at the same time intervals? Was the data collected using validated, calibrated and reliable tools/measuring equipment? Were all participants followed up at the end of the study? Was there any loss to follow up? Were the outcomes of the participants analysed according to the groups to which they were originally allocated? (i.e. was an intention to treat analysis used). Has any bias been evident in the data collection?
What are the results of the study?	How are the results presented? This could be as: - a measurement e.g. a median or mean difference - a proportion of people experiencing a particular outcome - as a graph - as a bar or pie chart Are the results clearly labeled and accurately presented? Are the results precise? Are the results large enough? Are the results both clinically and statistically significant? Can a decision be made from the results? Has a confidence interval been reported? If yes, would your decision about whether to use this intervention be the same at the highest as well as the lowest limit of the confidence interval? Has a p-value been stated? Can the results be clearly stated in one sentence?
Have high ethical standards been adhered to at all stages of the study?	Has appropriate ethical approval been sought and given prior to commencement of the study? Have the dignity and rights of all participants been respected throughout the trail and in the planned dissemination of the results?
How relevant are the outcomes of the trial?	Are the trial results generalisable to the wider population or are they just relevant to the participants in the study? Are the outcomes relevant to other people surrounding the trial participants e.g. family members, carers, policy makers, other health care professionals? Are there any cost benefits to the trails results? Are there any cost implications?
Discussion of the study findings.	Does the discussion of the results relate to the research question? If not, why not? Have the results been interpreted correctly according to the results presented? Have the results been placed in an appropriate context?
Are the references accurate?	Do the references match the citations in the text?

Could the study be improved if it was repeated?	What could be done to improve the design of the study?

Research Methods and Basic Statistics

A number of different methods can be employed depending on the type of research being used. Research methods can be divided into qualitative and quantitative approaches. A very brief overview will be given and a more in depth look at each method will be described in later articles.

Quantitative research methods include:

Case reports

This describes the medical history of a patient and is communicated in a narrative fashion. This is a useful way to communicate details about unusual patients. Writing a case report can be described as the first step in communicating patient information. Further information on writing a case report can be found at http://careerfocus.bmjjournals.com/cgi/content/full/327/7424/s153-a

Case series

This can be the natural sequel to a case report. A case series is comprised of information concerning a number of patients who experience a particular condition. Various aspects of their care can be examined including their treatment regime or any reactions (adverse or other) to that treatment.

Case control studies

In this type of study patients with a particular condition or disease are identified and are matched with a control group of patients who may have no disease or a different disease, alternatively the control group can be composed of patients' relatives. Information concerning past medical history is recorded from examination of medical records or by verbal reporting of past medical history. A relationship between a past exposure to a causal agent of a certain disease is then explored from this information. Case control studies are fundamentally examining the aetiology of a disorder or what makes a particular patient group different; they are not concerned with the therapeutic intervention for a disease.

Cohort studies

Cohort studies can take a considerable period of time to conduct. They examine at least two (or more) groups of subjects and find out what happens to them in the future. The follow up time in cohort studies has generally been measured in years. Subjects in cohort studies may or may not have a disease when the group is selected for

monitoring; the cause of a disorder or disease is usually the main concern of this type of study.

Cross-sectional Studies.

This is used to estimate the prevalence of a disease or the prevalence of an exposure to risk factors or both. It is important to distinguish between prevalence and incidence. Prevalence describes the overall proportion of a population that experience a disease; incidence describes the number of new cases of a disease each year.

Randomised controlled trials (RCTs)

Randomised controlled trials are described as the "gold standard" in medical research. They are suitable for testing interventions concerned with treatment or prevention, but give no information about the context of a trial or the patients' experience of treatment.

Participants in RCTs are **randomly** assigned to one treatment intervention (e.g. osteopathic treatment) or another (e.g. taking non-steroidal anti-inflammatory medication). The random assignment can be achieved in a number of ways e.g. patients can be given an envelope containing the type of intervention they will receive or, more appropriately, they can be assigned by telephoning an allocation centre. Interventions can be assigned according to a number of **blinding**/masking regimes: Single blind: Patients do not know the type of treatment they are receiving.

Double blind: Patients and investigators do not know the type of treatment being received.

