Introducing critical appraisal

Huw TO Davies

All health-care professionals are being encouraged to use research evidence more in developing their clinical practice. To do this they need to inculcate the skills of 'critical appraisal'. Such skills involve assessing the appropriateness of research designs for answering clinical questions, critiquing the quality of the data gathering in individual studies, and assessing the implications of research findings.

INTRODUCTION

All health-care professionals are enjoined to engage with research evidence in delivering their services. Health-care managers too need to be able to understand the implications of research for service re-configuration and development. Indeed, as health services research has burgeoned over the past decade, many more health-care professionals are becoming actively involved in research projects large and small. While not everybody within health care needs to be a researcher, the great majority need to be able to read and understand research, and apply the findings in practice.

Critical appraisal is a structured way of reading research reports, which lays bare their key features. As these are scrutinized, the reader can ask whether the basic design is appropriate to the issues in hand, how well the study was carried out, and what the implications of the findings are. The critical appraisal approach involves asking a series of questions as a means of interrogating written research reports. These questions structure our thinking and greatly speed the task of reading and assimilating published research.

BEGINNING WITH A QUESTION

Research aimed at influencing clinical practice can fall into a number of distinct categories. First, research studies may help to elucidate the causes or risk factors for disease. Second, research may describe the natural history of disease, allowing estimates of likely prognosis if the condition is untreated. Third, research may help to appraise the usefulness of some kind of test for a disease. And finally, research may help inform on the likely benefits of interventions.

The crucial point is that different research designs are differently able to provide reliable guidance in each of these areas. Thus the first key question to ask is 'what do I want to know and what research design is best able to provide this information?'

THE RIGHT DESIGN

Different research designs address different clinical questions. Exploring disease aetiology is done using either cohort studies or case-control studies, depending on whether the disease of interest is common with early onset or rare with a long lag between exposure to pathogen and onset of disease. Although cross-sectional surveys are sometimes used to explore associations with disease they are a weak and inadequate design for this purpose.

In describing prognosis, cohort studies provide the best basis for estimating the likelihood of specific outcomes, with qualitative studies providing richer descriptions of these outcomes and their meaning to patients. Simple surveys of sufferers are again inadequate for understanding natural history as we have no way of knowing who has been omitted from the sample (for example deaths or those who have recovered).

It is in appraising diagnostic or screening tests that cross-sectional studies may be helpful — so long as there is a 'gold standard' against which the new test is compared (in blinded fashion) and a full spectrum of individuals are included. Such studies cannot, of course, address the issue of whether or not better diagnostic information will improve health outcomes. For this, longitudinal studies are required.

Assessing the impact of interventions on health outcomes is the area where most methodological development has been seen. To answer questions of whether specific treatments or whole programmes of care are better than the alternatives requires very careful attention to design. Although cohort and case-control studies have been used to suggest therapeutic benefit, without doubt the most appropriate designs are large, well-conducted randomized control trials with other safeguards against bias such as blinding. Better still are systematic reviews with meta-analysis of several randomized control trials. Before-and-after studies, or uncontrolled designs, may suggest fruitful areas for more rigorous study but do not of themselves usually provide definitive guidance.

THE DESIGN DONE RIGHT

All research designs are prone to bias during their design, implementation, analysis or interpretation. Observational studies (such as surveys, cohort studies or case-control designs) are particularly susceptible to confounding (Davies and Williams, 1999), but have other potential pitfalls as well which have been explored in previous papers in this series (Davies, 1999a; Davies and Crombie, 2000a,b). For assessing prog-

Dr Huw TO Davies is Reader in Health Care Policy and Management, Department of Management, University of St Andrews, St Andrews, Fife KY16 9AL

nosis, a crucial consideration is that any cohort study has gathered an inception cohort and has achieved full follow-up.

Controlled trials have evolved in sophistication over the past 50 years in an attempt to eliminate bias. Blinding (single, double or even triple), intention-to-treat analyses and careful assessment of balance at baseline are all now key features of rigorous studies (Davies, 1998a, 1999b). Systematic reviews have also now developed their own tradition of rigour and have taken over from the traditional narrative review (Davies and Crombie, 1999). Assessing the implications of chance remains essential for all of these research designs (Brennan and Croft, 1994; Davies, 1998b).

For each of the major research designs there is now a wealth of literature which develops step-by-step guides or checklists for the appraisal of published research. To go with these checklists, several useful texts also provide considerable exploration of the key ideas underlying critical appraisal and evidence-based practice (Crombie, 1996; Greenhalgh, 1997; Sackett et al, 1997). Further guidance can be found in the users' guides developed by the Evidence-Based Medicine Working Group led by Gordon Guyatt; formerly published in the *Journal of the American Medical Association* these guides are now available on the Internet (http://www.cche.net/).

