LITERATURE REVIEWS: EVOLUTION OF A RESEARCH METHODOLOGY

David Evans, RN, DipN, BN, MNS, is Coordinator of Reviews at The Joanna Briggs Institute For Evidence Based Nursing and Midwifery, Adelaide, South Australia

Inge Kowanko, BSc(Hons), PhD, is a Senior Research Fellow at Flinders University, Adelaide, South Australia

Accepted for publication March 2000

Key Words: literature review, systematic review, meta-analysis

ABSTRACT

Literature reviews summarise all past research efforts, and because of this, are increasingly being used as a way of addressing the rapid accumulation of health care information. Because reviews, rather than primary research, are now being used as the basis for many health care decisions, it is important that they are conducted with the same rigour of the primary research. To ensure this rigour, the past two decades have seen a progressive evolution in review methodology, to the point where reviews are now considered research in their own right. Despite this development, the standard of many published reviews remains poor. This paper summarises the development of the literature review methodology and discusses some of the implications for nursing.

INTRODUCTION

Literature reviews play an important role in the advancement of a discipline, because they accumulate past endeavours, summarise major issues and are an important way to disseminate the information generated by a large number of individual studies. There are many different types of literature reviews. Some present a simple discussion of a topic, while others provide a comprehensive and systematic summary of past research. Regardless of their scope, reviews differ from other research endeavours in that they are based on a body of completed works, rather than new research.

As the aim of the review is to present the 'state of the science' on a specific health care topic, it follows that the rigour that is required of the original primary research must also be applied to the review of this research literature. Without this rigour, summarised information is at risk of the same errors that threaten the credibility of primary research. Unfortunately, the commonly used approaches for reviewing the research literature have often lacked scientific rigour and as a result have not been held in high regard. Literature reviewers have also had to cope with an ever-increasing volume of health care research. The quality of this primary research varies considerably and results are often contradictory. This situation means that a literature review can be used to promote the opinions of the reviewer, in that only research that supports a particular view point is included, rather than representing an accurate and balanced summary of all the research on the topic.

To address this problem there has been a progressive evolution in review methods. This evolution started in the 1960s and 1970s, and is still occurring today. This development has seen the status of reviews change to the point were some reviews are now seen as research in their own right (Mulrow and Cook 1997). This paper summarises this evolution of the methods for reviewing the health care literature, and discusses some of the implications for nursing.

A concern for standards of reviews

Due to the accumulative nature of scientific knowledge, trustworthy accounts of past research are necessary (Cooper 1984). In the 1960s Price described the role of the literature review as replacing those papers lost from sight by the research front (Price 1965). Primary research information is digested, sifted, classified and synthesised (Manten 1973). Because reviews deliberately aim at cumulating knowledge in a specific area, they play a major role in the progress of a discipline in that they bring together previous work, identifying past achievements and possible future directions (Feldman 1971). Without these reviews of the literature, it is likely that a large amount of the research would be lost to the profession as the research front continues to move to new ground, or rediscovers past areas of interest on which to start fresh investigations.

Literature reviews have also been likened to primary research, and as such seek to make accurate generalisations about phenomena from limited information (Jackson 1980). Cooper and Rosenthal (1980) suggested that literature reviews also have great information-gatekeeping potential because knowledge is communicated to undergraduates and the lay public through reviews, rather than through the primary research reports. However, a criticism of the reviews of the 1970s was that they made no attempt at rigorous definitions or standardisation of the techniques used, and therefore the findings could be biased by factors that would be unforgivable in primary research (Glass et al 1981).

Prior to the 1980s there were no explicit methods for conducting reviews and so reviews varied widely in both quality and scope, which made assessment of their quality difficult (Jackson 1980). Light and Smith (1971) used a four category typology to characterise the approach taken by reviewers of this period by which they;

- listed any factor that had shown an effect on a dependant variable;
- excluded all studies except those that supported a particular point of view;
- averaged statistics across relevant studies in one form or another; or
- counted studies with similar results.

Even in the late 1980s, many scholarly publications had no specific or formal definition of a literature review (Cooper 1988). As the volume of research continued to grow, the traditional literature review which cited a small number of studies no longer did justice to the literature, and prompted the term 'the misinformation explosion' (Glass 1976). The volume of literature has continued to grow dramatically with the annual publication of 20,000 to 30,000 biomedical journals and 17,000 biomedical texts (Booth 1996). This makes presentation of an accurate summary of the research a daunting challenge, for not only must the relevant studies be found in this mass of publications, but issues such as the variable quality of studies and contradictory result, must be addressed. Booth described the current dilemma facing health care workers as 'Drowning in information, thirsting for evidence' (Booth 1996 p. 25).