Randomised controlled trials can also utilise a **placebo** intervention. A placebo is an inactive compound which looks, tastes and smells the same as an active compound in a pharmacological study. Placebo or sham interventions can also be used when researching complex interventions e.g. acupuncture.

The patients in RCTs are followed for a designated period of time and specified outcomes are measured e.g. changes in levels of pain or mobility.

Qualitative research:

Judith Preissle described this as "a loosely defined category of research designs or models, all of which elicit verbal, visual, tactile, olfactory and gustatory data in the form of descriptive narratives like field notes, recordings, or other transcriptions from audio and videotape and other written records and pictures or films."

Qualitative research uses a variety of methods e.g. open-ended interviews, naturalistic observation, focus groups, self-reflective exercises, document analysis, life histories and descriptive analysis. The researcher is often described as the "instrument" in qualitative research – present to facilitate the process, rather than to conduct measurements and make evaluations to a pre-agreed format.

Fewer people tend to be studied in qualitative research since it can be very time consuming, not only in terms of contact time with a subject but also taking into account time to transcribe the recorded data. Data can be less generalisable than with qualitative studies.

Statistical Analysis of the Literature

When data has been gathered, it needs to be analysed statistically. Qualitative and quantitative data is examined differently. The basic examination of quantitative data will be considered here. Certain characteristics of a set of numerical data can be summarised in a succinct numerical form; the values produced are described as summary statistics. Different types of data (or variables) will be encountered in statistics. They will differ in their 'scale of measurement' – i.e. in terms of just what can be ascribed to any numerical values they have. Different types of analysis are appropriate for different types of variable; it is important, therefore, to identify the correct type of variable. Statistical analyses always appear in published research papers; consideration will be given here to quantitative data (also known as interval or scale or metric data).

Quantitative (or interval or scale or metric) discrete variables.

This describes a quantity that is measured on a well-defined scale with some clear units of measurement e.g. numbers of cars crossing a bridge in a minute.

Quantitative (or interval or scale or metric) continuous variables

This describes a measurement that is not restricted to taking certain numbers alone - e.g. whole numbers – but the value can be measured to any degree of precision and any two values can be differentiated. Examples are birth weight, height and blood pressure.

Overlap in Definition

It could be argues that certain discrete variables which can take a very large number of possible values are better thought of as continuous for the purpose of analysis. Just where to draw the line between the discrete and continuous data is not always easy.

Measures of Central Tendency

One of the basic measures that will be applied to research will come under the category of a measure of central tendency. This encompasses:

The mode: the most common reading. This is not used very often as it is not particularly useful. However, it is the only measure for summarising categorical data.
The median: the value which splits a sorted set of data in the middle so that half the values are smaller than the median value and half are larger than the median. It is a resistant measure that is unaffected by unusual data values.
The mean: the value obtained when the sum of all values is divided by the number of values. The mean can be affected by an occasional atypical value in a set of data.

Spread or Dispersion

There are a number of ways to measure the spread of data values.

The Range

This is the simplest measure to calculate, but probably the least useful. It focuses on the most unusual values in a set of data differentiating between the minimum and maximum values present and expressed as a single digit. The value is also dependent on the size of the sample; as the sample size increases, the range is likely to increase.

The Interquartile Range

This shows a range of data values spilt into four equal parts. The lower and upper quartiles express the smallest quarter of values in a set of data and the largest quarter of values respectively. This approach can be used when outliers are present in a set of data.

The Variance

This is the average squared deviation of the data points from the mean. It is usually expressed as $\sigma 2$

The Standard Deviation

The standard deviation is used to describe data. It can be calculated using the value obtained for the variance:

Standard Deviation = $\sqrt{}$variance

Alternatively, a scientific calculator with a statistics mode can be used to calculate the standard deviation using '*s' or 's* n-1' or *σ* n-1'. The total of all standard deviations will be zero.

