Critical Appraisal - a tool to make sense of research papers

Phil Wiffen, *Editor, Cochrane Pain, Palliative and Supportive Care Group, Pain Research, Churchill Hospital, Oxford; Visiting Professor, Department of Pharmacy & Pharmacology, University of Bath.*

Correspondence to: phil.wiffen@ndcn.ox.ac.uk

Abstract

Title
Critical Appraisal - a tool to make sense of research papers

Author list
Wiffen P

Summary
In spite of the specialisms now common among clinical pharmacists, it is increasingly difficult to stay up to date with published work. This requires us to read smartly and to carefully choose what we read. Our interaction with patients and the medical round will raise questions which need to be converted into clinically relevant, answerable questions to inform our searching. Papers identified then need to be appraised to assure us of their validity or not and to extract useful results which can then be applied in practice.

This paper covers some of the knowledge required in the appraisal process but also points to useful tools that are available to pharmacists to promote the skilful evaluation of potentially relevant clinical papers in day-to-day practice. In particular, appraisal tools for systematic reviews and randomised controlled trials are presented in full.

Keywords: evidence-based medicine, formulary, GRADE, medicines management, systematic review.

Background

Pharmacists play an important role in medicines management and are called upon to make key decisions affecting the care of patients. As practitioners, we have a duty to stay up to date and to practice using current best evidence. Both aspects present challenges for busy professionals. I remember talking to a professor of medicine. He would state that, for him to stay up to date with his particular medical speciality, he needed to read around 20-25 papers every day of the year. He recognised that this was impossible and we know from other research that the amount of reading professionals undertake tends to fall off as they go through their careers. He used the challenge to suggest that we focus on systematic reviews and randomised trials to inform practice as these are likely to be the most reliable papers.

The trend towards specialisation in clinical pharmacy makes good sense in the light of the volume of literature we have to deal with. The major bibliographic databases list over 20 million papers and they only cover a proportion of the world's medical literature. Material relevant to pharmacy practice is often hidden in less well-known databases such as International Pharmaceutical Abstracts (IPA) and the Cumulative Index to Nursing and Allied Health Literature (CINAHL).

Some time ago I happened to be travelling to an event with a consultant liver specialist. We got to chatting about keeping up to date. I asked if he knew all the key literature on hepatitis. He responded by stating that he was confident he knew the key papers around hepatitis B but not hepatitis C. Even a specialist consultant understood that he not could keep up to date with all the literature on liver disease.

Muir Gray, in his book on Evidence-Based Health Care and Public Health,[1] postulates that around a half of medical interventions do more good than harm, around a fifth probably do more harm than good and for the rest we don't really know. If true, that is worrying, though I suspect many of us hold beliefs based on what we were taught that have little in the way of underpinning evidence.

We live in a world that often seeks to define quality in terms of personal experience and feedback. Many retail websites and booking sites rely on customers to provide feedback on purchases and bookings as a guide to (hopefully) encourage

others to purchase. However, many do not take a similar approach to the research that underpins the choices and decisions they make in day-to-day practice. Many either believe what is stated by others or maybe assume the abstract is a good summary of the findings. Both approaches can be misleading. On the other hand, many pharmacists have expressed concerns about tackling what they see as complex papers and see critical appraisal as a specialist technique. Common objections include a sense of it being too time consuming, too difficult, needing expert knowledge or requiring expertise in medical statistics. In reality, critical appraisal can be a mechanism to get us to examine papers in a greater depth and come out with some firm conclusions. In practice, it is not as difficult as we might fear.

Quoting Sackett,[2] Muir argues the need to:

- convert the need for information into clinically relevant, answerable questions (see section on Patient Intervention Comparison Outcome (PICO) below)
- find, in the most efficient way, the best evidence with which to answer these questions (whether this evidence comes from clinical examination, laboratory tests, published research, or other sources)
- critically appraise the evidence for its validity (closeness to the truth) and usefulness (clinical applicability)
- integrate the appraisal with clinical expertise and apply the results to clinical practice
- evaluate your performance.

