



Beyond critical appraisal

Michael P Lowe, Clinical Dean, Northern Territory Clinical School, Royal Darwin Hospital, Darwin, Northern Territory, and Bradleigh D Hayhow, Resident Medical Officer, Kimberley Population Health Unit, Broome, Western Australia

Summary

All doctors require skills to critically appraise medical research. Critical appraisal is important but limited by its focus on the internal logic of research publications. A broader knowledge of the context in which studies are generated is sometimes necessary to understand their conclusions and their implications for clinical practice.

Key words: clinical trials, evidence-based medicine.

(*Aust Prescr* 2006;29:122–4)

Introduction

A key foundation of evidence-based medicine (EBM) is that clinicians with appropriate training can critically appraise research papers. Techniques of critical appraisal are taught to students and have been explained in several publications.^{1,2}

Critical appraisal has at least one major limitation – it suggests that by examining the content of publications alone one can assess the truth of their conclusions. The difficulty with this lies in a fundamental distinction between 'validity' and 'soundness'. Validity relates to the methodology used in a study, whereas soundness relates to the truth of the original data and its interpretation. Critical appraisal examines the validity of scientific studies to determine whether the evidence that is cited supports the conclusions, but it is unable to vouch for the soundness of those conclusions.

We tend to rely on researchers' assertions that their data are true. However, medical research occurs in an environment in which there are many conflicts of interest and powerful influences on researchers. We therefore need to know about the context in which evidence is generated before a true picture of the research findings can be made.

Academic fraud

Deliberate academic fraud represents one situation in which published data may be misleading. In one study almost 5% of medical authors reported fabrication or misrepresentation of results within the previous 10 years, and 17% of authors personally knew about a case of fraud in the previous 10 years.³

For example, in 1990 Werner Bezwoda appeared to begin trials of high-dose chemotherapy with bone marrow transplantation

in high-risk breast cancer patients. His published results showed markedly improved outcomes with this technique and therefore exerted a substantial influence on clinical practice worldwide. In 2000, a site visit to his laboratory revealed that the original results could only have been obtained by fraud.⁴

Academic fraud cannot necessarily be detected by critical appraisal. It is occasionally revealed by whistleblowers or by the discrepant results of subsequent studies but may take many years to come to light. Researchers do not often wish to repeat previous studies and ethics committees may say it is unethical to do so. Yet our vulnerability to academic fraud can only be reduced by independent corroboration of findings.

Inappropriate sub-group analysis

When evaluating clinical results one must be careful about the results of inappropriate sub-group analysis. Comparisons of multiple sub-groups can easily result in exaggeration of differences that are found. To avoid such problems, researchers are urged to clearly state their major hypotheses before their study begins.⁵ This requires researchers to be honest about their intentions as well as their data.

An example of inappropriate sub-group analysis occurred in the reporting of the CLASS trial.⁶ This trial was reported as a three-arm trial comparing the effects of celecoxib with two older non-steroidal anti-inflammatory drugs (NSAIDs) over a time period of six months. It showed a decrease in gastrointestinal complications for people treated with celecoxib. These results led to a marked rise in celecoxib prescribing around the world.

One year after the CLASS publication, it was revealed that the original intention of the trial had been very different, with a planned follow-up of 12–15 months, not six months.⁷ The trial had shown no difference in gastrointestinal adverse effects over the longer period, but when results had been restricted to six months a difference had emerged. To the original readers of the CLASS trial, none of this was evident and critical appraisal of the original article could only conclude that celecoxib was beneficial.

Leading medical journals now require all major trials to be registered at their onset and all Australian trials must be registered with the Australian Clinical Trials Registry (www.actr.org.au). However, doctors will continue to be bombarded with information from poorer quality trials in which problems of inappropriate analysis will be undetectable.

Not including all relevant outcomes

When analysing clinical study data all relevant outcomes should be considered, however it may not be clear which outcomes are important. Often clinical trials do not have the statistical power to detect important adverse events.