Types of reviews

One early attempt to describe the potential foci of literature reviews identified three different components and suggested that a comprehensive review will likely address two or more of these areas (Cooper 1984). These components were;

- integrative reviews - summarise past primary research, draw overall conclusions, highlight unresolved issues and provide direction for future research;
- theoretical reviews - present theories to explain phenomena, compare them in terms of breadth, internal consistency and the nature of their predictions; and
- methodological reviews - examine and critique the research methods and operational definitions that have been applied to a problem, and address rigour and the risk of bias.

In the late 1980s Cooper (1988) developed a taxonomy which provided a framework for reviews that had previously been lacking. This taxonomy addressed six characteristics of reviews.

Focus: primary research outcomes, research methods, theories or the application of findings of research.

Goals: integration, criticism or the identification of central issues.

Perspective: neutral, where the reviewer attempts to present all arguments, espousal were the perspective is a particular argument or issue, paying little attention to other views.

Coverage: may range from exhaustive to only a representative cover of the available research.

Organisation: how the literature is presented in the review, for example historically in chronological order, conceptually with literature relating to the same ideas presented together, or methodologically in which literature is grouped according to the methods employed in the primary research.

Audience: the differing audiences to whom the reviewer is directing the work.

These developments provided a structure for reviews and acknowledged the differing methods used by reviewers. With this taxonomy came the recognition that, like primary research, the scope of literature reviews also differed, from a simple discussion to a comprehensive coverage of the literature using methods with a rigour similar to that of the research it summarised.

Meta-analysis

The traditional approach to summarising the results from more than one study was by narrative discussion, such as highlighting studies with similar or contradictory findings. These narrative reviews of research have existed for as long as there has been scientific literature (Petitti 1994). However, this approach has certain limitations. Firstly, while a narrative discussion is possible with a small number of primary research studies, it is inadequate for large numbers of studies addressing the same topic. Secondly, the lack of rigour of the approach makes an accurate and credible summary difficult. For example, if individual primary research studies report contradictory results, it is difficult to determine which results to include in the review. Glass has suggested that the traditional approach used when inconsistent findings were encountered was to eliminate all but a few studies, then advance these as the truth (Glass 1976). He described narrative summaries as 'airy speculation, unbefitting an empirical science' (Glass 1976). Finally, attempting to summarise the results of many studies in a short narrative discussion inevitably means that large amounts of information are lost due to the difficulty of dealing with this magnitude of information in a simple discussion. To overcome these difficulties, statistical methods were developed that were able to deal with large amounts of numerical data.

Statistics have been used to combine the results from multiple studies since the 1930s (Petitti 1994). Meta-analysis has been described as a quantitative approach for systematically combining the results of previous research in order to reach conclusions about the body of research (Petitti 1994). Glass suggested that the analysis of research data occurs at three different levels (Glass 1976). The first level, primary analysis, is the original analysis of the results of a single research study. The next level, secondary analysis, is the re-analysis of the results of a single research study to better answer the research question or to answer new questions. The third level is meta-analysis, which is a method for the integration and statistical analysis of data from a number of separate studies. Meta-analysis provides a logical framework to a research review, in that similar measures from comparable studies are listed systematically and, when possible, the measures of the effect of an intervention are statistically combined (Dickersin 1992). Meta-analysis has now been used extensively in reviews as a way of dealing objectively with the large amounts of data generated by a group of similar individual studies. More recently, in response to the variability of results when there is more than one review utilising meta-analysis for a topic, the term 'meta meta-analysis' has emerged (Katerndahl and Lawler 1999). This term has been used to represent the combined analysis of two or more meta-analyses.

Systematic reviews

While meta-analysis brought greater rigour to the summary of research data, many other aspects of literature reviews were still conducted in a haphazard manner. The term 'Systematic Review' came into use in the early 1990s to emphasise the importance of an extensive, systematic process of identifying, appraising and summarising all research on the review topic. These systematic reviews have been described as scientific tools which are used to summarise and communicate the results and implications of otherwise unmanageable quantities of research (NHS Centre for Reviews and Dissemination 1996). In contrast to the traditional narrative literature review, systematic reviews use explicit rigorous methods to identify, critically appraise and synthesise relevant studies (Mulrow 1987). When appropriate, meta-analytic techniques are used to combine statistical data from individual studies. To date, the major research design summarised by systematic reviews is the randomised-controlled trial (RCT).

