9

Appraising the evidence

Andrew Booth and Anne Brice

Introduction

This chapter outlines the principles of critical appraisal for different types of research study. It introduces generic schemes and discipline-specific checklists for use with information research. Outputs from the Critical Skills Training in Appraisal for Librarians (CriSTAL) programme are introduced and reviewed. Teaching scenarios from the CriSTAL programme are presented. Resources to assist in presenting and interpreting useful statistics are briefly identified.

	1 Define the problem
	2 Find evidence
✓	3 Appraise evidence
	4 Apply results of appraisal
	5 Evaluate change
	6 Redefine problem

Figure 9.1 The evidence-based practice process

What is critical appraisal?

'How do *you* decide whether an article is worth reading?'. Although many practitioners can identify the features of a good research article, these bear little

resemblance to the factors that determine what we read. First in the list is *interest* – does the title or abstract address a current preoccupation or some burning professional concern? We devour descriptions of new and exciting technologies more eagerly than prosaic, but worthy, descriptions of enquiry desk or interlibrary loan procedures.

Next come *extrinsic factors*: Have *I* heard of the author? Is it published in a peer-reviewed journal? Does the article originate from a reputable institution? Although these factors bear some relation to the quality of an article they are not automatic indicators of research quality.

Finally come more rigorous *intrinsic factors* that relate to research design and aspects of methodology. These are the focus of 'critical appraisal', described by David Sackett, a founder of evidence-based medicine, as: 'To weigh up the evidence critically to assess its validity (closeness to the truth) and usefulness (clinical applicability)' (adapted from Sackett and Haynes, 1995, **1**, 4–5).

In other words, we put aside our prejudices regarding the source or nature of a research study and judge it entirely on its own merits. We thus take into account the three important factors of *validity*, *reliability* and *applicability* (Booth and Haines, 1998).

Validity 'refers to the extent to which the results of the research are likely to be free from bias' (Reynolds, 2000). In other words, is there some flaw in the way the research has been done that might 'explain away' its findings? Consider if you were to stand in your library with a clipboard, taking notes. How would users react? How would the staff respond? Their observed behaviour would probably differ, even if almost imperceptibly, from that when they are not observed. Similarly, if we conduct an experiment using a more rigorous design, such as a controlled trial, the study itself is likely to affect the environment within which it takes place. The question we ask is: 'How much have the methods used to obtain the results thrown into question the findings themselves?'

'Reliability' relates to the 'trustworthiness of results'. In other words, what is the likelihood that this study reports something that is reproducible as opposed to being a 'fluke' or chance result? The presence of such a result can be ascertained by statistical techniques that relate to the anticipated frequency of chance and the uncertainty surrounding a particular observation. It should be mentioned at this point that these statistical approaches do not relate to some 'absolute truth' but, rather, stem from arbitrary decisions on likelihood. Statisticians regard 5% (or 1 in 20) as a threshold for chance – if something occurs more frequently than this they consider it unlikely to have happened by chance.

'Applicability' relates to the extent to which the results are likely to impact on practice (See Chapter 10). It is often contrasted with 'statistical significance'. Practitioners are not concerned with whether you can *measure* a difference between the effect of two choices of action. Rather, they want to know whether

the chosen action will *make* a difference to the users of the service. Is it worth doing? As Einstein reminds us: 'Not everything that can be counted counts, and not everything that counts can be counted' (Albert Einstein, 1879–1955).

It is the added dimension of 'applicability', relating research to practice, that makes critical appraisal different from, and more relevant than, 'critical reading' commonly encountered on an undergraduate course.

Critical appraisal in practice

Critical appraisal commonly starts with a problem or scenario. Having identified a likely source for the evidence (Chapter 7) and then searched the literature (Chapter 8) for a research study that addresses the problem, the next stage is to assess the quality of what we have found. Of course we may already have our own ideas on how to judge quality. However, it is usually more efficient to use a pre-existing checklist so that we do not overlook some important considerations. Multiple checklists exist – some originate from the pre-evidence-based era (Fowkes and Fulton, 1991), a series of User Guides was developed in support of evidence-based medicine, and others have appeared subsequently. Within the context of evidence-based information practice, contributions of note have come from our own unfunded Critical Skills Training in Appraisal for Librarians (CriSTAL) project) (Booth and Brice, 2003) and from the work of the Evidence Based Information Systems team in New Zealand (See Chapter 18).

