

**Title**

Systematic reviews of nursing research: development of a conceptual framework

**Student**

David Evans

**Supervisor**

Professor Alan Pearson

Department of Clinical Nursing

Adelaide University

**Submission Date**

August 2001

# Thesis Index

Abstract	pg. 9
Statement of Original Work	pg. 11
Acknowledgments	pg. 12
<b>Chapter 1. Introduction</b>	pg. 13
<b>Chapter 2 Background</b>	
Evidence Based Health Care	
• introduction	pg. 19
• what is EBHC	pg. 20
• why EBHC	pg. 25
• the evidence	pg. 34
Reviews of the Research	
• introduction	pg. 44
• evolution of the literature reviews	pg. 45
• reviews as gatekeepers of knowledge	pg. 54
• summarising what we know	pg. 58
<b>Chapter 3. Methods Used to Develop an Expanded Review Process</b>	
Conceptual Framework	
• the different perspective of evaluations	pg. 64
• multi-dimensional concept of evidence	pg. 67
• justification of approach	pg. 68
Developing an Expanded Review Process	
• introduction	pg. 72
• method	pg. 73
<b>Chapter 4. The Expanded Review Process</b>	
Introduction	pg. 78
Review Protocol	
• introduction	pg. 79
• the role of the protocol	pg. 79
• proposed method for the protocol	pg. 80
Review question	
• introduction	pg. 82

• questions for interpretive research	pg. 84
• proposed method for review questions	pg. 86
<b>Selection of Studies</b>	
• introduction	pg. 88
• inclusion criteria	pg. 89
• exclusion criteria	pg. 91
• selection of interpretive studies	pg. 95
• selection of descriptive studies	pg. 97
• proposed method for selecting studies	pg. 97
<b>Search Strategy</b>	
• introduction	pg. 99
• search strategies	pg. 99
• publication bias	pg. 109
• searching for qualitative studies	pg. 110
• proposed search strategy	pg. 116
<b>Critical Appraisal</b>	
• introduction	pg. 118
• causes of bias and error	pg. 120
• appraisal of experimental studies	pg. 123
• appraisal of observational studies	pg. 126
• appraisal of interpretive studies	pg. 128
• appraisal of descriptive studies	pg. 134
• proposed method of appraisal	pg. 135
<b>Hierarchy of Evidence</b>	
• introduction	pg. 138
• current approaches	pg. 139
• ranking different types of research	pg. 143
• levels of evidence	pg. 152
• justification of this approach	pg. 154
• proposed hierarchy of evidence	pg. 155
<b>Data Collection</b>	
• introduction	pg. 157
• missing information and results	pg. 160
• data extraction from interpretive studies	pg. 162
• proposed method of data collection	pg. 164
<b>Data Analysis</b>	
• quantitative data synthesis	pg. 165

- qualitative data synthesis pg. 175
- descriptive data synthesis pg. 186
- proposed methods of synthesis pg. 187

<b>Chapter 5.</b>	<b>Evaluation of the Expanded Review Framework: The Protocols</b>	
	Introduction	pg. 193
	Overview of expanded review topic	pg. 194
	Music systematic review protocols	pg. 195
	Definitions	pg. 196
	Systematic review protocol I : effectiveness	pg. 198
	Systematic review protocol II : appropriateness	pg. 206
	Systematic review protocol III : feasibility	pg. 215
<b>Chapter 6.</b>	<b>Evaluation of the Expanded Review Framework: Review Results</b>	
	Demographics of identified studies	pg. 224
	Results of systematic review I	pg. 230
	Results of systematic review II	pg. 250
	Results of systematic review III	pg. 265
	Integration of the evidence	pg. 269
	Future research	pg. 271
<b>Chapter 7.</b>	<b>Comparison Review : Physical Restraint</b>	
	Introduction	pg. 273
	Overview of physical restraint review	pg. 274
	Physical restraint review results	
	• effectiveness	pg. 276
	• appropriateness	pg. 280
	• feasibility	pg. 282



	Summary	pg. 290
<b>Chapter 8.</b>	<b>Discussion</b>	
	The findings of the reviews	
	• integrative review	pg. 292
	• methodological critique	pg. 295
	• theoretical critique	pg. 298
	The conceptual framework	pg. 299
	The expanded review method	pg. 301
	Limitations	pg. 316
<b>Chapter 9.</b>	<b>Conclusion</b>	pg. 319
<b>Appendix 1</b>	Search strategy	pg. 324
<b>Appendix 2</b>	Critical appraisal checklists	pg. 329
<b>Appendix 3</b>	Data collection form	pg. 336
<b>Appendix 4</b>	Summary tables	
	• included RCTs - Hospital Patients	pg. 341
	• included RCTs - Procedure Patients	pg. 343
	• excluded clinical trials	pg. 345
	• interpretive and descriptive studies	pg. 349
<b>References</b>		pg. 352

## List of Figures

1	Components of the evaluation of interventions	pg. 65
2	Conceptual framework : dimensions of the evidence	pg. 66
3	Systematic review process	pg. 73
4	Ranking of research evidence evaluating health care interventions.	pg. 144
5	Systematic review on physical restraints	pg. 275
6	Type of restraints used in acute care setting	pg. 286
7	Graphical presentation of reasons for restraining people	pg. 289

## List of Graphs

1	Hospital patients: music and heart rate	pg. 231
2	Procedure patients: music and heart rate	pg. 231
3	Hospital patients: music and systolic blood pressure	pg. 232
4	Procedure patients: music and systolic blood pressure	pg. 233
5	Hospital patients: music and respiratory rate	pg. 234
6	Procedure patients: music and patients rating of pain	pg. 236
7	Sub-group analysis: intra-operative procedures - music and heart rate	pg. 240
8	Sub-group analysis: intra-operative procedures - music and systolic blood pressure	pg. 240
9	Sub-group analysis: post myocardial infarction - music and heart rate	pg. 245
10	Sub-group analysis: myocardial infarction patients - music and systolic blood pressure	pg. 246

11	Sub-group analysis: post myocardial infarction - music and respiratory rate	pg. 246
12	Hospital patients: music and anxiety	pg.252
13	Procedure patients: music and anxiety	pg. 252
14	Sensitivity analysis: all hospital patients - music and anxiety	pg. 253
15	Hospital patients: music and mood	pg. 255
16	Myocardial infarction patients: music and anxiety	pg. 260
17	Association between restraint and death	pg. 279
18	Cognitive impairment and physical restraint	pg. 289

## List of Tables

1	Taxonomy of literature reviews	pg. 47
2	The indexing of the same paper in MEDLINE and CINAHL	pg. 115
3	Categorising studies risk of bias according to individual criteria	pg. 125
4	Minimising bias in observational studies	pg. 126
5	Critical appraisal of observational studies	pg. 127
6	Critical appraisal for expanded review process	pg. 136
7	Levels of evidence	pg. 141
8	Quality of evidence ratings	pg. 142
9	Scales used to measure anxiety	pg. 251
10	Restraint reduction programs in the residential care setting	pg. 277
11	Stocktake of restraint reduction programs	pg. 278
12	Patient reactions to restraint	pg. 281

13	Family reactions to restraint	pg. 282
14	Proportion of residents restrained	pg. 283
15	Duration of restraint for acute and residential care patients	pg. 284
16	Type of restraints used in acute care setting	pg. 285
17	One sub-theme from patient-oriented reasons for restraining acute care patients	pg. 288

## **Abstract**

### **Background**

The past two decades has seen an increasing emphasis placed on basing health care on the best available evidence. However, existing research has come under increasing scrutiny, which suggests its quality was often poor. This problem has been exacerbated by the ever increasing volume of health care literature.

To address these difficulties systematic reviews have emerged as one of the most important ways by which research is summarised and communicated to its end-users. However, as these reviews have been primarily concerned with effectiveness, they have focused almost exclusively on randomised controlled trials. As a result, systematic reviews have excluded much of the research of nurses.

### **Purpose**

The purpose of this study was to develop a process to systematically collect, appraise, summarise and synthesise the findings of a range of different types of research.

### **Conceptual Framework**

To aid in the development of these expanded review methods, a conceptual framework was developed that addressed effectiveness, appropriateness and feasibility.

### **Method**

A search of the literature was undertaken to identify published reviews of different types of research, and discussions in the health care literature related to the conduct of research reviews. These reviews and discussion papers served as the basis for developing the expanded review methods.

## **Evaluation**

To evaluate the expanded review methods, two systematic reviews were conducted. The protocol and results of the first review on the use of music in hospitals are presented to demonstrate how the conceptual framework and expanded review methods enabled a broader evaluation of the topic. Selected results from the second review on the use of physical restraint are presented to demonstrate how the findings from a number of methodologically different types of research were incorporated into a systematic review.

## **Conclusion**

The conduct of the two systematic reviews clearly demonstrated that the proposed expanded review process was able to rigorously collect and summarise a range of different types of research. Additionally, the conceptual framework underpinning these reviews enabled each of the studies to be located logically and coherently during the synthesis of data

## Statement

This thesis contains no material or publications which has been accepted for the award of any other degree or diploma in any university and that, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan or photocopying.

Signed

Date

2/11/01

## **Acknowledgments**

I would like to acknowledge the support and assistance of several people that enable the completion of this thesis. Firstly, I would like to thank my supervisor Professor Alan Pearson. I would like to give special thanks to Dr. Mary Fitzgerald for her critique of the draft manuscript and the many valuable comments. Finally I would like to acknowledge and thank the assistance of Kate Cameron for her comments, proof reading and suggestions.



# **CHAPTER 1**

## **Introduction**

## **Introduction**

There is an increasing emphasis on ensuring that health care has a sound scientific basis. This interest started in medicine and has since spread to encompass all health care disciplines. However, finding the evidence to answer questions relating to specific clinical problems is becoming increasingly difficult because of the massive volume of literature. Additionally, when such evidence is located, issues related to variability in the quality of studies, contradictory findings and small samples, add to the complexity of determining what evidence should be used.

As a result of these issues, literature reviews have become an important part of the process by which research evidence is located, evaluated and then summarised. However, it has long been recognised that the quality of these literature reviews is also highly variable. To address this, systematic reviews have emerged as an important way to identify the best available evidence on a topic of interest. As these reviews represent the best evidence, they are starting to replace primary research as the basis of health care decisions.

During the past decade there has been a dramatic improvement in systematic review methodologies as witnessed by the many publications addressing different aspects of reviewing research. However, as this activity has primarily been undertaken by medicine, it is not surprising that the methods developed are those that are best suited to the questions posed by medicine. More recently nursing has also become interested in evidence-based health care, and in conjunction with this, in systematic reviews.

The methods developed by medicine have been adopted by nursing and as a result there is a growing number of systematic reviews published each year addressing nursing issues. These nursing systematic reviews continue to focus primarily on the

randomised controlled trial (RCT). The concern with this focus is that these systematic reviews may not necessarily answer all the questions posed by nursing. It can be argued that some areas of interest to nurses concern issues related to the psychosocial aspects of caring, which are less readily quantified, measured and evaluated by experimental research methods. Psychosocial aspects of care are an important part of nursing, and as such, should be part of the evidence on which nursing practice is based. Yet this type of evidence is not normally included in systematic reviews, and as a consequence, this evidence is also excluded from most health care decisions. On this basis it is argued that some of nursing's research has remained outside the evidence-based movement, and because of this, it has failed to influence health care delivery.

### **Purpose of Study**

The purpose of this study was to address the limitations in current methods of conducting systematic reviews by expanding the existing review methodology to enable a broad range of research to be incorporated into a review while still maintaining the rigour expected of these systematic reviews. Because of nursing's considerable investment in interpretive research, incorporation of this research into systematic reviews was a major focus of this methodological development.

The approach taken during this study involved two stages; development of an expanded review process, and a subsequent evaluation of this proposed method:

- Developing an expanded review method

This stage of the study investigated current approaches for reviewing research, the debates in the professional literature and how different types of research have been summarised in reviews. Because the development of an expanded review process was the central focus of this study, this represents the largest component of this study.

- Evaluating the expanded review method

This component evaluates the expanded review process developed for this study. To evaluate the proposed method a systematic review was conducted. The expanded review process was used to develop systematic review protocols and a review was conducted using these developed protocols.

## Structure of Document

In presenting this work, the structure differs from that of a thesis reporting primary research. The layout used is as follows:

### 1. Background

The aim of this section is to demonstrate the current limitations of existing methods and to justify the need for an expanded systematic review process to collect, appraise and summarise best evidence. The specific issues addressed include:

- the growing importance of evidence-based health care,
- reviews of the research literature, and
- limitations of current methods.

### 2. Development of the Expanded Review Process

This section discusses the current review methods and alternative approaches that enable the incorporation of a broader range of research evidence. This discussion provides the basis for the development of the expanded review process. As previously stated, because this is the major focus of this study, this section is the largest chapter in this thesis. Specific activities were:

- development of a conceptual framework to underpins this study,
- methodological discussion of each phase of review, and
- identification of the methods to be used in the expanded systematic review.

### 3. Evaluation of the Expanded Review Process

To allow the proposed expanded review process to be evaluated, a clinical topic was selected (music for hospital patients), systematic review protocols were developed, and then a review conducted guided by these protocols. This systematic review provided the means by which the proposed review method could be subject to an initial evaluation. Specific activities were:

- development of expanded systematic review protocol (using the music review as the example), and
- conduct of the expanded systematic review.

As a single review would not utilise all aspects of the proposed expanded review process, select findings of a second review are presented. This review addressing the use of physical restraint was conducted concurrently to this study, and provided opportunity to further develop and refine the proposed methods. The purpose of presenting small samples of the findings of these reviews is to demonstrate how a number of different types of research can be summarised in a coherent framework.

### 4. Discussion

Following the conduct and reporting of the music systematic review, the results and methods are discussed and critiqued. This section addresses:

- results of the expanded systematic reviews,
- the conceptual framework, and
- the expanded review process.

### 5. Conclusion

The discussion concludes the thesis by highlighting the major arguments and summarising the key issues that emerged from this study.

# **CHAPTER 2**

## **Background**

# Evidence Based Health Care

## Introduction

During the past two decades there has been an increased emphasis placed on basing health care decisions on the best available evidence. This interest in evidence based health care (EBHC), evidence based medicine (EBM) and evidence based nursing (EBN) started in medicine, but has since spread to encompass most health care disciplines (Richards 1996). While EBHC means different things to different people, the most widely used definition is;

*Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external evidence from systematic research (Sackett 1996 et al., p. 71).*

EBHC can best be described as a process, where appropriate information is collected, analysed, implemented and then the impact of its implementation evaluated. It has been described as a merging of clinical expertise and best available evidence (Sackett et al. 1996), and an attempt to place existing skills and methods on a sounder, more systematic footing (Booth 1996). This approach to health care has been claimed to be a new paradigm to guide clinical medicine (Sackett et al. 1996). However, this claim has been refuted and it has been argued that at best, it is a useful set of rules that aid in critically selecting evidence (Benitez-Bribiesca 1999).

There have been few proposed new concepts in health care in modern times that have generated such interest and activity. There are now few health care publications that

do not at some point proclaim its virtues, attack its limitations, or address a specific component of EBHC. It is a common feature of health care conferences, has dedicated medical and nursing journals, and is now starting to be integrated into post-graduate education. Its proponents come not only from the ranks of clinicians, but also from the ranks of economists, administrators, researchers and policy makers. It has generated its own industry, ensuring employment for professionals such as information technologists, librarians, clinical epidemiologists, statisticians, reviewers and researchers. EBHC has led to the development of specialist research centres that produce and summarise the best evidence research has to offer. The concept of EBHC and evidence based practice (EBP) is now starting to appear in the mission statements of some health care agencies and in job descriptions of health care workers. It has seen a shift in the direction of clinical practice, whereby best practice is determined not at from the bedside, but from the backroom. This description of the growth and spread of EBHC is even more impressive, given that most of this activity has occurred within a period of little more than twenty years.

Therefore it is important to describe the health care environment in which this study is located. This first section of the background describes what EBHC is and presents an overview of the health care climate that enabled the evidence-based movement to become such a dominant force. Following this, issues surrounding the complexities of determining which evidence should be used as the basis for practice are addressed. Finally, nursing is located within this movement and issues of specific importance to nursing are discussed.

### **What is Evidence Based Health Care**

Evidenced based health care can best be described as a framework that can be used to ensure clinical practice is based on the best available evidence. Discussions in the literature suggest it can be applied to individual patients, or used on an organisational



or health system level. Muir-Gray suggests that it provides an opportunity to bridge the gap between research and practice (Muir Gray 1997) and Thompson describes it as doing the right things right (Thompson 1998). While EBHC is often considered only in the context of clinical practice, it has wide implications for all aspects of the health care organisation (Hewison 1997).

EBHC is about doing “the right things, at the right time, for the right people and doing them right the first time” (Garbett 1998, p. 28). Firstly it is about doing the right thing, in terms of an appropriate and effective intervention. This highlights that interventions and treatments must be available at the right time if they are to achieve the intended outcomes. They must also be delivered to ‘the right people’, those who have been shown to benefit from the intervention. Finally, it is about implementing the intervention the right way.

An evidence based approach involves a cyclic process, starting with the identification and development of an answerable question and continuing through to the evaluation of the impact of implementing the evidence on clinical practice. It’s cyclic in that evaluation of its impact may well necessitate the beginning of the process again as new issues are identified. The steps in this process are to:

- identify a clinical problem then convert it to an answerable question,
- collect the evidence in a systematic and comprehensive manner,
- critically appraise studies to evaluate the rigour with which they were conducted and their risk of bias or error,
- summarise the evidence in an appropriate manner, which for some research will involve statistical synthesis, but for many clinical questions, will be achieved by a narrative discussion,

- determine the clinical effectiveness of the intervention, its cost effectiveness, and the benefits and harm associated with its use (including those related to personal preferences and circumstances of the consumer).
- disseminate and implement the evidence, and
- evaluate the impact of the intervention on practice.

The evaluation phase of this approach clearly distinguishes EBHC from the older research implementation models, in that implementing the evidence is not the end of the process. It is this evaluation that makes EBHC a cyclic process, in that the process may need to begin again with a new, or different, clinical problem.

A range of potential benefits has been attributed to the EBHC approach and these include that it;

- leads to improvements in clinicians reading habits,
- improves clinicians knowledge,
- provides a framework for teaching,
- enables junior team members to contribute to clinical decisions,
- allows better communication with patients,
- enables patients and health care workers to better make informed decisions,
- facilitates better organisation of information,
- allows ineffective and harmful methods of health care to be identified, and
- allows more effective use of resources (Bastian 1994; Hope 1995).

While this description of EBHC paints a somewhat 'rosy picture', it does represent the perspective of its proponents and highlights potential benefits to be gained by a shift from ritualistic and tradition based practice to a more rigorous scientific practice. However, the EBHC movement also has many critics, commentators and detractors.

The phenomenal rise in the status of EBHC has seen partnerships form among traditional adversaries. While clinicians, administrators and policy makers often clash on many issues of health care, EBHC has enjoyed support from all of these groups. Spodich uses a quote from Lewis Carroll's "Through the Looking Glass" to highlight one of the potential problems of EBHC:

*"When I use a word Humpty Dumpty said in a rather scornful tone, it means just what I choose it to mean - neither more nor less (Spodich 1996, p. 608)".*

This quote highlights the point that EBHC appears to have become 'all things to all people', in that each sees benefits for their own professional group. One editorial noted that EBHC has grown over the past 25 years or so, from a subversive whisper to a strident insistence, and suggested that revolutionaries notoriously exaggerate their claims (Editorial 1995). This editorial also criticised attempts "to foist evidence-based medicine on the profession as a discipline in itself. Charlton and Miles argue that EBM has labelled itself rational, objective and altruistic, and to oppose this implies promotion of practice that is illogical, self indulgent and opposed to the evidence (Charlton and Miles 1998).

It has also been suggested that some clinicians talk of EBHC with missionary zeal, and risk alienating others, and that a backlash to EBHC is not surprising because of the inflated expectations of outcomes-orientated and evidence based medicine (Batstone 1996). This has seen a division in nursing between those who support EBHC as a positive development and those who see it as a threat to nursing. One comment is that by uncritically following in medicine's footsteps we risk relegating

much of nursing's knowledge to the scrap heap of opinion and subjective speculation (Walker 1997). Ironically, while EBHC is based on the principle that interventions must be supported by evidence of their effectiveness, Goodman has noted that there is still no evidence to show that EBM provides better health care (Goodman 1999).

From the perspective of nursing, the critics of the evidence-based movement point out that much of clinical practice entails professional judgement often based on years of experience (Hunter 1999). Little is known about these many decisions that impact on recipients of nursing care and whether there is conscientious, explicit and judicious use of evidence (Lang 1999). Additionally, because of the relative newness of a nursing research tradition, much of the work of nurses is hidden (Hunter 1999). Upton suggests that nursing has been known to react enthusiastically to the call of contemporary concepts with little initial assessment of their effectiveness to nursing or patients (Upton 1999). In the past, nursing has adopted developments such as the nursing process, nursing diagnosis, research based nursing and nursing competencies, with evangelical zeal. The suggestion here is that uncritical acceptance of EBHC by the nursing profession may not necessarily be to the benefit of nursing practice.

With the advent of new academic nursing departments in the 1970s there was a plethora of nursing theories, usually of the 'armchair kind', on how nursing should be practiced rather than how it is actually practiced (Luker 1997). Adding to this, those who taught and researched nursing retained little continuing clinical commitment and responsibility (Mulhall 1995). When one also considers that much of nursing research has been conducted as part of higher degrees or as one-off studies (Luker 1997), the concept of evidence-based practice becomes more rhetoric than reality.

Therefore, for the nursing profession, how EBHC is perceived differs from that of medicine. For some, EBHC is seen as the next 'fad' that will exert a small influence on

practice as it is implemented to different degrees by indifferent clinicians. Others see EBHC as another example of nursing being condemned to forever follow in medicine's foot steps (Walker 1997). An alternative perspective is one of a profession having to underpin its practice on scientific evidence that is still at a developmental phase. In an environment where claims of 'no evidence of effectiveness' are often confused with 'evidence of no effectiveness', how EBHC is perceived by nursing becomes a very important issue. However, to a large extent, nursing has accepted the methods developed primarily for medicine, with little modification to suit nursing's unique knowledge. What EBHC should be for nursing is not to show other disciplines how nursing is different, but rather to demonstrate how nursing contributes to best patient outcomes and cost effectiveness (Hunter 1999). To do this, nursing must actively participate, define and develop EBN rather than simply be users of frameworks developed by others.

### **Why Evidence Based Health Care**

EBHC has become a major issue in health care as a consequence of many interrelated factors. Each has increased the complexity of care delivery, making an evidence-based approach to practice seem attractive. Ironically, the attractiveness has crossed the boundaries of traditional rivalry and has resulted in the clinician, administrator, researcher, policy maker and even the consumer of the service actively supporting and promoting the ideal of having health care decisions based on the best available evidence. The factors that have produced this collaborative call for evidence are diverse, but include:

- increased complexity of health care delivery,
- variation in practice,
- changing body of knowledge,
- separation of clinician and researcher,

- poor quality of research,
- increasing volume of literature, and
- rising cost of the delivery of health services.

### Increased Complexity

The complexity of health care delivery has progressively increased over the past few decades as new technology and demand for services grow. Muir Gray suggests that the growing need for evidence based health care is a result of factors such as the ageing population, rising expectations of consumers of health services in terms of ease of access, quality of service, and an increased demand for redress and compensation should there be a failure of service (Muir Gray 1997). As a greater proportion of the community enters the age where their use of health care services increases, organisations have been forced to ration services, accept long waiting lists or develop innovative approaches to health care. Community views of health care have also changed considerably, and many users now have a better understanding of illness and treatments, and as a result, have greater expectations. This greater understanding and knowledge of consumers has made Muir Gray's suggestion, of an expectation for compensation when service fails, a reality for many areas of the health industry.

Muir Gray also suggests the sophisticated product promotion and claims of effectiveness of new technologies, procedures, products and pharmaceutical agents add to the complexity of health decisions (Muir Gray 1997). The development of these products and services has been based more on commercial opportunities than on consumer need. Because of the increased cost of these products, decisions must be made regarding which products will be implemented, and what existing technology will be replaced. Comparing the benefits to be gained from these products is difficult because decisions must often be made on limited information. Additionally, improvement to health outcomes associated with any new technology is generally

small and so desirable outcomes, cost and potential harm must be considered (Muir Gray 1997).

### Variation in Practice

From another perspective, EBHC has gained support because one of its aims is to minimise variations in clinical practice. This variation is important because of its potential implication for patient outcomes, and also because poor quality of care results in unnecessary costs (Moore 1998). The rationale is that if practice is based on best evidence, and is similar between organisations, there will be minimal variation in practice and what variation there is will likely be the result of population differences rather than process differences.

One of the major causes of variation in practice is the weakness of current health care evidence (Smith 1991). This lack of rigorous evidence means clinicians must continue to base practice predominantly on expert opinion. However as the complexities of health care increase, and new technology continues to promote claims of improved outcomes, expert opinion becomes less reliable. Lopez-Jimenez and Lamas suggest that simple observation may disguise the truth, and that even with many years of experience, clinical experts can still be wrong (Lopez-Jimenez and Lamas 1999). Examples of incorrect decisions that have been made based on experience and observation include:

- In the 1950s people with stable angina were treated with clamping of the internal mammary artery based on the observable improvements post-operatively. These improvements of reduced pain, milder episodes of chest discomfort and improved electrocardiographic abnormalities were many years later shown to be present with a 'sham' operation (Chalmers 1972).

- Class 1a antiarrhythmic drugs were used for the treatment of ventricular arrhythmias, however some of these drugs have been shown to actually decrease survival (Echt et al. 1991).
- The Trendelenburg position is used by critical care nurses during hypotension emergencies despite a lack of supporting evidence that it produces any significant changes in arterial blood pressure or cardiac output (Taylor 1991).
- The use of saline instillation prior to endotracheal tube aspiration is used despite a lack of supporting evidence, and suggestions that it may do harm (Thompson 2000).
- The use of chlorhexidine to prevent and treat oral mucositis related to cancer treatment continues despite studies clearly demonstrating that it offers no additional benefits over saline solutions, and may do harm by removing normal oral flora (Kowanko et al. 1998).

Despite this acknowledged problem, expert opinion remains the major factor in the variation in practice, as opinion leaders impose their beliefs over local, regional or national practice. While EBHC is seen as a means to address this problem, it has been suggested that this variation in clinical practice remains one of the greatest challenges for EBHC (Cook 1996).

As an extension of this variation in practice, the evidence-based movement has also been seen as a tool for curbing excessive medical power and poor practice (Colyer and Kamath 1999). This aspect of EBHC has been well supported by most non-medical professional practitioners. It is seen as a means for professional groups to better direct and control their sphere of interest and limits the authority of medical opinion as the sole arbitrator of best practice. This view of EBHC would be well supported



by health care administrators and policy makers who, armed with agreeable evidence, could cut through the barriers of medical opinion. Yet as previously stated, the weakness of current evidence has made this more a wish than a reality.

### Knowledge Base

The body of knowledge on which health care is based is rapidly changing. Much of what is taught to the medical and nursing student remains relevant for only a small proportion of their professional lives. However, the rapid advances of research have not been matched by clinical practice which lags many years behind. Keeping up to date with current knowledge has never been more difficult. Antman *et. al.* found delays of 13 years from when thrombolytic drugs could have been shown effective by meta-analysis for the people with acute myocardial infarctions, till being recommended in most text books (Antman et al. 1992). There was a delay of nine years before the majority of the authors of texts recommended nitroglycerin and nitroprusside for routine use, and lignocaine has been recommended for the past 25 years for prophylaxis against ventricular fibrillation despite a lack of evidence (Antman et al. 1992). From this perspective, EBHC has offered a process that promises a shortening of the time between discovery to clinical implementation.

However, despite the expectations that EBHC will have an impact on the knowledge base of clinical practice, there are a number of issues that threaten this. Kleinert warns that EBM can not be used to make the hard choices however desirable it would be to shift responsibility (Kleinert 1998). Hunter also warns that EBHC can offer tidy solutions to matters of a more sensitive nature requiring the application of clinical judgement (Hunter 1996). The risk, Hunter suggests, is that politically subjective concerns may be unintentionally turned into technical objective ones for which the rational scientific approach is mistakenly seen as the answer. Perhaps one example of this is a recent proposal for evidence based organ allocation (Zenios et al. 1999), that

attempts to address issues of efficiency and equity under a broad 'evidence based' framework. This example highlights the risk of inappropriately simplifying complex issues.

Naylor highlights another difficulty confronting modern health care related to the knowledge base, in what has been described as the 'grey zones' where information is incomplete or contradictory, and where even the boundaries of these areas are themselves uncertain (Naylor 1995). To many clinicians, EBHC appeared to offer the promise of answering many of these difficult questions. Additionally, the offer by EBHC of producing the best available evidence for practice would have a considerable appeal to many clinicians.

#### Separation of Clinician and Researcher

Another area that may have contributed to the support of EBHC is the separation between the person that produces the evidence and its consumer. Because of the complexities of the research process, and the need for specialised knowledge and expertise in this area, most researchers are affiliated with the university rather than the health care setting. Because of the separation in nursing of researcher and clinician, the researcher is commonly seen as being divorced from the real world of practice (Mulhall 1995). However, there are fundamental differences between the clinician and the academic researcher. The academic researcher is able to pose good research questions without ever having to face the necessity of answering them (Mulhall 1995). In contrast, the practitioner requires answers. Contradictory results are a problem for practitioners, and a challenge for the academic. It has even been suggested that clinicians seek a state of 'optimal ignorance', knowing enough to be effective, but not so much so as to become constrained by uncertainties and ambiguities (Mulhall 1995). As a consequence, many nursing studies are only read by researchers (Mulhall 1995). Supporting this, an Australian survey found only 10% of nurses frequently

applied research findings to their practice, and that few regularly read journals (Nagy and Crisp 1992). The appeal for clinicians is that EBHC promises to summarise and disseminate the best evidence.

### Quality of Research

It has been claimed that a large proportion of published research is invalid (Rosenberg 1995), and for some studies, the risk of error or bias limits the usefulness of the findings. Evaluations of the standard of statistical analysis of published research have demonstrated that errors are common, vital information is often omitted, and the sample size used is commonly inadequate (Altman 1982; Lynn 1985; Murray 1988; Mills 1993). A study by Colditz *et al.* demonstrated that failure to randomise subjects in RCTs increased the likelihood of the results favouring the experimental intervention (Colditz *et al.* 1989). A study by Schulz *et al.* found that trials that had inadequate concealment of the allocation of participants to study groups resulted in the odds ratio being exaggerated by 41% in favour of the experimental treatment (Schulz *et al.* 1995). This study also found that trials that were not double blinded resulted in the odds ratio being exaggerated by 17%. Issues such as these make determination of what interventions are effective very difficult. A systematic review of fall prevention research found poor quality a problem for the majority of studies identified, with many failing to use the optimal research methods to evaluate the interventions, failing to report the methods used during the study or failed to provide adequate information about the findings (Evans *et al.* 1999).

However critical appraisal of research is difficult for many clinicians because they are not research experts and have had little training in the appraisal of research (Fowkes and Fulton 1991). As part of the systematic review process, all research is subject to some form of critical appraisal. Part of the success of EBHC may be related to this offer of providing clinicians with sanitised evidence, which releases them from the

need to undertake their own appraisal before selection of which interventions should be implemented.

### Volume of Literature

Booth suggests the rise in the popularity of EBHC is a result of the information explosion (Booth 1996). What he describes as “drowning in information, thirsting for evidence”, is a result of the annual publication of 20,000 to 30,000 biomedical journals and 17,000 biomedical text (Booth 1996, p. 25). For the novice literature searcher confronted with such a large volume of information, deciding what information should form the basis of clinical practice is difficult. Antman *et al* suggests that the number of clinical trials being produced in every specialty is now too large for the clinician to digest on an ongoing basis (Antman et al. 1992). Additionally, while research may have been conducted on a topic of interest, locating this research within this vast volume of literature is also difficult.

Electronic databases have been developed to aid the location of research papers, yet no single database covers all the health care literature. More recently the number of databases has also started to increase, each with its own area of coverage or specialty focus. As a result, determining which databases should be searched is also starting to become a challenge. To deal with this, systematic review search strategies have been developed to increase the likelihood that all relevant studies will be identified. The promise, again, is to free the clinician of the need having to search for relevant papers because they are presented as summaries compiled by the reviewer.

### Cost Efficiency

The potential for cost control and value for money that has been commonly linked to EBHC, would have appeal for health administrators and policy makers. However, this value for money ideal is difficult to define because of its different meaning for each

stakeholder. For policy makers it will likely represent cost, for the provider it would concern both cost and quality of care, while for patients it would relate to quality of life issues in terms of physical, psychological and social costs (Gerrish and Clayton 1998).

However, not all have supported EBHC and some have suggested that there are a number of hidden agendas. Some concerns arise from the fact that ineffective activities will not receive funding. Kleinert argues that EBHC has advanced to evidence based decision making and is now also evidence based rationing (Kleinert 1998). He suggests that the evidence-based 'cut-off', above which treatment will be funded and below which it will not, is a hopeless myth (Kleinert 1998). Hunter notes that a large proportion of health care interventions are unproven, and so the idea of funding only proven interventions is very seductive to politicians in terms of cost cutting and rationalisation (Hunter 1996). However, a lack of evidence does not indicate lack of effectiveness, a fact often lost in the debates surrounding EBHC.

This potential for cost reduction has seen government agencies playing a very active role by stimulating interest and helping to shape EBHC (Colyer and Kamath 1999). Charlton and Miles note that tremendous advances have been made in establishing the EBM brand, obtaining massive government funding and manoeuvring to a position of unchallenged authority (Charlton and Miles 1998). The outcome of this process, they argue, has been the takeover of clinical consultation. The managers and statistical technocrats have been empowered to define best practice, but have not accepted the responsibility for the clinical consequences (Charlton and Miles 1998).

This serves to highlight the potential power of the concept of EBHC, and Clarke suggests that the evidence carries authority, and so is a source of potential tension between the clinician and politician (Clarke 1999). However, despite this tension,

both clinicians and government advocate for EBHC, although perhaps for quite different reasons.

In summary, the popularity of EBHC is multifactorial, in that it has been seen as a way to deal with a range of different issues that impact on practice. Interestingly, it means a different thing for each of the stakeholders in health care, and while their support is offered, each will likely anticipate different outcomes. This has important implications for the actual processes of the evidence-based movement, because not all these differing expectations will be achieved. Additionally, the optimal processes of one health care stakeholder are not necessarily best for others.

### **The Evidence**

The definition of EBHC by Sackett *et al* suggests that it is the integration of clinical expertise with the best external evidence (Sackett 1996). This implies that the external evidence will have a major impact on the nature of clinical practice, therefore, what constitutes this external evidence is critical. However, determining what type of evidence should be used is both difficult and complex, and currently there are many competing opinions.

Part of the difficulty relates to the many different agendas within the evidence-based movement. Ray suggests that there are many stakeholders in EBHC, such as practitioners, consumers, policy makers, researchers, insurers, administrators, economists, and the medical industry (Ray 1999). The agenda of each of these stakeholders influences what they perceive to be relevant evidence. Examples of these agendas suggested by Ray include:

- Fiscal accountability: which involves managing within limited budgets and maximising profits, where evidence of positive outcomes is used as a competitive tool.
- Risk and liability: which focuses on harm reduction research, where achieving a positive outcome is replaced by avoiding negative ones.
- Quality care: whereby quality care is framed as evidence-based care, and so clinician expertise and patient preference are noted.
- Social and moral responsibility: where the standpoint is more closely aligned to the public and consumers of health care, focusing on consumer/professional alliances.
- Professional effectiveness: whereby research evidence is used to demonstrate the cost effectiveness and unique contribution of the profession (Ray 1999).

To this list could be added the different agendas of medicine and nursing. While both disciplines would share the 'professional effectiveness' agenda, their other agendas differ. For medicine it would primarily focus on 'cure', while nursing would put greater emphasis on 'care'. While it must be acknowledged that the evidence related to cure and care are not mutually exclusive, there are sufficient differences to argue that some of the evidence required by each of these disciplines must also be different.

The perspective of the health care consumer and health care provider towards illness would also differ. For the patient and family the experience of illness is highly contextualised, with interpretations stemming from a narrative model of disease (Ray 1999). Policy makers and epidemiological researchers, in contrast, work in aggregates of experience that limits the attention given to individual experience. The clinician perhaps shares both these worlds, conceptualising experience in both the narrative and

aggregative models. Finally, selection of one particular research model over other models, also limits what can be counted as evidence. Ray argues that the researcher's perspective is limited from the start of any investigation, suggesting:

*"Each investigator, by virtue of his or her disciplinary, historical, political and economic situatedness starts from a pre-figured range of research options"*  
(Ray 1999, p. 1022).

For nursing, the most vital issue in the many debates surrounding EBHC is what should constitute the evidence for practice. If nursing is to demonstrate its value, and to compete successfully in the health care market place, then like other professional groups, the evidence must answer the questions of the profession. Unfortunately, the literature clearly demonstrates that there are many different views regarding the nature of the best external evidence.

#### Views of the Evidence

Until recently, the major focus in the search for evidence was the RCT, with the results from research utilising other methodologies given a low status. This ranking of research evidence is based on the degree of bias of the different methods for evaluating the effectiveness of an intervention (NHMRC 1995). Within this framework, the best evidence is considered to be that generated by the systematic review and RCT. The evidence produced by observational studies is commonly considered to be at a lower level than that of the RCT (NHMRC 1995). The lowest level of evidence is considered to be that generated by descriptive studies, reports of expert committees and clinical experience (NHMRC 1995). However, this approach fails to acknowledge the valid contribution of a range of research methods.



The RCT was originally proposed and used to test the average efficacy of agricultural interventions, to remove subjective judgement and therefore the risk of bias (Feinstein 1994). Since this time, the RCT has found favour in health care as the best method to evaluate the cause and effect relationships of an intervention (Dawson-Saunders 1994). The researchers achieve this by randomly allocating participants to either the treatment or comparison, having strict inclusion criteria, and measuring the outcomes of both study groups in a similar manner. As a result of these processes, both groups are equal for all characteristics, other than the intervention, and so any differences in the outcome between groups can more easily be attributed to the intervention. The robustness of the finding of RCTs is the major reason for its wide spread support.

Nursing has not utilised the RCT research design to any great extent, and as a result, the evidence produced by the profession has fitted poorly within the evidence-based framework. Kitson noted that the rules of the evidence-based movement relate to medical diagnosis, single clinical interventions, RCTs and meta-analysis (Kitson 1997). She suggests that while there are limits to nursing evidence conforming to these criteria, nurses must not be excluded from the evidence-based movement because their research is poor, or insufficient in rigour or size. Bonell suggested that nursing's hesitance in applying quantitative and experimental methods has the potential to marginalise nursing research (Bonell 1999). It has even been suggested that the striving for distinction in nursing research methods may, in part, reflect a desire by some nursing researchers to distance themselves from the more dominant medical model (Poole and Jones 1996). Smith suggested that the allowable evidence is now extending beyond the RCT, but still remains numerically based (Smith 1996). He suggests that denial of social and psychological aspects may be detrimental, and that ignoring the less readily measurable dimensions of health care may be dangerous.

Concern regarding the evidence generated by meta-analyses has also been raised. One comparison of the findings of meta-analyses with the findings of large RCTs on the same subject identified only fair to slight agreement between the two (LeLorier et al. 1997). Despite agreement in direction of effect, the estimated size of the treatment effect was quite different. A similar problem was identified by Prins and Buller who noted the different conclusions reached by four meta analyses of the same group of studies evaluating optimal frequency of aminoglycoside administration (Prins and Buller 1996). This led to the question that if the RCT, considered the gold standard, disagrees with a meta-analysis which one should be believed (Bailar 1997). One suggestion for clinicians when choosing between competing meta-analyses was to select the meta analysis that most closely agrees with their own beliefs (Prins and Buller 1996).

Bailar notes that it is the last step of the systematic review process, determining the single estimate of effect from multiple studies, that has attracted the most criticism (Bailar 1997). He suggests that this is because there is often statistical evidence or biological reasons, or both, to show that studies included in the meta-analysis have in fact measured different things. Bailar argues, that while we can accept the results of a well done meta-analysis as a way to represent disparate studies on a common scale, any attempt to reduce the results to a single value with confidence limits, will likely lead to conclusions that are wrong (Bailar 1997). In response to this, he supports the traditional narrative review of the literature as the preferred option. Others have adopted a slightly broader view of systematic reviews by undertaking meta-analysis on the findings from observational studies such as case control and cohort (Jones 1992; Stroup et al. 2000). However, meta-analysis of these studies has been questioned because of their lack of reliability and spurious results (Shapiro 1994). As a result of this, Shapiro proposes that meta-analysis of non-experimental data should be abandoned (Shapiro 1994).

The view of what constitutes best evidence has broadened in recent years. Health care disciplines are starting to use different forms of evidence to provide answers to the diverse range of clinical problems. As part of this broadening of the perception of what constitutes the evidence, the RCT has been subject to a number of attacks in the literature. Hopkins suggests that the results of RCT cannot always be translated to the general population by the health care professional because they are based on a carefully selected and defined population, often have an upper age limit, and the intervention has usually been carried out by senior clinicians with a high degree of skill (Hopkins 1995). Feinstein notes that the information provided by the RCT is only the average rather than the individual response (Feinstein 1994). This means that regardless of the effectiveness of an intervention, as demonstrated by the RCT, individual response will vary around the mean. That is, some people will benefit from receiving the intervention, others may not. The risk from this perspective is that the very specific inclusion and exclusion criteria of the RCT, may result in the study population not truly representing the real world of practice. One commentator supporting this criticism, even suggested that the evidence underpinning EBHC does not provide information about individual patients and that no amount of “statistical jiggery pokery” will change that fact (Charlton 1997).

Poole suggested that the opposition to the RCT as best evidence relates to concerns regarding the rigid objectification, extreme rationalism, detached fact acquisition and the manipulation of humans (Poole and Jones 1996). Benitez-Bribiesca attacked the rigid criteria used to select ‘best evidence’, because information that does not conform to these criteria is lost to practice (Benitez-Bribiesca 1999). To rely dogmatically on this type of evidence, Benitez-Bribiesca suggests, ignores the fact that randomisation in clinical trials is intended to produce a false sense of homogeneity. Carr-Hill suggested that acceptable evidence is that which can be summarised on one graph, because complexity is devalued and context denied (Carr-Hill 1995). Black argued that the complexity of some clinical problems makes them insoluble by the RCT (Black

1998). Dunn suggested that scientific rigour, RCTs and statistical significance run the risk of tyrannising our understanding and appreciation of evidence-based practice (Dunn 1998). She argued that nursing must look beyond the RCT to have a balanced appreciation of the contribution of various research approaches. This view was shared by Clarke who argued that evidence from any source can be helpful in improving the effects of practice, and so best evidence means using a wide range of evidence, from anecdote and experience, to meta-analysis and systematic reviews (Clarke 1999).

Hewison raised some fundamental issues concerning epistemology in terms of what is the most suitable scientific procedure for the acquisition of knowledge (Hewison 1997). The higher levels of evidence proposed in these hierarchies of evidence imply a view of the world from an objectivist framework of what can be measured, touched, felt or counted. However this eschews alternative and complementary explanations of reality arising from data generated from methods other than the RCT (Hewison 1997). The concern is that qualitative data, about feeling, attitude and experience, appear to have no place in the hierarchies of evidence and so the risk is that vital information will be excluded from the decision making process.

From a slightly different perspective, Kerridge *et al* noted the large quantities of trial data needed to meet the standards of evidence based health care, means that few treatments or interventions have the evidence to support their effectiveness (Kerridge *et al.* 1998). However, this limited availability of evidence is also influenced by inequities in research funding. The result of this inequity is that rigorous evidence is available on some interventions such as new pharmaceuticals, but lacking for others such as palliative care (Kerridge *et al.* 1998). It has also been noted that the modern trend is to search for precise answers in the form of numbers and probabilities, however these can only have a limited role in human sciences (Kleinert 1998). Sorrow,

grief and despair, or relief, comfort and joy are poorly addressed by statistical methods.

In conjunction with this critique of the RCT, others have highlighted the important contribution of other sources of evidence. Colyer and Kamath argued that qualitative approaches allow exploration of patients' beliefs and understandings, and without such insights, they suggest that clinical practice is unlikely to be either cost or clinically effective (Colyer and Kamath 1999). Smith noted that while qualitative and anecdotal evidence have developed considerably, they are still not mentioned as evidence (Smith 1996). Another commentator suggested that much of what nurses do is not amenable to the RCT (Walker 1997). He argued that blindly following medicine into their version of the evidence-based movement threatens the efforts of many nursing scholars who have been working to re-define and reposition nursing to something more like a sophisticated, deeply context dependent, human craft. MacNaughton also suggested that much of what the clinician does is not amenable to testing by trial, and that it is based on knowledge obtained in other, no less valid ways. (MacNaughton 1995). He suggested it is anecdotal evidence, rather than empirical, that is important in diagnosis, giving explanations or planning treatment, because this evidence allows the clinician to see patients in the context of their lives.

Bhopal and Tonks suggested that letters to the editor are also an important part of the research process and so should constitute part of the evidence (Bhopal and Tonks 1994). They argue that it is only after a research report has been published that it can be exposed to critical appraisal and comment from the research community, and that this appraisal is communicated to others via letters to the editor. Unfortunately, these letters carry little weight, and are often difficult to locate.

In Australia, what constitutes evidence has been greatly complicated by the legal decision from a 1997 court case between the National Health and Medical Research Council (NHMRC) and the Tobacco Institute of Australia (Jamrozik et al. 1997). During the development of recommendations related to passive smoking, NHMRC excluded literature written by people known to have been associated with tobacco companies. This exclusion included reports that were acknowledged as having been prepared at the request of the tobacco industry and so were not subject to the normal peer review process like other research. However, the legal challenge by the tobacco industry was successful, with the court ruling that the NHMRC had erred significantly in regard to the consultative procedures related to the evaluation of available evidence. This finding significantly adds to the complexities of defining what is the best evidence for health care.

In summary, it is likely that each discipline determines what constitutes their evidence, and it is shaped according to the context of its use. For medicine, their major activity is the administration of varying forms of treatments to achieve a beneficial physiological outcome, and so their evidence must focus on 'cause and effect' relationships. From this perspective, evidence generated by RCTs will be vital for the medical profession. However, it can be argued that while the focus of nursing involves 'cause and effect' relationships, it also includes a large number of psychosocial outcomes as part of the care provided. This caring involves not only the patient, but also their family, and that some of the important outcomes of nursing care relate to the intangible psychosocial aspects of the person. However, as previously discussed, many of these aspects of nursing care do not lend themselves to trial and statistical analysis. So in terms of the best evidence for nursing practice, the contextual-narrative perspective of qualitative research is as important as the cause and effect evidence of experimental research. However, despite its importance to nursing, this evidence has largely remained outside the boundaries of the evidence-based movement.

Nursing must develop and communicate its own definitions of best evidence if it is to maintain its independence and unique focus in the health care system. This evidence will not only provide the foundation of practice, it will aid in the evaluation of clinical practice and help demonstrate the value of nursing to other professions, consumers and policy makers. Unfortunately, despite considerable literature addressing the inadequacies of current approaches in the context of EBHC, there have been few rigorous attempts to change the processes to better answer nursing's questions.

## **Reviews of the Research**

### **Introduction**

Literature reviews play an important role in the advancement of a discipline because they accumulate past research endeavours, summarise major issues and are an important way by which the information generated by a large number of individual studies can be disseminated. 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 all past research. Regardless of their scope, they differ from other research endeavours because they are based on completed work and processed data, rather than being new primary 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 (Evans and Kowanko 2000). Without this rigour, summarised information is at risk of the same errors that threatens the credibility of primary research. Unfortunately, the traditional approach to reviewing the research literature has lacked scientific rigour and as a result has not been held in high regard. In addition to this, literature reviewers have had to cope with an ever increasing volume of health care research. The quality of this 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 studies supporting a particular view point are included, rather than the review representing an accurate and balanced summary of all the research on the topic.

To address these problems, there has been a progressive evolution in review methods (Evans and Kowanko 2000). This development has seen the status of reviews change to the point where some reviews are now seen as research in their own right.



## **Evolution of The Literature Review**

Because of the cumulative nature of science, 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). This information is digested, sifted, classified and synthesised (Manten 1973). Because reviews deliberately aim to accumulate 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 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 primary research reports (Cooper 1980). 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 because of this, the findings could be biased by factors that would be unforgivable in primary research (Glass 1981).

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

- 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
- used vote taking, or counted studies with similar results.

Even in the late 1980's, 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, citing a small number of studies no longer did justice to the literature, and prompted the term "the misinformation explosion" (Glass 1976). This volume of literature has continued to grow with an annual publication of 20,000 to 30,000 biomedical journals (Booth 1996). This makes presenting 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 results, must also be addressed.

### Types of Reviews

One attempt to describe the potential focus of literature reviews by Cooper identified three different types of reviews (Cooper 1984):

- Integrative Research Review - summarise past research, draws overall conclusions, highlights unresolved issues and can provide direction for future research.
- Theoretical Review - presents theories to explain phenomena, compares them in terms of breadth, internal consistency and the nature of their predictions.
- Methodological Review - examines and critiques the research methods and operational definitions that have been applied to a problem.

Cooper suggested that a comprehensive review would likely address two or more of these areas (Cooper 1984). The integrative component of a review addresses the

results generated from primary research. The theoretical component move beyond integration of findings and attempts to explain the phenomena and predict with theories. The methodological component of a review addresses rigour, methods and the risk bias.

In the late 1980s Cooper (Cooper 1988) developed a taxonomy of reviews, which was an important development because it provided a framework for reviews that had previously been lacking (see table 1).

**Table 1**  
**Taxonomy of Literature Reviews**  
 (Cooper 1988, p.109)

Focus	Research Outcomes Research Methods Theories Practices or Applications
Goals	Integration Generalisation Conflict Resolution Linguistic Bridge building Criticism Identification of Central Issues
Perspective	Neutral Representative Espousal of Position
Coverage	Exhaustive Exhaustive with Selective Citation Representative Central or Pivotal
Organisation	Historical Conceptual Methodological
Audience	Specialised Scholars General Scholars Practitioners or Policy Makers General Public

In Cooper's taxonomy the first characteristic of a review, the "Focus", relates to the material that is central to the review's focus and includes research outcomes, research methods, theories or the application of research. It is noted that it is rare for a review to have a single focus, but most will have two or three foci given varying degrees of attention. The second characteristic, the "Goal", concerns what the author hopes to accomplish with the review. Integration and synthesis are common to most reviews to some degree. The goal may also be the critique of previous research or the identification of central issues. Cooper suggests these central issues may involve the identification of questions that have been dominant in the past, questions that should be a focus for the future or methodological problems that have not been resolved.

"Perspective", the third characteristic, may be neutral, where the reviewer attempts to present all arguments, or the more active espousal position that presents a particular argument, paying only limited attention to other views. Cooper notes that regardless of neutral or espousal perspective, the reviewer may still take a strong position based on the accumulated evidence. "Coverage" of the review may range from exhaustive to only a representative cover of the available research.

The fifth characteristic, "Organisation", is how the literature is presented in the review. It may be presented historically with topics presented in chronological order, conceptual order with literature relating to the same ideas presented together, or methodological order that groups literature according to the methods employed by the primary researcher. Finally, "Audience" covers the differing audiences for whom the reviewer is directing the work.

These early 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, ranging from a

simple discussion to a comprehensive coverage of the literature using methods that have a rigour similar to that of the research they summarised.

Since the development of Cooper's taxonomy of literature reviews, review methodologies have continued to develop. This development has included:

1. meta-analysis,
2. structured reviews,
3. qualitative reviews, and
4. systematic reviews.

#### 1. Meta-analysis

The traditional approach to summarising the results from more than one study was narrative discussion, using this to highlight studies with similar or contradictory findings. These narrative reviews of research have been around for as long as there has been scientific literature (Petitti 1994). However this approach brought with it certain limitations. Firstly, while a narrative discussion is possible with a small number of studies, it is inadequate when there are large numbers of studies addressing the same topic. As the number of published studies increased, the difficulties of using this approach also increased. Secondly, because of the lack of rigour in this approach, creating an accurate summary of the findings from many studies is difficult. For example, if studies report contradictory results, it is difficult to determine which results to accept. Glass even 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 also went as far as to suggest that narrative summaries were "airy speculation, unbecoming an empirical science" (Glass 1976, p. 4). Finally, attempting to summarise the results of many studies in a short narrative discussion means a large amount of information will be lost because it is impossible to deal with this magnitude of information in a simple discussion. To

overcome these difficulties, statistical methods have been developed that are able to deal with large amounts of numerical data.

The term meta-analysis was proposed by Glass in 1976 as a method for the integration and statistical analysis of data from a number of independent studies (Glass 1976). However, 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 used to systematically comb the results of research in order to reach conclusions about the body of research (Petitti 1994). Glass suggested that the analysis of research data is at three levels (Glass 1976). The first level, primary analysis, is the original analysis of the research. The next level, secondary analysis, is the re-analysis of research data to better answer the research question or to answer new questions. The third level is meta-analysis, which 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 combined (Dickersin and Berlin 1992). Meta-analysis has now been used extensively in reviews as a way of dealing objectively with the large amounts of data. More recently in response to the variability of results between competing meta-analyses, the term “meta meta-analysis” has emerged (Katerndahl and Lawler 1999). This term has been used to represent the analysis of two or more meta-analyses.

## 2. Structured Reviews

Another approach to summarising the literature is the conceptually broader structured review, sometimes also termed a comprehensive review or an integrative review (Evans and Kowanko 2000). 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 (Stevens and Milne 1998) that structured reviews are needed because end users, such as planners of healthcare, often require a reliable perspective on a topic that goes wider than that provided by a systematic review. These reviews are often needed on a short time scale

to address immediate problems, rather than the lengthy process of the systematic review. Structured reviews are often used by organisations, institutions and government bodies as the basis for planning. The need for these reviews is also a response to 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 a limited literature and an element of modeling to provide useful conclusions (Stevens and Milne 1998). Structured reviews are appropriate for some nursing topics because of the limited research available on which to base recommendations, and as many issues of importance to nursing are broader than the single intervention. Opposing these benefits is the fact that basing recommendations on limited evidence brings with it a greater risk of error. However, despite the potential usefulness of structured reviews, there is very little published information and even less attempt at ensuring current methods used in these reviews are appropriate.

### 3. Qualitative Reviews

Recently there has been growing discussion in the health care literature addressing the synthesis of qualitative research (Estabrooks et al. 1994; Jensen and Allen 1996; Sandelowski et al. 1997; Popay et al. 1998; Sherwood 1999; Evans and Pearson 2001a). A small number of reviews of interpretive studies have been published addressing issues such as chronic illness (Thorne and Paterson 1998), the concept of caring (Sherwood 1997), wellness and illness (Jensen 1994) and stressors in families with a child with a chronic condition (Ogden-Burke et al. 1998). However, methods for reviewing, summarising and synthesising interpretive research are still evolving and optimal approaches have yet to be determined (Evans and Pearson 2001a).

#### 4. Systematic Reviews

Systematic reviews are an extension of the traditional literature review and are used to review and summarise past research efforts using the same principles and rigour that underpins primary research. As the name suggests, they are systematic in their approach, and in contrast to the traditional literature review, they use explicit rigorous methods to identify, critically appraise and synthesise relevant studies (Mulrow 1997). These methods are pre-planned and documented in the systematic review protocol (Evans and Kowanko 2000). On completion of the review the methods used are reported, as with all research, to allow its validity to be evaluated (Evans 2001). The steps in the systematic review are listed below.

- Systematic Review Protocol

To ensure the systematic review is conducted with the same rigour as primary research, a predetermined plan is used (Jones and Evans 2000). This review protocol is similar to any proposal for primary research in that it carefully describes each step in the review process (NHS Centre for Reviews and Dissemination 1996; Mulrow and Oxman 1997).

- Review Question

Systematic reviews seek answers to specific questions, and the development of this question provides the basis for the search, selection and synthesis of studies. The questions addressed by systematic reviews are the same as those posed by the primary researcher.

- Selection of Studies

The study population for systematic reviews is all past primary research on the topic of interest. The selection of which of these studies should be included in the review is guided by eligibility criteria that forms part of the review protocol. These criteria operationalise the review question by explicitly stating all its components in a form



that can be used during the selection process (Petitti 1994; NHS Centre for Reviews and Dissemination 1996).

- Search for Studies

A comprehensive and unbiased search is one of the major differences between traditional literature reviews and a systematic review (Mulrow and Oxman 1997). The aim of the search is to minimise the number of relevant references that are missed, while also minimising the retrieval of non-relevant documents (Critical Reviews Advisory Group 1996). To increase the likelihood of finding all relevant studies, a search strategy is developed and used during the review.

- Critical Appraisal

As part of the systematic review process, all studies to be included in the review are assessed for methodological rigour (Evans 2001). This critical appraisal aims to discover if the methods and results of the research are sufficiently valid for the findings to be considered useful information (Fowkes and Fulton 1991). The rationale for this appraisal is that the use of only high quality studies should lead to better and more realistic estimates of treatment effect (Moher et al. 1995; Lohr and Carey 1999).

- Data Collection

The data that are used during the analysis phase of the systematic review are the results from individual studies. To analyse this data it must first be collected from the individual study reports, a process that brings with it the potential for errors during the transcription (Jones and Evans 2000). The aim is to collect data that are reliable, valid and free from bias (Petitti 1994). To aid in this, a data collection form is developed specifically for each review.

- Data Synthesis

The aim of this phase of the review is to synthesise the findings from

individual studies to provide an estimate of the effectiveness of the intervention. However it also allows the reviewer to investigate whether the effect is roughly the same in different studies, settings and participants, and if the effect is not the same, to investigate the differences (NHS Centre for Reviews and Dissemination 1996). The synthesis is achieved by a narrative discussion of studies or through statistical synthesis (Evans and Kowanko 2000).

There has been considerable methodological development of systematic review methods to increase the validity of findings. However, despite this activity, two issues limit the value of systematic reviews for nursing. Firstly, as the development of review methods has been undertaken by health care professions other than nursing, these reviews in their current format are not well suited to answering all the questions posed by the nursing profession. From this perspective, these reviews have become gatekeepers of nursing research. Secondly, the focus of systematic reviews has been to determine the effectiveness of interventions, yet for topics that have had little evaluation, determining what is not known is also very important. These two issues are addressed below.