Like primary research, systematic reviews have distinct stages which cover the review question, the search for relevant studies, critical appraisal, data extraction and analysis (Mulrow and Oxman 1997; NHS Centre for Reviews and Dissemination 1996).

Review question: each review has one or more specific questions which will likely encompass four components: the population of interest, the intervention, a comparison or control and an outcome of interest (Counsell 1997; NHS Centre for Reviews and Dissemination 1996). Inclusion criteria are developed from these, which operationalises the review question, putting it into practical terms that enables relevant studies to be selected for inclusion in the review.

Search: the aim of the search is to locate as much research on the topic as possible, while minimising the number of relevant studies missed. A comprehensive systematic review includes both published and unpublished research from around the world. The search strategy is documented in sufficient detail to allow others to replicate the search and to critique its quality (Counsell 1997; Jadad and McQuay 1993).

Critical appraisal: this involves making judgements about the validity of certain aspects of primary research studies, such as determining whether the methods used in studies can be trusted to provide a genuine, accurate account of the phenomena being studied (Avis 1994). The appraisal focuses primarily on bias as a result of selection of participants, differences in treatment of participants, differences due to losses of subjects from a study and differences in measuring the outcomes of interest (Mulrow and Oxman 1997).

Data extraction: the data used by systematic reviews are results produced by individual studies. This data is

extracted from the primary research reports using a specially developed form to minimise the risk of error while results are being transcribed.

Data analysis: the objective of the systematic review is to summarise the results from different studies, and when appropriate, this will be achieved using meta-analysis. It is only appropriate to use meta-analysis when the studies involved the same type of populations, used the same interventions and measured the same outcomes.

Each of these stages is documented before commencement in a systematic review protocol, which is the equivalent of a research proposal. This protocol minimises the risk of bias as a result of subjective decisions made during the conduct of the review. It also allows the review to be critically evaluated or replicated.

Hierarchy of evidence

To address the fact that primary research has many possible study designs with different risks of bias and credibility of findings, hierarchies of evidence have been developed. These scales rank the evidence according to its risk of bias and error. The most commonly used scales are those ranking research into the effectiveness of interventions, and while there are a variety of these scales, one example is listed below (NHMRC 1995).

Level I	Evidence obtained from a systematic review of all relevant randomised controlled trials.
Level II	Evidence obtained from at least one properly designed randomised controlled trial.
Level III. 1	Evidence obtained from well designed controlled trials without randomisation.
Level III. 2	Evidence obtained from well designed cohort or case control analytic studies, preferably from more than one centre or research group.
Level III. 3	Evidence obtained from multiple time series with or without the intervention, or dramatic results in uncontrolled experiments.
Level IV	Opinion of respected authorities based on clinical experience, descriptive studies or reports of expert committees.

Other scales have been developed and used for studies addressing other issues, such as diagnosis, prognosis, and clinical practice guidelines (Meltzer et al. 1998). However, a limitation of the hierarchies of evidence for the effectiveness of interventions and treatments is that qualitative research is ranked at the level of expert opinion, the lowest level in the hierarchy. This ranking ignores the rigour of qualitative research, which may be comparable to other research designs. The major difference is that qualitative studies seek answers to different types of questions. This ranking also fails to recognise the important contribution qualitative studies can make, particularly by ensuring that the perspective of the recipient of the treatment is part of any evaluation of its effectiveness and appropriateness.

Assessing the quality of reviews

Based on the growing importance of literature reviews, the increasing number that are published each year, and the fact that not all reviews are of equal quality, criteria have been developed to assess their scientific quality. A variety of assessment tools are available, one good example developed by Oxman and Guyatt (1991) consists of 10 questions:

1. Were the search methods reported?
2. Was the search comprehensive?
3. Were the inclusion criteria reported?
4. Was selection bias avoided?
5. Were the validity criteria reported?
6. Was validity assessed appropriately?
7. Were the methods used to combine studies reported?
8. Were the findings combined appropriately?
9. Were the conclusions supported by the reported data?
10. What was the overall scientific quality of the overview?