The following scenario and corresponding article are used in the CriSTAL workshops:

Scenario – keeping a finger on the pulse

You have been invited to join a local implementation group looking at how to improve information services to those working in primary healthcare. The group is in agreement that there is a need to improve access to, and use of, information resources. However, there is considerable disagreement as to the best way of spending the available finances.

One of the medical staff on the team comes into the group's next meeting and places a sheaf of photocopies in the middle of the table. 'There you are,' she says triumphantly, 'this article from the Medical Journal of Australia (MJA) website is all the evidence we need'. You pick up a copy of the article in question: Jane M. Young and Jeanette E. Ward (Young and Ward, 1999), General practitioners' use of evidence databases, *MJA*, **170**, 56–8.

The group decides that, at the very least, it should consider the implications of this article at its meeting. Using the checklist provided, answer the following questions:

1 Would you consider this article to be suitable evidence to inform the group in making its decision? YES NO DON'T KNOW
2 Should the local implementation group purchase the evidence databases for its primary healthcare teams? YES NO DON'T KNOW
3 If your answer to question 2 is either 'NO' or 'DON'T KNOW' what other information would you need in order to make your decision?

Several features of the scenario are worth comment. The first question relates to *strength of evidence*. In other words, what is the validity and reliability of the article under consideration? Does it support a definite course of action? We should be able to agree whether this is a good or bad research study. Admittedly, different groups may choose to identify or give prominence to different features of the same study. However, particularly if a standard checklist is used, they should agree about its quality. The second question, however, addresses issues regarding *strength of recommendation*. What action will the group recommend based on the study? It is at this point that the values and preferences of the users enter the picture. You may make a different decision because of such considerations as available resources, the skill mix of staff, local policies and procedures and the wider political, social and cultural environment (See Chapter 10).

Finally, the scenario asks: 'If your answer to the previous question is either 'NO' or 'DON'T KNOW', what other information would you need in order to make your decision?'. This question encourages participants to consider other forms of evidence that might inform their decision. Local surveys or audits may help establish how the local library compares with that in the study. Are the users younger and more computer-literate? Are they more prosperous or better educated? Data on costs or technical specifications for equipment may have a bearing on the eventual decision. Alternatively you may look for additional literature to reinforce the findings of the study, to address a different but related question or to provide a different perspective on the same question.

Any decision seeks to optimize the balance between three perspectives:

- A *librarian* brings a professional perspective relating to whether or not the service works (the effectiveness of an intervention)
- A *manager* adds a consideration of whether the service is affordable (cost-effectiveness)
- The *user* perspective, finally and most importantly, will consider whether the service is acceptable and fair.

No research study is likely to address all three dimensions equally. Indeed, aspects such as user views may require other evidence, such as that from qualitative research.

Critical appraisal will not yield a single quantitative estimate for a context-laden and messy reality. However it can:

- reduce uncertainty
- allow you to focus on important issues
- help unravel complex problems
- harness group perspectives.

Increasingly, an information professional's role in supporting evidence-based practice requires familiarity with critical appraisal skills, resources and techniques (Landrivon and Ecochard, 1992; Dorsch et al., 1990; Scherrer and Dorsch, 1999). Involvement in critical appraisal has not been without controversy – not only do many other professional groups consider that librarians are ill-equipped to exploit research literature but even librarians themselves have concerns about adopting such a mantle.

The CRItical Skills Training in Appraisal for Librarians (CriSTAL) project aimed to establish whether it is practical for librarians to apply critical appraisal skills in their day-to-day practice (Booth and Brice, 2003). In doing so it sought to introduce a rudimentary knowledge of research design and to present necessary statistics in a way that is meaningful and non-threatening. The use of critical appraisal checklists for assessing the relevance and rigour of research findings is established in all disciplines that claim to pursue evidence-based practice and has led to development of guidelines for surveys, cohort studies, clinical trials and case-control studies (Crombie, 1996). Qualitative research, economic analyses and systematic reviews have also been targeted for a checklist approach. However, within the field of information practice two types of information literature were identified as particularly important and yet lacking an appropriate checklist: use studies and information needs analyses (Julien, 1996). (See Figure 9.2 in this chapter and Box 19.2 in Chapter 19.)