### **Reviews as Gatekeepers of Knowledge**

It is not surprising that as a result of the large volume of research, and the difficulties in summarising their findings, the importance of reviews is growing. The single document of the review replaces the many individual study reports, and so frees the decision maker from having to rely on the mass of published research. These reviews aid in the translation of research evidence into practice because the identification, selection, appraisal and summary do not have to be performed by the end-user of the research (Evans and Kowanko 2000). As previously discussed, there is a range of

approaches to reviewing the literature, however, the systematic review is now seen to be the most reliable (Evans 2001).

As reviews are starting to replace primary research as the basis for clinical decisions, it means that any studies excluded from the review are also excluded from the decision making processes. The assumption from this perspective is that current best evidence is represented by the research summarised in the systematic review. By default this also means that poor or untrustworthy research is that which is excluded from the review. While this approach should not be a problem, because only quality research is included in a systematic review, how quality is defined becomes a critical issue (Evans and Pearson 2001b). Unfortunately, current systematic review methods, and the definitions of what constitutes quality evidence, are better suited to the questions raised by medicine than by nursing.

### The Nature of Reviews

Throughout the evolution of review methods, the systematic review has focused almost exclusively on experimental studies, epitomised by the RCT. The aims of these reviews have been the evaluation of the effectiveness of interventions. Because descriptive, observational and interpretive research generally seeks answers to different questions, they have found little place in the systematic review. So, as previously discussed, while systematic reviews answer questions related to what interventions work, they fail to adequately address some of nursing's questions related to caring or to the impact of illness and treatment (Evans and Pearson 2001b).

It can be argued that the approach of the systematic review should be used for all summaries of the research, ensuring that the care and rigour that was utilised by the primary researcher is maintained by the reviewer. However, this concept has not received wide acceptance and so the focus of most systematic reviews has remained predominantly fixed on the RCT. As a result of this, nursing research is collected,

sorted, appraised and synthesised according to medicine's concept of what constitutes quality evidence.

As a result of these processes, a good proportion of nursing research has been classified as 'lower level' evidence and therefore excluded from systematic reviews. While the methods needed to answer some of nursing's questions related to caring would differ from those used to address effectiveness, they are no less valid. From this perspective, systematic reviews in their current form have become gatekeepers of nursing knowledge (Evans 2001). As gatekeepers, these reviews are presenting health care decision makers with only part of nursing's research.

### The Challenge for Nursing

It has been argued that systematic reviews represent the 'gold standard' in research summaries (Evans 2001). On this basis the methods used should therefore be able to address all types of health care research. Additionally, these reviews should also be able to summarise research evaluating interventions from a range of different perspectives. That is, systematic reviews should not simply be for the synthesis of RCTs to determine the effectiveness of an intervention, rather they should provide a framework by which the findings generated by many independent studies can be synthesised to provide valid evidence on a topic of interest.

While the nursing profession has a considerable investment in interpretive research, these studies have had little impact on the evidence-based movement (Evans and Pearson 2001a). One factor in this marginalisation has been their exclusion from systematic reviews. Yet, if this research is not incorporated into reviews, it will fail to have any significant influence on health care decisions. Some early work has begun on how best to synthesise the findings of qualitative studies, based on the argument that single isolated studies do not in themselves contribute significantly to our understanding of phenomena of interest (Jensen and Allen 1996). Termed meta-

synthesis (Jensen and Allen 1996; Sandelowski et al. 1997; Sherwood 1999), it does not attempt the accumulative approach utilised by meta-analysis, rather it is more an interpretive exercise (Noblit and Hare 1988). However, to date, meta-analysis and meta-synthesis have been seen to be mutually exclusive, in that they are never both used in the same review (Evans and Pearson 2001a).

Nurses have also made extensive use of descriptive studies, yet these are not commonly incorporated into systematic reviews. These studies provide a simple description of specific events and so the strength of this evidence would be greatly increased by the synthesis of findings from many individual studies. This increased strength would be the result of findings being derived from a broader range of populations, settings and circumstances. Additionally, descriptive studies would benefit more from a summary using rigorous methods than by the unstructured approach commonly utilised in the past. However, these studies have remained beyond the scope of systematic reviews.

Another challenge for nursing is the existing hierarchies of evidence that are used to rank evidence produced by the differing research methods. While many scales exist, they have been developed primarily for medical research (Sackett 1986; Woolf et al. 1990; Cook et al. 1995; Guyatt et al. 1995; NHMRC 1995). This medical focus means that the RCT continues to be hailed as the gold standard despite its inappropriateness in many situations. While hierarchies have now been developed for issues such as diagnosis, prognosis and economic evaluations (Ball et al. 1998), they remain medically driven.

The findings of studies evaluating interventions from other perspectives, such as the reaction of people to the intervention, continue to be ranked as lower level evidence. Even though a study may utilise the method most appropriate for the research question, they are ranked as low level evidence because the standard by which they are appraised is in terms of their ability to provide evidence on effectiveness. As in all

areas of research endeavour, the highest level evidence is that which is generated by the research method best able to reliably answer the question. Despite this, current methods of ranking research disadvantage some of nursing's research.

As systematic review methods are still evolving, there are many areas in need of further exploration and development (Evans and Kowanko 2000). Additionally, while the review process that is used to synthesise the findings of RCTs is inadequate for non-RCT research, there are no other methods that are currently used with systematic reviews (Evans and Pearson 2001a). These processes are important because, like all research, systematic reviews must be pre-planned and be able to demonstrate the decision trail. Without this decision trail it becomes impossible to distinguish the systematic review from traditional reviews, and as part of this, to determine its rigour.

### **Summarising What We Know**

Systematic reviews focus predominantly on issues of effectiveness, and it is argued that they only present a summary of part of the total available evidence. Systematic reviews also have the potential to play a crucial role in bringing together all that is known on issues where research is still in the developmental stage, yet this has been largely ignored.

### **Reviews and the Development of Knowledge**

As previously discussed, the cumulative nature of science means trustworthy accounts of past research are an important part of the development of scientific knowledge (Cooper 1984). These reviews link past research endeavours to the ever advancing research front, and in so doing, prevent earlier work from being lost to the profession. From this perspective, the review provides a stocktake of all that is known on a topic providing the foundation for future research efforts. This

foundation is in the form of a link between past and current knowledge that contributes to the knowledge development by allowing a progressive growth of scientific knowledge as new research builds on what is already known.

This periodic stocktaking of knowledge has become even more important as the volume of literature continues to increase. As the number of health care journals is now numbered in the tens of thousands, and there are millions of citations in databases, it is reasonable to suggest that much of the published research will be lost in this massive amount of literature. This loss is in terms of studies no longer being identified once the journal is removed from the 'recent acquisition' shelves of the health care library. The risk, in terms of knowledge development, is that with each new generation of researchers many will simply repeat those research endeavours that have been lost from sight, and so not participate in the progressive accumulation of knowledge. However, the current approach to systematic reviews commonly makes little attempt to summarise what is known on the review topic. For example, a systematic review on seclusion and restraint for people with serious mental illness identified 2155 citations during the literature search, retrieved 35 RCTs and then excluded all from the review (Sailas and Fenton 2000).

In developing an approach that can extend the systematic review beyond the synthesis of RCTs, the three types of research reviews proposed by Cooper in 1984 are still relevant today (Cooper 1984). The three types of reviews are:

1. integrative,
2. theoretical, and
3. methodological.

These provide a framework for exploring the contribution reviews can make, beyond evaluating effectiveness, in areas still developing a body of research.

## 1. Integrative Review

The integration of individual studies through meta-analysis is the most well recognised activity of systematic reviews. This integration is the basis for evaluating an intervention. However, often meta-analysis is not feasible or possible, and so findings may also be summarised through a narrative discussion. This integration enables overall conclusions to be made, and the findings to be generalised to other populations with greater confidence. Currently, this is the major activity for most reviews, and they are actively promoted as 'reviews of effectiveness'. However, when effectiveness can not be determined, the value of these reviews is limited in that they only report that there is insufficient evidence to determine effectiveness. While this obviously highlights the need for further research, it offers little in terms of prioritising important issues, identifying what may work or exploring the broader theoretical or methodological aspects of the topic.

## 2. Theoretical Review

A theoretical review addresses the concepts and theories underpinning the review topic. It provides an opportunity to explore the 'taken-for-granted' issues, and to compare and contrast these beliefs, concepts and boundaries. Conflicts across studies may be also identified and explored. While resolution of conflicts may not always be possible, identifying and recording any conflicts are important components of a review. Through this process it may be possible to bridge concepts or even re-conceptualise the topic, and in so doing, refocus future research endeavours. These theoretical reviews are important in areas of developing knowledge because they allow accepted beliefs to be scrutinised and challenged. For example, a recent systematic review of the monitoring of vital signs during hospitalisation found hundreds of papers addressing issues related to the accuracy of temperature measurement, but little on the broader and more important issues such as the purpose of measuring temperature or optimal frequency (Evans et al. 2000). Another review on the nursing



management of chest drains found that this area of practice had been subject to little evaluation, and that there was no scientific basis for any of the recommendations for practice that were identified in the literature (Charnock and Evans 2001). The theoretical component of these reviews provides an opportunity to alert the profession and researchers of areas omitted from past research endeavours.

### 3. Methodological Review

While part of the systematic review process is the critical appraisal of research, this is usually limited to the inclusion or exclusion of studies. Many systematic reviews are completed and published with little information provided on the critical appraisal other than how many were excluded. However, for areas of developing research, methodological quality may at times be poor and the definitions used during the study maybe inappropriate or incomplete. Methodological reviews can therefore provide important information to better inform future research endeavours. For example, a systematic review on falls prevention research found that rigorous methods had rarely been used to evaluate preventative interventions, and it highlighted the need for preventative interventions to be evaluated by RCTs (Evans et al. 1999). This review also found that both of the two identified RCTs were small and so lacked the power to identify any beneficial outcomes, drawing attention to the need for studies to use larger populations when evaluating interventions. A systematic review of the prevention of catheter related urinary tract infections drew attention to the many varied, and at times poor, definitions of what constituted a urinary tract infection that have been used in research (Dunn et al. 2000). This review also highlighted to the need for studies to standardise the definitions used to improve the evaluation of interventions and to facilitate comparisons across different studies.

In summary, reviews play a pivotal role in the development of knowledge. They provide a stocktake of knowledge that imposes some form of order over random,

independent, pieces of information. As part of this summary of knowledge, new interpretations may arise that prompt professional debate and reflection. This process not only helps determine what is known about a topic, it also helps identify what is not known. The argument presented here is that the review provides the foundation of the next generation of advances in knowledge, in that they uncover, re-define and re-focus a profession's pursuit of knowledge on a specific issue.

# **CHAPTER 3**

## **Methods used to Develop An Expanded Review Process**

## **Conceptual Framework**

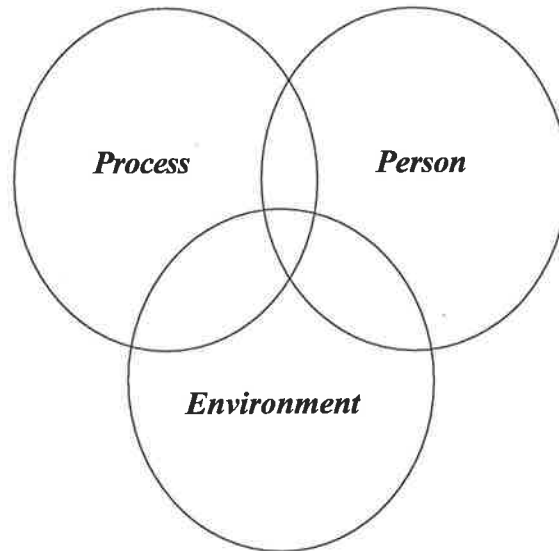
### **Introduction**

In developing a broader approach to summarising evidence within the framework of a systematic review, the concept of evidence must be re-conceptualised. As previously discussed, the common view of the evidence for systematic reviews addresses questions related to whether an intervention works. This focus has dominated these reviews since their inception in the early 1990s and the term 'reviews of effectiveness' emerged. The focus on effectiveness has matched the financial imperative that dominates health care, and provides the basis for investigation of issues related to efficiency. This fits well with policy makers, and as previously discussed, has seen government invest heavily in the evidence-based movement. It can be argued that as effectiveness and efficiency became the dominant paradigm, the importance of other views of health care diminished. However, while evidence on effectiveness is obviously vital, there are other questions that must also be answered before decisions related to funding and implementation can be determined.

### **The Different Perspectives of Evaluations**

A limitation of current approaches to systematic reviews is the major focus on effectiveness. While effectiveness is important, the scope of any evaluation should be broader. For example, it is also important to know whether the intervention is appropriate for its consumer. From this perspective the focus is on acceptability, and is concerned with issues such as whether the intervention would be used by the consumer and whether it also meets their needs. A third dimension relates to the feasibility of implementing an intervention and is concerned with broader organisational and resource issues. These different perspectives broaden the evaluation from the processes of the intervention to one that also encompasses the person and the implementation setting. See figure 1 for diagrammatic representation of this broader view of health care evaluation.

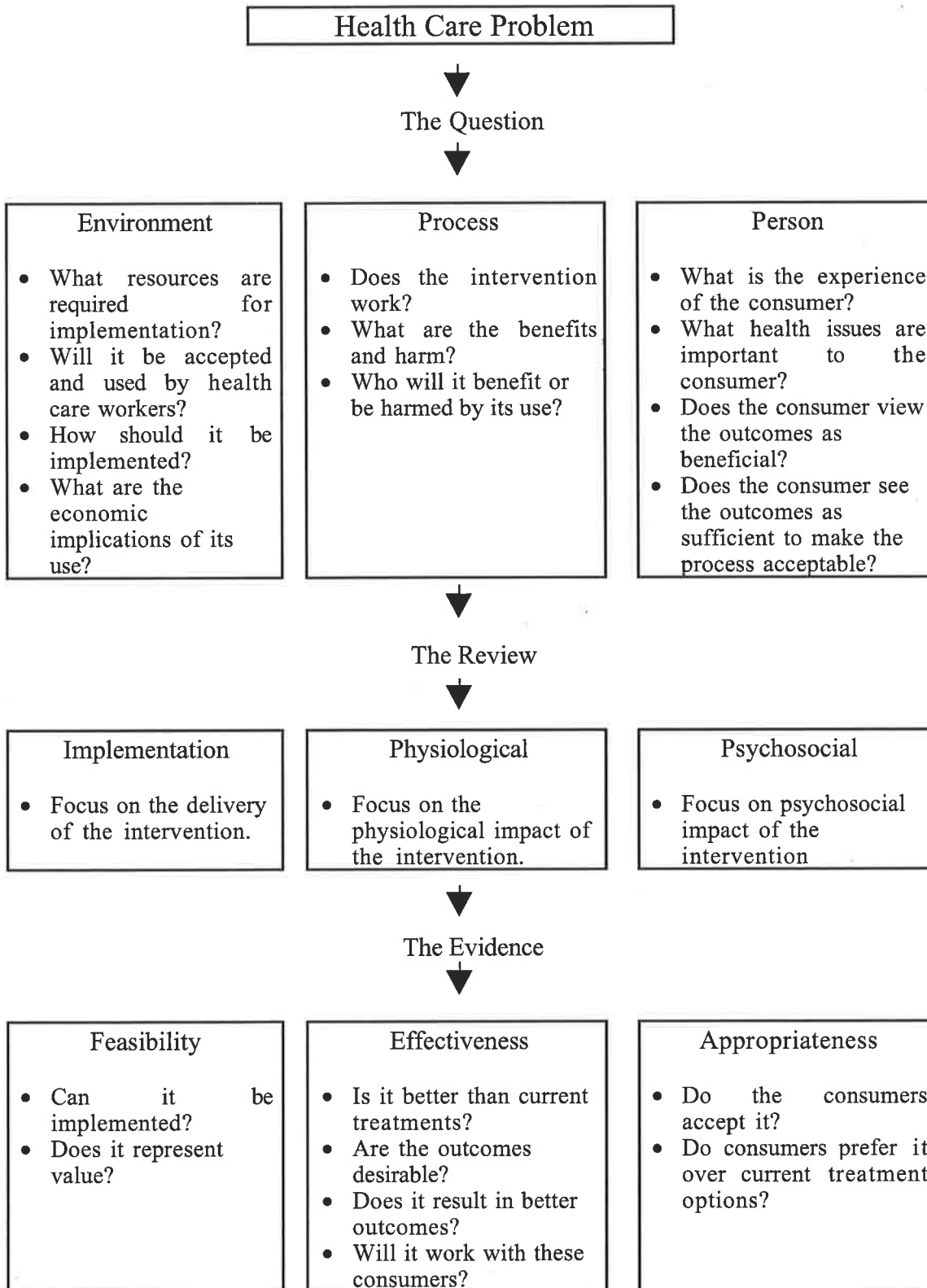
**Figure 1**  
**Components of the Evaluation of Interventions**



Evidence on effectiveness, appropriateness and feasibility provides a sounder base for health care evaluation, in that it acknowledges the many factors that can have an impact on the success of health care delivery. This means that no matter how effective an intervention is, if it can not be adequately implemented, or is unacceptable to the consumer, its value is questionable. This broader conceptualisation of the evidence needed to evaluate health care has formed the basis for the development of the expanded systematic review methodology in this study.

For systematic reviews, this model provides a framework that can be used during the planning phase of the review (see figure 2). The concepts of effectiveness, appropriateness and feasibility highlight questions of importance, and in so doing, gives the review direction. This framework attempts to address the multi-dimensional nature of the evidence and while it does not diminish the importance of evidence of effectiveness, it helps bring other questions of importance into the review process.

**Figure 2**  
Conceptual Framework : Dimensions of the Evidence



## **Multi-dimensional Concept of Evidence**

The concept of intervention evaluation on which the expanded review process was developed focuses on evidence from multiple perspectives. In developing this conceptual framework, effectiveness, appropriateness and feasibility were used to ensure the focus encompassed the processes of the intervention, its recipient and the environment in which it was to be implemented.

### **1. Effectiveness**

Effectiveness has been the most common concern of systematic reviews and relates to whether the intervention achieves the intended outcomes. Effectiveness is concerned with issues such as:

- Does the intervention work?
- What are the benefits and harm?
- Who will benefit from its use?

For the purposes of this conceptual framework, effectiveness is seen to be concerned with the physiological impact of interventions. That is, does the intervention achieve the desired therapeutic change.

### **2. Appropriateness**

Appropriateness addresses the impact of the intervention from the perspective of its consumer. As such it is concerned with the person's experience of the intervention or illness, and also personal preference issues. Appropriateness is reflected in questions such as:

- What is the experience of the consumer?
- What health issues are important to the consumer?
- Does the consumer view the outcomes as beneficial?
- Does the consumer see the outcomes as sufficient to make the process acceptable?

For the purposes of this conceptual framework, appropriateness is concerned with the psychosocial aspects of care, such as, is the intervention acceptable to its intended consumer. This perspective is an important part of the evidence in that enforced health care decisions will likely meet resistance and poor compliance. Some issues go beyond simple effectiveness, and the preferences of individuals may influence final outcomes. The argument underpinning this component of the conceptual framework is that the intervention must be wanted, at least accepted.

### 3. Feasibility

Feasibility addresses the broader environment in which the intervention is situated and involves determining whether the intervention can, and should be, implemented. This focus acknowledges that the process of intentional change in large organisations is complex and involves a great deal more than modification in attitudes and behaviour (MacGuire 1990). In this context, feasibility is reflected in questions such as:

- What resources are required for the intervention to be effectively implemented?
- Will it be accepted and used by health care workers?
- How should it be implemented?
- What are the economic implications of using the intervention?

The feasibility of an intervention must form part of any health care evaluation. This evidence will relate to barriers that inhibit change, the beliefs of the health care workers and work force training issues. The argument underpinning this component of the conceptual framework is that the intervention must be implementable.

### **Justification of Approach**

The benefit of this framework for summarising research evidence is that it moves beyond having a single focus on the RCT. This broader focus is vital as the RCT is



unlikely to be able to answer all the questions needed to fully evaluate the usefulness and value of health care. From this perspective, it is acknowledged that a variety of research methods can be used to contribute valid evidence. The framework also legitimises the perspective of the consumer and the pivotal role their opinions should have in health care decisions.

During the development of this conceptual framework an alternative approach was initially considered, although it was rejected after consideration of its implications. This alternative approach would summarise the evidence according to the research method. That is, there would be systematic reviews of RCTs, and reviews of observational or interpretive research. Indeed, it may even be possible to break interpretive research into more focused reviews of phenomenological studies or of ethnographies. However the concern with utilising a methodologically grounded conceptual framework are twofold. Firstly, this division would reinforce the traditional research boundaries that have plagued health care, and particularly nursing, for many years. Secondly, and more importantly, it relates to whether the question or the research method should lead the review process.

### Methodological Divisions

Research methods have long been an issue of debate within health care, and quantitative and qualitative methodologies have been divided according to discipline. That is, medicine has made extensive use of quantitative methods while nursing has had a major focus on qualitative research. Additionally, epidemiologists have used observation methods to a greater extent than any other health care disciplines. While these divisions still remain to some extent, it can be argued that in recent years there have been some shifts in these divisions as the disciplines concerned take a broader view of research.

This quantitative/qualitative divide has also dominated EBHC and in particular, systematic reviews. Within the evidence-based movement, systematic reviews have

been seen to be summaries of RCTs. While other methods, such as cohort and case control studies, have been incorporated into reviews, they are viewed with caution (Egger et al. 1998). Qualitative studies have also been incorporated into reviews (Jensen 1994; Sherwood 1997), however they have largely remained outside the evidence-based movement.

From this perspective, using the research method as the basis for delineating the domain of the systematic review only serves to entrench these divisions. While these different methodologies obviously cannot be merged in a single meta-analysis, each provides important information that helps inform practice. On this basis, systematic reviews may need to summarise a range of methodologically different research to include all the relevant evidence.

#### Question Versus Method

The second issue relates to the organisation of the systematic review. Currently it is by research method, and as previously stated, systematic reviews are linked to RCTs. There is a certain appeal with this approach in that it simplifies the review processes. Additionally, it helps the reviewer manage topics with a large volume of research because only studies using the chosen method are included in the review.

However, the concern with this approach is that it is the research method that leads the review process and delineates the area of interest rather than the question. The method taking precedence over the question has been a point criticism of primary research. Ray suggests that clinicians and researchers planning to evaluate health outcomes often start with the question 'What instruments can I use?' (Ray 1999). In doing this, Ray suggests the researcher puts the 'methodological cart before the horse', with purpose, conceptual congruence and validity becoming secondary. From this perspective it can be argued that factors such as statistical significance, available instruments and short time lines continue to structure and define much of current research.

The argument of question versus method is also relevant for systematic reviews. That is, systematic reviews most commonly start with RCTs, and because of this, are limited to effectiveness. However, this conceptual framework uses the question as the starting point for the review. Selection of the research method that provides the best evidence follows, rather than precedes, the questions. For some reviews the RCT will provide the best evidence, while for others it will be interpretive or observational studies. Some questions will require a range of research designs to provide complete evidence, and a small number of published reviews have incorporated both quantitative and qualitative research in the same review (Neill 2000; Suikkala and Leino-Kilpi 2001). This is not to suggest that data generated by studies using different methods should be synthesised together. As with meta-analysis, studies are synthesised only when they are similar in all important characteristics. For reviews of RCTs this would mean they have similar populations, interventions and outcomes. In a same manner, studies must also have a similar method. For example to merge data from experimental studies with that of interpretive would be meaningless. Yet investigating an intervention using a range of different methods would provide a rich source of data. These data would not only provide precise estimates of effect, it would also reveal the contextual dimensions of the intervention.

This conceptual framework means that systematic reviews may have a more extensive results section, as different research methods are grouped for successive syntheses. For some topics this would increase the work of the review, or a series of smaller reviews may be required to address each component of the review. However the result of these endeavours would provide a comprehensive evaluation of the intervention. In summary, while there is a certain neatness and aesthetic appeal in developing systematic reviews by research method, their reduced scope means the value of the review is also reduced.

## **Developing an Expanded Review Process**

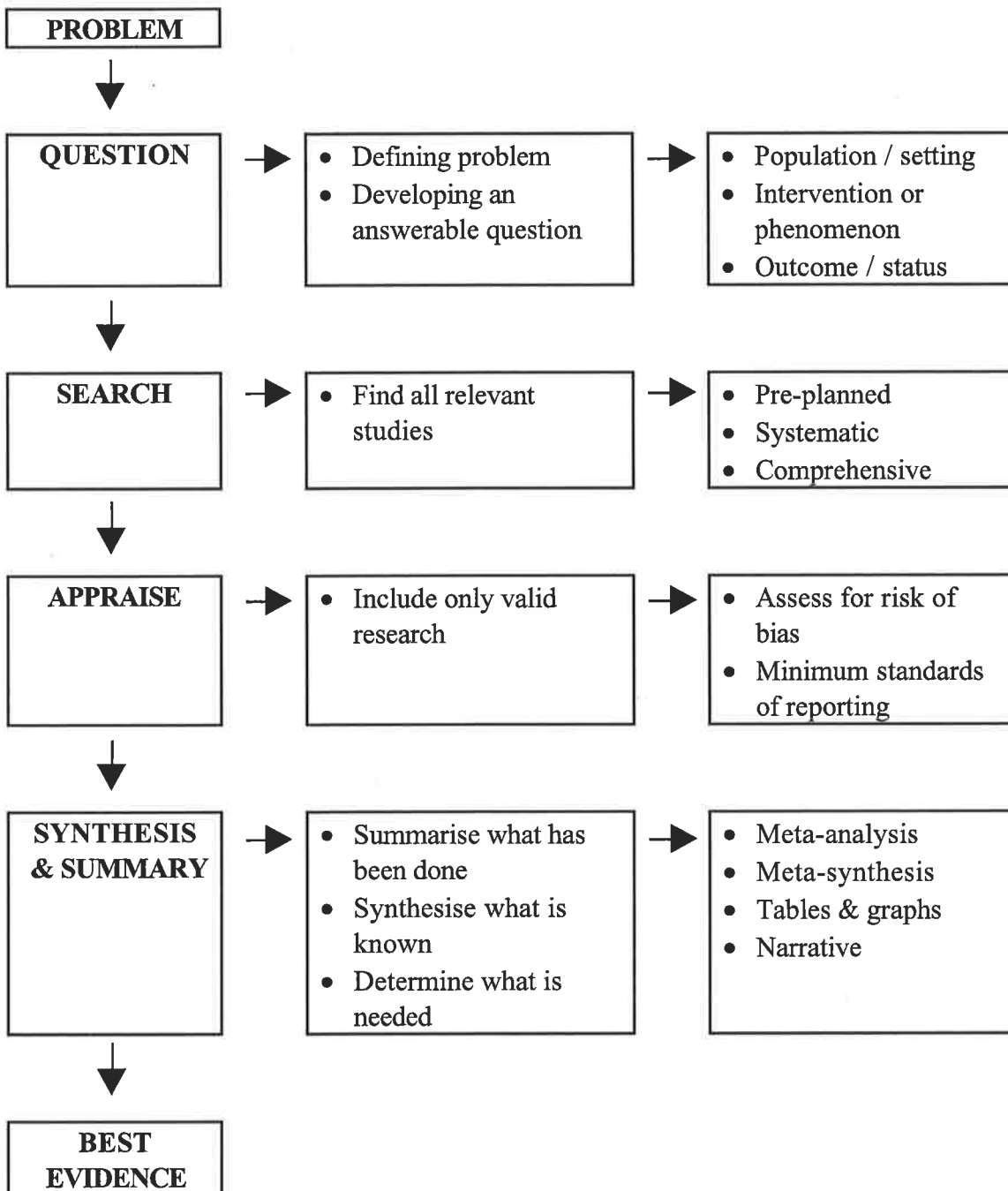
### **Introduction**

Systematic review methods for summarising RCTs have undergone significant development since the early 1990s. These methods follow a clearly documented process and each phase of this process has been subjected to considerable investigation and debate. Therefore, in developing an expanded review process the existing approaches to systematic reviews were used as the template on which to build. This review template was based on the work of the Cochrane Collaboration (Mulrow and Oxman 1997), the National Health Service Centre for Reviews and Dissemination (NHS Centre for Reviews and Dissemination 1996) and the Joanna Briggs Institute (Evans 2000), and was considered to represent the standard approach to undertaking a systematic review (see figure 3).

In expanding this review template, it is acknowledged that there has been growing interest in this area and this interest is reflected in the professional debate in the health care literature. There has also been some development of methods to synthesise the findings of non-RCT research and a small number of these reviews have been published. However, to a large extent, this work has been both isolated and fragmented, with little agreement on optimal methods. These exploratory reviews were used as the basis for broadening the systematic review process.

It should be noted, that while the developmental work reported in this thesis addresses experimental, observational and interpretive research, the interpretive component of this process was the major focus because of nursings' extensive use and investment in this research method.

**Figure 3**  
The Systematic Review Process



### Method

The methods used to develop an expanded review process consisted of three distinct phases; identification of existing approaches to reviewing the research literature; development of the review framework; and evaluation of the review framework.

1. Identification of existing review method.

An extensive search of the literature was undertaken to identify papers addressing some aspect of systematic reviews, with an emphasis placed on papers focusing on non-RCT research. As part of this literature search, examples of existing reviews that summarised studies other than RCTs were also sought. The databases searched to identify this literature were:

- CINAHL
- MEDLINE
- Current Contents
- Expanded Academic Index
- Psyclit

Keywords used during this database search were:

- meta-synthesis
- meta-ethnography
- meta-analysis
- research and synthesis
- (qualitative and review) in title
- (systematic and review) in title
- (integrative and review) in title
- (epidemiological and review) in title
- (observational and review) in title
- (cohort and review) in title
- (case and control and review) in title

A second search was undertaken of CINAHL to identify nursing reviews that summarised interpretive research. The keywords used during this search were:

- (experience and review) in title

- (qualitative and review) in title
- (interpretive and review) in title
- (content and analysis) and (review in title)
- (thematic and analysis) and (review in title)

All identified reviews were examined to determine the methods used during the review process. Specific areas of interest were:

- review question
- inclusion / exclusion criteria
- search strategy
- methods used to critically appraise the identified studies
- methods of data synthesis
- what descriptions were provided of the studies included in the review
- what details of the review methods were provided in the published review report

These papers, when combined with the standard approach to systematic reviews, provided the basis for the development of an expanded review framework.

## 2. Development of an expanded review process

The standard systematic review methods (see table 3) were compared to the discussions and recommendations in the literature, and to the methods reported in published reviews to identify:

- similarities in methods used to review non-RCT studies compared to those used in systematic reviews of RCTs.
- the different methodological approaches used to review non-RCT research.

The aim of this process was to determine what areas existed where there was consensus on optimal approaches to reviewing non-RCT research. When no

consensus was present, methods were developed based on the information derived from the identified reviews and discussion papers.

### 3. Evaluation of expanded review process

The third phase of this project was the evaluation of the expanded review process. This was achieved through the conduct of a review utilising the proposed methods. To facilitate the evaluation, the review was separated into the three components outlined in the conceptual framework, effectiveness, appropriateness and feasibility. The evaluation of this framework was in terms of:

- protocol development
- the review process
- review results
- practicality and usefulness

The clinical topic selected for this evaluation phase of the study was the use of music for hospital patients. This topic was chosen because the initial search of the literature suggested that the impact of this intervention had been subjected to extensive investigation using a variety of research methods.

To ensure an adequate evaluation of the expanded review process, selected results from a second review were presented. This review addressed the use of physical restraint in the acute and residential care settings. These selected results were used to further support the methods proposed in this thesis. These findings were used to demonstrate how a range of methodologically different studies can be used to determine the effectiveness, appropriateness and feasibility of a broad clinical issue. Additionally, these results also provided a practical example of how the findings from experimental, observational, interpretive and descriptive research were pooled.



# **CHAPTER 4**

## **The Expanded Review Process**

# The Expanded Review Process

## Introduction

The purpose of this chapter is to present an in-depth discussion of the different methods used to review the research literature. The approach taken in this chapter is to present a discussion of the standard approach to systematic reviews, then the approaches for reviewing non-RCT research. As previously discussed, these expanded methods will be drawn from the literature and existing reviews. At the end of each section of this chapter the methods to be used as part of the expanded review process are summarised. The areas to be addressed are:

- review protocols,
- review questions,
- selection of studies,
- search strategies,
- critical appraisal,
- hierarchies of evidence,
- data collection, and
- data synthesis.

The product of this chapter will be the expanded review process. While this process will differ in some aspects from approaches used to summarise RCTs, it will represent the current optimal method for broadening the scope of these reviews.

## **Review Protocols**

### **Introduction**

Systematic reviews are scientific investigations that aim to review and summarise past research efforts and utilise the same principles and rigour which underpins all research endeavours. However, as reviews are retrospective observational research they are therefore at risk of systematic and random error like any other type of research endeavour (Cook et al. 1998). Because of this, their quality depends on the extent to which scientific methods are employed (Cook et al. 1998). Like all research, pre-planned methods are developed that detail each step of the review process and so help minimise these risks.

### **The Role of the Protocol**

To help achieve the rigour expected of the systematic review, a review protocol is developed that specifies the question and methods prior to commencement, and as a result, the review is less likely to be biased (NHS Centre for Reviews and Dissemination 1996; Evans 2001). Without this predetermined plan there is a risk that the review could be driven by anticipated findings and subjective decisions. The components of the protocol are:

- questions or hypotheses of interest,
- method of selecting studies for inclusion in the review,
- strategy for identifying studies,
- strategy for critically appraising studies,
- method for collecting the data from individual studies, and
- method for summarising, analysing and synthesising the results from individual studies (NHS Centre for Reviews and Dissemination 1996; Mulrow and Oxman 1997).

For reviews with multiple objectives the protocol development will be more complex (NHS Centre for Reviews and Dissemination 1996). This is because the review may focus on a range of specific questions, require different search strategies and involve multiple analyses. During the conduct of the review, the aim is to follow each step in the protocol. However like primary research, sometimes modification of the protocol may be needed to adapt to unanticipated circumstances (Mulrow and Oxman 1997). These changes may be the result of factors such as a search that identifies too many or too few studies, or when new questions arise during the review (NHS Centre for Reviews and Dissemination 1996). If changes must be made, then when the results of the review are reported it will be important to distinguish the hypotheses that were posed *a priori* from *post hoc* hypotheses which were generated during the review (NHS Centre for Reviews and Dissemination 1996).

In terms of the development of a protocol for non-RCT research, there is little available information. During the literature search all review protocols that were identified addressed RCTs rather than non-RCT research. However, the principles underpinning these protocols, like those for primary research, are not restricted to any single approach. The aim of the protocol is to document the intended review process prior to commencement, and in so doing, limit the need for subjective decisions during the conduct of the review. From this perspective, the type of research to be reviewed has little impact on the protocol. That is, while the question or method of data collection and synthesis may differ, the principle of documenting each step of the process remains unchanged.

### **Expanded Review Process: Protocol**

For the expanded review framework the standard approach to the development of a review protocol was used. This protocol outlined the objectives and questions of the review, study selection, search strategy, critical appraisal, data collection and data synthesis. It is acknowledged that the specific components of this protocol will differ from those used for RCTs, however, as previously stated, the principles of a pre-

planned approach and rigorous method need not be influenced by changes in the review focus or in the type of research to be summarised.

## Review Questions

### Introduction

A well formulated question gives the review direction, and like primary research, review questions are generated from such areas as clinical encounters with the patient, variations in patient management, or as a result of the introduction of new treatments (Counsell 1997). The suggested format for a systematic review question involves a number of specific components (Critical Reviews Advisory Group 1996; NHS Centre for Reviews and Dissemination 1996; Counsell 1997; Mulrow and Oxman 1997):

- population and the health problem to be investigated,
- intervention, treatment or exposure of interest,
- comparator (control) under scrutiny,
- outcomes of concern, and
- research method that best answers the questions.

The population component of the question defines the specific participants, setting, condition or disease of interest. For example, cancer patients with chemotherapy or radiotherapy induced oral mucositis (Kowanko et al. 1998). Any restriction with respect to specific populations or settings, should be based on sound reasoning (Mulrow and Oxman 1997). For example, a systematic review addressing falls in hospitals excluded children on the basis that falls involving children during hospitalisation were likely to result in fewer adverse outcomes, and their prevention would involve different interventions than those used for adults (Evans et al. 1999).

The intervention or treatment, and the comparison, must also be clearly stated. Interventions should be defined so that the scope of the review is unambiguous. The comparison in the question enables the effect of the intervention to be measured along side current treatments. A critical component of the question is the outcome that will

be used to determine the effectiveness of the intervention and for many reviews, this will involve multiple outcome measures (Mulrow and Oxman 1997). One major constraint in the development of the review question is that primary research must exist on the topic, and while it is possible to synthesise the results of two studies, the value of such a review would be limited (Cooper 1994).

The scope of the question, and therefore the review, can be narrow or relatively broad covering a whole health care topic (Counsell 1997; Jones and Evans 2000). With a broad question, the research summary would likely consist of a narrative discussion rather than a statistical analysis. The use of these narratives has been the traditional approach to summarising research and has been subject to criticism because of their lack of rigour. However, narrative reviews are appropriate when a broad perspective rather than an in-depth look at a particular issue is required (Cook et al. 1997). This broad approach is necessary for issues such as the history or development of a problem, cutting edge developments if research is scant, and if studies are limited by flawed design or execution (Cook et al. 1997). These narrative reviews are less useful in providing quantitative answers to specific clinical questions. Counsell also warns that broad questions increase the likelihood of finding heterogeneity, or significant variation in results (Counsell 1997).

For systematic reviews addressing issues of interest to nurses, the nature of the existing body of research will have a significant influence on the review. Only a small number of nursing topics have been extensively researched, and for many of these topics, the RCT design has not been commonly used. This means that many systematic reviews of nursing issues will be broad representing a stocktake of knowledge and the identification of future research needs rather than determining the effect size of an intervention.

## **Review Questions for Interpretive Research**

While there has been little attempt to explore the formulation of review questions addressing issues other than effectiveness, it is likely that the principles guiding primary research will also be appropriate for systematic reviews. That is, regardless of focus and type of research to be summarised, it is still necessary to define the area of interest to ensure the boundaries of the review have been clearly determined and documented prior to commencement.

Reviews of interpretive research are also used to seek answers to specific questions. These questions must be broad enough to be of interest but small enough to be manageable (Sherwood 1999). Delineating a clear focus is particularly important with interpretive reviews because the data to be summarised in narrative form which does not lend itself to concise summary and synthesis in the same manner as numerical data. Too broad a review question will significantly increase the time and work required to complete the review. Too broad a question may also result in the review attempting to summarise studies that are fundamentally different in one of their major components, such as population, area of interest or method (Evans and Pearson 2001a), however there is little guidance on how best to strike this balance.

In exploring the type and structure of questions posed in published reviews of interpretive research, most are broad in nature. Many reviews give a purpose rather than a question. Some of these questions and purposes are so broad that it is argued that they fail to give the review a clear direction and do not delineate the specific area of interest. For example:

“To describe and develop a deeper understanding of each concept by means of research synthesis of qualitative caring in nursing studies.” (Fredriksson 1999, p. 1168), and

“...to inductively develop a theory of wellness-illness through the synthesis of qualitative research” (Jensen 1994, p. 350).



In some reviews no question or purpose is stated, and the reader must attempt to determine the specific focus. For example, the following quotes are as close as two reviews came to stating a question or purpose for the review of interpretive research:

“This paper, however, confines itself to comment and analysis regarding these studies that concerned cardiac ailments more generally and without presupposition about the aspect of experience to be the main focus of study” (Clark et al. 1998), and

“...this paper takes a look at the relatively little qualitative research that has specifically focused on self-care activities employed by people as a means of managing illness.” (Chapple and Rogers 1999).

However, in other reviews the purpose is clearly stated leaving little ambiguity. For example:

“...to discover whether women with benign breast disorders suffer similar amounts of anxiety and psychological distress to women with breast cancer in the time between the discovery of the problem and receiving a diagnosis...” (Woodward and Webb 2001, p. 30), and

“...to examine how qualitative research has helped to advance an understanding of childhood cancer from the perspective of children who have cancer.” (Woodgate 2000).

One review summarised a range of different types of research, including interpretive, and clearly defined the area of interest:

“How do parents identify acute illness in their child?  
How do parents respond to acute illness in their child?  
What are parents’ experiences of the health services when their child is acutely ill at home?” (Neill 2000, p. 822).

These examples suggest that both broad and narrowly focused questions may be used during reviews of qualitative research. It also appears that some reviews of interpretive research were conducted without the guidance of a well formed question or clearly stated purpose. This limits the value of this type of review for two reasons. Firstly, without the clearly defined focus the review has no distinct boundaries and is therefore at risk of becoming a rambling narrative. Secondly, the lack of a clear question or purpose increases the difficulty of determining the worth of the review, that is, whether it fulfilled its purpose and whether only relevant studies were included.

Broad questions will also increase the complexity, time and expense of the review because of the larger volume of research that must be summarised. Additionally, as the question is the basis for the development of the review method, ambiguous questions may also result in ambiguous protocols. For example, selection of studies for inclusion is more difficult when the boundaries have not been adequately defined. Synthesis of data will also be more difficult when there is no clear description of what is expected from the review. It is likely that the issues surrounding the development of a question will be as important for the reviewer as they are for the primary researcher. On the basis of this discussion, questions for reviews of interpretive studies must provide sufficient detail to define the area of interest and the expected product of the review. Whether this is expressed as a question or a purpose will likely be of little importance.

### **Expanded Review Process: Review Question**

For the expanded review framework, the questions used for reviews of RCTs or observational studies are similar, and address the population, the intervention and the outcomes of interest. This approach was used for the expanded review process, and represents the current methods used for systematic reviews.

To address broader issues, such as experience, opinions and preferences, a clearly stated question was considered the optimal approach to guide the review process. As the question must delineate the area of interest of the review, it must outline the population, setting, circumstances or phenomenon and the outcomes or focus of interest. For the expanded review process all questions, regardless of whether they address issues answered by experimental, observational or interpretive research were viewed as not differing substantially from these posed by primary researchers.

## **Selection of Studies**

### **Introduction**

The study population for systematic reviews is all past primary research on the topic of interest and so like any research, selection of the studies to be included in the review must be free from bias. To achieve this, the selection process should be based on clear scientific reasoning (Petitti 1994). As with other components of the review, these criteria must be developed and documented prior to commencement, to protect the review from allegations of investigator bias (Petitti 1994; Jones and Evans 2000). Investigator bias is when the investigator consciously or unconsciously selects studies for inclusion in a review based on the findings of the individual studies. Documenting the selection criteria also makes the review processes transparent, so that others using the criteria should reach the same conclusions. Use of criteria also allows others to evaluate the comprehensiveness of the review and provides a record of the judgements made about each study (Meade and Richardson 1998)

The process of selecting studies for inclusion in a systematic review has two distinct phases. The first assessment of eligibility occurs during the search for studies, and involves comparing the title and abstracts that are retrieved from electronic databases to the criteria. At this level, the assessment is to determine if the study appears to meet the inclusion criteria and so whether it is worthy of retrieving the full study report. The second level of assessment is based on the information provided in the actual study report. At this level, assessment of eligibility is undertaken with the care and precision of any research activity. During the conduct of some reviews this assessment is undertaken with all identifying information removed from the study report so factors such as high profile researchers or journals will not influence the selection process. However, the gain in validity has to be measured against the considerable increase in time and effort to disguise any identifying information (NHS Centre for Reviews and Dissemination 1996).

For experimental and observational research, the selection criteria are derived from the question and so address the intervention, population, outcomes of interest and the comparison of interest. To these four components is added the study method that will provide the most valid evidence. Like primary research, exclusion criteria may also be needed to clearly document when studies will not be included in a review. At times it may be necessary to make revisions to the criteria if the search produces a large number of references, and so may necessitate redefining the description of such things as the population, intervention, outcomes or time-frame (Critical Reviews Advisory Group 1996).

When only a small number of references is identified, it may indicate a need to revise the characteristics listed above, or that the effectiveness of the literature search was poor. However, these types of revisions bring with them a risk of bias, and so should not be made based on the actual results gained from identified studies.

### **Inclusion Criteria**

The inclusion criteria transform the review question into its components by defining the population, intervention, outcomes of interest, and the study design that provides the best evidence (Mulrow and Oxman 1997; Evans 2001). In developing these criteria, the boundaries and limits of the review are clearly defined.

#### **1. Population**

The inclusion criteria must describe the population of interest such as the elderly, adults or children. The description of the population may also be in terms of the presence of a disease (Meade and Richardson 1998). Like other components of the criteria, it does not differ significantly from those used by the primary researcher.

## 2. Intervention

The intervention must be clearly described and the comparison treatment stated. The description will vary between interventions, as it may be necessary to address issues such as dose, frequency, duration or special techniques. For reviews with a narrow focus, the intervention will be described in the same detail that is used by primary researchers. For reviews with a wider focus, this description will reflect the broader nature of the question. The aim is to strike a balance between making it too specific in that it excludes most studies, and making it too broad to be useful (Counsell 1997).

## 3. Outcome of Interest

Some form of outcome must be stated when investigating the effectiveness of an intervention. The outcomes may include adverse effects, to allow investigation of both the benefits and harm. These outcomes must be defined and then translated into criteria (Meade and Richardson 1998). The effectiveness of some interventions is determined through multiple outcome measurements. Selection of outcomes can be difficult for some topics. It has been suggested that surrogate outcome measures should be avoided when possible (Counsell 1997). Surrogate outcomes are measures that indirectly evaluate the effectiveness of an intervention. For example, blood pressure may be used to measure the effectiveness of antihypertensive medications, when the major outcomes of interest are long term morbidity and mortality. The use of these indirect measures can result in errors in the evaluation, for example one antihypertensive medication was shown to lower the blood pressure but also cause an increase in deaths (Fleming and DeMets 1996). Surrogate outcomes are used because many outcomes are long term, and therefore would be very costly for researchers to measure, or because some direct outcome measures are expensive or difficult to measure reliably.

## 4. Study Design

Reviews should aim to summarise the best available evidence, and so the criteria should define which research methods provide the most valid results (Meade and

Richardson 1998). Decisions related to which studies are to be included in the review should reflect the reliability and risk of bias of the different methods (NHS Centre for Reviews and Dissemination 1996). However, the study designs that are selected will ultimately be determined by the review questions. For some topics the optimal research design may not have been used to investigate the intervention, and so it may be necessary to consider what information is available and to modify the eligibility criteria accordingly (Meade and Richardson 1998). For topics with little research and where the aim of the review is a systematic stocktake of what is known, many different study methods may be considered. Petitti suggests that for reviews of non-experimental studies, it may be more appropriate to consider all research designs (Petitti 1994). However, when best evidence has been generated by methods at risk of bias, the findings provide a poor basis for clinical practice.

### **Exclusion Criteria**

In addition to inclusion criteria, exclusion criteria may also be needed. This is to ensure the studies included in the review are homogenous in terms of their major characteristics. The exclusion criteria help to achieve this homogeneity. The following are examples of possible exclusion criteria.

#### **Publication Date**

Some reviews limit the inclusion of studies according to year of publication. A common cut off date of many reviews is 1966, which is the date when the MEDLINE database began (Petitti 1994). While this adds to the ease of identifying studies, it may not be an adequate justification as potentially relevant papers may be missed.

However for some topics, it may be appropriate to exclude studies based on year of publication. For example, during a systematic review of the prevention of urinary catheter related urinary tract infections, studies prior to 1980 were excluded because of the many improvement in urinary catheter equipment that occurred after this date

(Dunn et al. 2000). Inclusion of studies that evaluated equipment such as urinary catheters with open drainage would provide little useful information for clinical practice where such practices no longer occur.

### Multiple Publications

Some studies are used to generate multiple publications based on the same data. Repeat reports based on the one population are not independent, and so including more than one estimate from the same study population violates the statistical assumptions that underlie the procedures for aggregating data (Petitti 1994). Information from the same study should only contribute once to the summary estimate of effect and failure to exclude this type of information may bias the results of the systematic review. However, it has been noted that some authors are good at disguising duplicate publications and it can be difficult to identify serial publication reporting accumulating numbers of participants or increasing length of follow up (NHS Centre for Reviews and Dissemination 1996).

### Language

Some reviews exclude studies on the basis of language, commonly including only English language publications. This is often due to the difficulties involved in accurate translation of non-English studies in terms of accuracy, cost and time to complete the review. This difficulty is increased by issues such as which studies should be translated, because until the abstract is translated it is not clear whether the paper meets the inclusion criteria. Language restrictions have been reported to occur in 78% of systematic reviews which results in the exclusion of some RCTs (Gregoire et al. 1995). A comparison of English language and non-English RCT found no significant difference in completeness of reporting between the two types of reports, questioning the basis for the exclusion (Moher et al. 1996). It has also been suggested that because health care differs across countries, the inclusion of research from around the world can offer important insights (NHS Centre for Reviews and Dissemination 1996). However, incorporating the findings of studies that were conducted in vastly different



health care systems may also threaten the usefulness of the findings. Despite this, the inclusion of all research, regardless of language, significantly increases the cost, complexity and duration of the review. Therefore if the review includes only English language publications, this should be clearly stated and acknowledged as a potential limitation (Petitti 1994).

### Study Size

Some reviews exclude studies on the basis of study size, excluding small studies (Petitti 1994). This is done to avoid the problem of small studies carrying too much weight in the meta-analysis and thereby overestimating the effect size. However contradicting this, systematic review and meta-analysis are commonly seen as a way by which the results of many small studies can be combined and so overcome the problem of studies having insufficient power to detect beneficial effects.

### Intervention

One of the most important issues regarding the eligibility of studies is the similarity of treatments or interventions evaluated (Petitti 1994). Even when the study evaluates the same treatment, differences in implementation may prevent synthesis of study results. For example, a systematic review evaluating the effectiveness of chlorhexidine for oral mucositis found many differences in the administration of chlorhexidine between studies, such as:

- strength of chlorhexidine solutions,
- frequency of administration,
- amount of solution administered,
- method of delivering the chlorhexidine solution, and
- concurrent treatments in addition to chlorhexidine (Kowanko et al. 1998).

As a result of these differences, evaluating the effectiveness of chlorhexidine was difficult and determining the optimal method of administering chlorhexidine was not possible.

## Outcome Measures

Differences in the outcomes used to measure the effect of a treatment can also influence decisions about the eligibility of studies for inclusion in a review (Petitti 1994). For example, the mucositis review mentioned above found a vast range of different outcomes had been used to evaluate treatments, including:

- severity of mucositis,
- number of ulcers,
- condition of the patients gums,
- dental plaque,
- ability to eat or drink,
- a range of different mucositis assessment scales,
- assessment by patient,
- pain, and
- infection (Kowanko et al. 1998).

This means that while a number of studies can evaluate the same treatment, the use of different outcome measurements may result in little similarity between findings. Additionally, while a range of potential outcomes may exist, the review may have a focus on only one or two of these outcomes. On this basis, some studies evaluating the intervention of interest may be excluded from the review because of the outcome measures used.

## Quality of Methods

As part of the systematic review process, all studies are subject to critical appraisal. The rationale is that only quality studies should be included in the review.

## Completeness of Reporting

A final area that influences the eligibility of studies for inclusion in a review is the completeness of information (Petitti 1994). Some reports provide incomplete results.

While this is commonly encountered in unpublished studies, conference abstracts, brief reports and studies reported in letters to the editor, it may also occur in published research reports. Lack of information in key areas such as the study population, intervention, outcomes, methods or results will mean that it is difficult to evaluate the study and may be impossible to include the findings in a meta-analysis.

### **Selection of Interpretive Studies**

While the selection of experimental and observational studies is achieved through similar processes, selection of interpretive studies differs. Some of the common components of the inclusion criteria are inappropriate for this type of research. For example, there is no comparison group or outcome measures for the evaluation of an intervention. While an intervention may be the focus of some interpretive studies, it is not evaluated in terms of its effectiveness. Therefore the development of inclusion criteria for interpretive studies differs from that used for experimental and observational studies.

Important issues related to the selection of interpretive studies suggested by Estabrooks *et al* include;

- all studies should focus on similar populations or themes,
- studies should have a similar research approach, as mixing methods may lead to difficulty in developing theory because of the differences in the epistemological foundations of different qualitative research methods, and
- the reporting of themes and labels must be clear and must be grounded in the data to be able to demonstrate comparability of themes and labels across studies (Estabrooks et al. 1994).

Evaluation of published reviews of interpretive research highlights the range of different approaches. The inclusion criteria used for the review of interpretive

research by Thorne and Paterson addressing chronic illness were (Thorne and Paterson 1998, p. 174).

- Focus           Qualitative interpretive research in which the subjective experience of living with chronic illness from the perspective of the individual with the disease was investigated.
- Participants    Individuals or aggregates with a diagnosis of a chronic illness.
- Time-frame     Studies published between January 1980 and July 1996.
- Language       English language journal, book, dissertation or thesis.
- Method         Studies that provided sufficient evidence of data trail to demonstrate how data were analysed.
- Outcome        Studies that provided demographic data about participants.

In addition to these comprehensive criteria, the selection of studies for inclusion in this review was made independently by three members of the review team. However, few other existing qualitative reviews appear to have used such a comprehensive approach to the selection process. A qualitative review by Jensen and Allen utilised far more simple criteria, which was:

*"...studies reporting use of a qualitative approach concerning health, disease, wellness, and illness among adults"* (Jensen 1994, p. 350).

A review by Sherwood, simply stated

*"...qualitative nursing research literature investigating caring from the perspective of clients..."* (Sherwood 1997, p. 33).

The suggestions by Estabrooks et al. (Estabrooks et al. 1994) and the criteria used by Thorne and Paterson (Thorne and Paterson 1998, p. 174) highlight the similarities with inclusion criteria used for RCTs. It can be argued that regardless of method, the principles of inclusion criteria remain the same. That is, the selection criteria define

the components of the question. As such they remove any ambiguities from the selection process making this phase of the review transparent.

### **Selection of Descriptive Research**

There is little information related to the selection of descriptive research for inclusion in a systematic review. Additionally, no examples of systematic reviews that incorporated descriptive research were identified during the literature search. This highlights the previous discussion that these reviews have been limited to RCTs. However, as with reviews summarising other types of research, questions posed by the researcher can also be posed by the literature reviewer. These questions would define the area of interest of the review in terms of setting, population and description of interest.

### **Expanded Review Process: Selection of Studies**

In terms of the expanded review framework, the following components of inclusion criteria for experimental and observational research will be addressed:

- Population - specific population of interest,
- Intervention - the intervention of interest,
- Outcome - outcome measures to be used to evaluate the intervention, and
- Method - the optimal research design that provides the best evidence.

These criteria will allow the review question to be operationalised, through the process of explicitly defining each area of interest. The selection criteria for interpretive and descriptive research will address:

- Population - specific population of interest,
- Focus - phenomenon or situation of interest,
- Outcome - sufficient description of themes and labels, or description of the issue of interest, and
- Method - sufficient description of method used to gather data.

The population component of the criteria, like that used for the selection of RCTs, defines the specific population of interest. The focus describes the circumstance, phenomenon or situation of interest, for example living with chronic illness. This component of the criteria is similar to the intervention of the RCT selection criteria, in that it documents the context of the studies included in the review. The outcome is concerned with supporting data to enable the validation of themes and labels across studies. Finally, as is expected of RCTs, interpretive and descriptive studies must provide sufficient information to determine how data were collected and analysed.

## **Search Strategy**

### **Introduction**

It has been suggested that a comprehensive and unbiased search is one of the major differences between traditional literature reviews and a systematic review (Mulrow and Oxman 1997). The purpose of this search is to attempt to identify all relevant studies on the review topic.

Scientific articles can be seen as clustering into independent sets of literature, with each set addressing common problems and advancing common arguments (Swanson 1990). These sets of articles interact by citing one another. From this perspective, the function of the reviewer is to uncover all the articles belonging to a set. However, discovering these articles is difficult because the full complement of articles is never known, and the articles belonging to a set change with each topic. Discovering all these articles is a time consuming task and it has been suggested that this is one reason why there have been so few critical reviews of the research (White 1994). Additionally, identifying these papers is becoming increasingly difficult as the volume of health care literature continues to increase. It is now likely that many published papers will quickly be lost to the profession as they are hidden in the vast volume of health care literature. It is also likely, that because these studies are lost, their impact on practice is unlikely to extend beyond the local area of the researchers who completed the study. Therefore, just the identification of all research on a substantive area is, in itself, an important contribution to health care.

### **Search Strategies**

To increase the likelihood that all relevant papers are identified, a search strategy is developed and documented in the protocol. This plan outlines the intended approach and documents all sources to be searched. This pre-planning of the search allows

critique and amendment of the process prior to commencement. While there has been a variety of approaches that can be taken, whichever approach is chosen must be reported in sufficient detail to allow others to appraise its quality and comprehensiveness and to allow replication of the search (Counsell 1997; Evans 2001).

For reviews on topics of interest to nurses, Lynn suggests that it is more difficult to locate nursing research because of factors such as the limited referencing and abstracting services, fewer research publications, and nursing's hesitancy to publish (Lynn 1989). When these are combined with publication bias, she suggests that it may be inaccurate to assume that acquired studies are truly representative of all research in that substantive area.

The aim of the search strategy is to minimise the number of relevant references that are missed, while also minimising the retrieval of non-relevant document (Critical Reviews Advisory Group 1996). However, while a rigorous procedure reduces the risk of error, it also increases the time and resources involved in preparing a review (Mulrow and Oxman 1997). Jadad and McQuay suggest that because of time and cost constraints, many reviewers are more realistic and attempt to identify the maximum possible number of relevant studies (Jadad and McQuay 1993). Locating these studies is mainly achieved through searches of electronic databases and bibliographies, but may also include other mediums such as conference proceedings, government reports and the internet. It is also common to hand search key journals related to the review topic to identify missed studies (Jones and Evans 2000). While superficially these activities appear simple, the apparent ease of the search belies the complexities that take place during the search process (Booth 1996). It has been suggested that there is a risk that the reviewer will find enough citations to be satisfied but be unaware of the many studies they may have missed (Booth 1996). The reason these papers are missed is most often the result of a flawed search strategy (Greenhalgh 1997).



## Databases

The most important way in which studies are identified is through searches of electronic databases. There are hundreds of different databases that document and record the health care literature. While there is considerable overlap of coverage between these databases, each database has its own specific focus. Some databases are very general, in that they cover mainstream journals, for example MEDLINE and CINAHL. Other databases are more focused, for example Current Contents focuses on the most recently published papers, and EMBASE covers the European literature. Some have a very narrow focus such as a specific areas of health care (Cancerlit), research methodology (Database of Abstracts of Reviews of Effectiveness) or setting (Rural). Databases are now available covering conferences, higher degree research, government reports, law and newspapers. During systematic reviews no attempt is made to search every available database, rather the aim is to identify those that are likely to yield relevant studies.