Rofecoxib was withdrawn after showing an increase in cardiovascular deaths with sustained use. Trials of rofecoxib (such as the VIGOR trial⁸) had noted but not emphasised this outcome, and attempted to explain it away. As a result, approval of the drug in world markets was based purely on equivalence of pain-relieving effects and decreased gastrointestinal adverse effects. Yet the possibility of adverse cardiac outcomes was apparent to experts soon after the drug's release.⁹

When trials are stopped early it may also be difficult to assess all relevant outcomes. Rules for stopping trials tend to rely on only one outcome (such as improvements in mortality) and may lead to other outcomes being ignored.

Placebos and semi-placebos

Trials need to ask the right question. Testing a drug against an inappropriate comparator or an inappropriate dose of a comparator can mislead practitioners. While there continues to be a place for placebo-controlled trials, there is no justification for use of 'semi-placebos' such as an inappropriately small dose of a competitor's drug.

A 1994 article entitled 'The continuing unethical use of placebo controls' suggested that wherever an established treatment existed, it should be used in trials in place of a placebo.¹⁰ Avoidance of placebos began to be seen as an important ethical principle and led to increasing numbers of so-called 'equivalence trials' in which new drugs were shown to be equivalent to older drugs rather than superior to placebos. Such trials may not always be clinically useful, and they assume that the established treatment has previously been shown to be significantly superior to a placebo.¹¹ Critical appraisal of any drug trial that is not placebo-controlled must therefore rely on expert knowledge of the evidence for the comparator drug.

Conflation and other complexities

An excellent summary of the problems encountered in critical appraisal warns about the issues that arise from 'conflating' trials.¹² It uses the example of the PROGRESS trial – which purported to show the benefits of ACE inhibitors after stroke.¹³

In fact, the PROGRESS trial actually shows a benefit from indapamide as a second-line agent, or from combinations of antihypertensives, rather than from an ACE inhibitor alone. Although the problem was noted by the editorial that accompanied the trial¹⁴, the result was so obscured within the paper that we believe only expert epidemiologists could come to the correct conclusion.

Evidence-based medicine downplays the role of experts, suggesting that we can all undertake critical appraisal. Yet an expert view of trials such as VIGOR would have differed from that of a general medical reader, not because of differing skills in critical appraisal, but because of a different knowledge of background issues. High levels of expertise in critical appraisal are also required for the interpretation of some trials in which key features may be deliberately hidden.

Until 2003, the Medical Journal of Australia published a series called 'EBM in action' in which the authors attempted to answer clinical questions by using techniques of critical appraisal. At the end of the series the authors appeared somewhat bemused by the reactions they had received:

There was a side effect that we did not anticipate. Content experts often disagreed with the evidence that we found – a collision between the findings of evidence expertise and content expertise. This often spilled over into the columns of the Journal's 'Letters to the Editor', generating about two letters for each 'EBM in action' article.¹⁵

This should not have been surprising. The content of the medical literature can really only be interpreted within the context of clinical medicine. Specialists in the field are 'content experts' who are ideally placed to assess the value of trials within this context. For this reason we believe that it is important to continue to emphasise the role of the content expert in augmenting the process of critical appraisal. However, we must be aware that experts may have conflicts of interest or be subject to influences that affect their views.

We believe that clinicians, in addition to paying attention to the method and results sections of a paper, should take note of editorials and any non-biased expert commentary that is available.

Conclusion

Critical appraisal uses techniques for analysing the validity of published evidence, however it is far less attuned to the soundness of that evidence. A solution to this problem is to pay greater attention to the context in which data are generated, but it seems unlikely that this will fall within the scope of most busy practising clinicians.

We believe that some simple rules can help prevent general medical readers from being misled by unreliable evidence. These include:

- not changing practice on the basis of single trials or trials from a single research centre
- sourcing information from trials that have been registered at their inception
- seeking expert opinion and commentary from content specialists as well as 'critical appraisal' specialists
- remaining aware of the possibility of biased original data.