A. Is the study a close representation of the truth?
   1. Does the study address a clearly focused issue?
   2. Does the study position itself in the context of other studies?
   3. Is there a direct comparison that provides an additional frame of reference?
   4. Were those involved in collection of data also involved in delivering a service to the user group?
   5. Were the methods used in selecting the users appropriate and clearly described?
   6. Was the planned sample of users representative of all users (actual *and* eligible) who might be included in the study?

Figure 9.2 Twelve questions to help you make sense of a user study

(*continued*)

B. Are the results credible and repeatable?
   7. What was the response rate and how representative was it of the population under study?
   8. Are the results complete and have they been analysed in an easily interpretable way?
   9. Are any limitations in the methodology (that might have influenced results) identified and discussed?
C. Will the results help me in my own information practice
   10. Can the results be applied to your local population?
   11. What are the implications of the study for your practice?
   - in terms of current deployment of services?
   - in terms of cost?
   - in terms of the expectations or attitudes of your users?
   12. What additional information do you need to obtain locally to assist you in responding to the findings of this study

Figure 9.2 (*continued*)

The CriSTAL project utilized workshops modelled on the Critical Appraisal Skills Programme (CASP) format used to deliver appraisal training to health professionals. Participants were presented with the scenario from a library setting presented above, a research article and the corresponding checklist required to resolve the scenario.

Research design

Central to understanding critical appraisal is the so-called 'hierarchy of evidence' (Earl-Slater, 2001). This attempts to convey graphically the respective merits of different research designs according to their validity. It is important to note that such a hierarchy can be used only where you are considering the 'effectiveness' of interventions (i.e. whether a particular service or technology works). Issues around the acceptability of a service to users and other concerns of user preference are best addressed through qualitative research designs such as focus groups or Delphi processes. For this reason, other approaches, such as the concept of a 'signal-to-noise ratio' (Edwards et al., 1998), have received favour in some circles.

Eldredge (2002, 2003) has made two attempts at defining levels of evidence within librarianship. Within evidence-based healthcare a traditional hierarchy of evidence is given in Figure 9.3, and definitions of study types are given in Box 9.1 on page 112.

It is often effective to take librarians through such a hierarchy as a narrative whereby each successive design is seen to improve on its predecessor. In the absence of test-tube research and animal experiments(!) the entry point for a social science such as librarianship, is at the level of ideas, editorials and opinions.

In providing a literature searching training programme to users I might

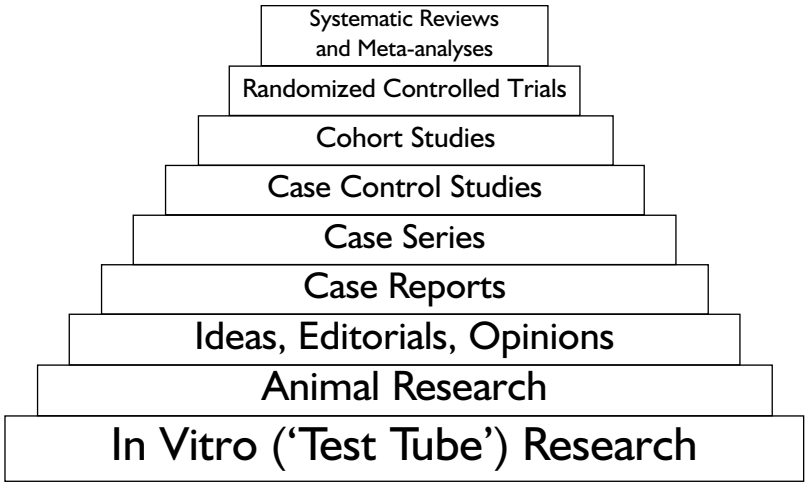

Figure 9.3 Conventional hierarchy of evidence in healthcare

ponder: 'I wonder if it would make a difference if I gave feedback to users on how good their searches are?'