In developing the search strategy, the reviewer must decide what databases are to be searched. While MEDLINE is often seen as the most comprehensive of all available health care databases, comparisons of searches of MEDLINE and CINAHL have produced contradictory results. One study comparing the usefulness of these databases to nursing post-registration students found that MEDLINE should be regarded as the first choice, except when topics relate specifically to the organisation of nursing (Brazier and Begley 1996). Another study concluded that both databases covered unique journal titles and yielded unique relevant results (Watson and Perrin 1994). One evaluation of the sensitivity and precision of MEDLINE for identifying RCT, found that even when the search was conducted by a trained searcher, they yielded only 51% of RCTs with a range of 17 to 82% (Dickersin 1994). The authors conclude that a systematic review relying only on MEDLINE to identify RCTs would miss approximately half of the available studies.

Given the fact that the results of these comparisons are contradictory it is likely that search strategies should encompass a range of databases, and for nursing related topics, this would be of particular importance. The reviewer is also faced with what has been termed diminished marginal returns (Mulrow and Oxman 1997). This means that after each phase of the search process is complete, additional searches return fewer studies that are relevant to the review. It has also been suggested that a point is reached where any effort to increase the number of references identified during database searching will likely be at the expense of precision and that a large number of irrelevant references will need to be reviewed to identify any additional references (Dickersin 1994). So while many hundreds of databases exist, the search strategy will only cover those likely to identify worthwhile studies. The general databases likely to be included in most reviews are MEDLINE, CINAHL, Psyclit, Current Contents and the Cochrane Library.

#### 1. MEDLINE

MEDLINE was created, and is maintained, by the National Library of Medicine (NLM) in the United States. It lists references back to 1966 (Lowe and Barnett 1994) and includes approximately one third of all medical articles (Greenhalgh 1997). The focus of MEDLINE is predominantly medical, although it does include some non-medical publications. It covers the health care literature of western countries, primarily English language publications, and so has only a limited cover of European publications.

#### 2. CINAHL

The Cumulative Index of Nursing and Allied Health Literature (CINAHL) is produced by the Cinahl Information Systems company in the USA, and covers nursing and allied health publications. It also includes some higher degree theses, however the listing is incomplete and it focuses predominantly on those produced in the USA. While CINAHL covers some medical journals, this coverage is limited. Traditionally,

the CINAHL database has been seen as the first choice for nurses searching the literature.

### 3. Psyclit

Psyclit is a small database covering the literature related to psychology, and psychological aspects of other disciplines such as nursing, medicine and sociology. The database is produced by the American Psychological Association.

### 4. Current Contents

The Current Contents database is produced by the Institute of Scientific Information in the USA and covers a number of fields, including health care. While Current Contents is another useful database to search for studies, its most important role is the identification of recently published research. This is possible because it is updated on a weekly basis, rather than just several times per year like some databases.

### 5. Cochrane Library

A number of specialised databases are found on the Cochrane Library CD ROM, and include the Cochrane Controlled Trials Register (CCTR) and the York Database of Abstracts of Reviews of Effectiveness (DARE) (The Cochrane Collaboration The Cochrane Library (database on disk and CDROM)). This database provides a comprehensive listing of RCTs and systematic reviews.

### Search Filters

One of the main challenges in searching databases for research is not to retrieve all citations, but rather to filter only the most relevant papers (Booth 1996). To aid this process, search filters have been developed that identify studies that utilised a specific research method, such as the RCT. These methodological filters are predetermined strategies that, when combined with subject terms, aim to identify high quality articles (Booth 1996). To identify all studies in a database that may be RCTs, these filters

would use terms such random, controlled, trial, placebo or blinded. When all RCTs have been identified, subject terms such as falls, music or catheter are then used. Evaluation of this approach suggests it improves the results of database searches (Jadad and McQuay 1993).

However, in using these filters, many papers addressing the review topic are not identified by search because of their lack of any specific methodological terms. This means that some relevant papers may be missed. Importantly, in choosing to use these filters, the assumption is that there is a large volume of published papers making it impossible to check every citation manually. For topics with only small numbers of publications, manually checking each citation identified during the database search would still be the most reliable method of ensuring relevant studies are not missed. This applies to many areas of nursing practice where the body of literature is not large enough to prevent the check of every citation on the topic.

#### Developing a Search Strategy

Developing a database search strategy involves the identification of appropriate terms related to the topic and combining these with methodological search filters. The most common approach is the use of repeated searches on a restricted area of the database, such as within a single year. With each successive search there is greater refinement of the terms used and better yield of retrieved references. Dickersin *et al.* recommend a two stage strategy (Dickersin 1994). The first stage entails a preliminary limited search of databases and journals, which allows the identification of optimal search terms. The second stage of the search incorporates all the identified optimal search terms. A similar approach using a strategy with increasing complexity has also been suggested by Counsell (Counsell 1997). The initial search uses terms that refer to the disease or condition of interest, then depending on the number of references identified, the search is modified by combining other terms such as exposure or study design.

The Cochrane Collaboration suggest a good approach to a comprehensive search strategy is to begin with multiple terms that describe the disease or condition and join them together with the Boolean “OR” operator (Mulrow and Oxman 1997). The search can be narrowed by joining these terms to an intervention with the operator “AND”. They note that while the review question may address a population, setting and specific outcomes, these concepts are often poorly indexed in electronic databases. The search can also be narrowed using methodological terms such as “clinical trial” (Mulrow and Oxman 1997). Thus there are three options for the final search structure, these bring:

- condition *AND* intervention,
- condition *AND* methodology,
- condition *AND* intervention *AND* methodology.

#### Barriers to Successful Searches

There are barriers to successful searches and these relate to issues such as the specialty languages and indexing practices. Specialist fields have their own sub-vocabulary and this can hinder effective identification of studies (Lowe and Barnett 1994). That is, the terms used by the specialty group to describe a topic may differ from those used by the reviewer. This means that even though a comprehensive search is conducted, the use of differing terms means many relevant papers are missed. Ways to overcome this include using the MeSH Tree to find appropriate search terms from the general categories, or to use known articles to view the indexing terms (Lowe and Barnett 1994).

Another difficulty relates to the accuracy of indexing terms used to identify references. Many RCTs have not been indexed, or indexed incorrectly in databases, and one search of the British Medical Journal and The Lancet for the years 1948 to 1965 identified 1916 clinical trials not indexed as RCT (McDonald et al. 1996). The accuracy of MEDLINE indexers in selecting correct Medical Sub-headings (MeSH) for articles has also been questioned (Farby 1993). One search using free text words

uncovered 801 citations, of which 80% were relevant, that were not identified in a search using only the MeSH terms. This suggests that both free text and MeSH terms should be used for MEDLINE searches (Farby 1993).

A comparison of the indexing practices used in MEDLINE and CINAHL found significant differences between the two databases that could influence the results of searches (Brenner and McKinin 1989). It was found that MEDLINE used twice as many descriptors per citation than CINAHL, and as a result, had a far greater depth of indexing. This study found that while 70% of CINAHL controlled vocabulary is the same as that used by MEDLINE indexers, there was little common agreement in the descriptors assigned to citations between the two databases. These findings suggest that a common search strategy for all databases will likely miss many relevant references (Brenner and McKinin 1989).

These issues have significant implications for the reviewer developing a search strategy. They suggest that searches must be planned and evaluated prior to beginning the search. In developing the strategy a series of limited searches will help identify the optimum terms. Evaluation of the results of these searches will also help minimise the risk of a search being either too broad and retrieving large numbers of irrelevant references, or too focused and retrieving too few. Finally, a search strategy will need to be developed for each database, as a common strategy for all will likely result in many relevant references being missed.

### The Extended Search

While a large proportion of the studies will be identified in the electronic databases, many studies may be missed if only databases are used. Some studies are not published in mainstream health journals, and because of this are very difficult to find. Other studies are never published. To help identify these papers a number of strategies have been developed including hand searching journals, checking

bibliographies of articles, contacting experts and checking for studies reported at conferences.

### 1. Hand Searching

Manual searching of journals is undertaken as part of the search strategy of many reviews (Mulrow and Oxman 1997). Hand searching is manually checking each volume of select journals to identify a range of publications, such as studies not indexed in databases, studies that have been incorrectly indexed, letters to the editor which report research results, or conference abstracts. However, because this type of searching is very time consuming, it is usually restricted to key journals on the review topic (Critical Reviews Advisory Group 1996).

### 2. Bibliographies

Searching reference lists and bibliographies of all identified papers is a useful way to identify studies missed during the database search (NHS Centre for Reviews and Dissemination 1996; Evans 2001). The reference lists of systematic reviews, literature reviews and discussion papers that address the topic of interest provide an efficient way to identify missed studies. However, this search is complicated by the number of errors found in bibliography lists. Evaluations of the accuracy of reference lists in medical journals have found that major errors are common (McLellan et al. 1992; Benning and Speer 1993; Goldberg et al. 1993). Evaluations of nursing journals have identified similar problems (Foreman and Kirchhoff 1987; Taylor 1998). These inaccuracies increase the difficulty of the retrieval of identified papers.

### 3. Contacting Experts

The validity of the statistical analysis in a research study depends on the validity of the data, and so it is recommended that attempts be made to find both published and unpublished studies (Dickersin 1994; Counsell 1997). For reviews of nursing topics, a major issue is the poor publication rate of research by nurses, adding to the difficulties of the search. One survey found that of the nurses who had conducted research, 58%

had written up the research but only 10% had actually submitted it for publication (Hicks 1995). Identifying these unpublished studies is very difficult, because by their very nature, there is no public record of their existence.

It has been suggested that contacting experts in the field or centres of excellence may be useful (Critical Reviews Advisory Group 1996). One evaluation of the value of contacting experts during a systematic review identified an additional 300 references of which 40 were found to be eligible for inclusion in the review (McManus et al. 1998). This evaluation found that 24% of studies would have been missed if experts had not been contacted. These experts have been described as the “invisible college”, or research community and are an important source of both published and unpublished work (Rosenthal 1994). However, for topics without an easily identifiable expert body, making contact with experts can be difficult and time consuming.

#### 4. Conference Abstracts

Researchers often communicate their findings initially to colleagues at conferences, scientific meetings and symposia. These meetings are part of the publication cycle, in that they provide opportunity for feedback on the work allowing modification before eventual publication in a journal (Kelly 1998). Conference presentations represent the most current research, and by the time they reach publication the findings are often old news to researchers (Kelly 1998). Many of these presentations never reach publication, and so the conference abstract is the only remaining record of the work. Identifying these abstracts is difficult given that the National Library of Medicine stopped indexing the abstracts from conference and symposiums in 1981 as a result of concerns regarding their quality (Funk et al. 1988). Adding to this difficulty, some journals do not allow authors to cite conference papers in bibliographies (Kelly 1998). Databases, such as ProceedingsFirst (FirstSearch), now exist to help identify these abstracts, however coverage of conferences is currently extremely limited, and appears to be predominantly confined to non-nursing conferences.



The use of conference abstracts in reviews is controversial because of the lack of peer review and subsequent issues concerning their quality. One editorial even argued that allowing abstracts to be cited in bibliographies devalues the strength of the biomedical literature whose basis is the peer review process (Lichter 1993). Another concern is that of industry sponsored symposia, which are promotional in nature (Bero et al. 1992). These concerns related to quality are compounded because these abstracts are difficult to critically appraise due to the limited information normally provided. However, for some topics conference proceedings may represent the extent of current knowledge and so identifying their existence and summarising their findings may make an important contribution to the field.

### **Publication Bias**

One of the risks to the findings of a systematic review is that of publication bias (Mulrow and Oxman 1997; Jones and Evans 2000). Publication bias refers to studies that have statistically significant results, are interesting, or are large well funded studies, being more likely to be submitted and published than studies without such characteristics (Easterbrook 1991; Sutton et al. 2000). This publication bias can occur on two levels. Firstly, researchers are more likely to submit studies for publication if they demonstrate a positive result, or alternatively, they may submit only the parts of the results with positive findings. Supporting this, one exploration of reasons why research was not published found it was most commonly due to a lack of significant results (Easterbrook 1991). The second level of publication bias is the preferential selection by journals of studies that demonstrate significant findings (Dickersin et al. 1992). As a result of this bias published studies may tend to overestimate the treatment effect of an intervention. This overestimation is a result of those studies with non-significant results not being as well represented in the literature. This risk adds support to efforts to identify unpublished studies, and highlights the importance of ensuring the review includes a representative sample of all research (Dickersin 1994; Counsell 1997).

Begg suggests a preliminary analysis should be undertaken on completion of the search to assess the chance of publication bias (Begg 1994). To evaluate this, Begg recommends what has been termed a funnel plot. To do this, the treatment effect estimate (such as relative risk or odds ratio) of individual studies are plotted on the  $x$  axis, against the study's sample size on the  $y$  axis (Mulrow and Oxman 1997). In the absence of publication bias, the plot would normally be funnel shaped (NHMRC 1999). The graph is funnel shaped because the precision of the results increases as the study's sample size increases, meaning that studies near the top of the graph are clustered together around the mid-point (the mean) (Mulrow and Oxman 1997). The less precise results of small studies will be scattered more widely at the base of the graph on either side of the mid point. In the absence of publication bias the plot should be symmetrical because results are scattered evenly around the mean. If the plot is not symmetrical possible causes could include problems with the location of studies or excessive reliance on reference lists for study identification (Mulrow and Oxman 1997). When bias is suspected, caution should be used when interpreting the findings of the meta-analysis. However, the power of this method for assessing the risk of bias is quite limited, and the potential causes of publication bias should have been addressed by the searching strategy (Mulrow and Oxman 1997).

### **Searching for Interpretive Research**

The major focus and methodological development of systematic reviews to date has focused on experimental research, but how useful these methods are in identifying interpretive research is not known. The risk will likely be similar to that encountered by the database searcher looking for RCTs, where evaluations have demonstrated that these searches may only find 30% to 80% of published studies (Dickersin 1994). That is, the reviewer may undertake the search and find enough studies to be satisfied but be unaware of the many studies missed. The percentage of interpretive studies

missed during database searches may be much greater because of the number of research designs that are categorised under the broad title of interpretive research.

Evaluation of the existing reviews of interpretive research reveals that some provided scant information of the search processes (Jensen 1994; Sherwood 1997), and one review limited the search to only the CINAHL database (Fredriksson 1999). However others have conducted searches that appear to be of a standard similar to that expected of systematic reviews (Paterson et al. 1998; O'Neil and Morrow 2001; Woodward and Webb 2001). It would appear that the literature searches used in existing reviews of interpretive research are highly variable, and additionally, that reporting of the methods is inadequate at times (Evans and Pearson 2001a). However, in attempting to develop an interpretive research search strategy, the terms used in the titles and abstracts by the researchers and descriptors terms used by the databases add to the difficulty and complexity of the process.

#### Title Searches

An important component of any search for RCTs has been to use specific keywords to search the titles of studies listed in databases. However, this approach to searching assumes that the title of the article clearly describes its contents. For example, the title "Intravenous heparin for the prevention of stroke progression in acute partial stable stroke: A randomised controlled trial" (Duke et al. 1986), provides a clear description of both the population, condition, outcome and research method. Identification of this study during a database search would be relatively simple. This means that this study would likely be identified during any database search addressing the use of heparin in people with stroke.

In comparison to this precise terminology, some interpretive research articles use titles that could best be termed descriptive. These descriptive titles clearly describe the focus of the study and are for most situations appropriate. However, during

database searches they can add to the difficulty of the search because of their lack of specific keywords. Examples of descriptive titles include:

- “How can I put this?” Exaggerated self-disparagement as alignment strategy during problematic disclosures by patients to doctors (du-Pre and Beck 1997).
- Unbearable incidents: failure to endure the experience of illness (Dewar and Morse 1995).
- Setting boundaries: a strategy for precarious ordering of women’s caring demands (Wuest 1998).

Each of these titles provides a rich description of the study that is understandable to the reader, while at the same time making selection of optimal search terms very difficult. The basis of this difficulty is that the author and database searcher may differ in how they define a concept. It has been suggested that this differing definition of concepts is an important factor contributing to failed database searches (Lowe and Barnett 1994). While these differences in how a topic is defined can also occur during RCT searches, it is a greater problem during searches for interpretive studies. As a result of these difficulties a search of the titles using specific terms is likely to miss many relevant studies. From this perspective, studies with titles that do not have terms that clearly identify the subject of the study will be more likely to be missed during the searches. Unlike the example of the heparin RCT, these interpretive studies have a greater chance of being lost to the profession in the large volume of health care literature.

### Abstract Searches

Electronic databases include abstracts as part of the information provided about articles, and searching these abstracts using keywords is also a common part of the search strategy for RCTs. These abstract searches can focus on research method using keywords such as random or controlled, or on the study subject using keywords such

as heparin and stroke. However, the success of these searches will be influenced by the completeness of the information provided in the abstract.

The structure of abstracts for RCTs has come under some scrutiny, and there has been a call for more informative abstracts (Haynes et al. 1990; Froom and Froom 1993; Taddio et al. 1994; Wilczynski et al. 1995). Better structure of abstracts not only assists searching, but will also make it easier to determine if a study is relevant to the review. While there has been no formal evaluation of the content of abstracts from interpretive studies, their contents appear to vary considerably, and not all provide a clear description of the method or subject. Once again, while this does not necessarily reflect on the quality of the study, it increases the difficulty of finding this research.

#### Index Term Searches

Another method used to identify studies cited in databases is through the use of the index terms. Each index term represents a single concept and they are used to describe the contents of articles listed in the database. In MEDLINE there are approximately 18,000 index terms and CINAHL has over 9000 (Lowe and Barnett 1994; CINAHL Database 1999). These index terms describe both the subject of an article and the research methods.

Index terms are used to good effect during RCT searches, and so could also be used during the database search for interpretive studies. However, there is little information on the usefulness of this approach. Additionally, while it has been reported that many RCTs have been indexed incorrectly (McDonald et al. 1996), there is no information on the accuracy of the indexing of interpretive research. Adding to these difficulties, it has been reported that while 70% of the controlled vocabulary used in CINAHL is the same as that used by MEDLINE indexers, there is little common agreement on which index terms are assigned to articles (Brenner and McKinin 1989). As interpretive research has not been a central component of the evidence-based movement, there has

been little attempt to investigate the indexing practices of databases for this type of research.

In comparing the indexing of interpretive research articles in MEDLINE and CINAHL significant differences can be noted. Interpretive papers in MEDLINE appear to be indexed under what can perhaps best be described as a “quantitative framework”. That is, the terms used are better suited to research such as the RCT. In comparison, the same papers indexed in CINAHL have terms that appear to more accurately reflect the interpretive methodology. As an example of this difference, Table 2 lists the indexing terms used by these two databases to describe the same research article. In this example, CINAHL clearly utilises a greater number of indexing terms focusing on the interpretive methods used during the research. While this issue has not been addressed in the literature, this difference in indexing is apparent in the citations of many studies found in both MEDLINE and CINAHL.

This single example clearly demonstrates that for interpretive research, the indexing practices of CINAHL are far more comprehensive than those of MEDLINE. Indeed, in this example MEDLINE has not assigned any index terms related to the methods used in the study. This is significantly different to the indexing of RCTs, where a number of terms are used to aid in the easy identification of all RCTs. From this perspective, while it would be possible to search for interpretive studies in CINAHL using keywords related to both subject and research method, this approach would be likely to fail in MEDLINE. This issue is supported by a comparison of the two databases, which found that the index terms used in CINAHL were more focused towards nursing topics than those used in MEDLINE (Brenner and McKinin 1989).

**Table 2**

The Indexing of the Same Paper in MEDLINE and CINAHL  
“Infant feeding choices among first time mothers” (Keith 1997).

<b>CINAHL</b>	<b>MEDLINE</b>
<u>Subject Index Terms</u>	<u>Subject Index Terms</u>
<ul style="list-style-type: none"><li>• infant-feeding</li><li>• adult</li><li>• pregnancy</li><li>• female</li><li>• breast-feeding</li><li>• infant-newborn</li><li>• decision-making-patient</li><li>• maternal-attitudes-evaluation</li><li>• maternal-attitudes</li></ul>	<ul style="list-style-type: none"><li>• adult</li><li>• child-preschool</li><li>• infant</li><li>• infant-nutrition</li><li>• child-nutrition</li><li>• interpersonal-relationships</li><li>• decision-making</li><li>• mothers-psychology</li><li>• United States</li></ul>
<u>Research Method Index Terms</u>	<u>Research Method Index Terms</u>
<ul style="list-style-type: none"><li>• qualitative-studies</li><li>• grounded-theory</li><li>• field-notes</li><li>• interviews</li><li>• thematic-analysis</li><li>• nursing-models-theoretical</li><li>• convenience-sample</li><li>• audiorecording</li><li>• constant-comparative-method</li><li>• theory-construction</li><li>• theoretical-sample</li></ul>	<ul style="list-style-type: none"><li>• models-psychological</li></ul>

## Summary

This section has highlighted the importance of developing systematic strategies if all relevant studies are to be identified. For most searches, the use of free-text and index terms addressing both the review subject and the study method should be used. A range of electronic databases should be searched to increase the scope of the search. The recommended approach for searches involves a series of steps and these steps include such things as reference list searches and hand searching key journals.

In terms of searching for interpretive studies, while much of the discussion was based on anecdotal information, it does suggest that if all studies are to be identified, a range of strategies must be used. This type of search would have to include both free-text and index terms, and it would be important to develop search strategies for each specific database. Failure to address these issues during the development of the strategy would mean that many studies may be missed.

## **Expanded Review Process: Search Strategy**

In terms of the expanded review process, the principles used to identify RCTs were used as the basis for the development of the search strategy. This strategy entailed a series of searches, each increasing the breadth of previous searches. Attention was given to ensure all relevant databases were searched and that a comprehensive range of keywords was used to increase the likelihood that all interpretive research were identified. This was supplemented by hand-searches of journals to identify those interpretive studies poorly indexed in databases. The specific steps of this search were:

- initial search of literature and databases to gain an understanding of the issues of importance related to identifying this group of studies,
- determined what databases should be searched and the optimal search terms for each,



- comprehensive search that progressively incorporated all selected databases,
- manual search of reference lists and bibliographies of relevant papers,
- manual search of selected journals, and
- search of published conference abstracts and conference databases.

This strategy was based on approaches used to identify RCTs and so increased the likelihood of finding all relevant studies. Increasing the range of keywords to fit with those used in interpretive research should also help to identify this group of studies. While it is impossible to identify every study ever conducted on the review topic, this approach ensured that all reasonable efforts were made to identify all studies of importance.

# Critical Appraisal

## Introduction

The aim of systematic reviews is to provide a summary of the best evidence. However, the quality of primary research is highly variable because of issues such as the use of inappropriate research methods, poor conduct of the research and inappropriate data analysis. At times, because of systematic errors or bias, the research findings can provide misleading or incorrect information. To minimise these risks, studies to be included in a systematic review are appraised for methodological rigour. This appraisal aims to discover if the methods and results of research are sufficiently valid for the findings to be considered useful information (Fowkes and Fulton 1991). The use of valid studies should lead to better and more realistic estimates of treatment effect and more accurate and reproducible estimates of treatment efficacy (Moher et al. 1995). This critical appraisal also helps the review gain insights into potential comparisons and can guide interpretation of the review findings (Mulrow and Oxman 1997). Critical appraisal has three objectives:

- to understand the validity of studies;
- to uncover reasons for differences among study results; and
- to provide readers of the review with sufficient information with which to judge for themselves the applicability of the systematic review to their clinical practice (Meade and Richardson 1997).

Appraisals do not focus on presentation or writing style, rather they are concerned with the hard facts of the research (Fowkes and Fulton 1991). However, the research can only be appraised indirectly through the research report. Moher *et al.* suggest that it is important to distinguish between assessing the quality of the research and the quality of the research report (Moher et al. 1995). They define the quality of the research as the degree to which the trial design, conduct, and analysis has minimised or avoided biases, whereas the quality of the research report is defined in terms of the

provision of information about the design, conduct, and analysis of the trial (Moher et al. 1995). Unfortunately it is difficult at times to separate the report and the research, in that a poorly designed or conducted study that is well reported can be assessed as high quality (Moher et al. 1995). Conversely, a well designed and conducted study that is poorly reported will likely be assessed as being of low quality.

This critical appraisal of studies is important because, despite the peer review process of health care journals, a good proportion of the published research is invalid (Rosenberg 1995). Evaluations of the standard of statistical analysis of published research have demonstrated that errors are common, that vital information is often omitted and sample sizes are often inadequate (Altman 1982; Lynn 1985; Murray 1988; Mills 1993). A similar evaluation of statistical methods used in nursing research also found that errors were common, and that most of these errors were quite elementary (Anthony 1996). As a result of these problems, research findings are often contradictory and the failure of studies to demonstrate a beneficial outcome has more to do with the study sample size than the actual intervention. These studies, if used as the basis for clinical decision making, can result in ineffective, inappropriate and even dangerous health care practices.

Appraisal of studies can influence the review process from four different perspectives (Mulrow and Oxman 1997). Firstly, it can serve as a threshold for inclusion of studies in the review. In this situation, studies must meet a predetermined minimum standard before they can be included in the review. Secondly, the appraisal can also provide an explanation for differences between studies. Thirdly, the appraisal can be used in sensitivity analyses, where studies of differing quality can be initially excluded and then be included in the analysis to determine their impact on findings. Finally, critical appraisal has been used to weight studies during the statistical analysis to avoid all studies having an equal contribution to the meta-analysis regardless of their size.

## **Causes of Bias and Errors**

In experimental studies, bias or errors occur most often as a result of the methods used during four critical stages of the study and so have been termed:

1. selection bias,
2. performance bias,
3. attrition bias, and
4. detection bias (NHS Centre for Reviews and Dissemination 1996; Mulrow and Oxman 1997).

These four stages are critical because poor methods may influence the findings studies.

### **1. Selection Bias**

Selection bias relates to inadequate randomisation. Randomisation is important because it removes the human influence from the process of selecting subjects and allocating them to study groups. It has been argued that this selection bias is the most important source of bias in clinical trials (Mulrow and Oxman 1997). An evaluation of 250 controlled trials compared the findings on the basis of the method of randomisation, classifying randomisation as adequate, unclear or inadequate (Schulz et al. 1995). In controlled trials where the method of randomisation was unclear, the treatment effect was exaggerated by 30%, and by 41% for studies with inadequate randomisation when compared to the results of trials with adequate randomisation. Another investigation of controlled trials found studies that used sequential assignment showed a significantly higher likelihood that patients would do better with the experimental intervention than with standard therapy (Colditz et al. 1989). However, despite the importance of randomisation during controlled trials, another evaluation of 206 RCTs found that only 32% of studies used an adequate method for randomisation (Schulz et al. 1994).

## 2. Performance Bias

Bias can also be a consequence of differences in care provided to participants in the experimental and control groups, other than the intervention being evaluated (Mulrow and Oxman 1997). As a result of these differences, results may be influenced by factors other than the intervention. The risk from this source of bias is minimised by blinding study subjects and care providers to the treatment group (Mulrow and Oxman 1997). This blinding means that health care workers do not know whether participants are in the control or experimental groups and so similar care is more likely to be provided to all. While blinding is effective for such studies as pharmaceutical evaluation, for many interventions, blinding is difficult or impossible.

## 3. Attrition Bias

Differences between groups in the number of participants who do not complete the study can also influence the results (Mulrow and Oxman 1997). The basis of this concern, is that these differences may be the result of the intervention. Outcomes such as death, intolerance of the intervention, or inability to adhere to the treatment protocol could all result in participants withdrawing from the study and so being lost to follow-up. These losses can bias the findings because these unfavourable outcomes are not attributed to the intervention. To minimise this risk, all participants who are entered into the study should be accounted for in the results. Additionally 'intention to treat' analysis, whereby participants who are lost to follow-up are included in the statistical analysis, helps determine if inclusion of these participants could have changed the findings of the study.

## 4. Detection Bias

Finally, detection bias results from differences in the assessment of outcomes (Mulrow and Oxman 1997). The risk of this bias is greatest when outcome assessment involves some form of subjective judgement as found with the many scales used in health care evaluation. In this situation the person assessing the outcome is influenced by knowledge of whether the participants are in the

experimental or control group. The risk of detection bias is minimised by blinding the assessors to study allocation. However, for objective outcomes, such as mortality rates or length of stay, there is less risk of bias.

Other sources of bias reported in the literature include inadequate sample size and inappropriate statistical analysis. It has been suggested that the most common design error encountered in studies is that of having too small a sample size to get reliable or useful results (Altman 1982). The most important consequence of having too small a sample size is that the study lacks the power to detect important effects, and so can incorrectly find that an intervention has no beneficial outcomes. Yet it is common to find undue emphasis placed on the negative findings of small studies, that is, claiming that the intervention is ineffective while in reality it is an issue of a lack of evidence (Altman 1982). These negative findings are important because they can stop the introduction of effective treatments, continue those that are ineffective, or prevent new research evaluating the intervention from being initiated. From the perspective of systematic reviews and critical appraisal, small sample size is not normally a criteria for exclusion because the process of combining studies in a meta-analysis helps address this problem.

The other common problem with research reports is that of inappropriate or incorrect statistical analysis. This has led one commentator to suggest

*“If you torture your data long enough, they will tell you whatever you want to hear”* (Mills 1993, p. 1196).

The two major types of data torturing have been termed opportunistic and manipulative data analysis (Mills 1993). In opportunistic data analysis the researcher continuously analyses the data until a significant association is found. Manipulative data analysis involves manipulating data so that they prove the hypothesis. From the perspective of the systematic review, inappropriate statistical analysis is less of a problem because the results are analysed independently during the meta-analysis.

## **Appraisal of Experimental Studies**

A large range of tools has been developed to aid in the appraisal of research. An investigation into these tools identified 25 scales and 9 checklists, with the number of items per tool ranging from 3 to 57 items (Moher et al. 1995). They found that only 4% of these scales had been rigorously developed and that few reported inter-rater reliability. Based on this investigation, they recommended caution when using any of the scales (Moher et al. 1995), a recommendation that has been echoed by others (Mulrow and Oxman 1997).

An example of one appraisal tool, developed using a delphi technique, to rate the methodological quality of RCTs prior to meta-analysis addressed over 40 items which included:

- the use of a control group,
- method of randomisation,
- outcomes measurements,
- description of study design,
- conclusions,
- intention to treat,
- statistical analysis,
- adherence to protocol,
- blinding of outcome assessors, and
- miscellaneous factors (Sindhu et al. 1997).

However, the large number of items in this appraisal tool would involve considerable time to complete. While the interrater reliability was reported to be good, its validity has not been evaluated, and there is no information on whether having a large number of items to appraise studies is better than an appraisal tool with a small number of items.

Another tool developed and evaluated by Jadad *et al.* (Jadad *et al.* 1996), claimed to be simple to complete while also being both reliable and valid. This 3 item assessment tool focused on the major issues related to bias in clinical trials:

- Was the study described as randomised?
- Was the study described as double blinded?
- Was there a description of withdrawals and dropouts?

This tool allocated a score that ranged from 0 to 5, although the use of scoring systems in appraisal tools have yet to be shown to be useful. Jadad's appraisal tool gave an extra point for trials which were double blinded, despite it being noted that this approach discriminated against trials where double blinding was not possible (Moher *et al.* 1995).

The Cochrane Collaboration approach to critical appraisal involves rating studies against criteria, in terms of whether each item was met, un-met or unclear, then the risk of bias is ranked against this criteria (see table 3) (Mulrow and Oxman 1997). It has been suggested that this approach will likely be appropriate if only a few criteria are used, and that these criteria address only important threats to the studies' validity (Mulrow and Oxman 1997).

The Centre for Reviews and Dissemination at the University of York, has developed a critical appraisal tool for experimental studies which focuses on issues such as:

- method and concealment of randomisation,
- comparability of groups at entry,
- completeness of follow-up
- whether drop-out and missing data is handled appropriately,
- blinding and objective assessment of outcomes, and
- appropriate statistical analysis (NHS Centre for Reviews and Dissemination 1996).



**Table 3**

Categorising studies risk of bias according to individual criteria (Mulrow and Oxman 1997, p. 39).

Risk of Bias	Interpretation	Relationship to Individual Criteria
A. Low risk of bias	Plausible bias unlikely to seriously alter results.	All criteria met.
B. Moderate risk of bias	Plausible bias that raises some doubt about results.	One or more criteria partly met.
C. High risk of bias	Plausible bias that seriously weakens confidence in the results.	One or more criteria not met

In summary, while a considerable number of tools exists to appraise RCTs, there is little agreement on what is the optimal method. While the concept of critical appraisal has wide spread support, available tools vary significantly. The areas of the review process that are at considerable risk of bias relate to the selection of participants, performance, attrition and detection. When undertaking these appraisals a balance is needed, because at one end of the spectrum is the risk of including poor quality research in the systematic review, at the other end is the risk of critically appraising all the research out of the review. Both these risks bring serious problems. Inclusion of poor quality studies may mean ineffective treatments are supported by the findings of the systematic review. Excluding most identified studies on the basis of the critical appraisal also excludes the valuable contribution these studies offer. The risk in this situation is that searching for only those few methodologically perfect studies may be as dangerous to health care practice as is including poor quality research.

## Appraisal of Observational Studies

Less attention has been given to how best to appraise observational studies compared with the attention given to RCTs. However, while observational studies are at risk of many of the same sources of bias as the RCT, they may also be more vulnerable to threats to validity than RCT (NHS Centre for Reviews and Dissemination 1996). Suggested ways in which the risk of this bias can be minimised are listed in table 4.

**Table 4**

Minimising bias in observational studies (Mulrow and Oxman 1997, p. 41).

Source of Bias	Cohort Studies	Case Control Studies
Selection bias	Control for confounders	Matching
Performance bias	Measurement of exposure	Measurement of exposure
Attrition bias	Completeness of follow-up	Completeness of follow-up
Detection bias	Blinding	Case definition

Selection relates to having two equal groups, although the lack of randomisation means that unknown or unmeasured confounders can not be controlled (Mulrow and Oxman 1997). Selection in observational studies relates to efforts to ensure both groups are derived from the same population and that they are comparable for all factors other than the intervention or characteristic of interest. Assessing performance bias in these studies is more difficult than for RCTs, as exposure to the intervention of interest must be measured to ensure differences between groups do not exist (Mulrow and Oxman 1997). Risk of bias through attrition is similar for both cohort and case control studies, and is minimised by ensuring all participants are appropriately accounted for. In terms of detection bias, the risk is different for cohort and case control studies. For cohort studies, this risk can be minimised through appropriate case definitions and by the blinding of those who assess the outcome of interest, as is done during RCTs. However, as case control studies are retrospective, blinding of assessors is not

possible because the starting point of the case control study is the outcome event. Therefore the risk of this bias is avoided through appropriate case definitions.

Like RCTs, problems related to poor research design have also been reported in observational studies. An assessment of 85 case control studies against accepted methodological standards, found the standards were commonly violated and the authors highlighted the need for these types of studies to follow more closely conventional research procedures (Horwitz and Feinstein 1979). To address this problem, an appraisal tool was developed by the Centre for Reviews and Dissemination at the University of York (NHS Centre for Reviews and Dissemination 1996) (see table 5).

In summary, less attention has been paid to the appraisal of observational studies compared to that given to the RCT. However, in terms of risk of bias, observational studies are at similar risks as threatens the RCT. These risks relate to the selection of participants, performance, attrition and detection. However, as with RCTs, searching for those few perfect observational studies creates as many problems as including poor quality research in a review.

**Table 5**  
Critical Appraisal of Observational Studies (NHS Centre for Reviews and  
Dissemination 1996, p. 75 - 76)

**Cohort**

- Are exposed people representative of the standard users of the intervention?
- Was the non-exposed cohort selected from the same population as the exposed?
- Was exposure reliably ascertained and verified?
- What factors (other than exposure) may affect outcome?
- Were the cohorts comparable on these important confounding factors?
- Was there adequate adjustment for the effects of these confounding variables?
- Was dose-response relationship between exposure and outcome demonstrated?
- Was outcome assessment blind to exposure status?
- Was follow-up long enough for the outcomes to occur?
- Was an adequate proportion of the cohort followed-up?
- Were drop-out rates similar in the exposed and unexposed groups?

## Case Control

- Has the disease state of the case been reliably assessed and validated?
- Are the case representative of a series, or is there a potential for selection bias?
- Were the controls selected from a similar population?
- Is there evidence that the controls are free from disease?
- How comparable are the case and controls with respect to potential confounding factors?
- Were hazards and interventions assessed in the same way for cases and controls?
- Was the response rate adequate?
- Were the non-response rates the same in both groups?
- Is it possible that over-matching has occurred in that cases and controls were matched on factors related to exposure?
- Was an appropriate statistical analysis used?

## Appraisal of Interpretive Studies

The critical appraisal of interpretive studies is more difficult and has been the subject of considerable controversy. Even the terms 'interpretive' and 'qualitative' add to this difficulty because they are used to refer to range of different methods such as phenomenology, grounded theory and ethnography. These differing methods are often not differentiated when issues related to quality in interpretive research are addressed (Evans and Pearson 2001a).

One of the issues currently being addressed is how interpretive research can be assessed. From the perspective of the reviewer, standard criteria for quality would aid in the appraisal, summary and synthesis of interpretive research. To date there has been no agreement on these criteria, and it is unlikely that a single tool will be sufficient for the many different methods. This ongoing professional debate has prompted Sandelowski to suggest that:

*"The problem of rigour in qualitative research continues to arouse, beguile and misdirect." (Sandelowski 1993, p. 1)*

Schwandt suggests that if we are to judge the goodness of the undertaking and product of social inquiry, then an alternative to the traditional criteriology must be sought (Schwandt 1996). Engel and Kuzel see the debate on how to judge the quality of

interpretive inquiry as part of the larger discussion about the efficacy of alternative paradigms in the human sciences (Engel and Kuzel 1992). They suggest that the pendulum has now swung so that the burden of proof has been placed on the proponents of these alternative paradigms. However, this alternative paradigm argument proposed by Engel and Kuzel is reminiscent of the qualitative-quantitative debate that plagued nursing for much of the 1980s and a good part of the 1990s. While the issue of qualitative versus quantitative has to some extent been resolved, in that the research method that best answers the question can now be confidently used, these issues have re-emerged during the debate surrounding critical appraisal and systematic reviews. Supporting this is the fact that a criticism of early appraisal tools for qualitative research was that they used criteria developed for experimental research (Burns 1989). It is reasonable to expect that the optimal criteria for judging the standard of interpretive research will, through necessity, have to be developed specifically for this purpose and that the translation of criteria from other branches of research will simply confuse the issue.

Lincoln refers to this debate as a dialogue about emerging criteria, because the entire field of interpretive inquiry is itself still emerging and being defined (Lincoln 1995). Because of this there are fewer fixed regulations than are encountered in other forms of inquiry. However, the debate on how best to appraise this research is similar to that surrounding the appraisal of RCTs, as witnessed by the large number of checklists and scales that have been developed (Moher et al. 1995). Currently, how best to judge the quality of any research method, based on the published report, is unclear.

### Criteria for Interpretive Research

While much has been written on the critical appraisal of interpretive research and a few tools developed (Burns 1989; Forchuk and Roberts 1993), the individual criterion that should be used is not clear. Estabrooks *et al.* suggest that when examining qualitative research findings for the purpose of aggregation, it is critical that labels are

closely and explicitly tied to supportive data or exemplars (Estabrooks et al. 1994). This is because categories or themes used across different studies may not always have the same label. These authors suggest that studies that are weak due to superficial analysis, or research biases due to deductive contamination from other theories, will stand out because the supporting data in the form of description or quotation is missing (Estabrooks et al. 1994). They recommend that these “weak studies” be excluded from the analysis. However, while this argument of evidence of a data trial appears to be a common theme in appraisal of qualitative studies, it fails to address poorly written reports. Methodologically sound studies that are poorly reported in a journal, or articles that are condensed due to word limitations of a journal, would be assessed as being a weak study due to a lack of supporting descriptions.

However, this is a problem that is also faced when appraising RCTs. Estabrooks *et al.* suggest that this type of appraisal needs to be undertaken by an experienced researcher, able to understand multiple qualitative methods, have an appreciation for epistemological issues and have the quality of scholarly vigilance (Estabrooks et al. 1994). However, this elitist approach means that users of interpretive research must rely on the opinions of experts to determine what studies are of quality high enough to serve as a basis for nursing practice. Additionally, these experts who will judge the quality of studies are also likely to be the producers of the research and so to obtain an unbiased appraisal may prove difficult.

It has also been suggested that it is a mistake to assume that the problem of what constitutes quality work must be resolved in one way for all paradigmatic contexts (Engel and Kuzel 1992). This ‘all or nothing’ solution encourages an ‘in principle’ style argument which is designed to cover all cases, and is therefore so abstract it is difficult to judge the applicability of the argument. Appraisal of experimental and observational studies has involved a variety of tools and so it is also likely that to appraise interpretive research, a single tool to cover all approaches will not be

feasible. It has been suggested that the criteria for judging the quality of interpretive studies would need to be specific for each method, and so these criteria may differ considerably across approaches (Smith 1987). This view has been supported by others, noting that different traditions might require different criteria (Lincoln 1995).

Interpretive research has been described as context sensitive, in that human acts are shaped by the physical material and social environment. Smith argues that an implication of this belief in context sensitivity is a de-emphasis of standardised or general research methods (Smith 1987). That is, standardised methods have little utility and are not guarantors of truth. This theme has been extended to the culturally context-dependent nature of knowledge, in that different communities of knowledge makers have sanctioned different criteria of goodness (Sandelowski et al. 1997). Because interpretive research lacks a catalogue of designs or certified methods, the researchers must describe what they did in detail (Smith 1987). It might therefore be suggested that while the debate continues on how best to appraise this research, minimum quality criteria would be a clear description of the what was done during the study. This expectation of describing what has been done has already been used during reviews of interpretive research, where evidence of a decision trail formed part of the inclusion criteria (Thorne and Paterson 1998).

An inspection of existing reviews of interpretive research to determine if critical appraisal was part of the review process, highlighted great variation in the practices used and many inconsistencies. Thorne and Paterson, during a review of interpretive research on chronic illness, appraised all studies “to assess the rigour, epistemologic soundness, and fruitfulness of the research methods used” (Thorne and Paterson 1998). One review of interpretive research addressing ‘wellness-illness’ placed no restriction on the scientific merit of studies to avoid eliminating data germane to the purpose of the investigation (Jensen 1994). Another review failed to mention critical appraisal or study quality in the description of the review method (Paterson et al. 1998). A review of the stressors in families with a child with chronic illness, did not

critically appraise identified studies and also noted in their method section that studies published in clinical journals that did not give details of the analytic procedures were also included in the review (Ogden-Burke et al. 1998).

There is no common agreement on how to deal with critical appraisal during the conduct of reviews of interpretive studies. The critical appraisal tool used during the review of chronic illness by Thorne and Paterson (Thorne and Paterson 1998) was based on the work of Burns (Burns 1989). This appraisal tool has also been utilised during a review of caring (Sherwood 1997) and during a review by Barroso and Powell-Cape (Barroso and Powell-Cope 2000). Unlike the tools used to appraise experimental or observational studies, Burns attempts to appraise all types of interpretive research with a single tool based on five standards (Burns 1989):

1. Descriptive Vividness
2. Methodological Congruence
  - rigour in documentation
  - procedural rigour
  - ethical rigour
  - auditability
3. Analytical Preciseness
4. Theoretical Connectedness
5. Heuristic Relevance
  - intuitive recognition
  - relationship to existing body of knowledge
  - applicability

However, because of the abstract nature of these criteria, each appraiser would likely reach differing conclusions about the quality of a study. Additionally, it has yet to be demonstrated whether the different interpretive methods can all be adequately appraised using a single tool.



Forchuk and Roberts developed an appraisal tool for interpretive studies as an aid for undergraduates and those unfamiliar with this type of research (Forchuk and Roberts 1993). This appraisal tool was based on seven questions.

1. Is the research design appropriate for the research topic or question?
2. Was the specific qualitative research method chosen appropriate?
3. Was the literature review done and used appropriately?
4. Were the descriptions of informants/participants, context and researcher all adequate?
5. Were information gathering and information analysis both described adequately?
6. Were researcher's conclusions appropriate?
7. Was the importance and relevance of the research to the intended progression indicated?

Both these proposed tools do not appear to have undergone any form of evaluation in terms of validity or inter-rater reliability. In comparing these tools, there are few similarities. While Forchuk and Roberts focus on issues such as an adequate literature review, appropriate information gathering and conclusions, Burns has taken a more abstract approach in terms of descriptive vividness, analytic preciseness and heuristic relevance. Neither tool is intuitive or user friendly, and so if they were to be used by reviewers, they would be likely to produce widely differing appraisals of study quality.

In summary, there is currently no rigorously developed and evaluated tool available to appraise interpretive research. However, from the previous discussion, some issues emerge as important factors in any attempt to appraise these studies:

- clear documentation of methods used during study,

- evidence of a decision trail, and
- supportive data or exemplars for themes, categories and labels.

### **Appraisal of Descriptive Studies**

Descriptive studies can also provide valuable information about health care and the use of specific interventions. Descriptive studies also provide important information on the circumstances surrounding the intervention, conditions related to its use, and the opinions of consumers. However, there is little information available on the quality of published descriptive studies. Yet based on the evaluations of other types of research, it is likely that some, or possibly many, descriptive studies will also suffer from a lack of rigour.

The potential lack of rigour in this research lends support to the need to also appraise descriptive studies before inclusion in a review. However, there is little information currently available on how this can best be achieved. Additionally, like interpretive research, a number of different methods are broadly classified under the title of descriptive research. While no validated critical appraisal tool currently exists, there is a minimum of information that must be provided about the descriptive study before its findings can be used. This information is similar to that needed for an appraisal of interpretive studies and would focus on a description of what was done, who was involved and the results. This amount of information is essential if a study is to be judged appropriate for inclusion. For example, without a description of the methods used, even a limited assessment of the trustworthiness of the study is impossible. Without a description of the population, setting and circumstances, the relevance of the study to the review is difficult to determine. Finally, without supportive data, the findings generated from the study must be accepted on trust.

### **Expanded Review Process: Critical Appraisal**

While no universally accepted method of appraisal exists for any of the different types of research, for the purposes of this study, the common themes emerging from the review of the literature were used. For experimental and observational studies the major threats to validity arise at four critical points of the research which are:

- selection,
- performance,
- attrition, and
- detection.

It is reasonable therefore that assessment of the validity of both experimental and observational research should primarily focus on these issues (see table 6). While many other aspects of the studies could also be appraised, the value of these in the context of critical appraisal is less clear. Firstly, other than for randomisation in controlled trials, there is no evidence to demonstrate that many of these other components used in appraisal tools actually influence the findings of the studies. That is, these criteria have a theoretical base rather than factual. Secondly, there is a risk that during the critical appraisal it is possible to appraise all identified research out of the review. From this perspective, the use of an appraisal tool with a standard so high that few studies can achieve it, is questionable. The use of an excessively high standard also means that the time and effort taken to identify all research on a topic is lost when potentially useful evidence is excluded. However a balance is needed as too low a standard would mean studies of questionable rigour could distort the findings of the review.

**Figure 6**

**Appraisal of Research for Expanded Review Process**

**CRITICAL APPRAISAL**

<b>RCT</b>	<b>Cohort Studies</b>	<b>Case Control Studies</b>
• Allocation bias	• Selection	• Selection
• Performance bias	• Exposure	• Exposure
• Detection bias	• Detection	• Detection
• Attrition bias	• Attrition	• Attrition

**MINIMUM STANDARD**

<b>Interpretive</b>	<b>Descriptive</b>
• Method appropriate	• Method appropriate
• Reporting of method	• Reporting of method
• Description of participants	• Description of participants
• Supportive data of themes & labels	• Supportive data provided

For interpretive research, there is currently no validated method to critically appraise these studies. While it can reasonably be expected that key stages of the research process bring some degree of risk, these risks differ considerably from those of experimental and observational studies. However, from the discussion it is obvious that minimum criteria exist. These criteria focus on the information provided in the report rather than the methods used during the study (see table 6). However, without this information the findings of interpretive research are of little use. The minimum criteria would include:

- clear description of study population,
- evidence of decision trail, and
- themes and labels supported by data.

One final criterion that could be added to these would be that the method was appropriate to answer the research question. While this addresses issues other than reporting, it is a critical issue for all research regardless of method.

Similarly for descriptive studies no appraisal criteria exists, but like interpretive research, a minimum standard can reasonably be expected before a study is included in a systematic review (see table 6). Like the appraisal criteria for interpretive studies, it would focus on the report rather than the study method. These criteria would include:

- clear description of study population,
- clear description of methods used, and
- supportive data provided.

Once again, a final criterion would be that the descriptive design was appropriate for the study question.

These criteria seek to balance the risk of needlessly excluding studies from the review with the risks associated with inclusion of poor quality research. These criteria, summarised in table 6, formed the basis for the development of the appraisal checklist used during the conduct of the review (see appendix 2). This appraisal checklist helped with the appraisal of studies and also served as a permanent record of those decisions related to appraisal, and so became part of the decision trail for the systematic review.

## **Hierarchies of Evidence**

### **Introduction**

In addition to the problems related to quality, it has also long been recognised that not all research designs are equal in terms of risk of error and bias in their results. When seeking answers to specific questions, some research methods provide better evidence than other methods. That is, the validity of the results of research varies as a consequence of the different methods used. For example, when evaluating the effectiveness of an intervention, the RCT is considered to provide the most reliable evidence (Muir Gray 1997; Mulrow and Oxman 1997; Sackett 1997). It is considered the most reliable evidence because the processes used during the conduct of the RCT minimise the risk of confounding factors influencing the results. This reduction in risk is attributed to the fact that all participants have equal chance of being in either the control or treatment groups, are treated exactly the same for all aspects of care except the intervention under investigation, and outcomes are measured in the same manner for each participant. As a result of this, the findings generated by RCTs will likely be closer to the true effect than those generated by other methods.

This confidence in the findings of research has important implications for those developing practice guidelines and clinical recommendations, or implementing the results of research into their area of practice. The aim during this process of development and implementation is to use the best evidence. It has been suggested that when recommendations are developed based on inadequate evidence they often require reversal when sufficient evidence becomes available (Guyatt et al. 1995).

The problem that arises from this situation is how to identify the evidence that best answers the question. To address this, hierarchies of evidence have been developed to allow research recommendations to be graded. These hierarchies or levels of evidence are used to grade primary studies according to their design and reflect the degree to which different study methods are susceptible to bias (NHS Centre for Reviews and

Dissemination 1996). By ranking research designs according to their internal validity, not only is the strength of the evidence graded, but also the confidence the end-user can have in the findings is also discovered.

In terms of systematic reviews, the hierarchy of evidence is used at two stages of the review. Firstly it is used to determine which studies designs best answer the review questions. From this perspective, the hierarchy informs the development of the inclusion criteria to ensure optimal research designs become the focus of the review. Secondly, the hierarchy is used to grade the strength of any recommendations based on the findings of the review. From this perspective, it informs the assessment of the systematic review findings.

### **Current Approaches**

Hierarchies of evidence were first popularised by the Canadian Task Force on the Periodic Health Examination in the late 1979 (Canadian Task Force on the Periodic Health Examination 1979). Until recently, these hierarchies focused on effectiveness, and for this reason, the RCT was most commonly listed as providing the highest level of evidence. A large number of different hierarchies have been developed and used. One hierarchy developed in 1986, and modified over the following ten years, was used during the summary of research related to antithrombotic agents (Sackett 1986; Cook et al. 1992; Cook et al. 1995). This grading system focused on the power of the studies, and the resultant risk of error. Levels I and II of this scale focused on the size of studies:

- Level I      Large randomised trials with clear-cut results (and low risk of error).
- Level II     Small randomised trials with uncertain results (and moderate to high risk of error).

By 1990, the approach to ranking evidence had expanded to include well designed controlled trials without randomisation (Woolf et al. 1990). However, as with the appraisal of research, there was no agreement on the optimal approach. For example, one hierarchy for clinical recommendations used levels A1 through to level C2 and included 'number needed to treat' (NNT) and confidence intervals.

- A1 RCTs, no heterogeneity, CI all on one side of threshold NNT.
- A2 RCTs, no heterogeneity, CI overlap threshold NNT.
- B1 RCTs, heterogeneity, CI all on one side of threshold NNT.
- B2 RCTs, heterogeneity, CI overlap threshold NNT.
- C1 Observational studies, CI all on one side of threshold NNT.
- C2 Observational studies, CI overlap threshold NNT (Guyatt et al. 1995, p. 1801).

Another complex hierarchy was developed by Carruthers *et al.* during a consensus conference addressing hypertension (Carruthers et al. 1993) and was later used during the development of practice guidelines for the management of diabetes (Meltzer et al. 1998) (see table 7).

Carruthers' hierarchy incorporated both sample size and quality assessment components as part of the criteria. Despite its complexity, this hierarchy allowed a range of evidence to make some contribution to the findings of the review, while at the same time cautioning the reader when recommendations had been based on lesser quality research.



**Table 7**  
Levels of evidence (Carruthers et al. 1993, p. 290-1)

Level I	A RCT that demonstrates a statistically significant difference in at least one important outcome (eg survival or major illness), or, If the difference is not statistically significant, an RCT of adequate sample size to exclude a 25% difference in relative risk with 80% power, given the observed results
Level II	An RCT that does not meet the level I criteria
Level III	A non-randomised trial with contemporaneous controls selected by some systematic method (eg not selected by perceived suitability for one of the treatment options for the individual), or, Sub-group analysis of a randomised trial
Level IV	A before-after study or case series (of at least 10 patients) with historical controls or controls drawn from other studies
Level V	Case series (at least 10 patients) without controls
Level VI	Case report (fewer than 10 patients)

An Australian developed hierarchy used a scale of level I through to level IV (NHMRC 1995) (see table 8). This scale was similar to the earlier versions developed in Canada, but suggested the evidence generated by a systematic review was equal to that of a multi-centred RCT.

**Table 8**  
Quality of evidence ratings (NHMRC 1995, p. 39)

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. Dramatic results in uncontrolled experiments (such as results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.
Level IV	Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

Many other approaches to the grading of research evidence have been developed, and have used other rankings, such as 'IA, IB, II and No Recommendations' (CDC 1996), and combinations such as 'Grade A recommendations supported by Level I evidence' down to 'Grade C recommendation supported by level III, IV or V evidence' (Sackett 1986; Cook et al. 1992; Meltzer et al. 1998). Systematic reviews are increasingly replacing the RCT as the best source of evidence (NHMRC 1995). More recently, one hierarchy listed 'N of 1' randomised trial, where a single patient is randomly allocated the treatment and comparison intervention, as the highest level of evidence (Guyatt et al. 2000). Hierarchies have now been developed to address a range of other areas including prevention, diagnosis, prognosis, harm and economic analysis (Carruthers et al. 1993; Ball et al. 1998; Meltzer et al. 1998).

Ultimately, these hierarchies aim to provide a simple way by which to communicate a complex array of evidence generated by variety of research methods. From the perspective of the health care decision maker, they provide a measure of the trust that can be placed in the recommendations, or alert the user when caution is required. Yet, it could be argued that the interpretation of some of the hierarchies is more complex than the findings from primary research. Additionally, the optimal format and order of rank for research designs within these hierarchies has yet to be determined.

### **Ranking Different Types of Research**

A limitation of current hierarchies is that most focus solely on effectiveness. Effectiveness is concerned with whether an intervention works as intended. While this is obviously vital, the scope of any evaluation should be broader. For example, it is also important to know whether the intervention is appropriate for its recipient. From this perspective, the evidence on appropriateness would focus on the psychosocial aspects of the intervention, would address questions related to its impact on a person, its acceptability and whether it would be used by the consumer. A third dimension of evidence relates to whether the intervention can be used, and so relates to issues concerning the impact it would have on an organisation and the resources or support that would be required to ensure its successful implementation. The feasibility of an intervention encompasses the broader environmental issues related to implementation, cost and practice change.

Evidence on effectiveness, appropriateness and feasibility provides a sounder base for evaluating health care interventions, in that it acknowledges the many factors that can have an impact on success. This highlights the range of dimensions that evidence should address before health care interventions can adequately be appraised. This means that no matter how effective an intervention is, if it can not be adequately implemented, or is unacceptable to the consumer, its value is questionable. The risk with available hierarchies is that because of their narrow focus on effectiveness,

research methods that generate valid information on the appropriateness or feasibility of an intervention may be seen to be lower level evidence.

In response to the limitations of existing frameworks, a hierarchy of evidence was developed as part of this study that acknowledges the legitimate contribution of a range of research methodologies for the evaluation of health care interventions (see figure 4).

**Figure 4**  
Ranking of research evidence evaluating health care interventions.

	<b>Effectiveness</b>	<b>Appropriateness</b>	<b>Feasibility</b>
<b>Excellent</b>	<ul style="list-style-type: none"> <li>• Systematic review</li> <li>• Multi-centre studies</li> </ul>	<ul style="list-style-type: none"> <li>• Systematic review</li> <li>• Multi-centre studies</li> </ul>	<ul style="list-style-type: none"> <li>• Systematic review</li> <li>• Multi-centre studies</li> </ul>
<b>Good</b>	<ul style="list-style-type: none"> <li>• RCT</li> <li>• Observational studies</li> </ul>	<ul style="list-style-type: none"> <li>• RCT</li> <li>• Observational studies</li> <li>• Interpretive studies</li> </ul>	<ul style="list-style-type: none"> <li>• RCT</li> <li>• Observational studies</li> <li>• Interpretive studies</li> </ul>
<b>Fair</b>	<ul style="list-style-type: none"> <li>• Uncontrolled trials with dramatic results</li> <li>• Before and after studies</li> <li>• Non-randomised controlled trials</li> </ul>	<ul style="list-style-type: none"> <li>• Descriptive studies</li> <li>• Focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Descriptive studies</li> <li>• Action research</li> <li>• Before and after studies</li> <li>• Focus groups</li> </ul>
<b>Poor</b>	<ul style="list-style-type: none"> <li>• Descriptive studies</li> <li>• Case studies</li> <li>• Expert opinion</li> <li>• Studies of poor methodologic quality</li> </ul>	<ul style="list-style-type: none"> <li>• Expert opinion</li> <li>• Case studies</li> <li>• Studies of poor methodologic quality</li> </ul>	<ul style="list-style-type: none"> <li>• Expert opinion</li> <li>• Case studies</li> <li>• Studies of poor methodologic quality</li> </ul>

This approach addresses the multi-dimensional nature of the evidence and accepts that valid evidence can be generated by a range of different types of research. It does not attempt to diminish the value of the RCT or the importance of determining effectiveness, rather it accepts that the RCT answers only some of the questions. Importantly, this framework acknowledges the contribution of interpretive and observational research.

