Evidence Based Medicine – A Critical Analysis

By:

Joseph W Vere

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Academic Unit of Dental Public Health
School of Clinical Dentistry

Submission Date: June 2018

Abstract

Evidence based medicine uses hierarchies of evidence to justify knowledge claims that are made. These knowledge claims are important because they dictate which treatment interventions are provided and funded, medicolegal standards and the medical research agenda. It is therefore interesting that different hierarchies can be used as this suggests that knowledge claims can be justified in different ways. This thesis presents a critical analysis of evidence based medicine, using the method of analytical philosophy, to improve understanding of the concept. The thesis is divided into two sections. In the first section a systematic review and thematic analysis of hierarchies of evidence is presented; the arguments used to rank systematic reviews, randomised controlled trials and expert opinion within hierarchies are analysed, and the properties used to rank different study designs are analysed. Five factors, independent of study design, that have influenced the development of hierarchies are then presented and it is argued that a lack of theoretical support for hierarchies has led to their proliferation. In the second section the claims that evidence based medicine is rational, science and a new Kuhnian paradigm are analysed. It is argued that evidence based medicine can be substantively rational but this means that knowledge claims can be both rational and inconsistent dependent upon any value commitments that are held. It is then argued that evidence based medicine cannot be science because it does not use scientific method and it cannot be a new Kuhnian paradigm because it is not science, it was not preceded by a revolutionary crisis and it is not incommensurable with previous versions of

1

medicine. The analysis presented strips evidence based medicine of power and has important implications for the status of knowledge claims that are made.

Table of Contents

List of Tables

Chapter 1: Introduction

1.1 Evidence Based Medicine:

Evidence based medicine emerged from the discipline of clinical epidemiology, and the work of British epidemiologists such as Archie Cochrane, in the early 1990's (Daly 2005). The term first appeared in the medical literature in 1991 (Guyatt 1991). However, it was not until the succeeding year that the concept of evidence based medicine was first fully articulated by the Evidence Based Medicine Working Group (1992). The Evidence Based Medicine Working Group was a collection of medical professionals and clinical epidemiologists centred on McMaster University in Canada (Daly 2005). The Working Group claimed that evidence based medicine was a new paradigm that would revolutionise the practice of medicine.

Evidence based medicine has become increasingly important since the concept was first articulated. Every year vast numbers of evidence-based guidelines are produced by institutions such as the National Institute for Health and Care Excellence (NICE), in the United Kingdom, and the Agency for Healthcare Research and Quality (AHRQ), in the United States. These evidence-based guidelines dictate which treatment interventions clinicians can prescribe for patients, which treatment interventions are funded and the

legal standard of care (Batchelor 2000, Timmermans and Berg 2003, Tanenbaum 2014). Evidence based medicine also influences the medical research agenda and the pharmaceutical industry (Hyde 2004, Mebius 2014, Bingeman 2016). Research that does not conform to the principles of evidence based medicine is unlikely to obtain funding, difficult to publish and can be detrimental to career progression (Hammersley 2013). New drugs that are not supported by the right type of 'evidence' are unlikely to obtain marketing approval. Evidence based medicine is clearly an important concept.

The concept of evidence based medicine was originally developed for use within the specialty of internal medicine (Evidence Based Medicine Working Group 1992). However, as time has progressed the underlying principles have been adopted by all of medicine and other medical professions such as nursing, midwifery and dentistry (Swinkels et al 2002, Needleman et al 2005, McNeill 2006, Satterfield et al 2009). Although it is possible to make a distinction between evidence based medicine, evidence based nursing, evidence based midwifery and evidence based dentistry the term 'evidence based medicine' is usually used as an overarching umbrella term to encompass all medical professions. The term 'evidence based medicine' is used in this overarching sense in this thesis. Recognition of the importance of evidence based medicine has also spread beyond the medical world and the concept made a list of great ideas published in the New York Times in 2001 (Hilt 2001). This provides further evidence of the importance of evidence based medicine.

Evidence based medicine is presented as a tool to improve patient care and prevent unnecessary patient harm (Dickersin et al 2007, La Caze 2011)[1]. Many medical interventions are associated with significant adverse effects and patients may be harmed if they receive the wrong intervention or do not receive the right intervention. Proponents of evidence based medicine argue that a number of emotive medical tragedies could have been avoided if the principles of evidence based medicine had been followed. Examples of these tragedies include the increased incidence of birth defects in pregnant women prescribed Thalidomide to alleviate morning sickness and the increased risk of breast cancer associated with the use of hormone replacement therapy (Evans et al 2010). Evidence based medicine is attractive to the medical profession and patients because it purports to improve patient care and prevent unnecessary patient harm.

Evidence based medicine is also presented as a tool that can be used to identify variation in medical practice (Dickersin et al 2007, La Caze 2011). Variations in medical practice are considered undesirable because they imply that some patients receive inferior care. Variations in medical practice may be indicative of health inequalities, misuse of healthcare resources or patient harm (Cochrane 1972, Wennberg and Gittelsohn 1973, Rosenberg and Donald 1995). In recent years it has been increasingly recognised that the

---

[1] These are not the only benefits have been attributed to evidence based medicine in the literature. Other benefits include improvements in study conduct and reporting, computer literacy, critical appraisal skills and improved accessibility to research evidence. However, improved patient care and identification of variation in medical practice are seen as the primary benefits of evidence based medicine (Rosenberg and Donald 1995, Bluhm 2005, La Caze 2011).

medical profession may harm patients. In the UK this recognition has been driven by a number of high profile scandals such as the failings at Mid-Staffordshire NHS Trust, the murderous activities of Harold Shipman and the Bristol Heart Hospital Inquiry. These scandals have led to greater oversight of medicine to assure the quality of care and reduce the risk of patient harm (Storey et al 2011). Evidence based medicine is important in the governance of medicine because it provides standards that medical care can be judged against. This facilitates the identification of variation in medical practice. Evidence based medicine is attractive to the medical profession, patients, regulators and purchasers of healthcare because it may provide a tool that can be used to identify such undesirable variations in practice.

Although widely accepted by the medical profession as a way to improve patient care and identify variation in medical practice, evidence based medicine has not been universally accepted. Critics have claimed that there is no empirical evidence to support the use of evidence based medicine. This claim has not been disputed (Haynes 2002, Djulbegovic et al 2009) although proponents of evidence based medicine frequently claim that it would be impossible or unethical to undertake a study demonstrating the superiority of evidence based medicine. Every-Palmer and Howick (2014) recently argued that there was no empirical evidence to support evidence based medicine because it had been poorly implemented following manipulation by the pharmaceutical industry.

Evidence based medicine has also been criticised because it has weakened the professional jurisdiction of the medical profession. Timmermans and Kolver (2004) argued that evidence based medicine made medical knowledge more transparent so that the knowledge was no longer esoteric and exclusive to the medical profession. It has been suggested that this deprofessionalisation of medicine has facilitated increased oversight, control and regulation of the medical profession by employers, regulators and purchasers of healthcare (Kelleher et al 2006)[2]. It has also been claimed that deprofessionalisation has allowed greater control of healthcare costs as evidence of ineffectiveness, or the absence of evidence of effectiveness, can be used to deny treatments and disinvest in health care services (Greenhalgh and Russell 2009, Tanenbaum 2014, Storey et al 2011). If evidence based medicine has weakened the professional jurisdiction of the medical profession it is likely to be less attractive to the medical profession, but more attractive to patients, regulators and purchasers of healthcare.

Interestingly, evidence based medicine has also been presented as a tool that reinforces and expands the professional jurisdiction of the medical profession. Mykhalovskiy and Weir (2004) argued that evidence based medicine buttressed the medical profession against patients, regulators and purchasers of healthcare because it emphasised the scientific character of medical practice[3]. They argued that the development of evidence based

---

[2] Increased managerial control over medical professionals has been referred to as the proletarianisation of medicine (Kelleher et al 2006).
[3] Mykhalovskiy and Weir (2004) also suggested that evidence based medicine had divided the medical profession with an academic elite controlling rank and file medical professionals.

medicine had allowed the medical profession to control medicine and maintain professional dominance. Evidence based medicine has also been implicated in the medicalisation of normal life processes (Ballard and Elston 2005, Maturo 2012). Once a life process has been medicalised it becomes amenable to medical treatment (Moynihan and Smith 2002). This medicalisation has expanded the professional jurisdiction of the medical profession and led to further increases in the cost of healthcare. On this interpretation evidence based medicine is likely to less attractive to regulators and purchasers of healthcare because it reinforces the professional dominance of the medical profession.

Evidence based medicine has also been criticised because it requires medical professionals to learn new skills, there is an unmanageable volume of evidence, studies report aggregate values that cannot be applied to individual patients and the results of studies have poor external validity (Tonelli 1998, Straus and McAlister 2000, Tanenbaum 2014). These are important practical considerations that relate to the interpretation and application of evidence.

Evidence based medicine is an important concept that dictates which treatment interventions are provided and funded, legal standards of care and the medical research agenda. Evidence based medicine has been widely accepted by the medical profession, patients and healthcare regulators as a tool to improve patient care and identify undesirable variation in medical

practice. However, it has also been criticised because it lacks empirical support. There is also controversy as to whether evidence based medicine strengthens or weakens the professional jurisdiction of the medical profession. This has important implications for the control of medical knowledge and medical resources. There is therefore some justification for examining the concept and practice of evidence based medicine in greater depth.

1.2. Defining Evidence Based Medicine:

In order to investigate evidence based medicine we must first develop a working definition. This is surprisingly difficult as evidence based medicine has been defined in a number of different ways in the literature (Sackett et al 1996, Bruce 1999, Straus and McAlister 2000, Straus et al 2007, Straus et al 2011) and there does not appear to be a single accepted definition (Satterfield et al 2009). Opponents of evidence based medicine are often criticised for attacking a straw man because they fail to capture the true essence of the concept (Tonelli 2009, Turner 2011). It is therefore important to consider several definitions of evidence based medicine, both early and more contemporary, to draw out the essential features of this concept, before it is analysed in greater detail.

An early, often quoted, definition of evidence based medicine was provided by Sackett et al (1996):

> '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 experience with the best available external clinical evidence from systematic research'. (Sackett et al, 1996, page 71)

This definition emphasises the importance of both clinical experience and evidence from systematic research although it does not clarify which types of research evidence should guide decision-making. This definition appears to restrict evidence based medicine to the care of individual patients and suggests a paternalistic approach where clinical experience and research evidence are the sole determinants of the care received by patients.

Straus and McAlister (2000) provided a later definition. They defined evidence based medicine as:

> 'The process of systematically finding, appraising and using contemporaneous research findings as the basis for clinical decisions' (Straus and McAlister, 2000)

This definition also emphasises the importance of research evidence without clarifying the type of research evidence that is preferred. This definition does not restrict evidence based medicine to medical decision-making and

presents a much wider concept that also involves accessing and critically appraising the medical literature. This is more consistent with the original articulation of the concept of evidence based medicine (Evidence Based Medicine Working Group 1992) and articulations in modern evidence based medicine textbooks (Guyatt et al 2008, Straus et al 2011). This definition does not limit evidence based medicine to the care of individual patients and allows the concept to be used at a population level in the development of clinical guidelines and healthcare policy.

A more contemporary definition of evidence based medicine is provided in *'Evidence-Based Medicine How to Practice and Teach it'*:

> 'Evidence based medicine requires the integration of the best research evidence with our clinical expertise and our patient's unique values and circumstances' (Straus et al, 2011, page 1)

This definition again emphasises the importance of research evidence but does not clarify which types of research evidence are preferred. It also emphasises the importance of clinical expertise and patient values rebutting the criticisms of paternalism that can be made against the definition provided by Sackett et al (1996).

It is clear from the definitions of evidence based medicine presented above that the use of research evidence is central to the concept. However, none of the definitions specify which types of research evidence should be used to

guide medical decision-making. This creates a problem if we want to critically analyse the concept of evidence based medicine. Fortunately this problem can be resolved if we consider the literature relating to evidence based medicine and the development of evidence-based guidelines. This reveals that evidence based medicine uses hierarchies of evidence to determine the type of research evidence that should be used to guide medical decision-making. These hierarchies of evidence generally value the results of randomised controlled trials and systematic reviews and do not value expert opinion and pathophysiological rationale (Cochrane 1972, Spitzer et al 1979, Trout 1981, Evidence Based Medicine Working Group 1992, Guyatt et al 2008, Straus et al 2011).

Proponents of evidence based medicine are clear that research evidence alone, as dictated in the hierarchy of evidence, is never sufficient to make a clinical decision (Sackett et al 1996, Guyatt et al 2008, Satterfield et al 2009, Straus et al 2011). Any interpretation of evidence based medicine, where the hierarchy of evidence is used as the sole determinant of clinical decision making, is a misrepresentation of the concept. Attacks based on this assumption might be considered as attacks on a straw man version of evidence based medicine (Tonelli 2009). This is clear from the definitions of evidence based medicine provided by both Sackett et al (1996) and Straus et al (2011). However, this does not mean that the hierarchy of evidence is not a fundamental component of evidence based medicine.

The importance of hierarchies of evidence can be illustrated by the large numbers of evidence-based clinical guidelines that are produced every year by institutions such as NICE in the UK and the AHRQ in the USA. These guidelines are subsequently used to determine the treatment interventions that are funded, and which clinicians can prescribe for patients. They also often determine legal standards of care. Consistent with the definitions of evidence based medicine presented above, evidence derived using hierarchies of evidence is not the sole determinant of the conclusion of the guideline development process, but it is a fundamental component of the process. The importance of hierarchies of evidence to evidence based medicine is further supported by the absence of evidence-based guidelines that do not use hierarchies.

Another important feature of evidence based medicine was that it was presented as a revolutionary change in the way that medicine was practised when the concept was first articulated (Evidence Based Medicine Working Group 1992). However, medical decision-making was surely influenced by research evidence, clinical experience and patient values prior to the development of evidence based medicine. The publication of large numbers of patient opinion surveys between 1950 and 1980 (French 1981) suggests that patient values clearly influenced medical decision-making prior to the advent of evidence based medicine. It could be argued that the only thing significantly different about evidence based medicine was the use of hierarchies to determine the research evidence that should be used. On this interpretation the hierarchies of evidence are not just a fundamental feature

of evidence based medicine they are the defining feature[4]. In essence, the hierarchies of evidence put the 'evidence' in evidence based medicine.

In this section different definitions of evidence based medicine have been considered. These definitions all highlight the importance of research evidence but do not clarify the preferred types or research evidence. Deeper investigation has revealed that evidence based medicine uses hierarchies of evidence to determine the types of research evidence that should inform medical decision-making. Hierarchies of evidence are not the sole determinant of decision-making within evidence based medicine but they are an important component and may be the defining component. The hierarchies of evidence should therefore by investigated in greater detail if we are to critically analyse evidence based medicine.

1.3. Hierarchies of Evidence

A hierarchy of evidence can be defined as any system within medicine that ranks the importance of evidence primarily based upon the study design that is used to produce that evidence (Upshur 2009). It is generally acknowledged that the first hierarchy was published by the Canadian Task Force in 1979 (Spitzer et al 1979). However, the idea that different study designs could be ranked hierarchically appears to originate in the work of Cochrane (1972).

---

[4] This is discussed in detail in Chapter 9 when I consider whether evidence based medicine is a new Kuhnian paradigm,

Early hierarchies usually rank randomised controlled trials as the highest level of evidence, observational studies as an intermediate level of evidence and expert opinion as the lowest level of evidence. Later hierarchies usually rank systematic reviews and meta-analyses above randomised controlled trials, observational studies and expert opinion.

Within evidence based medicine the term 'hierarchy of evidence' is usually used in a singular sense. This disguises the fact that there are actually many different hierarchies. A systematic review published by West et al (2002) identified 40 different systems for grading the strength of a body of evidence with many of these systems using different hierarchies. Since this time a significant number of new hierarchies have been published. The existence of many different hierarchies has been acknowledged by a number of different authors and this is not a new finding (West et al 2002, Upshur 2003, Atkins et al 2004, Rawlins 2008, Tonelli 2009, Gugiu and Gugiu 2010, Worrall 2010, Howick 2011, Turner 2011, Blunt 2015).

Most early hierarchies of evidence rank randomised controlled trials as the highest level of evidence (West et al 2002) However, there has been a significant discussion in the medical literature as to whether this status is deserved (Hyde 2004, Cartwright 2007, Thompson 2010, Worrall 2010, La Caze 2011). Randomised controlled trials are generally considered to provide the highest level of evidence because they control for known and unknown confounding factors, facilitate Fisherian significance testing, minimise

selection bias, permit blinding and provide an accurate estimate of treatment effect (Worrall 2007, Rawlins 2008, Worrall 2010). Many of these arguments originate in the work of Cochrane (1972).

Although randomised controlled trials often provide the highest level of evidence a number of theoretical and practical problems are recognised with this study design. From a theoretical perspective, randomised controlled trial have been criticised because they do not explain why a treatment intervention is effective (Thompson 2010, Chin-Yee 2014). They can produce inconsistent results, even when well conducted, and are considered unnecessary when the effects of a treatment intervention are obvious (Hyde 2004, Worrall 2007, Worrall 2010). From a practical perspective, randomised controlled trials have been criticised because they are expensive, inappropriate for rare conditions, inappropriate for the assessment of harm, fail to differentiate between clinical and statistical significance and have poor external validity (Hyde 2004, Rawlins 2008, Worrall 2010)[5]. These practical and theoretical problems have led some authors to question whether randomised controlled trials should provide the highest level of evidence in hierarchies (Hyde 2004, Cartwright 2007).

Later hierarchies of evidence often rank systematic reviews or meta-analyses as the highest level of evidence. Systematic reviews and meta-analyses are

---

[5] It is not my intention to provide an exhaustive list of all the practical and theoretical problems associated with randomised controlled trials at this stage. I simply aim to demonstrate that the ranking of randomised controlled trials as the highest level of evidence in any hierarchy of evidence is potentially problematic.

considered to provide the highest level of evidence because they produce results that are closer to the truth and allow assessment of the consistency of results (Cook et al 1992, Guyatt et al 1995). Systematic reviews and meta-analyses are usually ranked above randomised controlled trials but it is unclear how the arguments that are used to assert the superiority of these different study designs can be related. The ranking of systematic reviews and meta-analyses as the highest level of evidence has received less attention in the medical literature although it is notable that a number of hierarchies do not include these study designs.

Most hierarchies of evidence rank expert opinion as the lowest level of evidence. Expert opinion is considered to provide the lowest level of evidence because it does not involve measurement, does not consider large numbers of patients, is susceptible to bias and may be false. Many of the arguments that are used to assert the inferiority of expert opinion can be identified in the work of Cochrane (1972). The ranking of expert opinion as the lowest level of evidence has also been questioned in the medical literature (Thompson 2010, Hofmeijer 2014).

Some critics of evidence based medicine have questioned the utility of the hierarchies of evidence themselves (Glasziou et al 2004, Hofmeijer 2014). These criticisms move beyond arguments asserting the superiority or inferiority of different study designs and argue that hierarchies should be abolished. Sehon and Stanley (2003) argued that different forms of evidence

were complementary and could not be ranked hierarchically. Hyde (2004) and Cartwright (2007) argued that the most appropriate method should be determined by the question and the context not the study design.

Although the presence of different hierarchies of evidence has been recognised the reasons behind this have not been investigated in any great depth. West et al (2002) identified 40 different systems for grading the strength of a body of evidence but provided no explanation for this finding. Atkins et al (2004) proposed a new hierarchy of evidence but did not seek to explain variation amongst existing hierarchies. Gugiu and Gugiu (2010) criticised five different hierarchies because they lacked theoretical and empirical support, neglected methodological quality and defined terms imprecisely. However, although these criticisms could potentially explain some of the variation that was seen amongst hierarchies, this was not explored.

The existence of multiple hierarchies of evidence is confusing. Evidence based medicine claims to improve patient care and allow identification of variation in medical practice (Dickersin 2007). Yet the existence of different hierarchies suggests that decision-making may be justified in a variety of different ways. There is a risk that inconsistent conclusions may be considered evidence-based depending upon the hierarchy that is used to support decision-making. If this is the case, the hierarchy that is used may determine which treatment interventions can be offered to patients,

healthcare funding and the legal standard of care. This would have significant implications for the knowledge claims made by evidence based medicine. It is therefore important that the hierarchies of evidence are systematically investigated to understand why there are so many and to understand the impact this has had on evidence based medicine.

The existence of multiple hierarchies of evidence is not just a theoretical problem. It has important implications for medical professionals and patients. This is illustrated by the recent confusion surrounding the necessity for antibiotic prophylaxis prior to dental treatment in cardiac patients where different groups reviewed the same evidence base and reached contradictory conclusions[6]. The European Society of Cardiology concluded that antibiotic prophylaxis should be given to high risk cardiac patients (Habib et al 2015) whereas NICE concluded that antibiotic prophylaxis was unnecessary (NICE 2015).

In order to systematically investigate the hierarchies of evidence it is necessary to identify and critically analyse the arguments that have been used to support the different hierarchies. The arguments that have been used to rank randomised controlled trials, systematic reviews, meta-analyses and expert opinion within hierarchies of evidence are controversial and this may explain why there are so many different hierarchies. A critical analysis would make a significant contribution to our knowledge of evidence based medicine

---

[6] The reader is referred to the paper by Thornhill et al (2016) for a review of the necessity for antibiotic prophylaxis prior to dental procedures in cardiac patients.

as it would build upon previously published research to clarify areas of disagreement and improve understanding of this important concept. An analysis may also explain the variation that exists amongst different hierarchies.

1.4. The Implications of Different Hierarchies of Evidence:

The existence of multiple hierarchies of evidence has important implications for evidence based medicine because it suggests that the decision-making process may not be rational. One way to consider a decision-making process to be rational is if the conclusions that are derived from this process are done so in a consistent way (Newton-Smith 1981). However, if different hierarchies of evidence process the same information in different ways, inconsistent conclusions may be reached. This could mean that patients receive inconsistent clinical care, medicolegal standards are inconsistent, healthcare resources are allocated inconsistently and research priorities are judged inconsistently. It is therefore important to investigate whether the decision-making method that is used by evidence based medicine is rational. The rationality of the decision-making process used by evidence based medicine has not previously been explicitly considered in any real depth in the literature.

The existence of multiple hierarchies of evidence also has significant implications for the status of evidence based medicine as science. Evidence based medicine was originally going to be called 'scientific medicine' (Howick 2011) and it has been claimed to be science on a number of different occasions (Cochrane 1972, Spitzer 1979, Evidence Based Medicine Working Group 1992, Guyatt et al 2000, Djulbegovic et al 2009). It has also been claimed that the characterisation of evidence based medicine as science has allowed the medical profession to protect its professional jurisdiction (Mykhalovskiy and Weir 2004).

If evidence based medicine is science, knowledge claims should be derived using scientific method. Within the philosophy of science there are four main theories about the nature of science: inductivism (Ayer 1946), falsification (Popper 1963), Kuhnian paradigms (Kuhn 1996) and scientific research programmes (Lakatos 1970). None of these theories advocate a hierarchy of evidence or appear to value evidence derived from randomised controlled trials, systematic reviews or meta-analyses. Each of these theories does value explanatory theory and, although this may underpin the hypotheses that are tested in empirical studies, this form of evidence is not generally valued by the hierarchies of evidence (Thompson 2010). This suggests that there may be unanswered questions concerning the status of evidence based medicine as science.

The status of evidence based medicine as science and the rationality of the decision-making used by evidence based medicine are actually interlinked. This is because although different theories about the nature of science are recognised, they all derive conclusion using a rational process (the scientific method), (Ladyman 2002). Therefore, if evidence based medicine does not derive conclusions using a rational process it cannot be science and does not deserve the status of science (Sorell 1991). If evidence based medicine does derive conclusions using a rational process it does not necessarily follow that evidence based medicine is science. Rationality is not a necessary and sufficient condition for a discipline to be science and many disciplines are both rational and non-scientific.

If evidence based medicine is not science it may be more appropriately categorised as pseudoscience. Pseudoscience can be defined as a collection of beliefs or practices mistakenly regarded as being based on the scientific method (Oxford English Dictionary 2006, Pigliucci 2015). Categorisation of any discipline as pseudoscience is considered derogatory and has negative implications for the knowledge claims made by that discipline. Categorisation of evidence based medicine as pseudoscience would have significant implications because it would mean that the treatment interventions that were funded and provided to patients, medicolegal standards and the research agenda were determined by pseudoscience.

There has been a limited discussion in the medical literature about whether the method employed by evidence based medicine to derive conclusions can be best explained using inductivism or falsificationism (Senn 1991, Shahar 1997, Djulbegovic et al 2009, Silva and Wyer 2009, Goldenberg 2010, Sestini 2010, Thompson 2010, Kerry et al 2012)[7]. However, this discussion is predicated on the assumption that evidence based medicine is science. Therefore, although it is acknowledged that there are problems explaining the method used by evidence based medicine using both inductivism and falsificationism, this does not lead to consideration of whether evidence based medicine is science or pseudoscience[8]. There has been no discussion in the medical literature about whether the method used by evidence based medicine can be related to the theory of scientific research programmes.

The claim that evidence based medicine is a new Kuhnian paradigm has been discussed in greater depth in the literature (Couto 1998, Shahar 1998, Tonelli 1998, Greaves 2002, Sehon and Stanley 2003, Daly 2005, Lambert 2006, Djulbegovic et al 2009, Goldenberg 2010, Gaeta and Gentile 2016). However, this claim has proved contentious and considerable disagreement exists with a number of different arguments being used to both support and refute the claim. Some of these arguments relate to the concept of evidence

---

[7] A systematic review was undertaken to identify articles that related the method used by evidence based medicine to the theories of inductivism, falsificationism, Kuhnian paradigms and scientific research programmes. The search strategies that were used are included as Appendix 1.

[8] Charlton (2009) did categorise evidence based medicine as both pseudoscience and zombie science although no arguments were presented to support either of these claims. Schafranski (2016) also claimed that evidence based medicine was pseudoscience although no detailed analysis was presented. No other articles have been identified that explicitly categorise evidence based medicine as pseudoscience.

based medicine whereas others relate to the theory about the nature of science proposed by Kuhn (1996). It has been argued that evidence based medicine is not a new paradigm because it is not significantly different from the way that medicine was previously practised (Couto 1998, Sehon and Stanley 2003, Daly 2005, Goldenberg 2010), randomised controlled trials are poorly understood by the medical profession (Shahar 1998), the concept of evidence based medicine is poorly defined (Shahar 1998, Tonelli 1998, Sehon and Stanley 2003) and there are problems with the theory of Kuhnian paradigms (Greaves 2002, Sehon and Stanley 2003, Gaeta and Gentile 2016). Nobody has previously argued that evidence based medicine cannot be a new Kuhnian paradigm because evidence based medicine is not science[9]. Conversely, it has been claimed that evidence based medicine is a new paradigm because it is significantly different from the way that medicine was previously practised (Evidence Based Medicine Working Group 1992, Lambert 2006, Djulbegovic et al 2009).

There appear to be three main sources of confusion when the claim that evidence based medicine is a new Kuhnian paradigm has previously been considered. Firstly, there appears to be a poor understanding of the theory of Kuhnian paradigms. For example, it has been falsely claimed that evidence based medicine is not a new paradigm because the notion of incommensurability is incoherent (Sehon and Stanley 2003) and paradigms are not influenced by moral and cultural factors (Greaves 2002). Secondly,

---

[9] Gaeta and Gentile (2016) argued that evidence based medicine could not be a Kuhnian paradigm because it described a methodology not an underlying theory. However, they did not explicitly claim that evidence based medicine was not science.

some authors have struggled to define the concept of evidence based medicine. This is important because if evidence based medicine cannot be clearly defined it may be difficult to determine whether it is a new paradigm. Finally, there is disagreement as to whether evidence based medicine is incommensurable with a previous paradigm of medicine.

The claim that evidence based medicine is a new Kuhnian paradigm is also significant because it gives the concept power. This claim is used because it enhances the status of evidence based medicine, suggesting that it is somehow superior to the way that medicine was previously practised. Yet, if evidence based medicine is not a new Kuhnian paradigm this undermines its claim to a special status within medicine. It is therefore important to critically analyse such claims in order to improve our understanding of its status and resolve disputes about the underlying claims.

Howick (2011) and Blunt (2015) have previously analysed the hierarchies of evidence in depth. They both identified a number of problems with the epistemology of evidence based medicine and recognised that the existence of different hierarchies was confusing. Howick (2011) resolved these problems by favouring a particular hierarchical interpretation whereas Blunt (2015) advocated abolition of the hierarchies of evidence unless empirical support was provided. Neither Howick (2011) or Blunt (2015) sought to explain why there were so many different hierarchies or considered whether the decision-making process used within evidence based medicine was

rational or scientific. This thesis builds upon previously published work but considers the implications of different hierarchies of evidence from an alternative perspective.

The existence of different hierarchies of evidence has significant implications for evidence based medicine because they potentially allow knowledge claims to be justified in different ways. This leads us to question whether the decision-making process that is used by evidence based medicine is rational and whether the concept is science. These are important considerations because evidence based medicine dictates the treatment interventions that patients receive, healthcare funding, medico-legal standards and the research agenda. If the decision-making process that is used by evidence based medicine is not rational or scientific, we may question whether evidence based medicine should exert such a significant influence on modern medicine. The rationality of the decision making process used by evidence based medicine has not previously been explicitly considered in any great depth and the status of evidence based medicine as science is unclear. These claims require critical analysis to improve our understanding of the concept of evidence based medicine.

1.5 Aim and Structure of the Thesis

This thesis presents a critical analysis of evidence based medicine, using the method of analytical philosophy, to improve understanding of this important concept. The thesis is divided into two parts. The hierarchies of evidence are fundamental to the justification of the knowledge claims that are made by evidence based medicine and these will be analysed in detail in the first part. A systematic review is initially presented identifying the different hierarchies that can be used to support decision-making within evidence based medicine. These hierarchies are then compared and contrasted to identify underlying themes. These underlying themes allow the construction of a typology of hierarchies of evidence in Chapter 3. The arguments that are used to support the different hierarchies are critically analysed in Chapter 4. This analysis expands upon previously published work to increase clarity concerning the underlying arguments. Chapter 5 analyses the properties that are used to order different study designs within hierarchies and Chapter 6 considers factors, unrelated to study design, that have influenced the development of hierarchies. This completes the first part of the thesis.

The second part of this thesis considers the implications that different hierarchies of evidence have for evidence based medicine. In Chapter 7 the rationality of the decision-making process used by evidence based medicine is analysed. In Chapter 8 the claim that evidence based medicine is science is analysed and in Chapter 9 the claim that evidence based medicine is a

new paradigm, from the perspective of Thomas Kuhn, is analysed. These chapters draw on previously published work where this is available. The final chapter, Chapter 10, considers the implications of these analyses for the future of evidence based medicine and makes recommendations for future research.

The aim of this thesis is to critically analyse the arguments that are used to support the claim that evidence based medicine is rational and scientific. These arguments and claims have not previously been explicitly considered in great detail, and those that have been considered remain disputed. It should be pointed out that this thesis should not be interpreted as an attempt to undermine evidence based medicine. Rather this thesis draws on the techniques of analytical philosophy to explore argument and counterargument whilst evaluating the epistemological problems associated with evidence based medicine. This may ultimately lead to a re-evaluation and strengthening of the concept.

Chapter 2: Methodological Deliberations

2.1. Introduction:

Evidence based medicine is an important concept that dictates which treatment interventions are provided and funded, medicolegal standards and the medical research agenda. There is also controversy as to whether evidence based medicine strengthens or weakens the professional jurisdiction of the medical profession. Hierarchies of evidence are an important component of the decision making process used within evidence based medicine but the existence of different hierarchies means that knowledge claims may be justified in different ways. This has potentially important implications for the status of evidence based medicine as science, the rationality of the decision making process and the control of medical knowledge and resources.

In this chapter we will consider different methods that could have been used to investigate evidence based medicine: conceptual history, discourse analysis, qualitative methods and analytical philosophy. It should already be clear to the reader that analytical philosophy is used to analyse the hierarchies but the thesis could have taken a different direction. In the first part of this chapter, possible methods are outlined and challenges associated

with applying them to evidence based medicine are considered. In the second part, analytical philosophy is described in greater detail.

2.2. Possible Methods:

2.2.1 Conceptual History:

Conceptual history is a sociohistorical method that can be used to investigate how concepts change their meaning over time. It is argued that language has diachronic and synchronic aspects so, although the meaning of concepts changes with time, each concept has a definite meaning at any particular point in time. Conceptual history identifies these meanings and reveals how they are transformed over time. The method of conceptual history involves careful historical study so that concepts are interpreted within the correct context avoiding anachronism. Concepts cannot be reduced to language because language must always be interpreted. It is also important to investigate both related and opposite concepts and understand that concepts may exist prior to their articulation in language (Koselleck 1989, Hampsher-Monk et al 1998).

A conceptual history of evidence based medicine would chart the different meanings of the concept over time. Evidence based medicine first appeared

in the medical literature in 1991 (Guyatt 1991) but different meanings of the concept could be traced back, through clinical epidemiology, to epidemiologists such as Archie Cochrane and beyond. Moving forwards, different meanings could be identified as evidence based medicine was taken up by disciplines within medicine, dentistry, nursing and midwifery. Different meanings could also be identified as new hierarchies of evidence were used within evidence based medicine.

A conceptual history of evidence based medicine has not previously been undertaken and would be interesting. It is possible that changing meanings of the concept could be related to the attempts of different groups within medicine to establish power and professional dominance over time. However, a conceptual history would require a vast amount of reading to both contextualise evidence based medicine and identify related and opposing concepts at different points in time. This problem could be resolved by focusing on a particular time period but the thesis may be considered incomplete if it did not consider the entire history of the concept. It was therefore decided not to undertake a conceptual history of evidence based medicine.

2.2.2. Discourse Analysis:

Discourse can be understood as a continuous stretch of language. Discourse analysis refers to a range of methods that can be used to analyse language to understand the meaning that is conveyed. In order to undertake discourse analysis it is important to appreciate the sociohistorical context and reconstruct the beliefs and motives of writers to identify the full range of meanings that are conveyed. The researcher must differentiate between illocutionary acts, which are a resource of language, and perlocutionary acts which are designed to produce an effect. Discourse analysis can be challenging as meanings may not be explicitly stated and intended meanings can be contemporaneously misinterpreted (Skinner 2002).

A discourse analysis of evidence based medicine would be interesting as a number of terms with perlocutionary force, such as 'systematic review' and new 'paradigm', have been used to reinforce the power of the concept. Even the name of the concept, evidence based medicine, paints a powerful picture of any opposing concepts of medicine. A discourse analysis of evidence based medicine could focus on the way that the concept has been used by different groups to establish power and professional dominance within medicine. Power relationships could be interpreted using the theory of countervailing powers. This theory hypothesises that the power of any dominant professional group will be countered by other groups over time (Light 1991).

A discourse analysis could reveal how various groups, including the medical profession, patients, purchasers and regulators of healthcare, had struggled to control power and resources within medicine. An analysis could also clarify whether evidence based medicine has strengthened or weakened the professional dominance of the medical profession. The biggest challenge to undertaking a discourse analysis would be the identification and analysis of large bodies of discourse to determine the perspectives of all relevant groups at different times. I anticipated that it would be difficult to determine whether evidence based medicine had strengthened or weakened the professional position of the medical profession because of the existence of different groups within medicine and likely fluctuating power relationships. It was therefore decided not to undertake a discourse analysis of evidence based medicine.

2.2.3. Qualitative Methods in General:

Interviews, which may be structured or unstructured, can be used to improve understanding of concepts and explore the construction and negotiation of meanings. Structured interviews follow a rigid structure with predetermined questions whereas unstructured interviews follow a more flexible structure. Structured interviews are considered to have good reliability but the method does not allow exploration of unanticipated perspectives. Unstructured interviews can allow a plurality of perspectives to emerge but the method is

considered less reliable and replication of results is difficult. Unstructured interviews also provide the researcher with an opportunity to test out preliminary understanding (Kelleher et al 2006, Alshenqeeti 2014).

Unstructured interviews could be used to improve our understanding of evidence based medicine by exploring the meaning of the concept from the perspective of different groups. The importance of hierarchies of evidence and perceptions about the way evidence based medicine has been used to control power and resources within medicine could be investigated. The results that would be obtained using this method are difficult to anticipate because they would depend upon the groups that were sampled and the stories that emerged.

Unstructured interviews could provide an interesting insight into evidence based medicine although the researcher must recognise that they can influence interviewee responses. Furthermore, the method can be time consuming (Alshenqeeti 2014). In order to comprehensively investigate evidence based medicine the views of patients, medical professionals, purchasers and regulators of healthcare would need to be sampled but some of these groups, particularly purchasers and regulators of healthcare, would be difficult to access. It would be more feasible to interview individuals representative of different professional groups within medicine to explore their perceptions of how evidence based medicine had influenced power relationships. However, as several studies have suggested poor

understanding of evidence based medicine amongst rank-and-file medical professionals (Iqbal and Glenny 2002, Young et al 2002) it is unclear whether any results would be meaningful. This would need to be tested in a pilot study. Unstructured interviews then, whilst interesting, were determined to perhaps not be the best method to get at the heart of evidence based medicine.

2.2.4. Analytical Philosophy:

Analytical philosophy is a range of methods that can be used to identify and analyse concepts and arguments in order to understand philosophical problems (Smith 2003). This method is used to identify arguments and present them in a structured way to reveal the argument form, premises and conclusion. An argument is considered sound if the argument form is valid and the premises are true. Any argument that is sound has a conclusion that is necessarily true. This method allows detailed analysis of concepts, arguments and philosophical problems (Smith 2003, Walton 2008).

Evidence based medicine uses hierarchies of evidence to justify knowledge claims but the existence of different hierarchies allows knowledge claims to be justified in different ways. Analytical philosophy could be used to critically analyse the arguments that are used to support the different hierarchies. This may explain why there are different hierarchies of evidence and improve

understanding of evidence based medicine. This method differs from the other methods that have been outlined because it focuses on evidence based medicine itself rather than the way that evidence based medicine has been used to control power and resources within medicine. It is therefore a more direct evaluation of evidence based medicine.

Analytical philosophy could be used to provide an important insight into evidence based medicine. This method appealed to me because, as a practising clinician, I was interested in the way that the hierarchies influenced my clinical practice. A critical analysis could be restricted to a discrete written discourse so it solved some of the practical problems associated with undertaking a conceptual history or discourse analysis of evidence based medicine. It is primarily for these reasons that analytical philosophy was chosen as the method for this thesis.

2.3. The Method - Analytical Philosophy:

The method of analytical philosophy dates back to the time of Aristotle and can broadly be categorised into formal and informal approaches[10]. Formal approaches to analytical philosophy reformulate arguments and concepts in

---

[10] Modern analytical philosophy originated in the work of G.E. Moore and Bertrand Russell in the early twentieth century although it has evolved through a number of different phases with time. Formal approaches to analytical philosophy originated with Russell and informal approaches to analytical philosophy originated with Moore (Preston 2006). Formal approaches to analytical philosophy utilise truth tables, truth trees and indirect proofs but these are not relevant to informal approaches (Smith 2003).

ordinary language into an object language to remove ambiguities and clarify argument structure. Informal approaches to analytical philosophy analyse arguments and concepts within ordinary language (Smith 2003, Preston 2006). The approach used within this thesis is expressly informal in approach.

An argument consists of one or more premises and a conclusion where the premises are offered in support of the conclusion. Premises and conclusions are relative terms and the conclusion of one argument may be offered as a premise in a later argument. Theses that use the method of analytical philosophy should be envisaged as branching structures where the conclusions of arguments developed in the earlier chapters are used as premises in later chapters.

In order to undertake this critical analysis of evidence based medicine the arguments that are used to support the different hierarchies of evidence must first be identified. These arguments represent the commitment store[11] of the hierarchies of evidence. The different hierarchies of evidence can be identified by undertaking a systematic review. Recognition of the arguments is more complicated and involves iterative reading of the source material to identify premises and conclusions[12]. Arguments that are identified are

---

[11] The commitment store refers to all the arguments taken in their totality.
[12] Arguments are rarely clearly presented and it is often necessary to use indicator words such as 'therefore', 'because' and 'hence' to identify the premises and conclusions of different arguments (Walton 2008).

interpreted using the principle of charity. If arguments are not presented in the most favourable light the analyser may be guilty of attacking a straw man (Walton 2008).

Sometimes arguments that are identified may appear invalid because all of the premises are not explicitly stated. It is therefore important to interpret the arguments with charity rather than simply dismissing arguments because they are not clearly stated. Too hasty dismissal of an argument can result in the construction of a 'straw man' version of evidence based medicine. The result would undermine the credibility of this thesis. It is therefore important to give particular attention to supressed premises. These can be identified through wider reading of the literature pertaining to evidence based medicine[13]. Any suppressed premises that are added must be clearly identified and added in a way that is not misleading or detrimental to the argument (Walton 2008). Premises are often suppressed because they are widely accepted and their absence does not necessarily indicate that an argument is unsound. Nevertheless, suppressed premises should always be considered because, even when they are widely accepted, they may be false.

---

[13] The history of evidence based medicine, and its origins in the discipline of clinical epidemiology, is described in detail by Daly (2005). Further contextual information about the development of evidence based medicine is provided by books such as the *'Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice'* (Guyatt et al 2008) *'Evidence-Based Medicine How to Practice and Teach it'* (Straus et al 2011) and *'The Philosophy of Evidence Based Medicine'* (Howick 2011).

Once the arguments that are used to support the different hierarchies have been identified they are analysed to assess the validity of the argument form and the truth of the premises. An argument form is valid if there is no possible situation in which the premises can be true and the conclusion false. Examples of valid argument forms include modus ponens and modus tollens[14]. If an argument form is valid, and the premises are true, the argument is considered sound because the conclusion is necessarily true. The counterexample technique, where true premises lead to a false conclusion, can be employed to demonstrate invalid argument forms (Smith 2003). Within this thesis arguments are presented in a structured way to facilitate critical analysis.

An argument may be sound, but not persuasive to the reader, if the premises are unacceptable to them, they cannot comprehend the validity of the argument or the argument consists of a single premise restated as the conclusion[15]. Premises may be unacceptable if they do not reflect the value commitments of the reader. Value commitments cannot be shown to be true or false but this does not mean that they cannot be highlighted and explored. When arguments are analysed it is therefore sometimes necessary to comment on their persuasiveness (Shand 2000).

This thesis uses informal analytical philosophy to analyse the arguments that are used to support the different hierarchies of evidence. The arguments are

---

[14] Modus ponens: If A then C, A, hence C; modus tollens: If A then C, not C hence not A (Smith 2003).
[15] A single premise restated as a conclusion is an assertion not an argument (Smith 2003).

presented in a structured way to provide clarity and improve understanding. The structured approach that is used highlights the premises, conclusion and argument form and this is then used to facilitate critical analysis of the arguments.

2.4. Limitations of Analytical Philosophy:

In order to undertake this critical analysis of evidence based medicine the commitment store of the hierarchies of evidence must first be identified. This involves iterative reading of the source material to identify the arguments that are used to justify the different hierarchies. It is accepted that this is a subjective approach that involves interpretation of the literature. The arguments that are presented in this thesis have been developed with reference to multiple sources and this should reassure the reader of the robustness of the method that is used.

Some of the texts that have been used to develop the arguments and concepts presented in this thesis are over fifty years old. It is important to appreciate that our understanding of any historical text is limited by our preunderstanding[16]. We should not assume that the meaning that a contemporary reader would attribute to a text is the same meaning that the

---

[16] Preunderstanding reflects our values, expectations, resources and conception of the world (Hampsher-Monk et al 1998).

writer was attempting to convey at the time the text was written. This would imply a shared understanding (Hampsher-Monk et al 1998). Wider reading of the evidence based medicine literature has therefore been undertaken to place the arguments and concepts that are presented in the correct context. All of the arguments and concepts that are presented are supported by references and where different interpretations are plausible this has been explored. It is acknowledged that historical texts cannot be interpreted in isolation of later texts and it can be problematic to identify the meaning intended by the writer at the time of publication. This is a limitation of any research method that interprets historical texts and is not unique to analytical philosophy (Habermas 1988).

Analytical philosophy involves analysis of both the argument form and the truth of the premises. The concept of truth is therefore fundamental to the method of analytical philosophy. However, although truth is a concept that we routinely use, the concept is surprisingly difficult to define and a number of different theories of truth have been proposed. These include the correspondence theory, semantic theory, coherence theory and pragmatic theory of truth (Simmons 2005).

The correspondence theory of truth states that p is true if and only if p corresponds to a fact. This theory of truth presumes an external shared reality and this is not universally accepted. The semantic theory of truth describes a theory of truth which can be applied to an object language in

formal analytical philosophy but this theory cannot be extrapolated to informal analytical philosophy. This is because an ordinary language needs to refer to external 'facts' whereas an object language does not. The coherence theory of truth relates truth to our background knowledge. This theory allows truth to be relativistic so what is true for one individual may legitimately be false for another. This theory of truth creates problems for the method of analytical philosophy as the same argument could be both sound and unsound depending upon perspective. Pragmatic theories of truth claim that usefulness is the essential mark of truth. However, useful beliefs may be false and useless beliefs may be true. Similar to the coherence theory of truth, the pragmatic theory of truth can also be interpreted from a relativistic perspective (Simmons 2005).

Within this thesis the correspondence theory of truth is used. This theory of truth is compatible with the method of informal analytical philosophy as it cannot be interpreted from a relativistic perspective. This theory of truth reflects the way that the concept of truth is routinely used and it should be familiar to many proponents of evidence based medicine (Goldenberg 2006). It is acknowledged that the correspondence theory of truth is associated with some problems but these problems are considered to be beyond the scope of this thesis.

2.5. Summary:

Within this thesis the method of informal analytical philosophy is used to identify and analyse the concepts and arguments that underpin the different hierarchies of evidence and the claims that evidence based medicine is rational, science and a new paradigm. The arguments are presented in a structured way to highlight the argument form, premises and conclusion and facilitate critical analysis. It is accepted that the method that is presented uses a correspondence theory of truth and that the arguments have been developed following interpretation of the source material. In order to progress with this critical analysis the different hierarchies of evidence must first be identified. This is the focus of Chapter 3 where a typology of hierarchies of evidence is presented.

Chapter 3: Typology of Hierarchies of Evidence

3.1 Introduction:

Hierarchies of evidence are fundamental to the justification of the knowledge claims that are made by evidence based medicine. These knowledge claims are subsequently used to determine the treatment interventions that patient receive, funding for medical care, medicolegal standards and the medical research agenda. It is interesting that there are different hierarchies (West et al 2002, Gugiu and Gugiu 2010, Turner 2011) as this suggests that knowledge claims can be justified in different ways. It is therefore important to investigate why there are different hierarchies as this may have significant implications for the knowledge claims that are made by evidence based medicine. One way to investigate the hierarchies of evidence would be to first identify, and then analyse, the arguments that are used to support them.

This chapter presents a systematic review and thematic analysis of the different hierarchies used by evidence based medicine. These are first identified through a systematic review of the medical literature. A hierarchy of evidence is identified as any system within medicine that ranks the importance of evidence primarily based upon the study design that is used to produce that evidence (Upshur 2009). Once identified, the different hierarchies are compared and contrasted to identify emerging themes.

Attributes that are considered include the different study designs that provide the highest and lowest levels of evidence, the number of levels and the properties that are used to rank different study designs. Selected hierarchies are then presented in detail to illustrate how the hierarchies are used and have evolved with time. The chapter concludes with some reflections on the limitations of systematic reviews and thematic analyses. This chapter provides an initial insight into the complexity that exists amongst the hierarchies of evidence and provides a foundation for the critical analysis presented in later chapters.

3.2 Systematic Review:

Hierarchies of evidence were identified following a detailed search of MEDLINE via the OVID interface, CINAHL via the EBSCO interface and the Cochrane Methodology Register (Appendix 2). A search strategy was constructed with the assistance of a medical librarian and revised appropriately for each database. The literature searches were initially undertaken between November 2011 and January 2012. A further search of the MEDLINE database via the OVID interface was undertaken in April 2016 to identify any further hierarchies published in the intervening period. Once the hierarchies of evidence had been identified the literature was further searched for any studies relating to the development, implementation and modification of each hierarchy.

The abstracts of all studies identified by the search strategy were screened by a single reviewer using pre-defined inclusion and exclusion criteria. The full text of any study that potentially proposed a new hierarchy, or reviewed existing hierarchies, was obtained and examined. Additional hierarchies were identified from the reference lists of studies meeting the inclusion criteria, the reviews undertaken by West et al (2002) and Blunt (2015) and evidence based medicine textbooks (Daly 2005, Guyatt et al 2008, Howick 2011, Straus et al 2011).

3.3 Inclusion Criteria:

1. Studies presenting hierarchies of evidence that have been used within medicine to determine the effectiveness of treatment interventions.

2. Studies presenting hierarchies of evidence that have been used within medicine to determine the answers to questions that are not about the effectiveness of treatment interventions.

3. Studies that review or evaluate the use of hierarchies of evidence within medicine.

<u>3.4 Exclusion Criteria:</u>

1. Studies presenting hierarchies of evidence that were not available in the English language.

2. Studies presenting hierarchies of evidence where the full text was unavailable.

3. Studies presenting hierarchies of evidence that were not significantly different from previously published hierarchies of evidence.

A hierarchy of evidence was considered significantly different from other hierarchies if it could potentially interpret the same body of evidence in a different way. Any hierarchy that had a different number of levels, included different study designs, or ranked previously included study designs in a different way was designated as a new hierarchy of evidence. The reference lists of any potential new hierarchies were closely examined as they often revealed closely related hierarchies. As an example, the hierarchy described by the American College of Chest Physicians (Sackett 1989) was considered a new hierarchy, despite its similarity to the hierarchy used by the Canadian Task Force (Spitzer et al 1979), because it differentiated between randomised controlled trials depending upon the statistical significance of the results. These two hierarchies would therefore interpret randomised controlled trials with non-statistically significant results in different ways. The Canadian Task Force (Spitzer et al 1979) would rank this study design as the

highest level of evidence whereas the American College of Chest Physicians (Sackett 1989) would not.

3.5. Thematic Analysis

Once the different hierarchies of evidence had been identified the following information was collected: author, year of publication, whether there were different hierarchies for different questions, study designs providing the highest and lowest levels of evidence, number of levels, factors independent of study design that affected hierarchical position, the property that was used to rank different study designs and the existence of a separate grading process.

The hierarchies were then iteratively analysed to identify all of the different arguments that were presented to support them. The frequency with which different arguments were used was also recorded. These arguments were then grouped together within different themes to facilitate analysis. These themes emerged from the data but they were driven by my research aim which was to improve understanding of evidence based medicine. The themes that emerged therefore allowed me to understand, from the perspective of evidence based medicine, why hierarchies were required, why randomised controlled trials and systematic reviews were prioritised at the

expense of expert opinion and why there were so many different hierarchies of evidence.