In common with members of other pragmatic professions such as managers, social workers and teachers, my first recourse might be to ask a colleague for their *opinion*. While this might appear valid, my informant may be advancing an opinion without previous experience of the issue. My immediate concern would be, 'does an opinion advanced in theory hold up in actual practice?'

My next step might be to find a *case report* (or case study) where someone has actually tried this. Although this has the added virtue of being an actual occurrence I might have concerns about whether this is an isolated instance or whether it is likely to translate into practice (Fowkes and Fulton, 1991).

In accumulating a number of such cases we proceed to what is labelled a '*case series*'. This demonstrates the effect of an intervention in numerous different settings, some of which may even be exactly like my own. Nevertheless there may be considerable variation in the timing and duration of the intervention, the participants, the setting, and even the success factors or 'outcomes'. Published reports may differ in their definition of what exactly constitutes 'feedback'. We may be able to isolate a specific factor which may help to generate a hypothesis for further testing (Fowkes and Fulton, 1991).

The next stage would be to identify two broadly similar groups, one of 'successful searchers' and the other of 'unsuccessful searchers'. I would then ask all members of both groups whether they had received feedback on their searching. If one group had a higher proportion of those receiving feedback than the other then this would give an indication that such feedback does (or indeed does not) have an effect. At this stage matching cases (successful searchers) with controls

(unsuccessful searchers) in a *case-control study* is purely indicative and cannot establish cause and effect (Schulz and Grimes, 2002). There could be other explanations (confounding factors), the recall or memory of our informants may be imperfect or their records may be incomplete. Such are the limitations of studies that look back (retrospective) to a previous intervention.

At this point I would want to carry out a prospective (forward-looking) study where the record-keeping is both consistent and complete. In such a study – a *cohort study* – I follow an opportunistic group of those who have feedback and another group of those who do not have feedback (Grimes and Schulz, 2002). I capture a significant amount of detail on both groups so that I could identify any differences that might otherwise explain a difference in search performance (Eldredge, 2002).

An advantage of the cohort study is that it does not require participating users or libraries to change what they would have been doing anyway. This strength is also a limitation in that it opens up the possibility that both groups differ at the beginning of the study and therefore all that we are measuring upon completion of the study is that original difference. For this reason we might randomize a population into one of two groups as part of a *randomized controlled trial* (RCT). This ensures that, except for the prospect of chance – for which we make a statistical allowance (conventionally at 1 in 20) – the two groups are similar at the beginning (Eldredge, 2003). We can thus assume with a certain degree of confidence that any measurable difference is caused by a difference between the effect of the intervention and its comparison. In fact this is exactly what was done in the following published study: Bradley, D.R., Rana, G. K., Martin, P. W. and Schumacher, R. E. (2002) Real-time, Evidence-based Medicine Instruction: RCT in a neonatal intensive care unit, *Journal of the Medical Library Association*, **90** (2), 194–201.

Once a specific question has been studied rigorously in several comparable studies we can synthesize these findings through a *meta-analysis* or *systematic review*. A systematic review tries to answer a clear question by finding and describing all published and, if possible, unpublished work on a topic. A systematic review therefore uses explicit methods to perform a thorough literature search and critical appraisal of individual studies and uses appropriate statistical techniques to combine these valid studies. If a finding is replicated across a number of studies in a range of settings we have increased confidence that this finding is directly attributable to the intervention.

A well conducted systematic review helps *practitioners* avoid being overwhelmed by the volume of literature. Review articles help us keep up-to-date, define the boundaries of what is known and what is not known and can help us avoid knowing less than has been proven. By critically examining primary studies, systematic reviews improve our understanding of inconsistencies within diverse pieces of research evidence. By quantitatively combining the results of several small studies, meta-analyses can create more precise, powerful, and convincing