### Effectiveness

Effectiveness has been the most common concern of systematic reviews and clinical guidelines. Effectiveness relates to whether the intervention achieves the intended outcomes and so is concerned with issues such as:

- Does the intervention work?
- What are the benefits and harm?
- Who will it benefit or be harmed by its use?

It can be argued that multi-centred studies provide the best evidence for the effectiveness of an intervention because the results have been generated from a range of different populations, settings and circumstances. The findings from systematic reviews are generated in a similar manner and so also provide rigorous evidence (Mulrow 1987; Cook et al. 1998). As a result, the robustness and generalisability of evidence from both these approaches are better than those generated by other research designs. This means that for the evaluation of the effectiveness of an intervention, the best evidence would be that produced by both of these approaches.

However, this is not the only source of good quality evidence. A well conducted single centre RCT also produces results that are at low risk of error or bias, and so provides valid evidence of effectiveness. However, this evidence is ranked at a lower level than both the systematic review and multi-centre studies because its findings are based on a single population. As a result, factors unique to the study site, such as skill

mix, available resources, staffing levels or expertise, may have had an impact on the findings.

The place of observational studies, such as case control or cohort studies, within the hierarchy of research designs is less clear. Until recently they have been viewed as being at greater risk of systematic error than RCTs (Chalmers et al. 1983; Colditz et al. 1989; Miller et al. 1989). The concern with these studies is that it is believed they can distort the treatment effects, making them appear smaller or larger than they really are (Mulrow and Oxman 1997). Recently however, comparisons of the results of observational studies and RCTs evaluating the same intervention have questioned these claims (Benson and Hartz 2000; Concato et al. 2000). These investigations suggest that the findings of observational studies are similar to those produced by RCTs.

There are important differences between the RCT and the observational study relating to their internal and external validity. Internal validity in this context is a measure of how accurately differences in outcomes between comparison groups can be attributed to the intervention (Elwood 1998). External validity refers to the way in which the results of a study can be generalised to the wider population (Elwood 1998). The RCT minimises the risks posed by confounding variables through processes such as randomisation and strict inclusion criteria and as a result of this, the RCT has high internal validity. However, because of these very processes, only a narrow spectrum of patients may qualify for inclusion in the study. This means the external validity is low and so the generalisability of the findings of the RCT may be limited. Conversely, observational studies observe what is happening in practice and because of this have a lower internal validity due to potential differences between comparison groups. As a result of this, it is more difficult to attribute the outcome to the intervention. However, this lack of control means observational studies are firmly based in the real world, in that the comparison groups more closely reflect clinical practice. Therefore, it can be argued that observational studies have a higher external validity than RCTs.

Put more simply, the gains in the internal validity of the RCT are achieved at the expense of external validity, while the high external validity of the observational study is achieved at the expense of internal validity.

In some situations, observational studies may be more suitable than the RCT, such as when measuring infrequent adverse outcomes, evaluating interventions designed to prevent rare events or those evaluating long term outcomes (Black 1996). Legal or ethical issues may also prevent the conduct of RCTs. Observational studies may be the only option in situations where clinicians or patients are unwilling to accept randomisation as the mechanism for assignment of treatment (Horwitz et al. 1990). For some treatments, a sustained effort is required from the recipient and so their evaluation may require a different approach than the RCT (Brewin and Bradley 1989). Additionally, these studies cost less than RCTs and allow evaluation of a broader range of participants (Feinstein 1989). Finally situations in which the results of RCTs contradict consistent findings from observational studies serve to highlight the need for caution (Guyatt et al. 2000).

From this perspective it can be argued that both the RCT and the observational study can contribute important evidence related to the effectiveness of an intervention and therefore should have a role in any evaluation. The important difference between methods is that the RCT solely evaluates the intervention while the observational study measures the intervention in clinical practice. When differences in results exist, it cannot be assumed to be solely due to the presence or lack of randomisation (McKee et al. 1999). Factors such as differences in study populations, characteristics of the intervention or patient preferences may also be responsible for the differences in findings (McKee et al. 1999). These approaches can provide complementary evidence, however, end-users must be aware that both methods have their strengths and weaknesses (McKee et al. 1999).

In addition to the studies already discussed, evidence is also produced by other methods, however, their results are at greater risk of error (Dawson-Saunders 1994). With quasi-experimental designs, such as the non-randomised controlled trial, it is more difficult to show that differences in outcomes are the result of the intervention rather than because of differences between groups (Elwood 1998). These non-randomised studies differ from observational studies because the allocation to comparison groups is made by the researcher rather than health care workers who are independent of the study. For trials that use historical controls it is also difficult to demonstrate that changes to outcomes were the result of the intervention rather than changes over time. As a result of these factors, the risk of error or bias is high.

Uncontrolled trials may also be used to evaluate an intervention, but the lack of any comparison group makes interpretation of findings difficult. The exception to this would be studies with dramatic results, for example, as would be seen in the administration of oxygen to the hypoxic person or adrenaline to the person in shock. However, for most situations, the evidence generated by uncontrolled trials should be regarded with suspicion. As a result of these factors, evidence generated by these studies must be ranked at a lower level than the findings of RCTs or observational studies.

Finally, evidence about the effectiveness of an intervention may be generated through descriptive studies, expert opinion, case studies or poorly conducted studies. This evidence is at greatest risk of error and is inadequate for the evaluation of the effectiveness of an intervention. As a result of this, these methods provide the lowest level of evidence.

### Appropriateness

Appropriateness in the context of this hierarchy relates to the impact of the intervention from the perspective of the recipient. It also concerns the impact of



illness so that this information can be integrated into health care management and to assist in the prioritisation of care. Appropriateness is concerned more with the psychosocial aspects of care than with the physiological, and is reflected in questions such as:

- What is the experience of the consumer?
- What health issues are important to the consumer?
- Does the consumer view the outcomes as beneficial?

The range of research methods that can contribute valid evidence on the appropriateness of an intervention is broader than those addressing effectiveness. Firstly, as with effectiveness, results generated by multi-centred studies and systematic reviews would represent the best evidence on the appropriateness of an intervention. The systematic review need not be limited to synthesising the findings of RCTs, but should focus on all methods that can reasonably be used to evaluate the intervention from the perspective of appropriateness. Recommendations based on these sources of evidence would be at a low risk of error.

Good evidence for appropriateness can also be generated by a range of other research methods. As with effectiveness, a well conducted single centre RCT or observational study can also provide valid evidence about the appropriateness of an intervention. In this context, it would be through a focus on psychosocial outcome measures. As previously stated, while the experimental and observational studies evaluate the intervention from different perspectives, the evidence is complementary.

Interpretive studies can also contribute valid evidence in this area in that they represent the consumer's perspective of the treatment, illness or other such phenomenon. These interpretive studies help capture the subjective human experience that is often excluded from the experimental studies. This interpretive inquiry helps health care workers gain an understanding of everyday situations and the meaning of everyday experiences (Van Manen 1990; Van der Zalm 2000). While this information

differs considerably from that generated by experimental or observational research, it still contributes to our understanding of the impact of health care. However, despite this difference, the evidence generated by interpretive research is no less valid than that produced by other methods.

Evidence on appropriateness can also be generated by descriptive studies such as surveys, questionnaires and case studies. These studies contribute descriptive data related to interventions, their use and consumer responses. In addition to this, focus groups have emerged as an important method for gathering information on the feelings and opinions of small groups of people, and can aid in the evaluation of health care programs (Beaudin and Pelletier 1996). This information offers another perspective on appropriateness and as such is valid evidence. Finally, evidence can be generated by expert opinion or poor quality studies but is at greatest risk of error, and as with effectiveness, is ranked at the lowest level of the hierarchy of evidence.

### Feasibility

Feasibility addresses the broader environment in which the health care is situated and involves determining whether the intervention can and should be implemented. This focus acknowledges that the process of intentional change in large organisations is complex and involves a great deal more than modification in attitudes and behaviour (MacGuire 1990). In this context, feasibility is reflected in questions such as:

- What resources are required for implementation?
- Will it be accepted and used by health care workers?
- How should it be implemented?
- What are the economic implications of its use?

A broad range of research methods can reasonably be used to evaluate feasibility, and while each has a different focus, all offer important evidence. Again, results generated by multi-centred studies and systematic reviews should be considered the best

evidence for evaluating the feasibility of an intervention. The systematic review need not be limited to synthesising the findings of RCTs, but should focus on all methods that can reasonably be used to evaluate the intervention from the perspective of feasibility.

A well conducted single centre RCT can provide good evidence on the feasibility of an intervention. From this perspective, the RCT would be likely to focus on organisation, utilisation or implementation outcome measures, or on activities that support the intervention such as education programs. Both observational and interpretive studies can also generate valid evidence and would focus on issues such as implementation, acceptance, long term benefits, or the impact of the organisational culture on the implementation process.

Other approaches can provide useful evidence on feasibility. For example, action research is able to explore the relationships between attitudes and specific aspects of care, to identify barriers to practice change, and to systematically develop knowledge related to practice (Meyer 2000). As a result of this, action research can contribute legitimate evidence on which to influence and shape clinical practice. Action research studies have recently been summarised by a systematic review addressing factors influencing change in health care practice, highlighting the potential of this of this type of evidence (Meyer et al. 2000). As with appropriateness, focus groups can also gather valid information from small groups of people (Basche 1987; Beaudin and Pelletier 1996), and assist in the evaluation of health care programs (Robinson 1999). From the perspective of feasibility, this information would relate to such things as implementation, identifying barriers, determining health care priorities or determining what support is required. Descriptive studies can also provide information related to the feasibility of an intervention. However, because many issues unique to the study setting can influence the findings, this evidence would be ranked at a lower level than that produced from RCTs, observational studies and interpretive research.

Finally, as with both effectiveness and appropriateness, evidence can be based on expert opinion, case studies or poor quality studies. However, since this evidence is at greatest risk of error, it must be ranked at the lowest level of the hierarchy.

### **The Levels of Evidence**

The primary purpose of ranking research evidence is to provide an indication of its validity and its trustworthiness. This process assists in the selection of the best evidence to guide clinical practice. The description of each level follows.

- **Excellent:** This level of evidence provides the strongest scientific base for clinical practice. As this level of evidence is at the least risk of error it is optimal for the development of practice guidelines and clinical recommendations.
- **Good:** This level of evidence also provides a sound basis for clinical practice and is at low risk of error. However, as this evidence may be generated by single studies, it also highlights areas where replication of research is needed.
- **Fair:** This level of evidence will be at varying degrees of risk of error and so does not provide a strong evidence base for clinical practice. However, these studies represent initial exploration of interventions and health care programs and so assist in prioritising the research agenda. The rationale for this is that while the evidence is at greater risk of error than the previous levels, it allows identification of potentially beneficial interventions that require additional investigation and evaluation.
- **Poor:** This level of evidence provides a poor basis for clinical practice as it is at serious risk of error or bias. Additionally, while this level can help in determining research priorities, because this information may be wrong, it is ranked below that of other forms of evidence.

## Insufficient Evidence and Contradictory Findings

For some interventions there will be a lack of information, or the available evidence may be contradictory. Both these situations most commonly relate to small sample size, and therefore inadequate power of findings. This lack of power is often presented as a negative finding, that is studies suggest that the intervention does not produce any positive outcomes. As previously discussed, these negative findings have more to do with sample size than the intervention. In a similar manner, this lack of power can also result in contradictory findings between studies and so prevent the evaluation of the intervention.

In both these situations, determining the level of evidence is not justified and may lead to inappropriate confidence in the available evidence. Insufficient evidence or contradictory findings highlight the need for further investigation rather than serving as a guide for health care practices.

## A Guide

Finally, it must be acknowledged that the use of any hierarchy is, at best, a guide rather than a set of inflexible rules. A hierarchy provides the end-user of research with a framework to judge the strength of available evidence. Other issues, such as what outcome measures were used and which populations were studied also exert a major influence on the useability of the evidence. Finally, and most importantly, hierarchies cannot be used to rank evidence without some consideration of the quality of research. Regardless of the research method, if the processes used during the study were poor, then the findings must be regarded with suspicion. This hierarchy provides a guide to the evaluation of research, however factors such as research quality will also exert an important influence on the value of the available evidence.

## The Gold Standard

In the context of this hierarchy it could be argued that there are two interpretations of the label 'gold standard'. The common use of this term refers to the optimal research design to answer the question. Over the past decade, this label has most commonly been applied to RCTs evaluating effectiveness of interventions. However, as noted, this is being challenged in relation to observational studies. Also for research questions other than effectiveness, different methods will be needed. The optimal research method will be determined by the question, and it is the method that produces the most valid evidence that should become the standard to which others are compared.

Secondly, the use of this hierarchal structure for grading evidence provides another interpretation of what is meant by the gold standard. The concept of gold standard could move beyond research design and refer to evidence addressing all three dimensions of any evaluation of health care. That is, evidence demonstrates the intervention works, can be implemented and fulfils the needs of its consumers. Only when all these dimensions have been subject to investigation can an intervention be fully appraised. That evidence can then be considered to be of a gold standard.

## **Justification of This Approach**

The benefit of this approach for grading evidence evaluating interventions is that it enables reviewers to move beyond having a single focus on the RCT. This broader focus is important because the RCT is not able to answer all the questions needed for health care evaluation. From this perspective, it acknowledges that when evaluating an intervention a variety of research methods can contribute valid evidence. This hierarchy also recognises the greater strength of evidence when it has been generated from multiple populations, settings and circumstances. For this reason, evidence generated by properly conducted systematic reviews or multi-centred studies is considered to provide the strongest evidence.

This approach to ranking evidence also legitimises the perspective of the consumer of the intervention and so recognises the pivotal role consumers should have in health care decisions. This perspective is represented by the acknowledgment of the importance of the psychosocial impact of interventions and that the consumer's priorities of health needs may differ from that of the providers of the care. While the views of the consumer have long been part of the rhetoric, to date they have fitted poorly within the evidence-based framework. Through the use of this hierarchy evidence addressing these different aspects of the intervention can be ranked at a more appropriate level.

While an intervention may be effective, it must also be feasible to be able to be implemented in the health care setting. This aspect of the evidence relates to such things as cost, health care workers acceptance and what resources will be required to support the intervention. This evidence is particularly important given the rapid technological and pharmaceutical developments that have occurred in recent times, and as improvements to health are often small. This hierarchy recognises that evidence addressing the feasibility of an intervention is as important as evidence that addresses effectiveness. This broader approach to the ranking of evidence will provide a more robust scientific base for health care, in that it moves beyond the single dimension of effectiveness that has dominated the evidence based health care movement.

### **Expanded Review Process: Ranking the Evidence**

For the purposes of the expanded review process, the hierarchy of evidence was used to rank the findings of the systematic reviews. This ranking occurred after the findings of individual studies had been synthesised and recommendations developed. In accordance with the hierarchy, the evidence on which the recommendations were based was ranked as being either excellent, good, fair or poor.

Additionally, in some situations the available evidence was classified as being either insufficient or contradictory, thereby preventing the formulation of any recommendations.



## Data Collection

### Introduction

The data that are used during the analysis phase of the systematic review is the result of a number of independent studies. To analyse these data they must first be collected from the study reports. Like many other areas of the review, this brings with it the potential for error during the transcription and a number of strategies are used to minimise these risks. The aim of this phase of the review is to collect data that are reliable, valid and free from bias or error (Petitti 1994; Evans 2001). To this end, a data collection form is developed and used as part of the process. This data collection form is suggested to be the bridge between what is reported by the primary investigator and what is ultimately reported by the reviewer (Mulrow and Oxman 1997). The form has three purposes:

- it is directly linked to the formulated review question and planned assessment of included studies and so it provides a visual representation of these,
- it serves as the historical record of the multiple decisions that are made throughout the review, and
- it is the record of the data from which the analysis will emerge (Mulrow and Oxman 1997).

It has been suggested that a well developed form is more likely if the reviewer knows both the review topic and the integration methods (Stock 1994). The details that should be collected are those relevant to the key components of systematic review question and will vary depending on the subject and nature of the review. When determining what information is to be collected, consideration must be given to what analyses are to be made and what information will be required for the tables summarising the characteristics of identified studies.

There is a need to balance what data is collected as overly detailed collection of data can result in data collection forms that are longer than the reports they aim to summarise (Mulrow and Oxman 1997). Collection of unnecessary information also significantly increases the time needed to collect the data. Stock notes that if a specific piece of information adds five minutes to the data collection process, this results in an additional five hours work for every sixty papers included in the review (Stock 1994). Opposing this, if data collection forms contain insufficient data, the process will have to be repeated to obtain the missing information. However, it has also been suggested that to require formal justification for each item of information may restrict creative conjecture and hunches about the domain (Stock 1994). The data collected must therefore include only that relevant to the review, but this relevance must also be interpreted carefully.

To aid data collection, the sequence of the form should be logically related to the sequence of scientific publications (Petitti 1994). Stock recommends that the data collection form should mirror the order of information in the study reports (Stock 1994). Woodworth suggests that a poorly designed data collection form impedes data entry and greatly increase data entry errors (Woodworth 1994). He suggests that data items should be difficult to overlook, either because they are placed in a consistent predictable position, or are boxed or highlighted. While there are many potential approaches to the layout and structure of the data collection form, the type of information that will be required includes:

- Report Identification - items should include author, year and publication.
- Setting - addresses the general conditions of the study and items may include sampling (such as local, regional or national), special populations (such as hospital patients or nursing home residents), and characteristics of the organisation, community or business.
- Subjects - includes specific items about sample and sub-groups of subjects, demographic features, and any other relevant information to describe the subjects.

- Methodology - describes the research design, details of sampling assignment to groups, attrition and other relevant information on the conduct of the study and the collection of data.
- Treatment - includes specific components of treatment, the nature of the control group, duration of treatment and mode of delivery.
- Outcome Measures - items that describe the nature of outcome measures.
- Results - information related to the findings of the study (Stock 1994; NHS Centre for Reviews and Dissemination 1996; Mulrow and Oxman 1997).

Complete documentation of outcome data is necessary, however this is often difficult because of issues such as the use of multiple outcomes, the reporting of outcome data at different points in time, or the use of different measurement tools (Mulrow and Oxman 1997). Whenever possible, the aim should be to extract raw patient numbers, free from any manipulation or transformation. The data collection form should be pilot tested and revised as necessary prior to commencement (Orwin 1994; Mulrow and Oxman 1997). It has been suggested that this process provides the reviewer with first hand experience of the actual collection of data, tests the definitions and descriptions that are to be used during the collection process, identifies deficiencies in the form, and helps identify whether there is a need for additional categories (Orwin 1994).

To minimise the likelihood of error during transcription, double data collection is sometimes undertaken (Stock 1994; NHS Centre for Reviews and Dissemination 1996; Mulrow and Oxman 1997). This involves the independent collection of data by two reviewers, then comparison of the results for discrepancies. However, Orwin suggests that double entering may only actually be done for some of the studies, to provide a check on the accuracy of the extraction while limiting the impact on the time taken (Orwin 1994). Blinded data extraction is also sometimes used to help avoid the journal or author from influencing the data collection process (Petitti 1994; Mulrow and Oxman 1997). The basis for this is that collecting data from reports published in

prestigious journals or by internationally known researchers may influence the reviewer to devote greater attention in seeking out the relevant information and data than may be done for less prestigious reports (Petitti 1994). Finally, it has been suggested that knowledge of the results of a study may influence or bias a reviewer during the collection of data and so a separate form for study method and results has also been suggested (Petitti 1994). However, decisions related to the use of these approaches must be balanced against the fact that each additional activity, quality check or revision, further lengthens the process. Ultimately, like all other stages of the systematic review, the methods used to collect data from studies are a balance between rigour and practicality.

### **Missing Information and Results**

Problems can be encountered during the conduct of the review when studies fail to report relevant data or adequate description of methods (Pigott 1994). This missing data reduces the size of the sample of studies available for synthesis. This means that the remaining studies, that reported complete information, may no longer be representative of the population of all identified studies (Pigott 1994). Pigott also notes that in reviews involving a number of different outcomes, the problem of missing data may mean that the synthesis of each of these differing outcomes may each utilise a different set of studies (Pigott 1994). This may limit the generalisability of the findings, and may also increase the risk of error.

It has been suggested that virtually all write-ups of studies will report some information poorly, but at times some reports will be so vague that they obscure what took place entirely (Orwin 1994). The type of information commonly missing from reports includes;

- lack of information about subjects,
- vague or limited information regarding study method,
- lack of information about treatment regimens or comparison,

- only the test statistic is reported,
- only the p-value is given,
- total number of participants in each group is not provided, and
- the results are presented graphically.

Missing data may be the result of the researchers own preferences and writing style, but it may also be a result of restrictions related to publication word limits in journals. Word limits on abstracts and text mean that only information central to the study will be included. Some data may not be reported because of differences in reporting practices in the researchers related field (Pigott 1994). That is, each research will report the outcomes and data that is of importance to their discipline. However, data may also be missing because of the actual value of the outcome measure itself (Pigott 1994). For example, omitting to report small effect sizes because they represent insignificant differences between groups. This censoring of data is difficult to detect and harder to resolve than other types of missing data.

Contacting the authors may sometimes be useful in obtaining the missing data, although it can be a labor-intensive task if large numbers of studies are involved. However, the success of contacting the authors is influenced by a variety of factors including whether;

- the investigators are still alive,
- the investigators can be located,
- they still have the research information available, and
- they are willing and able to provide the information (Orwin 1994).

Occasionally, data reported in the abstract differs from that reported in the text of the paper. This poses difficulties for the reviewer, in terms of which data should be used. Once again, attempting to contact the authors may be only option.

As the aim of this study was to develop systematic review methods that could be used to summarise a range of different types of research, these methods must also enable the collection of data from this range of studies. The collection of data from experimental and observational studies is similar, in that both report numerical data. However, the findings of interpretive research is narrative and so different methods will be required to ensure the information collected is correct.

### **Data Collection from Interpretive Studies**

The aim of this phase of the review is to collect data from studies that is reliable, valid and free from bias. The data collection phase is the precursor to the data analysis / synthesis phase of the review. When all the results have been collected it is then analysed and synthesised. However during reviews of interpretive research the data collection phase is less clear. Unlike data collection during reviews of experimental, observational or descriptive studies, data collection from interpretive studies is generally reported as being part of the synthesis process (Noblit and Hare 1988; Jensen 1994; Paterson et al. 1998). This data collection occurs during the synthesis as studies are read and re-read. The findings from individual studies are treated more as a whole during initial readings to enable identification of major themes and categories, and to compare and contrast the themes of one study to those generated by other studies. This represents a significant change to the approach used during data collection from other study designs.

During an analysis of wellness-illness research, Jensen and Allen read and re-read the text of the articles to standardise the data through the use of common codes, outlines and reporting formats (Jensen 1994). Ogden-Burke collected each study's relevant findings (Ogden-Burke et al. 1998), while Thorne and Paterson collected evidence of the assumptions, preconceptions and presuppositions held by the researchers

(Thorne and Paterson 1998). In both these latter two reviews the collection of data was part of the synthesis process. However, clear descriptions of how reviewers track study characteristics, themes and metaphors is often not stated, and for some topics the management of the many themes that emerged from multiple studies would be a significant undertaking. For example, one review of interpretive studies addressing chronic illness identified over 400 studies of which 158 met the review inclusion criteria (Thorne and Paterson 1998). Clearly, some topics have a large body of research, and dealing with this volume of information would be impossible without a formalised process of collection and recording of relevant data and study characteristics. However, this component of these reviews has received little attention (Evans and Pearson 2001a).

While approaches appear to differ, some reviews of interpretive research have used formal data collection sheets to aid in the management of studies and record of data for the synthesis. Sherwood describes a data collection tool to record study demographics such as the study population, number of participants, geographic region, time frame, the setting and age of participants (Sherwood 1999). Sherwood also recommends the use of tables to summarise bibliographic information and other essential aspects, to help keep track of the large amount of information that must be managed. This tool also helps identify potential comparisons. Jensen and Allen recommend listing key metaphors, phrases, concepts, ideas and categories, although this is part of the synthesis process rather than a distinct phase of the review (Jensen and Allen 1996). Additionally, some reviews provide summaries of study characteristics in the form of tables (Jensen 1994; Sherwood 1997; Ogden-Burke et al. 1998), suggesting that specific data was collected at some stage during the review process. It could also be argued that this listing of study characteristics represents the data trail that is vital for primary interpretive studies and so should be part of systematic reviews. While the use of summary tables provides important information, it is not provided in all reviews.

This discussion highlights that some data must be collected from interpretive studies. This would include information such as population, gender, age, geographical location, time frame and study method. However, as the number of studies to be included in the review increases, the collection and management of this data will become increasingly difficult. This suggests that, like other reviews, a data collection tool is also needed during the review. This tool will aid in the management of the large amounts of data generated by these reviews and will also provide the basis for developing the summary tables of study characteristics. However, the main distinction between collecting data from RCTs and interpretive studies, is the stage of the review when data is collected. For reviews of RCTs, data collection precedes the synthesis phase, while with interpretive studies some of this data may be collected as part of the synthesis.

#### **Expanded Review Process: Data Collection**

In terms of the expanded review process, a data collection tool was developed as part of the systematic review protocol and this tool was used for all studies regardless of the research design used (see appendix 3). This tool provided the basis for developing the narrative and tabular summary. The use of this tool also helped minimise the risk of errors during the transcription process and served as a permanent record of data. As a result, it is argued that this tool helps maintain the rigour of the expanded systematic review.

Because the expanded review required data from a variety of studies, the actual data collection tool had to be adapted according to the nature of the data. The numerical data expected from experimental and observational studies, and some descriptive studies, required a very precise form. However for interpretive studies, the collection of narrative data did not permit a data collection tool to be as prescriptive as that used for numerical data. For these studies, demographic data was collected as is done for all studies, then major themes and categories were listed.



## **Data Synthesis**

### **Quantitative Data Synthesis**

#### Meta-analysis

Meta-analysis is a quantitative approach used to systematically combine the results of studies to produce a conclusion about a body of research (Petitti 1994). If used appropriately, meta-analysis can allow conclusions to be derived from the data and help minimise the risk of errors during interpretation (Mulrow and Oxman 1997). The aims of meta-analysis are to provide an estimate of the average effectiveness of an intervention, to investigate whether the effect is roughly the same in different studies, settings and participants, and if the effect is not the same, to investigate the differences (NHS Centre for Reviews and Dissemination 1996). The effect of an intervention is usually measured according to a change in an event rate or a change on a continuous scale (NHS Centre for Reviews and Dissemination 1996). This means that events such as mortality, or continuous data such as blood pressure, generated by individual studies can be pooled using a range of different methods. It is this pooling of results that is the basis of meta-analysis (Evans 2001).

Meta-analysis can only be undertaken when studies address the same question, use a similar population, administer the intervention in a similar manner, measure the same outcomes for all participants, and use the same research design (Jones and Evans 2000). When studies differ in one or more of these components, the use of meta-analysis is not appropriate (NHMRC 1999). Meta-analysis is also likely to be inappropriate when data are sparse, the data can not be obtained from studies, or when studies are too heterogeneous to sensibly combine (NHS Centre for Reviews and Dissemination 1996). Meta-analysis will also be inappropriate when critical appraisal of individual studies suggests that because of the methods used, the results must be interpreted with caution. In these situations, discussion of study findings,

differences in the methods used or the populations studied will be more appropriate. While meta-analysis is commonly used to synthesise the findings of RCTs, it can also be used for observational research such as cohort, case control and cross sectional studies.

### Comparisons

The comparisons that are made during meta-analysis are those that relate clearly and directly to the review question or hypothesis. These comparisons are determined as part of the process of developing the review protocol and provide the basis for determining the effectiveness of interventions. However at times it may be necessary to change or modify the comparisons because it may not be possible to determine what can be sensibly combined until after the data has been collected (Mulrow and Oxman 1997).

A difficulty in deciding what comparisons to make is the fact that there may be multiple outcome measures. These multiple measures can be the result of different outcome measurements, different sub-samples of people, or different times of measurement (Lipsey 1994). Many interventions have a range of potential measures of effectiveness. For example, evaluations of music therapy could focus on satisfaction, anxiety or tolerance of unpleasant procedures. In addition to reporting overall results, in some studies the results of specific sub-samples may also be reported. For example, while music therapy has been used for the broad population of hospital patients, it has also been used in more specific populations such as cardiac or pre-operative patients. Finally, many studies report findings of the same measure over time. For example, a review of management of oral mucositis found different studies measured the severity of mucositis at differing times, such as daily or only a few times each week (Kowanko et al. 1998). To effectively represent these multiple measures, sub-samples and time frames, it is necessary to establish conceptual categories for each of the dimensions (Lipsey 1994). These categories will group

components that are substantially similar, differentiate those that are importantly different, and ignore those that are irrelevant or not of interest to the review (Lipsey 1994). The comparisons that are finally made in the meta-analysis will be influenced by the data presented in the study reports. Some studies will contribute information to only a single category while others will contribute to many. As previously stated, this grouping of comparisons starts during development of the review protocol.

### Approaches to Meta-analysis

When it is appropriate that study results can be pooled, there is a range of possible outcome measures that can be utilised. In the past, methods such as vote counting or combining significance levels were used, although their value has been questioned. During vote counting, studies are tallied into one of three categories; positive relationship, negative relationship or no specific relationship (Sutton et al. 1998). However this method has been criticised because it fails to take into account the size of studies, does not provide an estimate of the size of the treatment effect and because it has low power it may fail to identify small to medium effects of treatments (Bushman 1994). Similarly, p-values have been used to pool studies but have been criticised as a result of their misuse, the misinterpretation of p-values, and the fact that the method used to combine p-values can influence results (Becker 1994; Sutton et al. 1998).

The choice of the measure of treatment effect depends on many things, the most important of which is the whether the measure is a continuous or dichotomous scale (NHMRC 2000). Common methods used to pool data on these scales include:

#### 1. Continuous Outcome Data

- Weighted Mean Difference

Averaged (pooled) difference between treatment and control groups in mean values across studies using the same scale of measurement, for example blood pressure in mmHg (NHMRC 1999). With this method

each study estimate of effect is given a weight directly proportional to its precision (inversely proportional to its variance) (Sutton et al. 1998). This method has been recommended for use whenever outcomes are measured in a standard way across studies (Mulrow and Oxman 1997). A weighted mean difference of 0 means there is no difference between groups.

- Standardised Mean Difference

In this approach differences between the means of the treatment and control group for each study are standardised using an estimate of the standard deviation of the measurements (NHMRC 1999). Because this approach uses the standard deviation, it can be used to pool estimates that have been measured on different scales. However, care is needed to ensure the pooling of these different scales is appropriate (Mulrow and Oxman 1997). A standardised mean difference of 0 means there is no difference between groups.

## 2. Dichotomous Outcome Data

- Risk

Many of the measures used in expressing the findings of studies that report dichotomous outcomes use risk. Risk expresses the frequency of a given outcome (McQuay and Moore 1997). They are probabilities, and so can vary from 0.0 to 1.0 (also expressed as 0% to 100%).

- Relative Risk (RR)

Relative risk is the ratio of the risk of a given outcome in the treatment group relative to the control group (McQuay and Moore 1997). A RR below 1 suggests the treatment reduces the risk of the event occurring (NHMRC 1999). A RR of 0 to 1 represents a lower risk in the population, and greater than 1 represents a higher risk (Fleiss 1994). For example, if 5% of people in the treatment group had the outcome

(such as migraine) compared to 30% in the control, relative risk is 0.17 ( $0.05 \div 0.30 = 0.17$ ) (McQuay and Moore 1997).

- Relative Risk Reduction (RRR)

The comparison of risk between groups can also be expressed in terms of its reduction. Relative risk reduction is the ratio between the decrease in risk (in the treatment group) to the risk in the control group (McQuay and Moore 1997). Using the migraine example given above, RRR is 0.83 ( $0.25 \div 0.30 = 0.83$ ). However, RRR can be misleading because it is influenced by the prevalence of the event. For example, a 25% reduction in relative risk of death is more compelling if it involves a reduction in mortality from 40 deaths per 100 to 30 deaths per 100, than a reduction from 4 deaths per 100 to 3 deaths per 100 (Wilson et al. 1995). So while relative risk is useful, it can be misleading if baseline risks are not reported (Wilson et al. 1995).

- Absolute Risk Reduction (ARR)

Absolute risk reduction is the absolute difference in the risk of a given outcome between the treatment and control. It is determined by subtracting the risk in the treatment group from that of the control group (McQuay and Moore 1997). Using the migraine example, the absolute risk is 25% ( $0.30 - 0.05 = 0.25$ ). The advantage of ARR is that it reflects the frequency of the outcome, and that for rare events the ARR would be very small.

- Odds Ratio (OR)

The odds of an event occurring can be calculated for a population by dividing the number of events by the number of non-events. For example, if 22 events occur in a population of 100, the risk is 0.22, and the odds of that event occurring is 0.28 ( $22 \div 78 = 0.28$ ) (McQuay and Moore 1997). The odds ratio is calculated by dividing the number of events that occur in the treatment group by the number occurring in the

control group (McQuay and Moore 1997). Like in the other outcomes, an odds ratio of less than 1 for an adverse event indicates the treatment reduces the risk. The odds ratio can be determined using data collected from both experimental and observational studies (Fleiss 1994; Davies et al. 1998; NHMRC 2000).

- Number Needed to Treat (NNT)

NNT is the number of patients who have to be treated to prevent one event occurring ( $NNT=1/ARR$ ) (NHMRC 1999). The NNT is calculated using the baseline risk for a specific outcome without treatment and the reduction in risk achieved with treatment (Smeeth et al. 1999). NNT are specific to a follow-up period, and so should only be calculated during the pooling of studies that used a similar follow-up (Mulrow and Oxman 1997). To be complete, NNT must specify the comparison, the outcome of interest and the duration of treatment to achieve the outcome (McQuay and Moore 1997). NNT is increasingly being calculated by pooling absolute risk differences in trials included in meta-analyses (Smeeth et al. 1999). In the migraine example, the ARR was 0.25 giving a NNT of 4 ( $1 \div 0.25 = 4$ ). As the baseline risk decreases, the NNT increases, when the NNT approaches 1 it means that a favourable outcome will occur in nearly every person receiving the treatment (McQuay and Moore 1997; NHMRC 2000). It is inappropriate to compare NNT across diseases. For example, a NNT of 30 for preventing deep vein thrombosis may be valued quite differently from a NNT of 30 for preventing death (McQuay and Moore 1997). When the outcome is an adverse event, then this summary measure is referred to as the Number Needed to Harm (NNH) (McQuay and Moore 1997). In some situations both NNT and NNH may be used to provide a method to quantify both the potential benefits and harm that accompany a specific treatment.

There are different approaches to combining data, however there is also disagreement on which is the best approach. The primary disagreement in meta-analysis is whether to incorporate between study variation, or heterogeneity, in the analysis (Mulrow and Oxman 1997). The two categories of summary estimates are the fixed and the random effects models (NHMRC 1999).

- Fixed Effect Model: Assumes there is a single 'true' value which all studies are attempts to measure (NHMRC 1999). The fixed effect summary is a weighted average with weights proportional only to each study's precision (NHMRC 1999).
- Random Effects Model: Assumes the 'true' value varies and therefore attempts to incorporate this variation into the weightings (NHMRC 1999). This model estimates the underlying study to study variation, which is then included in the weighting for each study. However, when significant heterogeneity exists, it is still necessary to examine studies to determine the reasons for the lack of homogeneity, as the random effects method does not correct for bias, or failure to control confounding (Petitti 1994).

If there is little between-study variation, as represented by a large p-value in the test for heterogeneity (addressed in next section), then the choice of random or fixed effects model will have little impact on the results (Mulrow and Oxman 1997). When there is considerable between-study variation using the fixed effects model, which ignores this variation, there will be a narrower confidence interval than the random effects approach which does not ignore the variation (Mulrow and Oxman 1997).

For many situations it may be appropriate to use both approaches, and both methods should basically agree when there is no heterogeneity between studies (Mulrow and Oxman 1997). The random effects model will tend to give a more conservative estimate and this is reflected in the wider confidence interval of the analysis.

However, where significant heterogeneity exists, then the appropriateness of undertaking a meta-analysis must be questioned.

### Investigating Heterogeneity

When carrying out a meta-analysis, it can be expected that the estimate of the effect size will differ to some degree between studies (Sutton et al. 1998). This can be expected and is partly due to the fact that the true effect size of the intervention will also vary to some extent from one population to another (Sutton et al. 1998). However this variation can also occur as a result of other factors, and it is therefore important to determine if the results in each study are similar enough to pool (Mulrow and Oxman 1997). If differences between results are greater than could be anticipated by chance, other factors related to differences in populations or the intervention or outcome measurement may be responsible. Additionally, if results are consistent across studies despite variation in populations and the methods used, this suggests the results are robust and transferable to other practice settings (NHMRC 1999).

The chi square statistic is commonly used to test for homogeneity (Mulrow and Oxman 1997; Sutton et al. 1998). With this technique the more significant the results, as demonstrated by a p-value of less than 0.1, the more likely it is that the observed differences were not due to chance alone (Mulrow and Oxman 1997). A suggested 'rule of thumb' to evaluate heterogeneity is that if the confidence interval of any study does not overlap the confidence interval of the summary estimate (confidence interval of the meta-analysis), it is likely that statistically significant heterogeneity exists (Mulrow and Oxman 1997; NHMRC 1999). In the presence of heterogeneity between studies, caution must be used in generalising the pooled results.

Testing for heterogeneity between observational studies is more difficult, because observational studies have different risks than RCT, and so the homogeneity



assumption is extremely unlikely to be satisfied (Greenland 1987). It has been suggested that the issue is more that the existing heterogeneity is small enough to be reasonably ignored (Greenland 1987). Heterogeneity in this situation is a result of factors such as:

- different observational designs utilised,
- differences in studies using the same study design,
- definitions of outcomes differ across studies,
- variation in populations between studies,
- differences in exposure levels between and within studies,
- different definitions of exposure levels,
- exposure measures involving surrogate measures, and
- differences in duration, intensity and frequency of exposure (Sutton et al. 1998).

While methods for testing for heterogeneity between RCTs can also be used for observational studies, dealing with this problem will be more difficult. As heterogeneity can be anticipated, testing will often simply show what was already known. Therefore an important consideration in the decision regarding whether meta-analysis should be undertaken is whether the studies can reasonably be combined, and whether the results of the synthesis are meaningful.

### Subgroup Analyses

Examination of specific subgroups is a common component of meta-analysis. Subgroup analysis is when the meta-analysis is conducted using only data generated from specific subgroups of the total population. However, caution is needed because the results can be misleading (Mulrow and Oxman 1997). It has been suggested that the risk with pooling of study results in this way is that ineffective treatments may be supported (false positive), effective treatments may not be supported (false negative), or the analysis may result in misleading recommendations about the direction of future

research needs (Mulrow and Oxman 1997). However, this type of analysis can provide important information about the effect of an intervention in specific populations. In deciding whether a subgroup analysis is appropriate, issues to be considered include:

- subgroup analyses should be kept to a minimum,
- it should be plausible to expect a difference between the subgroups,
- the hypothesis should precede rather than follow the analysis,
- the more analyses that are undertaken, the greater the likelihood of detecting a difference by chance alone, and
- conclusions are strengthened if the difference is observed across a number of studies (Mulrow and Oxman 1997).

To limit the risks associated with subgroup analyses, intended comparisons should be documented in the protocol.

### Sensitivity Analysis

Sensitivity analysis involves repeating the meta-analysis while making small changes to the data (NHS Centre for Reviews and Dissemination 1996). The process allows the reviewer to evaluate the impact of decisions made during the review process, and to determine their impact on the results of the meta-analysis. These decisions may relate to uncertainty or disagreements regarding the inclusion of certain studies, data extraction, missing data and in the statistical methods utilised (NHS Centre for Reviews and Dissemination 1996). For example, the sensitivity analysis allows the reviewer to repeat the meta-analysis and to incorporate those studies excluded during the critical appraisal to determine if these decisions have the potential to change the results. Other issues such as the impact of blinding on results or different outcome measurements can also be addressed during the sensitivity analysis. If results are not changed as a result of the sensitivity analysis, it suggests the decisions are not critical.

## Presentation of Results

The statistical method used to pool studies and the method used to present results can be considered separate issues (Mulrow and Oxman 1997). Display of the results of the meta-analysis should be by both tables and graphs. The graphic presentation enables easy interpretation of findings, and also allows the findings of individual studies to be compared to other studies or to the final summary of all studies. Presentation of data in tables allows specific results to be viewed and represents part of the data trail of the systematic review. Results are commonly presented using a forrest plot (also known as a Cochrane plot), with point estimate and 95% confidence interval (Sutton et al. 1998). However, a suggested disadvantage with this presentation is that the attention is drawn to the least significant studies because they have the widest confidence interval and are graphically more imposing (Sutton et al. 1998).

Meta-analysis can be used to pool the findings of studies that report numerical data derived from the comparison of interventions. However, when the studies to be pooled reported narrative data, meta-analysis can not be used. Therefore to pool the results of interpretive or descriptive studies, other methods are needed.

## **Qualitative Data Synthesis**

The increase in quantitative research has been mirrored by a similar increase in the number of interpretive studies. Yet, these isolated studies do not in themselves contribute significantly to our understanding of phenomena of interest (Jensen and Allen 1996). As knowledge has been viewed as accumulative, with each new work adding to existing knowledge, there is a growing interest in developing approaches to integrate these many independent interpretive studies. However, there are major differences between the integration of experimental and interpretive research. Reality

for the interpretive researcher, and therefore the reviewer, is viewed as multiple and constructed (Sandelowski 1993). Because of this, it has been suggested that no two reviews of interpretive studies will produce exactly the same results (Jensen and Allen 1996). While this approach to the synthesis of studies provides just one interpretation, it is argued that the aim is to capture the essence of the phenomenon (Sherwood 1999). The strength of this integration process is that it provides an understanding that is based on a range of populations, settings and circumstances (Evans and Pearson 2001a). This broad base for generation of evidence on a phenomenon allows for greater confidence in the evidence and increases its transferability to other settings.

In examining published reviews that included interpretive research, a number utilised meta-synthesis (Jensen 1994; Sherwood 1997; Ogden-Burke et al. 1998), while others use less formalised methods. In some reviews, details were provided about the methods used to locate, select and appraise studies, but little or no information about the methods used to synthesise data (Chapple and Rogers 1999; O'Neil and Morrow 2001; Woodward and Webb 2001). In one review the synthesis of data was described as content analysis, with the studies coded into predetermined categories (Suikkala and Leino-Kilpi 2001). Another reviewer collated major themes, then visually searched the list for key themes (Neill 2000). A review by Barroso and Powell-Cope described their synthesis process as a constant comparative analysis (Barroso and Powell-Cope 2000), and another used a continuous comparison approach derived from grounded theory methods (Kylma and Vehvilainen-Julkunen 1997). During the conduct of one review, NUD\*IST data handling software was used to aid in the coding of data from individual studies (Lemmer et al. 1999). Some reviews included both quantitative and qualitative studies (Jensen 1994; Neill 2000; Suikkala and Leino-Kilpi 2001). These examples serve to highlight the current variation in methods used to synthesise the findings of interpretive research.

## Meta-ethnography

Noblit and Hare proposed a framework for the summary of qualitative studies in 1988 termed meta-ethnography, as an analogy to meta-analysis (Noblit and Hare 1988). This term was proposed to highlight meta-ethnography as an interpretive alternative to research synthesis, and referred to the translation of interpretive studies into one another (Noblit and Hare 1988). They suggested that meta-ethnography was an attempt to develop an inductive and interpretive form of knowledge synthesis, and so provide a rigorous procedure for deriving substantive interpretations about a set of ethnographic or interpretive studies (Noblit and Hare 1988). The aims of meta-ethnography are to:

- produce more interpretive literature reviews,
- critically examine multiple accounts of an event or situation,
- systematically compare case studies to draw conclusions,
- provide a way of talking about a specific work and comparing it to the work of others, and
- synthesise ethnographic studies (Noblit and Hare 1988).

Meta-ethnography synthesises rather than aggregates the findings of individual studies and involves the translation of one study into another (Noblit and Hare 1988). This translation is not a 'word for word' translation, rather it focuses on the 'meaning of the text', the interpretations and explanations rather than the interviews and observations (Noblit and Hare 1988). The methods of meta-ethnography proposed by Noblit and Hare are described below (Noblit and Hare 1988).

- Reading the studies: Repeated reading of the accounts and the noting of interpretive metaphors, paying attention to the details of accounts and what they tell of the substantive concerns. This is not a clear and distinct phase, but continues throughout the synthesis.
- Determining how studies relate: This phase involves putting together the studies and determining the relationships between studies. This entails

creating a list of key metaphors, phrases, ideas and concepts used in each account. By the end of this phase the initial assumptions about the relationships between studies can be made.

- Translating studies into one another: This phase involves comparing both the metaphors or concepts and their interactions of one account with those of other accounts.
- Synthesising translations: The synthesis of translations can be on two different levels. The first level is that of the initial translation that occurs during the initial reading of studies. The second level occurs when there is a large number of studies and the number of translations are therefore also numerous. In this situation a second level of synthesis occurs by analysing types of competing interpretations and translating them into each other.
- Expressing synthesis: This phase involves communicating the findings of the synthesis.

However, there appears to have been very little critical discussion on this proposed technique for synthesising interpretive research. Additionally, the validity of the findings have yet to be determined. The major role of meta-ethnography has been as the precursor of meta-synthesis, a term more commonly used in the nursing literature.

### Meta-Synthesis

The term meta-synthesis has been used in the nursing literature, and while it shares many similarities to meta-ethnography, it also incorporates some of the components of the systematic review process (Evans and Pearson 2001a). Meta-synthesis has been described as a framework for the synthesis of multiple non-experimental studies relating to a phenomenon of interest. Sherwood describes it as the critical review and analysis of merged data (Sherwood 1999). That is, meta-synthesis focuses on the themes or descriptions from multiple interpretive studies rather than the numerical data produced by experimental and observational studies (Sherwood 1999). The

suggested benefit of this approach is that, because the findings have been generated by multiple studies involving a broader sample creating a composite of descriptions of the phenomenon, it increases the validity and power of the results (Sherwood 1999). The greater degree of abstraction from the inductive process of comparison and synthesis produces results that are more generalisable to nursing practice (Sherwood 1999).

The processes used during meta-synthesis are much the same as those proposed by Noblit and Hare. Studies are examined for homogeneity in terms of characteristics and circumstances of participants. Major results, concepts and propositions from each study are noted to facilitate meta-synthesis. This synthesis involves bringing together the data generated during the reduction of qualitative studies to their key components. The challenge during the synthesis process is to portray individual constructs of the phenomena accurately (Sherwood 1999). Studies are sorted by methodology, data collection instrument and by key metaphors, phrases, ideas and concepts (Jensen and Allen 1996; Sherwood 1999). It is suggested that different research methods, such as phenomenology, ethnography or grounded theory are not mixed in a single synthesis of all interpretive studies as it will be unclear what has been obtained from the synthesis. (Jensen and Allen 1996). Differences between studies are not merged, rather they are compared and contrasted and areas of commonality are identified as part of the synthesis (Paterson et al. 1998). This process allows data to be reduced to common denominators, as is done during meta-analysis (Sherwood 1999). That is, the studies and their results are grouped and categorised into areas of similarity. This similarity is in terms of study characteristics and results. Like in meta-analysis, differences between studies or their results are not merged.

The synthesis involves the translation of terms and metaphors of one study into those of another and focuses on meaning rather than a word for word translation. These processes are similar to those used during the original analysis by the primary researcher. Key metaphors, phrases, ideas or categories are listed and compared

among studies (Jensen and Allen 1996). When translating a study into another, the central metaphors of each account are maintained in relation to key metaphors in others (Sherwood 1999). The process continues with the clusters of metaphors becoming progressively more refined until a description of the phenomenon is achieved. This synthesised data is then analysed to identify common themes or specific descriptions from individual studies. Relationships are examined for key phrases and explanatory themes. This process results in a progressive refinement of the understanding of the phenomenon through a process of constant comparative analysis. The endpoint is the generation of a new construction on which there is consensus (Jensen and Allen 1996).

During a review on the experiences of parents of an acutely ill child, the reviewers described their approach to synthesis which is presented here as a practical example of the process (Neill 2000, p. 823).

- Preparation
  - key findings recorded on index card
- Data analysis
  - re-reading of most relevant papers to develop sense of the 'whole'
  - collation of main findings into single file
  - visual search for key themes
  - categorising themes using coloured pens
  - findings for each theme are collated maintaining reference to the original source
  - from these categories sub-themes are identified
- Data interpretation
  - themes and sub-themes re-examined to interpret content of each theme, and to identify consistencies and incongruities in the research reviewed



- each section written up referring back to the original source to check accuracy and to add quotes or additional detail.

In summary, based on this discussion it can be argued that there is little difference between the methods of data analysis used by the primary researcher and the reviewer, in that both entail a constant comparative analysis of content to identify recurring themes.

The major difference between the researcher and the reviewer is that one uses primary data, the other processed data. Compared to the discussions surrounding meta-analysis, qualitative data synthesis has seen only limited debate. While a small number of qualitative reviews using meta-synthesis have been published, they have generally been accepted uncritically by the nursing profession.

### Content Analysis

Content analysis has also been used as a method for synthesising interpretive research (Suikkala and Leino-Kilpi 2001). Content analysis has been described as a means to obtain simple descriptions of data (Cavanagh 1997), and as a means to systematically and objectively describe and quantify phenomena (Downe-Wambolt 1992). Nandy and Sarvel note that it is an established research tool and is used to gain knowledge, new insights and a representation of facts (Nandy and Sarvela 1997). Content analysis has been used to describe a large variety of topics in a number of different mediums, such as music videos (DuRant et al. 1997), women's magazines (Hill and Radimer 1996), professional journals (Armstrong and Standsfield 1996), and advertising (Pratt and Pratt 1995).

The descriptions of phenomena are achieved through the analysis of the meanings of words and phrases into fewer content-related categories (Cavanagh 1997). The

processes of content analysis have been described as a series of stages as follows (Cavanagh 1997; Nandy and Sarvela 1997).

- Selecting the unit of analysis:

The objects that is to be studied must be described, and could include such things as articles, documents, radio or television.

- Sampling:

It is necessary to determine which units of analysis (documents) will form the sample for the content analysis.

- Creating and defining categories:

To analyse the documents the coding rules must be defined to provide a means of describing the phenomena of interest.

- Pilot testing:

To gain an understanding of the data to be analysed and to ensure the coding system is unambiguous, a pilot test is conducted. The coding categories are then revised as needed and the pilot test may be repeated.

- Reliability:

The reliability of the coding rules and categories may also need to be considered in terms of the intra and inter-reliability. Intra-reliability refers to agreement in coding by the same coder over time, while inter-reliability is the agreement between coders.

- Validity:

The validity relates to the extent that the instrument measures what it claims to measure. The face validity of the instrument is at times considered in relation to the intra and intercoder reliability, that is when there is agreement this will likely not be an issue, however it is also assessed by a panel of experts.

- Data Collection:

All the units of analysis (documents) are carefully read and coded according to the coding rules. To aid this process a coding sheet may also be used. This

coding sheet helps in the collection of descriptive information that can be used to supplement the thematic content. Coded data are compiled into the coding categories and when the purpose is a descriptive analysis, may be reduced to count data or manageable units such as percentages, means or ranges.

- Data Analysis:

The coded data that has been reduced according to the coding categories can be counted and analysed to demonstrate frequency distribution and central tendencies and can also be subject to statistical analysis.

Qualitative content analysis is also undertaken and this differs from the numerically based analysis of the text. However it has been suggested that this is controversial because the divide between proponents of the qualitative and quantitative approaches (Morgan 1993). Differences in qualitative content analysis compared to the quantitative approach are listed below (Morgan 1993).

- Coding:

Coding categories are more likely to be generated from the text.

- Breadth of Codes

Coding categories are frequently broader and more subjective.

- Coding Procedure:

Search algorithms that apply codes automatically are less likely to be used, instead coding is achieved through the careful reading and re-reading of the text.

- Use of Counts

While in quantitative content analysis the coding of the text is counted, tabulated and analysed, for qualitative analysis the focus is on interpreting the pattern that is found in the coding.

This qualitative content analysis does not differ significantly from that used by primary qualitative researchers, in that the aim is to uncover patterns in the text.

These patterns emerge from the analysis of the coded text and so new insights and understandings are generated. While the units of analysis for the primary research will be interview data, for the reviewer it is published studies. However, both these data sources provide a means to generate new understandings of a phenomenon.

### Choice of Method

Based on this discussion, it can be argued that there would appear to be only relatively minor differences between the three approaches to the synthesis of interpretive research. However, because of the limited debate in the health care literature and the fact that there are only a small number of completed reviews, no single approach can currently be selected with confidence as the best option. Meta-ethnography, meta-synthesis and qualitative content analysis all aim to identify the themes emerging from multiple independent interpretive studies. However, it is proposed that the final choice of method will likely be no more important than the many other decisions made during the systematic review process.

In choosing an approach to interpretive data synthesis for the expanded review process, the aim of these three approaches must be examined. Both meta-ethnography and meta-synthesis aim to generate a new interpretation from the texts. The concern with this is whether the product will be meaningful. In undertaking this synthesis, the number of interpretations of the phenomenon increases with each activity. That is, the phenomenon is interpreted by the person, then by the primary researcher and finally by the reviewer. The risk is that the interpretive synthesis will produce a report that is readable, but too far removed from the phenomenon to be meaningful.

Additionally, the approach taken during meta-analysis of experimental and observational research does not create new interpretations, rather it is an aggregative process. Meta-analysis summarises the findings from individual studies to determine the average effect size of the intervention across studies. The mean effect size for the

treatment group is then compared statistically to the mean effect size of the control group. This summary is descriptive in nature rather than creating new interpretations from the data. New interpretations generated from this processed data are viewed with suspicion. This suspicion is evident in the cautions given in regard to subgroup analyses and the warning that these analyses may result in misleading information (Mulrow and Oxman 1997). Meta-analysis is sometimes used to generate new interpretations of the data across multiple studies. However, in these situations the raw data from individuals, rather than from the published reports, are used as the basis of the analysis (Mulrow and Oxman 1997).

The aim of the synthesis process should be a merging of the text rather than a re-interpretation. While new interpretations may be possible from individual studies, it is argued that like in meta-analysis, they should be generated using the primary interview data from each included study. Needless to say, this would be an extremely time consuming process.

The aim of the expanded review process, in relation to interpretive data, should be the identification of similarities and differences across studies, and to identify, group and categorise common themes as is done in all systematic reviews. This process should aim to determine the areas of agreement in the text and how these areas of commonality are described in the different reports. Through this synthesis process the essence of a phenomenon or event can be described. This description is strengthened because it has been generated from multiple populations, settings and circumstances. In addition to areas that are common across texts, areas of difference will also be identified. These similarities and differences will help define the boundaries of the phenomenon. Because of this composite description, it will be more robust and transferable to other similar populations.

For the purposes of the expanded review process, qualitative content analysis was used as the method for synthesising interpretive research. The rationale for this is that

because this approach aims to identify common themes and patterns from multiple independent text, it most closely relates to the purposes of systematic reviews. However, as with all research endeavours, some degree of interpretation is inevitable. Yet it is argued that merging of studies through this synthesis process will provide important information on which to base practice.

### **Descriptive Data Synthesis**

Descriptive studies can also be synthesised, although this is not commonly part of published systematic reviews. Yet if the findings of other research methods are strengthened by the use of multiple populations, settings and circumstances, this can also be applied to descriptive studies.

While there is little information in the literature on how this may be achieved, as previously described, other approaches to research synthesis always use methods similar to those used in the primary research. For RCTs the synthesis produces the mean treatment effect which serves as the basis for analysis. For interpretive studies, the approach involves the identification of recurring themes in the text. On this basis it is argued that the most appropriate approach to the synthesis of research findings generated by descriptive studies will be similar to those used during the original analysis by the primary researcher.

The nature of the data reported in descriptive studies will therefore influence the analysis. The data produced by some descriptive studies is narrative text. For example, descriptive studies have reported on the reasons why health care workers physically restrain patients (Koch 1993; Retsas 1998). For this type of data the aim would be to categorise, group and tabulate the reasons for restraining patients. As previously discussed, content analysis is suitable for this type of data and has been used to uncover common themes in a large number of differing mediums.

However, some descriptive studies report numerical results. For example, descriptive studies have reported on the use of physical restraints in terms of the frequency of their use, which patients are restrained, and the average age of these patients (Gaebler 1994; Karlsson et al. 1996). For this type of descriptive numerical data, the aim of the synthesis would be to determine the mean value across studies. This summary, like the primary data from studies, would be in terms of mean values, ranges and frequencies.

As a result of these syntheses, the findings from multiple independent descriptive studies could be pooled. The product of this type of synthesis would be more robust and generalisable than that generated by single studies because it is based on different populations, settings and circumstances. Despite this argument, it must be acknowledged that descriptive studies are not currently part of systematic reviews. Yet as previously argued, if these studies are to be prevented from being lost to the profession in the growing volume of health care literature, it will be reviews that will provide a record of their existence. On this basis, if the questions posed are best answered by descriptive research, then it is reasonable to include this evidence in a systematic review.

### **Expanded Review Process: Data Synthesis**

In terms of the expanded systematic review process, the approaches used were matched to the research methods and type of data of the primary research. Studies were therefore grouped according to their method; experimental or observational, interpretive and descriptive. This phase of the systematic review aimed to summarise studies and to synthesise their results.

## Summary of Studies

Firstly the review aimed to summarise all relevant studies on the topic. This was achieved using both tabular and narrative summaries. Study summary tables were developed for all relevant studies regardless of research method. These data were collected using data collection tool developed as part of the review protocol. These summaries recorded all important features of the studies. This section aimed to provide a record of all relevant information, including:

- populations and settings studied,
- variations in the use of the intervention,
- outcome measures used to evaluate the intervention,
- research methods used to evaluate the intervention, and
- any limitations of existing studies.

## Synthesis of Findings

The second aim was to synthesise the findings of independent studies using appropriate methods. The choice of the method of data synthesis was influenced by the type of research and the nature of the data.

### 1 Experimental and Observational Research

The method used to synthesise the findings from RCTs or observational studies was influenced by the nature of the data.

#### Data Synthesis:

- Dichotomous Data

The odds ratio and the 95% confidence interval was calculated for each study and these were then pooled in a meta analysis.

- Continuous Data



- For data generated using the same measurement scale the weighted mean difference and 95% confidence interval was calculated for each study and then pooled in a meta-analysis.
- For data generated using different measurement scales the standardised mean difference and 95% confidence interval was calculated for each study and then pooled in a meta-analysis.

#### Heterogeneity

- Heterogeneity between studies included in the meta-analysis was evaluated using the chi square test, with a level of significance of 0.1.