The development of evidence-based practice over the past 25 years or so have provided us with some useful tools to help us focus on what is reliable and how to interpret what we find. Also, concerns about the quality of published reports have stimulated a raft of reporting guidelines, which journal editors are encouraged to impose. They all seem to have unusual acronyms such as CONSORT, PRISMA, RAMASES and CHEERS!

Hierarchy of evidence

A number of hierarchies of evidence now exist, all of which have merit but demonstrate minor variations. An organisation called GRADE (Grading of Recommendations Assessment, Development and Evaluation) has grown up to inform thinking. One of the most useful hierarchies was from the Scottish Intercollegiate Guideline Network (SIGN),[3] as outlined in Table 1.

Ideally, we should look for the highest level of evidence that can be found. For major decisions such as additions to formularies, type 1 or type 2 evidence is important. However, for rare conditions, there may only be level 3 or 4. That is fine providing the elements of possible bias are understood. I was recently asked for evidence in managing treatment resistant Lennox-Gastaut syndrome, which is a rare condition in children. The few citations on Medline may give a steer on how to approach treatment but would not be classified as type 1 or 2 evidence.

The key point about levels of hierarchy is to identify the study designs that display the least bias. Bias is a common problem and it seems that all forms of bias in medical research make things look better that they really are. The Cochrane Collaboration has invested effort into examining these biases and this is incorporated into every Cochrane review. The following issues are commonly covered (from Moore et al[4]):

Random sequence generation (checking for possible selection bias)

Method used to generate the allocation sequence is assessed as: low risk of bias (any truly random process such as random number table or computer random number generator); unclear risk of bias (method used to generate sequence not clearly stated). Studies using a non-random process (so called quasi randomisation) are excluded (odd or even date of birth; hospital or clinic record number). We know that studies which are not randomised considerably overestimate treatment effects.

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well-conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1-	Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort or studies High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2+	Well-conducted case control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal
2-	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies e.g. case reports, case series
4	Expert opinion

Table 1: SIGN hierarchy of evidence

Allocation concealment (checking for possible selection bias)

The method used to conceal allocation to interventions prior to assignment determines whether intervention allocation could have been foreseen in advance of, or during recruitment, or changed after assignment. Methods used to assess are as follows: low risk of bias (telephone or central randomisation; consecutively numbered sealed opaque envelopes); unclear risk of bias (method not clearly stated).

Blinding (checking for possible detection bias)

Method used to blind study participants and outcome assessors from knowledge of which intervention a participant received are assessed as: low risk of bias (study stated that it was blinded and described the method used to achieve blinding e.g. identical tablets; matched in appearance and smell); unclear risk of bias (study stated that it was blinded but did not provide an adequate description of how it was achieved). Again, research shows that studies which are not blinded overestimate treatment effects.

Incomplete outcome data (checking for possible attrition bias due to the amount, nature, and handling of incomplete outcome data)

Methods used to deal with incomplete data are assessed as: low risk (less than 10% of participants did not complete the study or used 'baseline observation carried forward' (BOCF) analysis, or both); unclear risk of bias (used 'last observation carried forward' (LOCF) analysis); high risk of bias (used 'completer' analysis). BOCF goes back to the level at the start of a study and is a more conservative approach. LOCF (common in registration studies) uses the last report by the patient even if they could not tolerate the treatments).

Sample size (checking for possible biases confounded by small sample size)

Small studies have been shown to overestimate treatment effects, probably because the conduct of small studies is more likely to be less rigorous, allowing critical criteria to be compromised. Studies were considered to be at low risk of bias if they had 200 participants or more, at unclear risk if they had 50 to 200 participants, and at high risk if they had fewer than 50 participants. These numbers are somewhat arbitrary and may be too small. This is the subject of on-going work. The concept of size as a bias is not accepted by all statisticians.

Working up a Patient Intervention Comparison Outcome (PICO)

Sackett reminded us of the need to generate questions into a clinically relevant, answerable form.[2] Developing a PICO is a great way to do this. The acronym stands for:

P : patients or problem

I : intervention

C : comparison - if relevant

O : outcomes.