The Standard Error of the Mean

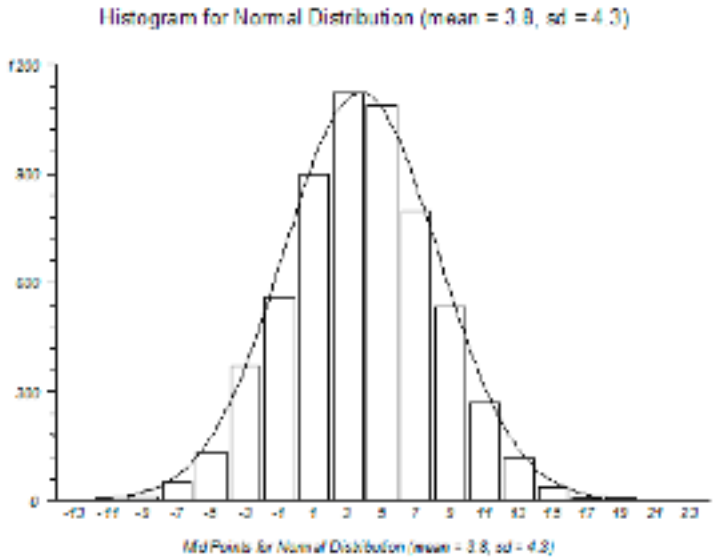
This can be used to estimate a characteristic in a sample population. It can be calculated:
Standard Error = standard deviation

$$\sqrt{n}$$

where 'n' is the value of the sample size.

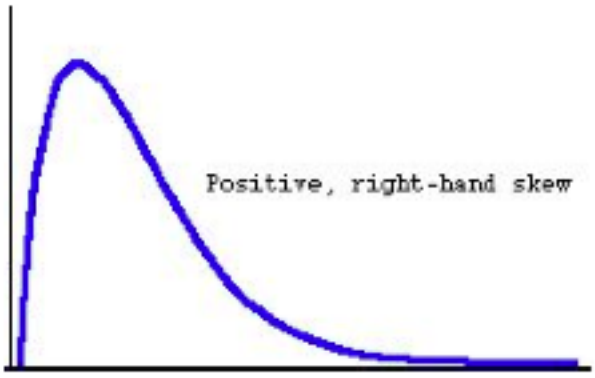
Visual Presentation of Data

The distribution of data can be expressed visually as shown below:



© www.statsdirect.com

A central peak will be seen at the top of symmetrically distributed data. This data can be said to be normally distributed. If a lack of symmetry is seen in the shape of a curve it is said to be skewed. If data is "positively skewed" the tail on the right hand side will be stretched out, as shown in the diagram:



© www.uwyo.edu

If the data is "negatively skewed" the tail on the left will be stretched out.

Probability

This is commonly described as a '*p'* value. It represents the probability that any particular outcome in a study could have occurred by chance. *P* values are commonly described in terms of having a value less than one in 20 which is expressed as $p < 0.05$; this is the level at which results are said to have gained "statistical significance." An alternative value for probability is les than one in one hundred and this is expressed as $p < 0.01$ which is described as "statistically highly significant."

Hypothesis Testing:

In any research study two hypotheses are described:

A null hypothesis i.e. that there is no difference or no relationship between what is being tested.

An alternative hypothesis i.e. that there is a difference or there is a relationship between what is being tested.

The null hypothesis will be believed until evidence can be found that shows that it is untrue or that there is a very low probability (i.e. very low *p* value) that it is true. The *p* value is the probability of observing a sample that is as extreme as or more extreme than the one being investigated given that the null hypothesis is true. An assessment is made whether the *p* value is smaller than some pre-determined small probability i.e. the significance level, which is typically pre-set at values of 0.05 and 0.01. The smaller the *p* value, the stronger the evidence against the null hypothesis (i.e. that the null hypothesis should be rejected).

Confidence Intervals

This expresses the range of values within which you are confident a particular characteristic of a population is expected to lie. The range is based on the estimate of that characteristic from the sample; it also takes into account the standard error of the estimate as an indication of the reliability of the estimate.

Number Needed to Treat (NNT)

This statistic is appearing more frequently in the analysis section of papers. It denotes the number of patients that need to be treated to obtain a positive outcome in one patient. The smaller the value for the NNT for a particular intervention indicates the effectiveness of that intervention.

Number Needed to Harm (NNH)

This statistic describes the number of patients that would need to be treated to get side effects from an intervention. If the NNH is smaller than the NNT then the intervention may be doing more harm than good.

An enormous variety of statistical tests are available for specific purposes and a vast array of computer software can assist with calculations. Further statistical tests will be covered in greater detail in later tutorials.

Author: Carol Fawkes, NCOR Research Development Officer.