Box 9.1 Definitions of research studies

Definitions:
Case report - a description of a particular service or event, often focusing on unusual aspects of the reported situation or adverse occurrences.
Case series - a description of more than one case.
Case-control study - An observational study in which the cases have the issue of interest (e.g. successful literature searching) but the controls do not.
Cohort study - An observational study of a particular group over a period of time.
Randomized controlled trial - An experimental study in which users are randomly allocated to one of two or more options, where some get the option of interest and others get another option (e.g. a standard service).
Systematic review - An approach that involves capturing and assessing the evidence by some systematic method, where all the components of the approach and the assessment are made explicit and documented.
Meta-analysis is a method of synthesizing the data from more than one study, in order to produce a summary statistic.

conclusions. *Researchers* use systematic reviews to summarize existing data, refine hypotheses and define future research agendas. Without systematic reviews, they may miss promising leads or embark on studies of questions that have been already answered.

Calls for systematic reviews within librarianship have been ongoing for over a decade. Trahan (1993) discussed the potential of meta-analysis for library and information science, concluding that meta-analysis can be an effective tool for library and information science research (for an example see Box 9.2). Booth (1998) reported on a feasibility study entitled 'Library LORE' (Literature Oriented Reviews of Effectiveness) which followed the review process recommended by the NHS Centre for Reviews and Dissemination. This required a systematic search of the literature, documentation of study characteristics and exploration of the practicalities of identifying and reviewing this literature. He concluded that it is possible to conduct a systematic review in an information setting with undoubted value to practitioners. Nevertheless, he observed that the potential of systematic reviews in librarianship is constrained by several factors:

- different methods used to address similar questions (technically known as 'heterogeneity')
- the poor quality of research designs
- deficiencies in indexing and abstracting that make identification and retrieval of candidate studies problematic.

More recently, Hjorland (2001) has questioned 'Why is meta analysis neglected by information scientists?'

Box 9.2 Published example of a meta-analysis (abridged)

Telemedicine versus face to face patient care: effects on professional practice and health care outcomes (Cochrane Review)

Background: Telemedicine is the use of telecommunications technology for medical diagnosis and patient care. From its beginnings telemedicine has been used in a variety of healthcare fields, although widespread interest among healthcare providers has only now become apparent . . .

Objectives: To assess the effects of telemedicine as an alternative to face-to-face patient care.

Search strategy: We searched the Effective Practice and Organisation of Care Group's specialised register, The Cochrane Library, MEDLINE (1966 to August 1999), EMBASE (to 1996), CINAHL (to August 1999), Inspec (to August 1996), Healthstar (1983 to 1996), OCLC, Sigle (to 1999), Assia, SCI (1981 to 1997), SSCI (1981 to 1997), DHSS-Data.

We hand-searched the *Journal of Telemedicine and Telecare* (1995-1999), *Telemedicine Journal* (1995–1999) and reference lists of articles. We also hand searched conference proceedings and contacted experts . . .

Selection criteria: Randomized trials, controlled before-and-after studies and interrupted time series comparing telemedicine with face-to-face patient care. The participants were qualified health professionals and patients receiving care through telemedicine.

Data collection and analysis: Two reviewers independently assessed trial quality and extracted data.

Main results: Seven trials involving more than 800 people were included . . .The studies appeared to be well conducted, although patient numbers were small in all but one. Although none of the studies showed any detrimental effects from the interventions, neither did they show unequivocal benefits and the findings did not constitute evidence of the safety of telemedicine. None of the studies included formal economic analysis . . .

Reviewers' conclusions: Establishing systems for patient care using telecommunications technologies is feasible, but there is little evidence of clinical benefits. The studies provide variable and inconclusive results for other outcomes such as psychological measures, and no analysable data about the cost effectiveness of telemedicine systems. The review demonstrates the need for further research and the fact that it is feasible to carry out randomized trials of telemedicine applications . . .

Citation: Currell, R., Urquhart, C., Wainwright, P. and Lewis, R. Telemedicine versus Face to Face Patient Care: effects on professional practice and health care outcomes (Cochrane Review). In: *The Cochrane Library*, Issue 2, 2002, Oxford, Update Software.