#### Transcription Error:

- Double data entry was utilised to minimise the risk of transcription error during data entry.

#### Sensitivity Analysis:

- To evaluate the impact of decisions made during the critical appraisal of studies, a sensitivity analysis was undertaken when practical by incorporating excluded studies in the meta-analysis and evaluating their impact.

#### Approach to Synthesis:

- To avoid the risk of error due to choice of approach to synthesis, the fixed effect and random effect models were both be used.

#### Missing Data

- When incomplete reporting was encountered, and data could not be obtained from researchers, studies were included in a narrative discussion.

#### Result Presentation:

- Results of the meta-analysis were displayed graphically and in tabular form.

## 2. Interpretive Research

To synthesise the findings of interpretive research a qualitative content analysis was used to identify the patterns in the text. During this analysis the following methods were used.

- Identify the units of analysis.
- Read and re-read the text.
- Identify the broad coding categories during the reading of the text.
- Develop data collection tool to record descriptive data and to list coded data.
- Pilot test the coding categories.
- Refine the coding schema.
- Code text and document categories on data extraction tool.
- Analyse by comparing and contrasting text.
- Grouping coded themes into categories, keeping links to the original source.
- Categories labelled and exemplars identified from the text.

## 3. Descriptive Research

To synthesise the findings of descriptive research that reported narrative data, a descriptive content analysis was used. The approach to this analysis is described below.

- Identify the units of analysis
- Read the text
- Formulate the coding categories
- Develop a data extraction tool to record descriptive data and to list coded data
- Pilot test the coding categories
- Refine the coding schema

- Code the text
- Count and group coded text

To synthesise the findings of descriptive studies that reported numerical data, outcomes of interest were averaged across studies. Whenever possible, these mean values were calculated from raw data rather than processed results.

# **CHAPTER 5**

## **Evaluation of the Proposed Expanded Systematic Review Framework**

### **The Protocols**

## Evaluating the Proposed Expanded Review Framework

### Introduction

In developing the expanded review process reported in this thesis, the methods were drawn from the professional literature and through examination of completed reviews. However, without some form of evaluation, this framework remains a theoretical construct rather than a practical approach to the conduct of a systematic review of nursing research. To evaluate both the methods and product of the expanded review process, a review was conducted. The methods used were based on the approaches described in the previous chapter of this thesis. As previously described, the evaluation of the framework was in terms of:

- its use in the development of a systematic review protocol,
- enabling the summary and synthesis of a range of methodologically different studies, and
- whether it produces valid and useful results.

The evaluation consisted of four phases:

- development of a systematic review protocol,
- presentation of the results of the reviews addressing music in hospitals,
- selected examples of the results of a comparison review on physical restraint, and
- discussion of the process and implications.

To aid in the evaluation of the process and results of the review, the systematic review addressing music in hospitals was treated as three separate reviews addressing the effectiveness, appropriateness and feasibility of music. The selected results of the physical restraint review are also presented according to their contribution to one of these three components of the conceptual framework.

## **Overview of Expanded Review Topic**

The topic used during this evaluation of the expanded review process was music as an intervention for hospital patients. The application of music as a specific therapeutic intervention is a development largely of the mid 20th century, although it has existed in various forms in most cultures for many centuries (Marwick 1996). In recent years the use of music as an intervention has increased, as witnessed by the evaluations undertaken by many health care disciplines. Music has been used for patients of all ages, including infants (Marley 1984; Caine 1991), children (Dun 1995; Malone 1996), adults (Whipple and Glynn 1992) and the elderly (Janelli and Kanski 1997). Music has also been used in many health care specialties such as intensive care (Johnston and Rohaly-Davis 1996), coronary care (Guzzetta 1989), cancer care (Standley and Hanser 1995), maternity units (Geden et al. 1989), geriatric units (Denney 1997), palliative care (O'Callaghan 1996) and outpatient departments (Dubois et al. 1995).

The effectiveness of music has been used to produce behavioural changes in physically restrained patients (Janelli and Kanski 1997), helping demented patients eat more calmly (Ragneskog et al. 1996), and improving the self-esteem, mood and depression in older adults (Hanser and Thompson 1994). It has been evaluated pre-operatively (Winter et al. 1994), intra-operatively (Kopp 1991) and as a post-operative intervention (Heiser et al. 1997). Music therapy has also been used as an adjunct to anaesthesia (Oddby-Muhrbeck and Jakobsson 1993; Tang et al. 1993).

During chemotherapy music may reduce the incidence, and time to onset, of nausea (Standley 1992). One study suggested it reduces pain in a group of cancer patients (Beck 1991), while another found it had no effect on the severity of side effects for patients experience during chemotherapy (Sabo and Michael 1996). Music has also been used to increase patients' tolerance of unpleasant, or uncomfortable procedures,

such as gastrointestinal endoscopic procedures (Bampton and Draper 1997), sigmoidoscopies (Palakanis et al. 1994) and bronchoscopies (Dubois et al. 1995).

In summary, music as an intervention for hospital patients has been evaluated in a diverse range of patient care situations. The focus of many studies related to the reduction of pain, anxiety or stress, improvement in comfort or patient satisfaction, or to increase patients' tolerance of unpleasant procedures. However, the results of these studies are contradictory. Additionally, many studies involved only small numbers of participants, and so may have lacked the power to detect beneficial outcomes. As a result of these issues, this topic was selected to evaluate the expanded review process.

### **Systematic Review Protocols**

As previously described, a protocol is developed for each systematic review and serves the same purpose as any research proposal. This protocol helped minimise the need for subjective decisions during the conduct of the review. The three music protocols followed the accepted conventions, and described the:

- objectives
- review questions
- inclusion criteria
- exclusion criteria
- search strategy
- assessment of validity
- data collection
- data synthesis

As previously stated, the protocols were treated to be three distinct reviews to allow a more comprehensive evaluation of the expanded review process. The three music in hospitals protocols were:

- Systematic review I      Effectiveness      The process
- Systematic review II      Appropriateness      The person
- Systematic review III      Feasibility      The environment

## **Definitions**

### **Music**

Music in the context of this review, was considered to be recorded music played via a tape recorder or compact disc player. Studies involving live music were considered beyond the scope of the review. Music as an intervention was defined as music played for a patient during a single episode of care to produce outcomes that were achievable during that session of music. Music played to patients prior to, and then following, surgery was considered to fit within this definition. Music played to patients as a series of sessions over an extended period of time, where the outcomes were achieved through participation in the program of sessions, was considered beyond the scope of this review.

### **Distracter**

Music in the context of this systematic review was viewed as a distracter. The basis for this is the fact that music is a form of communication and has been described as a universal language (Stevens 1990), so it can provide an escape through imaginative thought (Livingston 1985). This escape is through a shift in the focus of an individual from anxiety caused by an impending or occurring activity to the imaginative thought.



### Distinct Study Populations

During the initial search of the literature it became evident that there were two distinct populations. Firstly, the population of some studies consisted of hospital patients resting in bed. Typically these patients were awaiting operative procedures, recovering from surgery, or recovering from illness. However, a theme common to all of these study participants was that they rested quietly during the music intervention. This group of patients have been referred to as Hospital Patients in this review report.

A second group of studies involved hospital patients undergoing procedures during the music intervention. These procedures included such things as bronchoscopy, sigmoidoscopy, or surgical procedures under a regional anaesthetic. This group of patients have been referred to as Procedure Patients in this review report. Because of the important difference in circumstances between these two groups of studies, no attempt was made to combine the data across these categories.

## **Protocol I : Effectiveness**

### **Objectives**

The objective of this review was to summarise the best available evidence related to the effectiveness of music for patients with regard to its influence on physiological variables. The specific objective was to determine the effectiveness of music in minimising the physiological effects of anxiety and pain.

### **Review Questions**

The question this review sought to answer was:

- Does music reduce patient's perception of pain?
- Does music reduce the physiological effects of anxiety, such as increases in heart rate, blood pressure and respiratory rate?

In addition to all patients, the review also addressed these questions during specific events, which were:

- surgical procedures, covering the preoperative, intraoperative and postoperative period
- treatment of cancer
- recovery period following myocardial infarction

### **Inclusion Criteria**

Inclusion criteria were developed as part of the review protocol to aid in the selection of studies. These criteria assisted in the determination of the suitability and relevance of studies for the systematic review. These criteria were used on two specific occasions. During the database and reference list searches the criteria were used to

determine the suitability of studies for retrieval from the library. These criteria were used again when the complete research report had been retrieved to determine if the study should be included in the review. The criteria were:

1. Type of participants - adult hospital patients.
2. Type of interventions - the use of recorded music for a single episode of care.
3. Types of outcome measures - those relating to the patient's physiological response to music, including:
  - heart rate
  - blood pressure
  - respiratory rate
  - severity of pain
  - amount of analgesic used
  - amount of sedative used
4. Types of studies - randomised controlled trials.

### **Exclusion Criteria**

Exclusion criteria were also developed as part of the review protocol to assist in determining which studies would not be included in the review. Studies were excluded from the systematic review if:

- the report was in a non-English language
- the study involved a population other than hospital patients
- the study used live music rather than recorded music
- critical appraisal indicated that it was of poor methodological quality
- there was inadequate description in the study report to determine specific information about the participants, intervention, outcome measures or research method

## **Search Strategy**

The search involved a comprehensive and systematic search of electronic databases using the methods developed by the Cochrane Collaboration (Mulrow and Oxman 1997) and the NHS Centre for Review and Dissemination (NHS Centre for Reviews and Dissemination 1996). The search entailed a series of successive steps as recommended by Dickersin (Dickersin 1994). For the complete search strategy details see Appendix 1.

1. The initial step involved a limited search of the title, abstract and descriptor/MeSH sections of database citations to enable optimal search terms to be identified. This identification of optimal search terms was repeated for each database included in the search.
2. A comprehensive search of each database was then undertaken utilising all optimal search terms and no time limits were used to restrict the search.
3. The database Current Contents, which undergoes weekly updating of citations, was searched monthly to identify studies published during the conduct of the review.
4. Selected journals that focus specifically on both music and health care were hand searched.
5. The final step was a search of all bibliographies and reference lists of retrieved papers for additional studies not identified during the initial steps of the search.

The electronic databases searched were:

- MEDLINE
- CINAHL
- Cochrane Library
- EMBASE
- Psyclit

- DARE
- Expanded Academic Index
- Health Star
- AUSThealth, which included the following databases:
  - Rural, Australian Public Affairs Information Service - Health
  - Health & Society on Australian Health
  - Aboriginal & Torres Strait Islander Health
  - Australian Medical Index
- Current Contents.

Journals that were hand searched were:

- Journal of Music Therapy
- International Journal of Arts in Medicine
- Music Therapy
- The Australian Journal of Music Therapy

Studies were selected for retrieval from the library by comparing the information included in the title, abstract or descriptor/MeSH terms against the inclusion criteria. Studies identified during the searching of reference lists or bibliographies were selected for retrieval based on the information included in the citation title.

### **Assessment of Validity**

All identified studies that met the inclusion criteria were assessed for methodological validity prior to inclusion in the review. Assessment of validity was undertaken using the critical appraisal tool previously described (see Appendix 2). For RCTs the appraisal focused on the four sources of bias:

- selection bias
- performance bias

- attrition bias
- detection bias

Studies assessed as being of good methodological quality were considered for inclusion in the meta-analysis. Studies identified as having methodological limitations were excluded from the meta-analysis. The studies excluded from the meta-analysis were used in the narrative discussion to help document the current evidence related to the use of music for adult hospital patients and to determine future research needs.

### **Data Collection**

To minimise the risk of error during the transcription of data, a data collection tool was developed and used for this review (see Appendix 3). The collected data was then compared to the data reported in the study report to ensure its accuracy. The data collected included:

- demographic information about the study population
- description of the intervention
- description of the outcome measures
- the study method
- results data

### **Data Analysis**

The aim of the data analysis phase of the review was to both synthesise and summarise the findings of all studies addressing the use of music for hospital patients.

### **Research Synthesis**

The research synthesis component of the review was achieved using meta-analysis. When two or more comparable studies were identified the results were pooled in a

meta-analysis to determine the effectiveness of music therapy. During this review on effectiveness, meta-analysis was limited to RCTs. Comparability of studies was in terms of the specific study population, the intervention and the outcome measures utilised. Studies were grouped based on whether study participants were Hospital Patients or Procedure Patients. Initial analysis involved pooling all studies in each of the two review population groups for the specific outcomes of analgesic and sedative use, heart rate, blood pressure and respiratory rate.

Sub-group analysis was then undertaken to pool data from studies based on the different populations. The purpose of this analysis was to determine the effectiveness of music for specific populations. This sub-group analysis combined studies according to the following populations:

- cardiac patients
- cancer patients
- pre-operative patients
- intra-operative patients
- post-operative patients

Meta-analysis was undertaken using Review Manager, version 4.04 (Cochrane Collaboration 1999). Heterogeneity between comparable studies was assessed using chi square and visual inspection of the graphic presentation of results. Significant heterogeneity was considered present when the p-value was less than 0.05. Double data entry was utilised to minimise the risk of error during the data entry phase of the analysis.

The approach used during the meta-analysis was determined by the type of data collected. For dichotomous data, the odds ratio was used as the summary measure of effect. The 95% confidence interval was calculated for each study. For continuous data that used the same scale, such as blood pressure, the weighted mean difference was used as the summary measure of effect (Mulrow and Oxman 1997). For

continuous data that used different scales, such as occurred with the measurement of anxiety, the standardised mean difference was used (Mulrow and Oxman 1997; NHMRC 1999). The 95% confidence interval was calculated for each study. The results of the meta-analysis were presented graphically.

To test the robustness of the results of the review, a sensitivity analysis was performed. As some studies were excluded from the meta-analysis following critical appraisal, the results from these studies were added to the pooled data in order to evaluate their impact. In addition to this, a sensitivity analysis was used to evaluate the impact of using the fixed effects and the random effects models to pool data. The confidence in the pooled results was considered to be greater if they were not materially changed as a result of these sensitivity analyses.

## Research Summary

The research summary component of the review results was achieved using tabular and narrative summaries.

### 1. Tabular Summary

Summary tables were used to provide a brief description of the key study characteristics and their findings. Both included and excluded studies were documented in these summary tables.

### 2. Narrative Summary

Studies were also summarised through narrative discussion. The objective of this component of the review was to identify what approaches have been used, provide a methodological appraisal of studies, and provide a descriptive summary of demographic information of existing studies.



Through the activities of narrative and tabular summary, this systematic review aimed to create a record of all past research evaluating the effectiveness of music for hospital patients.

### **Implications For Practice and Research**

On the basis of the meta-analysis, recommendations for clinical practice and research were developed. The recommendations for clinical practice were graded according to the hierarchy previously described in this thesis. The recommendations for future research were generated based on areas where a lack of research evidence was identified.

## **Protocol II : Appropriateness**

### **Objectives**

The objective of this review was to summarise the best available evidence related to the appropriateness of using music for hospital patients. Appropriateness, as previously described, related to the psychosocial impact of music as an intervention. This section of the systematic review sought to summarise research addressing the recipient's perspective of music. The specific objects were to determine the appropriateness of music with regard to perception of pain, anxiety, tolerance, mood and satisfaction.

### **Review Questions**

The specific questions this review sought to answer were:

- Does music reduce the anxiety of patients?
- Does music improve comfort during unpleasant procedures?
- Does music improve patients' satisfaction with the health care provided?
- Does music improve patients' tolerance of unpleasant procedures?
- Does music improve the mood of patients?

Additionally, the review also evaluated the appropriateness of music during specific events, which were:

- surgical procedures, covering the preoperative, intraoperative and postoperative period
- treatment of cancer
- recovery period following myocardial infarction

## **Inclusion Criteria**

As described in Protocol I, inclusion criteria were developed as part of the review protocol to aid in the selection of studies to assist in determining the suitability and relevance of studies for the systematic review. For studies to be included in the review they had to meet all the criteria, which were:

1. Type of participants - adult hospital patients.
2. Type of interventions - the use of recorded music for a single episode of care.
3. Types of outcome measures - those relating to the patient's physiological response to music, including:
  - anxiety, such as measured with the State Trait Anxiety Inventory (STAI) or visual analogue scales (VAS)
  - satisfaction, such as measured with a VAS
  - mood, such as measured with a VAS

In addition to these quantitative outcomes, qualitative outcomes related to the appropriateness were also included. These outcomes included:

- the patient's perception of music,
  - reactions to the use of music.
4. Types of studies - a range of research designs could reasonably provide evidence related to the appropriateness of music for hospital patients, and these included:
    - RCTs that focused on psychosocial outcome measures
    - observational studies that focused on quality of life outcomes
    - interpretive studies that explored the topic from the perspective of the patient

## **Exclusion Criteria**

Exclusion criteria were also developed as part of the review protocol to assist in determining which studies would not be included in the review. Studies were excluded from the systematic review if:

- the report was in a non-English language
- the study involved a population other than hospital patients
- the study used live music rather than recorded music
- critical appraisal indicated that it was of poor methodological quality
- there was inadequate description in the study report to determine specific information about the participants, intervention, outcome measures or research method

## **Search Strategy**

The search involved a comprehensive and systematic search of electronic databases using the methods developed by the Cochrane Collaboration (Mulrow and Oxman 1997) and the NHS Centre for Review and Dissemination (NHS Centre for Reviews and Dissemination 1996). The search entailed a series of successive steps as described in Protocol I. For the complete search strategy details see Appendix 1.

1. A limited search of the title, abstract and descriptor/MeSH sections of database citations to identify optimal search terms. This identification of optimal search terms was repeated for each database included in the search.
2. A comprehensive search of each database utilising all optimal search terms and no time limits were used to restrict the search.
3. Weekly search of Current Contents to identify recently published studies.
4. Hand searching selected journals.
5. A search of all bibliographies and reference lists of retrieved papers for additional studies.

The electronic databases searched were:

- MEDLINE
- CINAHL
- Cochrane Library
- EMBASE
- Psyclit
- DARE
- Expanded Academic Index
- Health Star
- AUSThealth, which included the following databases:
  - Rural, Australian Public Affairs Information Service - Health
  - Health & Society on Australian Health
  - Aboriginal & Torres Strait Islander Health
  - Australian Medical Index
- Current Contents.

Journals that were hand searched were:

- Journal of Music Therapy
- International Journal of Arts in Medicine
- Music Therapy
- The Australian Journal of Music Therapy

Studies were selected for retrieval from the library by comparing the information included in the title, abstract or descriptor/MeSH terms against the inclusion criteria.

### **Assessment of Validity**

As described in Protocol I, all identified studies that met the inclusion criteria were assessed for methodological validity prior to inclusion in the review. Assessment of

validity was undertaken using the critical appraisal tool previously described (see Appendix 2).

For RCTs the appraisal focused on four sources of bias:

- selection bias
- performance bias
- attrition bias
- detection bias

For cohort and case control studies, a different checklist was used that focused on four potential sources of bias (see Appendix 2):

- selection bias
- performance bias
- attrition bias
- detection bias

The assessment of interpretive studies was undertaken using a checklist which focused on a minimum standard of reporting:

- a clear description of method used
- the method used was appropriate
- a clear description of the study population
- supportive data for themes, categories and labels

The assessment of descriptive studies was undertaken using a checklist which focused on a minimum standard of reporting:

- a clear description of method used
- the method used was appropriate
- a clear description of the study population
- adequate reporting of results

## **Data Extraction**

To minimise the risk of error during the transcription of data, a data collection tool was developed and used for this review (see Appendix 3). Data collected for the evaluation of the appropriateness of music were those measures related to psychosocial outcomes. For RCT and observational studies, the data collected included:

- demographic information about the study population
- description of the intervention
- description of the outcome measures
- the study method
- results data

Data collection from interpretive research was undertaken as part of the synthesis of studies. Data collected from these studies were narrative descriptions of major themes, metaphors, categories and labels. For descriptive studies, the data collected was influenced by the nature of the study, but included descriptive data such as frequency of events or mean scores.

## **Data Analysis**

The aim of the data analysis phase of the review was to both synthesise and summarise the findings of all studies investigating the use of music for hospital patients.

## **Research Synthesis**

### **1. Experimental and Observational Studies**

The research synthesis component of the review was achieved using meta-analysis. As described in protocol I, when two or more comparable studies were identified the results were pooled in a meta-analysis to determine the appropriateness of music

therapy. During this review on appropriateness, meta-analysis was limited to RCTs, cohort studies and case control studies. Comparability of studies was in terms of the specific study population, and how the intervention and the outcome measures were utilised. Studies were grouped based on whether study participants were Hospital Patients or Procedure Patients.

Sub-group analysis was then undertaken to pool data from studies based on the different populations. The purpose of this analysis was to determine the effectiveness of music for specific populations. This sub-group analysis attempted to combined studies according to the following populations:

- cardiac patients
- cancer patients
- pre-operative patients
- intra-operative patients
- post-operative patients

Meta-analysis was undertaken using Review Manager, version 4.04 (Cochrane Collaboration 1999). The approach used during the meta-analysis was determined by the type of data collected and was the same as described in Protocol I. As in Protocol I, heterogeneity between comparable studies was assessed using chi square, the measurement of effect was the odds ratio, weighted mean difference or standardised mean differences as appropriate. Similarly, to test the robustness of the results of the review, a sensitivity analysis was performed, when appropriate, to evaluate their impact to evaluate the impact of decisions made during the review.

## 2. Interpretive Studies

Interpretive studies focusing on the appropriateness of music were pooled using a qualitative content analysis. The steps in this synthesis were:



- Selected the text to be included in the synthesis using the inclusion criteria.
- Read and re-read the text.
- Selected the broad coding categories that were identified during the readings of the text.
- Pilot tested the coding categories and revised them as needed.
- Documented bibliographic details and other relevant information using data collection tool.
- Read and coded text according to the selected coding categories.
- Compiled the coded data using data the collection tool.
- Grouped and categorised similar themes and progressively refined the descriptions of the phenomena.
- Identified explanatory themes, specific descriptions and key phrases from the text.

Through this process a description of the phenomena was generated based on the findings of individual studies. The results of this analysis were presented using a narrative summary.

## Research Summary

The research summary component of the review results was achieved using tabular and narrative summaries.

### 1. Tabular Summary

Summary tables were used to provide a brief description of the key study characteristics and their findings. Both included and excluded studies were documented in these summary table.

## 2. Narrative Summary

Studies were also summarised through narrative discussion as described in Protocol I. The objective of this component of the review was to identify what approaches have been used, provide a methodological appraisal of studies and provide a descriptive summary of demographic information of existing studies. Through the activities of narrative and tabular summary, this systematic review aimed to create a record of past research evaluating the appropriateness of music for hospital patients.

### **Implications For Practice and Research**

On the basis of the meta-analysis, recommendations for clinical practice and research were developed. The recommendations for clinical practice were graded according to the hierarchy previously described in this study. The recommendations for future research were generated based on areas where a lack of research evidence was identified.

## **Protocol III : Feasibility**

### **Objectives**

The objective of Review III was to summarise the best available evidence related to the feasibility of using music for hospital patients. Feasibility, as previously described, related to the environmental aspects of music as an intervention. This section of the systematic review summarised research addressing the implementation and support of music as an intervention. The specific objects were to determine the feasibility of music with regard to such things as implementation, cost, benefit and harm. In addition to these, the objective was also to identify any factors that had an impact on the use of music in the hospital setting.

### **Review Questions**

The specific questions this review sought to answer were:

- Can music be implemented in the hospital setting?
- What is the potential benefit and harm from the use of music?
- What is the economic impact of the use of music?
- What issues impact on the implementation, or the use of music as a nursing intervention?

### **Inclusion Criteria**

As in protocols I and II, inclusion criteria were developed as part of the review protocol to aid in the selection of studies. These criteria assisted in determining the suitability and relevance of studies for the systematic review. For studies to be included in the review they had to meet all the criteria, which were:

1. Type of participants - adult hospital patients.

2. Type of interventions - the use of recorded music for a single episode of care.
3. Types of outcome measures - those relating to health care worker reaction and impact on the health care organisation, including those related to:
  - economic impact
  - organisation change
  - health care worker response to the use of music
  - acceptance

Qualitative outcomes related to the feasibility of implementing or utilising music in hospitals were included. These outcomes related to:

- the implementation of music for hospital patients
- issues related to music as a nursing intervention
- barriers to implementation

Descriptive outcomes related to the feasibility of implementing or utilising music in hospitals were included. These outcomes included:

- descriptions of the implementation of music as an intervention
- descriptions of the use of music in hospitals
- investigation of health care workers reactions to the use of music
- investigation of factors that act as barriers to the effective implementation of music as a nursing intervention
- investigations into factors that influence the use of music in hospitals

4. Types of studies - a range of research designs could reasonably provide evidence related to the feasibility of music for hospital patients, and these included:
  - RCTs that focused on outcome measures related to implementation

- observational studies that investigated the implementation of music
- interpretive studies that explored issues related to implementation, reactions by health care workers to the use of music, or barriers to implementation
- descriptive studies that described some aspect of implementation of music, reactions and perceptions to its use or barriers to its implementation

### **Exclusion Criteria**

Exclusion criteria were also developed as part of the review protocol to assist in determining which studies would not be included in the review, and these were:

- the report was in a non-English language
- the study involved a population other than hospital patients
- the study used live music rather than recorded music
- critical appraisal indicated that it was of poor methodologically quality
- there was inadequate description in the study report to determine specific information about the participants, intervention, outcome measures or research method

### **Search Strategy**

The search involved a comprehensive and systematic search of electronic databases using the methods developed by the Cochrane Collaboration (Mulrow and Oxman 1997) and the NHS Centre for Review and Dissemination (NHS Centre for Reviews and Dissemination 1996). The search entailed a series of successive steps as recommended by Dickersin (Dickersin 1994) and was the same as reported in Protocols I and II. For the complete search strategy details see Appendix 1.

As in protocols I and II, a large number of electronic databases were searched including:

- MEDLINE
- CINAHL
- Cochrane Library
- EMBASE
- Psyclit
- DARE
- Expanded Academic Index
- Health Star
- AUSThealth, which included the following databases:
  - Rural, Australian Public Affairs Information Service - Health
  - Health & Society on Australian Health
  - Aboriginal & Torres Strait Islander Health
  - Australian Medical Index
- Current Contents.

Journals that were hand searched were:

- Journal of Music Therapy
- International Journal of Arts in Medicine
- Music Therapy
- The Australian Journal of Music Therapy

Studies were selected for retrieval from the library by comparing the information included in the title, abstract or descriptor/MeSH terms against the inclusion criteria. Studies identified during the searching of reference lists or bibliographies were selected for retrieval based on the information included in the citation title.

## **Assessment of Validity**

As described in protocols I and II, all identified studies that met the inclusion criteria were assessed for methodological validity prior to inclusion in the review. Assessment of validity was undertaken using the critical appraisal tool previously described (see Appendix 2).

For RCTs the appraisal focused on four sources of bias:

- selection bias
- performance bias
- attrition bias
- detection bias

For cohort and case control studies, a different a checklist was used that focused on four potential sources of bias:

- selection bias
- performance bias
- attrition bias
- detection bias

The assessment of interpretive studies was undertaken using checklist which focused on a minimum standard of reporting:

- a clear description of the method used
- the method used was appropriate
- a clear description of the study population
- supportive data for themes, categories and labels

The assessment of descriptive studies was undertaken using a checklist which focused on a minimum standard of reporting (see Appendix 2):

- a clear description of the method used

- the method used was appropriate
- a clear description of the study population
- adequate reporting of results

### **Data Extraction**

To minimise the risk of error during the transcription of data, a data collection tool was developed as reported in Protocols I and II (see Appendix 3).

### **Data Analysis**

The aim of the data analysis phase of the review was to both synthesise and summarise the findings of all studies investigating the use of music for hospital patients.

### **Research Synthesis**

#### **1. Experimental and Observational Studies**

The research synthesis component of the review was achieved using meta-analysis. When two or more comparable studies were identified the results were pooled in a meta-analysis using Review Manager, version 4.04 (Cochrane Collaboration 1999). Heterogeneity between comparable studies was assessed using chi square and the measures of effect were the odds ratio, weighted mean difference or standardised mean difference as appropriate. The actual methods of meta-analysis were described in protocol I.

#### **2. Interpretive Studies**

Interpretive studies focusing on the feasibility of music were pooled using qualitative content analysis that sought to identify the patterns in the text. The methods used during the content analysis were as described in Protocol II.



### 3. Descriptive Studies

The findings from descriptive studies that focused on some aspect of the feasibility of music were synthesised when appropriate. The approach taken to the data analysis was influenced by the nature of the data presented in reports. For numerical data, mean values were averaged across studies to identify the mean for that variable and the range. For narrative data a descriptive content analysis was used to uncover the patterns in the text. The approach used was:

- Selected studies to be included in the analysis based on the review inclusion criteria.
- Read and re-read the text.
- Selected the coding categories during the readings of the text.
- Pilot tested the coding categories and revised them as needed.
- Documented bibliographic details and other relevant information using data collection tool.
- Read and coded according to the selected coding categories.
- Data were grouped into the coding categories.
- Coded data were reduced according to the coding categories and then counted and analysed to demonstrate frequency, distribution and differences across studies.

### Research Summary

The research summary component of the review was achieved using tabular and narrative summaries as described in protocol I.

#### 1. Tabular Summary

Summary tables were used to provide a brief description of the key study characteristics and their findings. Both included and excluded studies were documented in these summary table.

## 2. Narrative Summary

Studies were also summarised through narrative discussion. As described in Protocols I and II, the objective of this component of the review was to identify what approaches have been used, provide a methodological appraisal of studies and provide a descriptive summary of demographic information of existing studies.

### **Implications For Practice and Research**

As described in Protocols I and II, recommendations for clinical practice and research were developed. These recommendations were graded according to the hierarchy previously described in this thesis.

# **CHAPTER 6**

**Evaluation of the Expanded Systematic**

**Review Framework**

**Review Results**

## **Results of the Search and Critical Appraisal**

The results of the reviews are presented under the three specific areas, the process, person and environment. The results of reviews I (effectiveness) and II (appropriateness) are presented according to the two categories of patients, Hospital Patients and Procedure Patients. As there was limited research addressing the feasibility of music, these findings are presented as a general summary.

### **Demographics of Identified Studies**

#### **Randomised Controlled Trials**

The literature search identified a total of 39 clinical trials, of which 29 were RCTs that met the inclusion criteria (see appendix 4, tables 1 and 2). Ten RCTs were excluded from the review as a result of the critical appraisal (see appendix 4, table 3). Of the remaining 19 RCTs appraised as methodologically sound, 9 addressed the category of Hospital Patients (see appendix 4, table 1) and 10 RCTs addressed Procedure Patients (see appendix 4, table 2). One paper by Koch et al. reported two small RCTs in the same paper (Koch et al. 1998).

#### **Type of Music**

A wide range of music was used in the identified studies. However investigation of the effectiveness of specific styles of music used in studies, to determine if one style of music was better than other styles, was not possible because of issues such as too few studies, different circumstances of study populations and the different outcomes utilised across studies. The style of music used in the 19 RCTs were:

- selection from a variety of music styles (n = 9)
- classical (n = 4)

- instrumental (n = 2)
- tranquil / relaxing (n = 1)
- classical or sounds of nature (n = 1)
- piano (n = 1)
- new age (n = 1)

### Populations Studied

As previously stated the studies were grouped into two categories, Hospital Patients and Procedure Patients (appendix 4, tables 1 and 2). Within each of these two broad categories, music was evaluated in a range of specific patient groups which included:

- Hospital Patients
  - \* pre-operative (n = 2)
  - \* post-operative (n = 1)
  - \* cardiac surgical (n = 1)
  - \* post myocardial infarction (n = 3)
  - \* mechanically ventilated (n = 2)
- Procedure Patients
  - \* upper gastrointestinal investigations (n = 1)
  - \* cardiac surgery (intra-operative) (n = 1)
  - \* chest tube removal (n = 1)
  - \* first post-operative ambulation (n = 1)
  - \* bronchoscopy (n = 1)
  - \* cataract surgery (intra-operative) (n = 1)
  - \* urological procedures (intra-operative) (n = 1)
  - \* lithotripsy (intra-operative) (n = 1)
  - \* sigmoidoscopy (n = 1)
  - \* orthopaedic and plastic surgery (intra-operative) (n = 1)

### Incomplete Reporting of Results

Nine of the RCTs that met the inclusion criteria failed to provide complete results data, or data was presented graphically for some of the outcome measures (Walther-Larsen et al. 1988; Palakanis et al. 1994; Barnason et al. 1995; Chlan 1995; Miluk-Kolasa et al. 1996; Cruise et al. 1997; Chlan 1998; Koch et al. 1998; White 1999a). Following a search of the internet for contact details of researchers, missing data was obtained from 2 researchers (Chlan 1998; White 1999a). As a result of this missing data, the other 7 RCTs were not included in the meta-analysis.

### Critical Appraisal

Ten RCTs met the inclusion criteria but were subsequently excluded from the review as a result of the critical appraisal (appendix 4, table 3). The reasons for these exclusions were;

- multiple reasons (n = 3)
- attrition (n = 3)
- method of allocation (n = 2)
- other interventions were used in conjunction with music (n = 1)
- inadequate information provided in report (n = 1)

a) Multiple Reasons for Exclusion: Three RCTs were excluded from the meta-analysis for failing to meet two or more critical appraisal criteria. An RCT by Augustin utilised an inappropriate method of randomisation (alternation) and had significant differences in the treatment of groups (Augustin and Hains 1996). The significant difference was that the music group listened to music in a quiet darkened environment free from any stimulation, while the comparison group were permitted to read watch television, have visitors or walk about the pre-operative area. As a result of these factors, it was not possible to attribute differences in anxiety between the music and no-music group solely to the music intervention. In two other RCTs, the

method of randomisation was inappropriate in one and music was combined with another intervention in the second (Sabo and Michael 1996; Grey et al. 2000).

b) Attrition: Three RCTs were excluded from the meta-analysis because less than 80% of participants failed to complete the study. A study involving post-operative patients had 15 of 34 participants drop out (Heiser et al. 1997). A study of pre-operative patients had all the music group complete the study but only 19 of the 31 in the no-music group completed (Winter et al. 1994). Finally, a very small study of pre-operative patients involving only 6 patients in each of the two study groups, had 3 drop out of the no-music group (Szeto and Yung 1999).

c) Method of Randomisation: A number of studies were excluded because of the use of an inappropriate method of randomising participants to study groups, and for three of these studies other factors were also presented that would have resulted in their exclusion. These inappropriate methods included alternation (Augustin and Hains 1996), the day of the week (Cunningham et al. 1997; Grey et al. 2000), record number (Dubois et al. 1995), and by office (Sabo and Michael 1996).

d) Interventions Combined with Music: Some studies involved other interventions combined with music and because of this, differences between study groups could not be attributed to the music alone. Two of these studies also had other factors present that contributed to their exclusion. Other interventions used in combination with music included a recorded message (Sabo and Michael 1996), video (Miller et al. 1992), and a range of educational and supportive interventions (Grey et al. 2000).

e) Inadequate Information: One study, that appeared to be an RCT, failed to provide sufficient information of the method and so was not included in the meta-analysis (Kopp 1991). This differed from exclusion as a result of incomplete reporting, in that it was not clear what was done during the study or which patients

were involved. Additionally, there were no results data reported, with the paper simply stating the conclusions of the study.

### Study Size

A feature common to many of the RCTs addressing music, was the small number of participants in the studies. Below is a summary of the number of participants of studies:

- RCTs involving Hospital Patients
  - total number in all studies  $n = 483$
  - mean number of participants per study  $n = 48.7$
  - smallest number of participants in a single study  $n = 20$
  - largest number of participants in a single study  $n = 100$
- RCTs involving Procedure Patients
  - total number in all studies  $n = 594$
  - mean number of participants per study  $n = 59.4$
  - smallest number of participants in a single study  $n = 34$
  - largest number of participants in a single study  $n = 120$
- Excluded RCTs
  - total number in all studies  $n = 471$
  - mean number of participants per study  $n = 47.1$
  - smallest number of participants in a single study  $n = 12$
  - largest number of participants in a single study  $n = 100$
- Excluded clinical trials (non-RCTs)
  - total number in all studies  $n = 293$
  - mean number of participants per study  $n = 32.5$
  - smallest number of participants in a single study  $n = 15$
  - largest number of participants in a single study  $n = 63$



As a result of the small sample sizes many of these studies lacked the power to adequately detect beneficial outcomes.

#### Observational Studies

No observational studies addressing music in adult hospital patients were identified during the literature search.

#### Interpretive and Descriptive Studies

A single interpretive study addressing music in hospitalised adults was identified during the literature search (Stevens 1990). Two descriptive studies were also identified. One of these studies involved a survey of music therapists to explore their use of music for pain relief (Michel and Chesky 1995), the other provided a description of the use of music during surgery in one health care institution (Pratt 1999).

As a result of this extremely limited information a small number of clinical trials that had been excluded from the meta-analysis were presented in a discussion to attempt to identify issues of importance.

## Systematic Review I - Effectiveness

This section of the review addresses the effectiveness of music therapy and so focuses on the physiological outcome measures reported in RCTs. As previously described, RCTs were grouped into two broad categories, Hospital or Procedure Patients. The findings of the meta-analyses are reported firstly according to these two broad patient categories and then by specific sub-groups as described in the review protocol.

The outcome measures of interest in this review are:

- heart rate,
- blood pressure,
- respiratory rate,
- pain,
- sedation, and
- length of stay.

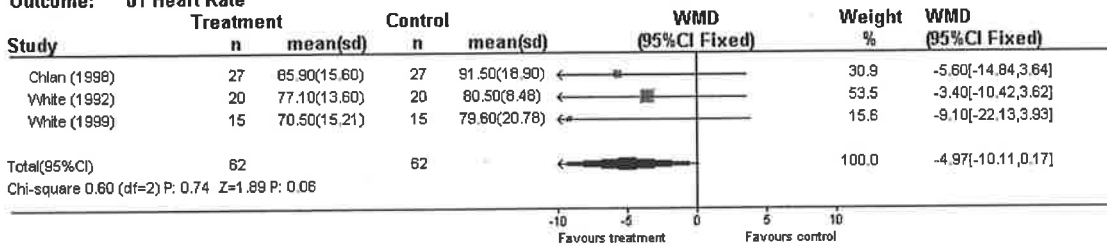
### **Heart Rate**

Eleven RCTs used heart rate as an outcome measure, however only six studies provided sufficient data to be included in a meta-analysis. For the category of Hospital Patients, three studies (White 1992; Chlan 1998; White 1999a) provided sufficient data and so were combined in a meta-analysis (see graph 1). The specific participants involved in these studies were:

- 54 mechanically ventilated patients (Chlan 1998)
- 40 patients post myocardial infarction (White 1992)
- 30 patients post myocardial infarction (White 1999a)

### Graph 1 Hospital Patients Music and Heart Rate

Comparison: 01 Vital Signs - Hospital Patients  
Outcome: 01 Heart Rate

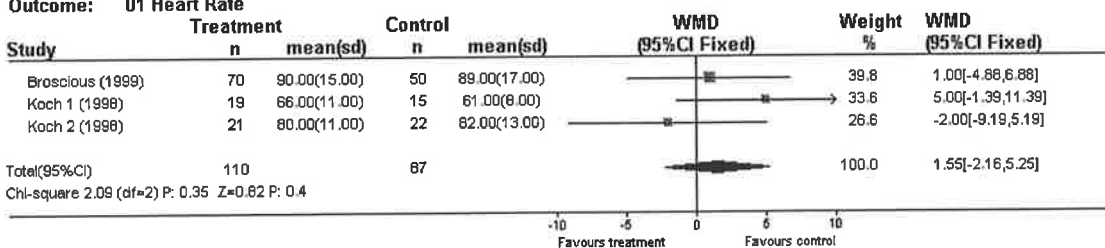


This meta-analysis showed no difference in heart rate between the music and no-music groups (WMD -4.97; 95%CI -10.11, 0.17). Of the RCTs not included in the meta-analysis because of missing data, two studies showed a reduction in heart rate in the music group (Chlan 1995; Miluk-Kolasa et al. 1996) and one found no difference (Barnason et al. 1995). Although some results are contradictory, on available evidence, it would appear that music has no impact on the heart rate of Hospital Patients.

Five RCTs evaluated music for Procedure Patients. Three studies evaluating music during chest tube removal (Broscious 1999), urological procedures (Koch et al. 1998, study 1) and renal lithotripsy (Koch et al. 1998, study 2) provided sufficient data and so were combined in a meta-analysis (see graph 2).

### Graph 2 Procedure Patients Music and Heart Rate

Comparison: 02 Vital Signs - Procedure Patients  
Outcome: 01 Heart Rate



Once again there was no difference in heart rate between the two groups (WMD 1.55; 95%CI -2.16, 5.25). Of the RCTs involving Procedure Patients that were not included in the meta-analysis because of missing data, one study involving the playing of music during sigmoidoscopies showed a reduction in heart rate in the music group (Palakanis et al. 1994), while another study involving patients undergoing intra-operative cataract surgery found no difference (Cruise et al. 1997). These studies suggest that music has no impact on the heart rate of patients during invasive or unpleasant procedures.

In summary, and based on a limited evidence, the use of music appears to have little impact on the heart rate of both Hospital Patients and Procedure Patients.

### Systolic Blood Pressure

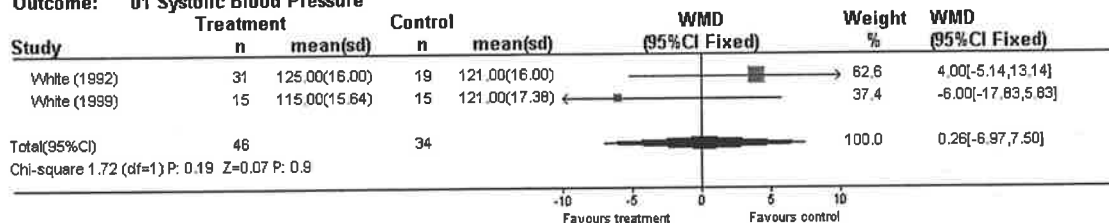
Eight RCTs used systolic blood pressure (SBP) as an outcome measure (see appendix 4, table 1 and 2). Four studies involved Hospital Patients, however only two provided sufficient data to combine in a meta-analysis (see graph 3). The participants of these two studies were:

- 40 patients post myocardial infarction (White 1992)
- 30 patients post myocardial infarction (White 1999a).

**Graph 3**  
Hospital Patients  
Music and Systolic Blood Pressure

Comparison: 03 Vital Signs - Hospital Patients

Outcome: 01 Systolic Blood Pressure

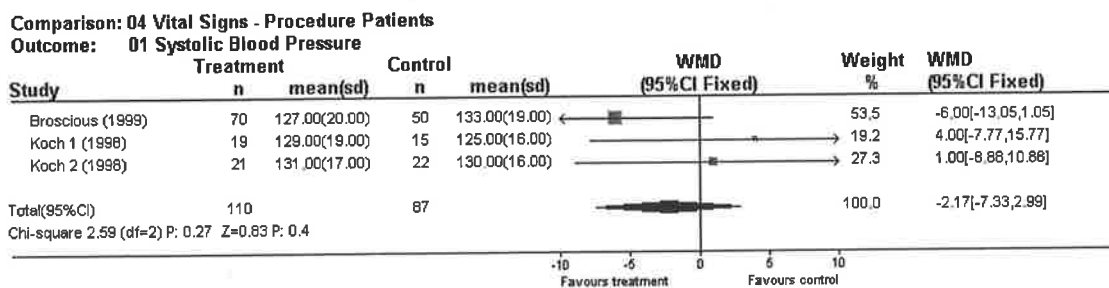


This meta-analysis found no difference in the SBP between the two groups (WMD 0.26, 95%CI -6.97, 7.50). Two other studies involving Hospital Patients that were excluded from the meta-analysis because of missing data, also found no difference in SBP between groups (Barnason et al. 1995; Chlan 1995).

Of the four studies involving Procedure Patients, only three provided sufficient data to be combined in a meta-analysis (Koch et al. 1998, studies 1 and 2; Broschious 1999). This analysis also found no difference in SBP between the groups (WMD -2.17 95%CI -7.33, 2.99) (see graph 4). Interestingly, the fourth study involving patients undergoing cataract surgery reported an increased SBP in the music group (Cruise et al. 1997). The participants of the three RCTs included in the meta-analysis were:

- 120 patients during chest tube removal (Broschious 1999)
- 34 patients during urological procedures (Koch et al. 1998, study 1)
- 43 patients during lithotripsy (Koch et al. 1998, study 2)

**Graph 4**  
**Procedure Patients**  
**Music and Systolic Blood Pressure**



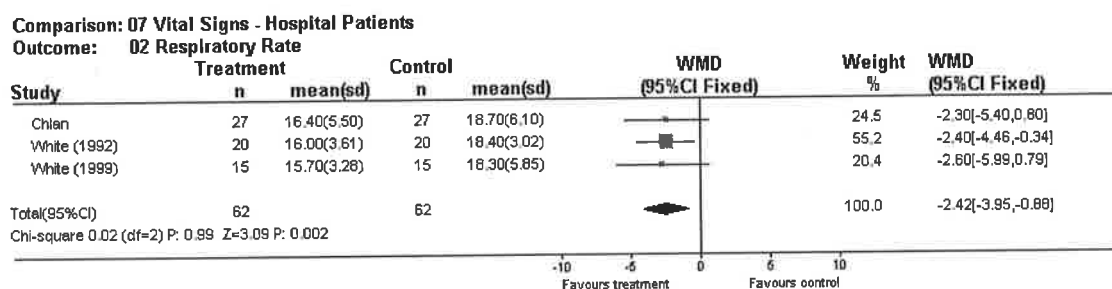
In summary, in these studies music had no impact on the SBP of Hospital Patients or Procedure Patients.

### Respiratory Rate

Five RCTs used respiratory rate as an outcome measure in Hospital Patients, however only three provided sufficient data to be included in a meta-analysis (White 1992; Chlan 1998; White 1999a) (see graph 5). The participants of the three RCTs included in the meta-analysis were:

- 54 mechanically ventilated patients (Chlan 1998)
- 40 patients post myocardial infarction (White 1992)
- 30 patients post myocardial infarction (White 1999a)

**Graph 5**  
Hospital Patients  
Music and Respiratory rate



These three studies showed a reduction in the respiratory rate of the music group as compared with that of the control group (WMD -2.42; 95%CI -3.95, -0.88) A fourth study in hospital patients also reported a reduction in the respiratory rate of the treatment group (Chlan 1995). However, this reduced respiratory rate was only 2 to 3 breaths per minute and so the clinical significance of this reduction is not clear.

For Procedure Patients, only a single RCT was identified that used respiratory rate as an outcome measure. This study involving intra-operative patients found no difference in respiratory rate between groups (Cruise et al. 1997).

In summary, four studies suggest the use of music produces a small reduction in the respiratory rate of Hospital Patients. Based on a single study, music has no effect on the respiratory rate of Procedure Patients.

## **Pain**

The evaluation of the impact of music on pain was determined by the use of two outcomes; a) patient's perception of the severity of pain, and b) the amount of analgesia required.

### **a. Severity of Pain**

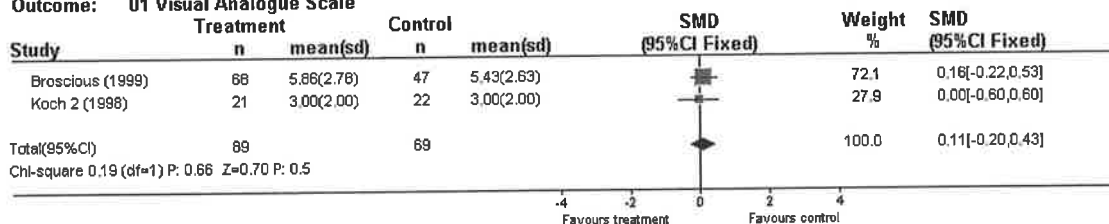
Only one RCT evaluated the impact of music on the perceived pain of Hospital Patients. This study of females during the post anaesthetic care following abdominal hysterectomy found no difference in the pain scores between groups (Taylor et al. 1998).

The meta-analysis of two studies evaluating the impact of music on Procedure Patients' assessment of pain using a visual analogue scale (VAS) (Koch et al. 1998, study 2; Broscious 1999) found no difference in pain scores between groups (SMD 0.11; 95%CI -0.20, 0.43) (see graph 6). The participants of the two RCTs included in the meta-analysis were:

- 115 patients during chest tube removal (Broscious 1999)
- 43 patients during lithotripsy (Koch et al. 1998, study 2)

**Graph 6**  
**Procedure Patients**  
**Music and Patients Rating of Pain**

Comparison: 05 Pain - Procedure Patients  
 Outcome: 01 Visual Analogue Scale



A third RCT which evaluated the impact of music on Procedure Patients provided insufficient data to be included in the meta-analysis (Good 1995). This study evaluated the effect of music played prior to the first post-operative ambulation on patients' perception of pain and also found no difference in the rating of pain between the two study groups.

**b Analgesic Use**

No RCTs evaluated the effect of music on the use of analgesics by Hospital Patients.

Two studies evaluated the impact of music on the amount of analgesia used by Procedure Patients during lithotripsy (Koch et al. 1998, study 2) and intra-operative cardiac surgery (Blankfield et al. 1995), but because of significant heterogeneity of results ( $\chi^2 = 45.59$ ,  $p < 0.05$ ), the use of meta-analysis was not appropriate. One of these studies reported no difference between groups (Blankfield et al. 1995), while the other reported significantly less narcotic analgesia administered during the procedure via a patient controlled device in the music group (Koch et al. 1998). It is likely that this difference in results may be a consequence of the different procedures the patients were undergoing at the time of the music intervention, and the different



analgesic used in the studies. Currently this is an area that requires further investigation.

In summary, the findings of these studies suggest that music has no effect when patients are asked to think about, and rate, the severity of their pain. However, while the evidence is limited and at times contradictory, music may be an effective diversion as demonstrated by the reduction in the use of analgesic in one study. This is an area that requires further investigation.

### **Sedation**

No RCTs evaluated the impact of music on the use of sedatives by Hospital Patients.

Two RCTs evaluated the impact of music on the sedation needs of Procedure Patients. Meta-analysis was not possible because of differences in the reporting of results. One study found the music group was administered significantly less sedative via a patient controlled device (mean total = 17 mg. propofol) than the no-music group (mean total = 94 mg. propofol) (Koch et al. 1998, study 1). The second study found significantly fewer patients asked for sedatives in the music group (4 of 32 patients) than in the no-music group (14 of 32 patients) (Walther-Larsen et al. 1988). However in this study, while fewer patients in the music group asked for sedatives, a greater number reported being anxious during surgery (12 of 32 patients) than in the no-music group (5 of 32 patients).

In summary, while evidence is limited, findings suggest music may act as a distracter as demonstrated by a reduced need for sedation during unpleasant procedures. However, current evidence is both limited and contradictory. This is an area that requires further investigation.

## **Sub-group Analyses**

As stated in the review protocol, studies were grouped into their specific populations for further analysis of the effectiveness of music within these groups. Studies were grouped according to the following headings:

- pre-operative patients
- intra-operative patients
- post-operative patients
- cancer patients
- myocardial infarction patients

The outcome measures of interest in these sub-group analyses were:

- heart rate,
- blood pressure,
- respiratory rate,
- pain, and
- sedation.

### **1. Pre-operative Patients**

#### **Heart Rate**

Only one included RCT evaluated the impact of music when pre-operative patients were told of their impending surgery (Miluk-Kolasa et al. 1996). This study found the heart rate of the music group returned to baseline levels more rapidly than the no-music group. One RCT, that was excluded because of attrition bias, evaluated the impact of music pre-operatively and found it had little impact on heart rate (Winter et al. 1994).

### Blood Pressure

One RCT evaluated the impact of music on blood pressure when pre-operative patients were told of their impending surgery and found the blood pressure of the music group returned to baseline levels more rapidly than the no-music group (Miluk-Kolasa et al. 1996). One RCT, that was excluded because of attrition bias, found music had no impact on the blood pressure of pre-operative patients (Winter et al. 1994).

### Respiratory Rate

No studies were identified that evaluated the impact of music on the respiratory rate of pre-operative patients.

### Pain

No studies were identified that evaluated the use of music and pre-operative pain. However, it is anticipated that for the majority of patients, pain will be a post-operative issue.

### Sedation

No studies were identified that evaluated the use of music and pre-operative sedation.

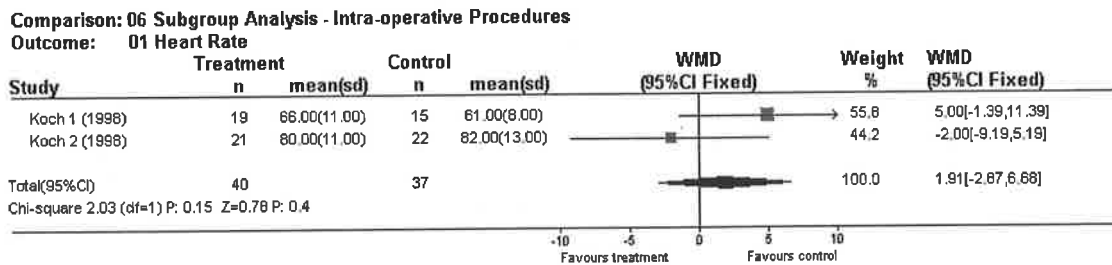
In summary, there has been limited evaluation of the use of pre-operative music. A single study suggests the use of music following patients being told about the impending surgery helps heart rate and blood pressure return to baseline levels more rapidly than when no music is played. There is little other information currently available in this area.

## 2. Intra-operative Patients

### Heart Rate

Two RCTs evaluated heart rate during surgical procedures (Koch et al. 1998, study 1 and study 2). Neither study demonstrated a difference in heart rate between the music and no-music groups (WMD 1.91; 95%CI -2.87, 6.68) (see graph 7).

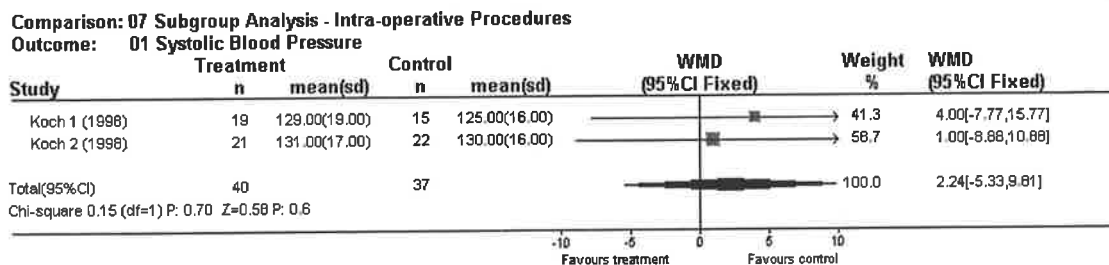
**Graph 7**  
Sub-group Analysis: Intra-operative Procedures  
Music and Heart Rate



### Blood Pressure

Two RCTs evaluated systolic blood pressure during surgical procedures (Koch et al. 1998, study 1 and study 2), however neither study demonstrated a difference in heart rate between groups (WMD 2.24; 95%CI -5.33, 9.81) (see graph 8).

**Graph 8**  
Sub-group Analysis: Intra-operative Procedures  
Music and Systolic Blood Pressure



### Respiratory Rate

A single RCT evaluating the impact of music on respiratory rate during cataract surgery found no difference between groups (Cruise et al. 1997).

### Pain

One RCT evaluated the effect of music on perceived pain during surgical procedures (Koch et al. 1998, study 2). This study found no difference in pain scores between the music and no-music group.

Two RCTs evaluated the impact of music on the amount of analgesic used by patients during surgical procedures. However, results were contradictory with one study showing a significant reduction in the amount of analgesia used by the music group (Koch et al. 1998, study 2) while the other showed no difference between groups (Blankfield et al. 1995).

### Sedation

Two RCTs evaluated the impact of music on sedative use during surgical procedures (Walther-Larsen et al. 1988; Koch et al. 1998, study 1). Both studies demonstrated a significant reduction in the use of sedatives by patients in the music group.

In summary, two studies demonstrated a significant reduction in the use of sedatives intra-operatively when music is played. One study suggests music also reduces analgesic use, however this was not supported by a second study. Music was not shown to have an impact on vital signs or the severity of pain during intra-operative procedures.

### 3. Post-operative Patients

#### Heart Rate

A single RCT evaluating the impact of music on the heart rate of post-operative patients found no difference between groups (Barnason et al. 1995).

#### Blood Pressure

A single RCT found no difference in SBP between the two groups of post-operative patients (Barnason et al. 1995).

#### Respiratory Rate

No RCTs were identified that evaluated the impact of music on the respiratory rate of post-operative patients.

#### Pain

A single RCT evaluating the impact of music on the perceived pain of patients following an abdominal hysterectomy found no difference in pain scores (Taylor et al. 1998).

#### Sedation

No RCTs were identified that evaluated the impact of music on post-operative sedative use.

In summary, while there is very limited information, it appears that music has no impact on the heart rate, blood pressure or severity of pain of post-operative patients.

#### 4. Cancer Patients

##### Heart Rate

No RCTs were identified that evaluated the impact of music on the heart rate of cancer patients.

##### Blood Pressure

No RCTs were identified that evaluated the impact of music on the blood pressure of cancer patients.

##### Respiratory Rate

No RCTs were identified that evaluated the impact of music on the respiratory rate of cancer patients.

##### Pain

One RCT was identified that evaluated the impact of music on pain in cancer patients, but was excluded because the music was evaluated over the course of a series of sessions (Beck 1991).

##### Sedation

No RCTs were identified that evaluated the impact of music on the use of sedatives by cancer patients.

## Discussion

One RCT evaluated the impact of music on anxiety and side effects associated with chemotherapy (Sabo and Michael 1996). This study was excluded from the meta-analyses because of an inappropriate method of randomisation, and the fact that a second intervention was combined with the music (personal message). This study found that music had no effect on the severity of side effects

Another RCT compared the use of music and the use of a sound consisting of a 60 second cycle hum, on pain in patients with cancer (Beck 1991). However this study failed to meet the inclusion criteria because the music intervention was repeated over a period of time rather than as a single intervention. This study found that while the effect of music on pain varied between individuals, there was a statistically significant decrease in pain from using either music or sound.

One RCT compared the impact of live music with taped on hospitalised adults with cancer (Bailey 1983), but failed to meet the inclusion criteria because of the use of live music. This study found that people who listened to live music reported significantly less tension and anxiety and more vigour than those who listened to taped music. A small non-randomised controlled trial evaluated the impact of music on nausea and vomiting, and findings suggested that music reduced the amount of nausea reported and that it also delayed the onset (Standley 1992).

