Critical Appraisal.

Critical appraisal forms part of a much larger entity – that of evidence-based practice, or evidence-based medicine (EBM). The strategy is that the clinician considers the best evidence available when making a decision regarding treatment, often in partnership with colleagues, the patient, or other service users. However, in order to identify the best evidence or practice, a large body of research and literature may need to be examined, sifted, and analysed, and this is what is determined as 'critical appraisal'.

Much of the research we read is flawed, and so it may be difficult to separate what is valuable from what is not applicable. It is the skill of critical appraisal that allows the reader to identify inconsistencies, assumptions, and flaws in the research process, and then balance these against the outcome of the research in order to decide whether the outcome has value and reliability, and can underpin a clinical decision.

The process of critical appraisal provides a systematic way of assessing the validity, results and potential usefulness of published research, providing the link between the somewhat isolated world of research and the real-life experience of clinical practice. The skills required are not difficult to acquire, and rely heavily on a straight-forward, common-sense approach to reading and thinking.

There are many different types of paper that could be appraised, and the approach changes slightly depending on whether the paper reflects a trial, a review, a case study, or a qualitative study. In some cases, there are well-accepted strategies for assessment, and it would be wise for the novice in critical appraisal to follow one of these.

As an example of how to follow the critical appraisal process, the example of the review process in a randomised controlled trial will be used.

Randomised controlled trials (RCTs).

Guyatt, Sackett & Cook (1993) suggested a checklist which could be used when examining an RCT. This type of design is frequently used in testing the efficacy or effectiveness of healthcare services or health technologies. As suggested by the name, RCTs involve the random allocation of different interventions (treatments versus placebo) to subjects. As long as numbers of subjects are sufficient, this ensures that both known and unknown variables are evenly distributed between treatment groups. In the hierarchy of evidence that influences healthcare policy and practice, RCTs are considered by most to be the top individual unit of research. They are considered the most reliable form of scientific evidence because they are thought to eliminate chance and bias. The use of this methodology can minimise accidental or intentional bias, but does not infer that every RCT is of good quality. This is where the process of critical appraisal should be employed, as only by examining the research method and process can the reader determine whether the outcome of the research is reliable and valid.

The process suggested by Guyatt et al is divided into sections, as follows:

Section 1. Are the results of the study valid ?

Screening questions;

1. Did the trial address a clearly focused research question?
2. Did the authors use the right type of study?

Detailed questions;

3. Was the assignment of patients to treatments properly randomised?
4. Were all the patients who entered the trial properly accounted for at is conclusion?
5. Were patients, health workers, and study personnel 'blind' to treatment?
6. Were the groups similar at the start of the study?
7. Apart from the intervention, were the groups treated equally?

Section 2. What are the results?

1. How large was the treatment effect?
2. How precise was the estimate of the treatment effect?

Section 3. Will the results help locally?

1. Can the results be applied to the local population?
2. Were all clinically important outcomes considered?
3. Are the benefits worth the harms and costs?

Adapted from :
Guyatt GH, Sackett DL, Cook DJ (1993) Users guide to the medical literature. II how to use an article about therapy on prevention. Journal of the American Medical Association 270: 2598-2607, and 271: 59-63

The first two questions of section 1 should be looked at as screening questions, allowing the reader to decide whether the study merits further consideration. Hopefully, the answers to these can be found very early on in the paper, and should really be made clear in the abstract. If the answer to either of these question is not a clear 'yes' then it is probably not worth spending further time on the paper.

All of the questions in section 1 are used to asses the trial method. If the methods are satisfactory, then it is more likely that the results will be useful – a poor or flawed research method is unlikely to yield meaningful results.

Use of randomisation aims to avoid the any possibility of selection bias. The test that randomisation has been successful is that different treatment groups have same characteristics at baseline, e.g. the same number of men and women, or older or younger people, or degrees of disease severity. Checking to see which method of randomisation has been adopted can give the reader an indication of the methodological rigour of the trial, and hence its likely validity. Correct randomisation also ensures that the groups are similar at the start of the trial.

Also of importance in relation to the questions in section one is checking to see if all the recruited participants are accounted for at the end of the trial. It's not unusual for some patients to drop out of a trial, for whatever reason, and whilst this is an accepted problem, as long as the dropped-out participants are discussed, then the trial can still be assessed as having validity. However, this changes if more than 15% of the participants drop out. In reality, the intention to treat analysis is more useful than actual completion of treatment analysis, as this covers all participants in each randomised group.