Facilitating critical appraisal

Healthcare librarians have a long history of involvement in facilitating critical appraisal activities. Increasingly they apply this expertise to examining the evidence base for their own professional practice. This may be in the form of a regular professional journal club (Doney and Stanton, 2003; Grant, 2003, Koufogiannakis, Dorgan and Crumley, 2003) or an ad hoc continuing professional development event. In either case the stages required to facilitate such an opportunity are as in Figure 9.4.

Stage 1	Identifying a topic area
Stage 2	Finding a relevant article
Stage 3	Devising a scenario
Stage 4	Choosing a checklist
Stage 5	Deciding on a workshop format

Figure 9.4 Stages of facilitating critical appraisal

The effectiveness of critical appraisal has been established by findings from CASP evaluations of their general appraisal programmes (Burls, 1997) and a recent systematic review of critical appraisal research studies (Hyde et al., 2000).

Interpreting statistics

Participants at critical appraisal sessions view prior knowledge of statistical techniques as a major impediment to appraising a paper. Library and information studies graduates do not generally receive a comprehensive grounding in statistical techniques and their interpretation. Difficulties encountered tend to be of three types:

1 the actual terminology used
2 the specialist nature of the tests and techniques employed
3 the actual measures used.

Suggested solutions involve use of tools, worksheets and glossaries of terminology to enable participants to get the most from learning possibilities in the workshop. Box 9.3 provides a brief list of statistics resources drawn from the library literature or from the wider field of evidence-based practice.

Box 9.3 Resources for understanding statistics

Library statistical texts

Hafner, A. W. (1997) *Descriptive Statistical Techniques for Librarians*, 2nd edn, Chicago, American Library Association.

Hernon, P. (1994) *Statistics: a component of the research process*, rev. edn, Norwood, NJ, Ablex.

Hernon, P. (1994) Determination of Sample Size and Selection of the Sample: concepts, general sources and software, *College and Research Libraries*, **55** (2), 171–9.

Janes, J. (2001) Categorical Relationships: chi-square, *Library Hi Tech News*, **19** (3), 296–8.

Janes, J. (2001) Causality, *Library Hi Tech News*, **19** (2), 191–3.

Janes, J. (2001) The logic of Inference, *Library Hi Tech News*, **19** (1), 96–8.

Janes, J. (1999) Descriptive Statistics: where they sit and how they fall, *Library Hi Tech News*, **17** (4), 402–8.

Janes, J. (2002) Comparing the Means of Two Groups – the *t*-test, *Library Hi Tech News*, **20** (4), 469–71.

(*continued*)

Box 9.3 (*continued*)

Osif, B. A. and Harwood, R. (2001) Statistics for Librarians, *Library Administration and Management*, **15** (1), (Winter), 50–5.
Osborn, C. E. (2000) *Statistical Applications for Health Information Management*, Gaithersburg, MD, Aspen Publishers.

Evidence-based practice statistical texts

Ajetunmobi, O. (2002) Chapter 1 – Basic Stats. In *Making Sense of Critical Appraisal*, London, Arnold, 1–57.
Banerjee, A. (2003) *Medical statistics: Made Clear: an introduction to basic concepts*, London, Royal Society of Medicine Press.
Greenhalgh T. (2001) Chapter 5 – Statistics for the non-statistician. In *How to Read a Paper: the basics of evidence based medicine*, 2nd edn, London, BMJ Books, 76–93.
Pereira-Maxwell, F. (1998) *A–Z of Medical Statistics: a companion for critical appraisal*, London, Arnold.

Critical appraisal tools and products

Critical appraisal takes time and specialist skills. Not all librarians can aspire to practising the full process of critical appraisal for all research studies they encounter. Indeed, advocates of evidence-based practice have refined their vision to acknowledge this (Guyatt et al., 2000). Instead they encourage the production of such products of appraisal as critically appraised topics (CATS) so that practitioners benefit from appraisals of others (Sauve et al., 1995; Wyer, 1997). A CAT from an information context (facetiously labelled a 'CAT-Log') is given in Box 9.4.

Box 9.4 CAT-Log

Touch screen no better than leaflets in improving understanding of prenatal tests.