In summary, there is little rigorous evidence on the effectiveness of music as an intervention for cancer patients. While the available studies provide an inadequate basis for clinical recommendations, they do suggest that there are some potential benefits that warrant further investigation. These areas in need of investigation include the use of music for cancer pain and for the reduction of side effects associated with the treatment of cancer.



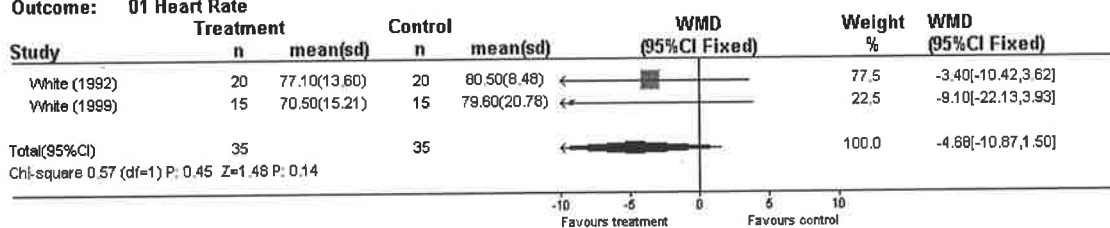
## 5. Post Myocardial Infarction

### Heart Rate

Two RCTs evaluated the impact of music on the heart rate of hospitalised patients recovering from a myocardial infarction (White 1992; White 1999a) (see graph 9). The findings of the meta-analysis suggest that music has little impact on the heart rate of this group of patients (WMD -4.68; 95%CI -10.87, 1.50).

**Graph 9**  
Sub-group Analysis: Post Myocardial Infarction  
Music and Heart Rate

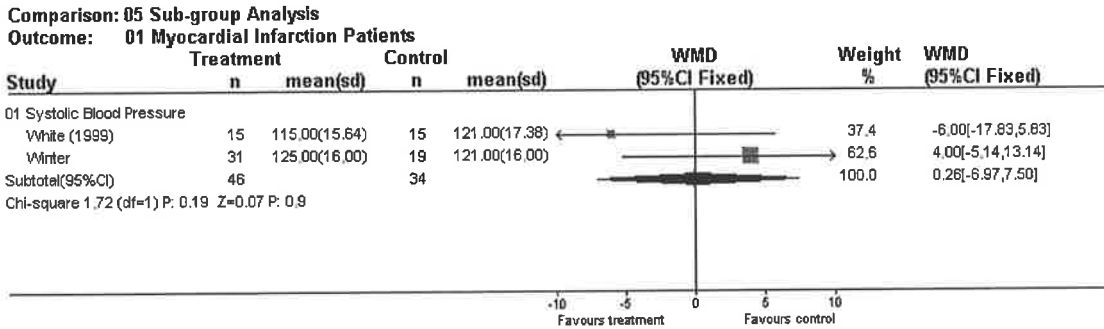
Comparison: 08 Subgroup Analysis - Post Myocardial Infarction  
Outcome: 01 Heart Rate



### Blood Pressure

Two RCTs evaluated the impact of music on the systolic blood pressure of patients recovering from a myocardial infarction (White 1992; White 1999a). This meta-analysis did not show any difference in blood pressure between groups (WMD 0.26; 95%CI -6.97, 7.50) (see graph 10).

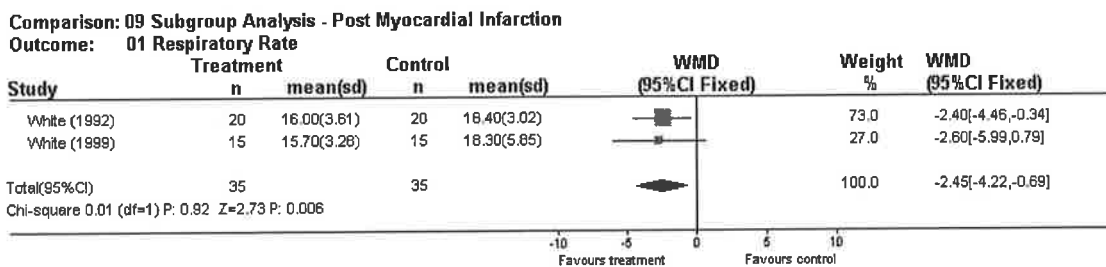
**Graph 10**  
**Sub-group Analysis: Myocardial Infarction Patients**  
**Music and Systolic Blood Pressure**



Respiratory Rate

Two RCTs evaluated the impact of music on the respiratory rate of patients recovering from a myocardial infarction (White 1992; White 1999a). This meta-analysis suggests music produced a small reduction in the respiratory rate of this group of patients (see graph 11).

**Graph 11**  
**Sub-group Analysis: Post Myocardial Infarction**  
**Music and Respiratory Rate**



Pain

No RCTs were identified that evaluated the impact of music on the pain of myocardial infarction patients.

## Sedation

No RCTs were identified that evaluated the impact of music on the use of sedation in myocardial infarction patients.

In summary, while music has not been shown to have any impact on heart rate and blood pressure of patients post myocardial infarct, it does produce a small reduction in respiratory rate.

## **Summary of Effectiveness of Music**

In summarising the studies addressing the effectiveness of music, and grading the strength of these findings according to the hierarchy of evidence previously described, the following can be stated.

### Hospital Patients

There is Good to Excellent evidence to demonstrate that music:

- has no impact on heart rate or systolic blood pressure of hospital patients, and
- produces a small reduction in respiratory rate in hospital patients.

Based on very limited evidence, there are suggestions that music:

- does not alter rating of the severity of pain by hospital patients.

### Procedure Patients

There is Good to Excellent evidence to demonstrate that music:

- has no impact on heart rate or systolic blood pressure of patients undergoing invasive or unpleasant procedures,

- has no impact on the respiratory rate of patients undergoing invasive or unpleasant procedures.

Based on very limited and at times contradictory evidence, there are suggestions that, music:

- may reduce the need for sedation during invasive and unpleasant procedures,
- may reduce the need for analgesia during invasive and unpleasant procedures, and
- does not alter the rating of the severity of pain by patients during invasive and unpleasant procedures.

### Sub-groups

There is a range of evidence related to the use of music as an intervention for specific hospital populations that can be ranked as either Good or Excellent, and this is summarised below.

#### 1. In pre-operative patients when music is played:

- following patients being told about the impending surgery, results in the heart rate and blood pressure returning to baseline levels more rapidly.

#### 2. In intra-operative patients the use of music:

- has no impact on vital signs or the severity of pain during intra-operative procedures, and
- significantly reduces the use of sedatives intra-operatively.

#### 3. In post-operative patients the use of music:

- has no effect on post-operative vital signs, and
- has no effect on the rating of the severity of post-operative pain.

#### 4. In cancer patients:

- there is no evidence available that can be ranked as good or excellent.
5. In myocardial infarction patients the use of music:

- has no effect on the heart rate or blood pressure, and
- produces a small reduction in respiratory rate.

Based on very limited and at times contradictory evidence, there are suggestions that music:

- may reduce analgesic requirements intra-operatively,
- may reduce the severity of pain in cancer patients,
- may reduce the nausea associated with chemotherapy, or delay its onset.

There is no evidence on :

- the impact of music on the respiratory rate, pain and sedation use of pre operative patients,
- the impact of music on the respiratory rate and sedation use of post-operative patients,
- the impact of music on the heart rate, blood pressure, respiratory rate, pain and sedation use of cancer patients, and
- the impact of music on the pain and sedation use of myocardial infarction patients.

## **Systematic Review II - Appropriateness**

This section of the review addresses the psychosocial outcome measures and includes anxiety, tolerance of unpleasant procedures, mood and satisfaction. As a number of research designs can contribute valid evidence on the appropriateness of an intervention, this section summarises evidence generated by:

1. randomised controlled trials,
2. observational studies, and
3. interpretive studies.

As with effectiveness, results are presented for all studies addressing either Hospital or Procedure Patients and then the findings from specific sub-groups are presented.

These sub-groups include:

- pre-operative patients,
- intra-operative patients,
- post-operative patients,
- cancer patients, and
- myocardial infarction patients.

### **Randomised Controlled Trails**

#### **Anxiety**

Anxiety was the most common psychosocial outcome measure used in studies evaluating music in hospital patients. A total of twelve RCTs evaluated the impact of music on anxiety. While the STAI was the most common tool used to measure anxiety, a range of approaches were used (see table 9).

Table 9  
Scales Used in Studies to Measure Anxiety

	STAI	VAS	STAI & VAS	Questions
Barnason, 1995 (Barnason et al. 1995)	✓			
Bolwerk, 1990 (Bolwerk 1990)	✓			
Chlan, 1998 (Chlan 1998)	✓			
Colt, 1999 (Colt et al. 1999)	✓			
Cruise, 1997 (Cruise et al. 1997)			✓	
Gaberson, 1995 (Gaberson 1995)		✓		
Good, 1995 (Good 1995)	✓			
Koch, 1998 (Koch et al. 1998, study 2)	✓			
Palakanis, 1994 (Palakanis et al. 1994)	✓			
Walther-Larsen, 1988 (Walther-Larsen et al. 1988)				✓
White, 1992 (White 1992)	✓			
White 1999 (White 1999a)	✓			

(STAI = State Trait Anxiety Inventory, VAS = Visual Analogue Scale)

Six of these RCTs evaluated the use of music with Hospital Patients and were combined in a meta-analysis. The specific populations involved in these studies were:

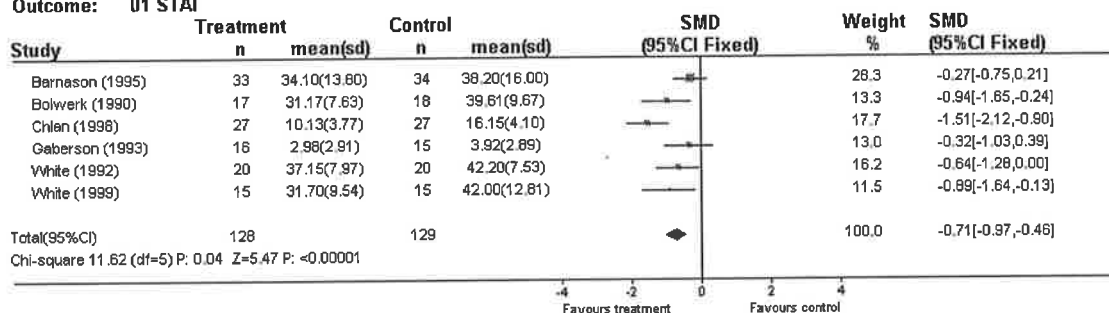
- post operative cardiac patients (Barnason et al. 1995)
- post myocardial infarction patients (Bolwerk 1990; White 1992; White 1999a)
- mechanically ventilated patients (Chlan 1998)

- pre-operative patients (Gaberson 1995)

The results of this meta-analysis (see graph 12), clearly demonstrate that music reduces anxiety of this group of patients (SMD -0.71; 95%CI -0.97, -0.46).

**Graph 12**  
Hospital Patients  
Music and Anxiety

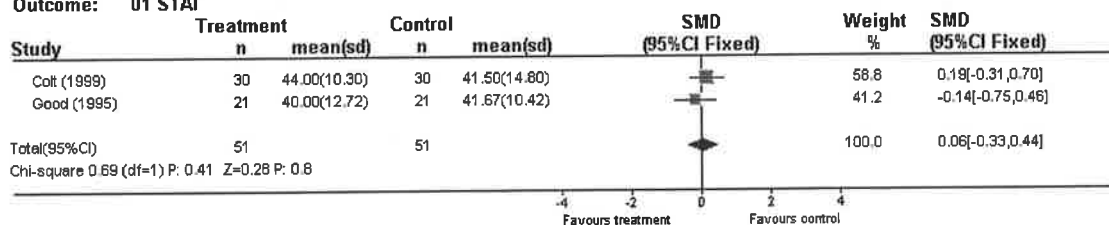
Comparison: 02 All Hostial Patients - Anxiety  
Outcome: 01 STAI



Six RCTs evaluated music in Procedure Patients. However only two studies involving patients undergoing a bronchoscopy (Colt et al. 1999) and patients during their first post-operative ambulation (Good 1995) provided sufficient data to allow meta-analysis. This meta-analysis showed music did not reduce anxiety (SMD 0.06; 95%CI -0.33, 0.44) (see graph 13).

**Graph 13**  
Procedure Patients  
Music and Anxiety

Comparison: 03 All Patients - Anxiety during Invasive Procedures  
Outcome: 01 STAI



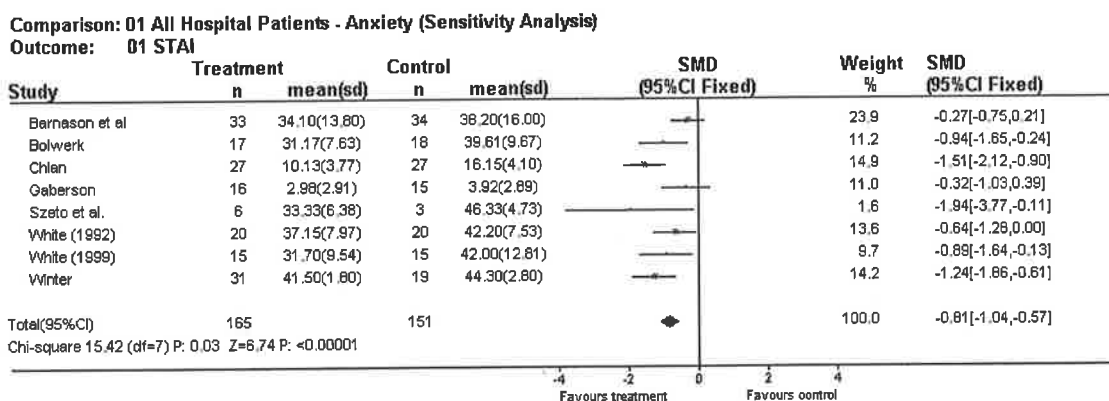


Of the remaining RCTs that could not be included in this meta-analysis because of missing data, two found no difference in anxiety between music and control groups (Cruise et al. 1997; Koch et al. 1998, study 2), a single study showed a reduction in anxiety in the music group (Palakanis et al. 1994), while another showed an increase in the anxiety of the music group in relation to the control group (Walther-Larsen et al. 1988). These findings suggest that music has no effect on the anxiety of patients undergoing invasive or unpleasant procedures.

### Sensitivity Analysis

To evaluate the impact of decisions made during the critical appraisal phase of the review, some excluded studies were added to the meta-analysis on anxiety to determine if their findings changed the results of the meta-analysis (see Graph 14). To the six studies involving Hospital Patients included in the original meta-analysis (see graph 12) a further two studies were added (Winter et al. 1994; Szeto and Yung 1999). These two studies had been excluded from the original meta-analysis because of attrition bias. A third excluded study had important differences in the treatment of the music and no-music groups and so was not included in this sensitivity analysis (Augustin and Hains 1996).

**Graph 14**  
Sensitivity Analysis - All Hospital Patients  
Music and Anxiety



This sensitivity analysis does not substantially change the findings of the first meta-analysis (graph 12), and continues to demonstrate that music produces a reduction in the anxiety of Hospital Patients (SMD -0.81; 95%CI -1.04, -0.57).

In summary, music is an effective intervention for reducing the anxiety of Hospital Patients. However it has little impact on the anxiety of patients undergoing unpleasant or invasive procedures.

### **Tolerance**

A single RCT evaluated the impact of music on Procedure Patients' tolerance of upper gastrointestinal investigations (Bampton and Draper 1997). Tolerance was rated by both patients and nursing staff using a VAS. The researchers reported no difference between groups for tolerance scores, although no data was presented. However, contradicting this finding, more patients in the no-music group rated the procedure as moderately unpleasant or worse (10 of 31 patients) than the music group (2 of 28 patients).

In summary, there is currently insufficient evidence to evaluate the impact of music on patients' tolerance of unpleasant procedures.

### **Satisfaction**

No studies were identified that evaluated the impact of music on the satisfaction of Hospital Patients.

Two RCTs involving Procedure Patients used satisfaction as an outcome measure to evaluate the impact of music during minor surgical procedures (Walther-Larsen et al. 1988; Cruise et al. 1997). These studies could not be combined in a meta-analysis as

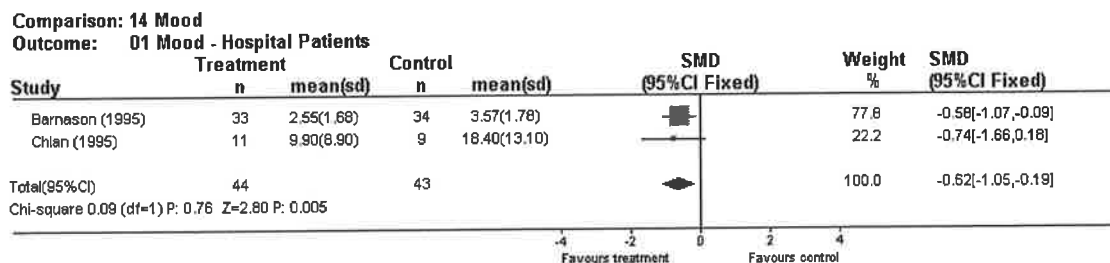
one study used dichotomous data (Walther-Larsen et al. 1988) and the other used continuous data [Cruise, 1997 #201. One study found that relaxing music was associated with the highest level of patient satisfaction (VAS score - music group = 92.0; control group = 82.2) (Cruise et al. 1997). The second study contradicted these findings, reporting no difference between groups (Satisfied - music group = 31 of 32, control group = 30 of 32) (Walther-Larsen et al. 1988).

In summary, there is currently insufficient evidence to evaluate the impact of music on patient satisfaction during unpleasant procedures and there is no evidence on whether it improves the satisfaction of Hospital Patients.

### Mood

Two RCTs involving post-operative cardiac patients (Barnason et al. 1995) and mechanically ventilated patients (Chlan 1995) evaluated the impact of music on patients' mood. One study used the Profile of Mood States (POMS) to measure mood (Chlan 1995). This scale is an instrument that measures negative mood states associated with psychological distress (McNair et al. 1992). The second study used a mood VAS, and the data was transformed to negative outcomes to perform the meta-analysis (see graph 15) (Barnason et al. 1995). The results of the meta-analysis demonstrates that playing music resulted in an improvement in patients' mood (SMD -0.62; 95%CI -1.05, -0.19).

**Graph 15**  
Hospital Patients  
Music and Mood



No studies were identified that evaluated the impact of music on the mood of Procedure Patients.

In summary, two RCTs demonstrated that the use of music as an intervention improves the mood of Hospital Patients, whether this also works for Procedure patients is not known.

### **Sub-group Analyses**

#### **1. Pre-operative Patients**

##### Anxiety

One RCT evaluated the impact of music on anxiety and found no difference between groups (anxiety VAS - music group = 2.98, control = 3.92) (Gaberson 1995). No other studies were identified investigating the use of music in pre-operative patients.

##### Tolerance

No RCTs evaluated the impact of music on the tolerance of pre-operative patients.

##### Satisfaction

No RCTs evaluated the impact of music on the satisfaction of pre-operative patients.

##### Mood

No RCTs evaluated the impact of music on the mood of pre-operative patients.

In summary, a single study found music had no impact on the anxiety of pre-operative patients. There is currently little evidence on the appropriateness of music in terms of its influence on the tolerance, satisfaction and mood of pre-operative patients.

## 2. Intra-operative Patients

### Anxiety

No RCTs evaluated the impact of music on anxiety during operative procedures.

### Tolerance

No RCTs evaluated the impact of music on the tolerance of patients of operative procedures.

### Satisfaction

Two RCTs evaluated the impact of music during minor surgical procedures (Walther-Larsen et al. 1988; Cruise et al. 1997). As previously discussed, one study found music was associated with the highest level of patient satisfaction (Cruise et al. 1997), while the second study found no difference between groups (Walther-Larsen et al. 1988).

### Mood

No RCTs evaluated the impact of music on the mood of patients during intra-operative procedures.

In summary, the only evaluations of music involving intra-operative patients focused on patient satisfaction, however the findings were contradictory. Currently there is little evidence available on the appropriateness of music during intra-operative procedures.

### 3. Post-operative Patients

#### Anxiety

A single RCT evaluated the impact of music on the anxiety of post-operative patients (Barnason et al. 1995). This study found no difference in the anxiety scores between the two study groups (STAI mean score - music group = 34.10; control group = 38.2).

#### Tolerance

No RCTs evaluated the impact of music on the tolerance of post-operative patients.

#### Satisfaction

No RCTs evaluated the impact of music on the satisfaction of patients during the post-operative period.

#### Mood

A single RCT evaluated the influence of music on the mood of post-operative cardiac patients (Barnason et al. 1995) This study, using a mood VAS, found that music significantly improved the mood of patients (Mood VAS mean score - music group = 7.45; control group = 6.43) (Barnason et al. 1995).

In summary, current evidence suggests that music does not reduce the anxiety of post-operative patients but can improve their mood. However, these findings are based on limited evidence.

#### 4. Cancer Patients

##### Anxiety

No RCTs evaluated the impact of music on the anxiety of cancer patients.

##### Tolerance

No RCTs evaluated the impact of music on the tolerance of cancer patients.

##### Satisfaction

No RCTs evaluated the impact of music on the satisfaction of cancer patients.

##### Mood

No RCTs evaluated the impact of music on the mood of cancer patients.

In summary, the appropriateness of music has not been evaluated in cancer patients.

#### 5. Post Myocardial Infarction

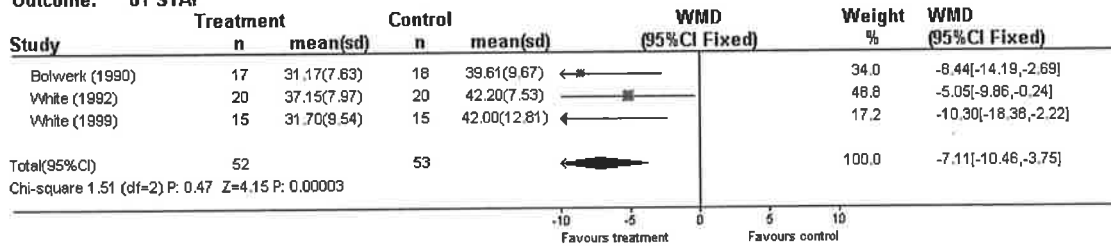
##### Anxiety

Three RCTs evaluated the impact of music on the anxiety of patients following a myocardial infarction (see graph 16) (Bolwerk 1990; White 1992; White 1999a). The meta-analysis clearly demonstrated the reduction in anxiety in the music group when compared to the control group (WMD = -7.11, 95%CI -10.46, -3.75).

## Graph 16 Myocardial Infarction Patients Music and Anxiety

Comparison: 15 Sub-group Analysis - Myocardial Infarction Patients

Outcome: 01 STAI



### Tolerance

No RCTs evaluated the impact of music on the tolerance of myocardial infarction patients.

### Satisfaction

No RCTs evaluated the impact of music on the satisfaction of myocardial infarction patients.

### Mood

No RCTs evaluated the impact of music on the mood of myocardial infarction patients.

In summary, in terms of the appropriateness of music for myocardial infarction patients, three RCTs demonstrated a reduction in anxiety. There is currently no evidence on its impact on the tolerance, satisfaction or mood in this sub-group of patients.



## **Observational Studies**

No observational studies addressing an aspect of the appropriateness of music were identified during the literature search.

## **Interpretive Studies**

An interpretive study involving structured interviews, and the use of an attitudinal scale, explored the impact of music played during surgical procedures with regional or local anaesthetic (Stevens 1990). This study found that 75% of participants rated the helpfulness of music as either 'very good' or 'excellent'. Specific statements from participants in support of music included (Stevens 1990, p. 1048):

"It takes your mind off what is going on."

"Before I had my eyes open but when the music came on I shut my eyes and went into my own little world."

In addition, participant comments also suggested that music relieved their anxiety and stress (Stevens 1990, p. 1048):

"I was rather distraught beforehand and the music certainly settled me down."

"It was very soothing. I could meditate."

Interestingly, this study also highlighted the importance of patient preference in music. Some participants indicated they would have liked to have brought their own music and three expressed disappointment at not having their requests (for a specific music selection) fulfilled (Stevens 1990). This finding is in contrast to some studies evaluating the effectiveness of music where selection of the music to be played was not an option during the study.

## **Summary of the Appropriateness of Music**

In summarising the studies addressing the appropriateness of music, and grading the strength of these findings according to the hierarchy of evidence previously described, the following can be stated.

### **Hospital Patients**

There is Good or Excellent evidence to demonstrate that:

- music reduces the anxiety of Hospital Patients, and
- music improves the mood of Hospital Patients,

There is no evidence on the impact of music on:

- the satisfaction of Hospital Patients.

### **Procedure Patients**

There is Excellent evidence to demonstrate that:

- music has no impact on patient anxiety during invasive or unpleasant procedures.

There is insufficient evidence to determine the impact of music on:

- patients' tolerance of unpleasant procedures, and
- patient satisfaction with invasive or unpleasant procedures.

There is no evidence on the impact of music on:

- the mood of patients during invasive or unpleasant procedures.

## Sub-groups

There is only limited evidence related to the use of music as an intervention for specific hospital populations that can be ranked as either Good or Excellent. This evidence is summarised below.

### 1. In pre-operative patients:

- based on the findings of a single RCT, music has no impact on the anxiety of pre-operative patients.

### 2. Intra-operative patients:

- there is no evidence available that can be ranked as good or excellent.

### 3. In post-operative patients:

- based on the findings of a single RCT, music does not reduce the anxiety of post-operative patients, and
- based on the findings of a single RCT, music can improve the mood of post-operative patients.

### 4. Cancer patients:

- there is no evidence available that can be ranked as good or excellent.

### 5. In myocardial infarction patients music:

- music can reduce the anxiety of myocardial infarction patients.

There is insufficient evidence to determine the impact of music on the satisfaction of patients during intra-operative procedures. The available evidence on the impact of music on the anxiety of intra-operative patients is contradictory

There is no evidence on:

- the impact of music on tolerance, satisfaction or mood of pre operative patients,
- the impact of music on the anxiety, tolerance or mood of intra-operative patients,

- the impact of music on the tolerance and satisfaction of post-operative patients,
- the impact of music on the anxiety, tolerance, satisfaction and mood of cancer patients, and
- the impact of music on the tolerance, satisfaction and mood of myocardial infarction patients.

Findings from a single interpretive study suggest that music reduces patients' intra-operative anxiety and is seen as a helpful intervention. This study also suggests that patient selection of music is preferable to selections made by hospital staff.

## **Systematic Review III - Feasibility**

This section of the review addresses research evaluating music from the perspective of the environment. The aim was to summarise research addressing issues such as the implementation of music in the hospital setting, the support required to achieve beneficial outcomes and other related influences. A number of research designs were considered to contribute valid evidence on the feasibility of music including:

1. randomised controlled trials,
2. observational studies, and
3. interpretive studies,

### **Randomised Controlled Trials**

No RCTs addressing the feasibility of music in the hospital setting were identified during the literature search.

### **Observational Studies**

No observational studies addressing the feasibility of music in the hospital setting were identified during the literature search.

### **Interpretive Studies**

No interpretive studies addressing the feasibility of music in the hospital setting were identified during the literature search.

## Other Types of Studies

The literature search failed to identify any studies specifically addressing the feasibility of music as an intervention. This may be a consequence of a range of factors such as:

- the relatively easy application of music,
- lack of harmful side effects,
- the acceptability of music to many people
- low cost (after initial purchase of equipment and music), and
- the common use of music in everyday life.

As a result of this lack of evidence on feasibility, other sources of information were summarised in an attempt to identify issues of importance or interest.

A survey of music therapists was used to investigate their use of music and found that of the 139 responders, 41% were using music for pain relief in clients (Michel and Chesky 1995). Of the music therapists using music, specific populations in which it is used included (Michel and Chesky 1995, p. 50):

- cancer patients (64%),
- the elderly (64%),
- hospice clients (21%),
- post-operative patients (12%),
- physically disabled clients (30%),
- neurologically impaired clients (21%), and
- obstetrics and gynaecology patients (3%).

The rationale given by music therapists for the use of music in pain relief included :

- relaxation (95%),

- distraction (91%),
- to change patients' mood (75%), and
- to stimulate imagery (30%) (Michel and Chesky 1995, p. 50).

In terms of the assessment of outcomes of those receiving music as part of their pain management, most reported using individuals' subjective response (31%) (Michel and Chesky 1995). Other methods of outcome measurement used by the music therapists included the use of observation of patient behaviours (12%) and the use of a visual analogue scale (8%).

One uncontrolled trial evaluated music in cancer patients undergoing chemotherapy (Weber et al. 1996). The researchers reported that after initial scepticism by staff, they accepted the use of music into normal hospital routine (Weber et al. 1996). Another paper discussed a 'Music During Surgery' program conducted at one institution (Pratt 1999). The aims of the Music During Surgery program were:

- to maintain tape players and a collection of varied styles of soothing music,
- to offer music to every patient admitted to the short stay surgery area, and
- to investigate other ways to promote the use of music by patients in surgery.

To assist in the follow-up of these patients, a sticker with a picture of an musical eighth note was placed on the patients' charts (Pratt 1999). This report also described patient information sheets and care plans on the use of music during same day surgery. However, there appears to have been no evaluation of this music program.

While there may be infection control issues related to the sharing of headphones between different patients this has not been investigated. One discussion paper suggested that the sharing of equipment may not be appropriate for patients on strict

isolation precautions (Chlan and Tracy 1999). Other than this discussion paper, no research evidence was identified addressing this issue.

### **Summary of Feasibility of Music**

There is currently no rigorous evidence related to the feasibility of music in the hospital setting.



## Integration of the Evidence

This review highlighted a considerable amount of evidence supporting the use of music for adults during hospitalisation. Most notable in these findings is that music is primarily beneficial for the psychosocial outcomes such as anxiety and mood. As a consequence of this effect, other outcomes such as the amount of sedatives used has also been reduced. However, music appears to have minimal impact on the physiological outcomes such as blood pressure and heart rate.

There are clear differences between people listening to music while resting in bed and those who were undergoing an unpleasant procedure. While some studies suggest beneficial outcomes for the Procedure Patients, such as a reduction in the need for analgesia and sedatives, further research is needed. In summarising the review findings, the evidence is graded according to the hierarchy of evidence proposed in this thesis.

### Hospital Patients

- Good to Excellent
- Reduces anxiety
  - Produces a small reduction in the patients' respiratory rate
  - Improves mood
  - Has no effect on heart rate or systolic blood pressure.

- Limited or Contradictory Evidence
- Does not alter the patients' rating of the severity of pain.

### Procedure Patients

- Good to Excellent
- Has no effect on anxiety.
- Limited or Contradictory
- May reduce the need for sedation during

- Evidence      invasive and unpleasant procedures,
- May reduce the need for analgesia during invasive and unpleasant procedures, and
  - Does not alter the rating of the severity of pain by patients during invasive and unpleasant procedures.

### **Integration of Sub-Groups Analysis**

Music can not be recommended for many of the specific sub-groups investigated during this review, but this is due to inadequate evaluation of music rather than a demonstrated lack of efficacy. However, there is Good to Excellent evidence to support the following statements:

1.      Pre-operative Patients

- Following explanations about impending surgery, listening to music results in the patients' heart rate and blood pressure returning to baseline levels more rapidly.
- Music has no impact on the anxiety of pre-operative patients.

2.      Intra-operative Patients

- Listening to music during surgery significantly reduces the amount of sedatives patients use.
- Listening to music has no impact on vital signs or the severity of pain during intra-operative procedures.

3.      Post-operative Patients

- Listening to music post-operatively improves the patient's mood.
- Listening to music has no impact on post-operative vital signs.

- Listening to music has no impact on the rating of the severity of post-operative pain.

4. Cancer Patients

- No evidence.

5. Myocardial Infarction Patients

- Listening to music produces a small reduction in the respiratory rate of myocardial infarction patients.
- Listening to music reduces the anxiety of myocardial infarction patients.
- Listening to music has no impact on the heart rate or blood pressure of myocardial infarction patients.

There is little research based evidence available on issues related to the implementation and support of music as an intervention in the hospital setting.

### **Future Research**

As many of these findings are based on limited evidence, further replication of studies is needed to fully evaluate the effectiveness of music. Additionally, the use of music in specific populations also requires further investigation. While this review focused on music delivered as a single episode of care, there appears to have been only very limited evaluation of the impact of music when provided on a daily or ongoing basis. This would suggest that further exploration into the potential cumulative effects of music during hospitalisation is also warranted.

# **CHAPTER 7**

## **Comparison Review: Physical Restraint**

## **Physical Restraint Expanded Systematic Review**

### **Introduction**

To fully evaluate all aspects of the conceptual framework and expanded review process a number of systematic reviews using this approach would be needed. This is because it is unlikely that a single topic would have the diversity of research required to utilise all components proposed in this thesis. Therefore to facilitate a more extensive evaluation of the proposed methods, a second review is presented.

This review was conducted concurrently with the work reported in this thesis and so served as a practical test during the development of the review methods. The aims of this chapter are twofold. Firstly, it provided an opportunity to examine some of the proposed expanded review methods that were not evaluated during the music review because of the limited breadth of research. Secondly, these examples provide a demonstration of the use of the proposed methods for summarising the findings of a number of different types of research. The intent of this chapter is not to present the entire results of this review, rather selected examples are used to support the proposed methods.

Physical restraint review was chosen as the second demonstration review because, unlike music in hospitals, it has had little evaluation using RCTs, and existing research has utilised a broad range of research methods. The focus of interest in presenting these restraint review examples is both the processes that were used to pool results and the product of the process. However, because this review was conducted concurrently with the development of the expanded review process, the review is currently unpublished.

## **Overview of Physical Restraint**

This systematic review aimed to provide a rigorous summary of research addressing the use of physical restraint in the acute and residential care setting. However, unlike many interventions, the issues of importance related to restraint devices do not concern the effectiveness.

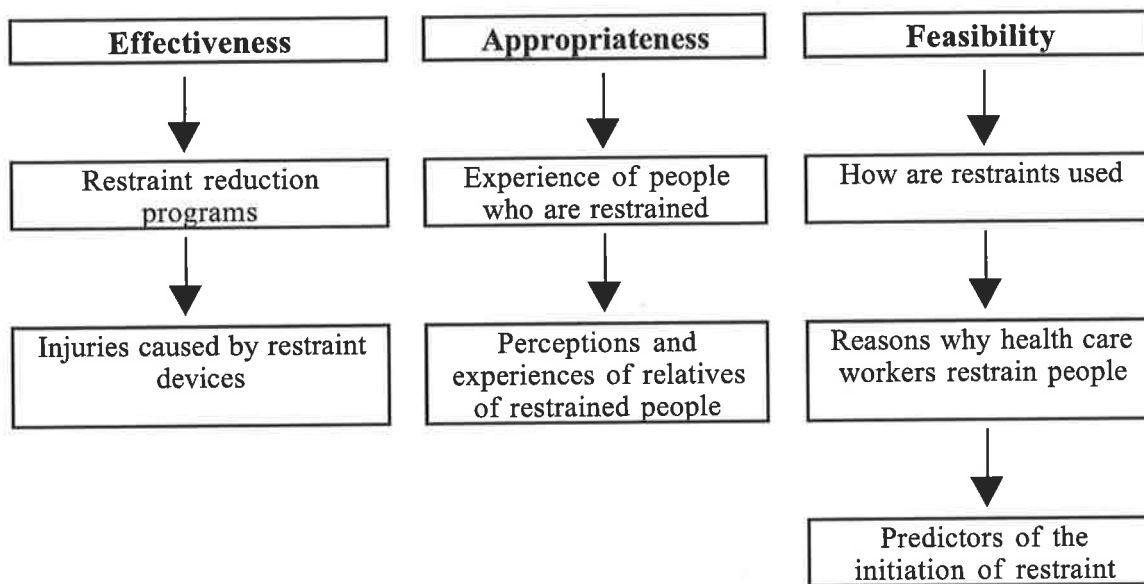
The major focus in the health literature has been restraint minimisation and whether this is accompanied by an increase in adverse events that restraints are intended to prevent. However, a considerable body of literature suggests that the use of restraints can cause a number of adverse events such as increases in mortality, length of hospitalisation and pressure areas. This highlights the contradictions accompanying the use of physical restraint, that they aim to prevent adverse events but this may be at the cost of other negative outcomes. As a consequence of these issues, the restraint review sought to summarise the effectiveness of restraint minimisation programs and investigate injury caused by these devices.

Another area of interest in the health care literature has been the experiences and perceptions of the people who are subject to restraint and those of their family. While a number of studies have investigated this issue, because of the methods used, they have remained outside of the current health care debate. A second focus of the restraint review was therefore to summarise the impact of physical restraint on people.

Other studies have provided a description of the current practice of restraining people in health care facilities, but once again this information has had minimal impact on the professional debate. These studies have investigated issues such as what devices are used, the duration of restraint, what characteristics increase the risk of being restrained, and why people are restrained. Therefore the third component of this review was to determine how restraints have been used.

The conceptual framework presented in this thesis was used to enable this diverse body of literature to be logically grouped and categorised (see figure 5). The expanded review process was used to provide the methods by which this research could be summarised and synthesised.

**Figure 5**  
**Systematic Review on Physical Restraints**



The conceptual framework for the restraint review addressed effectiveness, appropriateness and feasibility, allowing research that addressed similar issues to be grouped. While this research utilised a range of different research designs, the common theme across all studies was that they all contributed evidence to one of these three framework perspectives. As previously acknowledged, this type of research could not be synthesised in one grand meta-analysis, and so a series of discrete focused reviews was undertaken. To demonstrate the value of this approach, selected examples from each of the three perspectives are presented. The purpose in presenting these examples is to provide a demonstration of the methods used to pool the findings of individual studies and the value and usefulness of the product of this activity.

## Physical Restraint Review Results

### Effectiveness

Effectiveness in the context of physical restraint relates to restraint minimisation programs. Outcomes of interest for effectiveness were the reduction in the number of people restrained and changes in the number of adverse outcomes, such as falls and injury.

### Restraint Minimisation Programs

This component focused on whether restraint minimisation programs were effective. These programs have generally been implemented across organisations, in both acute and residential care settings, to reduce the number of people restrained. In conjunction with this, these programs also aimed to prevent increases in adverse events, such as falls. The optimal methodological approach for the evaluation of the effectiveness of this type of program is the RCT. However, unlike the systematic review of music, only a single RCT was identified that evaluated a restraint minimisation program. This study demonstrated that an education program supported by unit-based consultation by a specialist gerontological nurse effectively reduced restraint use by 56% in the residential care setting (Evans et al. 1997). However, it is not clear whether the findings of this single study can be generalised to all other residential care and acute care facilities. From this perspective, the findings of a systematic review focusing only on RCTs would have little to offer clinical practice regarding the minimisation of physical restraint.

However, other studies have also evaluated restraint minimisation programs using historical-comparison groups. This 'before and after' research design helped overcome many of the difficulties in attempting to randomise participants to study groups when the intervention was conducted across an entire organisation. While this design is at



greater risk of error than the RCT, has been argued in this thesis that consistent results from a number of studies suggest that the intervention is likely to be effective.

To evaluate the findings of these before and after studies, all results were summarised in a table (see Table 10). In this example of three of these studies, restraint minimisation programs are shown to consistently reduce the proportion of residents subject to restraint. Importantly, these studies also showed that falls and falls related injury do not increase in response to this physical restraint reduction. These findings are supported by many other similar studies which were also summarised in the restraint review. The conclusions drawn from this evidence is that physical restraints can be safely reduced.

**Table 10**  
**Restraint Minimisation Programs in the Residential Care Setting**  
 (Summary table of a sample of before and after studies )

**Residential Care Setting**

Citation	Duration of Evaluation	Intervention	Change in Restraint Rate	Change in Fall Rate	Change in Injury Rate
Kramer (Kramer 1994)	12mths	- Organised by committee. - Education program. - Clinical support. - Started with easy cases first.	57% to 10%	↔	↔
Levine (Levine and Marchello 1995)	3 years	- In-service education - Policy change - Procedural innovation	39% to 4%	↔	↔
Neufeld (Neufeld and Libow 1999)	2 year	- Education program across 16 nursing homes. - Telephone & on-site clinical consultation.	41% to 4.05%	not reported	↓ (variable across sites)

(↔ = no change, ↑ = increase, ↓ = decrease)

Another aim of the restraint review was to provide a summary of research. Therefore this review also documented the different restraint minimisation programs reported in the literature, creating a record of the structure of programs (see Table 11). This summary provided a stocktake of all published programs and so provides important information for clinicians attempting to initiate similar programs within their organisation.

**Table 11**  
**Stocktake of Restraint Minimisation Programs**  
 (One example of the stocktake summary data)

Citation	Evans (Evans et al. 1997)
Year	1997
Type of Study	RCT (three study groups)
Setting	Nursing homes
Population	643 residents
Structure of Program	Two Education interventions groups (plus one control group) 1) 6 month comprehensive Restraint Education Program (RE) - 30 to 45 minute education session Aimed at - highlighting hazards, - resident assessment - alternative interventions for specific behaviour problems 2) Restraint Education Program - plus unit based consultation (REC) - consultation provided by Gerontological Nurse Specialist 3) Control - no education or consultation
Outcome Measures	- Restraint status (at start, 6 mths, 9 mths & 12 mths) - Serious injuries, psychoactive drug use, staff levels
Results	- Only small difference between RE and Control - Significant reduction in restraint use in education & consultation nursing home Over all decline in restraint use (over 12 mths) were: - Control 11% decline - RE 23% decline - REC 56% decline No increase in psychoc-active drug use, staffing levels or injuries
Conclusion	Education program combined with unit based, resident centred consultation by a Gerontological Nurse Specialist can reduce the use of physical restraints over a 12 month period.

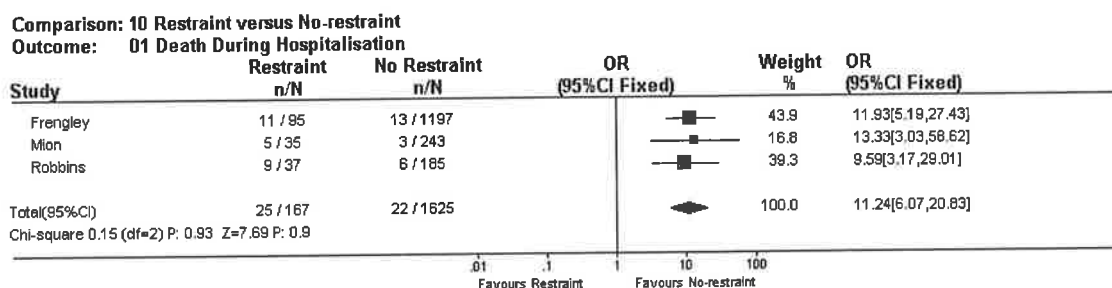
### Injury Caused by Restraints

Studies have also reported injury as a consequence of people being physically restrained, and this was therefore one of the areas of physical restraint addressed by the review. This component of effectiveness highlighted the importance of preventing

adverse outcomes in addition to producing positive outcomes. A number of studies using a range of research designs contributed evidence to this area of the review. Descriptive studies, such as case reports and investigations of Coroner's reports, highlighted the risks of strangulation (Dube and Mitchell 1986), nerve damage (Scott and Gross 1989), ischaemic injury (McLardy-Smith et al. 1986) and sudden death (Robinson et al. 1993). While a narrative summary of this type of evidence would commonly be ranked at the lowest level in terms of its contribution to the evaluation of effectiveness, it provides vital information regarding the negative consequences of physical restraint.

Evidence on injury was also been generated by observational studies. These studies investigated such aspects as changes in mortality, pressure sores and length of hospitalisation for patients who were restrained. These findings were pooled in a meta-analysis to determine the odds of adverse events occurring in restrained patients compared to those unrestrained (see Graph 17). The findings of this meta-analysis suggest that restrained patients are more likely to die during their hospitalisation than those who are not restrained.

**Graph 17**  
Association Between Restraint and Death



This evidence on injury provides important information to guide practice. As this review clearly demonstrates the use of physical restraint brings with it a real risk of a number of poor outcomes, it lends support to the need to minimise the use of these

devices. Despite the value of this evidence, it would normally be beyond the scope of systematic review.

### **Appropriateness**

Appropriateness as described in the conceptual framework relates to the psychosocial impact of physical restraints. This aspect of restraint concerns both the person being restrained and their family.

#### **Experience of Being Restrained**

Studies have explored the experience of being physically restrained in a hospital or nursing home. While this information does not address effectiveness, it offers a valid contribution to the professional debate and helps inform clinical practice through the insights provided. This experience was investigated by interpretive studies and so these were summarised using qualitative content analysis (see Table 12). In this example from the acute care setting, the use of content analysis allowed the themes to be categorised and grouped. Through this process, the negative impact of restraint was clearly demonstrated.

**Table 12**  
Patient Reactions to Restraint  
(qualitative content analysis)

**Negative Responses**

<b>Loss of Control</b>	"Like a caged bird, like I'm in a jail, stuck." (Hardin and Magee 1993)
	"The hospital is worse than a jail." (Strumpf and Evans 1988)
	"Just let me have my freedom." (Hardin and Magee 1993)
	"I can't give orders now, only take them" (Hardin and Magee 1993)
	"After a while I gave up, I became a mouse." (Strumpf and Evans 1988)
	"It's horrifying. It is just like being harnessed up like a mule." (Hardin and Magee 1993)
<b>Restriction</b>	"I can't get out of the chair." (Hardin and Magee 1993)
	"I can't turn over in bed." (Hardin and Magee 1993)
	"I couldn't move at all to do what I wanted or needed. I couldn't even bring my hands together."(Strumpf and Evans 1988)
	"I can't get to the toilet." (Hardin and Magee 1993)
<b>Discomfort</b>	"They are too tight." (Hardin and Magee 1993)
	"I feel it across my chest and I feel pain."(Strumpf and Evans 1988)
<b>Anger</b>	"I have done nothing to deserve this." (Strumpf and Evans 1988)
	"That's for crazy people, I was never like that. You must be mistaken, maybe the nurse had me mixed up with someone else." (Strumpf and Evans 1988)
<b>Humiliation</b>	"I felt like a dog and I cried all night."(Strumpf and Evans 1988)
	"Embarrassed." (Hardin and Magee 1993)

**Perceptions of Relatives**

Some studies explored the perceptions of people when a family member is restrained during admission to a hospital or a nursing home. Once again, this information was generated by interpretive research and studies were summarised using qualitative content analysis (see Table 13). In this example, negative responses related to restraint were also identified, and this information clearly highlighted a perspective of physical restraint not normally found in systematic reviews.

**Table 13**  
**Family Reactions to Restraint**  
 (qualitative content analysis)

**Negative Responses**

<b>Anger</b>	"I'm angry over the restraint order." (Kanski and Janelli 1996)
	"I take that thing off and throw it in the draw." (Newbern and Lindsey 1994)
	"Appalled" (Hardin and Magee 1993)
	"It made me mad —they just walked in, put on the restraint and never said a word." (Kanski and Janelli 1996)
<b>Disagreement</b>	"I intended to make sure my mother was not restrained." (Newbern and Lindsey 1994)
	"I cried, then felt guilty." (Hardin and Magee 1993)
	"She didn't need to be restrained. She couldn't move her right arm and uses her left hand to position her right arm. My mother started to cry when the tied her wrists." (Kanski and Janelli 1996)
	"I just could not stand seeing my husband tied to a chair." (Newbern and Lindsey 1994)
	"She didn't do anything and she's tied up." (Kanski and Janelli 1996)
<b>Degrading</b>	"I would rather die than be like that." (Newbern and Lindsey 1994)
	"I was uncomfortable to see my mother restrained." (Kanski and Janelli 1996)
	"It's degrading to see him tied." (Newbern and Lindsey 1994)
	"I hated the idea that it was necessary." (Hardin and Magee 1993)
	"I still don't like to see it, and I cover it up." (Newbern and Lindsey 1994)
<b>Loss of Hope</b>	"When I saw the restraint, I lost all hope." (Newbern and Lindsey 1994)
	"He expected me to do something about it. I didn't like how he felt nor my helpless feeling." (Hardin and Magee 1993)
	"Seeing the restraint makes it so real to me. It is so real, that we can never do the things we planned." (Newbern and Lindsey 1994)

**Feasibility**

Feasibility as described in the conceptual framework relates to the health care environment and this component of the review addressed issues such as how restraints are used, the reasons why people are restrained and what characteristics predict the initiation of restraint.

## The Use of Restraints

This review focused on the use of restraints in hospitals and nursing homes and pooled the findings of observational and descriptive studies. This review aimed to describe how restraints are used in these settings. Specific areas addressed included the proportion of people that are restrained and the duration of restraint (see Tables 14 and 15). Given the nature of these data, individual studies were summarised in tables and the mean values averaged across studies. While this approach differed from that commonly used in meta-analysis, it provides a valid means of pooling results to produce an accurate description of the phenomenon. Additionally, as the findings from each study are presented in a table format, it allows independent verification of the pooled results by the user of the review.

**Table 14**  
Proportion of People Restrained in Acute and Residential Care Facilities  
Summary Table

<b>Residential Care Setting</b>		
<b>Citation</b>	<b>Percentage Restrained</b>	<b>Number Restrained</b>
Tinnetti (Tinetti et al. 1991)	30.7%	122 of 397
Karlsson (Karlsson et al. 1996)	24%	312 of 1325
Burton (Burton 1992)	47.8%	209 of 437
Magee (Magee 1993)	32%	55 of 173
Lever (Lever 1994)	12%	44 of 367
Koch (Koch 1993)	26.4%	428 of 1620
<b>Total Restrained</b>		1170 of 4319 (27%)

<b>Acute Care</b>		
Frengley 1986 (Frengley and Mion 1986)	7.4%	95 of 1292
Robbins 1987 (Robbins et al. 1987)	17%	37 of 222
Mion 1989 (Mion 1989)	13%	35 of 278
Lever 1994 (Lever 1994)	21%	91 of 437
Lofgren 1989 (Lofgren 1989)	6%	102 of 1661
Whitehead 1997 (Whitehead et al. 1997)	12.5% (this figure includes also chemical restraints)	-
Roberge 1988 (Roberge and Beausejour 1988)	25%	43 of 171
Minnick 1998 (Minnick 1998)	3.4%	-
<b>Total</b>		403 of 4061 (10%)

**Table 15**  
Duration of Restraint for Acute and Residential Care Patients  
Summary Table

Citation	Restraint Duration (mean)	Restraint Duration (range)
<b>Acute Care Setting</b>		
Frengley (Frengley and Mion 1986)	2.7 days	1 to 13 days
Robbins (Robbins et al. 1987)	3 days	1 to 35 days
Mion 1989 (Mion 1989)	4.5 days	1 to 18 days
	Mean Duration Across Studies = 3.4 days	
<b>Residential Care</b>		
Tinetti 1991 (Tinetti et al. 1991)	86.5	1 to 350
	Mean Duration Across Studies = -	

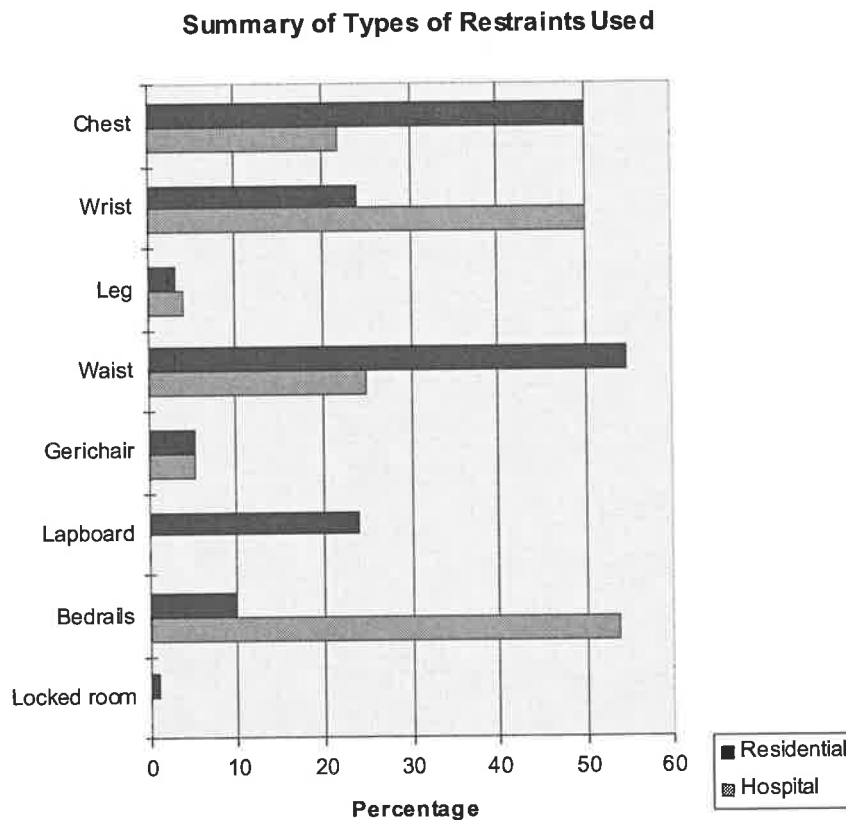
This review also sought to determine the specific restraint devices that are used in hospitals and nursing homes. The information was generated by observational and descriptive studies. The mean values were averaged across studies and these results were then presented in a table and graphically to better communicate this information (see Table 16 and Figure 6). As with the other summaries of descriptive research, the tabular summary of the findings of each study allowed independent verification of pooled results. While both presentation formats provided important information, the graphical presentation clearly communicates the differences in the types of restraint devices between acute and residential care settings. Importantly, this tabular and graphic presentation of data provided a structure to the data and allowed the patterns across many papers to be identified. This evidence highlighted the great variability in how restraints are used in each setting.



**Table 16**  
**Type of Restraints Used in Acute Care Setting**  
**Summary Table**

<b>Citation</b>	<b>Percentage</b> (% of total number of restrained)	<b>Number Restrained</b> (number of devices and total number restrained)
<b>Wrist Restraints</b>		
Minnick (Minnick 1998)	59%	471 of 799
Frengley (Frengley and Mion 1986)	13%	12 of 95
Robbins (Robbins et al. 1987)	72%	27 of 37
Mion (Mion 1989)	40%	14 of 35
Lofgren (Lofgren 1989)	14%	14 of 102
<b>Total</b>		<b>538 of 1068</b> (50%)
<b>Vest and Chest Restraints</b>		
Minnick (Minnick 1998)	16%	128 of 799
Frengley (Frengley and Mion 1986)	26%	24 of 95
Robbins (Robbins et al. 1987)	51%	19 of 37
Whitehead (Whitehead et al. 1997)	17%	8 of 45
Mion (Mion 1989)	40%	14 of 35
Lofgren (Lofgren 1989)	61%	62 of 102
Lever (Lever 1994)	5%	5 of 91
<b>Total</b>		<b>260 of 1204</b> (22%)
<b>Leg or Ankle Restraints</b>		
Minnick (Minnick 1998)	3%	24 of 799
Frengley (Frengley and Mion 1986)	5%	5 of 95
Mion (Mion 1989)	9%	3 of 35
<b>Total</b>		<b>32 of 929</b> mean = 3%
<b>Waist / Pelvic / Belt Restraints</b>		
Frengley (Frengley and Mion 1986)	30%	28 of 95
Robbins (Robbins et al. 1987)	31%	11 of 37
Mion (Mion 1989)	57%	20 of 35
Lever (Lever 1994)	15%	14 of 91
Whitehead (Whitehead et al. 1997)	9%	4 of 45
<b>Total</b>		<b>77 of 303</b> mean = 25%
<b>Gerichair</b>		
Frengley (Frengley and Mion 1986)	2%	2 of 95
Lever 1994 (Lever 1994)	7%	6 of 91
<b>Total</b>		<b>8 of 186</b> mean = 4%

**Figure 6**  
Type of Restraints Used in Acute Care Setting



### Reasons for Restraining People

A large number of studies have reported the reasons why health care workers use physical restraints and this type of information would help inform efforts to minimise their use. However, as this information has been generated by descriptive studies, it has remained beyond the scope of systematic reviews.

During the restraint review the findings generated by observational and descriptive studies were summarised using a descriptive content analysis that categorised, grouped and tabulated reasons for restraining people (see Table 17). To achieve this categorisation of reasons four coding categories and five sub-categories were used.

These categories allowed the many different reasons for restraining people to be grouped and so enabled the summary of this research. The coding categories were:

- Patient oriented reasons
  - \* safety
  - \* agitation
  - \* behaviour control
  - \* wandering
  - \* support
- Staff and organisation oriented reasons
- Social oriented reasons
- Treatment oriented reasons

Following coding and tabulation of data, the results were statistically analysed to compare differences between both settings. This analysis suggested that the only significant difference between acute and residential care settings in regard to the reasons for using restraints related to the facilitation of treatment (chi-square = 11.54,  $df = 1$ ,  $P=0.001$ ). Finally, the findings of the content analysis of observational and descriptive studies were presented graphically, to enable visual inspection and comparison of the reasons why people are restrained (see Figure 7). This graphical presentation highlighted the differences between acute and residential care facilities, particularly for the use of restraints to facilitate treatment. This graphic presentation also clearly communicates information that has been generated by many independent studies.