3.6 Results:

The search strategy detailed above identified 45 different hierarchies of evidence. 42 of these hierarchies were found to rank the importance of evidence for determining the effectiveness of treatment interventions (Cochrane 1972, Spitzer et al 1979, Trout 1981, Sackett 1989, Cook et al 1992, Carruthers et al 1993, Gross et al 1994, Braunewald et al 1994, Guyatt et al 1995, Hadorn et al 1996, Goldbloom 1997, Granados et al 1997, Guyatt et al 1998, Djulbegovic and Hadley 1998, Meltzer et al 1998, Stetler et al 1998, Wilkinson 1999, Chestnut et al 1999, Guyatt et al 2000, American Heart Association 2000, Briss et al 2000, Greer et al 2000, Harbour and Miller 2001, Ball and Phillips 2001, Guyatt et al 2001, Bandelow et al 2002, Evans 2003, Ebell et al 2004, Mitchell and Friese 2007, Newhouse et al 2007, Armola et al 2009, Brozek et al 2009, Grunze et al 2009, Merlin et al 2009, Hooper et al 2009, Gugiu and Gugiu 2010, Sayre et al 2010, Joanna Briggs Institute 2011, Howick et al 2011, Tomlin and Borgetto 2011 and Gugiu 2015). A further three hierarchies were identified which did not consider treatment interventions. These hierarchies were found to rank the importance of evidence for questions of aetiology (Ariens et al 2000) and diagnosis (Zhang et al 2006, Schunemann et al 2008). All of the hierarchies that were identified are listed chronologically in Table 1 overleaf. The 42

hierarchies that have been used to determine the effectiveness of treatment

interventions are detailed in Appendix 3 in chronological order.

Year	Hierarchy of Evidence
1972	Cochrane
1973	
1974	
1974	
1976	
1977	
1978	
1979	Canadian Task Force
1980	
1981	How to Read Clinical Journals'
1982	
1983	
1984	
1985	
1986	American College of Chest Physicians
1987	
1988	
1989	Canadian Task Force modification 1
1990	
1991	
1992	American College of Chest Physicians modification 1
1993	American Association of Critical Care Nurses, Canadian Hypertension Society
1994	Infectious Diseases Society of America, Brunewald et al
1995	Evidence Based Medicine Working Group, Australian National Health and Medical Research Council
1996	Agency for Healthcare Research and Quality
1997	EUR-ASSESS
1998	American College of Chest Physicians modification 2, Djulbegovic & Hadley, Canadian Diabetes Association, Stetler et al, Scottish Intercollegiate Guidelines Network
1999	Oxford Centre for Evidence Based Medicine, American Academy of Ophthalmology, Chestnut et al

Table 1a: Hierarchies of Evidence listed chronologically according to date of

first use.

Year	Hierarchy of Evidence
2000	Evidence Based Medicine Working Group modification 1, Cardiopulmonary Resuscitation Guidelines, Institute for Clinical Systems Improvement, United States Task Force on Community Preventive Services, Ariens et al.
2001	American College of Chest Physicians modification 3, Stetler modification 1
2002	World Federation of Societies of Biological Psychiatry
2003	Evans
2004	Strength of Recommendation Taxonomy, Grading of Recommendations Assessment Development and Evaluation
2005	Joanna Briggs Institute for Nursing
2006	European League against Rheumatism
2007	Oncology Nursing Society Putting Evidence into Practice Weight of Evidence Classification Schema, John Hopkins Evidence Rating Scale
2008	American Association of Critical Care Nurses modification 1, Grading of Recommendations, Assessment, Development and Evaluation (Diagnosis)
2009	Australian National Health and Medical Research Council modification 1, World Federation of Societies of Biological Psychiatry modification 1
2010	Cardiopulmonary Resuscitation Guidelines modification 1
2011	Oxford Centre for Evidence Based Medicine modification 1, Research Pyramid
2012	
2013	
2014	
2015	Hierarchy of Evidence and Appraisal of Limitations

Table 1(continued): Hierarchies of Evidence listed chronologically according

to date of first use.

Seven of the systems that were identified contained a number of different

hierarchies (Ball and Phillips 2001, Evans 2002, Merlin et al 2009, Sayre et al

2010, Joanna Briggs Institute 2011, Howick et al 2011, Tomlin and Borgetto

2011)[17]. In these systems, the question that is asked dictates the hierarchy that should be used to determine the most appropriate evidence. For example, Evans (2002) presented different hierarchies for questions of effectiveness, appropriateness and feasibility; Sayre et al (2010) presented different hierarchies for prognosis, diagnosis and treatment interventions and the Oxford Centre for Evidence Based Medicine (Howick et al 2011) presented different hierarchies for diagnosis, prognosis, treatment interventions, common harms, rare harms, screening and frequency of the problem. The different questions that were collectively considered by the hierarchies of evidence are shown in Table 2 overleaf. There appears to be little consistency amongst these hierarchies as they consider a variety of different questions.

When the 42 different hierarchies of evidence were considered collectively, a number of different study designs are found to provide the highest level of evidence. These are shown in Table 3 overleaf. Some hierarchies allow more than one study design to provide the highest level. Randomised controlled trials were found to provide the highest level in 24 hierarchies of evidence although in 15 of these hierarchies only randomised controlled trials meeting certain criteria provided the highest level. For example, in the hierarchy proposed by Evans (2003) only multi-centre randomised controlled trials provide the highest level of evidence; in the hierarchy used by the American

---

[17] The Canadian Hypertension Society (Carruthers et al 1993) and the Canadian Diabetes Association (Meltzer et al 1998) purported to present hierarchies for diagnosis and prognosis. However, these groups presented desirable methodological criteria not hierarchies of evidence. They have therefore been excluded.

Heart Association (2000) only prospective, statistically significant randomised controlled trials provide the highest level of evidence and in the Agency for Healthcare Research and Quality hierarchy (Hadorn et al 1996) only well-conducted, multi-centre randomised controlled trials with over 100 patients provide the highest level of evidence. In other hierarchies methodological criteria dictate which randomised controlled trials provide the highest level, although these methodological criteria do vary (Mitchell and Friese 2007, Brozek et al 2009, Gugiu 2015).

Question	Number of Hierarchies of Evidence
Aetiology	3
Appropriateness	2
Diagnosis	6
Economic analysis	2
Feasibility	2
Frequency of problem	1
Harm[18]	2
Meaningfulness	2
Outcomes research	1
Prognosis	4
Qualitative research	1
Screening	2

Table 2: Hierarchies of evidence that do not consider the effectiveness of treatment interventions.

Systematic reviews and/or meta-analyses provided the highest level of evidence in 20 hierarchies of evidence although in 9 of these hierarchies only

---

[18] Howick et al (2011) differentiate between common harms and rare harms.

systematic reviews/meta-analyses meeting certain criteria provided the highest level. For example, in the Strength of Recommendation Taxonomy (Ebell et al 2004) hierarchy only systematic reviews or meta-analyses with consistent findings provide the highest level of evidence. Other study designs or forms of evidence, including n-of-1 studies (Guyatt et al 2000, Howick et al 2011), observational studies (Brozek et al 2009), meta-syntheses of qualitative studies (Armola et al 2009) and expert opinion (Mitchell and Friese 2007), provided the highest level of evidence in 9 hierarchies.

Study design	Hierarchies of evidence
Randomised controlled trial	9
Randomised controlled trial meeting certain criteria	15
Systematic review/meta-analyses	11
Systematic review/meta-analyses meeting certain criteria	9
Other study designs	9

Table 3: Study designs that provide the highest level of evidence in hierarchies that are used to determine the effectiveness of treatment interventions.

When the 42 different hierarchies of evidence were considered collectively a number of different study designs were also found to provide the lowest level of evidence. These study designs are shown in Table 4 overleaf. Some hierarchies allow more than one study design to provide the lowest level of evidence. Expert opinion provides the lowest level of evidence in 19 hierarchies although only one of these hierarchies differentiates between the

opinion of an individual expert and a group of experts (Newhouse et al 2007). Observational studies provide the lowest level of evidence in 12 hierarchies (observational studies include cohort studies, cross-sectional studies, case-control studies, descriptive studies), case series provide the lowest level of evidence in 12 hierarchies and case reports provide the lowest level of evidence in 5 hierarchies. Studies that have been poorly conducted provide the lowest level of evidence in 7 hierarchies, with 3 of these ranking poorly conducted randomised controlled trials as the lowest level of evidence (Mitchell and Friese 2007, Brozek at al 2009, Gugiu 2015). The remaining hierarchies rank mechanistic reasoning (Sayre et al 2010, Howick et al 2011), manufacturer's recommendations (Armola et al 2009), rational conjecture, common sense (American Heart Association 2000) and the absence of evidence (Bandelow et al 2002, Grunze et al 2009) as the lowest level of evidence.

Form of Evidence	Hierarchies of Evidence
Expert opinion	19
Observational studies	12
Case series	12
Studies with flawed methodology	7
Case reports	5
Clinical experience	4
Common sense/rational conjecture	3
Physiology, bench research or first principles	3
Mechanistic reasoning	2
Absence of evidence	2

Table 4: Study designs that provide the lowest level of evidence in hierarchies that are used to determine the effectiveness of treatment interventions.

When considered collectively, the 42 different hierarchies contain a number of different levels between the highest and lowest level of evidence. This is shown overleaf in Table 5. The simplest hierarchies have only 3 levels of evidence (Cochrane 1972, Spitzer et al 1979, Djulbegovic and Hadley 1998, Ebell et al 2004) whereas the most complicated hierarchies have 10 or more levels and sub-levels (Hadorn et al 1996, Ball and Phillips 2001).

Number of levels	Number of hierarchies of evidence
2	0
3	9
4	10
5	6
6	7
7	1
8	6
9	1
10	1
11	0
12	1

Table 5: Levels of evidence in the different hierarchies that are used to determine the effectiveness of treatment interventions.

The property that is used to rank different study designs within the 42 different hierarchies is often unclear. Many of the hierarchies either failed to explain the property that was used or referred to nebulous qualities such as 'evidence strength' or 'quality of evidence' without explicitly defining these terms. Some hierarchies were clear about the property that was used although a number of different properties were mentioned. These properties

included truth (Cook et al 1992, Djulbegovic and Hadley 1998), bias (Guyatt et al 1995), validity (Hadorn et al 1996, Scottish Intercollegiate Guidelines Network 2011), confidence (Trout 1981) and trustworthiness (Joanna Briggs Institute 2011).

Twenty two of the hierarchies of evidence were presented in conjunction with a separate process for formulating treatment recommendations. The process of formulating treatment recommendations is often termed the grading process. These systems[19] use a hierarchy to determine the level of evidence supporting a treatment intervention but only make a final treatment recommendation following consideration of other factors. These factors may include financial costs, feasibility, external validity and adverse effects. It is therefore possible to have a treatment intervention that is ranked highly by a hierarchy but that is not recommended because it is too expensive or associated with adverse effects. The different factors that were considered during the formulation of treatment recommendations varied between different systems.

.

Hierarchies that are presented in conjunction with a separate process for formulating treatment recommendations include the systems proposed by the Canadian Task Force system (Spitzer et al 1979), Scottish Intercollegiate Guidelines Network (Harbour and Miller 2001) and the Joanna Briggs Institute for Nursing (2011). The hierarchies proposed by Sackett (1989),

---

[19] In this thesis the term 'system' is used as an overarching term to include both the hierarchy of evidence and any separate process that is used to formulate treatment recommendations.

Cook et al (1992), Carruthers et al (1993), Hadorn et al (1996) and Ball and Phillips (2001) purport to present a separate process for formulating treatment recommendations. However, in these systems the grade of treatment recommendation is directly determined by the hierarchy so the grading process is superfluous. These 5 hierarchies were therefore categorised as not having a separate process for grading recommendations.

3.6.1. Summary of Results:

The systematic review identified 45 different hierarchies of evidence that have been used within evidence based medicine. Most of these hierarchies are used to determine the effectiveness of treatment interventions although they have also been developed to support decision making in other areas of medicine. When considered collectively the hierarchies contain a variable number of levels and appear to use a variety of different properties to rank different study designs. Some hierarchies are accompanied by a separate process for formulating treatment recommendations. In the hierarchies that are used to determine the effectiveness of treatment interventions randomised controlled trials, systematic reviews and meta-analyses often provide the highest level of evidence. However, this is not a universal trend and randomised controlled trials can provide the lowest level of evidence when they are poorly conducted. Expert opinion and observational studies often provide the lowest level of evidence but this is also not a universal trend and they can provide the highest level of evidence in some hierarchies.

<u>3.7. Selected Hierarchies of Evidence:</u>

In this section, the hierarchies developed by Cochrane (1972), the Canadian Task Force (Spitzer et al 1979), American College of Chest Physicians (Sackett 1989, Cook et al 1992, Guyatt et al 1998, Guyatt et al 2001), Oxford Centre for Evidence Based Medicine (Ball and Phillips 2001, Howick et al 2011), GRADE Working Group (2004) and Oncology Nursing Society (Mitchell and Friese 2007) are summarised to illustrate how they have been used and evolved with time. This section expands upon the results of the systematic review and provides important contextual information. It is not my intention to summarise all hierarchies or describe any single hierarchy in great detail. This would result in a morass of information and contribute little to the overall thesis. The reader is referred to the original articles if they would like further information on particular hierarchies.

The hierarchies summarised in this section have been selected because they are early examples, commonly used, illustrate evolution or have been considered important, at different times, by proponents of evidence based medicine. It is accepted that the hierarchies that are presented have been subjectively selected and that different hierarchies may have been used to illustrate some of the points considered below.

<u>3.7.1. Cochrane (1972):</u>

The first hierarchy of evidence was described by Cochrane (1972) in Chapter 4 of *'Effectiveness and Efficiency: Random Reflections on Health Services'*. Cochrane (1972) believed that treatment provided under the auspices of the National Health Service should be cost-effective, beneficial to patients and equitable. He further believed that randomised controlled trials could be used to achieve these aims. Cochrane (1972) did not use the term 'hierarchy of evidence', but he very clearly ranked the importance of different study designs for determining the effectiveness of treatment interventions. Randomised controlled trials were considered to provide the most valuable evidence and clinical opinion was considered to provide the least valuable evidence with observational studies occupying an intermediate level. This was the first attempt to rank different study designs in a hierarchical manner.

Cochrane (1972) provided two arguments to justify his ranking of randomised controlled trials as the highest level of evidence: randomisation could be used to create comparable treatment and control groups and randomised controlled trials could be blinded to control measurement bias. These arguments are consistently used by many later hierarchies. Although Cochrane (1972) valued randomised controlled trials he also recognised that the study design were often impractical to implement, blinding was difficult when subjective outcome measures were used and randomisation could produce non-comparable treatment and control groups. Cochrane (1972)

was also clear that randomised controlled trials were unnecessary to support treatment interventions when the effects of a treatment intervention were obvious.

This hierarchy of evidence is considered important because it was the first hierarchy; recognises potential problems associated with the hierarchical ranking of knowledge claims derived from different study designs and provides the earliest arguments for the ranking of randomised controlled trials as the highest level of evidence.

3.7.2. Canadian Task Force (1979):

The Canadian Task Force (Spitzer et al 1979) used their hierarchy to determine whether specific conditions should be included in the Canadian annual health examination. This is widely cited as the first hierarchy (Atkins et al 2004) although it is clearly predated by the work of Cochrane (1972). This hierarchy ranked randomised controlled trials as the highest level of evidence and expert opinion as the lowest level of evidence although no arguments were presented to support this ranking.

The system presented by the Canadian Task Force is important because it is the first system to make a distinction between a hierarchy of evidence and a

grade of recommendation. In this system the grade of recommendation is determined following consideration of 3 factors: burden of suffering, the validity and acceptability of the manoeuvre used to detect or prevent the condition and the effectiveness of the treatment intervention. The effectiveness of the treatment intervention, as determined by the hierarchy, is the most important of these factors, as the Canadian task Force are clear that the highest treatment recommendation (Grade A) is unlikely to be made unless the intervention is supported by evidence from randomised controlled trials.

This hierarchy was used to support the conclusion that the routine annual health examination should be abandoned in Canada. The Canadian Task Force instead advocated a selective annual health examination targeting at-risk groups such as neonates and pregnant women. This illustrates the power that hierarchies possess as they can be used to support conclusions that may have huge implications for medicine. Abandoning routine health examinations may have meant that preventable conditions were not diagnosed (and therefore not treated) and would have had significant financial implications for patients, medical professionals and purchasers of health care. Spitzer et al (1979) appear to have been aware of the potential power inherent in their hierarchy as they warned against misuse or manipulation of their system.

3.7.3. American College of Chest Physicians (1986, 1992, 1998, 2001):

The American College of Chest Physicians have used a number of different hierarchies over the last 30 years to produce treatment recommendations following thromboembolic events (Sackett 1989, Cook et al 1992, Guyatt et al 1998, Guyatt et al 2001). Thromboembolic events, such as myocardial infarctions and strokes, are common events associated with significant mortality and morbidity. Antithrombotic medication can reduce mortality and morbidity following thromboembolic events but this medication can also impact on patient's lives[20] and the medication itself may cause bleeding complications. The hierarchies used by the American College of Chest Physicians have therefore been used to produce treatment recommendations which affect millions of people. The hierarchies used by the American College of Chest Physicians have been summarised because they demonstrate evolution over time.

The American College of Chest Physicians first used a hierarchy in 1986 (Sackett 1989). Expert opinion was excluded from this hierarchy because it was considered fallible and susceptible to statistical phenomenon such as regression towards the mean. At this time the grade of recommendation was directly determined by the hierarchy and evidence from any study design that was not a randomised controlled trial could only lead to the lowest grade of treatment recommendation.

---

[20] For example, patients who take warfarin have to have regular blood tests to monitor their bleeding potential.

The hierarchy used by the American College of Chest Physicians was modified in 1992 with the addition of meta-analyses (Cook et al 1992). This was the first hierarchy to include systematic reviews or meta-analyses. Meta-analyses were ranked as level I or II with a positive or negative suffix depending upon whether the results were clinically significant and consistent across included studies. This was the first hierarchy to consider clinical significance and heterogeneity, although no guidance was provided on how these factors should be interpreted. At this time the grade of recommendation was still directly determined by the hierarchy.

The hierarchy used by the American College of Chest Physicians was modified again in 1998 (Guyatt et al 1998). This hierarchy had only 3 levels and ranked randomised controlled trials with no heterogeneity as the highest level, randomised controlled trials with heterogeneity as the second level and observational studies as the lowest level of evidence. Meta-analyses were purposely excluded from this hierarchy because they could include poor quality studies[21] (Guyatt et al 2001). In this system the level of evidence did not directly determine the treatment recommendation and other factors, including risks and benefits, were considered.

---

[21] The inclusion of heterogeneity in this hierarchy suggests that meta-analyses undertaken by users are acceptable. It is actually unclear whether this hierarchy can be applied to individual study designs and it may be more appropriately considered a hierarchy of evidence bases (Blunt 2015).

The hierarchy used by the American College of Chest Physicians was modified again in 2001 to include extrapolations from randomised controlled trials and overwhelming evidence from observational studies (Guyatt et al 2001). This additional level was added because it was recognised that it may be difficult to undertake a randomised controlled trial when a particular treatment intervention was already widely accepted (Hirsch et al 2001)[22]. This hierarchy also included randomised controlled trials with methodological flaws. In the 2001 system, the importance of the hierarchy was reduced and consideration of risks and benefits assumed greater importance. The primacy of consideration of risks and benefits was illustrated by placing the recommended grade before the evidence level i.e. 1A and not A1. This system also differed from earlier systems because financial costs could now explicitly influence the final treatment recommendation. In 2006 the American College of Chest Physicians abandoned their own system and adopted the Grading of Recommendations, Assessment, Development and Evaluation system (Baumann and Gutterman 2006).

The hierarchies used by the American College of Chest Physicians are interesting because they changed 4 times over a 20 year period before they were finally abandoned. These hierarchies were the first to consider meta-analyses but they never included expert opinion. The evolution of these systems illustrates a decline in the relative importance of research evidence, as determined by the hierarchy, over time although this is not a consistent

---

[22] A randomised controlled trial is not considered possible in these circumstances because it would be unethical for the control group to not receive the treatment intervention.

finding amongst other hierarchies. It is important to appreciate that the hierarchies used by the American College of Chest Physicians are not the only hierarchies to have evolved with time. For example, the hierarchies proposed by the Canadian Task Force (Spitzer et al 1979, Goldbloom 1997) and the Oxford Centre for Evidence Based Medicine (Ball and Phillips 2001, Howick et al 2011) have also evolved.

3.7.4. Oxford Centre for Evidence Based Medicine Levels of Evidence (1999, 2011):

The system proposed by the Oxford Centre for Evidence Based Medicine (Ball and Phillips 2001) is important because it was the first to present different hierarchies of evidence for different questions. Each individual hierarchy ranked the systematic review as the highest level of evidence and the evidence level directly determined the grade of treatment recommendation. This system was modified in 2011 with the addition of new hierarchies for common and rare harms, screening and common problems and removal of the hierarchy for economic analyses. Howick et al (2011) reported that the hierarchy for economic analyses was removed because of uncertainty about what constituted good evidence. This would have had

important implications for any recommendations made using the previous hierarchy for economic analyses[23].

Following its development in 1999 the Oxford Centre for Evidence Based Medicine system was widely adopted by both medicine and dentistry in the United Kingdom. This was an interesting development as the system was originally designed for use in general acute medicine as part of the 'Evidence-Based On-Call' project and its potential applicability to other areas of medicine was questioned by the developers (Ball and Phillips 2001).

3.7.5. Grading of Recommendations, Assessment, Development and Evaluation (GRADE) System (2004):

The GRADE system was developed in 2004 by the GRADE Working Group following a review of existing hierarchies. Existing hierarchies were criticised because they missed important elements, had poor reproducibility and relied on implicit judgements. The GRADE system was proposed as a solution to these problems. The GRADE system can be used to determine the level of evidence supporting a particular treatment intervention or develop a treatment recommendation following a wider consideration of risks and benefits, patient values and resource implications in addition to an

---

[23] Howick et al (2011) are clear that they also expect the latest Oxford Centre for Evidence Based Medicine hierarchy to change with time. This has potentially important implications for any recommendations made using the existing system.

assessment of the supporting evidence (Atkins et al 2004, GRADE Working Group 2004, Atkins et al 2005).

The GRADE system initially ranks randomised controlled trials as the highest level and observational studies as the lowest level of evidence. However, randomised controlled trials and observational studies can be downgraded or upgraded if certain criteria are fulfilled. For example, randomised controlled trials can be downgraded if they are considered at risk of bias and observational studies can be upgraded when they demonstrate a large magnitude of effect (Guyatt et al 2008)[24].

The GRADE system does not include expert opinion or systematic reviews. Expert opinion is excluded because judgement is considered necessary for the interpretation of all evidence (Brozek et al 2009). Systematic reviews are excluded because they are considered a means of collecting evidence, not a form of evidence, and they are reliant on the quality of the primary studies (Atkins et al 2004). Nevertheless, although systematic reviews are explicitly excluded from the GRADE hierarchy the evidence appraisal process does involve a systematic review to identify relevant randomised controlled trials and observational studies before they are appraised.

---

[24] Guyatt et al (2011a, 2011b) have accepted that there is little empirical evidence to support the criteria that are used to upgrade and downgrade studies and that the threshold for upgrading and downgrading is often arbitrary.

The GRADE system is important because the Working Group behind the system recognised that the existence of multiple hierarchies created problems for evidence based medicine[25]. They aspired to create a universal hierarchy that would replace existing hierarchies and resolve confusion (Guyatt et al 2008). This attempt was unsuccessful as the GRADE system is not universally used and, as Table 1 reveals, it has not prevented the continued development of new hierarchies.

3.7.6. Oncology Nursing Society Putting Evidence into Practice Weight of Evidence Classification Schema (ONSPEP) (2007):

The ONSPEP hierarchy (Mitchell and Friese 2007) was developed to determine the effectiveness of nursing interventions in oncology patients. This hierarchy is primarily important because the opinion of a group of experts can provide the highest level of evidence alongside systematic reviews and randomised controlled trials. In order to attain the highest level, expert recommendations must synthesise any available evidence found following a thorough search and synthesis of the literature. This allows expert opinion alone to provide the highest level of evidence when a search of the literature reveals no evidence, evidence that is flawed or the experts disagree with any evidence that is found. This is the only hierarchy that places expert opinion at the highest level.

---

[25] Other groups have also recognised that multiple hierarchies create problems for evidence based medicine but they have responded in different ways. For example, Glasziou et al (2004) advocated that hierarchies should be abandoned not standardised.

The ONSPEP hierarchy was originally developed to support nursing interventions in oncology patients. However, it has subsequently been recommended as the most appropriate hierarchy to support perioperative nursing interventions in all patients (Steelman et al 2011). This recommendation was made because the ONSPEP hierarchy did not prioritise study designs, such as randomised controlled trials and systematic reviews, which were often unavailable to support nursing interventions. This hierarchy is important because it recognises that expert opinion can be as important as evidence derived from other study designs. The ONSPEP hierarchy also provides an example of a hierarchy targeted at a particular professional group within medicine.

3.7.7. Summary:

In this section, six hierarchies of evidence have been summarised to illustrate how they are used within evidence based medicine and have evolved with time. Some of the arguments that have been used to support the ranking of randomised controlled trials, systematic reviews and expert opinion, within different hierarchies, have been outlined although they have not been analysed. These arguments have led to the inclusion and exclusion of expert opinion and different study designs in different hierarchies. It is notable that the first hierarchy (Cochrane 1972) ranked the randomised controlled trial as the highest level of evidence despite acknowledging that evidence from this

study design was unnecessary in some situations. Problems associated with multiple hierarchies have been recognised within evidence based medicine, particularly by the GRADE Working Group, but attempts to create a universal hierarchy appear to have been unsuccessful.

3.8. Discussion:

All research methods have limitations and any results should be interpreted in light of these limitations. Within this thesis a systematic review has been undertaken to identify different hierarchies of evidence and thematic analysis has been used to identify emerging themes. It is therefore important to consider the limitations of these methods before analysing the arguments that have been used to support the different hierarchies. Systematic reviews identifying different hierarchies have previously been undertaken by West et al (2002) and Blunt (2015) and their findings should also be considered when the results of this systematic review and thematic analysis are interpreted.

The results of any systematic review are determined by the search strategy that is used. In this systematic review a search strategy was developed with the assistance of a medical librarian, building on the search strategy used by a previous systematic review (West et al 2002). The search strategy was designed to have high sensitivity, at the expense of specificity, in an effort to be as comprehensive as possible (Haynes et al 2005). The search strategy that was used actually identified 1178 potentially relevant studies when it was

used to search the MEDLINE database April 2016 (Appendix 2) although many of these studies were excluded following assessment of the abstract. Further hierarchies were identified from the reference lists of studies meeting the inclusion criteria and relevant evidence based medicine textbooks. It is however accepted that there are likely to be further hierarchies that have not been identified by the search process used in this systematic review.

Hierarchies of evidence can only be identified through systematic review if they are accurately coded in medical databases. West et al (2002) reported difficultly identifying different grading systems because of the insensitivity of thesaurus headings. Similar problems were encountered during this systematic review as the thesaurus headings for each study usually related to the treatment intervention that was considered not the hierarchy that was used. West et al (2002) advocated citation tracking and contact with guideline developers, in addition to a comprehensive literature search, as the most efficient means of identifying grading systems. This systematic review used citation tracking but did not contact guidelines developers as this was considered impractical. Accurate coding of any new hierarchies would facilitate any future systematic reviews in this area.

This systematic review was restricted to hierarchies published in the English language for practical reasons. However, this is likely to have introduced language bias. The hierarchy of evidence developed by Tomlin and Borgetto (2011) appears to have been published in German in 2006 but did not

appear in English until 5 years later. This is a clear example of a hierarchy published in a different language and it is possible that there are other hierarchies, in different languages, that have not been identified.

It was not possible to obtain the original articles for some of the earlier hierarchies because they were published in journal supplements and electronic copies were unavailable. In these situations, later studies that described these hierarchies were used. These hierarchies were only included if they were described in adequate detail. It is accepted that this may have introduced inaccuracies if the original hierarchies were modified following initial publication. It was felt important to include these hierarchies because they were of historical significance. Hierarchies that were not described in adequate detail were not included. Where the original studies were unavailable the year of first publication was always clear.

This systematic review excluded hierarchies that were not significantly different from earlier hierarchies. In general terms a hierarchy was considered significantly different if it contained different study designs or ranked the same study designs in a different way. This was a subjective decision made by the reviewer which involved comparing and contrasting different hierarchies. As an example, the hierarchy described by Eccles and Mason (2001) was not included because it was not considered significantly different from the first hierarchy used by the Australian National Health and

Medical Research Council (Appendix 4). Where hierarchies were not considered significantly different any later hierarchies were excluded.

This systematic review was designed to identify systems that ranked the importance of evidence based primarily upon the study design that was used to produce that evidence. However, several systems were identified that ranked pre-appraised sources of evidence within a hierarchical framework (Haynes 2001, Haynes 2006, DiCenso et al 2009). These systems include computerised decision support systems and interpretations of the results of systematic reviews and primary studies as different levels. These systems were not included as they did not rank evidence primarily according to study design.

This systematic review builds on the previous systematic reviews undertaken by West et al (2002) and Blunt (2015). West et al (2002) identified 40 different systems that could be used to grade the strength of a body of evidence within medicine. However, they did not identify 40 different hierarchies as many of the systems used the same hierarchies of evidence. Blunt (2015) identified 88 different hierarchies although his search strategy and inclusion criteria were not described. The systematic review undertaken by Blunt (2015) does not identify significantly more hierarchies than this systematic review as his total includes hierarchies that are very similar to one another and systems that rank pre-appraised evidence within a hierarchical framework.

Once the hierarchies of evidence had been identified they were compared and contrasted to identify recurrent themes and arguments. These recurrent themes and arguments were identified by iterative reading of the hierarchies and wider reading of the evidence based medicine literature. Common themes that were identified included the ranking of randomised controlled trials and systematic reviews/meta-analyses as the highest level of evidence and the ranking of expert opinion as the lowest level of evidence. The arguments that were used to support the respective ranking of randomised controlled trials, systematic reviews and expert opinion were also commonly repeated across different hierarchies. It is accepted that this systematic review may have failed to identify some hierarchies. However, as the themes and arguments that have been identified are consistent across many hierarchies it is unlikely that synthesis of further hierarchies would significantly affect the critical analysis presented in subsequent chapters.

Thematic analysis has been criticised because it lacks transparency and relies upon subjective interpretation of the source material (Pope et al 2007). These are valid criticisms but they can be levelled at any interpretivist approach. The recurrent themes identified amongst the hierarchies that are analysed in later chapters are supported by numerous references and quotes to assure the reader of the robustness of the approach that has been used.

With any qualitative research method the role of the researcher, and the influence they have on knowledge construction, should be considered. Reflexive evaluation acknowledges any assumptions and preconceptions that underpin the research process and strengthens the integrity of conclusions (Finlay 2002). It is therefore important to reflexively evaluate my influence, as the researcher, on the research process used in this thesis.

The reader should understand that I am a practising dentist who works as an NHS Consultant in a District General Hospital. Evidence based medicine was already well established when I qualified in 1999 and I have not known any other way to practice dentistry. I am motivated by a desire to explain why there are so many different hierarchies because they have direct implications for my clinical practice. I do not intrinsically believe that evidence based medicine is irrational and this may influence the way that I have collected and analysed the data. I may also have unconsciously favoured the arguments used to support the hierarchies that are commonly used within dentistry: The Oxford Centre for Evidence Based Medicine Levels of Evidence (Ball and Phillips 2001, Howick et al 2011) and the GRADE system (Brozek et al 2009). Any unconscious bias towards these arguments is unlikely to impact significantly on my analysis as the arguments used by <u>all</u> of the hierarchies have been considered and the hierarchies that are commonly used within dentistry are not supported by arguments that are not repeated by other hierarchies.

Blunt (2015) has described 4 phases in the development of hierarchies of evidence. In the first trend (1979-1991) hierarchies are generally restricted to different study designs and neglect expert opinion. The second phase (1991-1998) is characterised by experimentation with competition between different hierarchies. The third phase (1998-2004) is characterised by the prominence of pre-appraised evidence and the final phase (2004-present) is characterised by increasingly sophisticated hierarchies such as the GRADE system. However, this systematic review has identified a number of hierarchies that do not correspond to these different phases[26]. For example, the hierarchies described by Cochrane (1972) and Spitzer et al (1979) contain expert opinion and a number of recent hierarchies appear simple. The phases described by Blunt (2015) are not used because it is uncertain that they accurately reflect the development of the hierarchies and this thesis does not consider systems that rank pre-appraised sources of evidence.

This section has discussed the limitations of the systematic review and thematic analysis presented in this chapter. It is recognised that the systematic review may not have identified some relevant hierarchies because of language bias, problems with thesaurus headings and the need to balance sensitivity and specificity within any search strategy. The results of this systematic review appear similar to those presented by Blunt (2015) once systems considering pre-appraised evidence and similar hierarchies are discounted. This suggests that the approach that has been used is robust.

---

[26] Blunt (2015) does acknowledge that the 4 different phases represent trends and there may be some overlap between them.

Subsequent thematic analysis has revealed a number of consistent trends and common arguments across different hierarchies and it is considered unlikely that analysis of further hierarchies would affect these trends or arguments.

3.9. Summary:

The systematic review presented in this chapter has identified 42 different hierarchies of evidence that can be used to determine the effectiveness of treatment interventions. Hierarchies have also been identified that can be used to answer other questions that are of importance within medicine. The first hierarchy appeared in 1972 and they appear to have proliferated since this time. The GRADE Working Group has attempted to develop a universal hierarchy but this attempt appears to have been unsuccessful. The hierarchies provide justification for the knowledge claims made by evidence based medicine and these knowledge claims determine treatment decisions, medicolegal standards and healthcare funding. It is therefore important to understand why there are so many different hierarchies and why attempts to create a universal hierarchy of evidence have been unsuccessful.

When the hierarchies of evidence are analysed collectively some common themes emerge. Randomised controlled trials, systematic reviews and meta-analyses generally provide the highest level and expert opinion generally

provides the lowest level of evidence. However, these are not universal themes as some hierarchies rank randomised controlled trials as the lowest level and one hierarchy ranks expert opinion as the highest level of evidence. This is confusing and should be investigated. Some of the arguments that have been used to support the ranking of randomised controlled trials, systematic reviews and meta-analyses as the highest level of evidence, and expert opinion as the lowest level, have been outlined but these arguments require critical analysis.

The thematic analysis identified a number of common themes amongst the different hierarchies but also identified significant dissimilarities. The number of levels varies from 3 to 12 and the property that is used to rank different study designs is unclear. Properties that the hierarchies claim to use to rank different study designs include truth, bias, validity, confidence and trustworthiness. This raises important questions: What property should be used to rank different study designs within a hierarchy, how many different levels are required and do the hierarchies have utility if the property is unclear? This is another area that requires critical analysis.

The results of the systematic review and thematic analysis presented in this chapter provide a foundation for the analysis of the hierarchies of evidence that is presented in the next 3 chapters. In Chapter 4 the arguments that are used to support the claims that randomised controlled trials, systematic reviews and meta-analyses should provide the highest level, and expert

opinion should provide the lowest level, of evidence are analysed using the method of analytical philosophy. In Chapter 5 the properties that are purportedly used to rank different study designs within hierarchies are analysed. In Chapter 6 factors, independent of study design, that have influenced the development of hierarchies are presented. The critical analysis presented increases understanding of evidence based medicine and seeks to explain some of the variation that is seen amongst them.

Chapter 4: Randomised Controlled Trials, Systematic Reviews and Expert Opinion:

4.1. Introduction:

Evidence based medicine uses hierarchies of evidence to provide justification for knowledge claims that are made. These knowledge claims are important because they dictate which treatment interventions are prescribed and funded, medicolegal standards and the research agenda. It is interesting that there are so many different hierarchies because this suggests that knowledge claims may be justified in different ways. It is therefore important to critically analyse the arguments that are used to support the different hierarchies.

The systematic review presented in the previous chapter identified 42 different hierarchies that can be used to determine the effectiveness of treatment interventions. However, although significant variation was identified amongst these hierarchies, some common themes emerged when they were considered collectively. Randomised controlled trials provide the highest level of evidence in many early hierarchies, systematic reviews and meta-analyses provide the highest level of evidence in many later hierarchies, expert opinion usually provides the lowest level of evidence and observational studies usually provide an intermediate level of evidence. A hierarchy that ranks

randomised controlled trials above observational studies above expert opinion may be considered a standard hierarchy of evidence[27]. A hierarchy that ranks systematic reviews/meta-analyses above randomised controlled trials above observational studies above expert opinion may be considered a modified standard hierarchy of evidence[28].

This chapter, which is divided into three main sections, critically analyses the arguments that have used by evidence based medicine to support the standard hierarchy of evidence and the modified standard hierarchy of evidence. In the first section the design of the randomised controlled trial is explained. Then the two arguments that are used to support the claim that randomised controlled trials should provide the highest level of evidence are analysed: randomised controlled trials control for confounding factors and randomised controlled trials can be blinded. The section concludes by considering other arguments that have not been used by the hierarchies of evidence, but are relevant to the discussion (Worrall 2002, Worrall 2007, Howick 2011, Blunt 2015).

In the second section, the difference between systematic reviews and meta-analyses is explained. The two arguments that have been used to support the claim that systematic reviews or meta-analyses should provide the

---

[27] Examples of standard hierarchies of evidence include those proposed by Cochrane (1972), Spitzer et al (1979), Gross et al (1994) and Hadorn et al (1996).
[28] Examples of modified standard hierarchies of evidence include those proposed by Ball and Phillips (2001), Evans (2003) and the Joanna Briggs Institute (2011).

highest level of evidence are then analysed: systematic reviews and meta-analyses produce results that are closer to the truth and systematic reviews and meta-analyses allow assessment of the consistency of results.

In the third section, confusion surrounding use of the term expert opinion will be highlighted before a working definition of expert opinion is developed. This working definition is then used in the analysis of the five arguments that have been used to support the claim that expert opinion provides the lowest level of evidence: expert opinion does not involve measurement; expert opinion does not consider what happens to patients without treatment; expert opinion may be affected by measurement bias; expert opinion considers small numbers of patients and expert opinion may produce false conclusions[29].

4.2. Interpretation of the Hierarchies of Evidence:

The hierarchies of evidence can be interpreted in a number of different ways (Blunt 2015) and it is important to clarify some points before the arguments that are used to support the standard hierarchy of evidence and modified standard hierarchy of evidence are critically analysed. This will allow the reader to interpret the concepts and arguments that are analysed within the correct context.

---

[29] The reader will note that there is no section considering observational studies despite the inclusion of this study design in most hierarchies. This is because the arguments for the ranking of observational studies as an intermediate level of evidence are indirectly considered in the sections on randomised controlled trials and expert opinion.

Within this chapter a distinction is made between theoretical, practical and ethical considerations. This thesis is primarily concerned with the epistemology of evidence based medicine and critically analyses the *theoretical* arguments that are used to support the different hierarchies. The knowledge claims that are made by evidence based medicine have important implications for patients, medical professionals, healthcare regulators and purchasers of medical care. It is therefore reasonable to expect that these claims should have strong theoretical support. This does not mean that practical and ethical considerations are unimportant but they should not influence the epistemology of evidence based medicine.

Evidence based medicine is characterised by the use of a hierarchy of evidence to guide decision making but it is not committed to including any particular study design or form of evidence within the hierarchy of evidence. Systematic reviews, meta-analyses and expert opinion are included in many hierarchies but they are not universal features. For example, the systems proposed by GRADE (Brozek et al 2009) and the Australian National Health and Medical Research Council (Merlin et al 2009) exclude expert opinion but they are still hierarchies. All hierarchies do include randomised controlled trials but this does not necessarily have to be the case. This does not diminish the importance of the arguments that are analysed in this chapter as this reflects the way that evidence based medicine has historically been interpreted.

Evidence based medicine is not committed to the claim that certain study designs always provide superior or inferior evidence to other study designs within a hierarchy of evidence (La Caze 2009). This interpretation would be too strong and would not reflect later hierarchies, such as the GRADE system that allow upgrading and downgrading of different study designs. However, evidence based medicine cannot claim that the study designs that occupy the higher levels are simply no worse than the study designs that occupy the lower levels. This would be too weak an interpretation as it would imply parity of all evidence and the hierarchy of evidence would dissolve. Evidence based medicine is committed to the claim that certain study designs probably provide superior or inferior evidence to other study designs as this claim is inherent within the concept of a hierarchy. This has been termed the non-categorical interpretation (Turner 2011). Theoretical or empirical support is therefore required for the claim that certain study designs probably provide superior evidence to other study designs and this is the claim that is analysed in this chapter.

Blunt (2015) has suggested that the hierarchies of evidence make no claims about study designs that are not included. Under this interpretation the study design that provides the highest level of evidence may not necessarily provide the best evidence. This is an interesting interpretation but it is not how hierarchies are used within evidence based medicine. The hierarchies are used to provide justification for knowledge claims and they would have

limited utility if they excluded study designs that were considered to provide superior evidence[30].

In this section problems associated with the interpretation of hierarchies of evidence have been outlined. Evidence based medicine is committed to a non-categorical interpretation of hierarchies and should therefore provide theoretical or empirical support for the claim that some study designs probably provide superior evidence to other study designs. Hierarchies are not committed to including any particular study designs but they have historically included randomised controlled trials, systematic reviews and meta-analyses as the highest levels of evidence and expert opinion as the lowest level of evidence. Having clarified the interpretive framework that will be used in this chapter we can now consider the arguments that have been used to support the standard and modified standard hierarchies of evidence.

4.3. Randomised Controlled Trials:

A randomised controlled trial is a prospective study design where participants are randomly allocated to one of two groups[31]. The treatment group receive the treatment that is being tested and the control group receive no treatment, a placebo or standard treatment. The randomisation process is designed to

---

[30] This interpretation does not apply to hierarchies that explicitly exclude certain study designs. For example the GRADE system explicitly excludes systematic reviews and expert opinion.
[31] It is acknowledged that this is a simplistic interpretation of randomised controlled trials and that they can contain more than 2 groups.

produce treatment and control groups with identical characteristics (Hackshaw et al 2006, Gosall and Gosall 2012)[32]. The basic assumption is that if the treatment and control groups are identical, except for the treatment that is being tested, any difference in outcomes at the end of the trial can be attributed to the treatment intervention (Cartwright 2007, Cartwright 2011).

Randomised controlled trials were ranked as the highest level of evidence by over half of the hierarchies that were identified in the previous chapter and this study design provides the highest level of evidence in the standard hierarchy of evidence. The perceived importance of the randomised controlled trials is illustrated by this quote from the *'How to Read Clinical Journals'* series:

> 'Evidence from such a randomised controlled trial is the soundest evidence we can ever obtain about causation (whether it concerns aetiology, therapeutics or any other causal issue)' (Trout 1981 page 986).

The hierarchy proposed by Trout (1981) recognises that other study designs can provide evidence of treatment effectiveness because it includes cohort studies, case-control studies and case series in the lower levels. However, other hierarchies take a more radical view regarding the importance of randomised controlled trials. For example, the second hierarchy proposed by

---

[32] Evidence based medicine also uses randomisation to ensure that study participants are representative of the population from which they are drawn. This use of randomisation is not relevant to the analysis presented in this chapter.

the World Federation of Societies of Biological Psychiatry excluded all study designs that were not randomised controlled trials:

'Studies without randomisation and double-blinding are outdated in psychopharmacology' (Bandelow et al 2008, page 246).

Bandelow et al (2008) excluded non-randomised controlled trials because they felt that including them within their hierarchy only served to increase the importance of 'sub-optimal' study designs.

Only two arguments are used by the hierarchies to support the claim that randomised controlled trials should provide the highest level of evidence: randomised controlled trials balance confounding factors equally between the treatment group and the control group and randomised controlled trials can be blinded. These arguments are directly used to support the claim that randomised controlled trials provide superior evidence to observational studies but they are also used to assert the superiority of the randomised controlled trial study design over expert opinion. Both of these arguments were originally proposed by Cochrane (1972) and are restated by several hierarchies (Chestnut et al 1999, Guyatt et al 2000, Guyatt et al 2004). It is notable that many later hierarchies present no arguments to support the claim that randomised controlled trials should provide the highest level of evidence. This may be because the two arguments outlined above are so

fundamental to evidence based medicine that they have become implicit and unquestioned[33].

4.3.1. Randomised Controlled Trials Balance Confounding Factors:

The basic assumption underpinning the randomised controlled trial is that if the treatment and control groups are identical, except for the treatment that is being tested, any difference in outcomes at the end of the trial can be attributed to the treatment (Cartwright 2011). A confounding factor is unrelated to the treatment that is being tested but associated with the outcome of interest and differentially distributed between the treatment and control groups (Straus et al 2011). Confounding factors may be known or unknown (Worrall 2002). When a confounding factor is present it is not possible to determine whether any difference in outcomes at the end of the trial can be attributed to the treatment that is being tested or the confounding factor.

The importance of confounding factors can be illustrated with an example. Consider a randomised controlled trial undertaken to test whether intervention A was an effective treatment for condition X. All patients in the treatment group are treated with intervention A and all patients in the control group receive no treatment. It is well known that smoking makes condition X

---

[33] Latour and Woolgar (1986) have argued that once a 'fact' is accepted the social circumstances around its construction are forgotten.

worse. If smokers are not equally distributed between the treatment and control groups we will not know if any difference in outcomes can be attributed to intervention A or the differential distribution of smokers between the two groups. In this situation smoking acts as a known confounding factor. Known confounders can be identified by those with relevant background knowledge and can be controlled for by exclusion, matching or the use of appropriate statistical techniques (Straus et al 2011, Gosall and Gosall 2012).

Let us assume that condition X is also made worse by alcohol consumption, but this is not known. If alcohol consumption is not equally distributed between the treatment and control groups any difference in outcomes could be caused by intervention A or the differential distribution of alcohol consumption. In this situation, alcohol consumption acts as an unknown confounding factor. Unknown confounding factors cannot be identified because they are by definition unknown. Randomisation is proposed as a solution to this problem because it purports to balance unknown confounding factors, in this case alcohol consumption, equally between the treatment and control groups (Straus et al 2011).

The argument that is used to rank randomised controlled trials as the highest level of evidence because they control for confounding factors is shown below. This has been considered the most influential argument used by

evidence based medicine to assert the superiority of randomised controlled trials over other study designs (Worrall 2007).

The Confounding Argument:

If a study design controls for confounding factors it provides the best evidence for determining the effectiveness of a treatment intervention

Randomised controlled trials control for confounding factors

Randomised controlled trials provide the best evidence for determining the effectiveness of a treatment intervention

This is a valid argument of the form modus ponens. We therefore need to assess the truth of the premises in order to determine whether the argument is sound. However, before this argument is analysed several points need to be clarified. First, it should be clear from the structure of this argument that we are not assessing the claim that randomised controlled trials cannot provide evidence. It is my contention that all study designs included within hierarchies can provide evidence of treatment effectiveness. We are assessing the claim that randomised controlled trials probably provide better evidence than other study designs because they control for confounding factors[34]. Second, the term 'best evidence' is used in the first premise and

---

[34] This is consistent with a non-categorical interpretation of the hierarchies of evidence (Turner 2011).

the conclusion but the term is not defined. The properties that are used to rank different study designs within hierarchies are discussed in Chapter 5. If a study design provides the best evidence this must surely reflect the property that has been used to rank study designs in a hierarchy. For the purposes of the present discussion we will assume that it is possible to determine 'best evidence' although, as will be demonstrated in the next chapter, this assumption is problematic.

The first premise of the confounding argument states that if a study design controls for confounding factors it provides the best evidence for determining the effectiveness of a treatment intervention. Cartwright (2007) has argued that if all confounding factors are equally distributed between the treatment group and the control group any difference in the outcome of interest must be due to the treatment intervention. In these circumstances we have an ideal randomised controlled trial because, if a number of theoretical assumptions are met, the conclusion is derived using a deductive inference[35].

By definition, confounding factors are associated with the outcome of interest independent of the treatment intervention that is being tested. If confounding factors are not controlled the effect of the treatment intervention on the outcome of interest may be hidden. It is therefore clearly advantageous to control confounding factors in any study design. Whether control of

---

[35] Cartwright (2007) only claims that the conclusions are deductively derived for the group of subjects within that study and does not claim that they can be deductively applied to a different population.

confounding factors provides us with the best evidence will depend upon the property that is used to rank different study designs within a hierarchy. As this discussion has been side-lined until Chapter 5, for the purpose of the present discussion, we will accept that the first premise is true.

The second premise of the confounding argument claims that randomised controlled trials control for confounding factors. Randomisation is considered superior to other strategies that can be used to manage confounding factors, such as exclusion, matching and statistical techniques, because it controls for both known and unknown confounding factors (Straus et al 2011). The earliest hierarchies clearly claimed that randomisation could be used to create identical treatment and control groups. This is illustrated by the following quote from the *'How to Read Clinical Journals'* series:

> '(randomised controlled trials) are investigations in which identical groups of individuals, generated through random allocation, are or are not exposed to the putative causal factor' (Trout 1981 page 988).

If randomisation produced identical treatment and control groups we would have an ideal randomised controlled trial, the second premise would be true and the confounding argument would be sound. In order to determine whether the confounding argument is sound we therefore need to determine whether randomisation balances confounding factors between the treatment group and the control group.

In a randomised controlled trial study participants are randomly allocated to either the treatment group or the control group. Randomisation is a chance event and it is possible that a confounding factor will be equally distributed between the two groups. However, it is more likely that a confounding factor will be differentially distributed between the two groups. When there are multiple confounding factors it is unrealistic to expect randomisation to balance all confounding factors. It is far more likely that multiple confounding factors will be imbalanced between the treatment and control groups (Worrall 2002, Worrall 2007). A look at any published randomised controlled trial will reveal imbalances between the baseline characteristics of the treatment and control groups. If randomisation cannot balance baseline characteristics we should not expect it to balance confounding factors.

A number of different techniques are available to medical researchers, in addition to randomisation, to control known confounding factors (Straus et al 2011). Background knowledge can be used to identify potential confounding factors and then study participants with these confounding factors can either be excluded from the study or matched so they are evenly distributed between the treatment and control groups (Worrall 2002). When known confounding factors are unevenly distributed between the treatment and control groups this can be accounted for during data analysis if the confounding factors have been identified. Exclusion, matching and

appropriate statistical techniques provide greater control over known confounding factors than randomisation which is effectively a chance event.

Some proponents of evidence based medicine accept that non-randomised clinical trials (also known as pseudo-randomised or controlled clinical trials) can balance known confounding factors and provide evidence of equivalent value to randomised controlled trials (Howick 2011). It is considered more problematic to balance confounding factors in observational studies because the researcher does not determine which study participants receive the treatment intervention (Gosall and Gosall 2012). However, this does not mean that confounding factors cannot be balanced in observational studies. Exclusion, matching and statistical techniques can be employed in observational studies if sample sizes are large enough and confounding factors are identified and recorded.