Question: In women booking antenatal care (population) is a touch screen system alongside a leaflet (intervention) more effective in terms of uptake, understanding, satisfaction with information and levels of anxiety (outcomes) than a leaflet alone (comparison)?
Design: Randomized controlled trial; intervention group (touch screen and leaflet), control group (leaflet only)
Setting: Antenatal clinic in university teaching hospital
Subjects: 875 women booking antenatal care
Outcome measures: Informed decision making on prenatal testing as measured by uptake and understanding of five tests, satisfaction with information received and anxiety levels as measured by the Spielberger state-trait anxiety inventory (STAI).
Results: The only significant difference in uptake was that more women in the touch screen group underwent detailed anomaly scanning ($p=0.0014$). Both groups showed significant improvements in knowledge over baseline (16 weeks gestation) by time of second questionnaire (20 weeks' gestation). Both groups reported high levels of satisfaction with the leaflet,

(*continued*)

Box 9.4 (*continued*)

with over 95% of the touch screen group also reporting that they would recommend the touch screen to other pregnant women. Compared with the baseline questionnaire, anxiety had declined significantly in the touch screen group mainly amongst 'first-time' pregnancies.
Commentary: A major problem with any experimental information-based intervention is the high level of dropouts during the course of the study with the likelihood of this increasing with each successive round of questionnaires. This is clearly seen in the flowchart of progress of participants through the trial. So, of 1477 invited to participate, 280 declined and a further 147 were ineligible. Of the 1050 actually randomized a further 175 dropped out without filling in the baseline questionnaire. So nearly 41% of potential subjects had dropped out even before the first measurements were taken. This attrition continued, with a further 104 dropping out at the time of the second questionnaire and 37 dropping out at the third and final questionnaire. Clearly there must be concerns about the applicability (or indeed practicability) of such an intervention in practice. Another major limitation is that 47% of participants had received higher education, making the study population unrepresentative of the population at large. The authors' own statement is significant 'Like all new technologies, these devices should be subject to rigorous evaluation' (Graham et al., 2000) whilst, in the accompanying commentary Jeremy Wyatt concludes 'with limited evidence of benefit for these expensive tools over well designed leaflets they seem to fit best into the National Institute for Clinical Excellence (NICE) category C: for NHS use only in the context of rigorous research studies' (Wyatt, 2000).

References

Graham, W. et al. (2000) Randomised Controlled Trial Comparing Effectiveness of Touch Screen System with Leaflet for Providing Women with Information on Prenatal Tests, *BMJ*, **320**, 155–9.

Wyatt, J. (2000) Commentary. Evaluating Electronic Consumer Health Material, *BMJ*, **320**, 159–60.

Conclusion

It is not sufficient to improve the depth of critical appraisal skills in the profession. An associated challenge is to investigate better ways of getting appraised and synthesized research reports to the profession in more readily accessible formats (see Chapter 12). The success of critical appraisal as a foundation for evidence-based librarianship depends on the production of rigorous and useable research studies. As a CRISTAL participant observed: 'the frustrating thing is the gap between the ideal of how library research/writing should be and how most of it actually is . . . '.

References

Booth, A. (1998) Testing the Lore of Research, *Library Association Record*, **100** (12), 654.

Booth, A. and Brice, A. (2003) Clear-cut?: facilitating health librarians to use information research in practice, *Health Information and Libraries Journal*, **20**, (Suppl 1), 45–52.

Booth, A. and Haines, M. (1998) Room for a Review?, *Library Association Record*, **100** (8), 411–12.

Bradley, D. R., Rana, G. K., Martin, P. W. and Schumacher, R. E. (2002) Real-time Evidence-based Medicine Instruction: RCT in a neonatal intensive care unit, *Journal of the Medical Library Association*, **90** (2), 194–201.

Burls, A. J. E. (1997) *An Evaluation of the Impact of Half-day Workshops Teaching Critical Appraisal Skills*, Oxford, Institute of Health Sciences.

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

Doney, L. and Stanton, W. (2003) Facilitating Evidence-based Librarianship: a UK experience, *Health Information and Libraries Journal*, **20** (Suppl 1), 76–8.

Dorsch, J. L., Frasca, M. A., Wilson, M. L. and Tomsic, M. L. (1990) A Multidisciplinary Approach to Information and Critical Appraisal Instruction, *Bulletin of the Medical Library Association*, **78**, 38–44.