The process of 'blinding' means that as a participant you are unaware to which research group you belong – whether it is the active treatment group or the placebo group. If the participants were aware of their grouping, it may influence their reporting of results or outcomes. The same effect can occur with both health workers and study personnel. However, whilst it is not always essential to have health workers 'blind' it is essential to the rigour of the research process that the study personnel go through the blinding process, in order to prevent reporting bias. As it is essential that the groups are treated equally, the use of blinding takes on further importance to make sure that changes to outcomes are attributed to the intervention rather than differences in group management.

Of high importance, but which is not mentioned in Guyatt et al guidelines, are the ethical considerations of the study. There should be mention how issues with informed consent, confidentiality, risk factors, withholding or denying treatment, and participant distress, are dealt with. If the paper fails to mention these issues, then immediately its validity is called into question – this does not necessarily mean that the authors did not deal with the issues, but failure to discuss them in a paper is extremely poor, and in such cases these papers would not be published in the majority of scientific journals.

Moving to look more closely at section 2 questions, we then look to examine the results. This can sometimes be complex as different trials use different outcome measures, but commonly a measure of relative risk is adopted. The chance of a particular outcome being observed in an individual is the 'risk', which can be

either positive or negative. Comparing the risk in the intervention and control groups gives a measure of relative risk. A relative risk of 1 occurs when the incidences are the same in the two groups. If it is hoped that the intervention leads to more of the outcome being measured, such as an increase in ulcer resolution rates, then a relative risk of more than 1 is desirable. If it is hoped that that intervention creates less of the measured outcome, such as a reduction in lower limb ulceration, amputation, or death, then a relative risk of less then 1 is looked for.

However, because the trial can only examine a small percentage of the overall population, there is always some doubt about how the results can be applied to the population as a whole. This is where something known as a confidence interval (CI) is used, and indicates the range of doubt around the best estimate. Simply put, the confidence interval is the mean of the sample group which is used to estimate the mean of a larger population. The width of the confidence interval gives an idea about how much certainty can be attributed to the results - a very wide interval may indicate that more data should be collected before anything meaningful can be said about the results.

In addition to the confidence interval, the significance of any differences between the groups should be discussed, and is usually expressed as a p-value, used to indicated statistical significance. The standard p-value adopted which implies that the results have statistical significance is $p<0.05$. The lower the p-value, the less likely the result has occurred by chance, so the greater significance is attached to the result. A p-value of 0.05 corresponds to a 5% chance.

Finally, in section 3 the questions relate to the usefulness to the wider population, in particular to the groups of patients to which the reader relates. Large differences between the trial group, and the reader's group would make it difficult to apply the trial results locally. A further question to ask is whether the trial has addressed the issues and outcomes that relate to the reader's group – if not, then further evidence from other trials would be required in order to change local practice.

Finally, but no less importantly, do the costs involved in applying any changes suggested by the trial outcomes outweigh the potential negative costs? These costs not only relate financially, but also involve, time, quality if life, and social impact.

In addition to an RCT, there are numerous other types of experimental and research design, all of which provide varying degrees of evidence relating to outcomes. By learning the process of critical analysis relating to an RCT, one can apply these skills to the analysis of other methodologies, using similar strategies. Good strategies for the more commonly-encountered research processes can be found at : http://www.sign.ac.uk/methodology/checklists.html

On this webpage, provided by SIGN, the Scottish Intercollegiate Guidelines Network, there is also a tutorial, found by clicking the green 'online tutorials' tab at the top left of the page. Choosing to follow the 'asthma guideline' development tutorial takes the reader through the process of how to develop a treatment guideline by using evidence available through different types of trial and research process.

Task.

Select a situation you commonly encounter in your work, or a situation where you already have guidelines in place that require updating – it does not necessarily need to be a direct clinical question, but can also relate to management or personal development issues. Using the strategy suggested in the SIGN asthma tutorial, go through the process of developing new, or updating old, guidelines which can applied in your own area.

This activity can be carried out on your own, or as a group task. In the latter case, it may be beneficial to have an initial discussion with all group members before assigning individual tasks, followed by an on-going group effort to produce a set of guidelines. Don't forget that these new guidelines will need to be revisited on a regular basis, the interval for which can be decided by the group, and will contribute to on-going continual professional development activity.