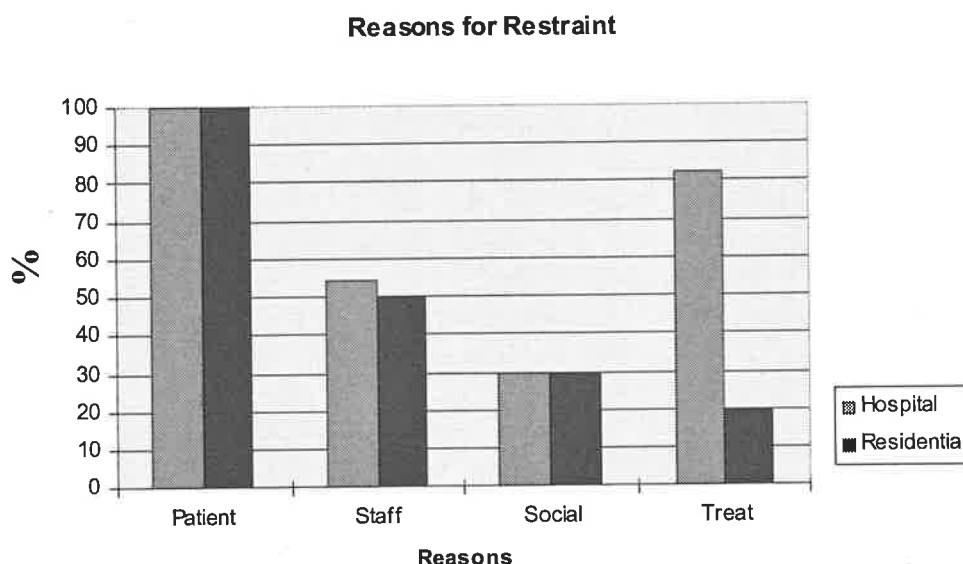
**Table 17**

The five sub-categories of 'patient-oriented' reasons for restraining acute care patients  
(descriptive content analysis)

<p>Restraint to Manage Patient Agitation</p>	<ul style="list-style-type: none"> <li>• Patient agitation / management of agitation (Schleenbaker 1994; Helmuth 1995; Kiat and Huan 1999; White 1999b)</li> <li>• Manage violent behaviour (Mion 1989)</li> <li>• Hitting nurse (White 1999b)</li> <li>• Safety of others (Strumpf and Evans 1988)</li> <li>• Protect staff and others (Helmuth 1995; Thomas et al. 1995; Kiat and Huan 1999; Lamb 1999)</li> </ul>
cited in 8 of 13 reports (62%)	
<p>Restraint to Ensure Patient Safety</p>	<ul style="list-style-type: none"> <li>• patient safety (Schleenbaker 1994; Helmuth 1995; Molassiotis and Newell 1996; Lee 1999; White 1999b)</li> <li>• prevent falls / protect from falls (Strumpf and Evans 1988; Molassiotis and Newell 1996; Minnick 1998; Lamb 1999; White 1999b)</li> <li>• previous falls (Schleenbaker 1994)</li> <li>• prevent falls from bed or chair (Helmuth 1995; Thomas et al. 1995; Kiat and Huan 1999)</li> <li>• unsteady gait / unsafe ambulation / impaired mobility (Helmuth 1995; Thomas et al. 1995; Whitehead et al. 1997; Kiat and Huan 1999; White 1999b)</li> <li>• safety when judgement is impaired (Thomas et al. 1995)</li> <li>• prevent patient obtaining dangerous supplies (Thomas et al. 1995)</li> <li>• prevent patient getting into dangerous places (Helmuth 1995; Kiat and Huan 1999)</li> <li>• inappropriate self-transfer (Schleenbaker 1994)</li> <li>• prevent injury secondary to wandering (Minnick 1998)</li> <li>• prevent self harm (Mion 1989)</li> </ul>
12 of 13 reports (92%)	
<p>Restraint for Behaviour Control</p>	<ul style="list-style-type: none"> <li>• confusion / altered mental status (Strumpf and Evans 1988; Schleenbaker 1994; Whitehead et al. 1997; Minnick 1998; White 1999b)</li> <li>• impulsive behaviour (Schleenbaker 1994)</li> <li>• manage behaviour (Minnick 1998)</li> <li>• restlessness / getting out of bed (Robbins et al. 1987)</li> </ul>
6 of 13 reports (46%)	
<p>Restraint to Prevent Wandering</p>	<ul style="list-style-type: none"> <li>• climbing out of bed or bed (Mion 1989; White 1999b)</li> <li>• prevent wandering (Robbins et al. 1987; Helmuth 1995; Kiat and Huan 1999)</li> <li>• manage wandering or hyperactivity (Mion 1989)</li> <li>• prevent injury secondary to wandering (Minnick 1998)</li> </ul>
6 of 13 reports (46%)	
<p>Restraint for Support</p>	<ul style="list-style-type: none"> <li>• provide quiet time for older person (Helmuth 1995; Kiat and Huan 1999)</li> <li>• maintain position / positional support (Molassiotis and Newell 1996; Minnick 1998)</li> <li>• maintain patients sitting balance (Mion 1989)</li> </ul>
5 of 13 reports (38%)	

**Figure 7**

Graphical presentation of the four categories of reasons for restraining people



**Predictors of the Initiation of Restraint**

A number of studies have investigated specific characteristics of the patient or resident and their relationship to the initiation of physical restraint. The results of studies that investigated these characteristics were pooled. As this involved a comparison of the characteristics in the restrained people to those in the unrestrained, observational studies provided the best evidence. Data were pooled to determine the odds of restraint being initiated with different characteristics (see graph 18).

**Graph 18**  
Cognitive impairment and Physical Restraint

Comparison: 04 Aged Care Setting - Predictors of Physical Restraint  
Outcome: 02 Cognitive Impairment

Study	Treatment n/N	Control n/N	OR (95%CI Fixed)	Weight %	OR (95%CI Fixed)
Burton	122 / 209	94 / 228	■	46.0	2.00[1.37,2.93]
Karlsson	291 / 312	523 / 1013	■	20.4	12.96[6.20,20.56]
Tinetti	61 / 122	89 / 275	■	33.6	2.09[1.35,3.23]
Total(95%CI)	474 / 643	706 / 1516	◆	100.0	4.27[3.40,5.35]

Chi-square 48.04 (df=2) P: 0.00 Z=12.56 P: <0.00001

.01   .1   1   10   100  
Restrained   Unrestrained

The findings from these three observational studies (Tinetti et al. 1991; Burton 1992; Karlsson et al. 1996) clearly demonstrated that cognitively impaired residents were more likely to be restrained than residents who were not cognitively impaired. Similar meta-analyses of observational studies demonstrated that residents who had bladder or bowel incontinence, or were unable to dress without assistance, were more likely to be restrained. This evidence highlighted the need for the development of strategies to care for cognitively impaired residents if the use of physical restraint is to be minimised.

### **Summary**

Through the conduct of this expanded systematic review, the best evidence on physical restraints was summarised. The product of this series of reviews differs from most systematic reviews because of its comprehensive cover of the topic. Additionally, unlike most systematic reviews, the pooling of results was achieved using a range of methods in addition to meta-analysis. However, it is argued that these methods were appropriate for the type of data summarised. This repertoire of summary methods enabled all types of research on this topic to be included in the expanded systematic review.

The restraint review clearly demonstrated that the conceptual framework combined with the methods proposed in the expanded review process provided a valid means to address broad clinical issues, while still adhering to the principles expected of all research and systematic reviews.

# **CHAPTER 8**

## **Discussion**

## **Discussion**

This section will discuss the findings of the two reviews and the methods used to produce the summary and synthesis of research. Specifically, it will address:

- The findings of the review
  1. the integrative review
  2. methodological critique of the research
  3. theoretical critique of the research
- The conceptual framework
- The expanded systematic review process

### **The Findings of the Reviews**

#### **1. Integrative Review**

##### **Music Review**

The use of music for people during hospitalisation has been subject to extensive evaluation. A considerable proportion of this research has been in the form of clinical trials addressing effectiveness. However, there has been only limited evaluation from other perspectives and utilising other research methodologies. In summarising the research findings, this review demonstrated that while music has a limited impact on physiological outcomes, it clearly has a positive impact on psychosocial outcomes. That is, while it does not change a patients' vital signs, it significantly reduces their anxiety. It may also act as a diversion during unpleasant procedures, as demonstrated by the reduction in analgesic and sedative use. These results support the use of music in hospitals, and highlight the need for further research into the use of music with specific hospital populations. There is little information on issues related to the implementation of a music program, or the perception of patients regarding its use.



The findings of this review also demonstrated the need for larger studies addressing many aspects of music in hospitals. The basis for this is that there may be many benefits from listening to music that have not been identified as a result of the lack of power of most studies.

The purpose of choosing music as the review topic was to evaluate the expanded review process. Music was selected after an initial search of the literature suggested that it had not only been subject to extensive evaluation, but also that it had been evaluated by a number of methodologically different studies. After completing the review, this proved not to be the case. This highlights the difficulty of evaluating the research, and research literature, using a limited 'snap-shot' of the literature. The views that are formed during the 'snap-shot' evaluation assume the sample will be representative of the total population of studies. However this will not always occur. For example, this snap-shot view would be influenced by the databases used to select the sample. A MEDLINE search would impose a medical perspective upon the topic that may differ from the predominantly nursing papers identified in the CINAHL database. These snap-shot literature evaluations are therefore at considerable risk of producing misleading or incorrect information. This risk would increase significantly for topics with extensive literature generated by a number of health care disciplines, where the snap-shot sample would be unlikely to be a representative sample of the total population of published studies.

From a slightly different perspective, this concept of an inaccurate 'snap-shot' of the literature has important implications for practice. It highlights how easy it is to select an unrepresentative sample of studies and use this sample as the basis for clinical practice change. The consequence of this could be the continuation of ineffective, inappropriate, or even dangerous practices. In reviewing discussion papers related to the use of music in hospitals, some recommendations are clearly wrong when compared to the findings of this systematic review. For example, music is recommended as an intervention for decreasing heart rate and blood pressure (Chlan

and Tracy 1999; White 2000). Similar concerns regarding clinical recommendations lacking any supportive evidence were also identified by a systematic review on the nursing management of chest drains (Charnock and Evans 2001). As the volume of literature continues to grow, these 'literature 'snap-shots' will become increasingly inappropriate, and be more likely to be detrimental to health care delivery.

### Physical Restraint Review

While only selected findings of the physical restraint review were presented in this thesis, these findings clearly demonstrate the many differences between topics. The restraint review provides a good example of one of the problems faced by the reviewer. As some clinical issues involve a very broad spectrum of care, they have fitted poorly within the existing systematic review framework. To date, this type of research has been summarised by the much criticised traditional literature review. An alternate approach has been for the systematic reviewer to impose a focus of effectiveness on the topic. As a consequence of these problems, systematic reviews have most commonly been limited to only those interventions that fit within the narrowly defined boundaries of these reviews.

Physical restraint served as an excellent contrast to the music review because of the many complexities faced when attempting to review the available research. Firstly, the effectiveness of physical restraints (as restraining devices) is not the issue of interest to clinicians. The evidence needed by clinical practice is more about not using restraints, and so restraint minimisation programs became a central theme of this review. However, clinical practice also needs evidence about many other aspects of the use of physical restraints. Secondly, the use of physical restraint in clinical practice has never been adequately described. When descriptive and observation studies are inspected, considerable variation is encountered. Without the structure provided by the conceptual framework and expanded review process, drawing conclusions from this body of research is difficult. Thirdly, while restraints are used

to protect patients and residents, a number of reports highlight the dangers that accompany their use. Weighing the benefit and harm from this diverse and poorly defined area is becoming increasingly difficult. Finally, this topic encompasses many ethical and moral issues related to the enforced immobilisation of people in the name of health care delivery. As a result, ensuring that the experience of those involved is represented along side other types of evidence is crucial.

In summarising the research addressing physical restraint, a number of issues of importance were addressed. While this research was summarised by a series of focused reviews, when integrated into a single report it provided the first comprehensive integration of all relevant evidence. This type of review endeavour has not previously been possible because the review methods were not available. The findings presented in this thesis demonstrate that it is possible to summarise primary research with the rigour expected of primary research.

## **2. Methodological Critique**

### **Music Review**

While studies were generally of good methodological quality, a number of methodological limitations were identified during this systematic review. This is in contrast to other systematic reviews of nursing research that have found most studies used inappropriate methods (Evans et al. 1999, Evans et al. 2000). The major criticism of studies evaluating music for hospital patients was that of small sample size. This was a feature common to the majority of the RCTs identified during this review. The sample size of the two specific populations investigated during this review were:

- mean of 48.6 participants per study for Hospital Patient studies
- mean of 59.4 participants per study for Procedure Patient studies

On this basis, it can be suggested that most studies lacked the power to adequately evaluate music as an intervention. The size of some studies was so small that it is questionable whether they should have been published. For example, Chlen's study had a total study population of only 20 (Chlan 1995), while Miller's had only 17 participants (Miller et al. 1992) and Standley's only 15 (Standley 1992). One study involving pre-operative patients had a population of only 12 people, of which 3 from the 6 patients in the control group dropped out of the study before its completion (Szeto and Yung 1999). These studies are potentially misleading as they lack the power to adequately evaluate the intervention and were likely to miss beneficial outcomes. This problem is not limited to music studies, and has previously been reported in the literature (Williams and Seed 1993). Additionally, other systematic reviews have reported similar findings (Kowanko et al. 1998; Evans et al. 1999; Dunn et al. 2000). Some of the studies identified during the review were so small that they failed to make any useful contribution to the body of knowledge on music in hospitals. Despite the inadequate sample size, many of these small studies were otherwise well designed and executed.

Some studies failed to report sufficient data to permit independent evaluation of the findings. This lack of data prevented the inclusion of these studies in the meta-analysis (Walther-Larsen et al. 1988; Palakanis et al. 1994; Cruise et al. 1997; Koch et al. 1998). This inadequate reporting of results is a limitation of many studies. For example, one study failed to provide information about the methods used, the population studied or the results obtained (Kopp 1991). The value of this type of research report is reduced as a consequence of this problem.

Ten RCTs that fulfilled the inclusion criteria were subsequently excluded from the meta-analysis as a result of the critical appraisal. Three of these studies were excluded when the critical appraisal identified multiple problems with the methods used (Augustin and Hains 1996; Sabo and Michael 1996; Grey et al. 2000). The most

common reason for excluding RCTs was inappropriate method of randomisation (Dubois et al. 1995; Augustin and Hains 1996; Sabo and Michael 1996; Cunningham et al. 1997; Grey et al. 2000), followed by attrition (Winter et al. 1994; Heiser et al. 1997; Szeto and Yung 1999). While the risk associated with inappropriate allocation of participants to study groups is well recognised, studies still use questionable methods such as alternation, day of week, hospital record number or treatment location. Similarly, it is also well recognised that if follow-up of participants is not adequate, the findings of the study may be at increased risk of error or bias (Mulrow and Oxman 1997). Yet in a small number of studies a considerable proportion of their population was lost to follow-up.

In summary, while the methodological quality of studies investigating music in hospitals was generally good, the processes used in some studies invalidated their findings. Additionally inadequate reporting of results by other researchers limited the contribution of their study. As a consequence of these factors, these studies offer little to health care delivery.

### Physical Restraint Review

Research addressing physical restraint was vastly different from the predominantly RCT research of music. The complex nature of this topic is reflected in the diverse research methods encountered. In terms of the methodological critique of this body of research, the major issue arising relates to the difficulty of utilising the RCT method. As previously discussed, the RCT is held to be the best design for the evaluation of the effectiveness of interventions. While the effectiveness of restraint minimisation programs was a component of the restraint review, this aspect of physical restraint lends itself poorly to investigation by RCT. As restraint minimisation programs are most commonly an intervention implemented across the entire health care organisation, randomisation by individual is difficult. If a ward, department or service is used as the unit of randomisation, this approach is then generally termed quasi-

experimental and ranked at a lower level of evidence than other types of RCTs. If the before and after design is used, as encountered in the restraint review, this evidence is ranked even lower in the current hierarchies of evidence. While this approach represents a practical compromise, and has limitations, it still provides some degree of evaluation of effectiveness. Despite this, these studies would not normally be included in a systematic review. However, it has been argued that consistent findings from a number of similar studies lends credence to this type of evidence and therefore increases the confidence clinicians can have in the results. Unfortunately, as these studies are often not part of the 'accepted' evidence, the findings are excluded by reviewers. As a consequence, complex topics like physical restraint go wanting for evidence, as a rich source of information is ignored by systematic reviewers. With the exclusion of this evidence from reviews, clinical practice must then be guided by expert opinion, which ironically, is ranked even lower in the hierarchies than the excluded 'before and after' studies.

### **3. Theoretical Critique**

#### **Music Review**

In terms of the theoretical construct of the area of music, there is some confusion over what constitutes music as an intervention. In the context of this study, it has been viewed as a distracter, something to divert an individual's attention during an unpleasant procedure or while they are resting in hospital. However in some studies music was structured quite differently. Music therapy in the literature also commonly involves a program of music sessions (Curtis 1986; Durham and Collins 1986; Hanser and Thompson 1994; Mornhinweg and Voignier 1995; Ragneskog et al. 1996; Purdie et al. 1997). At times the music was live (Bailey 1983; Malone 1996), and in other studies involved group music sessions (Ragneskog et al. 1996; Oliver 1999). In some studies the music involved singing or playing instruments (Goloff 1981; Purdie et al. 1997). In many of these studies music is an active process, or a group activity, rather

than a passive intervention addressed in the music review. The term 'music therapy' is more commonly associated with this active use of music, although these boundaries are not distinct in the literature. Clearly the processes and expected outcomes between these approaches differ considerably. However, these papers highlight the range of potential uses of music in addition to serving as a distracter.

### Physical Restraint Review

There was a consistent theme across most papers during the restraint review regarding the need to minimise the use of these devices. This is reflected in studies that provide evidence related to injury, negative impact, poor outcomes and the success of restraint minimisation programs. These studies clearly indicate that the major focus of future research is a continued effort to reduce the number of people restrained in health care facilities.

However, the definition of restraint devices is inconstant. Some studies viewed bedrails as restraint, while in other studies they were excluded. Mitts, gerchairs, and gerichairs also shared this ambiguous status. In one study, a single person was found to be restrained in a locked room, further challenging current definitions of what constitutes physical restraint. This is emerging as an area in need of further professional discussion to attempt to reach a consensus.

### **The Conceptual Framework**

The conduct of the music and physical restraint reviews allowed an evaluation of the expanded systematic review process using the structure provided by the conceptual framework. While there was a limited breadth of the research addressing music, this

was compensated by the restraint review. These two reviews provided the basis for discussing the potential value of using this approach to generate research summaries.

The conceptual framework proposed in this study provided a sound focus for the reviews. This clearer focus was a result of the differing research methods being located within a coherent framework. That is, the RCT and the interpretive study both had a logical place in the research summary. This framework also allowed the identification of the relationships between studies and of the potential contribution of each. As the major activity of reviews is to impose some form of order over diverse bodies of literature, this process of locating each study within a coherent framework is the first step in achieving this order. The process of imposing the order on the body of research was achieved through the grouping and categorising of research. However, throughout this activity the conceptual framework maintained the focus of the review.

For example, in developing the review questions, the conceptual framework mapped out three broad perspectives for the review, these being the process, person and health care environment. These perspectives helped define the review questions, and in so doing, provided the first level of order for the research. Continuing with this activity, the inclusion criteria provided the second level of order, allowing studies to be selected and grouped according to their potential contribution to the review. This process of imposing different levels of order over the literature was important because without this structure and clear direction, attempts to expand the scope of a systematic review would be very difficult.

From a slightly different perspective, the conceptual framework also allowed the question to lead the review process rather than the research method. As previously discussed, most systematic reviews address effectiveness and at times the impression in the literature has been that systematic reviews only summarise RCTs, with other methods beyond their scope. This ignores the fact that the systematic review process



is merely an approach to reviewing literature that helps maintain a rigour similar to that of primary research. As a consequence of this limited view, systematic reviews in their current format are led by the research method rather than the question. The physical restraint review clearly demonstrated that a range of research can be summarised while still maintaining the expected rigour.

### **The Expanded Systematic Review Method**

The purpose of conducting the expanded systematic reviews was to facilitate an evaluation of the proposed review methods. As previously described, the development of the expanded review process was informed by existing reviews of the research literature. In terms of reviews of non-RCT research, while some of the published reviews represented the initial exploration of this area, it appears there has been little critical debate. These reviews have been completed and published with little attention paid to the validity, reliability and usefulness of the finished product. This is particularly evident with reviews of interpretive research. This is in contrast to systematic reviews of RCTs which have been subject to ongoing critical debate, revision and refinement. While the interpretive reviews produce a readable story, their value and validity have yet to be determined. Additionally, current methods are also inconsistent and at times questionable. While more recent interpretive reviews have adopted some of the processes used during systematic reviews, other areas have advanced little beyond the traditional review of the research literature. The risk is that the increasing number of these reviews of interpretive research may result in the acceptance of the methods and so bypass any critical debate by the health professions.

Of greater concern is the poor level of reporting of the methods used during these reviews. This inadequate reporting limits the ability to critique the review and

hampers further methodological development. Reporting is highly variable, with some reviews providing a clear and comprehensive decision trail, while others provide scant information. From this perspective, it is argued that many of the existing reviews of interpretive research fail to meet the basic expectations of the primary research they summarise. As a result, they have failed to advance beyond the much criticised traditional literature review of twenty years ago.

### The Protocol

The format for protocols commonly used for systematic reviews of RCTs is also suitable for reviews of a broader range of research. The format can be likened to the research proposal, and as such, addresses all critical areas of the review. From this perspective, the standards that are expected of primary research can be applied to reviews through the process of developing a review protocol. These protocols will differ depending on the type of research to be summarised, yet like primary research, they outline each aspect of the proposed systematic review.

The protocols developed for the music review demonstrated how they can be adapted to accommodate a broader perspective. While three separate protocols were used to address the areas of effectiveness, feasibility and appropriateness, they could easily be incorporated into a single protocol.

The protocol addressing effectiveness had the soundest methodologic base as a result of the considerable development that has occurred for these reviews. However, the development of each of the three protocols, followed logically from the preceding components of the protocol. That is, when the question has been posed, the type of research that can provide valid evidence is easily determined. Additionally, once the best research design has been selected, the approach to the synthesis of results can also be determined. This development process is not unlike that used in primary research.

## The Question

The review questions provide the direction and determine the processes that should be used during the conduct of the review. For the music review, the questions did not differ from those posed by the primary researcher, for example:

- Does music reduce patients' perception of pain?
- Does music reduce the anxiety of patients?
- Can music be implemented in the hospital setting?

While the questions from the restraint review were not presented in this thesis, the broad nature of the topic meant a series of questions was needed to provide a clear direction and define the review boundaries. For example, questions in the effectiveness review addressing injury included:

- What proportion of patients and residents suffer restraint related injury?
- What injuries do physical restraint devices cause?
- What injuries are caused by specific restraint devices, such vests, wrist restraints and bedrails?

To determine how restraints are used in acute and residential care, the review questions included:

- What proportion of patients and residents are restrained?
- What is the duration of restraint for patients and residents?
- What type of physical restraints are used in acute and residential care settings?

Similar questions were used to investigate the relationship of specific characteristics, such as age, to the initiation of restraint and to determine why people are restrained. This represents a significant departure from current methods. During systematic reviews of RCTs, a small number of questions addressing a very narrow focus are

used. Because of this, these reviews summarise one component of the total available evidence. The methods proposed in this thesis are ideally suited to broad clinical issues. and because of this, will likely address a series of questions. However, the scope of the questions will be influenced by the available research. For topics where research is still at a developmental stage, broad questions permit a stocktake of current best evidence which can then be used to prioritise future research.

The scope of the review questions has even greater implications for reviews of interpretive studies. While these types of questions would differ from those of effectiveness, they must still provide a clearly define the area of interest. This is more important for these reviews because narrative data is more difficult to summarise than numerical data. However, as is evident by the questions used during the music and restraint reviews, the questions posed by the primary researcher are also suitable for the reviewer.

### Selection of Studies

If the selection process of reviews is to be free of the reviewer's subjective influence, there should be clearly stated selection criteria. For reviews of RCTs, the format and components of the criteria are well established. This format states, in precise terms, the interest of the review in terms of population, intervention and outcome. While the question defines the focus of the review, it is the inclusion criteria that operationalises this question, and in so doing, delineates the boundaries.

In terms of the methods proposed in this thesis, the inclusion criteria serve another purpose. Because this type of review may address broader clinical topics, the inclusion criteria serve as a template for the sorting and ordering of studies. This process allows a large volume of research to be concisely summarised. This grouping and sorting of studies starts with the development of questions and is continued by the development of the inclusion criteria. For example, in exploring the effectiveness

of music, the population was limited to adults thereby excluding studies addressing children and infants. This exclusion aimed to limit studies to a homogenous group. The selection of studies was then divided into hospital patients and patients undergoing procedures. This progressive categorisation enabled the large volume of research to be managed and grouped in a coherent structure.

While this imposing of order is important to manage large numbers of RCTs, it is vital if the review is to include interpretive studies. As interpretive studies produce narrative data, grouping and categorising this data is critical if the reviewer is not to be overwhelmed by the volume of information. The inclusion criteria used in the expanded review allowed the broader focus of the review to be managed. These criteria provided the means to determine whether to include or exclude studies. During the reporting of the review results, these processes became part of the decision trail of the systematic review.

### Search Strategies

Considerable attention has been devoted to the development of strategies to aid in the location of RCTs. However, similar efforts addressing other types of research have not been made. As previously discussed, as a result of the volume of literature and the difficulties in locating studies, a considerable proportion of the research of nurses could be lost to the profession. This loss is likely to be compounded by other issues, such as the failure of nurses to publish their research and poor indexing practices of databases. This difficulty will be greater for unpublished research, where the reviewer will be limited to searches of conference proceedings or contacting experts and professional groups. Unfortunately, this will likely only identify a small proportion of all completed research. While these issues present a challenge for the reviewer, they also highlight the important contribution of reviews in providing a stocktake of all research and thereby producing a public record of all identified research. This stocktake of studies is even more important in areas where the knowledge base is in an

early developmental stage. For example, despite a large volume of literature addressing physical restraint, many papers were not research based. The risk with this literature is that these many opinion papers may form the basis for clinical practice rather than the research publications.

In terms of the expanded review process, the general approach used to locate RCTs was utilised for other types of research. This involved multiple steps used to increase the likelihood of identifying all relevant studies. Despite the sound theoretical basis for this approach, it is difficult to evaluate its value and reliability. The basis for this difficulty is that without a list of all existing studies on a topic, it is impossible to evaluate the completeness of the search. However, it is argued that adapting successful RCT search strategies will increase the likelihood of identifying all relevant studies that used non-RCT research methods.

### Critical Appraisal

Critical appraisal of research is still a controversial area. As previously discussed, a vast array of tools have been developed for the appraisal of RCTs. How best to achieve the correct balance between under and over appraising studies has yet to be determined. The risk of too limited an appraisal is that research of poor methodological quality, and therefore with suspect results, will be included in the review. Over appraising risks excluding the evidence generated by valid studies. There is also considerable variation in the criteria that have been used to appraise these RCTs, but most appraisal tools include the method of randomisation. However, most appraisal tools have not been evaluated and are therefore only supported by theoretical arguments. This difficulty in determining which criteria to use during appraisal increases with interpretive research. Indeed, one of the current arguments is whether this type of research should be appraised, as these studies have been subject to peer review prior to publication.

However, the peer review process of journals has been repeatedly shown to be an inadequate filter for quality research (Rothwell and Martyn 2000). This is supported by the numerous papers describing the poor methodological quality of published research (Mills 1993; Williams and Seed 1993; Altman 1994; Schulz et al. 1994). Additionally, inadequate sample size is a common problem, and is one of the major reasons why studies fail to adequately evaluate the intervention. Yet many studies are still published using an inadequate sample size. Confusing matters still further, publication bias influences what is being published. With publication bias, studies are more likely to be published if they demonstrate positive or interesting outcomes. Publication bias has also been linked to the primary researchers, in terms of whether they submit their work for publication, and in the selective reporting of findings. These issues highlight the need to critically appraise research prior to utilising its findings. However, where interpretive studies are located within this discussion is less clear. The issue for interpretive research is whether the peer review process of journals accurately selects only studies of a high methodological quality. Additionally, there is currently no information on whether the findings of interpretive studies influence the likelihood of the report being submitted, and then accepted, for publication.

In this thesis it has been argued that having studies subject to the peer review process of a journal is not a reliable indicator of quality. On this basis, all research should be subject to some form of appraisal prior to inclusion in a systematic review. However, it is not clear how this can best be achieved. This argument has been applied to interpretive research, and the appraisal of these studies is part of the expanded review process. There is even less agreement on how interpretive research should be appraised, and developed appraisal tools are so abstract in nature that it is likely that each appraiser would reach a different conclusion. While it has been suggested that during the synthesis of interpretive studies no two reviewers will produce exactly the same findings, it would be of great concern if this view was also applied to the appraisal of interpretive research.

Contributing to these difficulties, it is not clear whether the published report provides a sound basis for determining the validity of the research. The indirect assessment of study quality means that a poorly written or incomplete reports will be judged as being of poor quality. Added to these difficulties, selective reporting of method and findings may also mean that poor or mediocre quality research is presented as scholarly endeavours.

The approach used for the expanded review process acknowledged the limitation of this aspect of reviews, while also providing a working alternative. For RCTs, four critical areas of the research process were used as the basis for the development of an appraisal checklist. Evaluation of the selection process, performance, outcome measurement and attrition of each RCT provided a sensible approach to critical appraisal. This approach did not aim to identify only those few studies that were methodologically perfect, rather it focused on studies whose findings were believable. For observational studies, these criteria were modified to suit the different methods of these studies.

Interpretive research does not follow a precise and consistent design to the same extent as experimental research. This consistency of experimental studies means that the method of one RCT is exactly the same as the next. That is, while the population, intervention and outcomes of the RCT may change, the study design remains constant. However, while all interpretive studies follow similar processes, there is greater variation in the methods. As a consequence, these studies do not lend themselves to appraisal by checklists as easily as does experimental and observational research.

While a checklist has been used in the expanded review process, it is acknowledged that by the very nature of this research, the approach has limitations and may not adequately distinguish good from poor quality research. In selecting appropriate



criteria to judge the validity of interpretive research, available appraisal tools provide only limited guidance. Some focus of these tools on specific components of the research process while others have taken a more abstract approach to the appraisal. Both these approaches are not without their limitations. Specific criteria addressing critical phases of the interpretive research process are likely to be the most useful approach to appraisal. However, the identification of these critical phases and their translation into appraisal questions has yet to be achieved.

The interpretive and descriptive appraisal tools proposed in the expanded review process focused on the research report rather than the research process. In proposing this, it is acknowledged that there is currently no valid tool available to appraise this type of research. However, this approach ensures that those studies considered for inclusion in a systematic review provide complete details of their methods. Additionally, this approach also highlights the importance of providing sufficient information of the methods and findings of studies to allow users to assess their value and usefulness.

### Hierarchy of Evidence

The ranking of research evidence, like critical appraisal is another area of controversy. This is reflected in the large number of hierarchies that have been proposed, most utilising an experimental framework. How the different research designs are ranked is highly variable, but most hierarchies rank the multi-centred RCT as the highest level of evidence. Systematic reviews have now joined these RCTs, and are seen to provide rigorous evidence at low risk of error. However, some hierarchies have included such things as confidence intervals, or have ranked 'N of 1' studies as best evidence.

The role of hierarchies of evidence is increasing in importance as part of the evidence-base health care and guideline development movements. However, the major focus on effectiveness has had a significant impact on what is held to be the best evidence. This

focus has also seen the RCT being viewed as the standard to which other research methods are compared. As part of this view, other approaches to research have often been seen as lower level evidence. This has important implications for nursing which has investigated many aspects of health care from a variety of different methodological perspectives. As a consequence of being outside the commonly accepted concept of best evidence, nursing research has not had a major impact on evidence-based health.

The focus of these hierarchies remains predominantly on effectiveness, thereby ignoring the contribution of other types of evidence. While observational studies have been ranked in the mid levels of these hierarchies, if the health care literature is an accurate gauge of opinion, this rank does not have the support of all. As previously discussed, examination of the processes of systematic reviews has seen the qualitative/quantitative debate revisited. As part of this debate, non-RCT research has been commonly seen to be supplementary evidence at best, never the basis of clinical decision making. This perspective is clearly entrenched in the economic framework that underpins current health care.

The argument proposed in this study is that all research has a unique perspective, and as a result, makes an important and valid contribution. This contribution is not necessarily as a support for empirical research, but through the insights not offered by empirical research. However, locating this research within the evidence-based framework has been difficult. As a consequence, reviews of interpretive research have remained outside the evidence-based movement. The hierarchy proposed in this study is an attempt to provide a logical framework for a range of methodologically different research. This approach aims to identify how evidence generated by different research methods can contribute to clinical practice and health care decision making. However, this proposal is based on theoretical arguments, and this limitation is acknowledged.

In the hierarchy proposed in this thesis, observational studies have been recognised as a valid method for evaluating interventions in the practice setting. As such, factors that influence the selection, delivery and evaluation of any intervention are part of the study processes. This differs in many important ways from the rigidly controlled and carefully selected population of the RCT. The basis for this premise is that observational studies evaluate interventions within the limitations of the real world of health care. As such, the evidence generated by these studies differs from that of the RCT. Despite this difference, it is argued that this evidence is no less valid. Additionally, this observational evidence has been seen to complement the evidence generated by RCTs.

However, as demonstrated in the physical restraint systematic review, when the focus of the review is to determine the relationship of specific characteristics between populations, the observational study provides the most valid evidence. This issue was highlighted during the investigation of characteristics that can be used to predict which people are likely to be restrained. Additionally, observational studies were considered to provide the best evidence when the negative outcomes of restraint were the focus of the review. During the restraint review, meta-analyses of observational studies were undertaken rather than of RCTs, and in those situations, observational studies were held to represent the best evidence.

The hierarchy proposed in this thesis also differs in how it ranks interpretive research. For research addressing appropriateness and feasibility of an intervention, interpretive research can provide valid evidence. Additionally, these studies address issues that are beyond the scope of the RCT or observational study. The ranking of interpretive research acknowledges that for some questions, interpretive research provides the best evidence. However, this is not to suggest that this evidence can be used to determine effectiveness, rather the type of evidence generated by this research provides a different perspective on health care.

The value of interpretive research was highlighted in the restraint review by the incorporation of information on perceptions and experience of people who were subject to restraint, and their families. This information provided an insight that is not normally part of the systematic review. It is argued that this information can make an important contribution and should form part of the evidence on which clinical practice is based.

Descriptive studies are most often seen to provide the lowest level of evidence. Yet the inclusion of this research contributes information that is beyond the scope of most other methods. The incorporation of descriptive research in the physical restraint review identified the great variation in current practice and also highlighted that some people are subject to extended periods of enforced immobility. This descriptive information also helped to determine the proportion of people restrained, the duration of restraint and the type of devices used. Additionally, single case reports described specific injuries caused by restraint devices. This information on specific injuries was not identified by any other type of research, and so despite the normally low ranking of this information, these reports made an important contribution to the review. On the basis of the two reviews reported in this thesis, it is argued that the proposed hierarchal framework enables the systematic review to be extended beyond RCTs.

### Data Collection

Data collection is one of the phases of the review that presents considerable risk of error. To minimise this during reviews of RCTs, data collection tools are developed and pilot tested. With interpretive research the approach used is more variable. During the evaluation of published reviews of interpretive research some of the data collection and data synthesis were undertaken concurrently. Yet if the review is to present a record of all the research on a topic, demographic data must be collected. Additionally, managing the themes and categories from multiple studies would be difficult without some form of systematic approach to data collection and

management. With the content analysis approach to data synthesis proposed in this study, data collection and synthesis occurs together.

During the physical restraint review, data was collected from interpretive studies through a process of reading and re-reading. However, using qualitative content analysis this process followed clearly defined steps, and as part of this, data collection forms were used to record the coded data. This formalised data collection was incorporated into the expanded review process to help ensure rigour of this phase of the review.

For the two reviews reported in this thesis, the formalised process of data collection assisted in maintaining the review focus. This was achieved through the use of data collection tools that were developed as part of the review protocol. Because of this, they operationalised the review questions and so directed the process of data collection. While specific details of interpretive data can not be predetermined in the same manner as is done with RCTs, the review focus can be maintained. During the restraint review on appropriateness of physical restraints, only data providing a description of the experience of being restrained, or having a relative restrained, was of interest. This allowed a more rapid collection of data from study reports.

#### Data-synthesis

The synthesis of data from independent studies is an area that is still undergoing development and refinement. Considerable attention has been devoted to the synthesis of RCTs, and to a much lesser extent, observational research. However the synthesis of findings from interpretive research has received only limited attention in the health care literature. There has been little interest in the synthesis of descriptive research.

During the review of the literature related to the conduct of reviews, considerable variation in methodological approaches, and many gaps in this debate were evident. For example, there has been ongoing professional debate and exploration related to all aspects of meta-analysis of RCTs, but reviews utilising meta-synthesis have appeared in the nursing literature with little professional discussion or critique.

The methods used for meta-analysis of RCTs during the expanded review process were based on currently accepted approaches. During the music review meta-analysis provided a means by which the findings from multiple RCTs could be pooled. For these studies, meta-analysis completed the process of grouping, categorising and tabulating the findings of a body of research. Similarly, meta-analysis also provided a means to pool the results of observational studies.

While reviews have not normally included non-randomised experimental studies, the summary of results from before and after studies evaluating restraint minimisation programs highlighted the potential contribution of this type of research. It was argued that in areas of limited research, or where the conduct of randomised trials is not possible, other types of experimental research can be used to help inform practice. This is not to suggest that these studies should be ranked at the level of RCTs, rather, that for some topics they represent the best available evidence.

Despite the existence of reviews of interpretive research, there is no clear method for the synthesis of this type of data. The proposed method of qualitative content analysis provides one way by which the findings of independent studies could be summarised. However, the product of this process differs from that produced by meta-synthesis. Meta-synthesis is used to generate new interpretations of interpretive data. It has been argued that the risk with this approach to data synthesis is that the findings are too far removed from the original phenomenon to be meaningful. Content analysis provided a means to identify themes and patterns in the narrative, and these findings were then grouped and categorised. The degree of

interpretation that occurs during this process was kept to a minimum. The findings presented in the restraint review highlighted how this process can be undertaken, while still maintaining the integrity and purpose of the original research. Additionally, the findings of this review demonstrated the important contribution of this type of evidence, as witnessed by the descriptions of the negative impact of physical restraint.

Descriptive studies have remained outside the scope of systematic reviews and there has been little interest in developing methods for the synthesis of this type of data. Yet as demonstrated in the restraint review, these data can make an important contribution with some topics. While descriptive studies are ranked as low level evidence on research hierarchies, they still provide important information when a description of an event or activity is one of the purposes of the review. This evidence is strengthened by pooling the findings with those generated by other similar descriptive studies. The methods proposed in the expanded review process provided a means by which this descriptive research could be pooled. For text, descriptive content analysis allowed emerging themes to be coded, grouped and tabulated. During the restraint review, this approach allowed the reasons for restraining people to be described. For numerical data, the determination of mean values across a number of studies provided one way by which these data could be pooled. Importantly, the methods used to pool these findings were very similar to those used by the primary researchers and so provided a meaningful way to talk about multiple descriptive studies.

The conceptual basis for the pooling of different types of studies is that the results of research are more robust and transferable / generalisable when generated from multiple settings, populations and by different researchers. From this perspective, the synthesis of these multiple studies allows the essence of a phenomenon, event or intervention to be determined. Equally important, it allows areas of difference between studies to be identified. It is argued that this pooling of research increases the

strength of results, and that the research method should not be a barrier to this process.

### Decision Trail

One of the difficulties in conducting a review that goes beyond RCTs, is that of demonstrating that the process was systematic, pre-planned and rigorous. Much of this difficulty relates to the poorly developed processes for conducting broad reviews, and as a consequence, for defending them. This conceptual framework, in conjunction with the expanded review process provides the basis for this defence. While it is acknowledged that further development is needed, it is argued that the methods proposed in this thesis provide an important first step.

In reporting the findings of both primary research and systematic reviews, it is expected that evidence of a decision trail is also provided. This decision trail is seen in the reporting of the many decisions made during their conduct. Yet this standard has not often been provided for the broader research review. However, if these reviews are to be valued and trusted, a decision trail must be provided. The conceptual framework, in conjunction with the expanded review process, provides the basis for this, and so helps the review meet the expectations of a systematic review.

### Limitations

The basis for the expanded review method in this thesis was the existing research reviews and the professional health care literature. However, although it is argued that reviews can go beyond effectiveness, this process is not without some limitations. In addition to the limitations outlined in the previous discussion, two other issues must be acknowledged.



### Limited Evaluation of Method

As already acknowledged, there are only limited existing reviews of interpretive research, and there has been little professional debate. The evaluation of the proposed expanded review methods were based two reviews. As a consequence, the evaluation of these proposed methods was very limited. While it has been argued that these processes represent a valid alternative to meta-analyses of RCTs, further exploration and evaluation is required.

### Accumulative Approach

An important assumption underpinning this developmental work, is that more is better. That is, evidence generated from multiple populations, settings and researchers, will provide more robust evidence than single studies. This assumption is the basis of systematic reviews of RCTs and observational studies. However, it has not commonly been applied to interpretive or descriptive research. In developing the expanded review process, the assumption is that this evidence will be more robust and generalisable than the findings produced by individual studies. However, this view of evidence will not necessarily be shared by all researchers and reviewers.

# **CHAPTER 9**

## **Conclusion**

## Conclusion

The argument presented in this thesis concerns the evidence on which nursing practice is based, and the many factors that have shaped it over the past ten to twenty years. These factors relate to the changing expectations of consumers and policy makers, the research on which practice is based and the issues surrounding the use of research. As a consequence of these many issues, the evidence-based movement has emerged as one of the most important influences on health care today. As part of this, the role of the research review has also changed significantly as they exert an increasing influence on health care decisions.

It has previously been noted that there is an increased demand for health care services as a result of such factors as the aging population and the higher expectations of consumers. In response to this, the scrutiny of health care delivery has increased and the focus has shifted to ensure only those services that are effective are provided. The interest of the health care policy maker and economist has become effectiveness and efficiency. The result of these many issues is a growing expectation that health care interventions and services have a sound scientific base.

Unfortunately, meeting these expectations has proven difficult. Firstly, it has been argued that the ever increasing volume of literature has made identifying those studies addressing a topic of interest very difficult. While electronic databases have improved the process of research identification, issues such as poor referencing and incomplete cover of health care journals all hinder the search process. An additional problem for nursing is the hesitancy of nurses to publish their research. As a consequence, it is suggested that some research will be lost to the profession as it lies hidden in the mountain of literature stored in libraries.

Quality of research has also emerged as a hindrance to ensuring a sound scientific basis for practice. The variable quality of research relates to such things as small sample sizes, inappropriate research methods, poor statistical analysis, and inadequate reporting of method and results. These factors have contributed to the many studies that are published with negative results or contradictory findings. This means that it is not only difficult to find the research, it is just as hard to decide if the findings are valid. Often, the end-user of the research must choose between many competing and contradictory findings. These issues are compounded by nursing's reluctance to use experimental research methods to investigate the cause and effect relationships of interventions. It has been argued in this thesis that deciding which research should form the scientific base for practice is a complex and difficult process.

Reviews have emerged as an important link between primary researcher and the end-user of the research. These reviews not only summarise available research, they also produce a 'stocktake' of all that is known about a topic. Increasingly, these reviews are replacing primary research in the decision process. However, the limitations of existing literature reviews have been noted. In response to these limitations the systematic review has emerged as a means to rigorously summarise best evidence on topics of interest.

Systematic reviews have been subject to extensive methodological development and as result, the findings are seen to be as rigorous as those produced by primary research. However, as the methodological development has been undertaken predominantly by medicine, it has been argued that current systematic review methods filter the research of nursing according to criteria defined by other disciplines. Currently, this filtering of research is in terms of effectiveness, and so research addressing perspectives other than effectiveness, have been seen to be beyond the scope of systematic reviews.

This need for evidence and the increasingly important role of reviews, combined with the inadequacy of current review methods to rigorously summarise research focusing

on issues other than effectiveness, provided impetus for the development of the expanded systematic review process reported in this thesis. The argument presented in this thesis is that the evaluation of health care services and interventions should be broader than effectiveness and should also incorporate the perspective of the person and the health care environment in which the intervention is located. The foci of effectiveness, appropriateness and feasibility was proposed as part of the conceptual framework to provide a more comprehensive evaluation addressing the many factors that impact on the success of an intervention. While these factors are acknowledged as having an impact on health care, they have fitted poorly into the evidence-based movement. This expanded review process and conceptual framework provides a means by which this evidence can be incorporated into the scientific basis of practice.

The methods used during the expanded systematic review to identify relevant research, appraise its validity then synthesise the findings were based on the processes used during systematic reviews of RCTs. However, as many areas of nursing practice have a relatively poor research base, the aim of this process was to both summarise and synthesise research findings. The summary component of the review was achieved through narrative and tabular summaries, while the synthesis was through meta-analysis and content analysis.

The development of these processes was based on approaches used for RCTs or observational studies, and informed by existing reviews of interpretive research. Merging the developmental work identified in the literature and in published reviews with the systematic approach used to summarise RCTs provided a viable and valid approach to summarise available evidence evaluating an intervention from a range of different perspectives.

As part of the review process, a new hierarchy of evidence was developed to better fit this broader evaluation of health care. This hierarchy acknowledges the valid contribution of a variety of research methods, and helps locate the contribution of

different types of research evidence within a logical review framework. It was argued that ranking the contribution of research according to appropriateness, feasibility and effectiveness allowed a more useful summary of the research. It was also argued that this approach to ranking research provided a valid framework for the conduct of the expanded systematic review.

While the evaluation of the proposed expanded review process can only be achieved through the conduct of a range of different systematic reviews, the findings of the music and physical restraint reviews suggest that it can successfully combine a broad range of research. The music systematic review demonstrated that the conceptual framework and expanded review process enabled a topic to be evaluated from a number of different perspectives. The physical restraint review demonstrated that the methods proposed in the expanded review process allowed different types of data to be synthesised.

In summary, health care literature benefits from a regular stocktaking and summary to bring together the current state of knowledge on a topic of interest. However focusing only on effectiveness excludes the important perspective generated by other research designs. To be useful to the nursing profession, reviews must summarise all valid research on a topic of interest. This evaluation of health care should not be limited to effectiveness, but must also incorporate the person's perspective, and that of the environment in which the care is to be implemented. The expanded review proposed in this thesis provides a template that helps achieve this. It is argued that this approach to reviewing the research provides information that better meets the demands and expectations of health care practice.

# **Appendix 1**

## **Search Strategy**

## Search Strategy Details

### Database Searches

#### CINAHL

#1	257	music in ti
#2	224	music in ab
#3	448	music in de
#4	71393	DT=RESEARCH
#5	178	(#1 or #2 or #3) and #4

#### MEDLINE

#1	389	music in ti
----	-----	-------------

#### Current Contents

#1	3533	music in ti
#2	2333	#1 and (la = english)
#3	15	#2 and (hospital* or patient*)
#1	3533	music in ti
#2	2333	#1 and (la = english)
#3	6	#2 and (surg* or pre?operative)
#1	3533	music in ti
#2	70056	cancer or chemotherapy or radiotherapy
#3	2	#1 and #2
#1	3533	music in ti
#2		myocardial or infarction*
#3	3	#1 and #2
#1	3533	music in ti
#2		procedure or invasive or insertion
#3	10	#1 and #2



#1	3533	music in ti
#2	2333	#1 and (la = english)
#3	275	yr = 2000

### **Cochrane Library**

music in ti (n = 61)

### **Expanded Academic**

#1 ? number music in ti

### **PsycLit**

#1	3125	music in ti
#2	376472	PT=EMPIRICAL
#3	1084	#1 and #2
#4	1155355	LA=ENGLISH
#5	1000	#3 and #4

### **Health Star**

- music in ti  
(n = 5 papers - but not all studies)

### **AUSThealth Databases**

1. Rural
  - music in ti  
(n = 0)
2. Australian Public Affairs Information Service - Health
  - music in ti  
(n = 2 papers)
3. Health & Society on Australian Health
  - music in ti  
(n = 0)
4. Aboriginal & Torres Strait Islander Health
  - music in ti  
(n = 157 papers)

5. Australian Medical Index
- music in ti  
(n = 5 papers)

### **Hand Searches**

Journal of Music Therapy

- from 1964 Vol. 1, No. 1 to 2000 Vol. 37, No. 2

International Journal of Arts in Medicine

- from 1991 Vol. 1 to 1994 Vol. 3.

Music Therapy

- 1991 Vol. 10 to 1994 Vol. 12

The Australian Journal of Music Therapy

- 1996 Vol. 7 to 1999 Vol. 10

### **Details of Researchers Contacted**

Ass Prof Routhieaux

paper - Music in waiting rooms

- has none of the data
- contacted 2nd researcher (Ass Prof David Tansik) and data obtained.

Dr Linda Chlan

paper - Effectiveness of a music therapy intervention on relaxation and anxiety...

Missing Data Provided

Assistant Professor Jill White

School of Nursing

University of Wisconsin-Milwaukee

paper - Effects of relaxing music on cardiac autonomic balance and anxiety after myocardial infarction

Data provided

### **Authors that could Not be Contacted**

Miluk-Kolasa, B. and Stupnicki, R. 1996, The effects of music listening on changes in selected physiological parameters in adult pre-surgical patients, *Journal of Music Therapy*, 3, 208-18.

Taylor, L.K., Kuttler, K.L., Parks, T.A. and Milton, D., 1998, The effect of music in the postanesthetic care unit on pain levels in women who have had abdominal hysterectomies, *Journal of PeriAnesthetic Nursing*, 13, 2, 88-94.

Koch, M.E., Kain, Z.N., Ayoub, C. and Rosenbaum, S.H. 1998, The sedative and analgesic sparing effect of music, *Anesthesiology*, 89, 2, 300-306.

Palakanis, K.C., Denobile, J.W., Sweeney, W.B. and Blanship, C.L., 1994, Effect of music therapy on state anxiety in patients undergoing flexible sigmoidoscopy, *Diseases of the Colon and Rectum*, 37, 478-481.

Cruise, C.J., Chung, F., Yogendran, S. and Little, D., 1997, Music increases satisfaction in elderly outpatients undergoing cataract surgery, *Canadian Journal of Anaesthesiology*, 44, 1, 43-8.

# **Appendix 2**

## **Critical Appraisal Checklists**

# Appendix 1

## Appraisal of Studies

### CRITICAL APPRAISAL

<b>RCT</b>	<b>Cohort Studies</b>	<b>Case Control Studies</b>
<ul style="list-style-type: none"><li>• Allocation bias</li><li>• Performance bias</li><li>• Detection bias</li><li>• Attrition bias</li></ul>	<ul style="list-style-type: none"><li>• Selection</li><li>• Exposure</li><li>• Detection</li><li>• Attrition</li></ul>	<ul style="list-style-type: none"><li>• Selection</li><li>• Exposure</li><li>• Detection</li><li>• Attrition</li></ul>

### MINIMUM STANDARD

<b>Interpretive</b>	<b>Descriptive</b>
<ul style="list-style-type: none"><li>• Method appropriate</li><li>• Reporting of method</li><li>• Description of Participants</li><li>• Supportive data of themes &amp; labels</li></ul>	<ul style="list-style-type: none"><li>• Method appropriate</li><li>• Reporting of method</li><li>• Description of Participants</li><li>• Supportive data provided</li></ul>

**Critical Appraisal Tool**  
**Randomised Controlled Trial Checklist**

Author  Year  Ref Number

1) Were the participants randomised to study groups

yes  no  not clear

2) Other than research intervention, were participants in each groups treated the same.

yes  no  not clear

3) Were the outcomes measured in the same manner for all participants

yes  no  not clear

4) Was there adequate follow-up of participants

yes  no  not clear   
(less than 80% followed up)

## Critical Appraisal Tool

### Observational Study Checklist : Cohort Studies

Author  Year  Ref Number

1) Were the non-exposed from the same population as the exposed.

yes  no  not clear

2) Was exposure reliably ascertained and verified.

yes  no  not clear

3) Were the outcomes measured in the same manner for both cohorts.

yes  no  not clear

4) Were the drop-out rates similar in the exposed and unexposed cohorts.

yes  no  not clear

## Critical Appraisal Tool

### Observational Study Checklist : Case Control Studies

Author  Year  Ref Number

1) Were the controls selected from a similar population as the case.

yes  no  not clear

2) Were the controls free from the condition / disease of interest.

yes  no  not clear

3) Was the case definition clear

yes  no  not clear

4) Were the non-response rates similar in both groups.

yes  no  not clear



**Critical Appraisal Tool**  
**Interpretive Study Checklist**

Author  Year  Ref Number

1) Was the research design appropriate for the research topic or question.

yes  no  not clear

2) Was there sufficient description of the method of information collection and analysis.

yes  no  not clear

3) Was there a clear description of the study informants / participants.

yes  no  not clear

4) Was there sufficient supportive data to adequately describe all themes, categories and labels.

yes  no  not clear

**Critical Appraisal Tool**  
**Descriptive Study Checklist**

Author  Year  Ref Number

1) Was the research design appropriate for the research topic or question.

yes  no  not clear

2) Was there sufficient description of the method of data collection and analysis.

yes  no  not clear

3) Was there a clear description of study population and setting

yes  no  not clear

4) Was sufficient data provided to support the findings of the study.

yes  no  not clear

# **Appendix 3**

## **Data Collection Form**

## Extraction Form

Author

Record Number

Journal

Year

**Method**

---

---

**Setting**

---

**Participants**

---

---

**Number of Participants**

Group A

Group B

Group C

**Interventions**

Intervention A

---

---

---

Intervention B

---

---

---

Intervention C

---

---

---

## Outcome Measures Used in Study

Anxiety

scale

---

---

Stress

scale

---

---

Pain

scale

---

---

Comfort

scale

---

---

Satisfaction

scale

---

---

Other

---

---

---

---

---

---

---

---

## Results

### Dichotomous Data

Outcome	Treatment Group number / total number	Control Group number / total number

### Continuous Data

Outcome	Treatment Group SD & mean (number)	Control Group SD & mean (number)

**Missing Data / Comments**

---

---

---

---

**Interpretive Data**

Theme

---

---

---

Theme

---

---

---

Theme

---

---

---

Theme

---

---

---

**Descriptive Data**

---

---

---

---

**Authors Conclusions**

---

---

---

# **Appendix 4**

## **Summary Tables**



**TABLE 1****Included RCTs Evaluating Music in Hospital Patients**

<b>Author</b>	<b>Population</b>	<b>Number</b>	<b>Music</b>	<b>Outcome Measures</b>	<b>Results</b>
Barnason et al (Barnason et al. 1995)	Post-op Coronary artery bypass graft.	music n = 33 no-music n = 34	Selected	STAI, anxiety VAS, mood scale, vital signs.	No difference in STAI, VAS or vital signs. Significant improvement mood.
Bolwerk (Bolwerk 1990)	Post MI with STAI score > 40	music n = 17 no-music n = 18	Classical	STAI	Music reduced STAI
Chlan (Chlan 1995)	Mechanically ventilated	music n = 9 no-music n = 11	Selected	Vital signs, cardiac rhythm, airway pressure, oxygen saturation, POMS.	Significant difference for HR, RR and POMS score. No difference in other parameters.
Chlan (Chlan 1998)	Mechanically ventilated	music n = 27 no-music n = 27	Selected	Modified STAI, vital signs.	Significant difference in STAI, HR and RR
Gaberson (Gaberson 1995)	Pre-operative	music n = 16 no-music n = 15	Tranquil	anxiety VAS,	No difference between groups

**TABLE 1 (continued)****Included RCTs Evaluating Music in Hospital Patients**

<b>Author</b>	<b>Population</b>	<b>Number</b>	<b>Music</b>	<b>Outcome Measures</b>	<b>Results</b>
Miluk-Kolasa et al (Miluk-Kolasa et al. 1996)	Pre-operative after being told about surgery	music n = 50 no-music n = 50	Selected	Vital signs, cardiac output, skin temp (every 20 minutes for 1 hour)	More rapid return to normal levels in music group
Taylor et al (Taylor et al. 1998)	Post-operative women	total n = 61 music n = ? no-music n = ?	Selected	Two pain scales	No difference in pain scores
White (White 1992)	Post MI	music n = 20 no-music n = 20	Classical	STAI, vital signs	Significant difference in HR, RR, STAI.
White (White 1999a)	Post MI	music n = 15 no-music n = 15	Classical	STAI, vital signs (taken every 20 minutes for 1 hour)	Significant difference in HR, RR, STAI.

(STAI = State Trait Anxiety Inventory, VAS = visual analogue scale, MI = myocardial infarction, POMS = profile of mood states, HR = heart rate, RR = respiratory rate)

**TABLE 2**

**Included RCT Evaluating Music During Invasive or Unpleasant Procedures**

<b>Author</b>	<b>Population</b>	<b>Numbers</b>	<b>Music</b>	<b>Outcome Measures</b>	<b>Results</b>
Bampton et al (Bampton and Draper 1997)	Upper GI Investigation or colonoscopy	music n = 28 no-music n = 31	new age	tolerance VAS (assessed by patient & nurse), asked would they have procedure again	No difference between groups for patient and nurse assessed tolerance. More in no-music group rated procedure moderately unpleasant or worse.
Blankfield et al (Blankfield et al. 1995)	Coronary artery bypass surgery (intra-operative)	music n = 32 no-music n = 29	classical	Duration of post-op stay, days in SICU, Post discharge measures = ADL, cardiac symptom scale.	No difference between groups for any outcomes.
Broschious (Broschious 1999)	Cardiac surgical patients during chest tube	music n = 70 no-music n = 50	Instrumental	Pain score VAS, vital signs.	No difference between groups.
Colt et al (Colt et al. 1999)	Fiberoptic bronchoscopy	music n = 30 no-music n = 30	piano.	STAI	No difference between groups.
Cruise et al (Cruise et al. 1997)	cataract surgery (intra-operative)	music n = 32 no-music n = 30	classical and sounds of nature.	STAI, anxiety VAS, satisfaction VAS, vital signs.	No difference for anxiety (STAI or VAS). Difference in BP.
Good (Good 1995)	Surgical patients during first post-op ambulation	music n = 21 no-music n = 21	instrumental	STAI, Sensations of Pain scale, Distress of Pain scale, McGill Pain Questionnaire, amount of narcotics.	No difference between groups.

**TABLE 2 (continued)**

**Included RCT Evaluating Music During Invasive or Unpleasant Procedures**

<b>Author</b>	<b>Population</b>	<b>Numbers</b>	<b>Music</b>	<b>Outcome Measures</b>	<b>Results</b>
Koch et al (Koch et al. 1998) <u>Study One</u>	Urological procedures (intra-operative) with patient controlled sedation device.	music n = 15 no-music n = 19	Selected	Level of sedation, vital signs, amount of sedation.	No difference between groups for vital signs or level of sedation. Music group used less sedation.
Koch et al (Koch et al. 1998) <u>Study Two</u>	Lithotripsy (intra-operative) with PCA	music n = 21 no-music n = 21	Selected	STAI, Pain scale, vital signs, amount of analgesia used.	No difference for STAI, pain or vital signs. Music group used less analgesia.
Palakanis et al (Palakanis et al. 1994)	Flexible sigmoidoscopy (intra-operative)	music n = 25 no-music n = 25	Selected	STAI, vital signs.	No difference in vital signs. Reduction in STAI.
Walther-Larsen et al (Walther-Larsen et al. 1988)	Orthopaedic & plastic surgery under regional anaesthesia (intraoperative).	music n = 32 no-music n = 32	Selected	Requests for sedatives, satisfaction, anxiety, would they have regional block again.	More in music group anxious. Music reduced requests for sedatives.

STAI = State Trait Anxiety Inventory, VAS = visual analogue scale, GI = gastrointestinal SICU = surgical intensive care unit, ADL = activities of daily living, BP = blood pressure.