Taking a clinical question such as 'is pregabalin effective in neuropathic pain?', the PICO may look like this:

P : Adults with neuropathic pain of at least three months duration who report their pain as moderate or worse.

I : Pregabalin given orally in any dose for any duration of time.

C : Other treatments for neuropathic pain or possibly placebo.

O : Patient reported pain relief using validated scales for either pain intensity or pain relief recorded over time. Adverse effects would also be part of the outcome assessment.

There are a few variations to PICO. Some add 'S' for 'studies' to define the study type being sought. In the example above, randomised controlled trials would be a good study type. Others add 'T' for 'time' so, again using the example above, it would be sensible to assess the effect of pregabalin over at least twelve weeks as this is a chronic condition. The other advantage of a PICO is that it greatly facilitates any searching that you want to carry out to find evidence.

Handling results

Meta-analysis is an optional part of a systematic review and is usually displayed as a forest plot. The power of a meta-analysis is in the increase of the sample size and a number of individual RCTs that are not statistically significant singly can provide a robust significant result.

The forest plot in Figure 1 shows the comparison of diclofenac potassium (fast acting) versus placebo to provide at least 50% pain relief at 6 hours.[4] While the confidence intervals on the individual studies are quite wide (due to size) the summary statistics (diamonds) have much narrower confidence intervals.

Turning statistics into meaningful numbers

Statistics seem to have the ability to turn the legs of some pharmacists to jelly! It is important to get to grips with some of the simpler concepts as these will help in interpretation. Generally, odds ratios are difficult to interpret so many results are presented as relative risk (also called relative benefit). These are somewhat easier to understand, though are frequently misused (especially by the BBC!). Results presented as relative risks can be easily converted to number need to treat (NNTs) which are much more widely understood. A good explanation of these terms can be found in 'Bandolier'.[6]

Using critical appraisal tools

- **10 questions to make sense of a review**
  The temptation on finding a paper of interest is to read it right through. A 'ten question' approach of critical appraisal tools takes the focus on to the key elements of a paper. For example, while the background might be interesting it is unlikely to inform decision making. The important aspects will be in the methods section (occasionally, the abstract may be sufficient) and in the results. The critical appraisal tools all use a similar