The second premise of the confounding argument claims that randomised controlled trials control for confounding factors but randomisation does not balance known confounding factors and these are surely better controlled by statistical techniques, exclusion and matching. These tools are not restricted to randomised controlled trials and can also be used by controlled clinical trials and observational studies. In order to use these tools known confounding factors must be identified and recorded but this uses background knowledge and is independent of study design. Known confounding factors may be easier to record when a prospective study

design is used but this is a practical consideration and would not differentiate between randomised controlled trials and any observational studies that used a prospective design.

It has also been claimed that randomised controlled trials should provide the highest level of evidence because they balance unknown confounding factors between the treatment and control groups. However, the truth of this claim is difficult to ascertain because unknown confounders are by definition unknown. Empirical evidence could be provided to support this claim if historical randomised controlled trials were found to have balanced previously unknown confounding factors between the treatment and control groups but this evidence is currently unavailable. If it is unlikely that randomisation balances known confounding factors it is also surely unlikely that unknown confounding factors will be balanced between the two groups. If randomisation does not balance known or unknown confounding factors the second premise is false and the confounding argument is unsound.

Howick (2011) has argued that evidence based medicine has never made the claim that randomisation balances known and unknown confounding factors between the treatment and control groups. However, as illustrated by the quote from the *'How to Read Clinical Journals'* series on the previous page, the early hierarchies very clearly made this claim. The following quote from the Evidence Based Medicine Working Group (2000) provides powerful indirect evidence that this idea has persisted:

'Observational studies are inevitably limited by the possibility that apparent differences in treatment groups are really due to differences in patient prognosis in the treatment and control groups' (Guyatt et al 2000, page 1291).

Howick (2011) claimed that randomised controlled trials did not need to rule out all confounding factors to provide superior evidence and only needed to rule out more confounding factors than other study designs[36]. He argued that randomisation prevented selection bias and allocation bias and that this ruled out some confounding factors that were not ruled out by other study designs:

'Randomised trials rule out the potentially confounding influence of self-selection bias and allocation bias while observational studies do not. Hence, randomised trials will usually provide stronger evidence than observational studies' (Howick 2011, page 59).

Allocation bias can occur when researchers are able to control whether study subjects are allocated to either the treatment group or control group. Self-selection bias can occur when study subjects are able to determine whether they are allocated to either the treatment or control group. Both allocation bias and selection bias can purportedly lead to the preferential allocation of study subjects with certain characteristics to a particular study group

---

[36] I would argue that the results of any study design are either confounded or they are not. The claim presented by Howick (2011) relates to the idea that the results of some study designs can be closer to the truth that other study designs. This is discussed in detail in Chapter 5 when the property that is used to rank different study designs within hierarchies is analysed.

(Hackshaw et al 2006). The argument presented by Howick (2011) is therefore subtly different to the earlier arguments for the benefits of randomisation. Randomisation does not control for confounding factors directly it purports to control confounding factors indirectly by controlling allocation bias and selection bias. This is not a new argument (Worrall 2002, Worrall 2007).

The argument proposed by Howick (2011) is associated with a number of problems. First, it assumes that if clinicians or patients are able to determine whether study subjects are placed in the treatment or control group confounding factors will be imbalanced between the two groups. This assumption is widely accepted as true but it is unclear whether it has any theoretical or empirical support. Second, randomisation may prevent researchers or study subjects determining whether study subjects are allocated to the treatment group or control group but it does not necessarily follow that this will balance confounding factors between the two groups. This is no different to the claim that randomisation balances confounding factors between the treatment and control groups. Third, even if allocation bias and selection bias can lead to an imbalance in confounding factors between study groups in interventional studies it does not automatically follow that randomisation is the only way, or even the best way, to prevent this. Surely we just need to prevent researchers and study subjects having any influence over group allocation. The allocation process does not need to be randomised it just needs to be concealed from patients and clinicians. Randomisation does not have any special epistemic property in this sense

and is no better than other methods that can be used to conceal the allocation sequence.

Proponents of evidence based medicine may accept that allocation concealment can be achieved in a number of different ways but argue that randomisation is the simplest way of ensuring that clinicians and patients have no influence over the allocation process. Allocation concealment is only relevant to interventional studies so this argument can only be used to claim that randomised controlled trials are superior to controlled clinical trials. However, although randomisation may be simple this is a practical consideration and it should not influence any theoretical arguments that are used to support hierarchies.

Observational studies cannot employ allocation concealment because the researcher does not determine which subjects receive the treatment intervention. The modified argument presented by Howick (2011) may provide support for the claim that randomised controlled trials provide superior evidence to observational studies if we accept that allocation and selection bias may lead to an imbalance of confounding factors between treatment and control groups and this imbalance can be prevented by randomisation. However, it has already been argued that randomisation does not prevent an imbalance of confounding factors so this argument would only appear to offer weak support if it offers any support at all.

There appears to be limited support for the claim that randomisation balances known and unknown confounding factors between the treatment and control group in a randomised controlled trial. It is therefore interesting to consider why has this idea has persisted. Randomisation could balance confounding factors in a probabilistic sense if cumulative frequencies were reported after an indefinite number of repetitions (Worrall 2007)[37]. However, this is a theoretical interpretation that has no practical use as each participant in the study must be in either the treatment or control group and cannot be spread across the different groups in some probabilistic sense. Repeated randomisation, rather than a single episode of a randomisation, has been suggested as a solution to the problem of confounders but repeated randomisation will not balance confounders. A mistaken probabilistic interpretation of randomisation may explain why the belief that randomisation balances known and unknown confounding factors has persisted.

In this section, the argument that randomised controlled trials provide the best evidence for determining the effectiveness of treatment interventions because they control for confounding factors has been considered. I have argued that randomisation is unlikely to balance known confounding factors and these can be better balanced by matching, exclusion or statistical analysis. These tools can also be used by observational studies. The claim that randomisation balances unknown confounding factors is unlikely and problematic to assess. The claim that randomisation balances confounding factors because it controls selection bias and allocation bias has also been

---

[37] This has been termed the limiting average effect.

analysed. However, it is unclear that randomisation prevents any imbalance of confounders that may occur when study participants and researchers are able to determine whether they are entered into the treatment group or control group. I would therefore argue that the confounding argument offers little support for the claim that randomised controlled trials should provide the highest level in the standard hierarchy of evidence.

4.3.2. Randomised Controlled Trials Can Be Blinded:

The second argument that is used to support the claim that randomised controlled trials should provide the highest level of evidence is that randomised controlled trials can be blinded. This argument was originally presented by Cochrane (1972) and has been restated in support of a number of later hierarchies (Sackett 1989, Guyatt et al 1995, Guyatt et al 2004).

In a clinical trial, study participants are allocated to either the treatment group or the control group. Blinding ensures that individuals involved with the clinical trial are unaware whether study participants are in the treatment group or the control group after they have been allocated. Seven groups of individuals can potentially be blinded in any clinical trial: patients, clinicians, data collectors, outcome assessors, data analysts, trial monitoring committee members and manuscript writers (Haynes 2006) although none of the hierarchies specify which of these groups of individuals need to be blinded

for blinding to be effective. Blinding is considered important because if individuals involved with a study are aware whether a study participant is in the treatment or control group this may influence the way that treatment is provided, outcomes are measured or the study is reported. Blinding is considered particularly important to prevent performance bias and measurement bias (Higgins and Green 2011). The blinding argument is presented below.

The Blinding Argument:

If a study design can be blinded it provides the best evidence for determining the effectiveness of a treatment intervention

Randomised controlled trials can be blinded

Randomised controlled trials provide the best evidence for determining the effectiveness of a treatment intervention

The blinding argument is a valid argument of the form modus ponens. We therefore need to assess the truth of the premises to determine whether the argument is sound. The second premise of the blinding argument claims that randomised controlled trials can be blinded. A randomised controlled trial can only be blinded if the interventions given to the treatment group and the control group are indistinguishable. If the interventions can be distinguished

blinding is broken. Blinding may be broken because the interventions are different, require different aftercare or have different adverse effects. Proponents of evidence based medicine acknowledge that blinding of randomised controlled trials is difficult in many trials (Haynes et al 2006). However, unless the treatment and control interventions are indistinguishable blinding is not difficult, it is impossible. Consider a randomised controlled trial where the treatment group receive a new surgical procedure. Even if the control group receive mock surgery, at some point, blinding must be broken so that the surgeon can perform the correct procedure[38].

Randomised controlled trials that investigate pharmacological interventions can potentially be blinded when the treatment and control interventions appear identical. However, even within this subset of randomised controlled trials, blinding is broken where the treatment drug and control drug have different side effects. The claim that all randomised control trials can be blinded is false. If this claim is restricted to randomised control trials where the interventions given to the treatment group and control group are indistinguishable the claim may be true. However, this restricts the claim to a small subset of randomised controlled trials and significantly weakens the argument.

The first premise of the blinding argument claims that if a study design can be blinded it provides the best evidence for determining the effectiveness of a

---

[38] It is beyond the scope of this thesis to consider the ethics of mock surgical procedures.

treatment intervention. Randomisation is not a necessary condition for a study to be blinded. Blinding simply means that individuals involved in a study are unaware whether study participants are in the treatment group or control group. As long as the method that is used to allocate patients to each group is concealed, blinding can be achieved. This is termed allocation concealment. For example, study participants could be allocated to the treatment group or control group based upon their date of birth, shoe size or hair colour. As long as none of the individuals involved with the study were aware of the allocation process that was used blinding could be achieved. This means that any controlled clinical trial can potentially be blinded. It may be argued that randomisation is the simplest method of generating a concealable allocation sequence but this is a practical consideration and should not influence any theoretical arguments underpinning the standard hierarchy of evidence.

Any study design where the investigator is able to control the treatment intervention that study participants receive has the potential to be blinded. Prospective case series are often considered to be observational studies but this study design has the potential to be interventional. A prospective case series could be envisaged where study participants were given a new drug but none of the individuals involved in the trial were informed. This study design could be blinded and, using the blinding argument, would be ranked as the highest level of evidence. It could be argued that a prospective case series conducted in this way was unethical. However, this is an ethical

consideration and we are considering the theoretical basis of the hierarchy of evidence.

The blinding argument is usually used to support the claim that randomised controlled trials provide superior evidence to observational studies. In an observational study the researcher does not allocate study participants to the treatment group or the control group and outcomes are simply observed. Observational studies cannot be blinded because the patient and clinician are already aware which group the study participant has entered. This leads us to an interesting question: should observational studies be ranked as a lower level of evidence than interventional studies because they cannot be blinded?

One reason that blinding is considered important is because it prevents knowledge of whether the patient has been allocated to the treatment group or the control group influencing the way that outcomes are measured. It is reasoned that if the outcome assessor is aware that the study participant is in the treatment group they may report outcomes more favourably. This is termed measurement bias and may be either conscious or unconscious (Haynes et al 2006, Straus et al 2011, Higgins and Green 2011). However, measurement bias is only problematic when the outcome assessor is able to interpret the outcome in different ways. Measurement bias is not a problem when objective outcome measures, such as tooth loss or death, are used. In these situations it does not matter if the outcome assessor has been blinded

as the tooth is either present or absent or the patient is alive or dead. This was recognised by Cochrane (1972):

> 'The best index in these sorts of comparisons is the fact of death where there is little possibility of bias due to observer difference' (Cochrane, 1972, page 20).

It is true that observational studies cannot be blinded but this is not a problem if objective outcome measures are used because these studies are not susceptible to measurement bias. Observational studies that do use subjective outcome measures are susceptible to measurement bias and would benefit from blinding. However, this does not mean that all observational studies need to be blinded to prevent measurement bias.

Blinding is also considered important because it prevents performance bias (Higgins and Green 2011). However, if observational studies are conducted according to strict protocols, and any deviations from these protocols are recorded and accounted for during statistical analysis of the results, performance bias should not be problematic. We trust researchers to observe best practice when they undertake randomised controlled trials so why should we not trust them to follow treatment protocols in observational studies that cannot be blinded?

Howick (2011) considered the importance of blinding in *'The Philosophy of Evidence Based Medicine'*. He was concerned that the effectiveness of many accepted interventions, for example adrenalin for the management of anaphylaxis, had never been confirmed by blinded studies. He termed this 'Phillip's paradox'. Howick (2011) resolved this paradox by claiming that dramatic treatment effects could override the importance of blinding. This provides further evidence that blinding is unnecessary to determine the effectiveness of treatment interventions and undermines the first premise of the blinding argument.

In this section, the argument that randomised controlled trials provide the best evidence for determining the effectiveness of treatment interventions because they can be blinded has been considered. I have argued that randomised controlled trials can only be blinded when the treatment and control interventions are indistinguishable and shown that the blinding argument does not separate randomised controlled trials from controlled clinical trials. Observational studies cannot be blinded but this is not always problematic as measurement bias, can be controlled through the use of objective outcome measures. The argument that all randomised controlled trials should provide the highest level of evidence because they can be blinded is therefore unsound. This argument could be modified. It could be restricted to the small subset of randomised controlled trials that use subjective outcome measures, where the treatment and control interventions are indistinguishable. However, this would significantly restrict the scope of

the blinding argument and would not support the standard hierarchy of evidence.

4.3.3. Further Arguments for Randomised Controlled Trials:

Two other arguments have been used to support the claim that randomised controlled trials should provide the highest level of evidence for determining the effectiveness of treatment interventions (Worrall 2002, Worrall 2007). Interestingly, neither of these arguments is directly used by any of the hierarchies that were identified by the systematic review presented in Chapter 3.

The first of these arguments claims that randomisation is necessary to guarantee the validity of classical significance testing (Worrall 2007). Classical significance testing is used to determine the probability that the observed difference in outcomes between the treatment and control groups would be observed by chance if the null hypothesis, that there is no difference between the treatment intervention and control intervention, was correct. Classical significance testing assumes that the treatment and control groups are identical. However, the treatment and control groups can only be identical if known and unknown confounding factors are equally balanced between the two groups. This argument is actually the confounding argument in a different guise and has already been considered.

The second argument claims that it is an empirical fact that observational studies overestimate the effectiveness of treatment interventions when compared with randomised controlled trials (Worrall 2002, Worrall 2007, Howick 2011). A number of historical examples are used to support this claim[39]. In each of these examples treatment interventions supported by observational studies are subsequently shown to be ineffective or harmful once randomised controlled trials are conducted.

The results of randomised controlled trials and observational studies conducted to answer the same question can vary substantially. However, the results of randomised controlled trials and observational studies conducted to answer the same question can also agree and the results of different randomised controlled trials conducted to answer the same question can vary substantially. The variability amongst the results of randomised controlled trials is the rationale for meta-analyses and is exemplified by the logo of the Cochrane Collaboration.

When observational studies and randomised controlled trials investigating the effectiveness of the same treatment intervention produce different results it does not necessarily follow that the observational studies have overestimated the treatment effect (La Caze 2009). It is also possible that the randomised

---

[39] Commonly used examples include the use of hormone replacement therapy in postmenopausal women to reduce the risk of adverse cardiovascular events and the use of the antiarrhythmic drugs encainide and flecainide following myocardial infarction (Guyatt et al 2008, Howick 2011).

controlled trials have underestimated the treatment effect or that neither study design provides a true estimate of treatment effect. In order to decide which of these interpretations is correct we need a standard that we can compare the results of the different studies against. In the standard hierarchy of evidence the randomised controlled trial provides this standard. However, the results of randomised controlled trials cannot be used to support the claim that observational studies overestimate the effectiveness of treatment interventions relative to randomised controlled trials because this would be a circular argument. This has been termed the calibration problem (Blunt 2015).

Blunt (2015) has argued that a number of different studies with consistent results could be used to provide an external standard[40]. However, if this evidence is available surely it should provide the highest level of evidence. Even if an external standard was available, and this could be used to show that observational studies had overestimated treatment effects in the historical examples highlighted in the evidence based medicine literature, this would not provide strong support for the argument that observational studies overestimated treatment effects in general. This is because there would still be a number of alternative explanations for the historical examples that are used. For example, the observational studies may have been poorly conducted, the study participants may not have been representative of the population or the data may have been misreported. It has been

---

[40] Hierarchies that include systematic reviews as the highest level of evidence are considered in the next section.

acknowledged that observational studies can produce similar results to randomised controlled trials when they are well conducted (Worrall 2007). If this claim is true the historical examples used by proponents of evidence based medicine carry little weight.

If the argument that observational studies overestimate treatment effects relative to randomised controlled trials is accepted, we should have less confidence in the results of an observational study, than a randomised controlled trial, that shows a positive treatment effect. We may have more confidence that the treatment intervention, investigated in the observational study, has some positive effect, if there is a large treatment effect, and this is recognised by the GRADE hierarchy of evidence (Guyatt et al 2008). However, this argument is only applied to studies that produce positive results. When randomised controlled trials and observational studies both show that a treatment intervention is not effective we have no reason to be more confident in the results of the randomised controlled trial. This is a paradox that is not explained by the hierarchies (Blunt 2015).

In this subsection we have considered two further arguments that have been advanced to support the claim that randomised controlled trials provide the highest level of evidence. The first argument claims that randomisation is necessary for classical significance testing. However, this is a restatement of the confounding argument and this argument has already been considered. The second argument claims that observational studies overestimate the

effectiveness of treatment interventions. However, this argument is problematic because it is a circular argument, there are alternative explanations for the historical examples that provide empirical support and it does not apply to observational studies that show a treatment intervention is ineffective.

4.3.4. Summary:

In this section we have considered the different arguments that have been used to rank randomised controlled trials as the highest level of evidence when determining the effectiveness of treatment interventions. The confounding argument is problematic because randomisation is unlikely to balance known confounding factors and this is better achieved by tools that are available to observational studies. There is also no theoretical or empirical evidence to support the claim that randomisation balances unknown confounding factors. The arguments that randomisation is necessary for the validity of classical significance testing and can prevent an imbalance in confounding factors by controlling selection bias and allocation bias are effectively the confounding argument in different guises. These arguments provide no further support for the claim that randomised controlled trials should provide the highest level of evidence.

The blinding argument offers no support for the claim that <u>all</u> randomised controlled trials should provide the highest level of evidence because many randomised controlled trials cannot be blinded. Blinding is primarily designed to prevent performance and measurement bias but these types of bias may be unproblematic in observational studies when treatment protocols are followed and objective outcome measures are used. The blinding argument may be sound if it is limited to randomised controlled trials where the treatment intervention and control intervention are indistinguishable but this severely restricts the scope of the argument and does not support the standard hierarchy of evidence.

It has also been argued that randomised controlled trials should provide the highest level of evidence because observational studies overestimate the effectiveness of treatment interventions. However, this requires an external standard of reference to avoid criticisms of circularity and this is not available. The results of randomised controlled trials and observational studies do sometimes disagree but this does not necessarily reflect a fundamental problem with the design of observational studies. The results of different randomised controlled trials can also disagree and this is the rationale for meta-analyses.

The different arguments that have been analysed in this section provide little support for the claim that randomised controlled trials should provide the highest level in hierarchies of evidence. This may explain why systematic

reviews and meta-analyses provide the highest level of evidence in many later hierarchies. The arguments that have been used to rank systematic reviews and meta-analyses atop the modified standard hierarchy of evidence are considered in the next section.

4.4. Systematic Reviews and Meta-Analyses:

The modified standard hierarchy of evidence ranks systematic reviews and/or meta-analyses above randomised controlled trials above observational studies above expert opinion. The American College of Chest Physicians (Cook et al 1992) were the first group to rank meta-analyses as the highest level of evidence. This group modified their original hierarchy because they recognised that meta-analyses, termed scientific overviews[41], provided superior evidence to randomised controlled trials:

> 'The (old) Levels of Evidence were developed prior to the recognition of the power of rigorous scientific overview'. (Cook et al 1992 page 305S-306S)

Systematic reviews first appeared as the highest level of evidence in 1995 in the hierarchy used by the Australian National Health and Medical Research Council (Gugiu and Gugiu 2010). Since this time meta-analyses and systematic reviews have increasingly appeared as the highest level of evidence in different hierarchies (Appendix 3).

---

[41] This provides another example of proponents of evidence based medicine using words with powerful perlocutionary force.

Different hierarchies rank meta-analyses (Cook et al 1992, Stetler et al 1998), systematic reviews (Ball and Phillips 2001, Merlin et al 2009) and both systematic reviews and meta-analyses (Meltzer et al 1998, Harbour and Miller 2001) as the highest level of evidence. Other hierarchies purposefully exclude systematic reviews and meta-analyses (Guyatt el al 1995, Atkins et al 2004). It is therefore important to clarify exactly what systematic reviews and meta-analyses are before analysing the arguments that are used to support the claim that they should provide the highest level of evidence.

A systematic review is a process that is used to identify, appraise and summarise individual studies addressing a focussed question in a structured way (Pope et al 2007, Guyatt et al 2008, Straus et al 2011). A meta-analysis is a statistical technique for quantitatively combining the results of individual studies measuring the same outcome to generate a pooled estimate for that outcome (Pope et al 2007, Guyatt et al 2008). A meta-analysis may be undertaken as part of a systematic review if the individual studies identified by the systematic review are considered sufficiently similar (Guyatt et al 2008).

Systematic reviews encompass a wide range of different processes including thematic analyses and meta-analyses (Pope et al 2007, Hammersley 2013). Systematic reviews are not restricted to randomised controlled trials and can be used to summarise the results of other study designs (Stroup et al 2000).

The technique of meta-analyses can also be used to generate a pooled estimate for a treatment outcome even when individual studies have been identified using a non-systematic process. However, when the hierarchies refer to systematic reviews it is explicitly clear that they refer to a process that can be used to identify, appraise and summarise the results of randomised controlled trials and that these systematic reviews should use meta-analysis when the randomised controlled trials are considered sufficiently similar. The hierarchies therefore consider meta-analyses to be a subset of systematic reviews of randomised controlled trials.

Some hierarchies (Guyatt el al 1995, Guyatt et al 1998, Atkins et al 2004) purposefully exclude systematic reviews. This is not because these hierarchies consider systematic reviews to be unimportant. On the contrary, the systematic review is integral to the way that these hierarchies are used. These hierarchies exclude systematic reviews so that users are able to control the randomised controlled trials that are included, appraised and summarised. Systematic reviews are effectively excluded because they can be poorly conducted. However, this position is paradoxical because any review process or study design can be poorly conducted. These hierarchies all include randomised controlled trials and observational studies despite the fact that these study designs can also be poorly conducted. Proponents of evidence based medicine may argue that it is more difficult to assess whether a systematic review has been well conducted than a randomised controlled trial. However, this is a practical consideration that is not relevant when we consider the theoretical support for hierarchies of evidence.

The systematic reviews valued by the hierarchies summarise the results of randomised controlled trials and undertake meta-analysis when the randomised controlled trials are sufficiently similar. A systematic review may identify no randomised controlled trials, one randomised controlled trial or multiple randomised controlled trials. A systematic review that identifies no randomised controlled trials provides no evidence although it may identify an area where research is required. A systematic review that identifies one randomised controlled trial provides no greater support for a treatment intervention than that one trial although it may appear to do so if ranked as a systematic review. Therefore, when we consider the claim that systematic reviews provide the highest level of evidence, we need to restrict the claim to systematic reviews that summarise the results of at least two randomised controlled trials.

When the hierarchies of evidence are considered collectively only two arguments are presented to support the claim that systematic reviews and/or meta-analyses should provide the highest level of evidence. The first argument claims that meta-analyses produce an estimate of treatment effect that is closer to the truth that the results of individual randomised controlled trials. The second argument claims that systematic reviews allow assessment of the consistency of treatment effect. These arguments are considered below.

4.4.1. Meta-analyses produce Results that are closer to the Truth:

The first argument that is used to support the claim that systematic reviews and/or meta-analyses should provide the highest level of evidence is that meta-analyses produce results that are closer to the truth than individual randomised controlled trials. This is illustrated by the following quote from the American College of Chest Physicians:

> 'The most compelling rationale for meta-analysis is its ability to generate more precise estimates of the true treatment effect' (Cook et al 1992 page 307S)

The truth argument is summarised below:

The truth argument:

If a process or study design produces an estimate of treatment effect that is closer to the truth that other processes or study designs it provides the best evidence for determining the effectiveness of a treatment intervention

Meta-analysis produces an estimate of treatment effect that is closer to the truth that other processes or study designs

Meta-analyses provide the best evidence for determining the effectiveness of a treatment intervention

This is a valid argument of the form modus ponens. In order to determine if the truth argument is sound we therefore need to assess the truth of the premises. However, before the truth argument is analysed in detail it is important to stress that this argument is restricted to meta-analyses and cannot be used to support the wider claim that systematic reviews should provide the highest level of evidence. This is because although systematic reviews summarise component studies only meta-analyses provide a pooled estimate of treatment effect. The truth argument only provides support for the claim that a small subset of systematic reviews should provide the highest level of evidence.

The truth argument is immediately problematic because it uses the concept of truth. Although a number of different theories of truth have been proposed within philosophy they all agree that propositions are either true or false (Simmons 2005)[42]. It is therefore unclear how the conclusions of any process or study design can take us closer to the truth. The conclusions of randomised controlled trials and meta-analyses are often expressed as numerical values which may produce the illusion that you are closer or further away from the true value but each conclusion can only ever be true or false. Unless proponents of evidence based medicine are able to articulate a theory of truth that allows conclusions to be closer to the truth the first premise of the truth argument is false and the argument is unsound.

---

[42] It is generally accepted that evidence based medicine uses a correspondence theory of truth.

An underlying principle of analytical philosophy is that any arguments should be interpreted using the principle of charity. We will therefore assume that the conclusions of different processes and study designs can be closer to the truth and that conclusions that are closer to the truth provide superior evidence. If these assumptions are accepted we then need to consider any empirical or theoretical support for the second premise that meta-analyses produce conclusions that are closer to the truth that randomised controlled trials.

Meta-analysis presumes that if we combine the results of several randomised controlled trials the pooled estimate of treatment effect will be closer to the true treatment effect. This is an inductive argument and does not provide strong support for the claim because it is not necessarily true. Empirical support for the claim could be provided if there was a standard that we could compare the results of meta-analyses against. However, other meta-analyses cannot provide this standard as this would be a circular argument. Another study design could potentially provide an external standard but this study design would then surely provide the highest level of evidence.

The inductive argument that is used to support the claim that meta-analyses provide the highest level of evidence is associated with additional problems. First, if the total number of observations is the most important factor then randomised controlled trials with large numbers of participants should be

ranked as the highest level of evidence alongside meta-analyses. This approach was adopted by Hadorn et al (1996) but this is unusual amongst hierarchies. Second, meta-analyses only quantitatively combine the results of randomised controlled trials but, as demonstrated in the previous section, there is little empirical of theoretical evidence to support the claim that randomised controlled trials provide superior evidence to observational studies. It is therefore not clear that we will get closer to the truth if we quantitatively combine randomised controlled trials and not observational studies.

In this subsection we have considered the argument that meta-analyses provide the highest level of evidence because they provide results that are closer to the truth. This argument is restricted to meta-analyses and uses one particular concept of truth. Even when the argument is interpreted charitably it derives no empirical support from an external standard and relies on inductive reasoning. Furthermore, the claim that randomised controlled trials provide superior evidence to other study designs underlies the truth argument and, as argued in the previous section, this claim has limited support. I would therefore argue that the truth argument is unsound.

4.4.2. Systematic Reviews Allow Assessment of Consistency of Treatment Effect:

The second argument that is used to support the claim that systematic reviews should provide the highest level of evidence is that systematic reviews can be used to assessment the consistency of treatment effect. Consideration of consistency of treatment effect was clearly important to the Evidence Based Medicine Working Group when they developed their first hierarchy:

'When differences in treatment effect across studies are greater than one would expect by chance alone, and varying populations, interventions, outcomes, or study methods cannot explain the differences, inferences become weaker' (Guyatt et al 1995 page 1800)

Consideration of consistency of treatment effects is important because treatment interventions usually produce variable effects in study participants. The results of randomised controlled trials and observational studies are presented as average treatment effects but this hides the fact that many patients do better or worse than average (Blunt 2015). When different randomised controlled trials or observational studies are undertaken investigating the same treatment intervention each study will usually produce a different average treatment effect. This inconsistency in treatment effect, both within and between studies, can be explained by a number of factors including differences in the study participants, clinicians, study designs and

outcome measures. Consistency in treatment effect is important from a clinical perspective because clinicians need to be confident that treatment will benefit patients, particularly when treatments have potential adverse effects. The consistency argument is summarised below:

The Consistency Argument

If a process or study design allows assessment of the consistency of treatment effects it provides the best evidence for the effectiveness of a treatment intervention

Systematic reviews allow assessment of the consistency of treatment effects

Systematic reviews provide the best evidence for determining the effectiveness of a treatment intervention.

This is a valid argument of the form modus ponens. We therefore need to analyse the truth of the premises to determine if the argument is sound. The second premise claims systematic reviews can be used to show that a treatment intervention has inconsistent treatment effects across different studies. However, what does it mean for treatment effects to be inconsistent? Unless the different studies produce identical treatment effects the results are inconsistent. Some hierarchies provide rules to guide interpretation of inconsistency but they provide no arguments or empirical evidence to support

these rules. For example, the Evidence Based Medicine Working Group claimed that inconsistent treatment effects became problematic when the point estimates of treatment effect in the two most disparate randomised controlled trials differed by greater than 20% (Guyatt et al 1995). Most hierarchies leave the interpretation of inconsistent treatment effects to the reviewer:

> 'By homogeneity we mean a systematic review that is free from worrisome variations in the direction and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant'. (Ball and Phillips, 2001, page 7)

The hierarchies appear to value systematic reviews because they allow the consistency of treatment effects to be assessed but these systematic reviews use arbitrary rules and opinion to determine when treatment effects are consistent. This makes it difficult to assess the second premise but we will assess it charitably and accept that systematic reviews can assess consistency of treatment effect.

The first premise of the consistency argument claims that we need to be able to assess the consistency of treatment effects in order to determine whether a treatment intervention is effective. Assessment of consistency is important but I would argue that explanation of any inconsistency is more important. When a treatment intervention produces inconsistent treatment effects, but

this inconsistency can be explained, the inconsistency is not troublesome and the treatment intervention can be used with confidence. This is implicit in the quote from Ball and Phillips (2001) on the previous page.

Systematic reviews may be able to identify inconsistency in treatment effects but they do not explain that inconsistency. The results of systematic reviews have to be interpreted and expert opinion, biological plausibility, pathophysiological rationale and findings from randomised controlled trials and observational studies are often used to explain any inconsistency. The explanation of any inconsistency is surely more important that the initial discovery of that inconsistency. It is therefore unclear why systematic reviews should be ranked as the highest level of evidence simply because they may identify inconsistent treatment effects. I would argue that the first premise of the consistency argument is false because identification of inconsistency is clearly less important than explanation of inconsistent treatment effects.

In order to assess the consistency of any treatment effect we need to assess treatment outcomes in large groups of study participants. However, systematic reviews are not the only processes or study designs that can provide this information. Narrative reviews are often considered inferior to systematic reviews but narrative reviews can nonetheless be rigorous. Indeed methodologies for narrative reviews have advanced significantly in the last ten years (Hammersley 2013). A single randomised controlled trial or observational study, involving a large number of subjects can also provide

this information. Proponents of evidence based medicine may argue that large studies are difficult to undertake (Howick 2011) but this is a practical consideration that should not influence the theory underpinning hierarchies. This has important implications because even if the consistency argument is sound it does not differentiate systematic reviews from narrative reviews and large primary studies.

The systematic reviews valued by the hierarchies of evidence only consider the results of randomised controlled trials. However, systematic reviews that are not restricted to randomised controlled trials can identify greater numbers of studies (Stroup et al 2000). This can provide more information about the consistency of treatment effects. If a systematic review that includes both randomised controlled trials and observational studies shows a consistent treatment effect this surely carries greater evidential weight than a systematic review based upon randomised controlled trials alone. The systematic reviews that are valued by the hierarchies of evidence may therefore be inferior to other types of systematic reviews at assessing consistency of treatment effect. If this is the case they should not be ranked as the highest level of evidence.

In this subsection we have considered the claim that systematic reviews should provide the highest level of evidence because they allow assessment of consistency of treatment effect. I have argued that the inconsistency argument is unsound because explanation of inconsistency is more important

than identification of inconsistency. The inconsistency argument is also problematic because it does not differentiate systematic reviews from narrative reviews and large primary studies. It therefore cannot be used to claim that systematic reviews are probably superior to these study designs.

4.4.3. Summary:

In this section the two arguments that have been used to rank systematic reviews and meta-analyses as the highest level of evidence have been considered: the truth argument and the consistency argument. The truth argument can only be applied to meta-analyses, uses one particular concept of truth and is underpinned by the problematic claim that randomised controlled trials provide superior evidence to observational study designs. The consistency argument is not restricted to systematic reviews and can be used to support the wider claim that narrative reviews and studies with large numbers of participants should provide the highest level of evidence. I have also argued that the consistency argument is unsound because explanation of inconsistent treatment effects is more important than identification of inconsistent effects. The claim that systematic reviews and/or meta-analyses should provide the highest level of evidence would therefore appear to have little theoretical support. In the next section we will move to the bottom of the hierarchies and consider the arguments that have been used to rank expert opinion as the lowest level of evidence.

<u>4.5. Expert Opinion</u>

When the hierarchies of evidence are considered collectively expert opinion is most commonly ranked as the lowest level of evidence. However, this is not a universal finding and many hierarchies either exclude expert opinion or rank a different form of evidence as the lowest level of evidence. For example, both of the hierarchies produced by the American Association of Critical Care Nurses (Armola et al 2009) rank manufacturer's recommendations below expert opinion. Other hierarchies attribute greater evidential value to expert opinion. For example, the Oncology Nursing Society hierarchy (Mitchell and Friese 2007) allows expert opinion to reach the highest level of evidence in some circumstances. In this section we will analyse the arguments that have been used to rank expert opinion as the lowest level of evidence in hierarchies.

Five arguments have been used to support the claim that expert opinion should provide the lowest level of evidence: expert opinion does not involve measurement; expert opinion is susceptible to measurement bias; expert opinion does not consider what happens to patients without treatment; expert opinion does not consider large numbers of patients and expert opinion may be incorrect. These arguments repeatedly appear in the evidence based medicine literature although no hierarchy uses all of the arguments and many hierarchies provide no arguments to justify the ranking of expert opinion as the lowest level of evidence. Each of the arguments can be identified in the

following quotes from *'Effectiveness and Efficiency'* (Cochrane 1972) and the Evidence Based Medicine Working Group (1995):

'The oldest, and probably still the commonest form of evidence proffered, is clinical opinion. This varies in value with the ability of the clinician and the width of his experience, but its value must be rated low, because there is no quantitative measurement, no attempt to discover what would happen if the patients had no treatment, and every possibility of bias affecting the assessment of the result.' (Cochrane, 1972, page 20).

'The unsystematic observations of the individual clinician constitute one source of evidence, and physiological experiments another. Unsystematic clinical observations are limited by small sample size, and more importantly, by limitations in human processes of making inferences. Predictions about intervention effects on clinically important outcomes from physiological experiments are usually right, but occasionally disastrously wrong' (Guyatt et al, 2000, page 1291)

The ranking of expert opinion as the lowest level of evidence has been considered confusing. Worrall (2007) could not understand why expert opinion was ranked below randomised controlled trials when randomised controlled trials had no place in established sciences such as physics. Thompson (2010) argued that explanatory theory, which would be categorised as expert opinion within most hierarchies, should provide the highest, not the lowest, level of evidence. It is therefore important to critically

analyse the arguments that have been used to rank expert opinion as the lowest level of evidence.

4.5.1. Defining Expert Opinion:

Before we can analyse the arguments that are used to rank expert opinion as the lowest level of evidence we need to clarify exactly what expert opinion is. This is surprisingly difficult because although hierarchies commonly include expert opinion few of them provide a definition of this term. The picture is further complicated by the use of terms such as clinical opinion (Cochrane 1972), clinical experience (Chestnut et al 1999), mechanism based reasoning (Howick et al 2011), rational conjecture and common sense (American Heart Association 2000) which may, or may not, be synonymous with expert opinion. For example, Cochrane (1972) ranks clinical opinion as the lowest level of evidence but does not define the term. Spitzer et al (1979), Harbour and Miller (2001) and Evans (2003) rank expert opinion as the lowest level of evidence but do not define the term and Chestnut et al (1999) differentiates between expert opinion and clinical experience, suggesting that they are different concepts, but neither term is defined and the distinction does not influence hierarchal position.

Some hierarchies are restricted to different study designs and exclude expert opinion altogether (Trout 1981, Sackett 1989, Cook et al 1992, Guyatt et al

1995, Brozek et al 2009). The GRADE system explicitly excludes expert opinion because expert opinion is considered integral to the interpretation of all evidence, including the results of randomised controlled trials and observational studies (Brozek et al 2009). None of the other hierarchies that exclude expert opinion provide an explanation for this decision.

When the hierarchies do provide a definition for expert opinion they often imply that expert opinion is unsystematic, uncritical and anecdotal. For example, the Evidence Based Medicine Working Group equate expert opinion with:

'Unsystematic observations from clinical experience' (Evidence Based Medicine Working Group, 1992, page 2421 and Guyatt et al 2000)

The Oxford Centre for Evidence Based Medicine hierarchy specifies that expert opinion has not undergone explicit critical appraisal (Ball and Phillips 2001) and Djulbegovic and Hadley (1998) equate expert opinion with anecdotal data. If expert opinion is unsystematic, uncritical and anecdotal it may deserve to be ranked as the lowest level of evidence or even excluded from hierarchies altogether.

Not all hierarchies characterise expert opinion as unsystematic, uncritical and anecdotal. The Oncology Nursing Society (Mitchell and Friese 2007) are explicit that expert opinion is produced using a systematic process. Their

hierarchy allows expert opinion to be ranked as the highest level of evidence. They define expert opinion in the following way:

> 'Recommendations from a panel of experts that derive from an explicit literature search strategy, and include thorough analysis, quality rating and synthesis of evidence' (Mitchell and Friese 2007)

Some hierarchies would categorise evidence produced in this way as a systematic review. However, the Oncology Nursing Society defines a systematic review as a process that collates information from research studies.

Some hierarchies differentiate between different types of expert opinion: Cochrane (1972) claimed that the value of expert opinion depended upon the ability and experience of the expert offering the opinion; the John Hopkins Evidence Rating Scale differentiated between the opinion of a group of experts and individual expert opinion (Newhouse et al 2007) and Braunwald et al (1994) excluded individual expert opinion and ranked expert consensus as the lowest level of evidence. These hierarchies provide no arguments to support the claim that the opinion of a group of experts should carry greater evidential weight than the opinion of an individual expert.

Expert opinion is included in many of the hierarchies although the term is often not defined. When expert opinion is defined a number of different definitions are provided. Many of these definitions have negative

perlocutionary force as they equate expert opinion with an unsystematic or uncritical approach. It has been recognised that expert opinion may vary in quality but this is not generally acknowledged by most of the hierarchies. Overall the hierarchies of evidence paint a confusing picture about expert opinion. This complicates analysis of the arguments. However, in order to progress with our analysis we need to develop a working definition of expert opinion.

The standard hierarchy of evidence contains randomised controlled trials, observational studies and expert opinion. A clear distinction is made between interventional and observational study designs and expert opinion. This allows us to define expert opinion in a negative sense: expert opinion is what remains once evidence from different study designs has been excluded. This definition is consistent with the way that the hierarchies are designed to be used. When clinicians are faced with a clinical problem, evidence based medicine instructs them to search the medical literature in a structured way for a solution. Expert opinion is only used if there are no interventional or observational studies to guide them (Guyatt et al 2000, Straus et al 2011).

Cochrane (1972) was concerned that expert opinion could vary in quality but this is true of any study design that can be included in a hierarchy. Systematic reviews, randomised controlled trials and observational studies can all vary in quality if they are poorly conducted. In this chapter we are analysing the theoretical basis of the hierarchy of evidence. We therefore

have a duty to interpret expert opinion in the most charitable light. Expert opinion does not have to be unsystematic, uncritical or anecdotal and when expert opinion is interpreted in this way it is easily denigrated. I would argue that there is a significant difference between the gut feeling of an expert based upon no evidence and the consensus opinion of a group of experts following detailed consideration of an evidence base. The characterisation of expert opinion as unsystematic and uncritical is unhelpful and may explain why this form of evidence is not valued by many hierarchies.

Fricker (2005) has argued that any expert opinion must be sincere and competent. Previous track record, credentials, reputation and possible distorting factors can provide a guide to sincerity and competence (Goldman 2001, Fricker 2002). However, these factors are generally ignored by the hierarchies. Manufacturer's recommendations provide the lowest level of evidence in the hierarchies developed by the American Association of Critical Care Nurses (Armola et al 2009) because they may be susceptible to distorting factors but no other hierarchies consider this factor. The *'How to Read Clinical Journals'* series did advocate consideration of the track record of study authors but their hierarchy did not include expert opinion (Trout 1981). Where the hierarchies make a distinction between individual expert opinion and expert group consensus this may reflect perceived sincerity and competence.

Expert opinion has been defined in a number of different ways by the hierarchies of evidence and this complicates our understanding of the concept. A number of the definitions characterise expert opinion in a negative way and this may have reinforced negative perceptions of this type of evidence. In order to proceed with this analysis expert opinion has been interpreted charitably and a working definition has been developed. Expert opinion is the evidence that remains once evidence from different study designs has been excluded. We will now consider the five different arguments that have been used to rank expert opinion as the lowest level of evidence.

4.5.2. The Measurement Arguments:

The first two arguments that are used to support the claim that expert opinion should provide the lowest level of evidence relate to the measurement of treatment effects. These arguments, which are summarised below, are considered together because they are inter-related.

<u>The Measurement Argument:</u>

If a form of evidence does not involve measurement it provides the worst evidence for determining the effectiveness of a treatment intervention

Expert opinion does not involve measurement

Expert opinion provides the worst evidence for determining the effectiveness of a treatment intervention

<u>The Measurement Bias Argument:</u>

If a form of evidence is affected by measurement bias it provides the worst evidence for determining the effectiveness of a treatment intervention

Expert opinion is affected by measurement bias

Expert opinion provides the worst evidence for determining the effectiveness of a treatment intervention

These are both valid arguments of the form modus ponens. We therefore need to consider the truth of the premises to determine if the arguments are sound. The second premise of the measurement argument claims that expert opinion does not involve measurement. However, experts in different medical fields routinely undertake measurements during the diagnosis, treatment and

follow-up of patients. These measurements form part of the clinical record, may be a medicolegal requirement and be required by purchasers of medical care. There measurements must influence the opinion of the experts and the claim that expert opinion does not involve measurement is surely false.

When a randomised controlled trial or observational study is undertaken measurements are undertaken in a standardised way according to the study protocol. The second premise of the first measurement argument could be modified to claim that expert opinion does not involve standardised measurement on a series of patients. However, there is no reason why an expert cannot make standardised measurements on a series of patients. Some hierarchies of evidence characterise expert opinion as unsystematic or anecdotal, but this does not have to be the case. It could be argued that standardised measurements systematically recorded on a series of patients should be categorised as an observational study. However, if the measurements are not published as a study, this evidence is surely expert opinion.

Proponents of evidence based medicine may accept that expert opinion can be informed by systematic measurements undertaken on a series of patient but argue that observational studies provide superior evidence because readers have access to the measurements themselves[43]. However, this modification fails to recognise the importance of interpretation of any

---

[43] We will charitably assume that any measurements are accurately reported.

measurements. The statement 'A knows p because A knows p' is generally seen as superior to 'A knows p because A knows that B knows p'. However, if A does not have the knowledge or opportunity to interpret p, knowledge acquired through B may be epistemically superior (Hardwig 1991). A series of systematic measurements that have been interpreted by an expert and presented as opinion may therefore provide superior evidence to the uninterpreted results of an observational study.

The first premise of the measurement argument claims that if a form of evidence does not involve measurement it provides the worst level of evidence. This is inconsistent with the claim that systematic reviews should provide the highest level of evidence as this process also does not involve measurement. Similarly an expert can offer an opinion following a rigorous and systematic assessment of an evidence base without making any measurements. This would be consistent with the definition of expert opinion offered by Mitchell and Friese (2007). Proponents of evidence based medicine may be uncomfortable with this definition of expert opinion because it equates expert opinion with a well conducted systematic review. However, it only becomes a systematic review when the process that has been used to identify and summarise the evidence is made explicit. Both premises of the measurement argument are false and the argument is unsound.

We will now turn our attention to the measurement bias argument. Both the measurement argument and measurement bias argument are originally

derived from the same passage in *'Effectiveness and Efficiency'* (Cochrane 1972). However, they are incompatible with each other because if expert opinion does not involve measurement it cannot be affected by measurement bias. The measurement bias argument can only be sound if the measurement argument is unsound.

The second premise of the measurement bias argument claims that expert opinion is affected by measurement bias. This restricts the scope of the argument to expert opinion that involves some form of measurement as it does not apply to expert opinion informed by a rigorous, systematic appraisal of any underlying evidence base. Within randomised controlled trials blinding is used to control measurement bias. Expert opinion that is informed by measurement cannot be blinded and is potentially susceptible to measurement bias. However, as discussed earlier in this chapter, observational studies and many randomised controlled trials, are also susceptible to measurement bias. The measurement bias argument does not therefore distinguish between expert opinion, observational studies and many randomised controlled trials. More importantly, proponents of evidence based medicine recognise that the use of objective outcome measures is more important than study design in the control of measurement bias (Cochrane 1972). There is no reason that objective outcome measures cannot be used to inform expert opinion.

In this subsection we have considered the arguments that expert opinion should provide the lowest level of evidence because it does not involve measurement and it may be affected by measurement bias. The measurement argument is unsound because expert opinion can involve measurement and measurement is not required for expert opinion to have value. The measurement bias argument limits the scope of the argument to expert opinion informed by measurement. However, this argument also provides little support for the claim that certain types of expert opinion should provide the lowest level of evidence. This is because observational studies and many randomised controlled trials are also susceptible to measurement bias and measurement bias is not problematic when objective outcome measures are used.

4.5.3. Expert Opinion does not Consider Outcomes in Large Numbers of Patients:

The third argument that is used to support the claim that expert opinion should provide the lowest level of evidence is that expert opinion does not consider treatment outcomes in large numbers of patients. This argument is summarised below:

If evidence does not consider treatment outcomes in large numbers of patients it provides the worst evidence for determining the effectiveness of a treatment intervention

Expert opinion does not consider treatment outcomes in large numbers of patients

---

Expert opinion provides the worst evidence for determining the effectiveness of a treatment intervention

This is a valid argument of the form modus ponens. We therefore need to consider the truth of the premises to determine whether the argument is sound. The first premise of the numbers argument claims that any study design that does not consider treatment outcomes in large numbers of patients provides poor evidence. However, there is only one study design that cannot consider treatment outcomes in large numbers of patients: the n-of-1 study. No other study designs or processes are defined by the number of patients that they involve. The first premise of the numbers argument does not therefore discriminate between systematic reviews, randomised controlled trials, observational studies and expert opinion. It could be argued that randomised controlled trials should provide the lowest level of evidence because they usually involve small numbers of patients. However, this would be unfair as low patient numbers within randomised controlled trials reflect

practical difficulties associated with the study design (Rawlins 2008). There is no theoretical reason that randomised controlled trials cannot include large numbers of patients.

The second premise of the numbers argument claims that an expert does not consider treatment outcomes in a large number of patients. An expert has been defined as a person who possesses a substantial body of truths in a target domain and has a skill set that allows them to exploit this information to form new true beliefs (Goldman 2001). According to this definition it is not necessary to observe treatment outcomes in large numbers of patients to become an expert. However, I would argue that it would be very difficult to become a medical expert without considering outcomes in large numbers of patients. There is certainly no reason that an expert cannot consider outcomes in large numbers of patients. The numbers argument is therefore unsound because the second premise is false and it fails to discriminate expert opinion from other study designs and review processes.

4.5.4. Expert Opinion may Produce False Conclusions:

The fourth argument that is used to support the claim that expert opinion should provide the lowest level of evidence is that expert opinion may produce false conclusions. This argument is summarised below:

The Error Argument:

If a method of producing evidence may produce false conclusions it provides the worst evidence for determining the effectiveness of a treatment intervention

Expert opinion may produce false conclusions

Expert opinion provides the worst evidence for determining the effectiveness of a treatment intervention

This is a valid argument of the form modus ponens. We therefore need to consider the truth of the premises to determine whether the argument is sound. The first premise of this argument is problematic because, although expert opinion can produce false conclusions, all other study designs and review processes can also produce false conclusions. Meta-analyses and randomised controlled trials are not infallible and randomised controlled trials have been published with conclusions that cannot possibly be true (Leibovici 2001)[44]. The error argument does not therefore discriminate expert opinion from other study designs included within hierarchies. If this argument is followed to its logical conclusion all forms of evidence would be ranked as the lowest level of evidence and there would be no hierarchy.

---

[44] Leibovici (2001) concluded that duration of fever and length of hospital stay were reduced in septicaemia patients when a group prayer was recited for the treatment group several years after the event. It is generally agreed that this conclusion cannot be true because the treatment intervention tested is biologically implausible. This randomised controlled trial appears to have been published as a joke but it demonstrates the fallibility of the study design.

The error argument could be modified to support the weaker claim that expert opinion is more likely to produce false conclusions than observational studies, randomised controlled trials and meta-analyses. However, this would require an external standard that the conclusions of expert opinion and different study designs could be compared against. Randomised controlled trials or meta-analyses cannot provide this standard as this would lead to a circular argument.

The conclusions of meta-analyses, randomised controlled trials and observational studies are reached using an inductive process and are not necessarily true. Only a deductive process produces a conclusion that is necessarily true (Smith 2003). If deductive processes were ranked using hierarchies they would be included within expert opinion. Paradoxically, expert opinion may be the only type of evidence that can produce a conclusion that is necessarily true. This does not mean that expert medical opinion derived using a deductive process will be true as this will depend upon the truth of the premises. However, expert opinion produced using a deductive process does at least have the potential to produce a conclusion that is necessarily true. I would therefore argue that the error argument is unsound because all forms of evidence may produce false conclusions and expert opinion is the only type of evidence that has the potential to produce conclusions that are necessarily true.

