Chiropractic Research Council

CRITICAL APPRAISAL SKILLS

One of the challenges with sourcing materials on the website is deciding which are valid and useful and which ones should be put to one side. The role of critical appraisal is to understand the process of determining sound and valuable research from that which is of poor quality and is not applicable to the question it is trying to answer.

What is critical appraisal?

Critical appraisal is a process by which a reader can evaluate the usefulness of a piece of written material and assess it for validity (how close it is to the truth) and applicability (whether it is clinically useful). Critical appraisal permits an objective analysis of the value of research and in so doing, the usefulness of its incorporation into clinical care.

The abstract of an article will give a good indication as to whether the research is relevant. It sets out the research question, i.e. what the research has attempted to investigate, the principal findings and the relative applicability to practice.

Research papers are usually organised in a format of Introduction, Method, Results and Discussion (IMRaD).

Introduction

The introduction usually sets the research in context. It may look at previous work in the same subject area, their outcomes and any gaps that still exist. It will also seek to set out the clinical importance of a subject area by including information about biological, clinical, cultural, epidemiological and economic factors. The introduction should end with an outline of the hypothesis that it intends to test. A hypothesis presented in a negative manner is a known as a null hypothesis.

Method

The quality of the method used in research is generally a reliable indicator as to its quality. The method informs the reader as to how the study was performed and how the results were analysed. It is important that three elements are considered:

- The research subjects
- How they were recruited
- The inclusion criteria.

Knowing how research subjects were recruited and which criteria were applied gives a valuable indication as to how valuable the research can be applied to a population. For example, the outcomes of a research study looking at those over 65 cannot be applied to an adolescent population.

The method should be widely referenced so that the reader could reproduce the conditions of the research if required. Therefore it is important to include information about measurement methods (equipment and parameters), the recording process and now the data was analysed. If questionnaires were used, a description of the questions asked must be included and itsreliability and validity outlined.

Research can be described as qualitative or quantitative, primary or secondary. Primary research explores an area for the first time; secondary research analyses previous studies. Examples of primary research are experiments, clinical trials and surveys. Secondary research consists of overviews (non-systematic and systematic reviews and meta-analyses), clinical guidelines, decision analyses and economic analyses.

A hierarchy of evidence exists and reflects how research value is perceived by the research community:

NICE (National Institute for Health and Clinical Excellence) further subdivides levels of evidence:

Level	Evidence
1a	Evidence from systematic review and meta-analysis of randomised controlled trials
1b	Evidence from at least one randomised controlled trial
2a	Evidence from at least one controlled study without randomisation
2b	Evidence from at least one other type of quasi-experimental study
3	Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case control studies.
4	Evidence from expert committee reports or opinions and/or clinical experience of expected authorities.

In critical appraisal, the reader must consider whether the most appropriate research method has been selected to address the hypothesis.

Results

The results section of any paper describe what the researchers found. Typically, this is presented in tabular or graphical form, but is accompanied by text which further explains the results. It is important when critically appraising articles to determine whether the results correlate with the question being investigated; if they don't, it could be because the researchers have gathered inappropriate data or that the findings were not as anticipated. Critical appraisal involves looking for inconsistencies in results.

Discussion

The discussion section of a paper explores the implications of the findings of the study and the value of the results that have been obtained. One of the criticisms of qualitative research is that it focuses on too narrow an area (for example, the environment) and is not generalisable to real practice. This is less often the case in quantitative research.

Many journals shy away from publishing the negative results of studies, although the implications of negative results can also be important. This is often a criticism of powerful pharmaecuetical companies, which only publish the results that support its hypothesis. There are ethical implications to not publishing the negative outcomes of studies.

Critical appraisal tools

Critical appraisal of qualitative and quantitative research varies, as does that of different types of study (e.g. case report and RCT). However, there are a number of tools used in critical appraisal that can be applied across the board. Considerations in randomised controlled trial appraisal can be seen in the table below:

Question	Consideration
Does the introduction and literature review place the research question in context?	Is the material in the literature review relevant to the research question?
Has the research hypothesis been clearly identified and is it related to the supporting literature?	Are the key terms in the study well defined?
Has the research study stated a clear and focused question?	Can the reader identify the population being studied? Is the proposed intervention clear? Are the outcomes clear?
Is the design of the study appropriate to the question?	Are there any alternative methods that could have been chosen?
Are the methods and procedures described in sufficient detail?	Is it possible to easily replicate the study from the information provided?
Who are research study participants?	What are the inclusion and exclusion criteria? Are the participants appropriate to the study? Are participant well-motivated and properly oriented? Do they understand the nature of their participation in the study? Are their instructions clear and precise? Is the sample size large enough to give the study sufficient statistical power? Was a power calculation performed to determine the sample size and minimise the results being due to chance occurrence? How have the participants been allocated to intervention and control groups? Has the selection process been truly random? What method of randomisation has been used (computer, telephone, envelopes)? Was a method used to balance the randomisation (eg. stratification)? Are there any differences between the groups at the beginning of the trial? Could any of these differences affected the outcomes? Has participant attrition occurred?
Consider the blinding processes that have been used	Were all participants in the trial blinded (researchers, support staff, participants)? 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 valirable being assessed appropriate to the research question? Are the levels of independent variable appropriate? Is the dependent variable appropriate in this study? Was all data in all groups collected in the same manner and at the same time intervals? Waas the data