ASSESSING LOCAL RELEVANCE

Once any given study has been assessed as offering trustworthy evidence (i.e. good internal validity), we turn to the question of whether the evidence will generalize to other settings or indeed particularize to our own setting (i.e. the question of external validity). Such questions are less those of checklists and more ones of judgments.

Four key considerations pertain. First, how similar are the patients in the study to patients seen elsewhere and locally? Second, are the interventions described readily replicable in other settings? Third, are the outcomes described the key outcomes of local interest? And finally, are there features of the setting where the research was carried out which suggest that the findings would not transfer well to other areas?

It is clear that none of the four questions above have simple answers. Yet consideration of each provides some structure around which to investigate the question of transferability of research findings from one setting to another.

ASSESSING THE IMPLICATIONS

It can be difficult at times to assess the importance of new research findings for clinical practice. Again, some simple questions can help clarify. For diagnostic tests, the key question is: will the findings from the test provide sufficiently reliable information to inform a treatment strategy?

For studies of prognosis, we are concerned about how patients feel about the range of possible outcomes and their attendant likelihood, and for studies of therapies the key issue is whether the potential benefits outweigh the possible harms. Thus information from several research studies may need to be combined with the individual patient's own preferences to develop a clinical management strategy.

Thus evidence on its own is incomplete. It needs to be particularized to individual patients. This process draws upon the collective experience and wisdom of clinicians known as 'clinical expertise'. It is this integration of best evidence with patient preferences facilitated by clinical expertise which is the hallmark of evidence-based practice. This integration cannot readily be reduced to a series of checklists: it draws on professional tacit knowledge accumulated through experience.

CONCLUSIONS

Evidence-based practice is about framing researchable questions as problems arise in clinical practice, searching for the best evidence available to answer those questions, appraising the quality of the evidence available and its applicability to local circumstance, and integrating the external evidence with clinical expertise and patient preferences to improve the quality of care.

This short article and others in this series have sought to provide insight and guidance into this process. Much fuller accounts are readily available to explore the issues raised in greater depth (Crombie, 1996; Greenhalgh, 1997; Sackett et al, 1997). HM

KEY POINTS

- Critical appraisal is a structured way of reading research reports to assess their appropriateness, trustworthiness and implications.
- The first key objective is to check that the right design has been used for the question in hand: cross-sectional blinded comparisons with a gold standard for the assessment of tests, inception cohorts with full follow-up for the elucidation of prognosis and randomized controlled trials (or systematic reviews of the same) for the assessment of interventions.
- The second objective is to ensure that the study under consideration has been performed to a high standard. Readily available checklists can assist with this assessment of internal validity.
- Deciding whether external evidence applies locally (external validity) involves consideration of differences between the study setting and local circumstances in terms of: the patients seen, the interventions available, the outcomes desired and the service contexts.
- Evidence alone is insufficient — it needs to be integrated with patients' preferences and values. Such a process utilizes clinical expertise and tacit knowledge over explicit checklists.

Brennan P, Croft P (1994) Interpreting the results of observational research: chance is not such a fine thing. *Br Med J* **309**: 727–30

Crombie IK (1996) *The Pocket Guide to Critical Appraisal.* BMJ Publishing, London

Davies HTO (1998a) Interpreting measures of treatment effect. *Hosp Med* **59**: 499–501

Davies HTO (1998b) Assessing chance variability in treatment trials. *Hosp Med* **59**: 650–2

Davies HTO (1999a) Bias in surveys. *Hosp Med* **60**(12): 898–900

Davies HTO (1999b) Bias in treatment trials. *Hosp Med* **60**(8): 599–601

Davies HTO, Crombie IK (1999) Getting to grips with systematic reviews and meta-analysis. *Hosp Med* **59**: 955–8

Davies HTO, Crombie IK (2000a) Bias in case-control studies. *Hosp Med* **61:** 279–81

Davies HTO, Crombie IK (2000b) Bias in cohort studies. *Hosp Med* **61**: 133–5

Davies HTO, Williams FLR (1999) Confounded by confounding: separating association from causation. *Hosp Med* **60**(4): 294–7

Greenhalgh T (1997) *How to Read a Paper: the Basics of Evidence Based Medicine*. BMJ Publishing Group, London

Sackett DL, Richardson WS, Rosenberg W, Haynes RB (1997) *Evidence Based Medicine: How to Practice and Teach EBM.* Churchill Livingstone, London