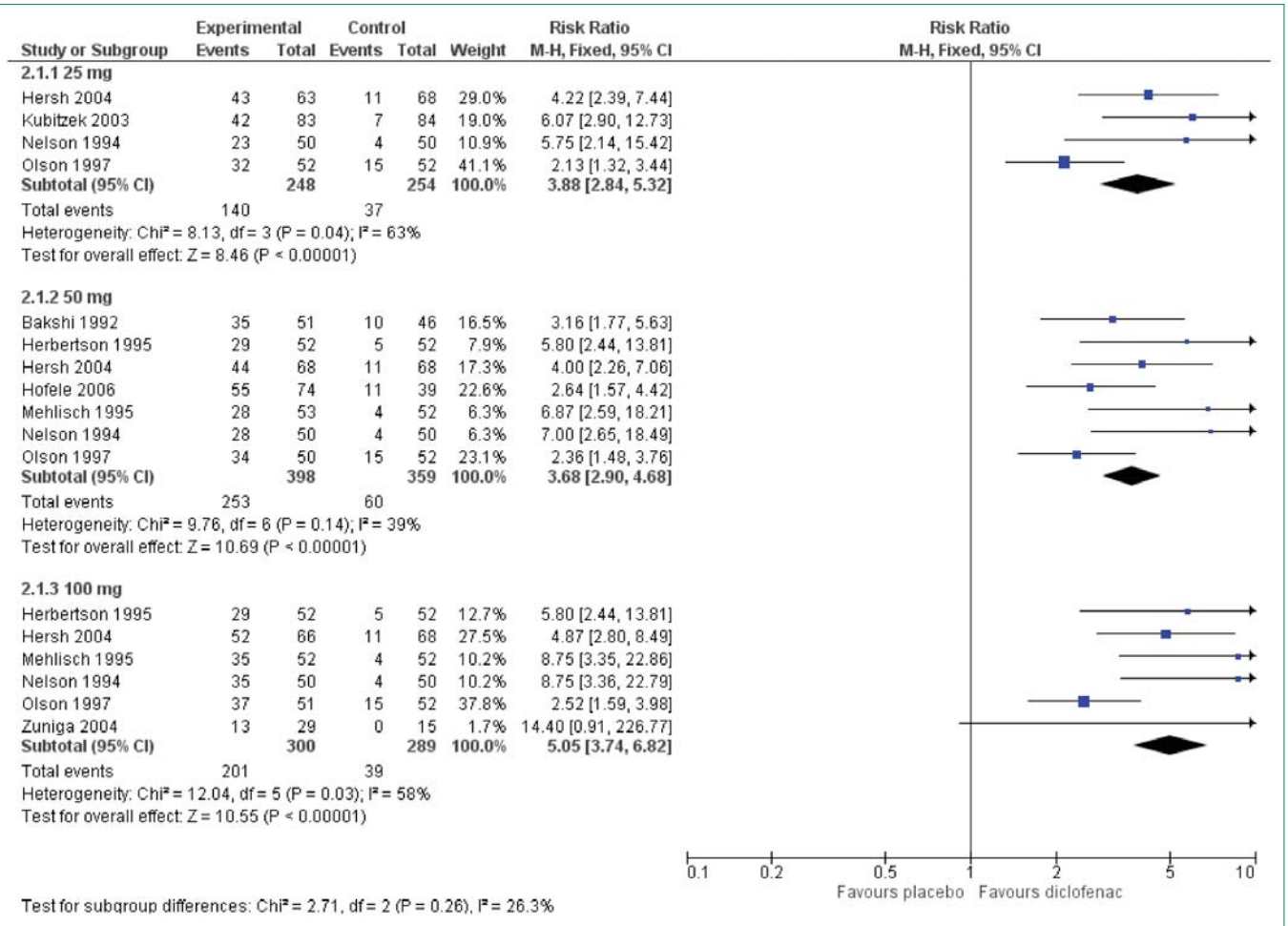

Figure 1: Forest plot for diclofenac potassium versus placebo

For each question answer 'Yes', 'Can't tell' or 'No'.

A. Are the results of the review valid?

1. Did the review address a clearly focused question?
   e.g. the population, intervention and or outcomes
2. Did the authors look for the appropriate sort of papers?
   Did they deal with the issues and have appropriate study design?

Is it worth continuing?

3. Do you think the important relevant studies were included?
   Look for search methods, reference list use, unpublished studies and non-English language
4. Did the authors do enough to assess the quality of included studies?
5. If the results of studies have been combined, was it reasonable to do so?

B. What are the results?

6. What is the overall result of the review?
   Is there a clear numerical expression?
7. How precise are the results?
   Are confidence intervals provided and are they reasonable?

C. Can I use the results?

8. Are the results likely to be useful in practice?
9. Were all the important outcomes considered?
10. Are the benefits worth the harms and costs?

Table 2: 10 questions to make sense of a review (adapted from CASP Systematic Review Checklist)

approach. The first point is to ensure that research methods reported are appropriate. If they are not, there is no point in wasting time on the paper - best to move on to something useful. Assuming that the methods are fine, then the appraisal tool asks key questions about the results. The final section is around deciding if the results have a local application in practice. Obviously this requires a value judgement and a balancing of the positive effects and adverse effects of any intervention. Cost may have a role but pharmacists need to think wider than the acquisition cost of a particular therapy. For example, the cost of antibiotic prophylaxis for elderly catheterised patients should also consider the morbidity, mortality and cost of a serious urinary tract infection. For the novice user the tool is likely to take around 30 minutes to complete but in that time a useful overview of the paper will be achieved. With use, that time commonly falls to around 10-15 minutes.