<u>4.5.5. Expert Opinion does not have a Control Group:</u>

The final argument that is used to support the claim that expert opinion should provide the lowest level of evidence is that expert opinion does not consider what happens to patients who do not receive the treatment intervention. Superficial consideration of this argument suggests that it is unsound. This is because in order to become a medical expert it is likely that an individual will encounter large numbers of patients with a particular medical condition. These patients will present in different ways, receive a variety of different interventions and some may even refuse treatment. It is therefore highly likely that an expert will be aware of the probable outcome if patients do not receive a particular treatment intervention and that this will influence their opinion. However, this would be a simplistic and uncharitable interpretation of the argument. What proponents of evidence based medicine are actually claiming is that expert opinion should provide the lowest level of evidence because it does not employ a control group. This argument is summarised below:

If a study design or process has no control group it provides the worst evidence for determining the effectiveness of a treatment intervention

Expert opinion has no control group

Expert opinion provides the worst evidence for determining the effectiveness of a treatment intervention

This is a valid argument of the form modus ponens. We therefore need to consider the truth of the premises to determine whether the argument is sound. This argument is subtly different from the other arguments considered in this section because it appears to make a distinction between study designs that employ a control group and both expert opinion and study designs that do not employ a control group.

Prospective case series do not employ a control group. However, it is well recognised that this study design can provide strong evidential support for the effectiveness of a treatment intervention when that treatment intervention is universally successful (Howick 2011). This is the 'all-or-none' effect. Cochrane (1972) claimed that clinical studies were not required to justify the use of penicillin for pneumonia because the treatment had such obvious beneficial effects and the original Oxford Centre for Evidence Based

Medicine hierarchy ranked 'all-or-none' studies as one of the highest levels of evidence (Ball and Phillips 2001). These examples immediately weaken the claim that study designs without a control group should provide the lowest level of evidence in any hierarchy.

Proponents of evidence based medicine may be able to solve the problem with the control group argument created by prospective case series by claiming that the results of 'all-or-none' studies only have strong evidential value because they are compared with a historical control group (Rawlins 2008). On this interpretation treatment intervention X may be successful in all patients with condition Y but this finding only has strong evidential value because all patients with condition Y died prior to the development of treatment intervention X. There is therefore a historical control group that the results of the prospective case series can be implicitly compared against. The control group argument can survive in this modified form so we need to consider why the hierarchies value study designs with control groups.

Proponents of evidence based medicine value control groups because they are concerned that outcomes observed in patients following a treatment intervention may be attributed to the intervention when they are actually caused by other factors. Evidence based medicine is particularly concerned by confounding factors, regression to the mean, the placebo effect and the tendency of many conditions to spontaneously improve with time (Howick 2011). Evidence based medicine would like to observe the different treatment

outcomes with the same group of patients in both the treatment group and the control group. However, this is a theoretical impossibility and the counterfactual position can never be observed (Shadish et al 2002). Study designs with a control group are used as a substitute for the counterfactual position but they can only serve this purpose if all confounding factors are balanced between the treatment and control groups[45]. This is problematic as was argued in the section on randomised controlled trials.

One reason that a control group is considered important is because measurements of treatment outcomes are susceptible to regression to the mean. Regression to the mean is the concept that every measurement varies by chance around the true measurement (Shadish et al 2002). Proponents of evidence based medicine are concerned that without a control group, any improvement in treatment outcome may be mistakenly attributed to the treatment intervention when the improvement actually reflects initial extreme values that have regressed to the mean.

The problem of regression to the mean is not prevented by employing study designs with a control group because any measurement may be affected. Outcomes in the treatment group of a randomised controlled trial are just as susceptible to regression to the mean as expert opinion based upon treatment outcomes observed in a series of patients. Proponents of evidence

---

[45] A control group can only ever approximate the counterfactual position because although all confounding factors could theoretically be balanced each patient must ultimately be in either the treatment or control group. There will therefore always be an imbalance with regard to patients.

based medicine may claim that regression to the mean is balanced across the treatment and control groups by the randomisation process but this appears similar to the confounding arguments and there is no empirical evidence to support this claim. Regression to the mean is best controlled by multiple pre-test measurements made prior to providing treatment (Shadish et al 2002). These multiple pre-test measurements can be made by an expert prior to providing treatment interventions or incorporated within different study designs. Expert opinion should not therefore be ranked as the lowest level of evidence because it is more susceptible to regression to the mean.

A control group is also considered important because some diseases can spontaneously resolve with time. This argument can be traced back to Archie Cochrane who recognised that soldiers in German prison camps in World War II often recovered despite limited medical supplies (Cochrane and Blythe 2009). However, expert opinion will not be misled by the spontaneous resolution of diseases if multiple pre-test measurements are recorded. Proponents of evidence based medicine may argue that multiple pre-test measurements are not possible for some conditions but this is a practical consideration not a theoretical consideration.

The final reason that a control group is considered important is because it allows the use of a placebo comparison. A placebo comparison is a situation in which we compare two groups that are identical in all but one therapeutically relevant respect (Howick 2011). A 'placebo' is usually

conceptualised as an intervention in itself but this is incorrect. The control group do not receive a 'placebo' they receive an intervention which is missing a particular component that is additionally present in the test group. A placebo comparison is simply a control group set up in a certain way (Turner 2012). The claim that expert opinion does not allow the use of placebo comparison therefore adds nothing to the claim that expert opinion does not use a control group.

A placebo comparison should not be confused with the placebo effect. The placebo effect recognises that patient expectations and beliefs can influence treatment outcomes independent of any 'true' underlying effect (Howick 2011)[46]. However, the use of a control group does not prevent patient expectations and beliefs influencing treatment outcomes. Patient expectations and beliefs can be controlled by blinding but this is only possible in a few randomised controlled trials where the treatment intervention and control intervention are indistinguishable. The use of objective outcome measures is considered a far better way of preventing patient expectations and beliefs influencing treatment outcomes (Cochrane 1972, Haynes et al 2006). Objective outcome measures are not dependent upon study design and can be used to inform expert opinion.

---

[46] It has been suggested that if expectations and beliefs can influence the treatment outcome they should be considered part of the treatment intervention and not separated from any underlying biological effect (Howick 2011).

In this subsection we have considered the claim that expert opinion should provide the lowest level of evidence because it does not consider what happens to patients without treatment. This claim has been interpreted charitably and reformulated as the claim that expert opinion does not have a control group. A control group is considered important because it balances confounding factors, allows placebo comparisons and controls for regression to the mean and spontaneous improvement. However, each of these arguments is problematic. Control groups do not ensure that confounding factors are controlled, placebo comparisons are effectively control groups and regression to the mean and spontaneous improvement are better controlled by multiple pre-test measurements. I would therefore argue that the control group argument provides little support for the claim that expert opinion should provide the lowest level of evidence in hierarchies of evidence.

4.5.6. Summary:

In this section we have considered the arguments that have been used to rank expert opinion as the lowest level of evidence in hierarchies of evidence. Critical analysis of these arguments has been complicated by the different definitions of expert opinion that have been used. Some of these definitions have been criticised because they characterise expert opinion as unsystematic or uncritical when this is not necessarily the case. A working

definition of expert opinion has been developed which defines expert opinion as any evidence that is not derived from a study.

The five arguments that have been used to support the claim that expert opinion should provide the lowest level of evidence were then considered: expert opinion does not involve measurement; expert opinion is susceptible to measurement bias; expert opinion does not consider treatment outcomes in large numbers of patients, expert opinion may produce false conclusions and expert opinion does not consider what happens to patients without treatment. The first argument is unsound because expert opinion can involve measurement. The second argument has restricted scope and does not discriminate expert opinion based upon measurement from other study designs and processes that are susceptible to measurement bias. The third argument is unsound because one would expect expert opinion to consider treatment outcomes in large numbers of patients. The fourth argument is unsound because all forms of evidence may produce false conclusions. The final argument was reformulated as expert opinion should provide the lowest level of evidence because it does not use a control group. However, the presence of a control group does not ensure that confounding factors are balanced, and regression to the mean and spontaneous improvement, are preferably controlled by multiple pre-test measurements. Collectively, these arguments provide little support for the claim that expert opinion should provide the lowest level of evidence in hierarchies of evidence.

<u>4.6. Conclusions:</u>

In this chapter we have critically analysed the arguments that have been used to rank meta-analyses, systematic reviews and randomised controlled trials as the highest level of evidence, and expert opinion as the lowest level of evidence, within hierarchies of evidence. The importance of practical considerations has been acknowledged but it has been reasoned that the hierarchies should have strong theoretical support because they are fundamental to the knowledge claims made by evidence based medicine. The theoretical arguments that have been analysed are either unsound or only survive critical analysis when their scope is restricted. This is worrisome as knowledge claims made by evidence based medicine have significant importance within medicine.

Inherent within the concept of a hierarchy of evidence is the claim that some study designs probably provide superior evidence to other study designs. The analysis presented in this chapter has revealed that evidence from meta-analyses, systematic reviews and randomised controlled trials is not epistemically superior, and evidence from expert opinion is not epistemically inferior, to evidence from other study designs. This does not mean that meta-analyses, systematic reviews, randomised controlled trials and expert opinion cannot provide evidence to support decision making. It is my contention that evidence from a variety of different processes and study designs can provide evidence to support decision making. However, no study design or form or

evidence is epistemically superior and there appears to be limited theoretical support for hierarchies. It would be more appropriate to view expert opinion and different study designs as a level playing field.

Epistemological problems associated with the ranking of different forms of evidence have been recognised, to a limited extent, by some hierarchies. This is exemplified by the GRADE system which excludes expert opinion and systematic reviews and ranks observational studies above randomised controlled trials under certain conditions (Guyatt et al 2008). Once we accept that no study design is epistemically superior this should not seem controversial. However, the GRADE system is still predicated on the assumption that evidence from randomised controlled trials is epistemically superior to evidence from observational studies.

Within hierarchies of evidence a number of apparently unconnected arguments are used to rank observational studies above expert opinion and below randomised controlled trials and systematic reviews/meta-analyses. The use of the term hierarchy implies that the different study designs are ranked according to some property. If this is the case the arguments that are used should reflect that property. In the next chapter we will therefore consider the different properties that the hierarchies claim to use to rank different study designs. This should improve our understanding of evidence based medicine and may explain some of the differences between the hierarchies that have been identified.

Chapter 5: Ranking Properties

5.1. Introduction:

As we have seen evidence based medicine uses a number of different hierarchies of evidence to determine the effectiveness of treatment interventions. These hierarchies most commonly rank systematic reviews and randomised controlled trials as the highest level of evidence, observational studies as the intermediate level of evidence and expert opinion as the lowest level of evidence. However, this is not a universal finding and some hierarchies rank these study designs in different orders or include other study designs. There is also considerable variation in the number of levels contained within different hierarchies. The simplest hierarchies have only 3 levels of evidence (Cochrane 1972, Spitzer et al 1979, Djulbegovic and Hadley 1998, Ebell et al 2004) whereas the most complicated hierarchies have 10 or more levels and sub-levels (Hadorn et al 1996, Ball and Phillips 2001).

The use of the term 'hierarchy' implies that the different study designs are ranked according to some property. It is important to investigate this property as this will improve our understanding of the hierarchies and may explain why apparently unconnected arguments are used at different levels. Knowledge of the property will also facilitate our understanding of the

arguments that were analysed in the preceding chapter. These arguments claimed that different study designs could provide superior or inferior evidence but did not explore what this meant. If different study designs can provide superior or inferior evidence this should relate to the property that is used to rank them within hierarchies. This property should also influence the number of levels contained within hierarchies.

When the hierarchies of evidence are considered collectively a number of different properties are used to rank study designs. Some hierarchies do not mention the property that is used whereas other hierarchies have claimed to use nebulous qualities, such as 'evidence strength', 'quality of evidence', 'international standards' or 'robustness of evidence', without explicitly defining these terms (Spitzer et al 1979, Armola et al 2009). These properties have powerful perlocutionary force but they are meaningless if they are not defined.

Other hierarchies are clear about the property that they use. However, they purport to use a variety of different properties. The properties that are most commonly used are bias, validity, confidence, trustworthiness and truth. Some hierarchies appear to use more than one property. This is exemplified by the hierarchy proposed by Evans (2003) that claims to use 11 different properties: risk of error, bias, validity, minimisation of confounding, closeness to the truth, confidence, trust, caution, generalisability, reliability and

robustness. Many of these properties can be identified from the following two quotations:

> 'It has long been recognised that not all research designs are equal in terms of the risk or error and bias in their results. When seeking answers to specific questions, some research methods provide better evidence than that provided by other methods. That is the validity of the results of research varies as a consequence of the different methods used.' (Evans 2003, page 78)

> '(The RCT) is considered the most reliable evidence because the processes used during the conduct of an RCT minimise the risk of confounding factors influencing the results. As a result of this, the findings generated by RCTs are likely to be closer to the true effect than findings generated by other methods. This confidence in the findings of research has important implications' (Evans 2003, page 78).

In this chapter, the five properties that are most commonly used by hierarchies of evidence to rank different study designs are considered. These properties are truth, bias, validity, confidence and trust. This chapter aims to further improve our understanding of the hierarchies and resolve some of the confusion that surrounds them.

5.2. Truth:

A number of hierarchies purport to use truth or closeness to the truth to rank different study designs (Cook et al 1992, Djulbegovic and Hadley 1998, Waugh 1999). This is clearly illustrated by the following quotes:

'Some evidence is closer to the truth than other evidence.' (Djulbegovic and Hadley 1998, page 313)

'The grading of evidence so far is mainly about how certain we can be that the evidence presented equates to truth.' (Waugh 1999 page 56)

It is important to appreciate that the hierarchies are designed to rank different study designs. Study designs are methodological tools and do not have any truth content. It is therefore not possible to rank different study designs according to the truth or falsity of the study design (Djulbegovic et al 2009). When the hierarchies claim to use truth they are not actually ranking the study designs themselves they are ranking the conclusions that are produced by the different study designs.

A number of different theories of truth have been proposed. These include the correspondence theory of truth, coherence theory of truth and pragmatic theory of truth. All theories of truth are considered problematic and none of them have been universally accepted. However, all theories of truth agree

that a proposition can only ever be true or false. Whether a particular proposition is true may depend upon whether it corresponds to a 'fact' about the external world (correspondence theory of truth), coheres with existing beliefs (coherence theory of truth) or has practical use (pragmatic theory of truth), but any proposition can only be true or false (Simmons 2005).

The different theories of truth dictate that any hierarchy that ranks the conclusions produced by different study designs according to truth can only have two levels: true and false. This is because any intermediate levels will be meaningless within theories of truth. This creates an immediate problem for all of the identified hierarchies because they all have at least three levels. For example, the Scottish Intercollegiate Guideline Network hierarchy claims to use truth but it contains 8 different levels (Harbour and Miller 2001). If the property used by this hierarchy is truth what do the intermediate levels of evidence represent? None of the hierarchies that claim to use truth consider this problem.

Proponents of evidence based medicine may argue that the intermediate levels in these hierarchies reflect closeness to the truth. This would imply that as we ascend the hierarchy the different study designs can be used to produce conclusions that take us closer and closer to the truth[47]. The idea that hierarchies may be ordered to reflect closeness to the truth has some similarities with the philosophical concept of verisimilitude.

---

[47] This assumes a correspondence theory of truth and an external reality that can be accessed.

Verisimilitude is the concept that scientific theories become increasingly true as old theories are replaced. Verisimilitude has been proposed as a potential solution to the pessimistic induction thesis. According to the pessimistic induction thesis we have good inductive grounds to believe that all current scientific theories, regardless of how successful they are, will eventually be shown to be false. The pessimistic induction theory is supported by numerous historical examples within science (Ladyman 2002, Chalmers 2010). Newton-Smith (1981) has argued, using inference to best explanation, that new scientific theories must have increasing truth content relative to the theories that they replace otherwise science does not progress. However, he was unable to explain what it meant to be closer to the truth or how we could determine closeness to the truth. This is a major problem for the concept of verisimilitude.

The concept of verisimilitude may be applicable to scientific theories but it is not applicable to study designs or the conclusions produced by those study designs. Scientific theories are used to explain phenomena and make predictions which can be confirmed or falsified through experiment. A scientific theory may appear to have greater truth content if it explains phenomena that were previously unexplained or it makes more successful predictions than a preceding theory[48]. However, the same reasoning cannot

---

[48] Unfortunately this is an illusion as all theories can be used to make an infinite number of predictions and explanations. A greater number of successful predictions and explanations do not therefore translate to greater truth content of a theory (Newton-Smith 1981).

be applied to the conclusion of a study because this can only be true or false and cannot have a variable degree of truth content.

It could be argued that study designs within hierarchies are ranked to reflect the probability that the conclusions are true. Probability has been defined as a quantitative estimate of the likelihood that an event will occur (Guyatt et al 2008). The concept of probability is routinely employed by evidence based medicine when null hypotheses are accepted or rejected and confidence intervals are presented. The truth of a conclusion can be more or less probable so this interpretation of hierarchies would allow intermediate levels. However, none of the hierarchies explicitly claim to rank study designs based upon the probability that the conclusions are true. It is also unclear how a probabilistic interpretation can be justified without returning to the arguments that were analysed in Chapter 4. If the arguments for the superiority of randomised controlled trials and meta-analyses are unsound, restricted in scope and lack empirical support how can these study designs produce conclusions that are more probably true?

If truth is the property that is used to rank different study designs within hierarchies the study designs at the top of the hierarchy should produce true conclusions. However, there are a several problems with this claim. Consider the standard hierarchy of evidence which ranks randomised controlled trials as the highest level of evidence. Empirical evidence shows that different randomised controlled trials investigating the same treatment intervention

can produce variable results. This is the rationale for meta-analyses. The conclusions of some of these randomised controlled trials must therefore be false. If randomised controlled trials can produce false conclusions, and the property that is used to rank different study designs is the truth of the conclusions, randomised controlled trials should not provide the highest level of evidence. Empirical research also shows that meta-analyses investigating the same treatment intervention can produce different conclusions. This means that meta-analyses should also not provide the highest level of evidence.

Proponents of evidence based medicine may argue that randomised controlled trials produce true conclusions when they are ideally conducted. However, this argument is unsound because randomisation does not guarantee that confounding factors are balanced between the treatment and control group. The results of any randomised controlled trial are always underdetermined because they may be attributed to the treatment intervention under investigation or an imbalance in confounding factors. Even ideally conducted randomised controlled trials can produce false conclusions.

I would argue that no study design that tests a null hypothesis should be ranked as the highest level of evidence in a hierarchy that uses the truth of the conclusion to determine hierarchical position. This is because the null hypothesis is rejected based upon the probability of the observed results if there really is no difference between the interventions given to the treatment

and control groups. Therefore even when the null hypothesis is rejected, there is always a small probability that the null hypothesis is actually true and that the conclusion of the study is false. Meta-analyses, randomised controlled trials, cohort studies and case-control studies are all designed to test null hypotheses based upon probability. The conclusions or all of these study designs are underdetermined and are not necessarily true when the null hypothesis is rejected.

Similar reasoning can be applied to the lower levels of hierarchies. Expert opinion is often ranked as the lowest level of evidence but expert opinion is not always false. Expert opinion may be informed by randomised controlled trials and meta-analyses and may agree with the conclusions of randomised controlled trials and meta-analyses. Expert opinion also includes deductive reasoning and this is the one process that can produce conclusions that are necessarily true. If the property that is used to rank different study designs within hierarchies is the truth of the conclusions, expert opinion does not deserve to be ranked as the lowest level of evidence.

Several hierarchies claim to rank different study designs using the property of truth. However, this claim is difficult to sustain because all hierarchies have at least 3 levels whereas the conclusion of any study can only be true or false. It is therefore unclear what any intermediate levels of evidence represent as the conclusions of different study designs cannot have increasing truth content. A hierarchy with two levels could use the property of truth to rank

different study designs but it is uncertain which forms of evidence would provide the highest level as no study design produces conclusions that are necessarily true. It would appear difficult for hierarchies to use truth to rank different study designs so we need to consider the other properties that have been proposed.

5.3. Bias:

A number of different hierarchies have claimed that study designs are ranked to reflect susceptibility to bias. These include the hierarchies proposed by the Evidence Based Medicine Working Group (Guyatt et al 1995), Institute for Clinical Systems Improvement (Greer et al 2000) and the Australian National Health and Medical Research Council (Merlin et al 2009). Cochrane (1972) was also concerned by bias although he never explicitly stated that his hierarchy was designed to reflect susceptibility to bias. The importance of bias is illustrated by the following quotations:

'The hierarchy of design types is fairly consistent amongst evidence grading systems and reflects the fact that different study designs vary in the likelihood that an individual study will be biased' (Greer et al 2000 page 5).

'In this hierarchy interventional study designs were ranked according to the likelihood that bias had been eliminated' (Merlin et al 2009 page 1)

In order to analyse the claim that hierarchies rank study designs to reflect susceptibility to bias we must first define what bias is. Evidence based medicine consistently defines bias as 'systematic deviation from the truth' (Sackett 1979, Guyatt et al 2002, Straus et al 2011). Therefore any hierarchy that uses bias ranks study designs based upon their ability to produce conclusions that are closer and closer to the truth. Any hierarchy measuring bias assumes that we are able to get closer to the truth if we employ certain study designs. Furthermore, the use of the term 'systematic' in the definition of bias implies that different study designs produce conclusions that deviate from the truth in a transparent or reproducible way.

It is not clear how the conclusion of any study can systematically deviate from the truth and this is not explained by any of the 14 hierarchies that claim to use bias. I would argue that if a conclusion has deviated from the truth it is false. This is because all accepted theories of truth agree that conclusions can only ever be true or false (Simmons 2005). Any hierarchy that uses bias therefore faces the same difficulties that were encountered in the previous section when we considered a hierarchy of truth. Unless evidence based medicine is able to articulate a theory of truth that allows varying degrees of truth a hierarchy based on bias is difficult to sustain.

Even if proponents of evidence based medicine were able to articulate a theory of truth that allowed varying degrees of truth a hierarchy measuring

the susceptibility of different study designs to bias would face further problems. Evidence based medicine recognises over 30 types of bias (Sackett 1979, Guyatt et al 2002) but the hierarchies are primarily concerned with selection bias and measurement bias. Other types of bias are rarely mentioned. The hierarchies appear preoccupied with selection bias and measurement bias because they believe that the randomised controlled trial study design can be used to control these types of bias. However, if bias is systematic deviation from the truth then all types of bias should be of equal importance.

It is also unclear how different types of bias should interact to determine evidence rank. As the hierarchy is ascended are the study designs susceptible to fewer types of bias or is overall bias reduced? Any attempt to quantify overall bias would be problematic because it would require access to an external standard of truth and some method of quantifying deviation from that truth. It would therefore make more sense if study designs were ranked based upon the number of different types of bias that they were susceptible to. This appears to be the rationale behind the hierarchy presented in the *'How to Read Clinical Journals Series'*:

> 'The case-control study is susceptible not only to the bias from the angina patient we noted in the cohort study, but also to several other sorts of bias'. (Trout 1981 page 987)

However, ranking study designs according to the number of different types of bias that they are susceptible to creates a problem for hierarchies that contain systematic reviews. This is because systematic reviews are susceptible to a number of biases, such as language bias, selective outcome reporting bias and publication bias that are not relevant to primary study designs (Higgins and Green 2011). Systematic reviews are actually susceptible to more types of bias than randomised controlled trials and some of these biases, particularly publication bias, are considered difficult to identify (Guyatt et al 2011a). If the number of different types of bias that a study design is susceptible to determines hierarchical position then systematic reviews would surely not be ranked above randomised controlled trials.

A number of hierarchies claim to rank study designs according to susceptibility to bias where bias is defined as systematic deviation from the truth. These hierarchies do not explain how conclusions can deviate from the truth without being false or how this deviation can be measured. It is also unclear whether these hierarchies of evidence are organised to reflect susceptibility to overall bias or different types of bias. A hierarchy organised in either way appears problematic. It is therefore not clear how a hierarchy of evidence can rank different study designs according to susceptibility to bias.

<u>5.4. Validity:</u>

Thirteen hierarchies of evidence claim to use validity to rank different study designs. These include the hierarchies proposed by the Agency for Healthcare Research and Policy and Research (Hadorn et al 1996) and the Evidence Based Medicine Working Group (Guyatt et al 2000):

> 'The present system is based on the tenet that flaws in research design are serious to the extent they threaten the validity of the study results.' (Hadorn et al, 1996, page 749)

> 'Underlying these steps are 2 fundamental principles. One, relating primarily to the assessment of validity, posits a hierarchy of evidence to guide clinical decision making.' (Guyatt et al 2000, page 1291)

Within analytical philosophy, validity is a property of deductive inferences. An argument form is valid if there is no possible situation in which the premises can be true and the conclusion false. An argument can only ever be valid or invalid and it is not possible to have increasing degrees of validity (Smith 2003). This would appear to restrict hierarchies that use validity to two levels: valid and invalid. Furthermore, the hierarchies of evidence are composed of empirical study designs which rely on inductive inferences and validity is not a property of inductive inferences. The hierarchies cannot therefore use validity to rank different study designs, where validity is defined as a property

of deductive inferences. We therefore need to consider how the hierarchies define validity.

Unfortunately, few of the hierarchies that claim to use validity actually define this term. Where this term is defined, validity is consistently equated with proximity to the truth. For example, the Strength of Recommendation Taxonomy claimed that validity was synonymous with bias (Ebell et al 2004). The Evidence Based Medicine Working Group defined validity as closeness to the truth (Oxman et al 1993) and the *'How to Read Clinical Journals Series'* defined validity as the absence of systematic deviation from the truth and the presence of precision (Haynes 1981)[49]. The *'Users' Guides to the Medical Literature'* defined validity in two different ways:

> 'In critical appraisal terms, validity reflects the extent to which the study results are likely to be subject to systematic error and thus be more or less likely to reflect the truth.' (Guyatt et al 2002, page 807)…In health status measurement terms, validity is the extent to which an instrument measures what it is intended to measure' (Guyatt et al 2002, page 807).

When validity is equated with proximity to the truth or systematic deviation from the truth we encounter the same problems that were discussed with regard to bias. These hierarchies do not explain how conclusions can deviate

[49] The definition of validity provided by Haynes (1981) is interesting because it suggests that validity is more than proximity to the truth. However, it is derived from an article on diagnostic tests and may not be applicable to hierarchies of evidence.

from the truth without being false or how this deviation can be assessed. The second definition provided by the *'Users' Guides to the Medical Literature'* is not enlightening unless the property that is being measured is specified.

Thirteen hierarchies of evidence claim to use validity to rank different study designs. However, validity is a property of deductive inferences and the hierarchies comprise empirical study designs that make inductive inferences. Several hierarchies have equated validity with closeness to the truth but, as analysed in the sections on truth and bias, no theories of truth allow degrees of truth. It is therefore unclear how a hierarchy can use validity to rank different study designs.

5.5. Confidence and Trust:

Eleven hierarchies of evidence claim to measure confidence or trust in the conclusions produced by different study designs. These include the hierarchies proposed by the *'How to Read Clinical Journals Series'* (Trout 1981), GRADE Working Group (2004) and the Joanna Briggs Institute (2011):

> 'Readers can place considerable confidence in estimates of strength from a randomised trial, fair confidence in an estimate of strength from a cohort study and only a little confidence in an estimate of strength from a case-control study' (Trout 1981 page 988).

'The quality of evidence indicates the extent to which we can be confident that an estimate of effect is correct' (GRADE Working Group 2004 page 1490).

'The aim of assigning levels of evidence is to provide an element of trustworthiness of the findings of the review.' (Joanna Briggs Institute, 2011, page 15)

The hierarchies often use the terms trust and confidence synonymously. However, Hawley (2012) has argued that trust and confidence are different concepts. Trust can be defined as the optimistic acceptance of a vulnerable situation in which the truster believes that the trustee will care for the truster's interests (Goudge and Gilson 2005). Trust requires a relationship between a truster and a trustee and is not a property associated with inanimate objects or processes. Confidence can be defined as the feeling or belief that one can have faith in or rely on someone or something. Confidence is a property that can be associated with inanimate objects or processes (Hawley 2012).

The hierarchies include a variety of different review processes, study designs and expert opinion. Review processes and study designs are inanimate processes that are used to generate conclusions. Therefore, if we accept the definition of trust proposed by Goudge and Gilson (2005), a hierarchy cannot measure trust because we cannot trust inanimate process. This does not

mean that we do not trust researchers to conduct and report studies accurately[50]. Simply that it is not possible to trust a study design per se.

Expert opinion may be trusted because there is a relationship between the expert (the trustee) and the user of the opinion (the truster). Some hierarchies differentiate between consensus expert opinion and individual expert opinion (Braunewald et al 1994, Newhouse et al 2007). This may be because the opinion of a group of experts is considered more trustworthy although Fricker (2005) has argued that sincerity and competence are more important determinants of trustworthiness than consensus opinion. However, once expert opinion and different study designs are combined within the same hierarchy the property that is used to determine hierarchical position cannot be trust.

Confidence is a property that can be associated with both individuals and inanimate processes (Hawley 2012). It is therefore possible to have confidence in both expert opinion and the ability of different study designs to determine the effectiveness of a treatment intervention. This means that confidence can potentially be used to determine hierarchical position. A hierarchy using confidence can have multiple levels because it is possible to be more or less confident about the effectiveness of a treatment intervention. The 12 levels and sublevels in the Agency for Healthcare Research and Quality hierarchy (Hadorn et al 1996) could therefore reflect varying degrees

---

[50] Lipton (1998) has argued that trust in researchers is essential to the progression of science because without this trust we would have to verify everything ourselves.

of confidence. Unlike a hierarchy that claims to use truth, a hierarchy that claims to use confidence to determine hierarchical position is not restricted to two levels.

If hierarchies are ordered to reflect confidence in the ability of different study designs to determine treatment effectiveness they should explain why we should have greater confidence in the results of some study designs. When we investigate this we encounter the same arguments that were analysed in Chapter 4. For example, The Joanna Briggs Institute (2011) claims that we should have greater 'trust' in study designs that minimise bias and (Trout 1981) places greatest confidence in randomised controlled trials because of features inherent to the study design. These hierarchies use confidence as a surrogate term for bias (Joanna Briggs Institute 2011) or purported ability to control confounding factors and be blinded (Trout 1981).

Study designs within the GRADE system are also ranked to reflect confidence in the estimate of treatment effect (GRADE Working Group 2004). This hierarchy initially ranks randomised controlled trials higher than observational studies because of factors inherent within the different study designs, although the final ranking is determined following consideration of a variety of factors including risk of bias, imprecision, inconsistency of results and publication bias (Balshem et al 2011). Guyatt et al (2011) accepted that there was little empirical evidence to support the use of these factors and emphasised that they should be interpreted subjectively. However, although

the GRADE system considers a number of different factors that may influence confidence in the overall effectiveness of a treatment intervention, it provided no additional reasons to support the initial claim that we should have greatest confidence in the results of randomised controlled trials.

The GRADE hierarchy clearly demonstrates that a number of different factors can influence our confidence in the estimate of treatment effect. These factors include consistency of results, biological plausibility and magnitude of treatment effect[51]. If this is the case why should we not primarily rank evidence according to consistency of results, biological plausibility or magnitude of treatment effect rather than study design? Such a hierarchy would not need to ignore study design as this could be included at a later stage. A hierarchy that ranked evidence primarily according to biological plausibility would be particularly interesting. Biological implausibility is usually recognised as providing strong evidence against a treatment effectiveness claim (Howick 2011, Blunt 2015). However, biological plausibility or implausibility is usually determined by laboratory based studies, pathophysiological rationale and expert opinion. These types or evidence are typically ranked in the lower levels of traditional hierarchies.

Any hierarchy that measures confidence in the results of a particular study design can be interpreted from a Bayesian perspective. This has important implications for hierarchies because confidence in the conclusion of a study

---

[51] Every-Palmer and Howick (2014) have proposed that hierarchies should downgrade evidence if there is industry bias although this is not done by any hierarchies at present.

is not just influenced by factors such as study design, biological plausibility and magnitude of treatment effect. Confidence in the conclusion is also influenced by our prior confidence in the conclusion before the study is considered. This prior confidence will be influenced by a variety of other factors including the existing evidence base and previous experience and will vary between individuals (Chalmers 2010). The confidence of different individuals in the conclusions of the same study may therefore vary depending upon prior confidence[52]. From a Bayesian perspective, confidence in the conclusion of a particular study cannot be determined in isolation based upon study design.

The importance of prior confidence to a Bayesian interpretation of hierarchies of evidence can be illustrated with an example. A well conducted randomised controlled trial is undertaken that concludes that intervention Y is an effective treatment for condition Z. Clinician A has successfully used intervention Y to treat a number of patients, he believes that intervention Y is effective and the conclusion of the randomised controlled trial reinforces this belief. He has a high degree of confidence in the effectiveness of intervention Y for condition Z following consideration of the randomised controlled trial. Clinician B has found intervention Y to be an unsuccessful treatment for condition Z, he believes that treatment Y is ineffective and although the randomised controlled trial may slightly increase his confidence in the effectiveness of

---

[52] When there is a large body of pre-existing evidence there may be convergence of belief so that most individuals have similar prior confidence but this is not necessarily the case and can never be true of a new intervention.

intervention Y, he still has little confidence in this intervention[53]. Both of these interpretations of the conclusion of the randomised controlled trial are correct but they result in differing degrees of confidence in the treatment intervention.

A Bayesian interpretation of hierarchies of evidence allows different interpretations of the conclusions of the same studies by different appraisers. Some readers may be concerned by this as evidence based medicine dictates the treatment interventions that can be prescribed and funded, medicolegal standards of care and the research agenda. On a practical level the problem created by Bayesian interpretations of hierarchies may be solved by attributing greater weight to the degree of confidence ascribed by an expert or group of experts. However, this uses expert opinion, the form of evidence that is usually least valued by hierarchies, to solve the problem created by a Bayesian interpretation. Consistency of decision making within evidence based medicine is considered in greater detail in Chapter 7.

A number of hierarchies claim to rank study designs based upon trust or confidence in the conclusions of the studies. Hierarchies cannot use trust because this is not a property that is associated with inanimate processes. Hierarchies can use confidence and these hierarchies can have multiple levels reflecting differing degrees of confidence. However, the hierarchies that do claim to use confidence do not explain why we should have greater confidence in certain study designs without using the arguments that were

---

[53] For the purpose of this example it does not matter if intervention Y is effective or ineffective.

analysed in Chapter 4. These arguments were either unsound, unsupported by empirical evidence or only survived critical analysis with restricted scope. Further empirical or theoretical evidence is therefore required to support the claim that we should have greater confidence in the conclusions of some study designs. Even if empirical or theoretical evidence could be provided for the superiority of some study designs it is not clear why study design is prioritised over other factors that may potentially affect confidence in any conclusions.

5.6. Conclusion:

In this chapter we have considered the properties that have been used to rank different study designs within hierarchies. Many hierarchies do not describe the property that they use. Where the property is described there is considerable variation with different hierarchies purporting to use truth, bias, validity, trust and confidence to rank study designs. Hierarchies that use truth, bias or validity are problematic because conclusions are either true or false and cannot have increasing truth content. Hierarchies that use trust are also problematic as this is not a property that is applicable to inanimate study designs. Hierarchies can potentially rank different study designs based upon confidence in the truth of conclusions. However, empirical or theoretical support for the claim that we should have increased confidence in conclusions produced by certain study designs is required. These hierarchies

also do not explain why they prioritise study design over other factors that can also influence our confidence in conclusions.

The analysis of the arguments that was presented in the previous chapter was complicated by a lack of clarity over the meaning of 'superior' and 'inferior' evidence. A hierarchy that uses confidence in the truth of conclusions to rank different study designs can produce 'superior' and 'inferior' evidence because we can have variable degrees of confidence in evidence claims. However, unless further theoretical or empirical support is provided, the arguments analysed in Chapter 4 appear circular. We are effectively expected to have most confidence in the conclusions of randomised controlled trials because the conclusion was derived using a randomised controlled trial study design. Similarly, we should have the least confidence in expert opinion because expert opinion is expert opinion.

In Chapter 4 we analysed the arguments that have been used to support the standard and modified standard hierarchies of evidence. These arguments were shown to be unsound, restricted in scope or lacking empirical support. This chapter aimed to further increase our understanding of hierarchies by analysing the properties that are used to rank study designs within them. However, this analysis has revealed a significant amount of confusion. Study designs can potentially be ranked to reflect increasing confidence in the truth of their conclusions but it is unclear why we should have greater confidence in the conclusions of certain study designs. If the property that is used to rank

study designs cannot be justified it is unclear whether hierarchies have any utility.

The analysis presented in Chapters 4 and 5 has revealed that there is little theoretical or empirical support for hierarchies of evidence. This analysis provides a foundation for the remainder of the thesis. In the next chapter I will consider a variety of factors that have influenced the development of hierarchies. I will argue that the absence of a firm theoretical foundation has allowed the hierarchies to proliferate under the influence of factors that are independent of study design. This may explain why we have so many different hierarchies of evidence.

6.1. Introduction:

The systematic review presented in Chapter 3 identified 45 different hierarchies of evidence. These hierarchies are used to provide justification for the knowledge claims that are made by evidence based medicine. Furthermore, these knowledge claims dictate the treatment interventions that can be used, healthcare funding, medicolegal standards and the research agenda. It is therefore important to understand why there are so many different hierarchies. The analysis presented in this chapter seeks to explain the variation that is seen amongst hierarchies and improve our understanding of evidence based medicine.

Hierarchies of evidence rank the importance of evidence primarily according to the study design that is used to produce the evidence (Upshur 2009). However, the hierarchies have also been influenced by a variety of others factors that are independent of study design[54]. These factors have led to an increase in the total number of hierarchies that can be used within evidence based medicine. Five factors that have had a significant influence on the hierarchies are considered in depth in this chapter. These factors are recognition of the importance of different questions; attempts to establish

---

[54] Factors independent of study design that have influenced the development of hierarchies of evidence have been termed conditions (Blunt 2015).

professional jurisdiction; failure to separate the hierarchy of evidence and grade of recommendation; practical considerations and standardisation. These are not the only factors that have influenced the development of hierarchies but it will be demonstrated that they have all had a significant influence.

Once the five factors that have influenced the development of the hierarchies have been analysed we will consider why these factors have been able to exert significant influence. It will be argued that the lack of theoretical and empirical support for the standard and modified standard hierarchies of evidence has allowed factors that are independent of study design to assume greater importance. This is because it is difficult for the standard and modified standard hierarchies of evidence to assume a normative role if they are undermined by a lack of theoretical and empirical support. This is advanced as a possible explanation but it is hoped that the analysis that is presented will persuade the reader to accept this argument.

6.2. Recognition of the Importance of Different Questions:

The first factor that has led to an increase in the number of different hierarchies of evidence is recognition of the importance of different questions within medicine. The earliest hierarchies all presented a single hierarchy and were designed to determine the effectiveness of a treatment intervention.

This changed in 1999 with the introduction of the Oxford Centre for Evidence Based Medicine Levels of Evidence (Ball and Phillips 2001) which presented different hierarchies for prognosis, diagnosis, economic evaluations and treatment interventions[55].

Since 1999 a number of other systems have been developed with different hierarchies for different questions (Evans 2003, Ebell et al 2004, Merlin et al 2009, Sayre et al 2010, Joanna Briggs Institute 2011 and Howick et al 2011). All of these systems present a hierarchy for determining the effectiveness of treatment interventions alongside other hierarchies addressing different questions. Collectively these systems consider questions of diagnosis, prognosis, harms, screening, aetiology, frequency of a condition, feasibility, appropriateness, meaningfulness and economic evaluations although no system considers all of these questions. When these systems are used to determine the effectiveness of a treatment intervention, systematic reviews[56] or randomised controlled trials often provide the highest level of evidence. When other questions are considered, systematic reviews and randomised controlled trials rarely feature. We find that these questions are best answered by a variety of different studies including cohort studies, systematic reviews of cohort studies, validated clinical decision rules or economic evaluations.

---

[55] The Canadian Hypertension Society had previously proposed 'hierarchies of evidence' for prognosis and diagnosis (Carruthers et al 1993). However, these 'hierarchies' simply recognised factors that could affect the methodological quality of studies and made no attempt to rank different study designs in a hierarchical manner. These were therefore not true hierarchies of evidence.

[56] When the term systematic review is used in an unqualified sense it is used to refer to a systematic review of randomised controlled trials. This was considered in detail in Chapter 4.

Prior to 1999 hierarchies only considered treatment interventions and they prioritised the results of systematic reviews and randomised controlled trials. However, there appears to have been an increasing realisation that randomised controlled trials could not be used to answer all questions that were important within medicine:

> 'Some health service activities (e.g. prognosis or diagnosis) may not be best evaluated using the randomised controlled trial approach and other evidence grading taxonomies may be appropriate in these instances' (Eccles and Mason 2001, page 8).

> 'The levels of evidence used by the Australian National Health and Medical Research Council (prior to 2005) for intervention studies, is restrictive for guideline developers, especially where the areas of study do not lend themselves to randomised controlled trials.' (Coleman et al 2005 page 1)

This created a problem for evidence based medicine as there was a risk that large areas of medicine would be considered non-evidence based. How could answers to questions regarding prognosis, diagnosis or appropriateness be considered 'evidence based' if they were not supported by the conclusions of systematic reviews or randomised controlled trials? Evidence based medicine solved this problem by developing different hierarchies for different questions with each individual hierarchy prioritising different study designs:

'The different axes allow for questions related to diagnosis, aetiology and prognosis to be considered as 'evidence-based' as well as traditionally intervention-orientated recommendations'. (Ball and Phillips 2001 page 2)

The development of systems with different hierarchies for different questions had interesting implications for evidence based medicine. Prior to the development of these systems the highest level of evidence could only be provided by evidence from systematic reviews or randomised controlled trials. After the development of these systems a wide variety of different study designs could provide the highest level of evidence. Any intervention, diagnostic test or economic decision could be considered 'evidence based' if it was supported by the results of the study design prioritised by the relevant hierarchy. Importantly, this study design no longer needed to be a systematic review or randomised controlled trial. This effectively changed the definition of 'evidence based' and allowed Ball and Phillips (2001) to claim that the answers to questions of diagnosis, aetiology and prognosis were 'evidence based' when they were not supported by randomised controlled trials or systematic reviews[57].

---

[57] An alternative perspective on the development of multiple hierarchies of evidence can be provided by systems theory (King 2009). Here the different hierarchies represent subsystems which each reinterpret the environment differently using their own rules which are encoded within their hierarchy. From this perspective different hierarchies of evidence are a product of functional differentiation.

The systems with multiple hierarchies consider a variety of different questions but no systems consider exactly the same questions. For example, the Joanna Briggs Institute (2011) presents hierarchies for feasibility, appropriateness, meaningfulness, effectiveness and economic evaluations whereas the Australian National Health and Medical Research Council present hierarchies for treatment, diagnosis, aetiology, screening and prognosis (Merlin et al 2009). It is unclear why many of the systems include or exclude particular hierarchies as little justification is provided. Where justification is provided it is rarely persuasive. For example, Howick et al (2011) included hierarchies for common and rare harms simply because they considered these questions to be clinically relevant and important.

Hierarchies for economic evaluations are particularly interesting as they have been excluded from the latest Oxford Levels of Evidence because of uncertainty over what constitutes good evidence (Howick et al 2011). Further research, in conjunction with economists and policy makers, is advocated before a new hierarchy is developed. This has uncertain implications for any economic evaluations made using the original Oxford Levels of Evidence (Ball and Phillips 2001) and the current Joanna Briggs Institute (2011) hierarchy for economic evaluations.

When the hierarchies that consider questions that do not relate to the effectiveness of treatment interventions, are considered collectively, it is notable that few arguments are provided to support the ranking of different

study designs within them. It is also unclear what property is used to rank different study designs. This is troublesome as these hierarchies provide justification for knowledge claims that are made. These hierarchies may be supported by sound arguments but if these arguments are not presented they cannot be analysed. Some of these hierarchies do appear questionable. For example, the first hierarchy for assessing economic evaluations has been withdrawn (Ball and Phillips 2001, Howick et al 2011) and the Joanna Briggs Institute (2011) included qualitative research in several hierarchies despite claiming that this type of evidence could not be ranked hierarchically:

> 'There is no hierarchy of evidence amongst methodologies for qualitative studies … The synthesis or "pooling" of the findings of qualitative research remains a contested field' (JBI Reviewers' Manual 2011 page 20).

Considerable variation is seen amongst hierarchies that are designed to answer the same question. This can be illustrated if we consider the different hierarchies that can be used to determine the accuracy of diagnostic tests. The highest level of evidence may be provided by a systematic review of studies of test accuracy with an independent blinded comparison and a valid reference standard (Ball and Phillips 2001, Merlin et al 2009); randomised controlled trial (Schunemann et al 2008); validating cohort studies, meta-analysis of validating cohort studies, validated clinical decision rule (Sayre et al 2010) or a systematic review of cross sectional studies with a consistently applied reference standard and blinding (Howick et al 2011). Only

Schunemann et al (2008) provided any justification for the study design that they ranked as the highest level of evidence. They argued that the randomised controlled trial should provide the highest level because diagnosis was only important if effective treatment was available. I would argue that this argument is unsound because accurate diagnosis can provide reassurance and allow planning for the future even if effective treatment is unavailable. None of the other hierarchies provide any justification for the study designs that they rank as the highest level of evidence. If arguments are not presented they cannot be analysed. It is therefore unclear which of the different study designs listed above should provide the highest level of evidence to determine the accuracy of diagnostic tests.

Recognition of the importance of different questions has led to a significant increase in the number of hierarchies to ensure that answers to these questions remain 'evidence based'. Many questions of importance within medicine cannot be answered by systematic reviews or randomised controlled trials and it was realised that answers to these questions could be devalued because they were not supported by evidence from these study designs. This problem was resolved by the creation of new hierarchies that prioritised other study designs. However, these hierarchies lack consistency, address different questions and, where they do address the same question, rank different study designs as the highest level of evidence. These hierarchies are also difficult to analyse as few arguments are presented to support them. When arguments are presented they appear problematic and it

is concerning that one hierarchy of evidence (Ball and Phillips 2001) has been withdrawn.

6.3. Establishment of Professional Jurisdiction:

The second factor that has led to an increase in the number of hierarchies of evidence is the attempt by some professional groups to establish their professional jurisdiction. Most hierarchies rank systematic reviews and randomised controlled trials as the highest level of evidence. However, not all professional groups rely on these study designs to inform their practice. This may be because randomised controlled trials and systematic reviews are unavailable to these groups or because they value other types of evidence such as qualitative research. These professional groups are threatened by evidence based medicine because if they do not engage with hierarchies they risk losing control over their professional domain. Nursing and occupational therapy have reacted to this threat by developing hierarchies that include the types of evidence that they value (Stetler et al 1998, Mitchell and Friese 2007, Armola et al 2009, Tomlin and Borgetto 2011)[58].

Stetler et al (1998) presented the first hierarchy that included qualitative research. This hierarchy did still rank meta-analyses as the highest level of evidence. However, qualitative studies with consistent findings were also

---

[58] Interestingly no hierarchies of evidence appear to have been developed for the dental profession.

able to lead to the strongest treatment recommendation. This group were concerned that many nursing interventions were not supported by the results of meta-analyses and randomised controlled trials:

> 'Neither this language or the routine reliance on large scale randomised controlled trials or meta-analyses was a fit for the division of nursing….In many cases within nursing, researchers have yet to accumulate a sufficient body of knowledge' (Stetler et al 1998 page 47).

Stetler et al (1998) claimed that meta-analyses and randomised controlled trials were impractical, poorly understood and often unavailable within nursing. However, they also recognised that nursing practice may not be funded and could be challenged because it was not supported by these study designs. This potential problem was solved by the introduction of a new hierarchy that was more applicable to nursing.

The hierarchies produced by the Oncology Nursing Society (Mitchell and Friese 2007) and the American Association of Critical Care Nurses (Armola et al 2009) are also aimed at nursing. The Oncology Nursing Society system is the only hierarchy that allows expert opinion to provide the highest level of evidence. The hierarchy proposed by Armola et al (2009) allows qualitative research to provide the highest level of evidence:

> 'Meta-syntheses should be placed at the highest hierarchical level with well-designed meta-analyses, thereby acknowledging the value that

qualitative studies provide about phenomenon of concern to professional nurses' (Armola et al 2009 page 407).

These hierarchies value expert opinion and qualitative research primarily because they are important to nursing (Steelman et al 2011). The second hierarchy produced by the American Association of Critical Care Nurses is actually termed an 'evidence-levelling system' (Armola et al 2009). This suggests that the group have purposefully created a hierarchy that values the forms of evidence that are important to nursing.

Nursing is not the only professional group that has developed novel hierarchies of evidence. The Research Pyramid (Tomlin and Borgetto 2011) was developed for use within occupational therapy because traditional hierarchies were considered to neglect the importance of qualitative research and external validity. Tomlin and Borgetto (2011) claimed that this hierarchy represented:

'A beginning attempt to order evidence-based practice in accordance with the epistemology of the profession.' (Tomlin and Borgetto 2011 page 189)

A consistent theme can be seen underlying the hierarchies that are described in this section. Nursing and occupational therapy would like to establish themselves as evidence-based. However, the treatment interventions that are used by these professional groups are rarely supported by systematic

reviews and randomised controlled trials. Support is instead often provided by expert opinion and qualitative research. This creates a problem when traditional hierarchies are used as treatment interventions may be considered non-evidence based. Nursing and occupational therapy have reacted to this challenge by creating novel hierarchies of evidence that value expert opinion and qualitative research[59].

Not all of the hierarchies of evidence used by the nursing profession have responded to the challenge posed by traditional hierarchies in the same way. Some of these hierarchies still prioritise systematic reviews and randomised controlled trials (Evans 2003, Joanna Briggs Institute 2011)[60]. These hierarchies encourage the nursing profession to undertake systematic reviews and randomised controlled trials in order to become more 'evidence-based'. However, the fact that some of the hierarchies used by the nursing profession still rank randomised controlled trials and systematic reviews as the highest level of evidence does not detract from the fact that other nursing groups have developed novel hierarchies in response to the challenge posed by the modified standard hierarchy of evidence.

---

[59] None of the hierarchies that include qualitative research explain how quantitative and qualitative research can be included in the same hierarchy of evidence. No arguments are presented to support the rank allocated to qualitative research and the property that is used to rank different study designs is unclear.

[60] Both of these systems recognise evidence as more than just evidence of treatment effectiveness and include different hierarchies for different questions.

In this section I have argued that some professional groups have developed novel hierarchies in order to protect professional jurisdiction. This has allowed these professional groups to claim that treatment interventions that are not supported by systematic reviews and randomised controlled trials are still supported by a high level of evidence. This has led to a demonstrable increase in the number of hierarchies and provides support for the claim that factors, unrelated to study design, have influenced the development of hierarchies of evidence.

6.4. Failure to Separate Study Design and Treatment Recommendation:

The third factor that has led to an increase in the total number of hierarchies is a failure to separate study design from other factors that can influence the final treatment recommendation. In order to consider this factor in detail it is important to clarify the distinction between the hierarchy of evidence, grade of recommendation and system that is used in this thesis. The hierarchy of evidence is primarily used to rank evidence according to study design. The grade of recommendation considers other factors, excluding study design, which may influence the final treatment recommendation. The term system is used as an overarching term to include all factors, including study design, which may influence the final treatment recommendation.