	collected using calibrated, validated and reliable 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? Has any bias been evidence in the data collection?
What are the results of the study?	How are the results presented (measurement, proportion, graph, bar/pie chart)? Are the results clearly labelled and accurately presented? Are the results precise? Are the results large enough? Are the results both clinically and statistically significant? Has a confidence interval been reported? As a p-value been stated? Can the results be clearly be stated in one sentence?
Have high ethical standards been adhered to at all stages of the study?	Has appropriate ethicval approval been sought and given prior to commencement of the study? Have the dignity and rights of the participants been respected at all stages of the study?
How relevant are the outcomes of the trial?	Are the results generalizable to the wider population or are they just relevant to the participants of the study? Are the outcomes relevant to other people surrounding the trial participants (family members, carers, policy makes, other health care professionals) Are there any cost benefits to the trial results? Are there any cost implications?
Discussion of the study findings	Doe 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 were repeated?	What could be done to improve the design of the study?

QUANTITATIVE RESEARCH METHODS

Case reports

A case report 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 the described as the first step in communicating patient information.

Case series

A case series is the natural sequel to a case report and describes a series of patients who are connected by exhibiting the signs and symptoms of a similar condition. In case series, many aspects of their care can be examined, including the treatment that was applied or their response (positive or negative) to treatment.

Case control studies

In this type of study, people with a particular condition or disease are identified ands 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 recording 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.

Cohort studies

Cohort studies can take a considerable amount of time to conduct. They examine at least two groups of subjects and find out what happens to them in the future. The follow up time in cohort studies is generally 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 disease or disorder 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 experiences a disease; incidence describes the number of new cases of a disease per year.

Randomised controlled trials
Randomised controlled trials (RCTs) are often 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 (eg. chiropractic treatment) or another (e.g. taking NSAIDs). The random assignment can be achieved in a number of ways eg. 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. In Single Blind studies, the participants do not know which type of intervention they are receiving; in Double Blind Studies, neither the patients or 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, like an active compound in a pharmacological study. Placebo or sham interventions can also be used when researching complex interventions, eg. acupuncture.

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

QUALITATIVE RESEARCH

Qualitative research uses a variety of methods eg. open-ended interviews, naturalistic observation, focvus 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 quantitative studies.

Statistical analysis of the literature

When data has been gathered, it needs to analysed statistically. Qualitative and quantitative data is analysed differently. The basis 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. Sifferent types of data (or variables) will be encountered in statistics. They will differ in their 'scale of measurement' – i.e. in terms of what can be ascribed to any numerical values that 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.

Overlap in definition

It could be argued that certain discrete variables which can take a very large number of possible values are better thought of as continuous for the purposes 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 basis measures that will be applied to research will come under the catergory 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 middleso 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 the 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 them between theminimum and maximum values present and expressed with a single digit. The value is also dependent on the size of the value; as the sample size increases, the range is likely to increase.

The Interquartile Range

This shows a range of data values split 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 of 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 in σ 2.

The Standard Deviation

The standard deviation is used to describe data. It can be calculate 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.

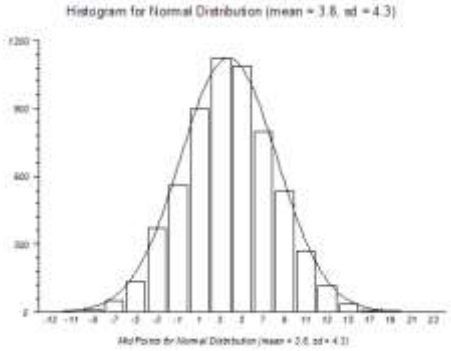
The Standard Error of the Mean

This can be used to estimate a characteristic in a sample population. It can be calculated:

Standard Error = $\frac{\text{standard deviation}}{\sqrt{n}}$ where 'n' is the value of the sample size

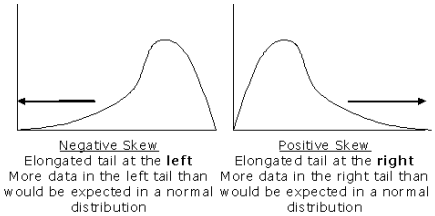
Visual presentation of data

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' that tail on the right hand side will be stretched out; conversely, 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 statistics are said to

have gained ‘statistical significance’. An alternative value for probability is less 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 of (i.e. very low p value) that it is true. The p valueis 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. As 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 the 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 (MMT)

This statistic is appearing more frequently in the analysis section of some 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, the more effective that intervention is considered to be.

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.