References

1. Sackett DL, Straus S, Richardson S, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd ed. London: Churchill Livingstone; 2000.
2. Greenhalgh T. How to read a paper. The basics of evidence-based medicine. London: BMJ Publishing Group; 1997.
3. Gardner W, Lidz CW, Hartwig KC. Authors' reports about research integrity problems in clinical trials. *Contemp Clin Trials* 2005;26:244-51.
4. Weiss RB, Rifkin RM, Stewart FM, Theriault RL, Williams LA, Herman AA, et al. High-dose chemotherapy for high-risk primary breast cancer: an on-site review of the Bezwoda study. *Lancet* 2000;355:999-1003.
5. Lagakos SW. The challenge of subgroup analyses – reporting without distorting. *N Engl J Med* 2006;354:1667-9.
6. Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, et al. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomized controlled trial. Celecoxib Long-term Arthritis Safety Study. *JAMA* 2000;284:1247-55.
7. Juni P, Rutjes AW, Dieppe PA. Are selective COX 2 inhibitors superior to traditional non steroidal anti-inflammatory drugs? *BMJ* 2002;324:1287-8.
8. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N Engl J Med* 2000;343:1520-8.
9. Mukherjee D, Nissen SE, Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. *JAMA* 2001;286:954-9.
10. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994;331:394-8.
11. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Ann Intern Med* 2000;133:455-63.
12. Scott IA, Greenberg PB. Cautionary tales in the clinical interpretation of therapeutic trial reports. *Int Med J* 2005;35:611-21.
13. PROGRESS Collaborative Group. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6,105 individuals with previous stroke or transient ischaemic attack. *Lancet* 2001;358:1033-41.
14. Staessen JA, Wang J. Blood-pressure lowering for the secondary prevention of stroke. *Lancet* 2001;358:1026-7.
15. Del Mar CB, Anderson JN. Epitaph for the EBM in action series. *Med J Aust* 2003;178:535-6.

Conflict of interest: none declared

Book review

Therapeutic Guidelines: Rheumatology, Version 1.

Melbourne: Therapeutic Guidelines Limited; 2006. 314 pages. Price: \$39, students \$25.30, plus postage

Paul Kubler, Rheumatologist/Clinical Pharmacologist, Royal Brisbane and Women's Hospital, Brisbane

The first edition of *Therapeutic Guidelines: Rheumatology* is a welcome addition to this series. The pocket-sized published version is well formatted and covers a broad spectrum of rheumatic complaints. In general, it provides clear, accurate and practical information.

The first chapter on 'Getting to know your drugs' is contemporary and succinct. It includes detailed information on analgesics, corticosteroids and disease-modifying antirheumatic drugs (especially the biologically active treatments), and many of the commonly used complementary medicines frequently taken or asked about by patients with musculoskeletal complaints.

As expected, there are chapters outlining the evaluation and management of common rheumatic disorders such as

rheumatoid arthritis, gout and autoimmune connective tissue diseases. Regarding the book's format, I particularly liked the way that topics are presented by pattern of arthritis (for example, recent onset arthritis) and regional pain. For a general practitioner, patients usually present with an undifferentiated symptom pattern rather than a clearly established diagnosis so the book may help the doctor navigate through the period of initial care until a clear diagnosis becomes apparent.

There are also very good sections on the assessment and management of spinal pain, and the care of musculoskeletal conditions in children, adolescents and pregnant women.

My only criticism of this publication is that some sections are textbook-like in their detail which somewhat diminishes the value of this book as a quick and easy therapeutic guideline for busy practitioners. A brief description with diagrams on the intra-lesional injection techniques for common conditions such as rotator cuff tendinitis would improve its appeal to general practitioners.

In summary, the first edition of *Therapeutic Guidelines: Rheumatology* comes with my strong recommendation as a useful resource for many practitioners in the assessment and management of a broad range of musculoskeletal conditions.