The basic principle underpinning a hierarchy of evidence is that certain study designs provide superior evidence to other study designs. However, this is not the only factor that is considered when evidence is evaluated to determine whether a treatment intervention is appropriate. Other factors that may be considered include, but are not limited to, methodological quality, biological plausibility, clinical significance, risks and benefits associated with treatment, applicability of the treatment intervention in a particular setting and financial costs. The importance of both study design and the risks and benefits associated with treatment, to recommendations made by the American College of Chest Physicians is illustrated by the following quote:

> 'Depending on the balance between benefits and risks, methodologically strong studies suggesting a benefit of one agent over a placebo or another agent may lead to a strong recommendation to administer the more effective agent, to conflicting recommendations and practice, or even to recommendations for administration of the less effective agent' (Guyatt et al 1998 page 441S).

Prior to 1998 most systems only presented a hierarchy of evidence. A few of these systems purported to present both a hierarchy and a grade of recommendation (Sackett 1989, Cook et al 1992, Carruthers et al 1993). However, in these systems the hierarchy led directly to the treatment recommendation without consideration of other factors. The grade of recommendation was therefore superfluous. The system proposed by the Canadian Task Force (Spitzer et al 1979) did allow three factors to influence

the final grade of recommendation[61]. However, the effectiveness of the treatment intervention, as determined by the hierarchy of evidence, was considered the most important factor:

> 'Because the effectiveness of treatment or of the preventive measure for a condition was of such importance to the task force, the final recommendation for each condition relied heavily on our assessment of the evidence for effectiveness of treatment. Thus a class A recommendation was rarely made in the absence of grade I evidence' (Spitzer et al 1979 page 1195-1196).

The first system to allow factors other than study design to significantly influence the final treatment recommendation was the system proposed by the Evidence Based Medicine Working Group (Guyatt et al 1995). Since 1998, most systems have allowed multiple factors, in addition to study design, to influence the final treatment recommendation (Greer et al 2000, Harbour and Miller 2001, Ebell et al 2004, Brozek et al 2009, Merlin et al 2009). A clear distinction between study design and other factors is actually valued by some of the later systems (Guyatt et al 1998).

Most of the systems developed prior to 1998 concentrated on study design. They either ignored other factors that could influence the final treatment recommendation or assigned them little significance. This was not because

---

[61] The 3 factors considered by the Canadian Task Force were the burden of suffering, validity and acceptability of the manoeuvre used to detect or prevent the condition and the effectiveness of the treatment intervention.

the developers of these systems did not recognise the importance of other factors. Cochrane (1972) was very clear that treatment provided by the National Health Service should be both cost-effective and beneficial to patients and Trout (1981) discussed different factors which could influence decision making. These early systems simply concentrated on study design and made no attempt to include other factors. This was not wrong but we need to be very clear about what these systems were designed to achieve. These systems could be conceptualised as trying to clarify one of the inputs that influenced treatment recommendations.

As time progressed systems became increasingly complex. They no longer restricted themselves to study design and they tried to account for other factors that could influence the decision-making process[62]. This was unproblematic when study design, as determined by a hierarchy of evidence, was retained as a discrete input, into the decision-making process. However, many systems did not retain study design as a discrete input and other factors became entwined within hierarchies. This led to a significant increase in the total number of hierarchies. This can be illustrated if we consider the ways that different systems have incorporated assessment of methodological quality into the decision-making process.

Methodological quality is a measure of the rigour with which a study is conducted. A study can be well conducted and have high methodological

---

[62] A detailed discussion regarding which factors should be considered and how they may be quantified is beyond the scope of this work.

quality or poorly conducted and have low methodological quality. Methodological quality is independent of study design: we can have a well conducted randomised controlled trial and a poorly conducted randomised controlled trial. The rigour with which any study is conducted is clearly important and should be considered when any evidence is evaluated. Methodological quality may even be the most important factor that should be considered when evidence is assessed.

Although many systems consider methodological quality we must appreciate that this is a problematic concept to measure. It is easy to envisage a study that is perfectly conducted that produces useful conclusions within any limitations imposed by the study design. It is also easy to envisage a study that is so poorly conducted that the conclusions are useless. However, how should we interpret the conclusions of studies that lie between these two extremes? It is unclear how well conducted a study needs to be before the results are useful and how this can be assessed.

It is also important to appreciate that we can only usually make an assumption about methodological rigour based upon the way that a study is reported. Reporting is influenced by the publication process and a well conducted study may be poorly reported whereas a poorly conducted study may be dishonestly reported. Measurement of methodological quality is clearly problematic. However, this is incidental to the present discussion as

we are interested in the way that methodological quality is incorporated into different systems not the way that it is measured.

A number of systems do not consider methodological quality. These systems have been criticised (Hadorn et al 1996, Waugh 1999) although this criticism may be unfair. The developers of these systems may think that methodological quality is unimportant. However, it is more plausible that they have chosen to concentrate on one discrete input into the decision making process or feel that it is self-evident that studies of poor methodological quality studies should be excluded.

Some systems explicitly exclude studies that are deemed to be of poor methodological quality. These systems usually present checklists or stipulate criteria that studies must fulfil before they are considered. In these systems, studies of poor methodological quality are never ranked by the hierarchy of evidence. This approach was used by the American College of Chest Physicians in 1992:

> 'An overview which incorporates low-quality studies is worse than useless, for it may mislead. An overview that is not rigorously conducted adds little to the evaluation of evidence by traditional means and should not be considered' (Cook et al, 1992, page 308S).

Some systems classify methodological quality independently of the hierarchy of evidence. The United States Preventative Services Task Force[63] (Harris et al 2001) categorised methodological quality as good, fair or poor. This information was considered alongside study design, as determined by the hierarchy of evidence, and a number of other factors when the final treatment recommendation was determined. In this system, methodological quality is retained as a discrete input into the decision making process.

Other systems have included assessment of methodological quality within the actual hierarchy of evidence (Hadorn et al 1996, Wilkinson 1999, Ariens et al 2000, Bandelow et al 2002, Ebell et al 2004). In these systems, study design and methodological quality, are no longer discrete inputs into the decision making process. Study design and methodological quality have become entwined. This approach is epitomised by the system proposed by the Agency for Healthcare Research and Quality (Hadorn et al 1996). In this system only randomised controlled trials that are well conducted can attain the highest levels of evidence. Randomised controlled trials with one or more major methodological flaws or three or more minor methodological flaws are relegated to the lower levels.

Some of the systems that combine study design and assessment of methodological quality within hierarchies have focused on particular aspects

---

[63] The hierarchy used by the United States Preventative Services Task Force was excluded from the systematic review reported in Chapter 3 because it was identical to the one used by the Canadian Task Force.

of study design. For example, the Strength of Recommendation Taxonomy (Ebell et al 2004) relegated any study that used surrogate outcomes[64] to the lowest level of evidence:

> 'Recommendations based only on improvements in surrogate or disease-oriented outcomes are always categorized as level C, because improvements in disease-oriented outcomes are not always associated with improvements in patient-oriented outcomes' (Ebell et al 2004).

Different systems have incorporated assessment of methodological quality in different ways. Some systems maintain a clear distinction between assessment of methodological quality and ranking of study design. This has not led to an increase in the number of hierarchies. However, other systems have entwined assessment of methodological quality and study design within their hierarchies. This has directly led to an increase in number of hierarchies because we now have some 'hierarchies of evidence' that consider methodological quality and others that do not. The problem has been exacerbated because methodological quality is problematic to assess and different 'hierarchies' have chosen to focus on different elements of study conduct.

---

[64] Surrogate outcomes do not directly measure how a patient feels, functions or survives and are usually considered to be less valuable than patient orientated outcomes. Surrogate outcomes are often used because they are easier, quicker, or cheaper to achieve, obtain, or monitor than patient orientated outcomes (Lassere, 2008). Richards (2005) has argued that, in some situations, surrogate markers can be directly predictive of patient orientated outcomes. If this argument is accepted studies that employ surrogate outcomes should not be automatically downgraded.

Methodological quality is not the only factor that has become entwined with study design within hierarchies. Some systems have incorporated assessments of clinical significance (Cook et al 1992, Guyatt et al 1995, Brozek et al 2009) and applicability of evidence (Wilkinson 1999, Brozek et al 2009) within their 'hierarchies of evidence'. For example, the first hierarchy proposed by the Evidence Based Medicine Working Group (Guyatt et al 1995) differentiated between studies where the results lay beyond a predetermined threshold number needed to treat and those that did not. Every time a different factor becomes entwined with study design within hierarchies they are likely to increase in number.

Hierarchies of evidence rank the importance of evidence according to the study design that is used to produce that evidence. Methodological quality, clinical significance of results and the applicability of evidence to a target population are clearly important considerations when any evidence is appraised. However, these factors are independent of study design and should not be incorporated within hierarchies when they are designed to rank evidence primarily according to study design. Once these factors become entwined with study design within 'hierarchies' they must increase in number. This is because you have some 'hierarchies' that include factors that are unrelated to study design and other hierarchies that do not. The problem is exacerbated by the fact that some of these factors are problematic to assess and can be incorporated in a variety of different ways. This provides further

support for my claim that factors unrelated to study design have influenced the development of hierarchies of evidence and led to their proliferation.

6.5. Practical Considerations:

The fourth factor that has influenced the development of hierarchies, and led to an increase in their number, is practical considerations. These practical considerations relate to the feasibility of using different hierarchies and specifically exclude theoretical considerations. The potential importance of practical considerations was suggested in Chapter 4 when the theoretical arguments that have been used to support the claim that randomised controlled trials should provide the highest level of evidence were analysed.

In order to understand how practical considerations have influenced the hierarchies of evidence we need to consider the three groups who have developed and used the hierarchies: clinical epidemiologists, guideline developers and clinicians. Hierarchies are developed by clinical epidemiologists; used by guideline developers to develop clinical guidelines; and used by clinicians to treat patients. These groups have different practical requirements and this has led to the development of different hierarchies.

Before practical considerations are considered in detail two important points need to be clarified. Firstly, the nomenclature of the groups is unimportant. It is not my intention to claim that only clinical epidemiologists can develop hierarchies or that only guideline developers can develop clinical guidelines. I only claim that there are three distinct groups that have different practical requirements. The nomenclature simply reflects the individuals traditionally associated with each of these groups. Secondly, an individual can belong to each group at different times. Indeed, many proponents of evidence based medicine have occupied each of these groups at different times. Individuals often belong to different social groups at different times and the fact that an individual can belong to each of the groups is not incompatible with the claim that different practical considerations are important to each of the three groups.

One group that is conspicuously absent from this discussion is patients. However, none of the hierarchies have been specifically designed for interpretation by patients. Atkins et al (2004) recognised that existing hierarchies were unsuitable for patients during the development of the GRADE system but this finding was not addressed.

Clinical epidemiologists are often concerned by the time and resources that will be required for the implementation of a hierarchy and the type of studies that are likely to be available. The hierarchy proposed by the United States Community Preventative Services Task Force ranked controlled clinical trials

without randomisation as the highest level of evidence even though randomised controlled trials were considered to provide superior evidence (Briss et al 2000). This was because randomisation was considered impractical in population based research:

> 'Randomisation is sometimes not feasible or ethical in population based research. Also, group randomised trials often cannot feasibly randomise sufficient numbers of units to ensure even distribution of potential confounders among groups' (Briss et al 2000 page 41).

The developers of this hierarchy have clearly allowed practical considerations to determine hierarchical position and as a consequence we have a new hierarchy.

Practical considerations have also influenced the development of other hierarchies. The EUR-ASSESS system (Granados et al 1997) allows quasi-experimental studies to attain the highest level because randomised controlled trials are considered too expensive and often impossible to conduct. The Scottish Intercollegiate Guideline Network (2011) openly acknowledge that their system had been influenced by time and resource requirements and Ariens et al (2000) include cross-sectional studies, despite reservations about this study design, because lots of these studies have been published:

> 'In cross-sectional research the temporal relationship between exposure and outcome, and thus causality, cannot be firmly established. The reason cross-sectional studies were included in this

review, despite this disadvantage, was that most research on risk factors for neck pain was actually based on a cross-sectional design. It would not have been acceptable to neglect the vast amount of information obtained from cross-sectional research' (Ariens et al, 2000, page 15-16).

These examples demonstrate that clinical epidemiologists must balance theoretical and practical considerations when hierarchies are developed. If hierarchies are developed using only theoretical considerations they may have no utility. In these circumstances clinical epidemiologists may choose to give greater weight to practical considerations. However, deciding whether different study designs are practical or impractical involves a value judgement. Value judgements are not true or false in the same way that premises are. In fact value judgements vary depending upon any value commitments that are held. Different clinical epidemiologists may therefore consider a study design impractical because it is impossible, difficult to undertake or simply requires significant resources. Once practical considerations are able to influence hierarchical position it is therefore highly likely that hierarchies will increase in number because clinical epidemiologists will have different value commitments.

The two other main groups that use hierarchies are guideline developers and clinicians. Both of these groups use hierarchies to answer questions about the effectiveness of treatment interventions. However, although we should be

wary about making sweeping generalisations, these two groups often have significantly different requirements. Clinical guidelines are usually developed by large groups of individuals comprising topic experts and experts in research methods. These groups do not have limitless time and resources but they can understand and utilise complicated systems and usually have sufficient time to consider the evidence base in detail. Clinicians are usually individuals or small groups of individuals working in clinical teams. These individuals are often topic experts, but they have limited time and may lack expertise in research methods. Evidence based clinical pathways have been developed to address these problems for clinicians but they are only available for common conditions and cannot cover every possible clinical situation.

Several hierarchies have clearly been designed to help clinicians make decisions about the care of individual patients (Greer et al 2000, Bandelow et al 2002, Howick et al 2011). Howick et al (2011) reported that the Oxford Levels of Evidence were revised because the original hierarchy was too complicated and did not reflect clinical decision making. The new system was described both as a hierarchy and a heuristic:

> 'The Levels should be designed in a way that could be used as a search heuristic for busy clinicians and patients[65] in real time in addition to serving as a hierarchy of evidence'. (Howick et al 2011 page 2)

---

[65] This is the only time that patients are mentioned as a group who may wish to use hierarchies.

Heuristics can be defined as processes that lead to conclusions that are only usually true because they are not supported by underlying theory (Blunt 2015). If heuristics produce conclusions that are only usually true they must sometimes produce false conclusions. This is accepted by Howick et al (2011) but they claim that their hierarchy is still valuable because it provides guidance to busy clinicians who need to make rapid decisions about the care of individual patients.

The World Federation of Societies of Biological Psychiatry hierarchy (Bandelow et al 2002) was developed because existing hierarchies did not make a distinction between inconsistent evidence, negative evidence and the absence of evidence. This distinction was considered important because a drug without supporting evidence could still reasonably be used in a patient unresponsive to standard treatment whereas it would be unreasonable to use a drug that was known to be ineffective (Grunze et al 2009). This hierarchy can also be viewed as a heuristic because it is unlikely that a drug without supporting evidence will be effective in all patients.

The hierarchies proposed by Bandelow et al (2002) and Howick at el (2011) have clearly been developed as practical tools to guide clinical decision making at the individual patient level. This demonstrates that the practical needs of clinicians have resulted in the development of further hierarchies. It is interesting that these hierarchies have been categorised as heuristics because this acknowledges that conclusions derived using an 'evidence-

based' process may be false. The use of hierarchies as heuristics will be considered in more detail in the next chapter when the rationality of decision making within evidence based medicine is considered.

Other hierarchies of evidence are very clearly aimed at guideline developers. This is exemplified by the GRADE system (Brozek et al 2009). This system values transparency and explicitness and seeks to standardise every stage of the evidence assessment process. This has resulted in a complicated system that has been adopted by some groups such as the World Health Organisation and the Cochrane Collaboration (Guyatt et al 2008) but rejected by other groups because it is time, training and labour intensive (Richards 2009, Howick et al 2011, Scottish Intercollegiate Guideline Network 2011).

Guideline developers often use the GRADE system because they value a transparent, comprehensive approach to guideline development. It has been suggested that the GRADE system is more accurate than other systems (Howick et al 2011) although no theoretical or empirical evidence has been provided to support this claim. Empirical evidence actually suggests that the GRADE system is difficult to apply and has poor reproducibility (Atkins et al 2005, Jaeschke et al 2008, Gillis et al 2011) although this does not mean that

the GRADE system cannot produce accurate conclusions if it used by experienced guideline developers[66].

Guideline developers and clinicians appear to have very different requirements of hierarchies. Guideline developers value accuracy and are prepared to use exhaustive, complicated processes to reach an accurate conclusion. Clinicians value heuristics and are prepared to accept some false conclusions if the process can be simplified and accelerated. It is only natural that these competing requirements have led to the development of very different systems resulting in an increase in the overall number of hierarchies of evidence.

In this section I have argued that hierarchies of evidence have increased in number because clinical epidemiologists, guideline developers and clinicians have different practical requirements. Clinical epidemiologists prioritise the study designs which they consider most practical but this allows value judgements to influence hierarchical position. Guideline developers prefer more complex systems as these are believed to produce more accurate conclusions whereas clinicians prefer heuristics, which allow rapid decision-making, but are more likely to produce false conclusions. A number of examples have been provided to support my argument. This provides further justification for my claim that factors, unrelated to study design, have

---

[66] Atkins et al (2005) reported that inter-examiner agreement with the GRADE system was worse than that expected by chance alone although this did not prevent them concluding that the GRADE system was clear, understandable and superior to existing systems.

influenced the development of the hierarchies and led to an increase in their number.

6.6. Standardisation:

The final factor that has influenced the development of hierarchies is standardisation. In this chapter a number of factors, independent of study design, that have influenced the development of hierarchies of evidence have already been identified. These factors have led to a demonstrable increase in the number of hierarchies. This proliferation of hierarchies has not gone unnoticed within evidence based medicine and concerns have been expressed that the situation is confusing:

> 'Journal readers do not have the time, energy, or interest to interpret multiple grading scales, and more complex scales are difficult to integrate into daily practice' (Ebell et al 2004, page 548).

> 'Guideline developers use a bewildering variety of systems to rate the quality of evidence underlying their recommendations. Some are facile, some confused, and others sophisticated but complex' (Guyatt et al 2008, page 995).

Several hierarchies have been developed with the stated aim of reducing the confusion created by multiple hierarchies of evidence. The developers of these hierarchies aim to create a universal hierarchy that can be used by everybody.

There has been a discussion in the evidence based medicine literature about whether a universal hierarchy of evidence should be complex or simple (Guyatt et al 1998, Ebell et al 2004, Atkins et al 2004). This discussion revolves around the factors that different groups feel should be included. Some groups advocate complex hierarchies that explicitly include many factors (Atkins et al 2004). Other groups advocate simpler hierarchies that consider fewer factors and allow implicit judgement to play a greater role (Guyatt et al 1998, Ebell et al 2004). If the different groups are unable to agree on the factors that should be considered attempts to create a universal hierarchy of evidence are likely to fail. As clinical epidemiologists, guideline developers and clinicians have different requirements it may be impossible to reach agreement. Attempts to create a universal hierarchy of evidence may lead to convergence on fewer hierarchies but it may also lead to an overall increase in number. This is because any new 'universal' hierarchies are only likely to appeal to a limited number of users.

Some groups do not aspire to create a universal hierarchy but do aim to ensure that any new hierarchy is compatible with previous hierarchies (Baumann and Gutterman 2006, Merlin et al 2009, Morley et al 2010). The second Australian National Health and Medical Research Council hierarchy was very clearly designed to ensure that it was compatible with the previous version:

'The objective of the first stage was to create a framework that aligned as closely as possible with the original evidence hierarchy – to minimise confusion for current users and maintain consistency with previous use of the hierarchy'. (Merlin et al 2009 page 3)

However, maintaining consistency with previous hierarchies can create tension. For example, Merlin et al (2009) expressed concern about the quality of evidence provided by systematic reviews because the studies included in any systematic review could be of poor quality. Despite this concern, systematic reviews maintained their place as the highest level of evidence to maintained consistency with the previous hierarchy.

It is interesting to consider why developers of new hierarchies seek to maintain consistency with older hierarchies. A new hierarchy is surely only needed when there is a problem with the old hierarchy. If there is a problem with the old hierarchy it is unclear why you would want to maintain consistency as this may perpetuate any failings. It is possible that developers of new hierarchies seek to maintain consistency so that clinical guidelines and treatment decisions made using older hierarchies remain justified. However, as existing hierarchies are significantly different, if new hierarchies are continually developed to maintain consistency with previous versions, they are unlikely to converge on a universal version.

A universal hierarchy of evidence is appealing because it would resolve the confusion that currently exists amongst hierarchies. However, attempts to

create a universal hierarchy are unlikely to be successful because there is disagreement about which factors should be included and different users have different requirements. Many newer hierarchies have been developed to maintain consistency with previous versions. However, this may perpetuate any failings and ensures that hierarchies remain dissimilar. These attempts at standardisation have therefore led to the development of yet more hierarchies. This provides further evidence to support my claim that factors unrelated to study design have led to an increase in the number of hierarchies of evidence.

6.7. Why have different Factors been able to Influence Hierarchies of Evidence?

The analysis presented in this chapter has demonstrated that five factors, unrelated to study design, have had a significant influence on the development of hierarchies of evidence. These factors are recognition of the importance of different questions; attempts to establish professional jurisdiction; failure to separate study design from other factors that can affect treatment recommendations, practical considerations and standardisation. These factors have been analysed in detail and it is clear that they have all repeatedly influenced the development of different hierarchies. The examples presented provide strong support for the claim that factors unrelated to study design have caused an increase in the total number of hierarchies of evidence. It is not my intention to claim that these are the only factors that

have influenced the development of hierarchies and other factors may also be important.

Hierarchies of evidence are designed to rank the importance of evidence primarily according to study design (Upshur 2009). It is therefore interesting to consider why factors that are independent of study design have come to exert such a significant influence. It is my contention that the lack of theoretical or empirical support for the standard and modified standard hierarchies of evidence has allowed factors that are independent of study design to assume greater importance. This is because these hierarchies are unable to provide a normative role as they are undermined by their lack of theoretical and empirical support. This argument is summarised below:

<u>The Normative Argument:</u>

The standard and modified standard hierarchies of evidence lack theoretical or empirical support.

If the standard and modified standard hierarchies of evidence lack theoretical or empirical support they will be influenced by factors that do not relate to study design.

Hierarchies of evidence are influenced by factors that do not relate to study design.

Throughout this chapter it has been argued that a number of different factors, independent of study design, have influenced the development of hierarchies. I would therefore claim that the conclusion of the normative argument is true. The normative argument also has a valid argument form. However, this does not necessarily mean that the argument is sound as this would make the mistake of affirming the consequent (Smith 2003)[67]. As with all arguments, we need to consider the truth of the premises in order to determine whether the argument is sound.

The first premise claims that the standard and modified standard hierarchies of evidence lack theoretical or empirical support. The arguments that have been used to support these hierarchies were considered in Chapter 4. These arguments were shown to be unsound, restricted in scope or lacking empirical support. The property used to rank different study designs within hierarchies was considered in Chapter 5. It was argued that hierarchies could potentially be ordered to reflect increasing confidence in the conclusions of certain study designs but it was unclear why we should have greater confidence in the conclusions of systematic reviews and randomised controlled trials. If these arguments are accepted the first premise of the normative argument is true.

---

[67] Modus ponens is a valid argument form: if A then C, A hence C. Affirming the consequent is an invalid argument form: if A then C, C hence A. This is because A is only a sufficient condition for C (Smith 2003).

The second premise claims that if hierarchies lack theoretical or empirical support they will be influenced by factors that are independent of study design. This is because they cannot provide a normative standard on theoretical or empirical grounds. If there was strong theoretical or empirical support for the standard or modified standard hierarchy of evidence any hierarchy that did not conform to this structure could be rationally rejected. For example, it would be difficult to justify a hierarchy that ranked expert opinion as the highest level of evidence if there were sound theoretical reasons to prefer evidence derived from systematic reviews and randomised controlled trials. Any decision that was made using a non-conforming hierarchy could therefore be rejected. However, if the standard and modified standard hierarchies of evidence lack theoretical and empirical support it is more difficult to challenge alternative hierarchies. This is because hierarchies that do not conform can be justified by appealing to the arguments that undermine the standard and modified standard hierarchies of evidence. As a consequence there is greater scope for hierarchies to be influenced by factors that are independent of study design.

The second premise of the normative argument could be challenged in two ways. First, it could be claimed that a normative standard does not necessarily require theoretical or empirical support and could be supported by practical, ethical or other reasoning. This means that hierarchies could resist the influence of other factors even if the normative standard lacked theoretical or empirical support. In order for a particular hierarchy to act as a normative standard it would simply need to be agreed that study designs

should be ranked in a particular order. However, I would argue that the standard and modified standard hierarchies of evidence should not perform a normative role without theoretical or empirical support. This is because hierarchies are fundamental to decision-making within evidence based medicine and these decisions have important implications. The hierarchies should therefore have strong theoretical or empirical support. It is acknowledged that this counter-argument involves a value judgement about the way that decisions within medicine should be justified. If this value judgement is not shared the reader will not be persuaded by the normative argument.

The second challenge to the normative argument is the claim that factors, independent of study design, would have influenced hierarchies regardless of the strength of theoretical and empirical support for the standard and modified standard hierarchies of evidence. However, I would argue that it would be unlikely for there to be so many different hierarchies if the normative standard had strong theoretical or empirical support. This counter-argument involves a similar value judgement about the importance of theoretical and empirical support for hierarchies. If this value judgement is not shared, the reader will not be persuaded that the normative argument can explain the variation that is seen amongst hierarchies.

The normative argument has attempted to explain how factors that are independent of study design have been able to influence the development of

hierarchies. This argument is problematic because it requires us to make a value judgement about the importance of theoretical and empirical support for any normative hierarchy of evidence. This means that the second premise is not necessarily true and the argument may be unsound. This problem is acknowledged and a lack of theoretical or empirical support is advanced as a possible explanation for the influence that factors, independent of study design, have had on hierarchies of evidence. The reader may be persuaded by the normative argument if they share the value judgements that have been highlighted.

6.8. Conclusion:

The analysis presented in this chapter has sought to explain the variation amongst hierarchies of evidence. Whilst hierarchies of evidence are designed to rank the importance of evidence primarily according to study design, they have also been influenced by a variety of others factors. Five such factors have been considered in depth in this chapter. Each of these factors has led to a demonstrable increase in the overall number of hierarchies of evidence. The influence of these factors can explain some of the variation that we see amongst the various hierarchies.

Having demonstrated that a number of factors have influenced the development of hierarchies, a possible explanation for this has been

advanced. There is a lack of theoretical and empirical support for the standard and modified standard hierarchies of evidence. This has prevented them from fulfilling a normative role and allowed factors that are independent of study design to assume greater importance. This claim is not advanced as the only possible explanation for the variation that is seen amongst hierarchies. However, the reader may be persuaded that it is a probable explanation if they share underlying value judgements about the importance of theoretical and empirical support for a normative hierarchy of evidence.

Hierarchies of evidence are fundamental to decision making within evidence based medicine. However, they are numerous and have been influenced by a number of different factors that are independent of study design. There is a possibility that different hierarchies could be used to support different conclusions following consideration of the same evidence. This would mean that the decision-making process within evidence based medicine was inconsistent. This is considered in detail in the next chapter when I analyse the rationality of evidence based medicine.

Chapter 7: The Rationality of Evidence Based Medicine

7.1. Introduction:

Evidence based medicine can use a number of different hierarchies of evidence to support decision-making. These hierarchies have limited theoretical support and have been influenced by a number of factors that are independent of study design. It is therefore possible that different hierarchies could be used to support different conclusions following consideration of the same evidence. This would mean that the decision making process used by evidence based medicine was not rational. If this is the case then decisions about patient care, healthcare funding, medicolegal standards and medical research may be made inconsistently. This would mean that patients with the same medical condition could receive inconsistent clinical care, healthcare professionals could be judged against inconsistent medicolegal standards, scarce healthcare resources could be allocated inconsistently and research priorities could be judged inconsistently. It is therefore important to consider whether the decision making method used by evidence based medicine is rational.

The decision making process used by evidence based medicine can be considered rational if conclusions are derived in a consistent way given any information that is available for consideration (Newton-Smith 1981). A

number of different types of rationality are described in the literature including practical rationality, theoretical rationality, formal rationality and substantive rationality (Kalberg 1980, Brubaker 1984, Audi 2001, Abulof 2015). The type of rationality that is used can have a significant influence on the rationality of any decision making process. This is particularly important when substantive rationality is considered as this is influenced by value commitments. It is therefore important to understand the concept of rationality in greater depth before the rationality of decision making within evidence based medicine is analysed.

In this chapter we will analyse the rationality of decision making within evidence based medicine. The concept of rationality will be elucidated and the importance of different types of rationality will be highlighted. Decision making within evidence based medicine will then be explored using a framework that acknowledges the importance of hierarchies. This framework will be developed as the chapter progresses. The existence of multiple hierarchies suggests that decision making within evidence based medicine may be irrational. This is because the use of different hierarchies may result in different conclusions following consideration of the same information. However, it will be argued that decision making within evidence based medicine can still be rational when it is considered from the perspective of substantive rationality. This has interesting implications for evidence based medicine because substantive rationality is dependent upon value commitments.

7.2. Rationality:

In order to analyse whether the decision making process that is used by evidence based medicine is rational we need to understand the concept of rationality. Reasoning can be considered rational if conclusions are derived in a consistent way given any information that is available for consideration (Newton-Smith 1981). Reasoning that is rational does not have to produce conclusions that are true or provide justification for any knowledge claims that are made (Audi 2001). This means that deductive reasoning, inductive reasoning and inference to best explanation can all be rational reasoning processes. Deductive reasoning is always rational because conclusions are necessarily true (Annas 2000). Inductive reasoning and inference to best explanation do not produce conclusions that are necessarily true, but these reasoning processes can still be rational, and often are rational, if conclusions are derived in a consistent way (Audi 2001).

The inclusion of inductive reasoning and inference to best explanation demonstrates that a rational reasoning process can produce false conclusions. This does not mean that a reasoning process that always produced false conclusions would be rational, even if the conclusions were derived in a consistent way. This is because any reasoning process should aim to produce true conclusions. It does mean that a decision making process that occasionally produced false conclusions could still be rational.

Evidence based medicine would therefore not necessarily be irrational if the decision making process, when correctly implemented, occasionally resulted in the use of the wrong treatment intervention.

A number of different types of rationality have been described in the literature including practical rationality, theoretical rationality, formal rationality and substantive rationality. The different types of rationality were originally identified by Max Weber (Kalberg 1980, Brubaker 1984, Audi 2001). It is important that the type of rationality under consideration is clearly defined because different types of rationality have significantly different meanings. A reasoning process that is rational from the perspective of formal rationality may be irrational from the perspective of substantive rationality and vice versa.

Practical rationality is a manifestation of means-ends rational action and considers whether ones actions are consistent with ones reasons for action. Theoretical rationality provides explanation and considers whether ones beliefs are consistent with ones reasons for believing (Audi 2001). Formal rationality is indifferent to value commitments and is exemplified by the blind following of rules with no consideration of the final outcome. Substantive rationality emphasises the importance of value commitments. An action or belief can only be considered substantively rational if it is consistent with underlying value commitments (Etzioni 1988, Kalberg 1980, Brubaker 1984).

Substantive rationality may be guided by individual values or clusters of values. These value commitments are essentially overarching principles that guide the reasoning process and they may influence both the desired end and the means through which that end is achieved (Etzioni 1988). Substantive rationality allows a relativist interpretation of rationality where the same actions or beliefs can be rational or irrational depending upon the value commitments that are held (Brubaker 1984). For example, a decision about the allocation of financial resources may be considered substantively rational or irrational depending upon whether you value maximising utility for the individual or the population. Value commitments are not true or false and rationality cannot be used to decide between value commitments that are followed in a substantively rational way (Audi 2001).

Irrespective of the type of rationality that is considered, the conclusions of any reasoning process are dependent upon the amount of information that is available for consideration (Newton-Smith 1981). Abulof (2015) has differentiated between minimal rationality, maximal rationality and bounded rationality to reflect information availability. Minimal rationality signifies calculative intentionality in the absence of any information. Maximal rationality requires complete information and produces conclusions that are necessarily true. Deductive reasoning is an example of a reasoning process with maximal rationality (Kukla 1991). Bounded rationality sits between these two extremes. Bounded rationality accepts that information is often incomplete and that the conclusions of reasoning processes are not necessarily true. Inductive reasoning and inference to best explanation are

examples of reasoning processes that use the concept of bounded rationality.

When information is incomplete, it may be possible to collect additional information to further inform decision making. However, additional information does not have to be collected for the reasoning process to be rational. The potential value of the additional information, the ease with which it can be collected, and time constraints, are important factors that are considered prior to deciding whether additional information should be collected (Kukla 1991, Abulof 2015). It may therefore be rational to make a decision based on information that is available, when collection of further information would be too time consuming, even if that further information, if collected, would lead to a different decision[68].

Maximal and minimal rationality are important concepts but they have limited usefulness when we consider the rationality of decision making within evidence based medicine. Maximal rationality requires complete information and conclusions are necessarily true. However, decisions about medical care are rarely guided by complete information[69]. If medical decision making was

---

[68] This type of situation can occur in medicine when a patient has an acute infection such as infective endocarditis. Blood cultures are required to identify the causative microorganism but these take time, and immediate treatment is required, so antibiotic therapy is initiated to target the most likely microorganisms. The antibiotic therapy may subsequently be changed once the results of the blood cultures are obtained (Habib et al 2015).

[69] Maximal rationality may exist when a treatment intervention is supported by an 'all-or-none' study, where all patients survive when treated with a particular treatment intervention when they all previously died. This would be a rare situation and requires acceptance of a number of background assumptions.

informed by complete information there would be little debate about the most appropriate treatment intervention. Minimal rationality is also not useful as we need to understand how information is processed by evidence based medicine to derive conclusions. We must therefore presume calculative intentionality[70]. The rationality of decision making within evidence based medicine should therefore be considered within the context of bounded rationality.

Rationality may pertain to an individual, a reasoning process or a particular decision. When rationality is considered it is essential that the subject that is analysed is clearly defined (Abulof 2015). This chapter considers whether the reasoning process that underpins decision making within evidence based medicine is rational. This chapter does not consider whether it is rational for individual clinicians to practice evidence based medicine. This would not be an interesting question because evidence based medicine is so widely accepted within medicine. It could therefore be argued that any individual who practiced evidence based medicine was simply conforming to the prevailing culture of the medical profession. If an individual did not practice evidence based medicine it is likely that their clinical decisions would be challenged, they may fall short of medicolegal standards of care and they would struggle to attain professional status or attract research funding. It could therefore be argued that it was rational for an individual clinician to practice evidence based medicine even if the underlying decision making process was irrational.

---

[70] Abulof (2015) has equated minimal rationality with human nature.

In this section we have explored some ideas about the concept of rationality to provide a foundation for the analysis that is presented in later sections. The importance of different types of rationality, information availability and defining the subject of analysis has been recognised. In the rest of this chapter the rationality of the decision making process that is used within evidence based medicine is analysed from the perspective of different types of rationality. In order to move forwards with this analysis we next need to understand in greater detail the decision making process that is used by evidence based medicine.

7.3. Decision Making in Evidence Based Medicine:

In order to analyse the rationality of any decision making process we need to understand the framework within which decisions are made. Newton-Smith (1981) used a framework to analyse the rationality of scientific method in *'The Rationality of Science'*. This framework attributed an aim, method and goal to any decision making process. The aim provides direction to the decision making process and the process is considered rational if the goal is consistently achieved when the method is used. Using this framework, Newton-Smith (1981) argued that science was rational because scientific method (the method) could be used to decide between competing scientific theories (the aim) to select scientific theories with increasing verisimilitude

(the goal). This framework can also be used to understand the decision making process that is used by evidence based medicine.

The framework developed by Newton-Smith (1981) requires us to attribute an aim, method and goal to evidence based medicine. Evidence based medicine uses hierarchies of evidence to decide between different treatment interventions in order to improve patient care. Therefore the aim of evidence based medicine is to decide between different treatment interventions; the method uses hierarchies of evidence and the goal is improved patient care. Using this framework evidence based medicine can be considered rational if patient care is consistently improved when hierarchies are used to decide between different treatment interventions.

At this stage, the reader needs to understand that the framework has been used to present a simplified model of the decision making process that is used. This model will be expanded in the next section. It is assumed that the aim ascribed to evidence based medicine is important and uncontroversial. The aim will therefore not be considered further. The method ascribed to evidence based medicine may be considered an oversimplification and will be elaborated upon as the chapter progresses. The goal of evidence based medicine is not clearly defined in the literature and the ascribed goal may be considered controversial. However, it is not possible to determine whether any decision making process is rational unless a goal is defined. The importance of alternative goals will be explored as the chapter progresses.

7.4. The Rationality of the Decision Making Process used by Evidence Based Medicine:

Using our framework, decision making within evidence based medicine can be considered rational if patient care is consistently improved when hierarchies of evidence are used to decide between treatment interventions. However, a number of different hierarchies can be used within evidence based medicine. The implication is that different hierarchies may produce different conclusions, following consideration of the same information, and this may influence the treatment intervention that is provided to patients. If there is a 'best' treatment intervention, and the treatment intervention that is provided to patients is dependent upon the hierarchy that is used, patient care may not be consistently improved. It could therefore be argued that decision making within evidence based medicine is not rational:

Hierarchies of evidence are central to decision making within evidence based medicine.

Different hierarchies of evidence can be used to inform decision making.

Different hierarchies of evidence produce different conclusions following consideration of the same information.

If different conclusions are produced following consideration of the same information patient care is not consistently improved.

If patient care is not consistently improved decision making within evidence based medicine is not rational.

Decision making within evidence based medicine is not rational

This is a valid argument form that builds on the analysis that has been presented in previous chapters. We therefore need to consider the truth of the premises to determine if the argument is sound. It is not my intention to argue that decision making within evidence based medicine is not rational. What I wish to explore is under what conditions decision making within evidence based medicine *can be* rational. What we shall see is that the soundness of the argument depends upon the type of rationality that is used.

The first premise of this argument claims that hierarchies of evidence are central to decision making within evidence based medicine. This does not mean that hierarchies are the only determinant of the conclusions that are produced by evidence based medicine. This would be a misrepresentation of the decision making process. It is acknowledged that other factors, specifically clinical experience and patient preferences, are also of fundamental importance to the decision making process. This is repeatedly emphasised by the definitions of evidence based medicine that appear in the literature (Sackett et al 1996, Straus et al 2011). However, decision making within evidence based medicine is not characterised by consideration of clinical experience and patient preferences. These factors were part of the medical decision making process prior to the introduction of evidence based medicine and are important to decision making within contemporary alternative medicine. It is the use of hierarchies of evidence that characterise decision making in evidence based medicine[71].

The second and third premises of this argument claim that different hierarchies of evidence can be used by evidence based medicine and that use of different hierarchies will produce different conclusions following consideration of the same information. Collectively, the hierarchies allow a variety of different study designs to provide the highest level of evidence. Randomised controlled trials, observational studies and expert opinion can provide both the highest and lowest level of evidence in hierarchies. If the

---

[71] The importance of the hierarchies of evidence to evidence based medicine is considered in greater detail in Chapter 9 when the claim that evidence based medicine is a new paradigm is analysed.

same study designs can provide both the highest and lowest levels of evidence the choice of hierarchy may lead to a different conclusion when a body of evidence is considered. Hierarchies that prioritise randomised controlled trials, observational studies or expert opinion can produce very different conclusions following consideration of the same information.

It is important to appreciate that use of different hierarchies will not necessarily lead to different conclusions following consideration of the same information. Where the results of different study designs and expert opinion all support the same treatment intervention the selected hierarchy is unlikely to affect the conclusion. However, a hierarchy becomes redundant when the results of different study designs all support the same treatment intervention. In these situations the treatment intervention that should be provided will be obvious. The hierarchy is far more important where the results of different study designs and/or expert opinion are conflicting. For example, a situation where expert opinion suggests we should use intervention A whereas evidence from cohort studies suggests that we should use intervention B. In this situation, the choice of hierarchy can very clearly alter the conclusion.

The systematic review presented in Chapter 3 identified 42 hierarchies that can be used to guide medical decision making. These hierarchies have variable levels and rank study designs differently. Similarities do exist between some hierarchies, illustrated by the concept of the standard hierarchy of evidence. However, even hierarchies that rank randomised

controlled trials as the highest level of evidence are often significantly different and may reach different conclusions following consideration of the same information. This is because some of these hierarchies rank all randomised controlled trials as the highest level of evidence whereas others downgrade randomised controlled trials when they are methodologically flawed, inadequately powered, or produce results that are not clinically significant.

It could be argued that some of the hierarchies of evidence are historical and have been superseded by later versions. For example, the hierarchy used by the American College of Chest Physicians was modified in 1992, 1998 and 2001 before it was finally abandoned in 2006 in favour of the GRADE system (Baumann and Gutterman 2006). However, even if older hierarchies are discounted, users of evidence based medicine are still able to choose between a large number of hierarchies at any time. These hierarchies are all different and may produce different conclusions following consideration of the same information.

Discounting historical hierarchies of evidence may be considered a charitable approach[72]. This is because there is limited theoretical support for any individual hierarchy. Although many earlier hierarchies have been superseded by later hierarchies there is limited justification for this. If there is

---

[72] When using analytical philosophy arguments should be presented in the most favourable light to avoid accusations of attacking a straw man (Walton 2008). This does not mean that the most favourable presentation is the correct or intended presentation.

no theoretical justification for later hierarchies it may be considered legitimate to use any hierarchy regardless of whether it has been superseded by a later version.

The fourth premise claims that if different conclusions are produced following consideration of the same information patient care is not consistently improved. Different hierarchies have the potential to produce different conclusions following consideration of the same information but this may not happen in practice. No empirical research appears to have been undertaken in this area which suggests that inconsistent conclusions are not perceived as a problem. It is therefore important to consider why inconsistent conclusions are not perceived as a problem as this suggests that the fourth premise may be false.

Different hierarchies may produce different conclusions, following consideration of the same information, but this may not lead to different treatment recommendations. This is because a number of other factors are considered, including methodological quality, biological plausibility and clinical significance, before treatment recommendations are made, as part of the grading process[73]. This way the conclusions of study designs valued by particular hierarchies may be upgraded or downgraded based upon methodological rigour, biologically plausibility or clinically significance in some systems. This allows the grading process to effectively iron out any

---

[73] The distinction between the hierarchy of evidence and the grading process was considered in Chapter 6.

inconsistency created by the use of different hierarchies. If the grading process ensures consistent treatment recommendations, even when different hierarchies are used, the decision making process used by evidence based medicine may still be rational from the perspective of practical rationality.

The grading process may ensure consistent treatment recommendations when different hierarchies are used despite the different hierarchies supporting different conclusions. However, this does not mean that decision making is rational from the perspective of practical rationality. This is because evidence based medicine is characterised by the use of hierarchies of evidence, not other factors that are considered during the grading process. Decision making would only appear rational, from the perspective of practical rationality, because inconsistency created by using different hierarchies was corrected by non-characteristic features of evidence based medicine. If the characteristic feature of evidence based medicine does not ensure consistent conclusions how can decision making be practically rational? I would actually argue that decision making within medicine (not evidence based medicine), is rational from the perspective of practical rationality, despite the use of hierarchies by evidence based medicine. The claim that the fourth premise is false because the grading process irons out the inconsistency created by using different hierarchies is therefore problematic for evidence based medicine.

Although a number of different hierarchies of evidence are available to guide decision making some hierarchies are more frequently used that others. For example, the Oxford Levels of Evidence (Ball and Phillips 2001) have been commonly used in dentistry and the GRADE system is commonly used by clinical epidemiologists (Guyatt et al 2008). If the same hierarchy is always used conclusions may be consistent because the same information is processed in the same way[74]. The fourth premise may therefore be false because only a few of the hierarchies are regularly used.

The analysis presented in Chapters 4 and 5 revealed that the hierarchies of evidence have limited theoretical support. Therefore although the Oxford Levels of Evidence (Ball and Phillips 2001) and the GRADE system (Guyatt et al 2008) are commonly used, there are no strong theoretical grounds for this decision. If there are no strong theoretical grounds for preferring one hierarchy over others, decision making cannot be considered rational from the perspective of theoretical rationality. It would therefore appear difficult for proponents of evidence based medicine to claim that the fourth premise is false because theoretical arguments dictate the use of particular hierarchies.

Although there are limited theoretical grounds to support the use of particular hierarchies they may be chosen because they reflect the value commitments of users. Value commitments are central to the concept of substantive

---

[74] Interestingly, several studies have suggested that individual hierarchies have poor reproducibility and can produce different conclusions when the same body of information is evaluated by different assessors (Atkins et al 2005, Turpen et al 2010). In these studies the poor reproducibility is attributed to misuse or misunderstanding of the hierarchies by the assessors not the hierarchies themselves.

rationality. In Chapter 6 a number of factors that have influenced the development of hierarchies were identified. Many of these factors are underpinned by value commitments. For example, the nursing profession uses hierarchies that value qualitative research and expert opinion; guideline developers use complex hierarchies that value transparency and clinicians favour simpler hierarchies because they value the ability to make rapid decisions. Different groups of users may therefore always use the same hierarchy because it is consistent with their value commitments. If the same hierarchy is always used, based upon value commitments, conclusions should not vary. This means that decision making within evidence based medicine may be rational from the perspective of substantive rationality.

Substantive rationality allows decision making within evidence based medicine to be interpreted from a relativistic perspective. This means that patient care can still be consistently improved when different conclusions are derived using different hierarchies. This is because different conclusions can legitimately be reached, with different implications for patient care, following consideration of the same information. As long as the treatment that is recommended is consistent with the value commitments of the user, patient care is consistently improved. I would therefore argue that the fourth premise is false and the irrationality argument is unsound.

Substantive rationality may explain why there are so many different hierarchies of evidence. The problem is however that on this interpretation

there would be no normative hierarchy of evidence and it is legitimate to use any hierarchy as long as it corresponds to the value commitments that are held. Individuals do not therefore use a hierarchy in an instrumental sense to conform to the prevailing culture within medicine. They use particular hierarchies because they correspond to the value commitments that they hold.

Several hierarchies have been presented as heuristics to facilitate rapid decision making (Greer et al 2000, Bandelow et al 2002, Howick et al 2011). Heuristics are designed to produce conclusions that are usually true when information and time are limited (Blunt 2015). These hierarchies are used by clinicians to guide urgent decision making. It is accepted that these hierarchies may sometimes produce false conclusions but this is outweighed by the need to provide urgent treatment when patients are acutely unwell. A decision making process does not have to produce conclusions that are necessarily true in order to be substantively rational. Bounded rationality acknowledges that information is often incomplete and that the conclusion of a rational reasoning process may be false. The use of hierarchies of evidence as heuristics is an explicit acknowledgement that bounded rationality is used.

So far in this section we have examined the decision making method used by evidence based medicine but we have not considered the goal of this decision making in any great depth. A decision making process can only be

considered rational if the goal is consistently achieved. It is therefore important to consider the goal of evidence based medicine in greater detail.

Within the literature a number of different goals have been ascribed to evidence based medicine. These include superior patient care, prevention of harm and control of financial costs (Cochrane 1972, Spitzer et al 1979, Evidence Based Medicine Working Group 1992, Waugh 1999, Guyatt et al 2008 page 924, Howick 2011). The Evidence Based Medicine Working Group (1992) claimed that the goal of evidence based medicine was superior patient care. However, the term 'superior patient care' was not defined. This term can be interpreted in a number of different ways. It could refer to maximisation of survival rates, minimisation of morbidity rates, prevention of harm, improved quality of life, more precise estimates of treatment effect, clinically significant differences in treatment effects, cost effective care or control of financial costs. None of these are unreasonable goals and the list is certainly not exhaustive. It is also easy to envisage scenarios where different potential goals conflict with one another. For example, a particular treatment intervention could maximise survival rates but be expensive, cause significant morbidity and have an imprecise estimate of treatment effect. This is precisely why decision making in medicine is so complex.

A lot has been written about evidence based medicine but a single goal has not been clearly defined. This is problematic when the rationality of decision making is considered from the perspective of practical rationality. This is

because, without a clearly defined goal, it is difficult to determine whether that goal has been consistently achieved. This does not mean that decision making within evidence based medicine cannot be rational. We simply have insufficient information to make a judgement unless a clear goal is defined. In this chapter we have resolved the problem by ascribing a generic goal to evidence based medicine. If the reader disagrees with this they may not be persuaded by the arguments that have been presented, and we would be back to square one. Indeed this might be why the hierarchies have proliferated: different groups disagree on the goals. More significantly, they might disagree on different goals because they are all seeking to achieve different things within medicine.

The absence of a clearly defined goal is less problematic when the rationality of decision making is considered using substantive rationality. This is because value commitments can influence both the reasoning process and the goal (Etzioni 1988). From the perspective of substantive rationality, decision making can be guided by a number of different goals. Each decision does need to be guided by a goal but this goal does not need to be immutable. Each of the possible definitions for 'superior patient care' that were previously listed could provide the goal for a particular decision. As long as the hierarchy that is used and the goal are consistent with value commitments, decision making can be substantively rational. This means that the fifth premise of the irrationality argument is also false because the goal of evidence based medicine does not have to be improved patient care.

In this section I have analysed the claim that the decision making process used within evidence based medicine is not rational. I have argued that the decision making process used is not rational from the perspective of theoretical or practical rationality but can be rational from the perspective of substantive rationality. This is because hierarchies may be consistently used when they reflect the value commitments of users. This does not mean that every decision made using the framework imposed by evidence based medicine is substantively rational but every decision has the potential to be rational. The substantive rationality of any decision can only be determined if the value commitments that influence selection of hierarchy of evidence and desired goal are clearly stated.

7.5. Conclusion:

In this chapter we have considered the rationality of the decision making process that is used within evidence based medicine. The concept of rationality has been explored and the importance of different types of rationality has been highlighted. Decision making can only be rational if conclusions are derived in a consistent way following consideration of the same information. However, the existence of different hierarchies may allow different conclusions to be reached following consideration of the same information. This suggests that decision making within evidence based

medicine may not be rational and has important implications for patient care, healthcare funding, medicolegal standards and medical research.

It has been argued that decision making cannot be rational from the perspective of theoretical rationality because the hierarchies have limited theoretical support. Decision making may be rational, from the perspective of practical rationality, when conclusions derived from different hierarchies are modified by the grading process. However, this would mean that decision making was only rational when conclusions derived using the characteristic feature of evidence based medicine, hierarchies of evidence, were overridden by factors that do not characterise the concept. This interpretation is problematic for evidence based medicine. Decision making can be rational from the perspective of substantive rationality but this allows a relativistic interpretation where treatment recommendations, medicolegal standards and decisions about healthcare funding and research can be both rational and inconsistent, depending upon the value commitments that are held. This has important implications for evidence based medicine and highlights the importance of clearly identifying the value commitments that underpin any decision.