Earl-Slater, A. (2001) Critical Appraisal of Clinical Trials: critical appraisal and hierarchies of the evidence, *Journal of Clinical Governance*, **6**, 59–63.

Edwards, A. G. K, Russell, I. T. and Stott, N. C. H (1998) Signal Versus Noise in the Evidence Base for Medicine: an alternative to hierarchies of evidence, *Family Practice*, **15** (4), 319–22.

Eldredge, J. D. (2002) Cohort Studies in Health Sciences Librarianship, *Journal of the Medical Library Association*, **90** (4), 380–92.

Eldredge, J. D. (2003) The Randomized Controlled Trial Design: unrecognized opportunities for health sciences librarianship, *Health Information and Libraries Journal*, **20** (Suppl 1), 34–44.

Fowkes, F. G. R. and Fulton, P. M. (1991) Critical Appraisal of Published Research: introductory guidelines, *BMJ*, **302**, 1136–40.

Graham, W. et al. (2000) Randomised Controlled Trial Comparing Effectiveness of Touch Screen System with Leaflet for Providing Women with Information on Prenatal Tests, *BMJ*, **320**, 155–9.

Grant, M. J. (2003) Journal Clubs for Continued Professional Development, *Health Information and Libraries Journal*, **20** (Suppl 1), 72–3.

Grimes, D. A and Schulz, K. F. (2002) Cohort Studies: marching towards outcomes, *Lancet*, **359** (9303), 341–5.

Guyatt, G. H., Meade, M. O., Jaeschke, R. Z., Cook, D. J. and Haynes, R. B (2000) Practitioners of Evidence Based Care, *BMJ*, **320**, 954–5.

Hjorland, B. (2001) Why is Meta Analysis Neglected by Information Scientists?, *Journal of the American Society for Information Science and Technology*, **52** (13), 1193–4.

Hyde, C., Parkes, J., Deeks, J. and Milne, R. (2000) *Systematic Review of Effectiveness of Teaching Critical Appraisal*, Oxford, ICRF/NHS Centre for Statistics in Medicine.

Julien, H. (1996) A Content Analysis of the Recent Information Needs and Uses of Literature, *Library and Information Science Research*, **18**, 53–65.

Koufogiannakis, D., Dorgan, M. and Crumley, E. (2003) Facilitating Evidence Based Librarianship: a Canadian experience, *Health Information and Libraries Journal*, **20** (Suppl 1), 76–8.

Landrivon, G. and Ecochard, R. (1992) Principles of the Critical Appraisal of Medical Literature, *Health Information and Libraries*, **3**, 29–34.

Reynolds, S. (2000) In Reynolds, S. and Trinder, E., *Evidence-based Practice: a critical appraisal*, Oxford, Blackwell Science.

Sackett, D. L. and Haynes, R. B. (1995) On the Need for Evidence-based Medicine [EBM Notebook], *Evidence-Based Medicine*, **1**, 5–6.

Sauve, S., Lee, H. N., Meade, M. O., Lang, J. D., Farkouh, M., Cook, D. J. and Sackett, D. L. (1995) The Critically Appraised Topic: a practical approach to learning critical appraisal, *Annals of the Royal Society of Physicians and Surgeons of Canada*, **28**, 396–8.

Scherrer, C. S. and Dorsch, J. L. (1999) The Evolving Role of the Librarian in Evidence-based Medicine, *Bulletin of the Medical Library Association*, **87**, 322–8.

Schulz, K. F. and Grimes, D. A. (2002) Case-control Studies: research in reverse, *Lancet*, **359** (9304), 431–4.

Trahan, E. (1993) Applying Meta-analysis to Library and Information Science Research, *Library Quarterly*, **63** (1), 73–91.

Wyer, P. C. (1997) The Critically Appraised Topic: closing the evidence-transfer gap, *Annals of Emergency Medicine*, **30**, 639–40.

Young, J. M. and Ward, J. E. (1999) General Practitioners' Use of Evidence Databases, *Medical Journal of Australia*, **170**, 56–8.