## TABLE 3

### Excluded Clinical Trials

Clinical trials that evaluated music with adults in the hospital setting that were excluded from the meta-analyses.

<b>Author</b>	<b>Population</b>	<b>Outcomes</b>	<b>Reason for Exclusion</b>
Augustin et al (Augustin and Hains 1996)	Pre-operative patients (n = 42)	STAI, vital signs	Method of randomisation (alternation). Performance (major differences in the treatment of the two study groups).
Bonny (Bonny 1983)	cardiac patients (n = 26)	Heart rate, BP, emotional rating scale, and observations of the patients reactions.	Not RCT (before and after study)
Bonny <i>et al.</i> (Bonny and McCarron 1984)	2 separate groups (n = 25) - patients undergoing light anaesthesia - patients undergoing regional block	comments by patients, & observations by personnel	Not RCT (2 uncontrolled trials)
Byers <i>et al.</i> (Byers and Smyth 1997)	Cardiac surgery patients (n = 40)	noise level, heart rate, blood pressure noise annoyance	Not RCT (cross-over study)

## TABLE 3 (continued)

### Excluded Clinical Trials

Author	Population	Outcomes	Reason for Exclusion
Cunningham et al 1997 (Cunningham et al. 1997)	Pre-operative patients (n = 50)	Questionnaire	RCT Method of randomisation (by day)
Daub <i>et al.</i> 1988 (Daub and Kirschner-Hermanns 1988)	Pre-operative patients (n = 90)	STAI	RCT Non English language report (& insufficient data provided in English abstract)
Dubois et al 1995 (Dubois et al. 1995)	Patients undergoing bronchoscopy (n = 49)	Borg scale, vital signs.	RCT Method of randomisation (record number)
Eisenman <i>et al.</i> 1995 (Eisenman and Cohen 1995)	Patients undergoing regional anaesthesia (n = 30)	questionnaire	Not RCT (uncontrolled trial) Minimal data provided
Grey et al 2000 (Grey et al. 2000)	Patients undergoing magnetic resonance imaging (n = 64)	STAI	RCT Music combined with other interventions Method of randomisation (by day)
Heiser et al 1997 (Heiser et al. 1997)	Post-operative patients (n = 34)	Pain VAS, anxiety VAS, vital signs, analgesia, satisfaction.	RCT Attrition (15 of 34 did not complete)

## TABLE 3 (continued)

### Excluded Clinical Trials

Author	Population	Outcomes	Reason for Exclusion
Kopp 1991 (Kopp 1991)	Local anaesthetic intra-operative surgical patients (n = 41)	Self developed rating scale.	RCT Inadequate information provided about population to allow assessment.. No results data provided.
Miller et al 1992 (Miller et al. 1992)	Patients receiving dressing to burn (n = 17)	STAI, McGill Pain Questionnaire	RCT Music combined with video.
Sabo et al 1996 (Sabo and Michael 1996)	Patients undergoing chemotherapy (n=100)	STAI, chemotherapy side effects	RCT Method of randomisation (by office) Music combined with message
Schuster 1985 (Schuster 1985)	Patients undergoing dialysis (n = 63)	BP	Controlled trial
Standley 1992 (Standley 1992)	Patients receiving chemotherapy (n = 15)	nausea	Not RCT (non-randomised controlled trial)
Stevens 1990 (Stevens 1990)	Patients undergoing surgery (n = 25)	Interviewed 20 hours after surgery.	Not RCT (uncontrolled trial)

## TABLE 3 (continued)

### Excluded Clinical Trials

Author	Population	Outcomes	Reason for Exclusion
Szeto et al 1999 (Szeto and Yung 1999)	Pre-operative patients (n = 12)	STAI, tension scale	RCT Attrition (3 of 6 in control group did not complete)
Thornby <i>et al.</i> 1995 (Thornby et al. 1995)	Patients with COPD during exercise tests (n = 36)	Pulmonary measures	Not RCT - all randomly under went the experimental interventions (music, grey noise and silence).
Weber <i>et al.</i> 1996 (Weber et al. 1996)	Cancer patients during Chemotherapy (n = 33) Two groups (both had music) 1) those having chemotherapy for 1st time 2) those who had previously had chemotherapy	Questionnaire STAI	No comparison group (both groups had music) convenience sample
Winter et al 1994 (Winter et al. 1994)	Pre-operative patients (n = 62)	STAI, vital signs	RCT Attrition (12 of 31 in control group did not complete)



**TABLE 4****Interpretive and Descriptive Studies**

Interpretive and descriptive studies that explored some aspect of music in the hospital setting.

<b>Author</b>	<b>Population</b>	<b>Music</b>	<b>Method</b>	<b>Results</b>
Stevens 1990 (Stevens 1990)	Hospital patients undergoing surgery with local or regional anaesthetic n = 25 (of which 20 had listened to music)	Via headphones during surgery	Interpretive Structured interviews and attitudinal scale.	<ul style="list-style-type: none"> <li>- Some wished for specific types of music</li> <li>- Only 1 did not like music</li> <li>- 75% of participants rated music as either very good or excellent</li> <li>- Findings supported by positive participant statements.</li> </ul>
Michel <i>et al.</i> 1995 (Michel and Chesky 1995)	Music Therapists using music for pain relief questionnaires sent n = 348 questionnaires returned n = 139	Music used specifically for pain relief	Survey	<ul style="list-style-type: none"> <li>- Description of clients who receive music for pain relief</li> <li>- Rationales for use of music</li> <li>- Description of outcome assessment.</li> </ul>
Pratt 1999 (Pratt 1999)	Interview with nurse and research director on the use of music during surgery in one institution.	Music during surgery	Interview with music therapists.	Description of music during surgery program



# References

- Altman, D.G. 1982. Statistics in medical journals. *Statistics in Medicine* 1:59-71.
- Altman, D.G. 1994. The scandal of poor medical research. *British Medical Journal* 308:283-284.
- Anthony, D. 1996. A review of statistical methods in the Journal of Advanced Nursing. *Journal of Advanced Nursing* 24:1089-94.
- Antman, E.M., Lau, J., Kupeinick, B., Mosteller, F., and Chalmers, T.C. 1992. A comparison of results of meta-analyses of randomised controlled trials and recommendations of clinical experts. *Journal of the American Medical Association* 268 (2):240-248.
- Armstrong, L., and Standsfield, J. 1996. A content analysis of the professional journal of the British Society of Speech Therapists I: the first 10 years. Spotlight on 'speech' 1935-1945. *European Journal of Disorders of Communication* 31 (2):91-105.
- Augustin, P., and Hains, A.A. 1996. Effect of music on ambulatory surgery patients' preoperative anxiety. *AORN* 63:750-758.
- Bailar, J.C. 1997. The promise and problems of meta-analysis. *New England Journal of Medicine* 337:559-561.
- Bailey, L.M. 1983. The effects of live music versus tape-recorded music on hospitalized cancer patients. *Music Therapy* 3 (1):17-28.
- Ball, C., Sackett, D.L., Phillips, B., Haynes, B., and Straus, S. 1998. Levels of evidence and grading recommendations: Centre for Evidence Based Medicine.
- Bampton, P., and Draper, B. 1997. Effect of relaxation music on patient tolerance of gastrointestinal endoscopic procedures. *Journal of Clinical Gastroenterology* 25 (1):343-5.
- Barnason, S., Zimmerman, L., and Nieveen, J. 1995. The effects of music interventions on anxiety in the patient after coronary artery bypass grafting. *Heart Lung* 24 (2):124-32.
- Barroso, J., and Powell-Cope, G.M. 2000. Metasynthesis of qualitative research on living with HIV infection. *Qualitative Health Research* 10 (3):340-353.

- Basche, C.E. 1987. Focus groups interview: an underutilized research technique for improving theory and practice in health education. *Health Quarterly* 14:411-418.
- Bastian, H. 1994. *The power of sharing knowledge. Consumer participation in the Cochrane Collaboration*. Oxford: UK Cochrane Centre.
- Batstone, G., and Edwards, M. 1996. Evidence based practice: converts, cynics and those in between. *Journal of Clinical Effectiveness* 1 (4):123.
- Beaudin, C.L., and Pelletier, L.R. 1996. Consumer-based research: using focus groups as a method for evaluating quality of care. *Journal of Nursing Care Quality* 10 (3):28-33.
- Beck, S.L. 1991. The therapeutic use of music for cancer-related pain. *Oncology Nursing Forum* 18 (8):1327-37.
- Becker, B.J. 1994. Combining significance levels. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Begg, C.B. 1994. Publication Bias. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Benitez-Bribiesca, L. 1999. Evidence based medicine: a new paradigm? *Archives of Medical Research* 30:77-79.
- Benning, S.P., and Speer, S.C. 1993. Incorrect citations: a comparison of library literature and medical literature. *Bulletin of the Medical Library Association* 81:56-8.
- Benson, K., and Hartz, A.J. 2000. A comparison of observational studies and randomised controlled trials. *New England Journal of Medicine* 342 (25):1878-1886.
- Bero, L.A., Galbraith, A., and Rennie, D. 1992. The publication of sponsored symposiums in medical journals. *The New England Journal of Medicine* 327 (16):1135-1140.
- Bhopal, R.S., and Tonks, A. 1994. The role of research letters in reviewing research. *British Medical Journal* 308:1582-3.

- Black, D. 1998. The limitations of evidence. *Journal of the Royal College of Physicians of London* 32 (1):23-26.
- Black, N. 1996. Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal* 312:1215-1218.
- Blankfield, R.P., Zyzanski, S.J., Flocke, S.A., Alemagno, S., and Scheurman, K. 1995. Taped therapeutic suggestions and taped music as adjuncts in the care of coronary-artery-bypass patients. *American Journal of Clinical Hypnosis* 37 (3):32-42.
- Bolwerk, C.A.L. 1990. Effects of relaxing music on state anxiety in myocardial infarction patients. *Critical Care Nursing Quarterly* 13 (2):63-72.
- Bonell, C. 1999. Evidence-based nursing: a stereotyped view of quantitative and experimental research could work against professional autonomy and authority. *Journal of Advanced Nursing* 30 (1):18-23.
- Bonny, H.L. 1983. Music listening for intensive coronary care units: A pilot project. *Music Therapy* 3 (1):4-16.
- Bonny, H.L., and McCarron, N. 1984. Music as an adjunct to anesthesia in operative procedures. *Journal of the American Association of Nurse Anesthetists* 52 (1):55-7.
- Booth, A. 1996. In search of the evidence: informing effective practice. *Journal of Clinical Effectiveness* 1 (1):25-29.
- Brazier, H., and Begley, C.M. 1996. Selecting a database for literature searches in nursing: MEDLINE or CINAHL. *Journal of Advanced Nursing* 24:868-875.
- Brenner, S.H., and McKinin, E.J. 1989. CINAHL and MEDLINE: a comparison of indexing practices. *Bulletins of the Medical Library Association* 77 (4):366-371.
- Brewin, C.R., and Bradley, C. 1989. Patient preferences and randomised clinical trials. *British Medical Journal* 299:313-315.
- Broscious, S.K. 1999. Music: an intervention for pain during chest tube removal after open heart surgery. *American Journal of Critical Care* 8 (6):410-5.
- Burns, N. 1989. Standards for qualitative research. *Nursing Science Quarterly* 2:44-52.

- Burton, L.C. 1992. Physical restraint use and cognitive decline among nursing home residents. *Journal of the American Geriatrics Society* 40:811-816.
- Bushman, B.J. 1994. Vote-counting procedures in meta-analysis. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Byers, J.F., and Smyth, K.A. 1997. Effect of a music intervention on noise annoyance, heart rate, and blood pressure in cardiac surgery patients. *American Journal of Critical Care* 6 (3):183-91.
- Caine, J. 1991. The effects of music on the selected stress behaviors, weight, caloric and formula intake, and length of hospital stay of premature and low birth weight neonates in a newborn intensive care unit. *Journal of Music Therapy* 28 (4):180-192.
- Canadian Task Force on the Periodic Health Examination. 1979. The periodic health examination. *Canadian Medical Association Journal* 121:1193-1254.
- Carr-Hill, R. 1995. Welcome to the brave new world of evidence based medicine. *Social Science Medicine* 41 (11):1467-1468.
- Carruthers, S.G., Larochele, P., Haynes, R.B., Petrasovits, A., and Schiffrin, E.L. 1993. Report of the Canadian Hypertension Society consensus conference: 1. introduction. *Canadian Medical Association Journal* 149 (3):289-293.
- Cavanagh, S. 1997. Content analysis: concepts, methods and applications. *Nurse Researcher* 4 (3):5-16.
- CDC. 1996. Guidelines for the prevention of intravascular device-related infections. *American Journal of Infection Control* 24 (4):262-293.
- Chalmers, T.C. 1972. Randomization and coronary artery surgery. *Annals of Thoracic Surgery* 14:323.
- Chalmers, T.C., Celano, P., Sacks, H.S., and Smith, H. 1983. Bias in treatment assignment in controlled clinical trials. *New England Journal of Medicine* 309:1358-1361.

- Chapple, A., and Rogers, A. 1999. Self-care and its relevance to developing demand management strategies: a review of qualitative research. *Health and Social Care in the Community* 7 (6):445-454.
- Charlton, B.G. 1997. Book review of evidence-based medicine. *Journal of evaluation of Clinical Practice* 3:169-72.
- Charlton, B.G., and Miles, A. 1998. The rise and fall of EBM. *Quarterly Journal of Medicine* 91:371-4.
- Charnock, Y., and Evans, D. 2001. Nursing management of chest drains: a systematic review. *Australian Critical Care Journal* in press.
- Chlan, L. 1998. Effectiveness of a music therapy intervention on relaxation and anxiety for patients receiving ventilatory assistance. *Heart Lung* 27 (3):169-76.
- Chlan, L., and Tracy, M.F. 1999. Music therapy in critical care: indications and guidelines for intervention. *Critical Care Nurse* 19 (3):35-41.
- Chlan, L.L. 1995. Psychophysiologic responses of mechanically ventilated patients to music: a pilot study. *American Journal of Critical Care* 4 (3):233-8.
- . Cinahl Information Systems, Glendale, California.
- Clark, A.M., Curzio, J., Lindsay, G.M., Fleming, V.E.M., and McIntosh, J. 1998. Exploring patients' perspective of coronary heart disease: discerning methods and a review of the qualitative literature. *Coronary Health Care* 2:118-128.
- Clarke, J.B. 1999. Evidence-based practice: a retrograde step? The importance of pluralism in evidence generation for the practice of health care. *Journal of Clinical Nursing* 8:89-94.
- Cochrane Collaboration. 1999. *Review Manager 4.0.4*: Wintertree Software Inc.
- Colditz, G.A., Miller, J.N., and Mosteller, F. 1989. How study design affects outcomes in comparisons of therapy. I: Medical. *Statistics in Medicine* 8:441-454.
- Colt, H.G., Powers, A., and Shanks, T.G. 1999. Effect of music on state anxiety scores in patients undergoing fiberoptic bronchoscopy. *Chest* 116 (3):819-24.



- Colyer, H., and Kamath, P. 1999. Evidence based practice. A philosophical and political analysis: some matters for consideration by practitioners. *Journal of Advanced Nursing* 29 (1):188-93.
- Concato, J., Shah, N., and Horwitz, R.I. 2000. Randomised controlled trials, observational studies and the hierarchy of research designs. *New England Journal of Medicine* 342 (25):1887-1892.
- Cook, D.J., Sibbald, W.J., Vincent, J.L. and Cerra, F.B. 1996. Evidence based critical care medicine: what is it and what it can do for us. *Critical Care Medicine* 24 (2):334-337.
- Cook, D.J., Guyatt, G.H., Laupacis, A., and Sackett, D.L. 1992. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 102 (4):305s - 311s.
- Cook, D.J., Guyatt, G.H., Laupacis, A., Sackett, D.L., and Goldberg, R.J. 1995. Clinical recommendations using levels of evidence for antithrombotic agents. *Chest* 108 (4):227s - 230s.
- Cook, D.J., Mulrow, C.D., and Haynes, B. 1998. Synthesis of best evidence for clinical decisions. In *Systematic reviews: synthesis of best evidence for health care decisions*, edited by Mulrow, C.D. and Cook, D. Philadelphia: American College of Physicians.
- Cook, D.J., Mulrow, C.D., and Haynes, R.B. 1997. Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of Internal Medicine* 126:376-380.
- Cooper, H. 1984. *The Integrative Research Review: A Systematic Approach*. Beverly Hills: Sage Publications.
- Cooper, H., and Hedges, L.V. 1994. *The Handbook of Research Synthesis*. New York: Sage Foundation.
- 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.
- Critical Reviews Advisory Group. 1996. Introduction to Systematic Reviews. Sheffield: School for Health and Related Research (ScHARR), Sheffield University.
- Cruise, C.J., Chung, F., Yogendran, S., and Little, D. 1997. Music increases satisfaction in elderly outpatients undergoing cataract surgery. *Canadian Journal of Anaesthesia* 44 (1):43-8.
- Cunningham, M.F., Monson, B., and Bookbinder, M. 1997. Introducing a music program in the perioperative area. *Aorn* 66 (4):674-82.
- Curtis, S.L. 1986. The effect of music on pain relief and relaxation of the terminally ill. *Journal of Music Therapy* 23 (1):10-24.
- Daub, D., and Kirschner-Hermanns, R. 1988. Reduction of preoperative anxiety. A study comparing music, Thalamonal and no premedication. *Anaesthesist* 37 (9):594-7.
- Davies, H.T.O., Crombie, I.K., and Tavakoli, M. 1998. When can odds ratios mislead. *British Medical Journal* 316:989-91.
- Dawson-Saunders, B., and Trapp, R.G. 1994. *Basic and Clinical Biostatistics*. London: Prentice Hall International.
- Denney, A. 1997. Quiet music. An intervention for mealtime agitation? *Journal of Gerontological Nursing* 23 (7):16-23.
- Dewar, A.L., and Morse, J.M. 1995. Unbearable incidents: failure to endure the experience of illness. *Journal of Advanced Nursing* 22 (5):957-64.
- Dickersin, K., Scherer, R., and Lefebvre, C. 1994. Identifying relevant studies for systematic reviews. *British Medical Journal* 309:1286-1291.
- Dickersin, K., and Berlin, J.A. 1992. Meta-analysis: state of the science. *Epidemiologic Review* 14:154-176.
- Dickersin, K., Min, Y., and Meinert, C.L. 1992. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *Journal of the American Medical Association* 267 (3):374-378.

- Downe-Wambolt, B. 1992. Content analysis: methods, applications and issues. *Health Care for Women International* 13 (3):313-321.
- Dube, A., and Mitchell, E. 1986. Accidental strangulation from vest restraint. *Journal of the American Medical Association* 256 (2725-2726).
- Dubois, J.M., Bartter, T., and Pratter, M.R. 1995. Music improves patient comfort level during outpatient bronchoscopy. *Chest* 108 (1):129-30.
- Duke, R.J., Bloch, R.F., Turpie, A.G.G., Trebilcock, R., and Bayer, N. 1986. Intravenous heparin for the prevention of stroke progression in acute partial stable stroke: A randomised controlled trial. *Annals of Internal Medicine* 105:825-828.
- Dun, B. 1995. A different beat: Music therapy in children's cardiac care. *Music Therapy Perspectives* 13 (1):35-39.
- Dunn, S., Pretty, L., Reid, H., and Evans, D. 2000. *Management of short term indwelling urethral catheters to prevent urinary tract infections: a systematic review*. Adelaide: The Joanna Briggs Institute.
- Dunn, S.V. 1998. Perspectives on evidence-based practice. *Australian Critical Care* 11 (1):2-4.
- du-Pre, A., and Beck, C.S. 1997. "How can I put this?" Exaggerated self-disparagement as alignment strategy during problematic disclosures by patients to doctors. *Qualitative Health Research* 7 (4):487-503.
- DuRant, R.H., Romes, E.S., Rich, M., Allred, E., Emans, S.J., and Woods, E.R. 1997. Tobacco and alcohol use behaviours portrayed in music videos: a content analysis. *American Journal of Public Health* 87 (7):1131-1135.
- Durham, L., and Collins, M. 1986. The effect of music as a conditioning aid in prepared childbirth education. *Journal of Obstetrics Gynecology and Neonatal Nursing* 15:268-70.
- Easterbrook, P.J., Berlin, J.A., Gopalan, R., and Matthews, D.R. 1991. Publication bias in clinical research. *The Lancet* 337 (8746):867-872.

- Echt, D.S., Liebson, P.R., and Mitchell, L.B. 1991. Mortality and morbidity in patients receiving ecainide, flecainide or placebo: The Cardiac Arrhythmia Suppression Trial. *New England Journal of Medicine* 324:781.
- Editorial. 1995. Evidence based medicine, in its place. *The Lancet* 346:785.
- Egger, M., Schneider, M., and Davey-Smith, G. 1998. Spurious precision? meta-analysis of observational studies. *British Medical Journal* 316:140-4.
- Eisenman, A., and Cohen, B. 1995. Practical innovations. Music therapy for patients undergoing regional anesthesia. *AORN Journal* 62 (6):947-50.
- Elwood, M. 1998. *Critical Appraisal of Epidemiological Studies and Clinical Trials*. 2nd ed. Oxford: Oxford University Press.
- Engel, J.D., and Kuzel, A.J. 1992. On the idea of what constitutes good qualitative inquiry. *Qualitative Health Research* 2 (4):504-10.
- Estabrooks, C.A., Field, P.A., and Morse, J.M. 1994. Aggregating qualitative findings: an approach to theory development. *Qualitative Health Research* 4:503-11.
- Evans, D. 2000. *Consultant Reviewers Manual*. Adelaide: The Joanna Briggs Institute.
- Evans, D. 2001. Systematic reviews of nursing research. *Intensive and Critical Care Nursing* 17:51-57.
- Evans, D., Hodgkinson, B., and Berry, J. 2000. Vital signs in hospital patients: a systematic review. *International Journal of Nursing Studies*:in press.
- Evans, D., Hodgkinson, B., Lambert, L., and Wood, J. 1999. Fall prevention: a systematic review. *Clinical Effectiveness in Nursing* 3:106-11.
- Evans, D., and Kowanko, I. 2000. Literature reviews: evolution of a research methodology. *Australian Journal of Advanced Nursing* 18 (2):31-36.
- Evans, D., and Pearson, A. 2001a. Systematic reviews of qualitative research. *Clinical Effectiveness in Nursing* in press.
- Evans, D., and Pearson, A. 2001b. Systematic reviews: gatekeepers of nursing knowledge. *Journal of Clinical Nursing*:in press.
- Evans, L.K., Strumpf, N.E., Allen-Taylor, L., Capezuti, E., Maislin, G., and Jacobsen, B. 1997. A clinical trial to reduce restraints in nursing homes. *The Journal of American Geriatrics Society* 45:675-681.

- Farby, R. 1993. Searching the literature: be creative with MEDLINE. *British Medical Journal* 307:66.
- Feinstein, A.R. 1989. Epidemiologic analysis of causation: the unlearned scientific lessons of randomised trials. *Journal of Clinical Epidemiology* 42:481-489.
- Feinstein, A.R. 1994. Clinical judgement revisited: the distraction of quantitative models. *Annals of Internal Medicine* 120:799-805.
- Feldman, K.A. 1971. Using the work of others: some observations on reviewing and integrating. *Sociology of Education* 4:86-102.
- Fleiss, J. 1994. Measures of effect size for categorical data. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Fleming, T.R., and DeMets, D.L. 1996. Surrogate end points in clinical trials: are we being misled. *Annals of Internal Medicine* 125 (605-13).
- Forchuk, C., and Roberts, J. 1993. How to critique qualitative research articles. *Canadian Journal of Nursing Research* 25 (4):47-56.
- Foreman, M.D., and Kirchoff, K.T. 1987. Accuracy of references in nursing journals. *Research in Nursing and Health* 10:177-83.
- Fowkes, F.G.R., and Fulton, P.M. 1991. Critical appraisal of published research: introductory guidelines. *British Medical Journal* 302:1136-40.
- Fredriksson, L. 1999. Modes of relating in a caring conversation: a research synthesis on presence, touch and listening. *Journal of Advanced Nursing* 30 (5):1167-76.
- Frengley, J.D., and Mion, L.C. 1986. Incidence of physical restraints on acute general medical wards. *Journal of the American Geriatric Society* 34 (8):565-568.
- Froom, P., and Froom, J. 1993. Deficiencies in structured medical abstracts. *Journal of Clinical Epidemiology* 46 (7):591-595.
- Funk, M.E., Reid, C.A., and McGoogan, L.S. 1988. The usefulness of monographic proceedings. *Bulletins of the Medical Library Association* 76 (1):14-21.
- Gaberson, K.B. 1995. The effect of humorous and musical distraction of preoperative anxiety. *AORN Journal* 62 (5):784, 786-7, 788 passim.
- Gaebler, S. 1994. Just who do we restrain. *Australian Nursing Journal* 2 (1):39-41.

- Garbett, R. 1998. Do the right thing. *Nursing Times* 94 (27):28-29.
- Geden, E.A., Lower, M., Beattie, S., and Beck, N. 1989. Effects of music and imagery on physiologic and self-report of analogued labor pain. *Nursing Research* 38 (1):37-41.
- Gerrish, K., and Clayton, J. 1998. Improving clinical effectiveness through an evidence based approach: meeting the challenge for nursing in the United Kingdom. *Nursing Administration Quarterly* Summer:55-65.
- 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*. Beverly Hills: Sage Publications.
- Goldberg, R., Newton, E., Cameron, J., Jacobson, R., Chan, L., Bukata, W.R., and Rakab, A. 1993. Reference accuracy in the emergency medicine literature. *Annals of Emergency Medicine* 22:1450-4.
- Goloff, M.S. 1981. The responses of hospitalized medical patients to music therapy. *Music Therapy* 1 (1):51-56.
- Good, M. 1995. A comparison of the effects of jaw relaxation and music on postoperative pain. *Nursing Research* 44 (1):52-7.
- Goodman, N.W. 1999. Who will challenge evidence-based medicine. *Journal of the Royal College of Physicians of London* 33 (3):249-51.
- Greenhalgh, T. 1997. How to read a paper: The MEDLINE database. *British Medical Journal* 315:180-183.
- Greenland, S. 1987. Quantitative methods in the review of epidemiological literature. *Epidemiological Review* 9:1-30.
- Gregoire, G., Derderian, F., and Le Lorier, J. 1995. Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias. *Journal of Clinical Epidemiology* 48 (159-63).
- Grey, S.J., Price, G., and Matthews, A. 2000. Reduction of anxiety during MR imaging: a controlled trial. *Magnetic Resonance Imaging* 18:351-5.

- Guyatt, G.H., Haynes, R.B., Jaeschke, R.Z., Cook, D.J., Green, L., Naylor, C.D., Wilson, M.C., and Richardson, W.S. 2000. Users' guide to the medical literature XXV. Evidence-based medicine: Principles for applying the users' guides to patient care. *Journal of the American Medical Association* 284 (10):1290-1296.
- Guyatt, G.H., Sackett, D.L., Sinclair, J.C., Hayward, R., Cook, D.J., and Cook, R.J. 1995. Users' guide to the medical literature: IX. A method for grading health care recommendations. *Journal of the American Medical Association* 274 (22):1800-1804.
- Guzzetta, C.E. 1989. Effects of relaxation and music therapy on patients in a coronary care unit with presumptive acute myocardial infarction. *Heart & Lung* 18 (6):609-16.
- Hanser, S.B., and Thompson, L.W. 1994. Effects of a music therapy strategy on depressed older adults. *Journal of Gerontology* 49 (6):265-9.
- Hardin, S.B., and Magee, R. 1993. Patient and family perceptions of restraint. *Journal of Holistic Nursing* 11 (4):383-397.
- Haynes, R.B., Mulrow, C.D., and Huth, E.J. 1990. More informative abstracts revisited. *Annals of Internal Medicine* 113 (1):69-76.
- Heiser, R.M., Chiles, K., Fudge, M., and Gray, S.E. 1997. The use of music during the immediate postoperative recovery period. *Aorn* 65 (4):777-85.
- Helmuth, A.M. 1995. Nurses' attitudes towards older persons on their use of physical restraints. *Orthopaedic Nursing* 14 (2):43-51.
- Hewison, A. 1997. Evidence-based medicine: what about evidence-based management. *Journal of Nursing Management* 5:195-198.
- Hicks, C. 1995. The shortfall in published research: a study of nurses' research and publication activities. *Journal of Advanced Nursing* 21:594-604.
- Hill, J.M., and Radimer, K.L. 1996. Health and nutrition messages in food advertisements: comparative content analysis of young and mature Australian women's magazines. *Journal of Nutrition Education* 28 (6):313-320.

- Hope, T. 1995. Evidence based medicine and ethics. *Journal of Medical Ethics* 21:259-60.
- Hopkins, A. 1995. Some reservations about clinical guidelines. *Archives of Disease in Childhood* 72:70-75.
- Horwitz, R.I., and Feinstein, A.R. 1979. Methodological standards and contradictory results in case control research. *American Journal of Medicine* 66:556-564.
- Horwitz, R.I., Viscoli, C.M., Clemens, J.D., and Sadock, R.T. 1990. Developing improved observational methods for evaluating therapeutic effectiveness. *American Journal of Medicine* 89:630-638.
- Hunter, D.J. 1996. Evidence based medicine and rational rationing. *Journal of Clinical Effectiveness* 1 (4):134-136.
- Hunter, S. 1999. Evidence-based nursing practice: a cautionary note. *Australian Nursing Journal* 6 (6):35.
- 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.
- Jadad, A.R., Moore, A., Carroll, D., Jenkinson, C., Reynolds, D.J.M., Gavaghan, D.J., and McQuay, H.J. 1996. Assessing the quality of reports of randomised clinical trials: is blinding necessary. *Controlled Clinical Trials* 17:1-12.
- Jamrozik, K., Chapman, S., and Woodward, A. 1997. How the NHMRC got its fingers burnt. *Medical Journal of Australia* 167:372-374.
- Janelli, L.M., and Kanski, G.W. 1997. Music intervention with physically restrained patients. *Rehabilitation Nursing* 22 (1):14-9.
- Jensen, L.A., and Allen, M.N. 1994. A synthesis of qualitative research on wellness-illness. *Qualitative Health Research* 4 (4):349-369.
- Jensen, L.A., and Allen, M.A. 1996. Meta-synthesis of qualitative findings. *Qualitative Health Research* 6 (4):553-60.



- Johnston, K., and Rohaly-Davis, J. 1996. An introduction to music therapy: helping the oncology patient in the ICU. *Critical Care Nursing Quarterly* 18 (4):54-60.
- Jones, D.R. 1992. Meta-analysis of observational epidemiological studies: a review. *Journal of the Royal Society of Medicine* 85:165-8.
- Jones, T., and Evans, D. 2000. Conducting a systematic review. *Australian Critical Care Nursing Journal* 13 (2):66-71.
- Kanski, G.W., and Janelli, L.M. 1996. Family reactions to restraint in an acute care setting. *Journal of Gerontological Nursing* 22 (6):17-22, 48-49.
- Karlsson, S., Bucht, G., Eriksson, S., and Sandman, P.O. 1996. Physical restraints in geriatric care in Sweden: prevalence and patient characteristics. *Journal of the American Geriatrics Society* 44 (11):1348-1354.
- 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.
- Keith, K.A. 1997. Infant feeding choices among first time mothers. *Scholarly Inquiry for Nursing Practice* 11 (3):199-224.
- Kelly, J.A. 1998. Scientific meeting abstracts: significance access and trends. *Bulletin of the Medical Library Association* 86 (1):68-76.
- Kerridge, I., Lowe, M., and Henry, D. 1998. Ethics and evidence based medicine. *British Medical Journal* 316:1151-1153.
- Kiat, T.K., and Huan, T.H. 1999. Attitudes toward use of restraints. *Singapore Nursing* 26 (2):25-29.
- Kitson, A. 1997. Using evidence to demonstrate the value of nursing. *Nursing Standard* 11 (28):34-39.
- Kleinert, S. 1998. Rationing of health care - how should it be done. *Lancet* 352:1244.
- Koch, M.E., Kain, Z.N., Ayoub, C., and Rosenbaum, S.H. 1998. The sedative and analgesic sparing effect of music. *Anesthesiology* 89 (2):300-6.
- Koch, S. 1993. Restraining nursing home residents. *Australian Journal of Advanced Nursing* 11 (2):9-14.

- Kopp, M.R. 1991. Music's effect on stress-related responses during surgery. *Kansas Nurse* 66 (7):4-5.
- Kowanko, I., Long, L., Hodgkinson, B., and Evans, D. 1998. *The effectiveness of strategies for preventing and treating chemotherapy and radiotherapy induced oral mucositis in cancer patients*. Adelaide: The Joanna Briggs Institute.
- Kramer, J. 1994. Reducing restraint use in a nursing home. *Clinical Nurse Specialist* 8 (3):158-162.
- Kylma, J., and Vehvilainen-Julkunen, K. 1997. Hope in nursing research: a meta-analysis of the ontological and epistemological foundation of research on hope. *Journal of Advanced Nursing* 25:364-371.
- Lamb, K.V. 1999. Help the health care team release its hold on restraint. *Nursing Management* 30 (12):19-23.
- Lang, N.M. 1999. Discipline-based approaches to evidence-based practice: a view from nursing. *Journal of Quality Improvement* 25 (10):539-544.
- Lee, D.T.F. 1999. Use of physical restraints on elderly patients: an exploratory study of the perceptions of nurses in Hong Kong. *Journal of Advanced Nursing* 29 (1):153-159.
- LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre, J., and Derderian, F. 1997. Discrepancies between meta-analyses and subsequent large randomised controlled trials. *New England Journal of Medicine* 337:536-42.
- Lemmer, B., Grellier, R., and Stevens, J. 1999. Systematic review of nonrandom and qualitative research literature: exploring and uncovering an evidence base for health visiting and decision making. *Qualitative Health Research* 9 (3):315-328.
- Lever, J.A. 1994. The use of physical restraint and their relationship to medication use in patients in four different institutional settings. *Humane Medicine* 10 (1):17-27.
- Levine, J.M., and Marchello, V. 1995. Progress toward a restraint-free environment in a large academic nursing facility. *Journal of the American Geriatrics Society* 43 (8):914-918.

- Lichter, P.R. 1993. Bibliographies and abstracts. *Ophthalmology* 100 (1):7-8.
- Light, R.J., and Smith, P.V. 1971. Accumulating evidence: procedures for resolving contradictions among different research studies. *Harvard Educational Review* 41:429-471.
- Lincoln, Y.S. 1995. Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry* 1 (3):275-89.
- Lipsey, M.W. 1994. Identifying potentially interesting variables and analysis opportunities. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Livingston, J.C. 1985. The use of music to reduce the pain of pregnancy and childbirth. In *Perspectives on Pain*, edited by Copp, L.A. Edinburgh: Churchill Livingston.
- Lofgren, R.P. 1989. Mechanical restraints on the medical wards: are protective devices safe. *American Journal of Public Health* 79 (6):735-738.
- Lohr, K.N., and Carey, T.S. 1999. Assessing "Best Evidence": issues in grading the quality of studies for systematic reviews. *Journal of Quality Improvement* 25 (9):470-9.
- Lopez-Jimenez, F., and Lamas, G.A. 1999. Evidence based medicine. *Archives of Medical Research* 30:80-1.
- Lowe, H.J., and Barnett, G.O. 1994. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Journal of the American Medical Association* 271 (14):1103-1108.
- Luker, K.A. 1997. Research and the configuration of nursing services. *Journal of Advanced Nursing* 6:259-267.
- Lynn, M.R. 1985. Reliability estimates: use and disuse. *Nursing Research* 34 (4):254-6.
- Lynn, M.R. 1989. Meta-analysis: appropriate tool for the integration of nursing research. *Nursing Research* 38 (5):302-305.

- MacGuire, J. 1990. Putting nursing research findings into practice: research utilisation as an aspect of the management of change. *Journal of Advanced Nursing* 15:614-620.
- MacNaughton, J. 1995. Anecdotes and empiricism. *British Journal of General Practice* November:571-572.
- Magee, R. 1993. Institutional policy: use of physical restraints in extended care and nursing homes. *Journal of Gerontological Nursing* 19 (4):31-39.
- Malone, A.B. 1996. The effects of live music on the distress of pediatric patients receiving intravenous starts, venipunctures, injections, and heel sticks. *Journal of Music Therapy* 33 (1):19-33.
- Manten, A.A. 1973. Scientific literature reviews. *Scholarly Publishing* 5:75-89.
- Marley, L.S. 1984. The use of music with hospitalized infants and toddlers: A descriptive study. *Journal of Music Therapy* 21 (3):126-132.
- Marwick, C. 1996. Leaving concert hall for clinic: therapists now test music's charms. *Journal of the American Medical Association* 275 (4):p267.
- McDonald, S.J., Lefebvre, C., and Clarke, M.J. 1996. Identifying reports of controlled trials in the BMJ and the Lancet. *British Medical Journal* 313:1116-1117.
- McKee, M., Britton, A., Black, N., McPherson, K., Sanderson, C., and Bain, C. 1999. Interpreting the evidence: choosing between randomised and non-randomised studies. *British Medical Journal* 319:312-315.
- McLardy-Smith, P., Burge, P.D., and Watson, N.A. 1986. Ischaemic contracture of the intrinsic muscles of the hand: a hazard of physical restraint. *Journal of Hand Surgery* 11:65-67.
- McLellan, M.F., Case, L.D., and Barnett, M.C. 1992. Trust, but verify: the accuracy of references in four anesthetic journals. *Anesthesiology* 77:185-8.
- McManus, R.J., Wilson, S., Delaney, B.C., Fitzmaurice, D.A., Hyde, C.J., Tobias, R.S., Jowett, R.S., and Hobbs, F.D.R. 1998. Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *British Medical Journal* 317:1562-3.

- McNair, D., Lorr, M., and Droppleman, L. 1992. *Manual for the Profile of Mood States (POMS)*. San Diego: EdITS.
- McQuay, H., and Moore, A. 1997. Using numerical results from systematic reviews in clinical practice. *Annals of Internal Medicine* 126:712-20.
- Meade, M.O., and Richardson, W.S. 1997. Selecting and appraising studies for a systematic review. *Annals of Internal Medicine* 127:531-7.
- Meade, M.O., and Richardson, W.S. 1998. Selecting and appraising studies for a systematic review. In *Systematic reviews: synthesis of best evidence for health care decisions*, edited by Mulrow, C. and Cook, D. Philadelphia: American College of Physicians.
- 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.
- Meyer, J. 2000. Using qualitative methods in health related action research. *British Medical Journal* 320:178-181.
- Meyer, J., Spilsbury, K., and Prieto, J. 2000. Comparison of findings from a single case in relation to those from a systematic review of action research. *Nurse Researcher* 7 (2):37-59.
- Michel, D.E., and Chesky, K.S. 1995. A survey of music therapists using music for pain relief. *Arts in Psychotherapy* 22 (1):49-51.
- Miller, A.C., Hickman, L.C., and Lemasters, G.K. 1992. A distraction technique for control of burn pain. *Journal of Burn Care and Rehabilitation* 13 (5):576-580.
- Miller, J.N., Colditz, G.A., and Mosteller, F. 1989. How study design affects outcomes in comparisons of therapy: II. Surgical. *Statistics in Medicine* 8:455-466.
- Mills, J.L. 1993. Data Torturing. *The New England Journal of Medicine* 329 (16):1196-1199.

- Miluk-Kolasa, B., Matejek, M., and Stupnicki, R. 1996. The effects of music listening on changes in selected physiological parameters in adult pre-surgical patients. *Journal of Music Therapy* 33 (3):208-218.
- Minnick, A.F. 1998. Prevalence and patterns of physical restraint use in the acute care setting. *Journal of Nursing Administration* 28 (11):19-24.
- Mion, L. 1989. A further exploration of the use of physical restraints in hospitalised patients. *Journal of the American Geriatrics Society* 37:949-956.
- Moher, D., Fortin, P., Jadad, A.R., Juni, P., Klassen, T., Le Lorier, J., Liberati, A., Linde, K., and Penna, A. 1996. Completeness of reporting of trials published in languages other than English: implications for the conduct and reporting of systematic reviews. *Lancet* 347 (363-6).
- Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P., and Walsh, S. 1995. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 16:62-73.
- Molassiotis, A., and Newell, R. 1996. Nurses' awareness of restraint use with elderly people in Greece and the UK. *International Journal of Nursing Studies* 33 (2):201-211.
- Moore, Y. 1998. All you need to know about clinical effectiveness. *Nursing Times* 94 (27):26-7.
- Morgan, D.L. 1993. Qualitative content analysis: a guide to a path not often taken. *Qualitative Health Research* 3 (1):112-121.
- Mornhinweg, G.C., and Voignier, R.R. 1995. Music for sleep disturbance in the elderly. *J Holist Nurs* 13 (3):248-54.
- Muir Gray, J.A. 1997. *Evidence-Based Healthcare*. New York: Churchill Livingstone.
- Mulhall, A. 1995. Nursing research: what difference does it make. *Journal of Advanced Nursing* 21 (3):576-583.
- Mulrow, C.D. 1987. The medical review article: state of the science. *Annals of Internal Medicine* 106:485-488.
- 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. 1997. *Cochrane Collaboration Handbook, In: The Cochrane Library [database on disk and CDROM]: The Cochrane Collaboration, Oxford: Update Software.*
- Murray, G.D. 1988. The task of a statistical referee. *British Journal of Surgery* 75 (7):664-667.
- Nagy, S., and Crisp, J. 1992. Journal reading practices of RNs in NSW public hospitals. *Australian Journal of Advanced Nursing* 9 (2):29-33.
- Nandy, B.R., and Sarvela, P.D. 1997. Content analysis reexamined: a relevant research method for health education. *American Journal of Health Behaviour* 21 (3):222-234.
- Naylor, C.D. 1995. Grey zones of clinical practice: some limits to evidence based medicine. *The Lancet* 345:840-842.
- Neill, S.J. 2000. Acute childhood illness at home: the parents perspective. *Journal of Advanced Nursing* 31 (4):821-32.
- Neufeld, R.R., and Libow, L.S. 1999. Restraint reduction reduces serious injuries among nursing home residents. *Journal of the American Geriatrics Society* 47 (10):1202-1207.
- Newbern, V.B., and Lindsey, I.H. 1994. Attitudes of wives towards having their elderly husbands restrained. *Geriatric Nursing* 15:3.
- NHMRC. 1995. *Guidelines for The Development and Implementation of Clinical Guidelines*. 1st ed. Canberra: Australian Government Publishing Service.
- NHMRC. 1999. *How to review the evidence: systematic identification and review of the scientific literature*. Canberra: AusInfo.
- NHMRC. 2000. *How to use the evidence: assessment and application of scientific evidence*. Canberra: AusInfo.
- 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.
- Noblit, G.W., and Hare, R.D. 1988. *Meta-ethnography: Synthesizing Qualitative Studies*. Vol. 11. Newbury Park: Sage Publications.

- O'Callaghan, C.C. 1996. Pain, music creativity and music therapy in palliative care. *American Journal of Hospital Palliative Care* 13 (2):43-9.
- Oddby-Muhrbeck, E., and Jakobsson, J. 1993. Recall of music: a comparison between anaesthesia with propofol and isoflurane. *Acta Anaesthesiol Scand* 37 (1):33-7.
- Ogden-Burke, S., Kauffmann, E., Costello, E., Wiskin, N., and Harrison, M.B. 1998. Stressors in families with a child with a chronic condition: analysis of qualitative studies and a framework. *Canadian Journal of Nursing Research* 30 (1):71-95.
- Oliver, J. 1999. Music in theatres. *British Journal of Theatre Nurses* 9 (10):460-463.
- O'Neil, E.S., and Morrow, L.L. 2001. The symptom experience of women with chronic illness. *Journal of Advanced Nursing* 33 (2):257-268.
- Orwin, R.G. 1994. Evaluating coding decisions. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Palakanis, K.C., DeNobile, J.W., Sweeney, W.B., and Blankenship, C.L. 1994. Effect of music therapy on state anxiety in patients undergoing flexible sigmoidoscopy. *Dis Colon Rectum* 37 (5):478-81.
- Paterson, B.L., Thorne, S., and Dewis, M. 1998. Adapting to and managing diabetes. *Image* 30 (1):57-62.
- Petitti, D.B. 1994. *Meta-analysis, decision analysis and cost-effectiveness analysis*. New York: Oxford University Press.
- Pigott, T.D. 1994. Methods for handling missing data in research synthesis. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Poole, K., and Jones, A. 1996. A reexamination of the experimental design for nursing research. *Journal of Advanced Nursing* 24:103-114.
- Popay, J., Rogers, A., and Williams, G. 1998. Rationale and standards for the systematic review of qualitative literature in health services research. *Journal of Qualitative Health Research* (in press).



- Pratt, C.A., and Pratt, C.B. 1995. Comparative content analysis of food and nutrition advertising in *Ebony*, *Essence* and *Ladies Home Journal*. *Journal of Nutrition Education* 27 (1):11-17.
- Pratt, R.R. 1999. Listening to music during surgery: a program of Intermountain Health Care. *International Journal of Arts in Medicine* 6 (1):21-30.
- Price, D. 1965. Networks of scientific papers. *Science* 149:56-64.
- Prins, J.M., and Buller, H.R. 1996. Meta-analysis: the final answer or even more confusion? *Lancet* 348:199.
- Purdie, H., Hamilton, S., and Baldwin, S. 1997. Music therapy: facilitating behavioural and psychological change in people with stroke: a pilot study. *International Journal of Rehabilitation Research* 20 (3):325-7.
- Ragneskog, H., Kihlgren, M., Karlsson, I., and Norberg, A. 1996. Dinner music for demented patients: analysis of video-recorded observations. *Clinical Nursing Research* 5 (3):262-77.
- Ray, L. 1999. Evidence and outcomes: agendas, presuppositions and power. *Journal of Advanced Nursing* 30 (5):1017-1026.
- Retsas, A. 1998. Survey findings describing the use of physical restraints in nursing homes in Victoria, Australia. *International Journal of Nursing Studies* 35 (3):184-191.
- Richards, D. 1996. A centre for evidence based dentistry. *Journal of Clinical Effectiveness* 1 (2):70.
- Robbins, L.J., Boyko, E., Lane, J., Cooper, D., and Jahnigen, D.W. 1987. Binding the elderly: a prespective study of the use of mechanical restraints in an acute care hospital. *Journal of the American Geriatric Society* 35 (4):290-296.
- Roberge, R., and Beausejour, R. 1988. Use of restraints in chronic care hospitals and nursing homes. *Canadian Journal of Ageing* 7 (4):377-381.
- Robinson, B.E., Sucholeiki, R., and Schocken, D.D. 1993. Sudden death and resisted mechanical restraint: a case report. *Journal of the American Geriatric Society* 41:424-425.

- Robinson, N. 1999. The use of focus group methodology: with selected examples from sexual health research. *Journal of Advanced Nursing* 29 (4):905-913.
- Rosenberg, W., and Donald, A. 1995. Evidence based medicine: an approach to clinical problem solving. *British Medical Journal* 310:1122-1126.
- Rosenthal, M.C. 1994. The Fugitive Literature. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Rothwell, P.M., and Martyn, C.N. 2000. Reproducibility of peer review in clinical neuroscience: is agreement between reviewers any greater than would be expected by chance alone. *Brain* 123:1964-1969.
- Sabo, C.E., and Michael, S.R. 1996. The influence of personal message with music on anxiety and side effects associated with chemotherapy. *Cancer Nursing* 19 (4):283-9.
- Sackett, D.L. 1986. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 89 (2):2s - 3s.
- Sackett, D.L., Rosenberg, W.M., Muir Gray, J.A., Haynes, R.B. and Richardson, W.S. 1996. Evidence based medicine: what it is and what it isn't. *British Medical Journal* 312:71-72.
- Sackett, D.L., Richardson, W.S., Rosenberg, W. and Haynes, R.B. 1997. *Evidence Based Medicine: How To Practice and Teach EBM*. New York: Churchill Livingstone.
- Sailas, E., and Fenton, M. 2000. Seclusion and restraint for people with serious mental illness. In *The Cochrane Library*. Oxford: Cochrane Collaboration.
- Sandelowski, M. 1993. Rigor or rigor mortis: the problem of rigor in qualitative research revisited. *Advances in Nursing Science* 16 (2):1-8.
- Sandelowski, M., Docherty, S., and Emden, C. 1997. Qualitative metasynthesis: issues and techniques. *Research in Nursing and Health* 20:365-71.
- Schleenbaker, R.E. 1994. Restraint use in inpatient rehabilitation: predictors, and implications. *Archives of Physical Medicine and Rehabilitation* 75 (4):427-430.

- Schulz, K.F., Chalmers, I., Grimes, D.A., and Altman, D.G. 1994. Assessing the quality of randomization from the reports of controlled trials published in obstetrics and gynecology journals. *Journal of the American Medical Association* 272 (2):125-8.
- Schulz, K.F., Chalmers, I., Hayes, R.J., and Altman, D.A. 1995. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association* 273 (5):408-12.
- Schuster, B.L. 1985. The effect of music listening on blood pressure fluctuations in adult hemodialysis patients. *Journal of Music Therapy* 22 (3):146-153.
- Schwandt, T.A. 1996. Farewell to criteriology. *Qualitative Inquiry* 2 (1):58-72.
- Scott, T.F., and Gross, J.A. 1989. Brachial plexus injury due to vest restraints. *New England Journal of Medicine* 320:598.
- Shapiro, S. 1994. Meta-analysis / shmeta-analysis. *American Journal of Epidemiology* 140 (9):771-8.
- Sherwood, G. 1999. Meta-synthesis: merging qualitative studies to develop nursing knowledge. *International Journal of Human Caring* 3 (1):37-42.
- Sherwood, G.D. 1997. Meta-synthesis of qualitative analyses of caring: defining a therapeutic model of nursing. *Advanced Practice Nursing Quarterly* 3 (1):32-42.
- Sindhu, F., Carpenter, L., and Seers, K. 1997. Development of a tool to rate the quality assessment of randomized controlled trials using a Delphi technique. *Journal of Advanced Nursing* 25:1262-8.
- Smeeth, L., Haines, A., and Ebrahim, S. 1999. Numbers needed to treat derived from meta-analyses: sometimes informative usually misleading. *British Medical Journal* 318:1548-51.
- Smith, B.H. 1996. Evidence based medicine: rich sources of evidence are ignored. *British Medical Journal* 313:169.
- Smith, M.L. 1987. Publishing qualitative research. *American Educational Research Journal* 24:173-82.

- Smith, R. 1991. Where is the wisdom: the poverty of medical evidence. *British Medical Journal* 303:798-9.
- Spodich, D.H. 1996. Evidence Based Medicine: terminologic lapse or terminologic arrogance. *The American Journal of Cardiology* 78:608-609.
- Standley, J.M. 1992. Clinical applications of music and chemotherapy: The effects on nausea and emesis. *Music Therapy Perspectives* 10 (1):27-35.
- Standley, J.M., and Hanser, S.B. 1995. Music therapy research and applications in pediatric oncology treatment. *Journal of Pediatric Oncology Nursing* 12 (1):3-8.
- 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.
- Stevens, K. 1990. Patients' perceptions of music during surgery. *Journal of Advanced Nursing* 15 (9):1045-51.
- Stock, W.A. 1994. Systematic coding for research synthesis. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., and Thacker, S.B. 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Journal of the American Medical Association* 283 (15):2008-12.
- Strumpf, N.E., and Evans, L.K. 1988. Physical restraint of the hospitalised elderly: perceptions of patients and nurses. *Nursing Research* 37 (3):132-137.
- Suikkala, A., and Leino-Kilpi, H. 2001. Nursing student-patient relationships: a review of the literature from 1984 to 1998. *Journal of Advanced Nursing* 33 (1):42-50.
- Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A., and Song, F. 1998. Systematic reviews of trials and other studies: Health Technology Assessment.

- Sutton, A.J., Duval, S.J., Tweedle, R.L., Abrams, K.R., and Jones, D.R. 2000. Empirical assessment of effect of publication bias on meta-analysis. *British Medical Journal* 320:1574-7.
- Swanson, D.R. 1990. Medical literature as a potential source of new knowledge. *Bulletin of the Medical Library Association* 78 (1):29-37.
- Szeto, C.K., and Yung, P.M.B. 1999. Introducing a music program to reduce preoperative anxiety. *British Journal of Theatre Nursing* 9 (10):455-9.
- Taddio, A., Pain, T., Fassos, F.F., and al., e. 1994. Quality of non-structured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical association Journal and the Journal of the American Medical Association. *Canadian Medical Association Journal* 150 (10):1611-1615.
- Tang, C.S., Ko, C.J., Ng, S.M., Chen, S.C., Cheng, K.I., Yu, K.L., and Tseng, C.K. 1993. "Walkman music" during epidural anesthesia. *Kao Hsiung I Hsueh Ko Hsueh Tsa Chih* 9 (8):468-75.
- Taylor, L.K., Kuttler, K.L., Parks, T.A., and Milton, D. 1998. The effect of music in the postanesthesia care unit on pain levels in women who have had abdominal hysterectomies. *Journal of Perianesthetic Nursing* 13 (2):88-94.
- Taylor, M.K. 1998. The practical effects of errors in reference lists in nursing research journals. *Nursing Research* 47 (5):300-3.
- Taylor, M.W. 1991. Effects of initial stress level, social support, and participation in an exercise or music condition on the posttreatment stress, depression, and anxiety of nurses. *St. John's University, ED.D. Thesis*.
- The Cochrane Collaboration. The Cochrane Library (database on disk and CDROM). : The Cochrane Collaboration, Oxford: Update Software; 1996.
- Thomas, A., Redfern, L., and John, R. 1995. Perceptions of acute care nurses in the use of restraints. *Journal of Gerontological Nursing* 21 (6):32-38.
- Thompson, D. 1998. Why evidence based nursing. *Nursing Standard* 13 (9):58-59.
- Thompson, L. 2000. *Suctioning adults with an artificial airway*. Adelaide: The Joanna Briggs Institute.

- Thornby, M.A., Haas, F., and Axen, K. 1995. Effect of distractive auditory stimuli on exercise tolerance in patients with COPD. *Chest: The Cardiopulmonary Journal* 107 (5):1213-7.
- Thorne, S., and Paterson, B. 1998. Shifting images of chronic illness. *Image: Journal of Nursing Scholarship* 30 (2):173-8.
- Tinetti, M.E., Liu, W.L., Marottoli, R.A., and Ginter, S.F. 1991. Mechanical restraint use among residents of skilled nursing facilities: prevalence, patterns and predictors. *Journal of the American Medical Association* 265 (4):468-471.
- Upton, D.J. 1999. How can we achieve evidence based practice if we have a theory practice gap in nursing today? *Journal of Advanced Nursing* 29 (3):549-55.
- Van der Zalm, J.E. 2000. Hermeneutic-phenomenology: providing living knowledge for nursing practice. *Journal of Advanced Nursing* 31 (1):211-218.
- Van Manen, M. 1990. *Researching lived experience: Human science for an action sensitive pedagogy*. London: State University of New York.
- Walker, K. 1997. Show me the evidence and I'll give you the truth: a critical philosophical snapshot. *Nursing Matters: Journal of the Nursing Council of South Eastern Sydney Area Health Service* 14:3-4.
- Walther-Larsen, S., Deimar, V., and Valentin, N. 1988. Music during regional anesthesia. A reduced need of sedatives. *Regional Anesthesia* 13:69-71.
- Watson, M.M., and Perrin, R. 1994. A comparison of CINAHL and MEDLINE CD-ROM in four allied health areas. *Bulliten of the Medical Library Association* 82 (2):214-216.
- Weber, S., Nuessler, V., and Wilmanns, W. 1996. A pilot study on the influence of receptive music listening on cancer patients during chemotherapy. *International Journal of Arts in Medicine* 5 (2):27-35.
- Whipple, B., and Glynn, N.J. 1992. Quantification of the effects of listening to music as a noninvasive method of pain control. *Scholarly Inquiry for Nursing Practice* 6 (1):43-58.

- White, H.D. 1994. Scientific Communication and Literature Retrieval. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- White, J.M. 1992. Music therapy: an intervention to reduce anxiety in the myocardial infarction patient. *Clinical Nurse Specialist* 6 (2):58-63.
- White, J.M. 1999a. Effects of relaxing music on cardiac autonomic balance and anxiety after acute myocardial infarction. *American Journal of Critical Care* 8 (4):220-30.
- White, J.M. 2000. State of the science of music interventions: critical care and perioperative practice. *Critical Care Nursing Clinics of North America* 12 (2):219-225.
- White, S. 1999b. A study into the use of physical restraint within Royal Adelaide Hospital. (*unpublished*).
- Whitehead, C., Finucane, P., and Henschke, P. 1997. Use of restraints in four Australian teaching hospitals. *Journal of Quality in Clinical Practice* 17:131-136.
- Wilczynski, N.L., Walker, C.J., and McKibbin, K.A. 1995. Preliminary assessment of the effect of more informative (structured) abstracts on citation retrieval from MEDLINE. *Medinfo* 8 (Pt. 2):1457-1461.
- Williams, H.C., and Seed, P. 1993. Inadequate size of negative clinical trials in dermatology. *British Journal of Dermatology* 128:317-326.
- Wilson, M.C., Hayward, R.S.A., Tunis, S.R., Bass, E.B., and Guyatt, G. 1995. Users' guide to the medical literature: VIII. How to use clinical practice guidelines; B. What are the recommendations and will they help you in caring for your patients. *Journal of the American Medical Association* 274 (20):1630-1632.
- Winter, M.J., Paskin, S., and Baker, T. 1994. Music reduces stress and anxiety of patients in the surgical holding area. *J Post Anesth Nurs* 9 (6):340-3.
- Woodgate, R. 2000. Part II: A critical review of qualitative research related to children's experiences with cancer. *Journal of Pediatric Oncology Nursing* 17 (4):207-228.

- Woodward, V., and Webb, C. 2001. Women's anxieties surrounding breast disorders: a systematic review of the literature. *Journal of Advanced Nursing* 33 (1):29-41.
- Woodworth, G. 1994. Managing meta-analytic databases. In *The Handbook of Research Synthesis*, edited by Cooper, H. and Hedges, L.V. New York: Russell Sage Foundation.
- Woolf, S.H., Battista, R.N., Anderson, G.M., Logan, A.G., and Wang, E. 1990. Assessing the clinical effectiveness of preventative maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. *Journal of Clinical Epidemiology* 43 (9):891-905.
- Wuest, J. 1998. Setting boundaries: a strategy for precarious ordering of women's caring demands. *Research in Nursing and Health* 21 (1):39-49.
- Zenios, S.A., Wein, L.M., and Chertow, G.M. 1999. Evidence based organ allocation. *American Journal of Medicine* 107:52-61.