Evidence based medicine may derive conclusions using a rational process when value commitments dictate the hierarchy of evidence that is used. Science also derives conclusions using a rational process (Newton-Smith 1981). Evidence based medicine has been claimed to be science on multiple

occasions. However, the limited empirical and theoretical support for hierarchies of evidence suggests that this claim may be false. The claim that evidence based medicine is science is therefore analysed in the next two chapters.

Chapter 8: Evidence Based Medicine as Science

8.1. Introduction:

Evidence based medicine has been claimed to be science on a number of different occasions (Cochrane 1972, Spitzer 1979, Evidence Based Medicine Working Group 1992, Guyatt et al 2000, Djulbegovic et al 2009). 'Scientific medicine' was originally proposed as an alternative name for the concept but this name was abandoned because of the implication that medicine had previously been unscientific (Howick 2011). Science is often considered epistemically superior to other forms of knowledge because it uses scientific method to provide justification for knowledge claims that are made (Ladyman 2002)[75]. This use of scientific method allows science to attain a privileged position in society (Sorell 1991). It is therefore important to analyse the claim that evidence based medicine is science because if it is not science it does not deserve the privileged status attributed to science.

Within philosophy of science four main theories about the nature of science are recognised: inductivism (Ayer 1946), falsificationism (Popper 1963), Kuhnian paradigms (Kuhn 1996) and research programmes (Lakatos 1970). If evidence based medicine is science the knowledge claims that inform decision making should be derived using a process that corresponds to one

---

[75] Knowledge can be defined as justified true belief. For a belief to constitute knowledge the belief must be true and that true belief must be justified (Ladyman 2002).

of these theories. However, it is notable that none of these theories propose hierarchies of evidence. Pseudoscience can be defined as beliefs or practices mistakenly regarded as being based on scientific method (Oxford English Dictionary 2006, Pigliucci 2015). If evidence based medicine claims to be science, but the knowledge claims that inform decision making are not derived using a process that corresponds to an accepted theory about the nature of science, evidence based medicine may be pseudoscience.

Categorisation of a discipline as pseudoscience is usually considered derogatory. Any claim that evidence based medicine is pseudoscience is therefore likely to be contested because it will have negative implications for knowledge claims that are made. Within philosophy of science a range of different criteria have been developed that can be used to demarcate science from pseudoscience (Ayer 1946, Popper 1963, Lakatos 1970, Thagard 1978, Ruse 1982, Kuhn 1996). These criteria do not provide necessary and sufficient conditions for this demarcation but they are considered useful (Curd, Cover and Pincock 2013). These demarcation criteria may be used to investigate any claim that evidence based medicine is pseudoscience.

In this chapter we will consider whether evidence based medicine is science or pseudoscience. In the first section the four main theories about the nature of science: inductivism, falsificationism, Kuhnian paradigms and research programmes will be introduced. In the second section the claim that evidence based medicine is science will be reinterpreted as the claim that the

knowledge claims that inform decision making within evidence based medicine are science. This claim has previously been considered and the existing literature will be reviewed. In the third section the justification for the knowledge claims that inform decision making within evidence based medicine will be considered from the perspective of different theories about the nature of science. It will be argued that these knowledge claims are not justified using scientific method. This suggests that evidence based medicine may be pseudoscience. In the final section, the claim that evidence based medicine is pseudoscience is considered using criteria that have been developed to demarcate science from pseudoscience.

8.2. Theories about the Nature of Science:

There is a common misconception that there is a single scientific method. However, within philosophy of science, four main theories about the nature of science are recognised: inductivism (Ayer 1946), falsificationism (Popper 1963), Kuhnian paradigms (Kuhn 1996) and research programmes (Lakatos 1970)[76]. It is important to appreciate that none of these theories are universally accepted and they are all considered problematic from an epistemological perspective. It is not my intention to discuss these theories in any great detail or argue for the superiority of any one theory. This would be

---

[76] Feyerabend (1975) has argued that there is no such thing as scientific method. On this interpretation it is meaningless to consider whether evidence based medicine corresponds to a theory about the nature of science. Newton-Smith (1981) has counter-argued that there must be scientific method because established sciences like chemistry and physics are so successful in making predictions and manipulating the physical world.

beyond the scope of this thesis and has been comprehensively undertaken elsewhere (Newton-Smith 1981, Ladyman 2002, Chalmers 2010, Curd, Cover and Pincock 2013). The different theories are simply presented in sufficient detail to provide a foundation for the analysis that is presented in the rest of the chapter.

The first theory about the nature of science that will be considered is inductivism. Inductivism places a high value on explanatory theory and proposes that scientific method involves the confirmation of theory by observation. In inductivism, theories are used to generate predictions and these are tested in experiments. When a predicted outcome is observed the theory is confirmed and the observation is considered to provide justification for the knowledge claim made by the theory (Ayer 1946). Theories may also be used to explain established phenomenon. Inductivism is considered a rational account of science because theories that are confirmed by observation are consistently accepted by scientists (Ladyman 2002). This theory about the nature of science is considered problematic because inductive argumentation is fallible.

The second theory about the nature of science is falsificationism. Falsificationism also places a high value on explanatory theory and proposes that scientific method involves the falsification of theory by observation. Falsificationism employs deductive reasoning therefore any conclusions are necessarily true if the premises are true (Popper 1963). Falsificationism is

considered problematic because theories are always tested in conjunction with auxiliary assumptions and background knowledge. As a result falsification does not prove that the theory being tested is false because it is also possible that an auxiliary assumption or background knowledge is false[77]. Falsificationism has also been criticised because it only leads to negative knowledge as hypotheses are refuted but never proven. Falsification is considered a rational theory of science because theories that are falsified are consistently rejected by scientists (Lakatos 1970, Ladyman 2002, Chalmers 2010).

The third theory about the nature of science is Kuhnian paradigms. The concept of a paradigm was introduced by Thomas Kuhn in *'The Structure of Scientific Revolutions'*. Kuhn (1996) argued that normal science occurred with an established paradigm and was characterised by puzzle solving. Puzzle solving is guided by the disciplinary matrix and exemplars. The disciplinary matrix includes the general theoretical laws, instruments and assumptions that scientists use and governs permissible concepts, problems and explanations. Exemplars demonstrate the problem solving techniques that can be used to extend and elaborate the scope of a paradigm. Normal science is considered a highly determined activity with conceptual, theoretical, instrumental and methodological commitments. This ensures that normal science is a rational activity.

---

[77] This is the Duhem-Quine hypothesis (Ladyman 2002).

Puzzle solving by scientists within an established paradigm had three important elements: the matching of facts with theory, the articulation of the consequences of theory and determination of significant facts (Kuhn 1996). The paradigm is assumed to guarantee the existence of a solution to every puzzle and failures within normal science are blamed upon individual scientists not the paradigm itself. However, as normal science progresses, experimental and theoretical anomalies accumulate and a crisis emerges. This crisis is resolved by the emergence of a new paradigm and a paradigm shift occurs. Paradigm shift is discussed in Chapter 9 when we consider the claim that evidence based medicine is a new paradigm.

The final theory that will be considered is research programmes. This theory of scientific method was developed by Imre Lakatos and also emphasises the importance of the framework within which science is undertaken. A scientific research programme consists of a hard core of fundamental principles surrounded by a peripheral protective belt of auxiliary assumptions. Work within any research programme is guided by the heuristic and may involve falsification and/or confirmation of theory through observation. Work undertaken within the confines of a research programme is therefore rational. The heuristic drives the research programme and allows the development and possible refutations of a research programme to be anticipated in advance. This theory solves some of the problems posed by the Duhem-Quine hypothesis because a methodological decision has been taken to unquestioningly accept the hard core (Lakatos 1970).

The theory of research programmes differentiates between progressive and degenerating research programmes[78]. A progressive research programme continues to produce novel facts. A degenerating research programme only accommodates known facts and does not produce novel facts. A new research programme is considered to supersede an old research programme when it has greater explanatory power than its rival and/or predicts novel phenomenon. All research programs have unsolved problems and it is often not possible to determine whether a research programme is progressing or degenerating contemporaneously (Lakatos 1970). As a consequence, extra-empirical factors may influence the decision to retain or reject a research programme. The theory of scientific research programmes is therefore considered an irrational account of science (Newton-Smith 1981, Ladyman 2002, Curd, Cover and Pincock 2013).

The four theories introduced in this section are all theories about the nature of science. However, it is important to make a distinction between inductivism and falsificationism on one hand and Kuhnian paradigms and research programmes on the other hand. Inductivism and falsificationism are theories of scientific method whereas Kuhnian paradigms and research programmes are broader theories that emphasise the importance of the framework within

---

[78] It is acknowledged that this is a simplistic explanation as proto-science and old science have also been recognised (Curd, Cover and Pincock 2013). However, this is unimportant to the analysis presented in this chapter.

which science is undertaken[79]. Falsificationism or inductivism can actually be used within a Kuhnian paradigm to solve puzzles. This distinction between the different theories about the nature of science is important when we reinterpret the claim that evidence based medicine is science.

Within philosophy of science four main theories about the nature of science are recognised: inductivism, falsificationism, Kuhnian paradigms and research programmes. These theories have been introduced and described in sufficient detail to provide a foundation for the analysis that is presented in later sections. If evidence based medicine is science knowledge claims should be derived using a process that corresponds to one of these theories. It is therefore interesting that none of these theories postulate hierarchies or appear to value randomised controlled trials and meta-analyses. This suggests that evidence based medicine may not be science. However, before we analyse the claim that evidence based medicine is science we need to establish that proponents of evidence based medicine have indeed made this claim.

---

[79] This also explains why Section 8.2 is entitled 'theories about the nature of science' not 'theories of scientific method'.

<u>8.3. Evidence Based Medicine as Science:</u>

The claim that evidence based medicine is a new paradigm provides the strongest evidence that proponents believe that it is science. Evidence based medicine was first claimed to be a new paradigm by the Evidence Based Medicine Working Group (1992) and this claim has been restated by Guyatt et al (2000) and Djulbegovic et al (2009). Each of these groups uses the term paradigm in a Kuhnian sense and the Working Group (1992) actually reference *'The Structure of Scientific Revolutions'* in their original paper. These groups all clearly identify evidence based medicine with the theory of Kuhnian paradigms.

The early hierarchies provide further evidence that proponents of evidence based medicine believed that the concept was science. Cochrane (1972) claimed that randomised controlled trials could be used to make scientific measurements and the Canadian Task Force claimed that their approach was based on the best available scientific evidence (Spitzer et al 1979). The *'How to Read Clinical Journals'* series presented rules of scientific evidence for the study of treatment (Sackett 1981) and the American College of Chest Physicians proposed a new hierarchy (Cook et al 1992) because the previous hierarchy did not include 'scientific' overviews (systematic reviews):

> 'Moreover, the Levels of Evidence were developed prior to the recognition of the power of rigorous scientific overview.' (Cook et al 1992 page 305S-306S)

Although many of the early hierarchies claimed that their approach was science these claims are rarely repeated by later hierarchies. This may reflect increasing scepticism, or tacit uncertainty, regarding the claim that evidence based medicine is science. Alternatively, the claim that evidence based medicine is science may be accepted unquestioned.

Superficial consideration of the claim that evidence based medicine is science suggests that the claim may be meaningless. Evidence based medicine is used to determine the most appropriate medical care whereas the aim of science is to provide explanation and understanding. However, the claim can be reformulated so that is meaningful. When proponents claim that evidence based medicine is science what they really mean is that the knowledge claims that inform decision making within evidence based medicine are science. These knowledge claims are derived from the study designs that are prioritised by the hierarchies of evidence: randomised controlled trials and meta-analyses.

If randomised controlled trials and meta-analyses are science the method that these study designs use to produce knowledge claims should correspond to an established theory about the nature of science. The hierarchies should not themselves be confused with science because they do not directly produce explanation and understanding. The hierarchies simply rank the knowledge claims produced by different study designs. However,

this does not mean that the hierarchies are irrelevant when we assess whether knowledge claims derived from randomised controlled trials and meta-analyses are science. The hierarchies are actually fundamental to this claim because they dictate the study designs that are preferred within evidence based medicine. The hierarchies are particularly important when the knowledge claims derived from randomised controlled trials and meta-analyses are considered within the context of theories that emphasise the importance of the framework within which science is conducted i.e. Kuhnian paradigms and research programmes.

In Chapter 4 it was argued that the hierarchies of evidence lack empirical and theoretical support and that randomised controlled trials and meta-analyses are not epistemically superior to expert opinion or other study designs. These arguments are not incompatible with the claim that randomised controlled trials and meta-analyses are science. Randomised controlled trials and meta-analyses can still be science if the method that they use to produce knowledge claims corresponds to an established theory about the nature of science.

There has been a limited discussion in the literature about whether randomised controlled trials use the scientific method of inductivism or falsificationism (Senn 1991, Shahar 1997, Djulbegovic et al 2009, Thompson 2010, Kerry et al 2012). However, this discussion is generally predicated on the assumption that randomised controlled trials are science. Senn (1991)

has argued that the method used by clinical trials corresponds to falsificationism. Shahar (1997) has counter-argued that randomised controlled trials cannot use falsificationism because the null hypothesis is not falsified. Shahar (1997) claimed that the method used by randomised controlled trials instead corresponded to inductivism although this claim was disputed by Kerry et al (2012) because the results of randomised controlled trials were presented in terms of probabilities. Thompson (2010) has claimed that randomised controlled trials cannot be science because they do not provide explanation. None of these authors have explicitly suggested that randomised controlled trials may be pseudoscience[80].

Several authors have considered whether evidence based medicine can be related to inductivism or falsificationism (Silva and Wyer 2009, Sestini 2010). Silva and Wyer (2009) argued that evidence based medicine could not be explained using inductivism because the hierarchies contained pathophysiological rationale[81]. Sestini (2010) claimed that evidence based medicine corresponded to falsificationism because a correspondence theory of truth was used and research evidence could be rejected. I would argue that evidence based medicine itself cannot be related to either inductivism or falsificationism because it is not scientific method. The claim made by Sestini (2010) is also problematic because use of the correspondence theory of truth

---

[80] Charlton (2009) and Schafranski (2012) suggested that evidence based medicine may be pseudoscience but no arguments were presented to support these claim. This claim has not been advanced by anybody else.

[81] This appears to relate to a particular interpretation of inductivism termed logical positivism. In logical positivism only statements that are empirically verifiable and deductive inferences are meaningful (Ayer 1946).

and rejection of research evidence do not differentiate falsificationism from other theories about the nature of science.

The theories of Kuhnian paradigms and research programme are broader theories than falsificationism and inductivism because they consider the framework within which science is undertaken. It is therefore meaningful to consider whether evidence based medicine itself is science in relation to these theories. The claim that evidence based medicine is a new Kuhnian paradigm has been discussed in detail in the literature (Couto 1998, Shahar 1998, Tonelli 1998, Greaves 2002, Sehon and Stanley 2003, Daly 2005, Lambert 2006, Djulbegovic et al 2009, Goldenberg 2010). This discussion assumes that evidence based medicine is science but this assumption is not explicitly considered. There has been no discussion in the literature about whether evidence based medicine corresponds to the theory of research programmes[82].

Proponents of evidence based medicine have claimed that evidence based medicine is science. However, this claim is problematic because evidence based medicine is used to determine the most appropriate medical care whereas the aim of science is to provide explanation and understanding. The claim has therefore been reinterpreted as the claim that knowledge claims derived from randomised controlled trials and meta-analyses are science.

---

[82] Kulkarni (2005) does acknowledge the theory of research programmes but does not relate this theory to evidence based medicine.

The hierarchies of evidence are fundamental to this claim because they prioritise randomised controlled trials and meta-analyses over other study designs. There has been a limited discussion in the literature about whether randomised controlled trials use inductivism or falsificationism but this is generally predicated on the assumption that they are science. In the next section we will therefore consider whether randomised controlled trials and meta-analyses are science from the perspective of the four main theories about the nature of science.

8.4. Randomised Controlled Trials and Meta-analyses as Science:

In this section we will consider the claim that randomised controlled trials and meta-analyses are science. Science is characterised by its use of scientific method to provide justification for knowledge claims that are made (Newton-Smith 1981, Ladyman 2002, Chalmers 2010, Curd, Cover and Pincock 2013). Therefore if randomised controlled trials and meta-analyses are science the method that they use to justify knowledge claims should correspond to one of the four main theories about the nature of science: inductivism, falsificationism, Kuhnian paradigms or research programmes. If the method that they use does not correspond to one of these theories they are not science. This argument is summarised below:

Knowledge claims can only be science if they are justified using a method that corresponds to a theory about the nature of science

The method used by randomised controlled trials and meta-analyses to justify knowledge claims does not correspond to a theory about the nature of science.

Knowledge claims justified by randomised controlled trials and meta-analyses are not science.

This is a valid argument form so we need to consider the truth of the premises to determine whether the argument is sound. The first premise claims that knowledge claims can only be science if they are justified by a method that corresponds to a theory about the nature of science. This premise will be considered uncontroversial and is accepted. We therefore need to consider whether the method used by randomised controlled trials and meta-analyses, within the framework imposed by hierarchies of evidence, corresponds to inductivism, falsificationism, Kuhnian paradigms or research programmes.

Inductivism places a high value on explanatory theory. In inductivism, theories are used to generate predictions and these are tested in experiments. When a predicted outcome is observed the theory is confirmed and the observation is considered to provide justification for the knowledge claim made by the theory (Ayer 1946). If the method use by randomised controlled trials and meta-analyses corresponds to inductivism these study designs should test underlying theory.

Randomised controlled trials and meta-analyses are set up to test the null hypothesis that there is no difference in outcome between two treatment interventions. If a difference is found, the probability of the observed difference occurring by chance, if there really is no difference between the two treatment interventions, is calculated. The null hypothesis is then accepted or rejected depending upon a preassigned probability value. Randomised controlled trials and meta-analyses do not use inductivism because they are not designed to confirm the hypothesis that a particular treatment intervention is effective. Randomised controlled trials and meta-analyses do not confirm theory by observation. They test the hypothesis that there is no difference between two treatment interventions.

It is important to appreciate that the treatment interventions that are tested in randomised controlled trials, and later collated in meta-analyses, are not selected at random. They are selected because they are supported by underlying theory and these treatment interventions must undergo rigorous laboratory and animal testing before clinical trials on human subjects. In this sense the interventions that are tested are supported by underlying theory and any claim that basic medical science did not use inductivism would be false. However, this underlying theory is not relevant when the null hypotheses is accepted or rejected. The null hypothesis is accepted or rejected based only on the preassigned probability value[83].

Although underlying theory is not relevant when the null hypothesis is accepted or rejected this information is considered when the conclusions of the study are interpreted. This was illustrated in Chapter 6 when different factors that have influenced the hierarchies were considered. Biological plausibility is informed by underlying theory and this is important when evidence is interpreted (Blunt 2015). However, this is not relevant to the present discussion as we are considering whether knowledge claims are derived from randomised controlled trials and meta-analyses using inductivism not whether underlying theory influences the interpretation of any conclusions. Interestingly no hierarchy or grading process differentiates between treatment interventions that are biologically plausible and

---

[83] This criticism is equally applicable to other study designs that compare outcomes in different groups such as controlled clinical trials and observational studies.

implausible. This may be because it is assumed that all treatment interventions that are tested are biologically plausible.

The hierarchies prioritise knowledge claims derived from randomised controlled trials and meta-analyses but they do contain a number of other study designs. One of these study designs, the prospective case series, does employ inductivism because it uses observations to confirm underlying theory. In a prospective case series all patients are given a treatment intervention and outcomes are recorded. The theory supporting the treatment intervention is then confirmed to a greater or lesser extent depending upon the outcomes that are observed. Inductivism can provide compelling evidence for the effectiveness of a treatment intervention when all subjects benefit from treatment. This is the 'all-or-none' study design. However, prospective case series are rarely valued by hierarchies. This study design is usually excluded or ranked amongst the lowest levels. Only the hierarchy developed by the Oxford Centre for Evidence Based Medicine (Ball and Phillips 2001) ranks the 'all-or-none' study design amongst the highest levels. Study designs that do employ inductivism are therefore not generally prioritised by the hierarchies of evidence.

Inductivism places a high value on explanatory theory. This is in direct contrast to the hierarchies that value evidence from empirical study designs. Few hierarchies include explanatory theory and those that do tend to devalue this form of evidence. For example, the American Heart Association (2000)

ranks rational conjecture as the lowest level of evidence. Most hierarchies would subsume explanatory theory within expert opinion ranking it as the lowest level of evidence. Only the hierarchy developed by Mitchell and Friese (2007) would allow explanatory theory to attain the highest level of evidence. The fact that expert opinion and prospective case series are not generally prioritised by hierarchies does not mean that randomised controlled trials and meta-analyses do not employ inductivism. However, if inductivism was important within evidence based medicine we would also expect explanatory theory and prospective case series to be prioritised. This provides indirect evidence to support the claim that randomised controlled trials and meta-analyses do not use inductivism.

Randomised controlled trials and meta-analyses do not use inductivism because they do not seek to confirm underlying theory through observation. They test the hypothesis that there is no difference between different treatment interventions. This is fundamentally different to inductivism and may be considered a negative characterisation of knowledge. The treatment interventions that are tested in randomised controlled trials and meta-analyses are supported by theory but this theory is not directly tested. Explanatory theory is fundamental to inductivism and prospective case series clearly employ inductivism. However, explanatory theory and prospective case series are rarely valued by hierarchies. This suggests that the hierarchies do not value the confirmation of theory by observation and provides further indirect support for the claim that randomised controlled trials and meta-analyses do not use inductivism.

8.4.2. Falsificationism:

Falsificationism also places a high value on explanatory theory. Falsificationism proposes that science involves the falsification of theory by observation. However, although the interventions that are tested in randomised controlled trials and meta-analyses are underlain by theory, this theory is not directly tested. Randomised controlled trails and meta-analyses are set up to test the null hypothesis that there is no difference in outcome between two treatment interventions. If the underlying theory is not directly tested it cannot be falsified.

Randomised controlled trials and meta-analyses do not even falsify null hypotheses. When the results of randomised controlled trials and meta-analyses are interpreted the null hypothesis is accepted or rejected depending on a preassigned probability value. If the null hypothesis is rejected the alternative hypothesis is accepted using inference to best explanation. Falsificationism is characterised by its rejection of theory using deductive inferences and inference to best explanation is not a deductive inference. Randomised controlled trials and systematic reviews cannot use falsification because the results of these studies are interpreted probabilistically. Even a conclusion that is highly likely to be true is not necessarily true.

The hierarchies prioritise knowledge claims derived from randomised controlled trials and meta-analyses but they do contain a number of other different study designs. The prospective case series study design can be used to falsify theory in the same way that it can be used to confirm theory. This can be illustrated if we consider infective condition X. Theory proposes that treatment A is an effective treatment for this condition in all patients and a prospective case series is undertaken. If some patients with infective condition X fail to recover after treatment with treatment A, the theory that this is an effective treatment for infective condition X in all patients is falsified, within the limits imposed by the Duhem-Quine hypothesis. However, the hierarchies do not usually value prospective case series. This study design is generally ignored or ranked as the lowest level of evidence. In short the hierarchies do not value the one study design that can be used to falsify theory.

Testing of theory is fundamental to the scientific method of falsificationism. However, as demonstrated above, when inductivism was considered, explanatory theory is not valued by hierarchies of evidence. Few hierarchies include explanatory theory and this form of evidence would normally be subsumed within expert opinion, as the lowest level of evidence. The fact that expert opinion and prospective case series are not prioritised by hierarchies provides indirect evidence that randomised controlled trials and meta-analyses do not employ falsificationism. This is because if falsificationism

was valued we would also expect explanatory theory and prospective case series to be prioritised by the hierarchies.

Randomised controlled trials and systematic reviews do not use falsificationism because the underlying theory is not directly tested and null hypotheses are rejected using inference to best explanation not deduction. Explanatory theory is fundamental to falsification and the prospective case series study can potentially be used to falsify underlying theory but explanatory theory and prospective case series are not valued by the hierarchies. Randomised controlled trials and systematic reviews do not use falsificationism and this scientific method is not valued by hierarchies.

8.4.3. Kuhnian Paradigms:

The theory of paradigms differs from inductivism and falsificationism because it emphasises the importance of the framework within which science is undertaken. The knowledge claims derived from randomised controlled trials and meta-analyses must therefore be considered within the framework imposed by evidence based medicine. Evidence based medicine has been claimed to be a new paradigm on a number of occasions (Evidence Based Medicine Working Group 1992, Guyatt et al 2000, Djulbegovic et al 2009) and it is important to note that it is evidence based medicine itself, not

randomised controlled trials and meta-analyses, that is claimed to be the new paradigm.

Superficial consideration of evidence based medicine suggests that it may conform to the definition of paradigm proposed by Kuhn (1996). Hierarchies of evidence are fundamental to decision making within evidence based medicine and these could be considered part of the disciplinary matrix. Evidence based medicine textbooks are replete with examples that show how the hierarchies should be used and these could be exemplars. However, we should be careful not to conflate tools that have been developed to guide decision making in medicine with the conduct of science. The hierarchies, and the examples provided in evidence based medicine textbooks, have been developed to guide medical decision making not the conduct of science.

When evidence based medicine is conceived solely as a tool to guide medical decision making it cannot be a paradigm but evidence based medicine has also driven the medical research agenda. Randomised controlled trials and meta-analyses are more likely to be undertaken, funded and published because they are valued by hierarchies (Hammersley 2013, Mebius 2014). The idea that certain study designs provide the 'best evidence' to support medical decision making and the idea that certain study designs simply provide the 'best evidence' have become conflated. When evidence

based medicine is conceived in this way there may still be potential for evidence based medicine to be a paradigm.

Evidence based medicine can only be considered a paradigm if it engages in puzzle solving. Puzzle solving within an established paradigm has three important elements: the matching of facts with theory, the articulation of the consequences of theory and determination of significant facts (Kuhn 1996). We therefore need to consider whether the study designs prioritised by hierarchies of evidence, randomised controlled trials and meta-analyses, can be used to match facts with theory, articulate the consequences of theory and determine significant facts.

Randomised controlled trials and meta-analyses are designed to test null hypotheses. The interventions that are tested are underlain by theory but this theory is not directly tested and these study designs do not use inductivism or falsificationism. If randomised controlled trials and meta-analyses do not directly test underlying theory it is difficult to see how these study designs can match facts with theory or articulate the consequences of underlying theory.

Evidence based medicine may still be considered normal science if the conclusions of randomised controlled trials and meta-analyses are significant facts. However, knowledge claims derived from these study designs are

based upon a probabilistic interpretation of the results. Null hypotheses are rejected when it is unlikely that the treatment intervention and control intervention are equally effective. I would suggest that knowledge claims derived in this way are not significant facts in the sense that Kuhn intended. This does not mean that knowledge claims derived from randomised controlled trials and meta-analyses cannot inform clinical decision making but this conflates evidence based medicine as a decision making tool with evidence based medicine as science.

Some hierarchies do include the prospective case series and this study design can be used to confirm or falsify underlying theory. Theory is rarely explicitly included within hierarchies but can be included within expert opinion. Evidence based medicine therefore has the potential to match facts with theory, articulate the consequences of theory and determine significant facts. However, prospective case series and expert opinion are not generally valued by hierarchies and these forms or evidence are clearly viewed as inferior to randomised controlled trials and meta-analyses.

The argument that evidence based medicine does not engage in puzzle solving can be developed further. Kuhn (1996) was clear that scientists engaged in normal science assume that the paradigm guarantees a solution to every puzzle. Failures in puzzle solving are blamed upon individual scientists not the paradigm itself. However, this is not how evidence based medicine is presented (Guyatt et al 2008, Straus et al 2011). Randomised

controlled trials and meta-analyses are not expected to produce significant facts, even when they are undertaken to a high standard, because of bias inherent within these study designs[84]. The hierarchies that purport to rank different study designs using bias rank randomised controlled trials and meta-analyses as the highest levels of evidence because they minimise not eliminate bias[85]. If bias is only ever minimised significant facts can surely not be established.

It is important to reiterate that I am not claiming that the treatment interventions that are tested in randomised controlled trials and meta-analyses are not developed using underlying theory. In fact, the testing of underlying theory is of fundamental importance in the development of these interventions. New pharmacological agents are often painstakingly developed over many years in laboratory and animal studies (Mebius 2014). During this development, experiments are repeated, underlying theory may be modified and new instruments may be developed. This is puzzle solving activity. However, laboratory and animal studies are not valued by the hierarchies and the hierarchies dictate the study designs that should be used to justify knowledge claims within evidence based medicine.

Evidence based medicine has been claimed to be normal science but the study designs prioritised by the hierarchies are not designed to match facts

---

[84] This does not mean that medical researchers are not blamed when randomised controlled trials are considered to be poorly conducted.
[85] Bias was considered in detail in Chapter 5

with underlying theory or articulate the consequences of theory. Moreover, knowledge claims derived from these study designs are not significant facts, because they are based on a probabilistic interpretation of the null hypothesis. Other study designs within the hierarchies can be used to solve puzzles but these study designs are rarely valued. I would therefore argue that evidence based medicine cannot use the theory of Kuhnian paradigms to justify knowledge claims that are made.

8.4.4. Research Programmes:

Similar to the theory of Kuhnian paradigms, this theory also emphasises the importance of the framework within which science is conducted. This means that randomised controlled trials and meta-analyses cannot be considered independently of evidence based medicine. All progressive research programmes have a hard core of underlying theory and a heuristic that guides research. Progressive research programmes also continue to produce novel facts. Therefore, if evidence based medicine is a research programme it should have a hard core of underlying theory, a guiding heuristic and continue to produce novel facts.

It could be argued that the hierarchies represent the hard core of an evidence based medicine research programme. This argument can only be made if the hierarchies are conceived as a tool for conducting science not a tool to

support medical decision making. The hierarchies do possess some of the features of a hard core. They are generally unquestioned and have developed with time. However, there are a number of problems with this interpretation. Firstly, it is not clear that the hierarchies should be conceived as a tool for conducting science. Secondly, although hierarchies have proliferated with time it is difficult to claim that they have evolved with time because the same study designs consistently provide the highest and lowest levels of evidence. Finally, the hard core of a research programme is a hard core of underlying theory but, as argued in Chapter 4, the hierarchies have limited theoretical support. If the hierarchies have limited theoretical support it is difficult to see how they can provide a hard core of underlying theory. I would therefore argue that the hierarchies cannot represent the hard core of an evidence based medicine research programme.

Hierarchies may not be able to provide the hard core of a research programme but they could conceivably provide the heuristic. Several hierarchies have been claimed to be heuristics to facilitate rapid decision making (Greer et al 2000, Bandelow et al 2002, Howick et al 2011) although this claim has not been extended to heuristics within a research programme. Heuristics are designed to produce conclusions that are usually true (Blunt 2015) therefore hierarchies could potentially act as a heuristic with limited theoretical support. Randomised controlled trials and meta-analyses do not provide superior evidence to other study designs or expert opinion but this does not necessarily mean that they provide inferior evidence. There may therefore be scope for hierarchies to act as a heuristic within an evidence

based medicine research programme although it is unclear how the development of such a research programme could be anticipated in advance.

If evidence based medicine is a research programme and the hierarchies provides the heuristic there would have to be a separate hard core of underlying theory. This hard core has not previously been articulated but this does not mean that it could not be articulated. The hard core could include axioms such as treatment interventions must be biologically plausible and treatment interventions that are tested in humans should be supported by laboratory and animal studies. However, even if a hard core could be developed it is not clear that it would characterise an evidence based medicine research programme. This is because evidence based medicine is characterised by the use of hierarchies. If the hierarchies do not form part of the hard core, evidence based medicine would surely lose its essence. I would therefore suggest that it would be challenging to develop a hard core, without the hierarchies, which was limited to evidence based medicine and not the broader concept of medical science.

Any research programme must have an underlying hard core, heuristic and produce novel facts. Even if a hard core of theory underlying evidence based medicine could be articulated, independent of the hierarchies of evidence which would provide the heuristic, novel facts would still need to be produced for there to be a research programme. This research programme would preferentially value evidence claims derived from randomised controlled trials

and meta-analyses. However, these study designs reject null hypotheses based upon probability and do not produce significant facts. This being the case it is difficult to claim that they can produce novel facts within any evidence based medicine research programme. If randomised controlled trials and meta-analyses do not produce novel facts evidence based medicine cannot be a progressive research programme.

In this section we have considered whether evidence based medicine may be a research programme. The hierarchies of evidence cannot provide the hard core because they have limited theoretical support although they could potentially provide a heuristic. This would require articulation of a separate hard core of underlying theory that was distinct from medical science. A characteristic feature of any research programme is that it should produce novel facts yet randomised controlled trials and meta-analyses do not produce novel facts. If the study designs that are prioritised by the hierarchies do not produce novel facts it is difficult to envisage how evidence based medicine can be a research programme.

8.4.5. Summary:

Randomised controlled trials and meta-analyses can only be considered science if knowledge claims are derived using a method that corresponds to a theory about the nature of science. We have therefore considered whether

the method used by randomised controlled trials and meta-analyses, within the framework imposed by hierarchies of evidence, corresponds to inductivism, falsificationism, Kuhnian paradigms or research programmes. Inductivism and falsificationism are theories of scientific method whereas Kuhnian paradigms and research programmes are broader theories about the nature of science.

The interventions that are tested in randomised controlled trials and meta-analyses may be supported by underlying theory but this theory is not confirmed or falsified when the null hypothesis is accepted or rejected. This means that randomised controlled trials and meta-analyses do not use inductivism or falsificationism. Treatment interventions that are tested in prospective case series can be confirmed or falsified but these study designs are not valued by hierarchies. This shows that the hierarchies themselves do not value inductivism or falsificationism.

Kuhnian paradigms and research programmes are broader theories about the nature of science. Knowledge claims derived from randomised controlled trials and meta-analyses have therefore been considered within the framework imposed by evidence based medicine when these theories have been considered. Evidence based medicine cannot be normal science or a research programme because the study designs prioritised by hierarchies, randomised controlled trials and meta-analyses, test null hypotheses not underlying theory. Therefore they do not match facts with theory, articulate

the consequences of theory, determine significant facts or produce novel facts.

The analysis presented in this section has revealed that knowledge claims derived from randomised controlled trials and meta-analyses, within the framework imposed by hierarchies of evidence, are not derived using a process that corresponds to an established theory about the nature of science. I would therefore argue that the second premise of the science argument is true and the argument is sound. This means that knowledge claims derived from randomised controlled trials and meta-analyses are not science. If the study designs prioritised by hierarchies of evidence, are not science, it is difficult to see how evidence based medicine can be science. If accepted by the reader, this conclusion has significant implications for evidence based medicine because it means that knowledge claims do not deserve the status of science. If evidence based medicine is not science it may be more appropriately categorised as pseudoscience. This will be considered in the final section of this chapter.

8.5. Evidence based medicine as pseudoscience:

If evidence based medicine is not science it may be more appropriately categorised as pseudoscience. Pseudoscience can be defined as a collection of beliefs or practices mistakenly regarded as being based on the scientific

method (Oxford English Dictionary 2006). Pseudoscience is essentially non-science masquerading as science. Pseudoscience can produce knowledge but these knowledge claims are not justified using scientific method. Categorisation of a discipline as pseudoscience is usually considered derogatory and has negative implications for the status of the discipline in society.

Evidence based medicine has been claimed to be science on a number of occasions. This claim is implicit in the claim that evidence based medicine is a new paradigm (Evidence Based Medicine Working Group 1992, Guyatt et al 2000, Djulbegovic et al 2009) and has been made by a number of the earlier hierarchies of evidence (Cochrane 1972, Spitzer et al 1979, Sackett 1981, Cook et al 1992). However, knowledge claims do not appear to be derived using a process that corresponds to an established theory about the nature of science. If evidence based medicine claims to be science, but does not use scientific method to justify knowledge claims, it may be pseudoscience. This would not mean that evidence based medicine could not produce knowledge but it would mean that patient care, healthcare funding, medicolegal standards and the medical research agenda were determined by pseudoscience. Any claim that evidence based medicine is pseudoscience is therefore likely to be resisted by proponents of evidence based medicine.

Within philosophy of science a number of different criteria have been developed to demarcate science from pseudoscience (Ayer 1946, Popper 1963, Lakatos 1970, Thagard 1978, Ruse 1982, Kuhn 1996, Resnik 2000)[86]. These demarcation criteria do not provide necessary and sufficient conditions for the demarcation of science from pseudoscience but they are still considered to have utility (Curd, Cover and Pincock 2013). It is generally agreed that some disciplines are science e.g. physics and chemistry, whereas other disciplines are pseudoscience e.g. homeopathy and astrology. The concepts of science and pseudoscience are therefore useful although we may need to accept that there is blurred demarcation area between the two concepts and that some disciplines may be difficult to categorise. It is not unusual to have a blurred demarcation area between concepts in analytical philosophy (Belohlavek et al 2009, Pigliucci 2015). These demarcation criteria can be used to support the claim that evidence based medicine is pseudoscience.

If evidence based medicine is pseudoscience it should not satisfy the criteria that have been used to demarcate science from pseudoscience. As the demarcation area between science and pseudoscience is blurred, evidence based medicine could still be pseudoscience even if it does satisfy some of the demarcation criteria. Failure to satisfy any of the demarcation criteria would obviously strengthen the claim that evidence based medicine was pseudoscience. Many of the demarcation criteria are related to the theories

---

[86] Resnik (2000) argued that science is simply what scientists do. However, this demarcation criterion is not considered helpful.

about the nature of science that have already been described and they will not be considered in great detail.

Ayer (1946) claimed that science produced meaningful statements whereas pseudoscience did not produce meaningful statements. A statement is considered meaningful if it is analytic or empirically verifiable. This relates to a particular interpretation of inductivism termed logical positivism. Evidence based medicine values knowledge claims derived from randomised controlled trials and meta-analyses. These study designs do not test underlying theory and the null hypotheses are accepted or rejected on the basis of probability. Prospective case series can be used to empirically verify knowledge claims but this study design is not prioritised by hierarchies. Analytic statements are also not prioritised by hierarchies of evidence as they would be subsumed within expert opinion. I would therefore argue that evidence based medicine does not value the production of meaningful statements and should be considered pseudoscience using this criterion.

Popper (1963) claimed that science made testable predictions and was prepared to abandon beliefs when these predictions were falsified. This was in direct contrast to pseudoscience that made vague predictions that could not be falsified. Kuhn (1996) claimed that science was characterised by puzzle solving whereas Lakatos (1970) claimed that science was characterised by the continued production of novel facts. These demarcation criteria relate to the theories of falsificationism, Kuhnian paradigms and

research programmes. It was argued in the previous section that randomised controlled trials and meta-analyses do not justify knowledge claims using falsificationism and that evidence based medicine does not value puzzle solving or the production of novel facts. I would therefore argue that evidence based medicine should be considered pseudoscience using the demarcation criteria proposed by Popper (1963), Kuhn (1996) and Lakatos (1970).

Thagard (1978) has claimed that science is simple whereas pseudoscience is complex. Pseudoscience effectively disguises the fact that it is not science by wrapping itself in multiple layers of complexity to hide its true nature. This demarcation criterion has been criticised because complexity is a relative term, a preference for simplicity often reflects practical concerns or ontological beliefs and some science, such a quantum mechanics, is complex (Newton-Smith 1981). This criterion is difficult to apply to evidence based medicine because it uses hierarchies of varying complexity. Some of the earlier hierarchies are simple whereas many later interpretations are complex. If evidence based medicine is interpreted charitably this demarcation criterion could potentially be used to provide weak support for the claim that evidence based medicine is not pseudoscience.

Ruse (1982) claimed that science was prepared to abandon long standing beliefs, whereas pseudoscience retained long standing beliefs, when confronted with contradictory evidence. This demarcation criterion has been criticised because it does not consider the justification for abandonment of

belief and the history of science reveals that scientists have often retained beliefs in the face of contradictory evidence (Ladyman 2002, Curd, Cover and Pincock 2013). Evidence based medicine will abandon long standing beliefs when this decision is supported by the results of randomised controlled trials and systematic reviews. This is exemplified by the logo of the Cochrane Collaboration. In some circumstances evidence based medicine will abandon long standing beliefs, based upon other evidence, when that evidence is considered particularly convincing. This possibility was first recognised by Cochrane (1972) and is implicit in the hierarchies that allow other study designs to provide the highest level of evidence.

Although evidence based medicine may abandon long standing beliefs based on evidence derived from expert opinion and other study designs, collective consideration of the hierarchies reveals a preference for evidence derived from randomised controlled trials and meta-analyses. This demarcation criterion can be used to support the claim that evidence based medicine is not pseudoscience. However, the reader may be troubled by the fact that study designs that do not justify knowledge claims using scientific method are preferentially used to justify abandonment of long standing belief. I would therefore suggest that this demarcation criterion only provides weak support for the claim that evidence based medicine is not pseudoscience.

In this section the claim that evidence based medicine is pseudoscience has been considered using criteria developed with philosophy of science to

demarcate science from pseudoscience. Evidence based medicine is categorised as pseudoscience using the demarcation criteria developed by Ayer (1946), Popper (1963), Lakatos (1970) and Kuhn (1996) because it does not make meaningful statements, value falsificationism, engage in puzzle solving or produce novel facts. This is not surprising as these demarcation criteria relate to the four established theories about the nature of science that were considered in the previous section. We therefore need to be careful about the weight that we attribute to these criteria to avoid 'double-counting' as they have already been used to argue that evidence based medicine is not science. The demarcation criteria proposed by Thagard (1978) and Ruse (1982) claim that science is simple and prepared to abandon long standing beliefs when faced with contradictory evidence. Evidence based medicine will abandon long standing beliefs and, when interpreted charitably, may be considered simple. However, these demarcation criteria are considered problematic from a philosophical perspective. These criteria only provide weak support for the claim that evidence based medicine is science.

The demarcation criteria outlined in this section do not provide necessary and sufficient condition for the demarcation of science from pseudoscience and it is accepted that the demarcation area between these concepts is blurred. Having considered the different criteria we need to make a decision about where evidence based medicine lies. Evidence based medicine does not justify knowledge claims using a method that corresponds to an established theory of science and it fails to satisfy most of the demarcation

criteria. The two criteria that do suggest that evidence based medicine may be science, are philosophically problematic, require a charitable interpretation of hierarchies and only offer weak support for the claim. I would therefore argue that evidence based medicine is not science and is more appropriately categorised as pseudoscience.

The claim that evidence based medicine is pseudoscience has negative implications for knowledge claims that are made using this concept. These knowledge claims determine patient care, healthcare funding, medicolegal standards and the medical research agenda. If evidence based medicine is pseudoscience it could be argued that knowledge claims derived using evidence based medicine should not have such an important role within medicine. The claim that evidence based medicine is pseudoscience is therefore likely to be resisted by proponents of the concept.

Proponents of evidence based medicine may respond to this claim in one of four ways. Firstly, the claim that evidence based medicine is not science may be challenged although it is my contention that the analysis presented in this chapter is robust. Secondly, the concept of evidence based medicine may be rearticulated so that knowledge claims are justified using a method that corresponds to an established theory about the nature of science. This would require a significant reworking of the concept and we would likely be left with something that was so far removed from the current concept of evidence based medicine that it was no longer 'evidence based medicine'. Thirdly, it

may be accepted that evidence based medicine is pseudoscience. Pseudoscience can provide justification for knowledge claims it just does not use scientific method to provide that justification. This response is unlikely as the term pseudoscience is generally considered derogatory. Finally, evidence based medicine could abandon claims to be science. It is notable that later hierarchies do not make this claim. If evidence based medicine is conceived solely as a tool to guide medical decision making it is neither science nor pseudoscience. I would suggest that this final response is the most likely response to the claim that evidence based medicine is pseudoscience.

The claim that evidence based medicine is not science should not have negative repercussions for the wider field of medicine. Medicine has been defined as the science or practice of the diagnosis, treatment and prevention of disease. However, although medicine may be an applied science, it cannot be reduced to science because it also requires empathy, ethics, compassion, tacit knowledge and clinical judgement (Saunders 2000, Panda 2006). The suggestion that medicine may still be an applied science may surprise the reader but we should remember that the interventions tested in randomised controlled trials are supported by basic medical science. Medical practice that is informed by evidence based medicine can therefore still be applied science in an indirect sense.

8.6. Conclusion:

In this chapter we have considered whether evidence based medicine is science or pseudoscience. This is an important consideration because science and pseudoscience hold very different positions in society. It has been argued that evidence based medicine is not science. This is because knowledge claims are preferentially derived from randomised controlled trials and meta-analyses and these study designs do not derive knowledge claims using a process that corresponds to an established theory about the nature of science. The claim that evidence based medicine is pseudoscience has been analysed using criteria that have been used to demarcate science from pseudoscience. These criteria provide weak support for evidence based medicine as science and suggest that evidence based medicine may be pseudoscience. Any claim that evidence based medicine is pseudoscience can be avoided if evidence based medicine abandons any claim to be science.

In the next chapter we will consider the claim that evidence based medicine is a new paradigm in greater detail. This claim is important because it gives evidence based medicine powerful perlocutionary force. We should immediately be sceptical about the claim that evidence based medicine is a new paradigm because if evidence based medicine is not science it cannot be a new paradigm. The theory of Kuhnian paradigms is also interesting because, although normal science is a rational activity, paradigm choice is

irrational. This suggests that the decision to adopt evidence based medicine

by the medical profession may have been irrational.

Chapter 9: Evidence Based Medicine as a new Kuhnian Paradigm

9.1. Introduction:

Evidence based medicine has been claimed to be a new Kuhnian paradigm on several occasions (Evidence Based Medicine Working Group 1992, Guyatt et al 2000, Djulbegovic et al 2009). This claim was initially made by the Evidence Based Medicine Working Group in 1992 when the concept of evidence based medicine was first fully articulated. Evidence based medicine was claimed to be a new Kuhnian paradigm because it recognised the limitations of certain types of evidence, required clinicians to search the literature efficiently and used formal rules of evidence:

> 'A new paradigm for medical practice is emerging. Evidence based medicine de-emphasises intuition, unsystematic clinical experience and pathophysiological rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research. Evidence based medicine requires new skills of the physician including efficient literature searching and the application of formal rules of evidence evaluating the clinical literature' (Evidence Based Medicine Working Group, 1992, page 2420).

The concept of paradigm was introduced by Thomas Kuhn in *'The Structure of Scientific Revolutions'*. Kuhn argued that science was characterised by

long periods of normal science[87] interspersed with revolutionary transitions. Normal science occurs within an established paradigm and is characterised by puzzle solving. The paradigm is assumed to guarantee a solution to every puzzle. However, as normal science progresses, experimental and theoretical anomalies accumulate and a crisis emerges. This crisis is resolved by the emergence of a new paradigm and a revolutionary transition occurs. Revolutionary transitions are called paradigm shifts and they are usually rare events (Kuhn 1996).

Kuhn (1996) argued that different paradigms were incommensurable because they defined terms and concepts in different ways, used different reference frameworks and disagreed about the legitimacy of problems and proposed solutions. As different paradigms are incommensurable a new paradigm cannot be shown to be superior to a previous paradigm in an objective sense (Ladyman 2002). Scientists must therefore be persuaded to adopt a new paradigm. The theory of paradigms is considered an irrational theory about the nature of science because the decision to adopt a new paradigm may be influenced by the ideology, cultural norms and values of scientists (Kuhn 1996).

The claim that evidence based medicine is a new Kuhnian paradigm is important for several reasons. Firstly, if evidence based medicine is a new Kuhnian paradigm, the adoption of evidence based medicine by the medical

---

[87] Normal science within an established paradigm was considered in greater detail in Chapter 8 when the claim that evidence based medicine is science was evaluated.

profession may not have been rational. If the adoption of evidence based medicine by the medical profession was not rational we need to consider which factors persuaded the medical profession to adopt evidence based medicine. Secondly, the claim that evidence based medicine is a new paradigm has powerful perlocutionary force. The claim enhances the status of evidence based medicine as it suggest that evidence based medicine is somehow superior to the way that medicine was previously practised. If evidence based medicine is not a new paradigm it does not deserve this status.

The claim that evidence based medicine is a new paradigm has previously been discussed in the literature (Couto 1998, Shahar 1998, Tonelli 1998, Greaves 2002, Sehon and Stanley 2003, Daly 2005, Lambert 2006, Djulbegovic et al 2009, Goldenberg 2010, Gaeta and Gentile 2016)[88]. However, this claim has proved contentious and considerable disagreement exists with a number of different arguments being used to both support and refute the claim. Some of these arguments relate to evidence based medicine itself whereas others relate to the theory proposed by Kuhn. It has been argued that evidence based medicine is not a new paradigm because it is not significantly different from the way that medicine was previously practised (Couto 1998, Sehon and Stanley 2003, Daly 2005, Goldenberg 2010), randomised controlled trials are poorly understood by the medical profession (Shahar 1998), the concept of evidence based medicine is poorly

---

[88] A systematic review was undertaken in April 2013 to identify papers considering the claim that evidence based medicine is a new paradigm. The search strategies that were used can be found in Appendix 1.

defined (Shahar 1998, Tonelli 1998, Sehon and Stanley 2003) and there are problems with the theory of Kuhnian paradigms (Greaves 2002, Sehon and Stanley 2003, Gaeta and Gentile 2016). Nobody has previously argued that evidence based medicine cannot be a new Kuhnian paradigm because evidence based medicine is not science[89]. Conversely, it has been claimed that evidence based medicine is a new paradigm because it is significantly different from the way that medicine was previously practised (Evidence Based Medicine Working Group 1992, Lambert 2006, Djulbegovic et al 2009).

There appear to be three main sources of confusion when the claim that evidence based medicine is a new paradigm has previously been considered in the literature. Firstly, there appear to be a poor understanding of the theory of paradigms. For example, it has been falsely claimed that evidence based medicine is not a new paradigm because the notion of incommensurability is incoherent (Sehon and Stanley 2003) and paradigms are not influenced by moral and cultural factors (Greaves 2002). Secondly, some authors have struggled to define the concept of evidence based medicine. This is important because if evidence based medicine cannot be clearly defined it may be difficult to determine whether it is a new paradigm. Finally, there is disagreement as to whether evidence based medicine is incommensurable with a previous paradigm of medicine. This third area of confusion is central

---

[89] Gaeta and Gentile (2016) argued that evidence based medicine could not be a Kuhnian paradigm because it described a methodology not an underlying theory. However, they did not explicitly claim that evidence based medicine was not science.

to consideration of the claim that evidence based medicine is a new paradigm.