Because clear and precise reporting of the methods employed is important for literature reviews, questions one, three, five and seven address reporting. Without this reporting of methods, judgement of scientific quality of reviews would be impossible. Other questions address the scope of the search, the process of selecting studies, their critical appraisal, and whether findings were combined appropriately. Question nine, 'were the conclusions supported by the reported data?' highlights that, like primary research, the findings of a review must accurately reflect the studies summarised. Finally, the review must have the rigour and the scientific quality that characterises all reputable research endeavours, and this is achieved through adherence to accepted review methods.

Structured reviews

Another approach to summarising the literature is the conceptually broader structured review. This type of review has a much broader scope than that of a systematic review and may cover more than one intervention, or even an entire disease or health care topic. It has been suggested by Stevens and Milne (1998) that structured reviews are needed because end users, such as planners of health care, often require a reliable perspective on a topic which goes wider than that provided by a systematic review. These reviews are often needed within a short time scale to address immediate problems, precluding the lengthy process of the systematic review. The need for structured reviews also results from the rapid rate of diffusion of new health care technologies (Stevens and Milne 1998).

Although many of these reviews follow the strict rules of the systematic review, they often have to be based on limited literature and an element of modelling to provide useful conclusions (Stevens and Milne 1998). Structured reviews are appropriate for some nursing topics where there is limited research on which to base recommendations. Additionally, many issues of importance to nursing are broader in nature than the single intervention, medication or surgical procedure commonly summarised in systematic reviews. However, basing recommendations on limited evidence brings with it a greater risk of bias and error. Despite the potential usefulness of structured reviews, there is very little published information about them and minimal attempts to ensure the appropriateness of current methods.

IMPLICATIONS FOR NURSING

Literature reviews are not static scholarly activities, but rather an evolving research method. As the volume of research literature increases, reliance on these summaries will also increase. This reliance will see literature reviews forming the basis of many health care decisions rather than the original primary research. It is therefore important that they are conducted with similar rigour to the studies they summarise.

Despite the importance of rigour in review methods, the quality of reviews have been questioned. An examination of reviews in the nursing literature in 1987 found that the majority fell short of the standards used by the primary researchers, and included problems such as failing to state how the research reports were identified and chosen, and inadequate reporting of the methods used by the reviewer to appraise and integrate studies (Ganong 1987). A similar lack of scientific quality was identified in a study of medical review articles (Mulrow 1987). This lack of quality and rigour in review articles is a problem yet to be addressed by nurses, as a more recent appraisal of 33 review articles published in one nursing journal found only two papers specified the methods used in the review process (Redfern 1997).

Another area of concern for nursing is that the methodological development for literature reviews has focused on empirical research, primarily the randomised controlled trial. How qualitative research can best be identified in the literature, critically appraised and then summarised has yet to be determined. For without the inclusion in literature reviews of qualitative nursing studies, this research will be lost to the profession as the research front moves on.

Hierarchies of evidence have been developed and used to help determine which study design best answers clinical questions. The 'best' in this context, are study designs at least risk of bias and error. In order to evaluate the effectiveness of a treatment or intervention, a well conducted randomised controlled trial is cited as the design that will provide the best evidence. Similarly, for determining risk factors for a specific condition, cohort and case control studies are seen as the best design (Counsell 1997; Fowkes and Fulton 1991). These hierarchies commonly rank qualitative research at the same level as the opinion of respected authorities and expert committees (NHMRC 1995). However, well designed qualitative research, like other research, is based on clearly identified populations; rigorous collection; analysis and interpretation of data; and carefully documented methods. Qualitative research provides important information on issues such as the appropriateness of an intervention, the behaviour and attitudes of clients and health care professionals, why clients fail to follow treatment protocols, and the client's experience of illness or treatment. As these hierarchies of evidence are mainly used for medical research, nurses must evaluate their appropriateness for research and literature reviews addressing nursing issues.

CONCLUSION

There has been a progressive evolution of review methodology and this has reached a point where reviews themselves are now considered research activities. Despite these dramatic developments, review methods are still evolving to meet the needs of health professionals. To date, nursing has had only limited participation in this process, instead borrowing methods developed primarily for medicine, or the social and behavioural sciences.

Some issues of importance have still to be addressed by nursing, and these relate to the quality of reviews and to the methods used. If these reviews are to move nursing knowledge from the researcher to the practitioner, they must be useful to the profession. This will mean developing review methods that allow the broad nature of some of nursing questions to be appropriately summarised. It will also mean incorporating research information into reviews that are meaningful to nurses. While randomised controlled trials will certainly answer some nursing questions, there are nursing research questions best approached by other means. To not include qualitative studies in literature reviews will have important consequences for nursing. Rigorous methods to identify, appraise, integrate and summarise qualitative research are yet to be developed

While there will always be a place for the brief discussion paper, reviews that aim to change clinical practice must be of the same standard as the primary research they summarise. This high standard must be applied during the conduct of the review and also in the reporting of the methods used.