The ten questions to help you make sense of a review have been adapted for Table 2 but can be found in full on the Critical Appraisal Skills Programme (CASP) website.[8]

- **11 questions to make sense of a randomised controlled trial**
  Eleven questions to help you make sense of a randomised controlled trial have been adapted for Table 3 but can be found in full on the on the Critical Appraisal Skills Programme (CASP) website.[9]

Conclusion

This paper has set out to describe some of the key elements needed by pharmacists to evaluate papers that may be relevant to their practice. Issues such as the reputation of the publishing journal have not been discussed as these are largely irrelevant. All journals publish good and bad material. The tools outlined are designed to help us reject the bad ones and make good use of the good ones.

Critical appraisal is a skill owned by many physicians and nurses but is not widely used by pharmacists. Such skills should be a part of every clinical pharmacist toolkit and not confined to medicine information services.

For each question answer: YES, NO or DON'T KNOW

A. Are the results of the trial valid?

1. Did the trial address a clearly focused question?
   e.g. the population, intervention and or outcomes (PICO)
2. Was the assignment of participants to treatments randomised?
   How was this carried out? Was the allocation concealed from investigators?

Is it worth continuing?

3. Were participants, health workers and study personnel blinded?
   Was blinding possible but not done. Were outcome assessors blinded?
4. Were groups similar at the start of the trial?
   Think about age, gender, also severity of illness
5. Apart from the experimental intervention, were groups treated equally?
6. Were all of the participants who entered the trial properly accounted for at its conclusion?
   Was the trail stopped early? Are withdrawals fully described?

B. What are the results?

7. How large was the treatment effect?
   What outcomes were measured? Is the primary outcome clearly specified and reported in results? Were results reported for every outcome even if not significant? Is there evidence of selective reporting?
8. How precise was the estimate of the treatment effect?
   Are confidence intervals provided and are they reasonable? Were results statistically significant? Were the results clinically significant?

C. Can I use the results?

9. Are the results likely to be useful in your practice?
   Are the people you treat similar to those in the trial?
10. Were all the important outcomes considered?
    Is there information which you would have liked to have seen?
11. Are the benefits worth the harms and costs?
    Important even if not addressed by in the trial report

Table 3: 11 questions to make sense of a randomised controlled trial (adapted from CASP Randomised Controlled Trial Checklist)

Declaration of interests

The author reports personal fees from Pharman Ltd during the writing of this paper.

References

1. Muir Gray. Evidence-Based Health Care and Public Health: How to Make Decisions About Health Services and Public Health, 3rd edn. 2009. Churchill Livingstone. ISBN 978 0 443 10123 6.
2. Strauss SE, Sackett DL. Using research findings in clinical practice. BMJ 1998;317:339. Available from: http://www.bmj.com/content/317/7154/339 .
3. SIGN. SIGN grading system 1999 – 2012. Available at: http://www.sign.ac.uk/guidelines/fulltext/50/annexoldb.html . [Accessed Oct 2016].
4. Moore RA, Derry S, Aldington D, Cole P, Wiffen PJ. Amitriptyline for neuropathic pain in adults. Cochrane Database of Systematic Reviews 2015, Issue 7. Art. No.: CD008242. DOI: 10.1002/14651858.CD008242.pub3. Available at: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008242.pub3/full .
5. Derry S, Wiffen PJ, Moore RA. Single dose oral diclofenac for acute postoperative pain in adults. Cochrane Database of Systematic Reviews 2015, Issue 7. Art. No.: CD004768. DOI: 10.1002/14651858.CD004768.pub3. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26151766 .
6. Bandolier. "Evidence based thinking about health care". Available at: http://www.bandolier.org.uk/ .
7. Oxman AD, Cook DJ, Gordon HG for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature VI. How to use an overview. JAMA 1994;272(17):1367-71. Available from: http://jamanetwork.com/journals/jama/article-abstract/381796 .
8. Critical Appraisal Skills Programme (CASP). CASP Systematic Review Checklist. 10 questions to help you make sense of a review. Available from: http://www.casp-uk.net/casp-tools-checklists . [Accessed October 2016].
9. Critical Appraisal Skills Programme (CASP). CASP Randomised Controlled Trial Checklist. 11 questions to help you make sense of a randomised controlled trial. Available from: http://www.casp-uk.net/casp-tools-checklists . [Accessed October 2016].