In this chapter I will consider the claim that evidence based medicine is a new Kuhnian paradigm in detail. In the first section I will explain what it means for different paradigms to be incommensurable and clarify some important elements of the theory of Kuhnian paradigms. This will resolve some of the confusion that currently exists in the medical literature and provide a foundation for later analysis. In the second section, several problems with the original claim that evidence based medicine is a new paradigm will be highlighted. These include the concurrent claim that evidence based medicine provides superior patient care and a failure to identify the relevant scientific community affected by paradigm shift. In the final section, I will argue that evidence based medicine cannot be a new paradigm because it is not science, it was not preceded by a revolutionary crisis and it is not incommensurable with the way that medicine was practised prior to its introduction.

9.2. Kuhnian Paradigms and Incommensurability:

The theory of Kuhnian paradigms is considered complicated and the concepts of paradigm and incommensurability can be difficult to understand. Furthermore, this theory of scientific method has been interpreted in a

number of different ways (Masterman 1970, Ladyman 2002, Chalmers 2010). This may explain why there appears to be poor understanding of the theory of Kuhnian paradigms within the evidence based medicine literature (Greaves 2002, Sehon and Stanley 2003). Kuhn (1996) was aware that his theory had been interpreted in a number of different ways by philosophers of science and he wrote the *'Japanese postscript'* to address these problems. In this chapter the definitions and theory about the nature of science presented in the *'Japanese postscript'*, will be used to minimise confusion.

The original version of *'The Structure of Scientific Revolutions'* was criticised because the term paradigm was used in a number of different senses (Masterman 1970). Kuhn (1996) later clarified, in the *'Japanese postscript'*, that the term paradigm was used in only two different senses. Firstly, the term paradigm was used to refer to exemplars. An exemplar is an example showing how puzzles can be solved within normal science. Exemplars are typically found in science textbooks. Second, the term paradigm was used to refer to the entire constellation of beliefs, commitments, values and techniques shared by a scientific community, including exemplars. In order to avoid further confusion, unless otherwise explicitly stated, the term paradigm will only be used in the second sense in this chapter.

Kuhn (1996) acknowledged that distinguishing paradigms was difficult and required close historical study. He was very clear that revolutions could involve both large and small communities of scientists. A revolution can

actually affect a community with fewer than 25 members. Kuhn was also clear that revolutions did not happen immediately and a community could take a generation to convert to a new paradigm. This is because new paradigms are usually taken up by the younger members of a scientific community who are more amenable to change. In order to study paradigms it is therefore necessary to isolate the relevant scientific community and reconstruct their beliefs, commitments, values and techniques over time.

Communities are fundamental to the study of Kuhnian paradigms. A community can be understood as a group of members that interact with each other and share common interests and goals (Fulcher and Scott 2011). Communities can be difficult to identify and this may complicate the study of paradigms in some sciences. This is a recognised problem with the theory of Kuhnian paradigms but may not be particularly important when the theory is applied to medicine. This is because medicine is a profession (MacDonald 1995). Different scientific communities in medicine can therefore be more readily identified because they are regulated and accredited by different professional bodies.

Different paradigms are considered incommensurable because they define terms and concepts in different ways, use different reference frameworks and disagree about the legitimacy of problems and proposed solutions. The theory of Kuhnian paradigms has been misinterpreted as claiming that reality changes with paradigm change. However, this is not the case as Kuhn

(1996) clearly states in the *'Japanese postscript'* that different paradigms are simply different perspectives or views of the same world.

Sehon and Stanley (2003) argued that evidence based medicine could not be a new paradigm because the notion of incommensurability was incoherent. They appeared to equate incommensurability with incomprehensibility but this demonstrates a misunderstanding of the theory of paradigms. Different paradigms can be comprehended by scientists and translation between paradigms is possible with care. Translation between paradigms is not the same as conversion to the new paradigm as conversion involves adopting a particular world view and thinking and working within the new paradigm. The difference between incommensurability and incomprehensibility can be illustrated with an example. The adoption of the heliocentric model of the solar system, during the Scientific Revolution, is commonly used to illustrate paradigm shift (Ladyman 2002). Geocentric and heliocentric models of the solar system are incommensurable because they represent different world views. However, they are not incomprehensible as we are clearly able to understand both models. What we cannot do is simultaneously interpret the solar system from both perspectives.

The theory of scientific method expounded by Kuhn (1996) in *'The Structure of Scientific Revolutions'* was mainly developed using historical examples from physics and it has been suggested that this theory may have no application outside of physics (Chalmers 2010). This potential problem is

acknowledged but should not prevent us analysing the claim that evidence based medicine is a new paradigm. This is because the theory of Kuhnian paradigms is not restricted to physics within contemporary philosophy of science textbooks (Ladyman 2002, Curd, Cover and Pincock 2013) and Kuhn (1996) himself considered Darwinian evolutionary theory. It is also important to remember that the claim that evidence based medicine is a new Kuhnian paradigm was originally advanced by proponents of evidence based medicine. If the theory of paradigms is only applicable to physics evidence based medicine cannot be a new paradigm because it is not a branch of physics.

In this section the theory of Kuhnian paradigms has been described in greater detail. The meanings of the terms paradigm and incommensurability have been clarified and elements of the theory that are commonly misunderstood have been explained. I have also acknowledged some potential limitations of the theory of paradigms. This section is important because it provides a framework for the analysis that takes place in the rest of this chapter. In the next section we will explore problems associated with the original claim that evidence based medicine is a new paradigm.

9.3. Evidence Based Medicine as a new Paradigm:

The Evidence Based Medicine Working Group have claimed on several occasions that evidence based medicine is a new paradigm (Evidence Based Medicine Working Group 1992, Guyatt et al 2000). The term 'paradigm' can be defined in a number of different ways (Oxford English Dictionary 2006) therefore it is important to clarify that the term is used in a Kuhnian sense. If the term 'paradigm' is not used in a Kuhnian sense it would be inappropriate to criticise the Working Group for claiming that evidence based medicine was a new Kuhnian paradigm. However, when we appraise the original paper it is very clear that the Working Group use the term 'paradigm' in a Kuhnian sense as they reference *'The Structure of Scientific Revolutions'* and state that:

> 'Thomas Kuhn has described scientific paradigms as ways of looking at the world'. (Evidence Based Medicine Working Group, 1992, page 2420)

It is therefore appropriate to consider whether evidence based medicine is a new paradigm within the theory of Kuhnian paradigms.

There are a number of problems with the original claim that evidence based medicine is a new paradigm. Firstly, the Working Group simultaneously claimed that evidence based medicine was both a new paradigm and resulted in superior patient care. This appears to demonstrate a

300

misunderstanding of the theory of Kuhnian paradigms because different paradigms are incommensurable. Incommensurability means that a new paradigm cannot be shown to be superior in any objective sense when compared with a preceding paradigm. If evidence based medicine is a new paradigm it cannot also be shown to result in superior patient care.

There may be an alternative explanation for the concurrent claim that evidence based medicine provides superior patient care. This claim may be interpreted as an attempt to persuade medical scientists to convert to evidence based medicine. The theory of paradigms is an irrational theory of scientific method and individuals must be persuaded to convert to the new paradigm. If medical professionals believe that evidence based medicine results in superior patient care they may be more likely to convert. This may also explain why the new paradigm is described as exciting and fun in the original paper.

The second problem with the claim that evidence based medicine is a new paradigm is that the relevant scientific community is not clearly identified. In order to establish whether evidence based medicine is a new paradigm we need to isolate the relevant scientific community and reconstruct their beliefs, commitments, values and techniques. If the relevant scientific community cannot be isolated we cannot determine if evidence based medicine is a new paradigm. The opening sentence of the original paper claims that a new paradigm for medical practice is emerging (Evidence Based Medicine

Working Group 1992). This suggests that evidence based medicine is a new paradigm for medical practice. Later in the same paper, the principles of evidence based medicine are restricted to internal medicine and other (un-named) clinical departments. This suggests that evidence based medicine is a new paradigm for communities within medicine rather than medicine as a whole. The original paper is actually entitled: *'Evidence Based Medicine a New Approach to Teaching the Practice of Medicine'*. This suggests that evidence based medicine is only a new paradigm for the teaching of medicine.

The failure of the Evidence Based Medicine Working Group (1992) to identify a relevant scientific community, makes the claim that evidence based medicine is a new paradigm, difficult to analyse. When we contemporaneously consider the claim that evidence based medicine is a new paradigm we tend to assume that the relevant scientific community is the medical profession. This is because, evidence based medicine has permeated all areas of modern medicine. However, this may not be the relevant scientific community that is referred to in the original paper. A lack of clarity over the relevant scientific community may partially explain the lack of agreement that exists in the literature when the claim that evidence based medicine is a new paradigm is considered (Couto 1998, Daly 2005).

The final problem with the claim that evidence based medicine is a new Kuhnian paradigm is that the concept of evidence based medicine is

simultaneously articulated, and claimed to be a new paradigm, at the same time. Paradigm choice is a consensus decision made by a scientific community and revolutions take many years. Revolutions do not occur instantaneously. It is therefore unclear how the Evidence Based Medicine Working Group (1992) can claim that evidence based medicine is a new paradigm at the same time as the concept is first fully articulated. This does not mean that the relevant scientific community will not convert to evidence based medicine in the future but this conversion will not occur instantaneously.

Evidence based medicine did not mysteriously appear in 1992 and the concepts underpinning evidence based medicine actually developed over many years within clinical epidemiology (Daly 2005). It is therefore conceivable that evidence based medicine was a new paradigm for the clinical epidemiology community in 1992. The relevant scientific community affected by paradigm change is not clear in the original paper so it could refer to the clinical epidemiology community. It is important to appreciate that this is a charitable interpretation of the claim because, although a number of different possible scientific communities are suggested, clinical epidemiology is not mentioned.

In this section we have considered three problems with the original claim that evidence based medicine is a new paradigm: evidence based medicine cannot be a new paradigm and provide superior patient care, the relevant

scientific community is not specified and evidence based medicine cannot be a new paradigm as soon as it is articulated. These problems suggest that the Evidence Based Medicine Working Group (1992) did not fully understand the theory of Kuhnian paradigms when they first advanced the claim that evidence based medicine was a new paradigm. However, the problems do not necessarily prevent evidence based medicine being a new paradigm. The first problem can be resolved if the claim that evidence based medicine provides superior patient care is withdrawn or acknowledged as an attempt to persuade individuals to convert to evidence based medicine. The second and third problems can potentially be resolved if a relevant scientific community is identified.

In the next section, the claim that evidence based medicine is a new paradigm is analysed in greater detail. I will argue that evidence based medicine cannot be a new paradigm because it is not science, it was not preceded by a revolutionary crisis and it is not incommensurable with the way that medicine was practised prior to its introduction. The relevant scientific community will be assumed to be the medical profession although other scientific communities will be considered where they are important to the arguments that are analysed.

9.4. Evidence based medicine is not a new Paradigm:

In order to analyse the claim that evidence based medicine is not a new paradigm we need to identify the relevant scientific community. The relevant scientific community is identified as the medical profession. This is because the medical profession is one of the communities mentioned in the original paper and evidence based medicine has influenced the whole medical profession with time. Different scientific communities, both inside and outside the medical profession, are considered when these influence the arguments that are analysed. It is acknowledged that the relevant scientific community may not have been correctly identified by applying the arguments to the entire medical profession. However, the arguments that are presented remain sound when they are applied to different scientific communities within medicine.

Three arguments are presented to support my claim that evidence based medicine is not a new paradigm: evidence based medicine is not science, evidence based medicine was not preceded by a revolutionary crisis and evidence based medicine is not incommensurable with the way that medicine was practised prior to its introduction. These arguments are summarised below:

Paradigm Argument 1:

If evidence based medicine is not science it is not a new Kuhnian paradigm

Evidence based medicine is not science

Evidence based medicine is not a new Kuhnian paradigm

Paradigm Argument 2:

If evidence based medicine was not preceded by a revolutionary crisis it is not a new Kuhnian paradigm

Evidence based medicine was not preceded by a revolutionary crisis

---

Evidence based medicine is not a new Kuhnian paradigm

Paradigm Argument 3:

If evidence based medicine is not incommensurable with the way that medicine was practised prior to its introduction it is not a new Kuhnian paradigm

Evidence based medicine is not incommensurable with the way that medicine was practised prior to its introduction

---

Evidence based medicine is not a new Kuhnian paradigm

These three arguments are all valid arguments of the form modus ponens. We therefore need to consider the truth of the premises to determine if the arguments are sound. The first paradigm argument claims that evidence based medicine cannot be a new Kuhnian paradigm because it is not science. The theory of Kuhnian paradigms describes how science evolves with time and cannot be applied to pseudoscience or knowledge claims justified in other ways. In Chapter 8 I argued that evidence based medicine was not science because it did not justify knowledge claims using a method that corresponded to an established theory about the nature of science. If the reader accepts that evidence based medicine is not science it cannot be a new paradigm. Importantly, this argument remains sound when it is applied to any 'scientific' community within medicine. If the method used to justify knowledge claims does not correspond to an established theory about the nature of science that 'scientific' community cannot claim to practice normal science. This argument provides strong support for my claim that evidence based medicine is not a new paradigm and we could stop at this point. However, the claim that evidence based medicine is not science may be disputed so it is still important to consider the other arguments.

The second paradigm argument claims that, if evidence based medicine is a new paradigm, there must have been a revolutionary crisis prior to its acceptance by the medical profession. Kuhn (1996) argued that all revolutionary crises had two universal features: blurring of the existing paradigm and loosening of the rules of normal research. Revolutionary crises can be identified because they are associated with the proliferation of

competing ideas, a willingness to try anything, debate over fundamentals and recourse to philosophy. However, there does not appear to have been a revolutionary crisis in the medical literature prior to the introduction of evidence based medicine (Gaeta and Gentile 2016). There was no period of time where competing ideas proliferated and different approaches were tried before evidence based medicine emerged. If evidence based medicine was not preceded by a revolutionary crisis it cannot be a new paradigm.

It would actually be very difficult to envisage a revolutionary crisis in any scientific community within medicine where the medical profession could try anything. This is because medicine involves the treatment of patients who may suffer morbidity and mortality if they receive inappropriate treatment. This highlights an important difference between medicine and established sciences such as astronomy, chemistry and physics. Wild theories can be developed and tested in science because falsification does not have important ramifications for the care of patients. However, it would be inappropriate and unethical to adopt this approach in medicine. This reinforces the argument that evidence based medicine cannot be a new paradigm because it is not science. The theory of paradigms is simply not applicable to evidence based medicine on a practical level because it is medicine and not science. The absence of a revolutionary crisis in medicine prior to the introduction of evidence based medicine provides further support for my claim that evidence based medicine is not a new paradigm.

The third paradigm argument claims that if evidence based medicine is a new paradigm then there must be an old paradigm and this must be incommensurable with evidence based medicine. In order to analyse this argument we need to identify and compare the relevant features of the different paradigms. This is complicated by the lack of clarity regarding the relevant scientific community affected by the paradigm shift. Several authors have argued that evidence based medicine cannot be a new paradigm because the concept of evidence based medicine is poorly defined (Shahar 1998, Tonelli 1998, Sehon and Stanley 2003). However, the Evidence Based Medicine Working Group (1992) clearly defined an old 'paradigm' and a new 'paradigm' in their original paper and they provided three arguments to support their claim that evidence based medicine was a new paradigm.

The Working Group first argued that evidence based medicine was a new paradigm because it recognised the limitations of certain types of evidence: common sense, pathophysiological rationale, expert opinion and clinical experience. It is important to appreciate that these forms of evidence are still used by evidence based medicine. They are simply attributed lesser importance than other forms of evidence, particularly randomised controlled trials and later systematic reviews. This is illustrated if we consider the large number of hierarchies that include expert opinion. Evidence based medicine will clearly use evidence derived from expert opinion to guide decision making when preferred forms of evidence are unavailable. Furthermore, hierarchies are available that attribute much greater importance to expert opinion. The claim that evidence based medicine always attributes lesser

importance to this type of evidence can therefore be disputed. For evidence based medicine to be a new paradigm it needs to be a change in world view for a community of scientists. I would argue that paradigm shift involves more than a reduction in the importance of evidence derived from expert opinion, pathophysiological rationale, common sense and clinical experience.

The Evidence Based Medicine Working Group also argued that evidence based medicine was a new paradigm because it required the skills of question formulation and literature searching. Medicine has surely always required healthcare professionals to formulate and answer questions so these skills alone cannot be used to argue that evidence based medicine is a new paradigm. However, what the Working Group actually mean is that evidence based medicine requires practitioners to search the literature in a specific way. This is illustrated in the *'How to Keep up with the Medical Literature'* series published in the Annals of Internal Medicine (Haynes et al 1986a, 1986b, 1986c, 1986d, 1986e, 1986f). This series expressed concern about the increasing volume of medical literature and used the presence or absence of randomisation in the study design to identify articles that should be appraised. This series also briefly discussed electronic searching of the literature although computers and appropriate software were not widely available at this time.

Within the context of this argument, evidence based medicine uses randomisation as a screening tool to identify its preferred study design, the

randomised controlled trial, amongst the ever increasing medical literature[90].

I would argue that screening the medical literature to identify certain study designs does not represent a change in world view and this does not make evidence based medicine a new paradigm.

The use of computers, both hardware and software, to show that evidence based medicine is a new Kuhnian paradigm is more interesting. Computers have allowed us to develop electronic databases and more efficient literature search strategies within evidence based medicine. However, computers have also had a profound effect on all science. Medicine would surely have started using computers with or without evidence based medicine. When we consider whether a paradigm shift has occurred we need to identify the relevant scientific community. It may be possible to argue that the use of computers has resulted in a paradigm shift but this shift has affected a much wider community than the medical profession. It has affected the whole science community. Therefore I would argue that when the relevant scientific community is identified as the medical profession, or any community within the medical profession, the use of computers cannot be used to demonstrate that evidence based medicine is a new paradigm.

The final argument claims that evidence based medicine is a new paradigm because it uses formal rules of evidence to determine the importance of

---

[90] The use of randomisation as a screening tool has recently been superseded and we now have the concept of a rapid review process which screens the literature to identify systematic reviews and meta-analyses (Manger et al 2017).

research. The Working Group did not define what they meant by the term 'formal rules of evidence' and this may explain why some authors have complained that the concept of evidence based medicine is poorly defined. However, although the term 'formal rules of evidence' is not clearly defined, if the original paper is placed in the correct historical context, we can see that the term is used to refer both to hierarchies of evidence that prioritise randomised controlled trials and methods to appraise different study designs. This is because by 1992 many of the pioneers of evidence based medicine had made strong claims about the merits of different study designs and study methodology (Sackett 1979) and already developed hierarchies of evidence (Cochrane 1972, Spitzer et al 1979, Trout 1981, Sackett 1989). Furthermore, the Evidence Based Medicine Working Group subsequently developed their own hierarchies (Guyatt et al1995, Guyatt et al 2000).

The Evidence Based Medicine Working Group (1992) argued that evidence based medicine was a new paradigm because it recognised the limitations of certain types of evidence, required clinicians to search the literature in an efficient manner and used formal rules of evidence. However, although the Working Group purport to present three separate arguments they actually make the same argument in three different ways. Each of these arguments effectively claims that evidence based medicine is a new paradigm because it values evidence from randomised controlled trials. The first argument devalues evidence that is not from randomised controlled trials, the second argument suggests that we should only search for randomised controlled trials and the third argument values the study design prioritised by early

hierarchies, the randomised controlled trial. The key question then is: Does a preference for evidence derived from randomised controlled trials make evidence based medicine incommensurable with the way that medicine was practised prior to its introduction?

Randomised controlled trials are not new. The first randomised controlled trials were conducted by Fisher in the field of agriculture in the 1920's (Fisher 1926). Prior to this randomisation had been used in the Nineteenth Century, albeit in a less sophisticated way, in empirical investigations into telepathy (Hacking 1988). The first medical randomised controlled trial, which showed that streptomycin was effective in the treatment of pulmonary tuberculosis, was published in 1948 (Marshall et al 1948) and the methodological advantages of this study design were well known to Bradford-Hill (1951), Feinstein (1964) and Cochrane (1972). The work of Bradford-Hill (1951) is particularly advanced in its understanding of randomised controlled trials and would not appear out of place in a contemporary evidence based medicine textbook.

Randomised controlled trials were clearly available to medicine for a number of years prior to 1992. If evidence based medicine originated in 1992 the continued use of randomised controlled trials would not make evidence based medicine incommensurable with the way that medicine was previously practised. However, although the concept of evidence based medicine was not articulated until 1992 this does not mean that the concept was not

developed before this time. Kuhn (1996) was clear that revolutions did not happen immediately and could take a generation to occur. It is therefore possible that Bradford-Hill (1951), Feinstein (1964) and Cochrane (1972) were early adopters of the unnamed concept that subsequently became evidence based medicine. It is also possible that, although randomised controlled trials were available prior to 1992, they were not the preferred form of evidence before this time. It is therefore still possible that evidence based medicine could be a new paradigm even though randomised controlled trials had been available within medicine since 1948.

Many of the later hierarchies rank systematic reviews above randomised controlled trials as the highest level of evidence. The first systematic review was published in 1974 (Bastian et al 2010) although evidence has been collated in medicine for several centuries (Evans et al 2010)[91]. As the first systematic review was published at around the same time that the concept of evidence based medicine was articulated evidence based medicine could argue that it was a new paradigm because it used evidence from systematic reviews. However, this argument would be difficult to justify because few hierarchies include systematic reviews or meta-analyses prior to 1995. Proponents of evidence based medicine would therefore need to explain this chronological discrepancy if they were going to use the use of systematic reviews as justification for the claim that evidence based medicine was a new paradigm.

---

[91] James Lind undertook a review of scurvy over 250 years ago (Evans et al 2010).

If evidence based medicine is a new paradigm it must have conceptual, theoretical, instrumental and methodological commitments that are incommensurable with the previous paradigm. It should also define terms and concepts in different ways; disagree about the legitimacy of problems and proposed solutions and use a different reference framework (Kuhn 1996). However, the only features that differentiate evidence based medicine from the way that medicine was previously practised are the hierarchy of evidence and the preferred use of evidence derived from randomised controlled trials. Even this is a charitable interpretation given the long history of use of randomised controlled trials in medicine. Evidence based medicine does not define terms and concepts in different ways and it does not disagree about the legitimacy of problems when it is compared with the way that medicine was practised prior to 1992.

If evidence based medicine only used evidence from randomised controlled trials and these had never been used in medicine prior to evidence based medicine, there may be a stronger case for arguing that evidence based medicine was a new paradigm. However, this is clearly not the case. Users of evidence based medicine are also happy to use evidence from other study designs when randomised controlled trials are unavailable and ignore the results of randomised controlled trials when they are considered implausible. This is illustrated by the response to the randomised controlled trial undertaken by Leibovici (2001). Evidence based medicine is not

315

incommensurable with the way that medicine was practised prior to its introduction and it is not a new Kuhnian paradigm.

The argument that evidence based medicine is not incommensurable with the way that medicine was practised prior to its introduction has been analysed with the medical profession identified as the relevant scientific community. However, the use of a hierarchy of evidence and a preference for evidence derived from randomised controlled trials do not represent a paradigm shift for any scientific community within medicine. I would argue that no scientific community within medicine has defined terms and concepts in different ways or used a different reference framework following the introduction of randomised controlled trials. If proponents of evidence based medicine would like to refute this claim they need to first specify a scientific community and then show how this scientific community has conceptual, theoretical, instrumental and methodological commitments that are incommensurable with a previous paradigm. I would suggest that this would be extremely challenging.

In this section, we have considered the claim that evidence based medicine is a new paradigm. I have argued that evidence based medicine cannot be a new paradigm because evidence based medicine is not science, evidence based medicine was not preceded by a revolutionary crisis and evidence based medicine is not incommensurable with the way that medicine was practised prior to its introduction. Difficulties associated with identifying a

relevant scientific community have been acknowledged. However, this is not detrimental to the arguments that are presented because they remain sound regardless of the scientific community that they are deemed to affect within medicine.

9.5. Conclusion:

In this chapter the claim that evidence based medicine is a new Kuhnian paradigm has been analysed. This claim is important because categorisation as a paradigm has given evidence based medicine powerful perlocutionary force. This claim that evidence based medicine is a new paradigm has previously been considered in the literature although considerable confusion exists with a number of different arguments used to both support and refute the claim. The analysis presented in this chapter has attempted to resolve this confusion.

The claim that evidence based medicine is a new paradigm is difficult to analyse because the relevant scientific community affected by paradigm shift is unclear. To facilitate analysis the relevant scientific community has been identified as the medical profession although it is acknowledged that this may not have been the intention of the original claim. Failure to clearly identify the relevant scientific community is not detrimental to the arguments that are presented because it has been argued that evidence based medicine is not a

new paradigm. However, any refutation of this claim would need to specify a relevant scientific community and demonstrate how this community has conceptual, theoretical, instrumental and methodological commitments that are incommensurable with a previous paradigm.

Three arguments have been presented to support my claim that evidence based medicine is not a new paradigm. Firstly, building on the analysis presented in Chapter 8, I have argued that evidence based medicine is not science. The theory of Kuhnian paradigms is a theory about the nature of science therefore if evidence based medicine is not science it cannot be a new paradigm. Secondly, I have argued that evidence based medicine cannot be a new paradigm because it was not preceded by a revolutionary crisis. Finally, I have argued that an increased reliance on evidence derived from randomised controlled trials and systematic reviews does not make evidence based medicine significantly different from the way that medicine was practised prior to the development of the concept. If evidence based medicine is not incommensurable with the way that medicine was previously practised it cannot be a new Kuhnian paradigm. These three arguments provide strong support for my claim that evidence based medicine is not a new Kuhnian paradigm and it should not be characterised as such.

Chapter 10: Implications for Evidence Based Medicine

10.1. Summary of Findings:

This thesis has presented a critical analysis of evidence based medicine using the method of analytical philosophy. The knowledge claims that are made by evidence based medicine determine which treatment interventions clinicians can prescribe for patients, healthcare funding, medicolegal standards and the medical research agenda. These knowledge claims therefore have significant implications for patients, clinicians, regulators and purchasers of healthcare. This thesis has analysed the justification for these knowledge claims with the aim of improving understanding of evidence based medicine.

Hierarchies of evidence are fundamental to the justification of knowledge claims that are made by evidence based medicine. These hierarchies are not the sole determinant of knowledge claims but, it has been argued that, as they are the characteristic feature of evidence based medicine they are fundamental to any knowledge claims that are made. It is therefore surprising that so many different hierarchies of evidence have been developed. This suggests a problem with the epistemology of evidence based medicine and provided the rationale for this thesis.

The hierarchies generally rank randomised controlled trials, systematic reviews and meta-analyses as the highest level of evidence and expert opinion as the lowest level of evidence. However, there is little theoretical or empirical support for the superiority of randomised controlled trials and the properties that have been used to determine hierarchical position are problematic. It has been argued that a lack of theoretical support has allowed factors that are independent of study design to influence the development of hierarchies. This may explain why there are so many different hierarchies of evidence.

The existence of multiple hierarchies led us to question the rationality of the decision making process used by evidence based medicine. Decision making is not rational from the perspective of theoretical or practical rationality but can be rational from the perspective of substantive rationality. This has important implications because knowledge claims are dependent upon underlying value commitments. This means that different conclusions can rationally be derived from the same evidence base. It is therefore importance to identify the value commitments that underpin any knowledge claim produced by evidence based medicine.

We have also considered the status of evidence based medicine as science. Evidence based medicine does not use scientific method to justify knowledge claims and it may be pseudoscience. If evidence based medicine is not science it cannot be a Kuhnian paradigm. These conclusions strip evidence

based medicine of some of its power and have important consequences for its status in society. This does not mean that evidence based medicine is not important or cannot produce knowledge. However, if evidence based medicine is not science knowledge claims are surely more likely to be robustly examined and challenged.

The analysis presented in this thesis has built on existing research, clarified areas of disagreement and improved understanding of evidence based medicine. Evidence based medicine has not been portrayed in a favourable light and some of the conclusions may be contested. In the final section my conclusions will be related to current ideas about evidence based medicine, possible responses will be anticipated and wider implications for evidence based medicine are considered.

10.2. Implications for Evidence Based Medicine:

Analytical philosophy progresses through argument and counterargument. This thesis should not therefore be considered a definitive analysis of the epistemology of evidence based medicine. The arguments and concepts that have been presented in this work should themselves be critically analysed and counterarguments are expected. Possible counterarguments have been anticipated as the thesis has progressed but it is not possible to anticipate every possible counterargument. The findings of this thesis will be published

to stimulate further debate in the literature about evidence based medicine. It is hoped that this will lead to a deeper understanding of the concept.

The method of analytical philosophy emphasises the importance of argument structure and requires clear definitions of any concepts that are analysed. It is acknowledged that the arguments can feel artificial and the style of writing is necessarily repetitive. Analytical philosophy may not be familiar to many medical professionals and it is uncertain how they will interpret the thesis. This thesis is envisaged as a contribution to medicine but it is accepted that the use of analytical philosophy may provide a barrier to the dissemination of the research findings to the medical profession. The medical profession may be more accepting of my research findings if they are supported by examples. The current controversy surrounding the necessity for antibiotic prophylaxis in cardiac patients prior to dental treatment (Thornhill et al 2016) may provide a useful case study.

This thesis represents a direct evaluation of evidence based medicine because it analyses the arguments that have been put forward by evidence based medicine to support the hierarchies. The arguments have been interpreted charitably and evidence based medicine has been given every opportunity to defend the hierarchies. This has allowed me to understand the hierarchies from their own perspective and is a major strength of the research method that has been used.

This thesis has not investigated how the hierarchies have been interpreted outside of evidence based medicine and this may be considered a weakness of my approach. We have not considered how the concept of evidence based medicine has evolved, how the hierarchies have been used to control power within medicine or how different hierarchies influence the construction of medical knowledge. Conceptual history, discourse analysis and qualitative methods were considered as possible research methods when my research question was developed and they remain important areas for further research. However, I would argue that I am now better placed to pursue these more indirect evaluations of evidence based medicine following my direct evaluation of the hierarchies themselves.

Some proponents of evidence based medicine hold strong views about the concept and may react negatively to the analysis that has been presented. It has never been my intention to denigrate evidence based medicine and this thesis has simply sought to improve understanding of the concept. Within philosophy of science the pessimistic induction thesis teaches us that all scientific theories, regardless of how strongly they are held by scientists at any particular time, are eventually replaced by new scientific theories. Evidence based medicine is not science but similar reasoning can be applied to this concept. It is therefore highly likely that, as time progress, evidence based medicine will be replaced by a different concept. Systems theory teaches us that we cannot anticipate how medicine will evolve but the concept of evidence based medicine will almost certainly be replaced. This

should be remembered when the arguments that have been presented in this thesis are interpreted.

Problems with the epistemology of evidence based medicine have previously been considered in detail by both Howick (2011) and Blunt (2015). It could also be argued that each new hierarchy reflects an implicit recognition of problems with the epistemology of evidence based medicine. Howick (2011) resolved these problems by favouring a particular hierarchical interpretation whereas Blunt (2015) advocated abolition of the hierarchies unless empirical support was provided. Neither Howick (2011) or Blunt (2015) sought to explain why there were so many different hierarchies or considered whether the decision making process used within evidence based medicine was rational or scientific. This thesis builds upon their previously published work but considers the implications of different hierarchies from an alternative perspective.

The hierarchies of evidence could be modified to include or exclude evidence derived from different study designs. Randomised controlled trials, systematic reviews and meta-analyses are not epistemically superior to other study designs and they could be excluded from hierarchies. However, if randomised controlled trials, systematic reviews and meta-analyses were excluded which study design should replace them as the highest level of evidence? Following on from the analysis that has already been presented I would be highly sceptical that any study design could be shown to be

epistemically superior to other study designs. This is because the most appropriate study design is dictated by a variety of different factors including the question of interest and any value commitments that are held. It is therefore unlikely that any modification that maintains a hierarchical structure will withstand epistemological analysis.

Abandonment of hierarchies of evidence by evidence based medicine has been suggested on several occasions (Glasziou et al 2004, Hofmeijer 2014 and Blunt 2015) and I would be broadly supportive of this idea. This does not mean that study design is unimportant but we should move away from the idea that some study designs are epistemically superior. Study design should still be considered, alongside other factors, when medical decisions are made. However, producers of evidence-based recommendations must be clear about all factors that are considered and any value commitments that underlie the decision making process. This will allow users of evidence based medicine to be better informed about any conclusions that are reached. If evidence based medicine did abandon the hierarchies it could be argued that it was no longer 'evidence based medicine' because it had lost its characteristic feature.

Hierarchies have clearly been influenced by a variety of factors that are independent of study design. This has been demonstrated by considering five factors in detail. It is also likely that other factors have influenced the development of hierarchies. These factors are important and should be

recognised when any conclusions that are derived using evidence based medicine are interpreted. It is not clear whether the hierarchies have naturally evolved to include these factors or they have been consciously manipulated to ensure the right sort of 'evidence based' conclusions. This may provide an interesting area for future research.

In this thesis it has been argued that evidence based medicine is not science and may be pseudoscience. It is important to appreciate that pseudoscience can provide justification for knowledge claims and pseudoscience is not necessarily epistemically inferior to science in this respect. Evidence based medicine could therefore accept that it was pseudoscience but argue that knowledge claims were still justified. I would suggest that this approach is unlikely because the term pseudoscience has such negative perlocutionary force. Proponents of evidence based medicine are more likely to abandon any claim to be science in which case it is neither science nor pseudoscience.

If evidence based medicine was accepted as pseudoscience there could be important implications for the control of professional jurisdiction and governance within medicine. There is an unresolved debate in the literature about whether evidence based medicine has strengthened or weakened the professional jurisdiction of the medical profession. Power relations within medicine are dynamic and complex and it is uncertain how the conclusions of this thesis, if accepted, will alter power relationships both within the medical

profession and between the medical profession, patients, regulators and purchasers of healthcare. Categorisation of evidence based medicine as pseudoscience would certainly be detrimental to the argument, presented by Mykhalovskiy and Weir (2004), that the concept buttressed the medical profession by reinforcing the scientific character of medical practice. The potential impact of categorisation of evidence based medicine as pseudoscience on power relationships within medicine would be another interesting area of future research.

This thesis has identified a number of problems with the epistemology of evidence based medicine. However, although it is easy to be critical of decision making within evidence based medicine, we do need to make decisions about medical care. It is also not clear that there are any alternatives to evidence based medicine at the present time. The analysis presented has drawn attention to the importance of value commitments and other factors that influence knowledge claims that are made using evidence based medicine. This improves our overall understanding of the concept and should make us more sceptical about any decisions that are justified using evidence based medicine.

The aim of this thesis has been to improve understanding of evidence based medicine. I have argued that evidence based medicine is not science but this does not mean that knowledge claims should not inform medical decision making with medicine. Randomised controlled trials and meta-analyses are

not epistemically superior to other study designs but they do produce knowledge. In many circumstances these study designs will actually provide the strongest evidence to support medical decision making. If we accept that randomised controlled trials and meta-analyses can produce knowledge should it matter that this knowledge is justified in a non-scientific way? I therefore envisage that, although the arguments presented in this thesis may lead to a re-evaluation of evidence based medicine, they should have limited impact on the wider practice of medicine.

<u>References</u>

Abulof U (2015). The Malpractice of 'Rationality' in International Relations. *Rationality and Society.* **27**(3), 358-384.

Alshenqeeti H (2014). Interviewing as a Data Collection Method: A Critical Review. *English Linguistics Research,* **3**(1), 39-45

American Heart Association (2000). Part 1: Introduction to the International Guidelines 2000 for CPR and EEC: A Consensus on Science. *Circulation.* **102**(Supplement 1): I1-I11.

Annas J (2000). *Ancient Philosophy: A Very Short Introduction.* Oxford: Oxford University Press.

Ariens G A M, van Mechelen W, Bongers P M and van der Wal G (2000). Physical Risk Factors for Neck Pain. *Scandinavian Journal of Work, Environment and Health.* **26**(1), 7-19.

Armola RR, Bourgault MA, Halm MA, Board RM, Bucher L, Harrington L, Heafey C, Lee RK, Shellner PK and Medina J. (2009). Upgrading the American Association of Critical-Care Nurses' Evidence Levelling Hierarchy. *The American Journal of Critical Care.* **18**, 405-409.

Atkins D, Eccles M, Flottorp S et al. (2004). Systems for Grading the Quality of Evidence and Strength of Recommendations I: Critical Appraisal of Existing Approaches The Grade Working Group. *BMC Health Services Research.* **4**, 38.

Atkins D, Briss PA, Eccles M et al. (2005). Recommendations for Grading the Quality of Evidence and the Strength of Recommendations II: Pilot Study of a New System. *BMC Health Services Research*. **5**, 25.

Audi R (2001). *The Architecture of Reason. The Structure and Substance of Rationality*. Oxford: Oxford University Press.

Ayer A J (1946). *Language, Truth and Logic*. New York: Dover Publications Inc.

Ball C M and Phillips R S (2001). *Evidence Based On-Call Acute Medicine*. London: Churchill Livingstone.

Ballard K and Elston M A (2005). Medicalisation: A Multi-disciplinary Concept. *Social Theory and Health*. **3**, 228-241.

Balshem H, Helfand M and Schunemann H J. (2011). GRADE guidelines: 3. Rating the Quality of Evidence. *Journal of Clinical Epidemiology*. **64**(4), 410-406.

Bandelow B, Zohar J, Hollander E, Kasper S, Moller H-J. (2002). WFSBP Guidelines for the Pharmacological Treatment of Anxiety, Obsessive-Compulsive and Posttraumatic Stress Disorders. *The World Journal of Biological Psychiatry*. **3**, 171-199.

Bandelow B, Zohar J, Kasper S, Moller H-J. (2008). How to Grade Categories of Evidence. *The World Journal of Biological Psychiatry*. **9**(4): 242-247.

Bastian H, Glasziou P and Chalmers I. (2010). Seventy-Five Trials and Eleven Systematic Reviews a Day: How will we ever keep up? *PLoS Medicine.* **7**(6), e1000326.

Batchelor P (2000). The Legal and Ethical Implication of Evidence Based Clinical Guidelines for Clinicians. *Evidence-Based Dentistry.* **2**, 5-6.

Baumann M H and Gutterman D (2006). American College of Chest Physicians Evidence-Based Guidelines- the Next Generation. Considering Resource Use and Evolution to a Single Grading System. *Chest.* **129**(1), 10-11.

Belohlavek R, Klir G, Lewis H and Way E. (2009). Concepts and Fuzzy Sets: Misunderstandings, Misconceptions and Oversights. *International Journal of Approximate Reasoning.* **51**, 23-34.

Bingeman E (2016). Evaluating Normative Epistemic Frameworks in Medicine: EBM and Casuistic Medicine. *Journal of Evaluation in Clinical Practice,* **22**, 490-495.

Bluhm R (2005). From Hierarchy to Network: A Richer View of Evidence for Evidence Based Medicine. *Perspectives in Biology and Medicine.* **48**(4), 535-547.

Blunt C J (2015). *Hierarchies of Evidence in Evidence-Based Medicine.* Ph.D. Thesis, London School of Economics and Political Science.

Bradford Hill A (1951). The Clinical Trial. *British Medical Bulletin.* **7**, 278-282.

Braunwald E, Mark D B, Jones R H (1994). *Unstable Angina: Diagnosis and Management*. Rockville (MD): Agency for Health Care Policy and Research. (AHCPR Clinical Practice Guidelines, No. 10).

Briss PA, Zaza S, Pappaioanou M et al. (2000). Developing an Evidence Based Guide to Community Preventive Services - Methods. *American Journal of Preventive Medicine*. **18**, 35-43.

Brozek JL, Akl E A, Alonso-Coello P et al. (2009). Grading Quality of Evidence and Strength of Recommendations in Clinical Practice Guidelines. Part 1 of 3. An Overview of the GRADE Approach and Grading Quality of Evidence about Interventions. *Allergy*. **64**, 669-677.

Brubaker R (1984). *The Limits of Rationality. An Essay on the Social and Moral Thought of Max Weber*. London: George Allen & Unwin.

Bruce M (1999). Evidence Based Learning in General Dental Practice. *Evidence Based Dentistry*. **1**, 20-21.

Carruthers GS, Larochelle P, Haynes R B, Petrasovits A, Schiffrin E L. (1993). Report of the Canadian Hypertension Conference: 1. Introduction. *Canadian Medical Association Journal*. **149**(3), 289-293.

Cartwright N (2007). Are RCTs the Gold Standard? *Biosocieties*. **2**, 11-20.

Cartwright N. (2011). The Art of Medicine A Philosopher's view of the long road from RCTs to effectiveness. *The Lancet*. **377**, 1400-1401.

Chalmers AF. (2010). What is this Thing Called Science? London: Open University Press.

Charlton BG (2009). The Zombie Science of Evidence Based Medicine: A Personal Retrospective. A Commentary on Djulbegovic B, Guyatt G H & Ashcroft R E (2009). Cancer Control, 16, 158-168. *Journal of Evaluation in Clinical Practice.* **15**, 930-934.

Chestnut R M, Carney N, Maynard H et al (1999). Rehabilitation for Traumatic Brain Injury. Rockville (MD): Agency for Health Care Policy and Research (US). Evidence Reports/Technology Assessments, No. 2.

Chin-Yee BH (2014). Underdetermination in Evidence Based Medicine. *Journal of Evaluation in Clinical Practice.* **20**, 921-927.

Cochrane AL (1972). *Effectiveness and Efficiency: Random Reflections on Health Services.* London: Nuffield Provincial Hospitals Trust.

Cochrane AL and Blythe M (2009). *One Man's Medicine. An Autobiography of Professor Archie Cochrane.* Cardiff: Cardiff University Publication.

Coleman K, Norris S, Weston A et al (2005). NHMRC Additional Levels of Evidence and Grades of Recommendations for Developers of Guidelines. www.nhmrc.gov.au/_files_nhmrc/file/guidelines/levels_grades05.pdf (accessed 30/9/14).

Cook DJ, Guyatt GH, Laupacis A, Sackett DL. (1992). Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents. *Chest.* **102**: 305S-311S.

Cook D J, Guyatt G H, Laupacis A, Sackett D L and Goldberg R J. (1995). Rules of Evidence and Clinical Recommendations on the use of Antithrombotic Agents. *Chest.* **108**, 227S-230S.

Couto J, S. (1998). Evidence based Medicine: A Kuhnian Perspective of a Transvestite Non-Theory. *Journal of Evaluation in Clinical Practice.* **4**(4): 267-275.

Curd M, Cover JA and Pincock C (2013). *Philosophy of Science the Central Issues.* 2nd edition. New York: W W Norton & Company.

Daly J (2005). *Evidence-based Medicine and the Search for a Science of Clinical Care.* Los Angeles: University of California Press.

DiCenso A, Bayley, Haynes BR. (2009). Accessing Pre-Appraised Evidence: Fine-tuning the 5S Model into a 6S Model. *Evidence Based Nursing.* **12**: 99-101.

Dickersin K, Strauss S E and Bero LA. (2007). Evidence Based Milestones: Increasing not Dictating Choice. *British Medical Journal*, **334**(supplement 1), 10.

Djulbegovic B and Hadley T (1998). Evaluating the Quality of Clinical Guidelines: Linking Decisions to Medical Evidence. *Oncology.* **12**, 310-314.

Djulbegovic B, Guyatt G. H., Ashcroft R. E. (2009). Epistemologic Inquiries in Evidence-Based Medicine. *Cancer Control.* **16**(2): 158-168.

Ebell M H, Siwek J, Weiss B D, Woolf S H, Susman J, Ewigman B, Bowman M (2004). Strength of Recommendation Taxonomy (SORT): A Patient Centered Approach to Grading Evidence in the Medical Literature. *American Family Physician.* **69**: 548-556.

Eccles M and Mason J (2001). How to Develop Cost-Conscious Guidelines. *Health Technology Assessment*. **5**(16), 1-78.

Etzioni A (1988). Normative-Affective Factors: Towards a New Decision-Making Model. *Journal of Economic Psychology*. **9**, 125-150.

Evans, D. (2003). Hierarchy of Evidence: A Framework for Ranking Evidence Evaluating Healthcare Interventions. *Journal of Clinical Nursing*. **12**, 73-84.

Evans I, Thornton H and Chalmers I. (2010). *Testing Treatments Better Research for Better Healthcare*. United Kingdom: Spartan Press Ltd.

Every-Palmer S and Howick J (2014). How Evidence Based Medicine is failing due to biased trials and selective publication. *Journal of Evaluation in Clinical Practice.* **20**, 908-914.

Evidence Based Medicine Working Group (1992). Evidence Based Medicine. A New Approach to Teaching the Practice of Medicine. *Journal of the American Medical Association.* **268**, 2420-2425.

Feinstein AR (1964). Scientific Methodology in Clinical Medicine III. The Evaluation of Therapeutic Response. *Annals of Internal Medicine*. **61**(5), 944-965.

Feyerabend P (1975). *Against Method*. New York: Verso Books.

Finlay L (2002). Outing the Researcher: The Provenance, Process and Practice of Reflexivity. *Qualitative Health Research.* **12(4)**, 531-545.

Fisher R. A. (1926). The Arrangement of Field Experiments. *Journal of the Ministry of Agriculture of Great Britain.* **33**, 503-513.

French K (1981). Methodological Considerations in Hospital Patient Opinion Surveys. *International Journal of Nursing Studies*. **18**, 7-32.

Fricker E. (2002). Trusting Others in the Sciences: A Priori or Empirical Warrant? *Studies in History and Philosophy of Science.* **33**, 373-383.

Fricker E (2005). Testimony. In: Borchert D M (Editor). *The Encyclopaedia of Philosophy, Volume 9,* 2nd Edition. USA: Thomson Gale. 400-403.

Fulcher J and Scott J (2011). *Sociology* (4th Edition), Oxford: Oxford University Press.

Gaeta R and Gentile N (2016). Evidence, Discovery and Justification: the Case of Evidence Based Medicine. *Journal of Evaluation in Clinical Practice*. **22**, 550-557.

Gerber A and Lauterbach KW (2005). Evidence Based Medicine: Why do Opponents and Proponents use the same Arguments? *Heath Care Analysis*. **13**(1), 59-71.

Gillis A M, Skanes A C and the CCS Atrial Fibrillation Guidelines Committee. (2011). Canadian Cardiovascular Society Atrial Fibrillation Guidelines 2010: Implementing GRADE and Achieving Consensus. *Canadian Journal of Cardiology*. **27**, 27-30.

Glasziou P, Vandenbroucke J and Chalmers I (2004). Assessing the Quality of Research. *British Medical Journal*. **328**, 39-41.

Goldbloom R (1997). Weighing the Evidence: the Canadian Experience. *American Journal of Clinical Nutrition*. **65**, 584S-586S.

Goldenberg M J (2006). On Evidence and Evidence Based Medicine: Lessons from the Philosophy of Science. *Social science & Medicine*. **62**, 2621-2632.

Goldenberg MJ. (2010). From Popperian Science to Normal Science. Commentary on Sestini (2009) 'Epistemology and Ethics of Evidence Based Medicine'. *Journal of Evaluation in Clinical Practice*. **16**, 306-309.

Goldenberg M J, Borgersen K and Bluhm R (2009). The Nature of Evidence in Evidence Based Medicine. *Perspectives in Biology and Medicine*. **52**(2), 164-167.

Goldman AI (2001). Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research*. **Vol. LXIII** (1), 85-110.

Gosall N and Gosall G. (2012). The Doctor's Guide to Critical Appraisal (3rd Edition). Lancaster: Carnegie Book Production.

Goudge J and Gilson L (2005). How Can Trust be investigated? Drawing Lessons from Past Experience. *Social Science & Medicine*. **61**, 1439-1451.

GRADE Working Group (2004). Grading Quality of Evidence and Strength of Recommendations. *British Medical Journal*. **328**, 1490-1494.

Granados A, Jonsson E, Banta HD et al (1997). EUR-ASSESS Project Subgroup Report on Dissemination and Impact. *International Journal of Technology Assessment in Health Care*. **13**(2): 220-286.

Greaves D (2002). Reflections on a New Medical Cosmology. *Journal of Medical Ethics.* **28**(2), 81-85.

Greenhalgh T and Russell J. (2009). Evidence based policymaking - A critique. *Perspectives in Biology and Medicine*. **52**(2): 304-318.

Greer N, Mosser G, Logan G, Halas G. (2000). A Practical Approach to Evidence Grading. *Joint Commission Journal on Quality Improvement*. **26**, 700-712.

Gross PA, Barrett TL, Dellinger P, Krause PJ et al. (1994). Purpose of Quality Standards for Infectious Diseases. *Clinical Infectious Diseases*. **18**(3), 421.

Grunze H, Vieta E, Goodwin GM, Bowden C, Licht R W et al. (2009). The World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for the Biological Treatment of Bipolar Disorders: Update 2009 on the Treatment of Acute Mania. *The World Journal of Biological Psychiatry*. **10**: 85-116.

Gugiu P C and Gugiu M R. (2010). A Critical Appraisal of Standard Guidelines for Grading Levels of Evidence. *Evaluations and the Health Professions*. **33**(3): 233-255.

Gugiu PC (2015). Hierarchy of Evidence and Appraisal of Limitations (HEAL) Grading System. *Evaluation & Program Planning*. **48**: 149-159.

Guyatt GH (1991). Evidence-based Medicine. *ACP Journal Club*. **114**(supplement 2), A-16.

Guyatt G, Sackett D, Sinclair J, Hayward R, Cook DJ and Cook R (1995). Users' Guide to the Medical Literature: IX. A Method for Grading Health Care

Recommendations. *Journal of the American Medical Association*. **274**, 1800-1804.

Guyatt G, Cook DJ, Sackett DL, Eckman M, Pauker S. (1998). Grades of Recommendation for Antithrombotic Agents. *Chest*. **114**(5), 441S-444S.

Guyatt G, Haynes RB, Jaeschke R et al. (2000). Users' Guides to the Medical Literature: XXV. Evidence Based Medicine: Principles for Applying the Users' Guide to Patient Care. *Journal of the American Medical Association*. **284**, 1290-1296.

Guyatt G, Schunemann H, Cook D, Jaeschke R, Pauker S and Bucher H (2001). Grades of Recommendation for Antithrombotic Agents. *Chest*. **119**(1), 3S-7S.

Guyatt G, Rennie D, Meade M O and Cook D J (2002). *Users' Guides to the Medical Literature A Manual for Evidence-Based Clinical Practice*. USA: McGraw Hill Medical.

Guyatt G, Schunemann H J, Cook D, Jaeschke R and Pauker S (2004). Applying the Grades of Recommendation for Antithrombotic and Thrombolytic Therapy. The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. **126**, 179S-187S.

Guyatt GH, Oxman A, Vist G et al. (2008). GRADE: What is 'Quality of Evidence' and why is it Important to Clinicians. *British Medical Journal*. **336**, 995-998.

Guyatt G, Rennie D, O' Meade M and Cook DJ. (2008). *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* Second Edition. USA: McGraw-Hill Companies.

Guyatt G, Oxman A, Akl EA et al. (2011a). GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables. *Journal of Clinical Epidemiology*. **64**, 383-394.

Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J et al. (2011b). GRADE Guidelines: 4. Rating the Quality of Evidence- Study Limitations (Risk of Bias). *Journal of Clinical Epidemiology*. **64**, 407-415.

Habermas J (1988). *On the Logic of the Social Sciences*. Cambridge: Polity Press.

Habib G, Lancellotti P, Antunes MJ et al (2015). 2015 ESC Guidelines for the Management of Infective Endocarditis. *European Heart Journal*. **36**, 3075-3123.

Hacking I (1988). Telepathy: Origins of Randomisation in Experimental Design. *Isis.* **79**(3): 427-451.

Hackshaw A, Paul E and Davenport E. (2006). *Evidence-Based Dentistry an Introduction*. Oxford: Blackwell Munksgaard.

Hadorn D C, Baker D, Hodges J S, Hicks N (1996). Rating the Quality of Evidence for Clinical Practice Guidelines. *Journal of Clinical Epidemiology*. **49**, 749-754.

Hammersley M. (2013). *The Myth of Research-Based Policy and Practice*. London: Sage Publications Ltd.

Hampsher-Monk I, Tilmans K and van Vree F (Editors), (1998). *History of Concepts: Comparative Perspectives*. Holland: Amsterdam University Press.

Harbour R and Miller J (2001). A New System for Grading Recommendations in Evidence Based Guidelines. *British Medical Journal*. **323**, 334-336.

Hardwig J (1991). The Role of Trust in Knowledge. *The Journal of Philosophy*. **88**(12), 693-708.

Harris R P, Helfand M, Woolf S H, Lohr K N, Mulrow C D, Teutsch S M, Atkins D (2001). Current Methods of the U.S. Preventative Services Task Force. A Review of the Process 2. *American Journal of Preventative Medicine.* **20**, 21-35.

Hawley K. (2012). *Trust A Very Short Introduction*. Oxford: Oxford University Press.

Haynes RB. (1981). How to read clinical journals: II. To learn about a diagnostic test. *Canadian Medical Association Journal*. **124**, 703-710.

Joanna Briggs Institute (2011). *Joanna Briggs Institute Reviewers' Manual*. Joanna Briggs Institute.

Haynes R B, McKibbon A K, Fitzgerald D, Guyatt G, Walker C J and Sackett DL. (1986a). How to keep up with the Medical Literature: I. Why try to keep up and how to get started. *Annals of Internal Medicine*. **105**(1), 149-153.

Haynes R B, McKibbon A K, Fitzgerald D, Guyatt G, Walker C J and Sackett D L (1986b). How to keep up with the Medical Literature: II. Deciding which journals to read regularly. *Annals of Internal Medicine*. **105**(2), 309-312.

Haynes RB, McKibbon AK, Fitzgerald D, Guyatt G, Walker CJ and Sackett DL. (1986c). How to keep up with the Medical Literature: III. Expanding the number of journals you read regularly. *Annals of Internal Medicine*. **105**(3), 474-478.

Haynes RB, McKibbon AK, Fitzgerald D, Guyatt G, Walker CJ and Sackett DL. (1986d). How to keep up with the Medical Literature: IV. Using the literature to solve clinical problems. *Annals of Internal Medicine*. **105**(4), 636-640.

Haynes RB, McKibbon AK, Fitzgerald D, Guyatt G, Walker CJ and Sackett DL. (1986e). How to keep up with the Medical Literature: V. Access by personal computer to the medical literature. *Annals of Internal Medicine*. **105**(5), 810-824.

Haynes RB, McKibbon AK, Fitzgerald D, Guyatt G, Walker CJ and Sackett DL. (1986f). How to keep up with the Medical Literature: VI. How to store and retrieve articles worth keeping. *Annals of Internal Medicine*. **105**(6), 978-984.

Haynes R. (2001). Of Studies, Syntheses, Synopses and Systems: the '4S' Evolution of Services for Finding Current Best Evidence. *Evidence Based Medicine*. **6**, 36-38.

Haynes RB (2002). What Kind of Evidence is it that Evidence Based Medicine Advocates want Health Care Providers and Consumers to pay Attention to? *BMC Health Services Research*. **2**, 3.

Haynes RB, McKibbon KA, Wilczynski NL et al (2005). Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *British Medical Journal*. **330**, 1179-1185.

Haynes R (2006). Of Studies, Syntheses, Synopses, Summaries and Systems: the '5S' Evolution of Information Services for Evidence Based Healthcare Decisions. *Evidence Based Medicine*. **11**, 162-164.

Haynes R B, Sackett D L, Guyatt G H and Tugwell P (2006). *Clinical Epidemiology: How to do Clinical Practice Research.* Philadelphia, USA: Lippincott Williams & Wilkins.

Higgins JPT and Green S (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration.

Hilt J (2001). The Year in Ideas: A to Z.; Evidence-Based Medicine. *New York Times*. Dec 9.

Hirsh J, Dalen JE, Guyatt G (2001). The Sixth (2000) ACCP Guidelines for Antithrombotic Therapy for Prevention and Treatment of Thrombosis. *Chest*. **119**, 1S-2S.

Hofmeijer J (2014). Evidence based medical knowledge: The neglected role of expert opinion. *Journal of Evaluation in Clinical Practice*. **20**, 803-808.

Hooper V, Chard R, Clifford T, Fetzer S, Fossum S, Godden B et al. (2009). ASPAN's Evidence-Based Clinical Practice Guideline for the Promotion of Perioperative Normothermia. *Journal of Peri-anesthetic Nursing.* **24**(5), 271-287.

Howick J, Chalmers I, Glasziou P et al. (2011). Explanation of the 2011 Oxford Centre for Evidence Based Medicine Levels of Evidence (Background Document). Oxford Centre for Evidence Based Medicine.

Howick J (2011). *The Philosophy of Evidence Based Medicine.* Chichester: Wiley-Blackwell.

Hyde P (2004). Fool's Gold: Examining the Use of Gold Standards in the Production of Research Evidence. *British Journal of Occupational Therapy.* **67**(2), 89-94.

Iqbal A, Glenny A-M (2002). General Dental Practitioners' Knowledge of and Attitudes towards Evidence-Based Practice. *British Dental Journal.* **192**, 587-591.

Jaeschke R and Colleagues (2008). Use of GRADE Grid to Reach Decisions on Clinical Practice Guidelines when Consensus is Elusive. *British Medical Journal.* **337**, 327-330.

Joanna Briggs Institute (2011). *Joanna Briggs Institute Reviewers' Manual*, Joanna Briggs Institute.

Kalberg S (1980). Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalisation Processes in History. *The American Journal of Sociology.* **85**(5), 1145-1179.

Kelleher D, Gabe J and Williams G (2006). *Challenging Medicine*. 2nd Edition. Oxford: Routledge.

Kerry R, Eriksen T E, Noer S A, Mumford S D and Anjum R L (2012). Causation and Evidence Based Practice: An Ontological Review. *Journal of Evaluation in Clinical Practice*. **18**, 1006-1012.

King M (2009). *Systems, not People, make Society Happen*. United Kingdom: Holcombe Publishing.

Koselleck R (1989). Social History and Conceptual History. *International Journal of Politics, Culture and Society*. **2**(3), 308-325.

Kuhn TS (1996). *The Structure of Scientific Revolutions*. 3rd Edition. Chicago: University of Chicago Press.

Kukla A (1991). Criteria of Rationality and the Problem of Logical Sloth. *Philosophy of Science*. **58**, 486-490.

Kulkarni A V (2005). The Challenges of Evidence Based Medicine: A Philosophical Perspective. *Medicine, Health Care and Philosophy*. **8**, 255-260.

Ladyman J (2002). *Understanding Philosophy of Science*. London: Routledge.

Lakatos I (1970). *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Lambert H. (2006). Accounting for Evidence Based Medicine: Notions of Evidence in Medicine. *Social Science & Medicine*. **62**, 2633-2645.

La Caze A (2009). Evidence Based Medicine Must Be… *Journal of Medicine and Philosophy*. **34**, 509-527.

La Caze A (2011). The Role of Basic Science in Evidence Based Medicine. *Biology and Philosophy*. **26**, 81-98.

Lassere M (2008). The Biomarker-Surrogacy Evaluation Schema: A Review of the Biomarker-Surrogate Literature and a Proposal for a Criterion-Based, Quantitative, Multidimensional Hierarchical Levels of Evidence Schema for Evaluating the Status of Biomarkers as Surrogate Endpoints. *Statistical Methods in Medical Research*. **17**(3), 303-340.

Latour B and Woolgar S (1986). *Laboratory Life the Construction of Scientific Facts*. USA: Princeton University Press.

Leibovici L (2001). Effects of Remote Retroactive Intercessory Prayer on Outcomes in Patients with Bloodstream Infection: Randomised Controlled Trial. *British Medical Journal*. **323**, 1450-1451.

Light H (1991). Professionalism as a Countervailing Power. *Journal of Health Politics, Policy and Law*. **16**(3), 499-513.

Lipton P. (1998). The Epistemology of Testimony. *Studies in the History and Philosophy of Science*. **29**(1), 1-31.

MacDonald K M (1995). *The Sociology of the Professions*, London: Sage Publications.

Manger D, Walshaw M, Fitzgerald R, Doughty J et al. (2017). Evidence Summary: The Relationship between Oral Health and Pulmonary Disease. *British Dental Journal*. **222**(7), 527-533.

Marshall G, Blacklock JWS, Cameron C et al. (1948). Streptomycin treatment of pulmonary tuberculosis. A Medical Research Council Investigation. *British Medical Journal*. **ii**, 769-782.

Masterman M. (1970). The Nature of a Paradigm in Lakatos I. and Musgrave A. Editors. *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Maturo A (2012). Medicalisation: Current Concepts and Future Directions in a Bionic Society. *Mens Sana Monographs*. **10**(1), 122-133.

McNeill T (2006). Evidence Based Practice in an Age of Relativism: Towards a Model for Practice. *Social Work*. **51**, 147-155.

Mebius A (2014). Corroborating evidence based medicine. *Journal of Evaluation in Clinical Practice.* **20**, 915-920.

Meltzer S, Leiter L, Daneman D, Gerstein HC, Yale J-F, Zinman B et al (1998). Clinical Practice Guidelines for the Management of Diabetes in Canada. *Canadian Medical Association Journal*. **159**(8), S1-S29.

Merlin T, Weston A, Tooher R (2009). Extending an Evidence Hierarchy to Include Topics Other than Treatment: Revising the Australian Levels of Evidence. *BMC Medical Research Methodology*. **9**, 34.

Mitchell SA and Friese CR (2007). www.ons.org/practice-resources/pep/evaluation-process, accessed 15/05/16.

Morley PT, Atkins DL, Billi JE, Bossaert L, Callawat CW, de Caen AR et al (2010). Part 3: Evidence Evaluation Process. *Resuscitation.* **81**(S), e32-e40.

Moynihan R and Smith R (2002). Too Much Medicine? *British Medical Journal.* **324**, 859-860.

Mykhalovskiy E and Weir L (2004). The problem of Evidence Based Medicine: Directions for Social Science. *Social Science and Medicine.* **59**(5), 1059-1069.

Needleman I, Moles DR, Worthington H. (2005). Evidence-based periodontology, systematic reviews and research quality. *Periodontology 2000.* **37**(1), 12-28.

Newhouse RP, Dearholt SL, Poe SS et al. (2007) John Hopkins Nursing Evidence-Based Practice Model and Guidelines. Sigma Theta Tau International.

Newton-Smith W. H. (1981). *The Rationality of Science.* Oxford: Routledge & Kegan Paul Ltd.

NICE (2015). Prophylaxis against Infective Endocarditis, Clinical Guideline 64.1.

Oxford English Dictionary (2006). Editors Soanes A and Stevenson A. Oxford: Oxford University Press.

Panda S C (2006). Medicine: Science or Art? *Mens Sans Monographs.* **4(1)**, 127-138.

Pigliucci M (2015). Scientism and Pseudoscience: A Philosophical Commentary. *Journal of Bioethical Inquiry.* **12**(4), 569-575.

Pope C, Mays N and Popay J (2007). *Synthesising Qualitative and Quantitative Health Evidence: A Guide to Methods.* England: Open University Press.

Popper K (1963). *Science: Conjectures and Refutations.* London: Routledge &Kegan Paul.

Preston A (2006). Analytic Philosophy. Internet Encyclopaedia of Philosophy.

Rawlins MD (2008). *De Testimonio on the Evidence for Decisions about the use of Therapeutic Interventions.* The Harveian Oration 2008, Royal College of Physicians.

Resnik D (2000). A Pragmatic Approach to the Demarcation Problem. *Studies in History and Philosophy of Science.* **31**(2), 249-267.

Richards, D. (2005). Outcomes, What Outcomes. *Evidence Based Dentistry.* **6**, 1.

Richards, D. (2009). GRADING - Levels of Evidence. *Evidence Based Dentistry.* **10**: 24-25.

Rosenberg W. and Donald A. (1995). Evidence Based Medicine: An Approach to Clinical Problem Solving. *British Medical Journal.* **310**, 1122-1126.

Ruse M (1982). Creation-Science is not Science. *Science, Technology and Human Values.* **7**(40), 72-78.

Sackett D (1979). Bias in Analytic Research. *Journal of Chronic Disease.* **32**, 51-63.

Sackett D (1981). How to read clinical journals: V. to distinguish useful from useless or even harmful therapy. *Canadian Medical Association Journal.* **124**, 1156-1162.

Sackett D (1989). Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents. *Chest.* **95**, 2S-4S.

Sackett D, Rosenberg W, Muir Gray J A, Haynes R B, Richardson W S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal.* **312**, 71-72.

Satterfield J, Spring B, Brownson RC et al. (2009). Towards a Transdisciplinary Model of Evidence Based Practice. *The Millbank Quarterly.* **87**, 368-390.

Saunders J (2000). The Practice of Clinical Medicine as an Art and as a Science. *Journal of Medical Ethics.* **26**, 18-22.

Sayre M R, O'Connor R E, Atkins D L, Billi J E, Callaway C W, Shuster M et al (2010). Part 2: Evidence Evaluation and Management of Potential or Perceived Conflicts of Interest. 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation.***122**(Suppl 3): S657-S664.

Schfranski M D (2012). Evidence-Based Medicine – New Paradigm or Pseudoscience. *Brazilian Journal of Cardiovascular Surgery.* **27**(2): 334-337.

Schunemann A, Oxman A, Brozek J et al. (2008). GRADE: Grading Quality of Evidence and Strength of Recommendations for Diagnostic Tests and Strategies. *British Medical Journal.* **336**: 1106-1110.

Scottish Intercollegiate Guidelines Network (2011): *SIGN 50 A Guideline Developers Handbook.* Scottish Intercollegiate Guidelines Network.

Sehon SR and Stanley D. (2003). A Philosophical Analysis of the Evidence Based Medicine Debate. *BMC Health Services Research.* **3**, 14.

Senn SJ (1991). Falsificationism and Clinical Trials. *Statistics in Medicine.* **10,** 1679-1692.

Sestini P. (2010). Epistemology and Ethics of Evidence Based Medicine: Putting Goal-setting in the Right Place. *Journal of Evaluation in Clinical Practice.* **16**, 301-305.

Shadish W R, Cook T D and Campbell D T. (2002). *Experimental and Quasi-experimental Designs for Generalised Causal Inference.* USA: Wadsworth Cengage Learning.

Shahar E. (1997). A Popperian Perspective of the Term 'Evidence Based Medicine'. *Journal of Evaluation in Clinical Practice.* **3**(2), 109-116.

Shahar E. (1998). Evidence-based Medicine: A New Paradigm of the Emperor's new Clothes? *Journal of Evaluation in Clinical Practice.* **4**(4), 277-282.

Shand J (2000). *Arguing Well.* London: Routledge.

Silva SA and Wyer PC. (2009). Where is the Wisdom? II – Evidence-based Medicine and the Epistemological Crisis in Clinical Medicine. Exposition and Commentary on Djulbegovic, B., Guyatt, G. H. & Ashcroft, R. E. (2009). Cancer Control, 16, 158-168. *Journal of Evaluation in Clinical Practice.* **15,** 899-906.

Simmons (2005). Truth in Borchert, D. M. (Editor in Chief). *Encyclopaedia of Philosophy* (2nd Edition) Volume 9, page 534-452. USA: Thomson Gale.

Skinner Q (2002). *Visions of Politics Volume 1 Regarding Method.* Cambridge: Cambridge University Press.

Smith, P. (2003). *An Introduction to Formal Logic.* Cambridge: Cambridge University Press.

Sorell T (1991). *Scientism: Philosophy and the Infatuation with Science.* London: Routledge.

Spitzer W O, Bayne R D, Charron K C et al (1979). The Periodic Health Examination. Canadian Task Force on the Periodic Health Examination. *Canadian Medical Association Journal.* **121**, 1193-1254.

Steelman V M, Pape T, King C A, Graling P, Gaberson K B (2011). Selection of a Method to Rate the Strength of Scientific Evidence for AORN Recommendations. *AORN Journal.* **93**, 433-443.

Stetler C, Brunell M, Giullano K, Morsi D, Prince L, Newell-Stokes V (1998). Evidence-Based Practice and the Role of Nursing Leadership. *Journal of Nursing Administration*. **28**, 45-63.

Storey J, Bullivant J and Corbett-Nolan A (2011). *Governing the New NHS Issues and Tensions in Health Service Management*. Oxford: Routledge.

Straus SE, Haynes RB, Glasziou P, Dickersin K and Guyatt G (2007). Misunderstandings, Misperceptions and Mistakes. *ACP Journal Club*. **146**, A8.

Straus S E, Glasziou P, Richardson W S, Haynes R B (2011). *Evidence-Based Medicine How to Practice and Teach it.* Toronto: Elsevier Ltd.

Straus SE and McAlister FA (2000). Evidence Based Medicine: A Commentary on Common Criticisms. *Canadian Medical Association Journal*. **163**, 837-841.

Stroup D F, Berlin J A, Morton S C et al (2000). Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *Journal of the American Medical Association*. **283**, 2008-2012.

Swinkels A, Albarran J W, Means R I, Mitchell T, Stewart M C (2002). Evidence-Based Practice in Health and Social Care: Where Are We Now? *Journal of Interprofessional Care*. **16**, 335-347.

Tanenbaum S.J. (2014). Particularism in Health Care: Challenging the Authority of the Aggregate. *Journal of Evaluation in Clinical Practice*. **20**, 934-941.

Thagard PR (1978). Why Astrology is a Pseudoscience. *Proceedings of the Philosophy of Science Association*. **1**, 223-234.

Thompson, R. (2010). Causality, Mathematical Models and Statistical Association: Dismantling Evidence Based Medicine. *Journal of Evaluation in Clinical Practice*. **16**, 267-275.

Thornhill M, Dayer M, Lockhart PB et al (2016). Guidelines on Prophylaxis to Prevent Infective Endocarditis. *British Dental Journal*. **220**, 51-56.

Timmermans S and Berg M (2003). *The Gold Standard. The Challenge of Evidence-Based Medicine and Standardisation in Health Care.* Philadelphia: Temple University Press.

Timmermans S and Kolver E (2004). Evidence –based medicine and the reconfiguration of medical knowledge. *Journal of Health and Social Behaviour*. **45**, 177-193.

Tomlin G and Borgetto B (2011). Research Pyramid: A New Evidence-Based Practice Model for Occupational Therapy. *The American Journal of Occupational Therapy*. **65**, 189-196.

Tonelli M. (1998). The Philosophical Limits of Evidence-based Medicine. *Academic Medicine*. **73**(12), 1234-1240.

Tonelli M R (2009). Evidence-Free Medicine: Forgoing Evidence in Clinical Decision Making. *Perspectives in Biology and Medicine.* **52**(2), 319-331.

Trout KS (1981). How to read clinical journals: IV. To determine aetiology or causation. *Canadian Medical Association Journal.* **124**, 985-990.

Turner A (2011). *'Placebos', Evidence Based Medicine and the Homeopathy Controversy.* Ph.D. Thesis: University of Nottingham.

Turner A (2012). 'Placebos' and the logic of placebo comparison. *Biology and Philosophy.* **27**, 419-432.

Turpen R M, Fesperman S F, Sultan S, Borawski K M, Norris R D, Klink J et al (2010). Levels of Evidence Ratings in the Urological Literature: An Assessment of Inter-observer Agreement. *BJU International.* **105**, 602-606.

Upshur R (2003). Are all Evidence Based Practices Alike? Problems in the Ranking of Evidence. *Canadian Medical Association Journal.* **169**, 672-673.

Upshur R (2009). Making the Grade, Assuring Trustworthiness in Evidence. *Perspectives in Biology and Medicine.* **52**(2), 264-275.

Walton D (2008). *Informal Logic: A Pragmatic Approach.* 2nd Edition. Cambridge: Cambridge University Press.

Waugh N (1999). Grading of Scientific Evidence a Discussion Paper from SHPIC. *Health Bulletin.* **57**(1), 53-63.

Wennberg J E and Gittelsohn A M (1973). Small Area Variations in Health Care Delivery. *Science.* **182**(117), 1102-1108.

West S, King V, Carey TS, et al. (2002) Systems to Rate the Strength of Scientific Evidence. Evidence. Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute–University of North Carolina Evidence-based Practice Centre under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality.

Wilkinson CP (1999). Evidence Based Medicine Regarding the Prevention of Retinal Detachment. *Transactions of the American Ophthalmological Society.* **XCVII**, 397-406.

Worrall J. (2002). What Evidence in Evidence-Based Medicine. *Philosophy of Science.* **69**(S3), S316-S330.

Worrall J (2007). Evidence in Medicine and Evidence Based Medicine. *Philosophy Compass.* **2**(6), 981-1022.

Worrall J (2010). Evidence: Philosophy of Science meets Medicine. *Journal of Evaluation in Clinical Practice.* **16**, 356-362.

Young J M, Glasziou P, Ward J E (2002). General Practitioners' Self Ratings of Skills in Evidence Based Medicine: Validation Study. *British Medical Journal.* **324**, 950-951.

Zhang W, Doherty M, Pascual E et al (2006). EULAR evidence-based recommendations for gout. Part 1: Diagnosis. Report of a task force for the

standing committee for international clinical studies including therapeutics.

Annals of Rheumatic Diseases **65**, 1301-1311.

Number	Search Terms	Results
1	*Evidence-Based Medicine/	24012
2	*Evidence-Based Dentistry/	521
3	*Evidence-Based Nursing/	2000
4	*Evidence-Based Practice/	3526
5	1 or 2 or 3 or 4	30008
6	*Philosophy/	6389
7	*Philosophy, Medical/	5571
8	*Knowledge/	5149
9	*Empiricism/	137
10	"epistemol*".ab,sh,ti.	3925
11	logical positivism.ab,sh,ti.	42
12	"falsification*".ab,sh,ti.	657
13	popper.ab,kw,ti.	417
14	"paradigm*".ab,sh,ti.	117177
15	"Kuhn*".ab,sh,ti.	1464
16	"research program*".ab,sh,ti.	8672
17	lakatos.ab,sh,ti.	48
18	rationalism.ab,sh,ti.	207
19	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18	147128
20	5 and 19	889
21	limit 20 to English language	771

Table 6: Medline Search using the Ovid Interface 12/08/17

Line	Search Term	Number of Hits
1	SU.EXACT("Medicine")	9297
2	SU.EXACT("philosophy of science")	4602
3	1 AND 2	55

Table 7: IBSS Search undertaken 16/04/13

	Search Term	Results
1	*Evidence-Based Medicine/	22020
2	*Evidence-Based Dentistry/	464
3	*Evidence-Based Practice/	2778
4	*Evidence-Based Nursing/	1604
5	evidence based medicine.ab.ti.kw.	10664
6	evidence based dentistry.ab.ti.kw.	455
7	evidence based practice.ab.ti.kw.	7178
8	evidence based nursing.ab.ti.kw.	639
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8	37664
10	levels of evidence.ab.ti.kw.	5277
11	hierarchy of evidence.ab.ti.kw.	235
12	grading of evidence.ab.ti.kw.	177
13	quality of evidence.ab.ti.kw.	7547
14	10 or 11 or 12 or 13	13083
15	9 and 14	1392
16	limit 15 to (abstracts and English language)	1178

Table 8: MEDLINE search strategy via the OVID interface (26/04/16)

	Search Term	Hits
S1	Evidence based medicine	2749
S2	Evidence based dentistry	150
S3	Evidence based practice	29893
S4	Evidence based nursing	5844
S5	MM "Medical Practice, Evidence-Based"	3352
S6	MM "Nursing Practice, Evidence-Based"	3237
S7	MM "Professional Practice, Evidence-Based"	4304
S8	MM "Evidence-Based Dental Practice"	6
S9	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8	31584
S10	"level* of evidence"	5584
S11	"grading of evidence"	137
S12	"quality of evidence"	4042
S13	"hierarchy of evidence"	130
S14	S10 OR S11 OR S12 OR S13	9235
S15	S9 AND S14	2019
S16	Limiters: abstract available; English language; Peer reviewed; Exclude MEDLINE records	332

Table 9: CINAHL Search Strategy (19/01/12)

"level* of evidence" ti.ab.kw
OR "hierarchy of evidence" ti.ab.kw
OR "grading of evidence" ti.ab.kw
OR "quality of evidence" ti.ab.kw

Table 10: Cochrane Methodology Register Search Strategy

Appendix 3: Hierarchies of Evidence listed in Chronological Order:

1. Cochrane (1972):

Level of Evidence	Study design
1	Randomised Controlled Trial
2	Observational Study
3	Clinical Opinion

2. Canadian Task Force (Spitzer et al 1979):

I	Randomised controlled trial
II-1	Cohort or case-control study
II-2	Historical control
III	Expert opinion

3. How to Read Clinical Journals (Trout 1981):

Methodological Strength	Study Design
Strongest	Randomised controlled trial
	Cohort Study
	Case-control Study
Weakest	Case Series

4. American College of Chest Physicians (1986), (Sackett 1989):

Level of Evidence	Study Design
I	Randomised controlled trials with low alpha and beta errors (high power)
II	Randomised Controlled Trials with high alpha and beta errors (interesting but not significant, may be useful in future meta-analyses)
III	Non randomised concurrent cohort study
IV	Non randomised historical cohort comparison
V	Case series

5: Canadian Task Force Modification 1 (1989), (Goldbloom 1997):

1	At least one randomised controlled trial
2-1	Well-designed controlled trials without modification
2-2	Well-designed cohort or case-control analytic studies preferably from more than one centre or group
2-3	Multiple time series with or without intervention
3	Expert opinion, clinical experience or descriptive studies

6. American College of Chest Physicians (modification 1), (Cook et al 1992):

LOE	Definition
I	Results come from a single RCT[92] in which the lower limit of the confidence interval for the treatment effect exceeds the minimal clinically important benefit
I+	Results come from a meta-analysis of RCTs in which the treatment effects from individual studies are consistent, and the lower limit of the confidence interval for the treatment effect exceeds the minimal clinically important benefit
I-	Results come from a meta-analysis of RCTs in which the treatment effects from individual studies are widely disparate, but the lower limit of the confidence interval for the treatment effect still exceeds the minimal clinically important benefit
II	Results come from a single RCT in which the confidence interval for the treatment effect overlaps the minimal clinically important benefit
II+	Results come from a meta-analysis of RCTs in which the treatment effects from individual studies are consistent and the confidence interval for the treatment effect overlaps the minimal clinically important benefit
II-	Results come from a meta-analysis of RCTs in which the treatment effects from individual studies are widely disparate and the confidence interval for the treatment effect overlaps the minimal clinically important benefit
III	Results come from non-randomised concurrent cohort studies
IV	Results come from non-randomised historic cohort studies
V	Results come from case series

---

[92] RCT = Randomised controlled trial.

7. American Association of Critical Care Nurses (1993), (Armola et al 2009):

Level	Description
I	Manufacturers recommendation only
II	Theory based, no research data to support recommendations; recommendations from expert consensus groups may exist
III	Laboratory data only, no clinical data to support recommendations
IV	Limited clinical studies to support recommendations
V	Clinical studies in more than 1 or 2 different populations and situations to support recommendations
VI	Clinical studies in a variety of populations of patients and situations to support recommendations

8. Canadian Hypertension Society (Carruthers et al 1993):

Level	Criteria
I	A randomised controlled trial (RCT) that demonstrates a statistically significant difference in at least one important outcome e.g. 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.
II	An RCT that does not meet level I criteria.
III	A non-randomised trial with contemporaneous controls selected by some systematic method (i.e. not selected by perceived suitability for one of the treatment options for individual patients) OR Subgroup analysis of a randomised trial.
IV	A before-after study or case series (of at least 10 patients) with historical controls or controls drawn from other studies.
V	Case series (at least 10 patients) without controls.
VI	Case report (fewer than 10 patients).

9. Infectious Diseases Society of America (Gross et al 1994):

Quality of Evidence	Definition
I	Evidence from 1+ properly randomised, controlled trial
II	Evidence from 1+ well designed clinical trial without randomisation, from cohort or case-controlled analytic studies (preferably 1+) , from multiple time series or dramatic results from uncontrolled experiments
III	Evidence from opinions of respected authorities, based on clinical experience, descriptive studies or reports of expert committees

10. Braunwald et al (1994):

Strength of Evidence	Definition
A	Randomised controlled trials
B	Well-designed clinical studies
C	Panel consensus, studies with methodological deficiencies

11. Evidence Based Medicine Working Group (Guyatt et al 1995):

Level of Evidence	Definition
A1	Randomised controlled trials, no heterogeneity, confidence intervals all on one side of threshold number need to treat.
A2	Randomised controlled trials, no heterogeneity, confidence intervals overlap threshold number needed to treat.
B1	Randomised controlled trials, heterogeneity, confidence intervals all on one side of threshold number needed to treat.
B2	Randomised controlled trials, heterogeneity, confidence intervals overlap threshold number needed to treat.
C1	Observational studies, confidence intervals all on one side of threshold number needed to treat.
C2	Observational studies, confidence intervals overlap threshold number needed to treat.

12. Australian National Health and Medical Research Council (1995), (Gugiu and Gugiu 2010):

LOE	Definition
I	Evidence obtained from a systematic review of all relevant randomised controlled trials.
II	Evidence obtained from at least one properly designed RCT.
III-1	Evidence obtained from well-designed pseudo-randomised controlled trial (alternate allocation or some other method).
III-2	Evidence obtained from comparative studies (including a systematic review of these studies) with concurrent controls and allocation not randomised, cohort studies, case-control studies or interrupted time series with a control group.
III-3	Evidence obtained from comparative studies with historical control, two or more single arm studies, or interrupted time series without a parallel control group.
IV	Evidence obtained from case series, either post-test or pre-test/post-test.

13. Agency for Healthcare Research and Quality (Hadorn et al 1996):

LOE	Definition
1	Supportive evidence from well conducted randomised controlled trials that included 100 patients or more
A	Evidence from a well conducted multicentre trial
B	Evidence from a meta-analysis that incorporated quality ratings in the analysis and included a total of 100 patients in its estimate of effect size and confidence intervals.
2	Supportive evidence from well conducted randomised controlled trials that included fewer than 100 patients
A	Evidence from a well conducted trial at one or more institution
B	Evidence from a meta-analysis that incorporated quality ratings in the analysis and included fewer than 100 patients in its estimate of effect size and confidence intervals.
3	Supportive evidence from well conducted cohort studies
A	Evidence from a well conducted prospective cohort study or registry
B	Evidence from a well-conducted retrospective cohort study
C	Evidence from a well conducted meta-analysis of cohort studies
4	Supportive evidence from a well conducted case-control study
5	Supportive evidence from poorly controlled or uncontrolled studies
A	Evidence from a randomised controlled trial with one or more major or three or more minor methodological flaws that could invalidate the results
B	Evidence from observational studies with a high potential for bias (such as a case series with comparison to historical controls)
6	Conflicting evidence with the weight of evidence supporting the recommendation
7	Expert opinion

14. EUR-ASSESS (Granados et al 1997):

Level of Evidence	Definition
1	Strong: based on empirical evidence, including experimental and quasi-experimental data
2	Moderate: clear consensus among committee members
3	Weak: insufficient evidence, but viewed worth considering by committee members

15. American College of Chest Physicians (modification 2), (Guyatt et al 1998):

LOE	Study design	Explanation
A	Randomised controlled trials, no heterogeneity	Methods strong, results consistent
B	Randomised controlled trials, heterogeneity present	Methods strong, result inconsistent
C	Observational studies	Methods weak

16. Djulbegovic and Hadley (1998):

Level of Evidence	Description	Relative Strength
Level I	Well-designed prospective randomised controlled trial	5
Level II	A single arm, prospective study	3
Level III	Retrospective/anecdotal data	1

17. Canadian Diabetes Association (Meltzer et al 1998):

Level	Criteria
1+	Systematic overview or meta-analysis of randomised controlled trials
1	1 randomised controlled trial with adequate power
2+	Systematic overview or meta-analysis of level 2 randomised controlled trials
2	RCT that does not meet level 1 criteria
3	Non randomised clinical trial or cohort study
4	Before-after study, cohort study with non-contemporaneous controls, case-control study
5	Case series without controls
6	Case report or case series of <10 patients

18. Stetler et al (1998):

Level	Type of Evidence
I	Meta-analysis of controlled studies
II	Individual experimental studies
III	Quasi-experimental studies such as non-randomised controlled single group, pre-post, cohort, time series or matched case-controlled studies
IV	Non-experimental studies such as comparative and correlational descriptive research as well as qualitative studies
V	Program evaluation, research utilisation or quality improvement projects or case reports
VI	Opinions of respected authorities or an expert committee including their interpretation of non-research based information

19. Scottish Intercollegiate Guideline Network (1998), (Harbour and Miller 2001):

LOE	Definition
1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well-conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1-	Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies. High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2+	Well-conducted case control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal
2-	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, e.g. case reports, case series
4	Expert opinion

20. Oxford Levels of Evidence (1999), (Ball and Phillips 2001) – Therapy, prevention, aetiology and harm hierarchy only:

Level	Definition
1a	SR[93] (with homogeneity) of RCTs
1b	Individual RCT (with narrow confidence interval)
1c	All or none study
2a	SR (with homogeneity) of cohort studies
2b	Individual cohort study (including low quality RCT e.g. <80% follow-up)
2c	Outcomes research
3a	SR (with homogeneity) of case control studies
3b	Individual case-control study
4	Case-series (and poor quality cohort and case-control studies)
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

21. American Academy of Ophthalmology (Wilkinson 1999):

Level	Definition
I	Data that provided strong evidence in support of the recommendation. The design of the study addressed the issue in question and the study was performed in the population of interest and executed in a manner that ensured production of accurate and reliable data, using appropriate statistical methods
II	Data that provided substantial evidence in support of the recommendation. The study had selected attributes of Level I support but lacked one or more of the components of Level I
III	Consensus of opinion in the absence of evidence that met Levels I and II

22. Chestnut et al (1999):

Class	Definition
I	Properly designed randomised controlled trials
II a	Randomised controlled trials that contain design flaws e.g. failure to blind or lack of follow up data. Multi-centre or population based longitudinal cohort studies
II b	Non-randomised controlled trials. Case-control studies. Case series with adequate description of the patient population, interventions and outcomes measured
III	Descriptive studies (uncontrolled case series). Expert opinion. Case reports. Clinical experience.

---

[93] SR = Systematic review.

23. Evidence Based Medicine Working Group (modification 1), (Guyatt et al 2000):

n-of-1 randomised trial
Systematic Review of randomised trials
Single randomised trial
Systematic review of observational studies addressing patient important outcomes
Single observational study addressing patient important outcomes
Physiologic study
Unsystematic clinical observation

24. Cardiopulmonary Resuscitation Guidelines (American Heart Association 2000):

Evidence Level	Definition
1. Positive RCTs (p<0.05)	A prospective RCT. Concluding that new treatment significantly better (or worse) than control treatment
2. Neutral RCTs (non-significant)	An RCT concluding new treatment no better than control treatment
3. Prospective, non-random	Non-randomised prospective observational study of a group that uses new treatment. Must have a control group for comparison
4. Retrospective, non-random	Non-randomised retrospective observational study where one group uses new treatment. Must have a control group for comparison
5. Case series	Series of patients received new treatment in past or will receive in future
6. Animal studies (A and B)	Studies using animals or mechanical models. A level studies are higher than B level studies
7. Extrapolations	Reasonable extrapolations from existing data or data gathered for other purposes; quasi-experimental designs
8. Rational conjecture, common sense	Fits with common sense; has face validity; applies to many non-evidence based guidelines that "made sense". No evidence of harm

25. Institute for Clinical Systems Improvement (Greer et al 2000):

A	Randomised Controlled Trial
B	Cohort study
C	Case-control study, non-randomised study with concurrent or historical control, population based study or study of the sensitivity/specificity of a diagnostic test
D	Cross-sectional study, case series or case report

26. United States Task Force on Community Preventive Services (Briss et al 2000):

Suitability	Attributes
Greatest	Concurrent comparison groups and prospective measurement of exposure and outcome
Moderate	All retrospective designs or multiple pre or post measurements but no concurrent comparison group
Least	Single pre and post measurements and no concurrent comparison group or exposure and outcome measured in single group at the same point in time

27. American College of Chest Physicians modification 3 (Guyatt et al 2001):

Grade	Methodological Strength of Supporting Evidence
1A	RCTs without important limitations
1C+	No RCTs but strong RCT results can be unequivocally extrapolated, or overwhelming evidence from observational studies
1B	RCTs with important limitations (inconsistent results, methodological flaws)
1C	Observational studies

28. Stetler (modification 1), (2001), (Hooper et al 2009):

Level of Evidence	Strength of Individual External Evidence
I	Systematic statistical review of multiple controlled studies (e.g. meta-analysis)
II	Systematic interpretive, tabular integrative review of multiple studies, primarily of quantitative research
III	Experimental studies, may be called randomised controlled trials
IV	Quasi-experimental studies, such as non-randomised control group studies or interventional time series
V	Systematic interpretive, tabular integrative review of multiple studies, primarily of qualitative research
VI	Non-experimental studies such as correlational and descriptive research, as well as qualitative research or systematically designed case studies
VII	Systematically obtained, verifiable quality or program evaluation data from the literature
VIII	Consensus opinion of respected authorities (e.g. a nationally known guideline group)

29. World Federation of Societies of Biological Psychiatry (Bandelow et al 2002):

Category of Evidence	Definition
A	Positive evidence 2 or more randomised double blind studies showing superiority to placebo **AND** one or more positive double blind study showing superiority to or equivalent efficacy as established comparator drug. In case of negative existing studies (studies showing non superiority to placebo or inferiority to comparator drug) these must be outweighed by at least two more positive studies. Studies must fulfil established methodological criteria.
B	Preliminary positive evidence.
B1	One or more randomised double-blind study showing superiority to placebo and no negative studies.
B2	One or more positive naturalistic open studies and no negative studies.
B3	One or more positive case reports and no negative studies exist.
C	Inconsistent results. Controlled positive studies are outweighed by an approximate equal number of negative studies.
D	Negative evidence. The majority of controlled studies show non superiority to placebo or inferiority to comparator drug.
F	Lack of evidence. Adequate studies providing efficacy or non-efficacy are lacking.

30. Evans (2003):

	Effectiveness	Appropriateness	Feasibility
Excellent	Systematic Review Multi-centre randomised controlled trials	Systematic Review Multi-centre randomised controlled trials	Systematic Review Multi-centre randomised controlled trials
Good	Randomised controlled trial Observational studies	Randomised controlled trial Observational studies Interpretive studies	Randomised controlled trial Observational studies Interpretive studies
Fair	Uncontrolled studies with dramatic results Before and after studies Non-randomised controlled trials	Descriptive studies Focus groups	Descriptive studies Action research Before and after studies Focus groups
Poor	Descriptive studies Case studies Expert opinion Studies of poor methodological quality	Expert opinion Case studies Studies of poor methodological quality	Expert opinion Case studies Studies of poor methodological quality

31. Strength of Recommendation Taxonomy (Ebell et al 2004):

Level 1	Systematic review or meta-analysis with consistent findings. High quality individual randomised controlled trial with allocation concealment, blinding if possible, intention to treat analysis, adequate statistical power and adequate loss to follow up or 'all or none' study.
Level 2	Systematic review of lower quality clinical trials or studies with inconsistent findings, lower quality clinical trial, cohort study or case control study.
Level 3	Consensus guidelines, extrapolations from bench research, usual practice, opinion, disease orientated evidence (surrogate outcomes), or case series for studies of diagnosis, treatment, prevention or screening.

32. Grading of Recommendations Assessment Development and Evaluation (2004), (Brozek et al 2009):

Quality of Evidence	Definition
High	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very Low	Any estimate of effect is very uncertain

33. Joanna Briggs Institute for Nursing (2005), (Joanna Briggs Institute 2011):

LOE	Effectiveness
1	Meta-analysis (with homogeneity) of experimental studies (e.g. RCT with concealed randomisation) OR one or more large experimental studies with narrow confidence intervals
2	One or more smaller RCTs with wider confidence intervals OR quasi-experimental studies (without randomisation)
3	A. Cohort studies. B. Case-control studies. C. Observational studies without control group
4	Expert opinion or physiology, bench research or consensus

34. Oncology Nursing Society Putting Evidence into Practice Weight of Evidence Classification Schema (Mitchell and Friese 2007):

Recommended for Practice	Supportive evidence from at least two well conducted randomised controlled trials that were performed at more than one institutional site and that included a sample size of at least 100 participants
	Evidence from a meta-analysis or systematic review of research studies that incorporated quality ratings in the analysis, and included a total of 100 patients or more in its estimate of effect size and confidence intervals
	Recommendations from a panel of experts, that derive from an explicit literature search strategy, and include thorough analysis, quality rating and synthesis of evidence
Likely to be Effective	Supportive evidence from a single well conducted randomised controlled trial that included fewer than 100 patients or was conducted at more than one institutions
	Evidence from a meta-analysis or systematic review of research studies that incorporated quality ratings in the analysis, and included fewer than 100 patients or had no estimates of effect size and confidence intervals
	Evidence from a synthetic review of randomised controlled trials that incorporated quality ratings in the analysis
	Guidelines developed largely by consensus/opinion rather than primarily based on evidence and published by a panel of experts, that are not supported by synthesis and quality rating of the evidence
Benefits Balanced with Harms	Supportive evidence from one or more randomised trials, meta-analyses or systematic reviews, but where the intervention may be associated, in certain patient populations, with adverse effects that produce or potentially produce mortality, significant morbidity, functional disability, hospitalisation or excess length of stay
Effectiveness not Established	Supportive evidence from a well conducted case control study
	Supportive evidence from a poorly controlled or uncontrolled study. Evidence from randomised clinical trials with one or more major flaws or three or more minor methodological flaws, Evidence from non-experimental studies with high potential for bias (such as case series with comparison to historical controls). Evidence from case series or case reports.
	Conflicting evidence, but where the preponderance of the evidence is in support of the recommendation or meta-analysis showing a trend that did not reach statistical significance
Effectiveness Unlikely	Evidence from a single well conducted randomised controlled trial with at least 100 participants or conducted at more than one site and which showed no benefit for the intervention
	Evidence from a well conducted case control study, a poorly controlled or uncontrolled study, a randomised trial with major methodological flaws, or an observational study (e.g. case series with historical controls) that showed no benefit and a prominent and unacceptable pattern of adverse effects and serious toxicities
Not Recommended for Practice	Supportive evidence from two or more well conducted randomised controlled trials with at least 100 participants or conducted at more than one site which showed no benefit for the intervention and excessive costs or burdens expected
	Evidence from a well conducted trial that showed a prominent and unacceptable pattern of adverse events and serious toxicities
	Evidence from a meta-analysis or systematic review of research studies that incorporated quality ratings in the analysis, included a total of 100 patients or more in its estimate of effect size and confidence intervals with demonstrated lack of benefit or prominent and unacceptable toxicities
	Intervention discouraged from use by a panel of experts in the related subject, after conducting a systematic examination, quality rating and synthesis of the available evidence

35. John Hopkins Evidence Rating Scale (Newhouse et al 2007):

Evidence Level	Definition
I	RCT or systematic review of RCTs (with or without meta-analysis)
II	Quasi-experimental study or systematic review of quasi-experimental studies (with or without meta-analysis)
III	Non-experimental study (or systematic review of non-experimental studies), qualitative study or meta-synthesis
IV	Opinion of nationally recognised experts based upon research evidence or expert consensus panel e.g. clinical guidelines
V	Opinion of individual expert based on non-research evidence (includes case studies, narrative reviews, quality improvement projects, personal experience/expertise)

36. American Association of Critical Care Nurses modification 1(2008), (Armola et al 2009):

Level	Description
A	Meta-analysis of multiple controlled studies or meta-synthesis of qualitative studies with results that consistently support a specific action, intervention or treatment.
B	Well-designed controlled trials, both randomised and non-randomised, with results that consistently support a specific action, intervention or treatment.
C	Qualitative studies, descriptive or correlational studies, integrative reviews, systematic reviews or randomised controlled trials with inconsistent results.
D	Peer-reviewed professional organisational standards with clinical studies to support recommendations.
E	Theory based evidence from expert opinion or multiple case reports.
M	Manufacturer's recommendation only.

37. Australian National Health and Medical Research Council (modification 1), (Merlin et al 2009):

Level	Intervention
I	A Systematic Review of Level II Studies
II	A Randomised Controlled Trial
III-1	A Pseudo-randomised Controlled Trial (i.e. alternate allocation or some other method)
III-2	A Comparative Study with concurrent controls: non randomised experimental trial, cohort study, case-control study, interrupted time series with a control group
III-3	A comparative study without concurrent controls: historical control study, two or more single arm studies, interrupted time series without a parallel control group
IV	Case series with either post-test or pre-test/post-test outcomes

38. World Federation of Societies of Biological Psychiatry (modification 1), (Grunze et al 2009):

Category of Evidence	Definition
A	Full evidence from controlled studies is based on: 2 or more double blind, parallel group randomised controlled trials showing superiority to placebo (or in the case of psychotherapy studies, superiority to a 'psychological placebo' with adequate blinding) and one or more randomised controlled trials showing superiority or equivalent efficacy with established comparator treatment in a three arm study with placebo control or in a well powered non-inferiority trial (only required if such a treatment exists). In case of negative existing studies (studies showing non superiority to placebo or inferiority to comparator treatment) these must be outweighed with at least two more positive studies or a meta-analysis of all available studies showing superiority to placebo and non-inferiority to a comparator treatment. Studies must fulfil established methodological criteria. The decision is based on the primary efficacy measure.
B	Limited positive evidence from controlled studies is based on: 1 or more randomised controlled trial showing superiority to placebo (or in the case of psychotherapy studies superiority to a "psychological placebo". A randomised controlled comparison with a standard treatment without placebo control with a sample size sufficient for a non-inferiority trial. In case of negative existing studies (studies showing non superiority to placebo or inferiority to comparator treatment) these must be outweighed by at least one more positive studies or a meta-analysis of all available studies showing superiority to placebo or at least one more randomised controlled comparison showing non inferiority to an established comparator treatment.
C1	Evidence from uncontrolled studies is based on: One or more positive naturalistic studies (with a minimum of 5 evaluable patients) or a comparison with a reference drug with a sample size insufficient for a non-inferiority trial and no negative controlled studies exist
C2	Evidence from case reports is based on: One or more positive case reports and no negative controlled studies exist
C3	Based on the opinion of experts in the field or clinical expertise
D	Inconsistent results Positive randomised controlled trials are outweighed by an approximate equal number of negative studies
E	Negative evidence The majority of randomised controlled trials or exploratory studies show non superiority to placebo (or in the case of psychotherapy studies superiority to a "psychological placebo") or inferiority to comparator treatment
F	Lack of evidence Adequate studies providing efficacy or non-efficacy are lacking

39. Cardiopulmonary Resuscitation Guidelines modification 1(Sayre et al 2010):

LOE	Intervention studies
1	Randomised controlled trials (RCTs) or meta-analyses of RCTs
2	Studies using concurrent controls without true randomisation (e.g. pseudo-randomised)
3	Studies using retrospective controls
4	Studies without a control group (e.g. case series)
5	Studies not directly related to the specific patient/population (e.g. different patient/population. Animal models, mechanical models etc.) or expert opinion

40. Oxford Centre for Evidence Based Medicine (modification 1), (Howick et al 2011)

Level	Description
1	Systematic review of randomised trials or n-of-1 trials
2	Randomised trial of observational study with dramatic effect
3	Non-randomised controlled cohort/follow-up study
4	Case-series, case-control studies or historically controlled studies
5	Mechanism based reasoning

41. Research Pyramid Experimental Research Face (Tomlin and Borgetto 2011):

Level	Definition
1	Meta-analysis of randomised controlled trials
2	Randomised controlled trial
3	Controlled clinical trial
4	Single-subject study

42. HEAL system (Gugiu 2015):

Research Design	Grade	Description
Randomised controlled trial	A	A well conducted RCT free of bias or confounding factors in which the randomisation process is able to ensure the equivalence of the treatment and control groups beyond statistical chance.
	B	A RCT with a flawed randomisation process but which properly used an ECT[94] method for ensuring the equivalence of the treatment and control groups beyond a reasonable doubt.
	C	A RCT with a flawed randomisation process, which cannot ensure the equivalence of the treatment and control groups beyond a reasonable doubt. Alternatively, a cohort RCT that failed to show the population from which the samples were drawn remained stable over time.
	D	An RCT with one or more fatal flaws (a number of these are listed).
Pseudo-RCT, matched comparison, stratified random sampling, paired comparison, regression discontinuity	B	A well conducted Equivalent Controlled Trial that demonstrated the statistical equivalence of the treatment and control groups on key baseline variables or the pre-post regression model.
	C	An Equivalent Controlled Trial which demonstrated baseline differences on key variables or that did not employ an adequate number of covariates, pairing or strata variables to sufficiently remove reasonable doubt regarding the statistical equivalence of the treatment and control groups on key baseline variables or a regression discontinuity design that failed to demonstrate the equivalence of the pre-post regression model. Alternatively a cohort ECT that did not demonstrate the population from which the samples were drawn remained stable over time.
	D	An ECT with one or more fatal flaws.
Cohort study, case-control	C	A controlled study that did not adequately establish the equivalence of the treatment and control groups beyond a reasonable doubt.
	D	A Non-Equivalent Controlled Trial with one or more fatal flaws.
Before-after, case series	D	Any study that did not employ a controlled comparison between two or more groups, including RCT, ECT and non-ECT studies that did not employ a true comparison group (i.e. compared two treatments against each other but not against a treatment as usual group).

---

[94] ECT = Equivalent controlled trial.

Appendix 4: Eccles and Mason (2001) Hierarchy of Evidence

Level	Definition
I a	Evidence from meta-analysis of randomised controlled trials
I b	Evidence from at least one randomised controlled trial
II a	Evidence from at least one controlled study without randomisation
II b	Evidence from at least one other type of quasi-experimental study
III	Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case-control studies
IV	Evidence from expert committee reports or opinions and/or clinical experience of respected authorities