REFERENCES

Avis, M. 1994. Reading research critically. An introduction to appraisal: Designs and objectives. *Journal of Clinical Nursing.* 3:227-34.

Booth, A. 1996. In search of the evidence: Informing effective practice. *Journal of Clinical Effectiveness*. 1(1):25-29.

Cooper, H. 1984. *The integrative research review: A systematic approach.* Sage Publications, Beverly Hills.

Cooper, H.M. and Rosenthal, R. 1980. Statistical versus traditional procedures for summarising research findings. *Psychological Bulletin.* 87(3):442-449.

Cooper, H.M. 1988. Organizing knowledge synthesis: A taxonomy of literature reviews. *Knowledge in Society.* Spring:104-126.

Counsell, C. 1997. Formulating questions and locating primary studies for inclusion in systematic reviews. *Annals of Internal Medicine.* 127:380-387.

Dickersin, K. and Berlin, J.A. 1992. Meta-analysis: State of the science. *Epidemiologic Review*. 14:154-176.

Feldman, K.A. 1971. Using the work of others: Some observations on reviewing and integrating. *Sociology of Education.* 4:86-102.

Fowkes, F.G.R. and Fulton, P.M. 1991. Critical appraisal of published research: Introductory guidelines. *British Medical Journal.* 302:1136-40.

Ganong, L. H. 1987. Integrative reviews of nursing research. *Research in Nursing and Health.* 10:1-11.

Glass, G.V. 1976. Primary, secondary and meta-analysis of research. *Educational Researcher*. 5(10):3-8.

Glass, G.V., McGaw, B. and Smith, M.L. 1981. *Meta-analysis in social research.* Sage Publications, Beverly Hills.

Jackson, G.B. 1980. Methods for integrative reviews. *Review of Educational Research*. 50:438-460.

Jadad, A.R. and McQuay, H.J. 1993. Searching the literature: Be systematic in your searches. *British Medical Journal.* 307:66.

Katerndahl, D.A., and Lawler, W.R. 1999. Variability in meta-analytic results concerning the value of cholesterol reduction in coronary heart disease: A meta meta-analysis. *American Journal of Epidemiology.* 149:429-41.

Light, R.J. and Smith, P.V. 1971. Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review*. 41:429-471.

Manten, A.A. 1973. Scientific literature reviews. *Scholarly Publishing.* 5:75-89.

Meltzer, S., Leiter, L., Daneman, D., Gerstein, H.C., Lau, D., Ludwig, S.,Yale, J., Zinman, B. and Lillie, D. 1998. 1998 clinical practice guidelines for the management of diabetes in Canada. *Canadian Medical Association Journal.* 159(8 Suppl):S1-S29.

Mulrow, C.D. 1987. The medical review article: State of the science. *Annals of Internal Medicine.* 106:485-8.

Mulrow, C.D. and Cook, D.J. 1997. Systematic reviews: Critical links in the great chain of evidence. *Annals of Internal Medicine.* 126:389-391.

Mulrow, C.D. and Oxman. A.D. 1996. *Cochrane collaboration handbook [updated 1 March 1997], In: The Cochrane Library [database on disk and CDROM]*: The Cochrane Collaboration, Oxford: Update Software; Updated quarterly.

NHMRC. 1995. *Guidelines for the development and implementation of clinical guidelines.* 1st ed: Canberra: Australian Government Publishing Service.

NHS Centre for Reviews and Dissemination. 1996. *Undertaking Systematic Reviews of Research on Effectiveness: CRD Guidelines for Those Carrying Out or Commissioning Reviews*. York: University of York.

Oxman, A.D. and Guyatt. G.Y. 1991. Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology.* 44(11):1271-8.

Petitti, D.B. 1994. *Meta-analysis, decision analysis and cost-effectiveness analysis.* New York: Oxford University Press.

Price, D. 1965. Networks of scientific papers. *Science.*149:56-64.

Redfern, S. 1997. Review articles in the Journal of Clinical Nursing. *Journal of Clinical Nursing.* 6:251-252.

Stevens, A. and Milne. R. 1998. A knowledge-based health service: How do the new initiatives work. *Journal of the Royal Society of Medicine.* 91(Suppl 35):26-31.