

The Limitations and Misuses of Evidence-Based Medicine: A Critical Evaluation

Abstract

Over the last 20 years, since the introduction and elaboration of the principles of evidence-based medicine (EBM) within the medical profession, we have seen its broad scale acceptance and rise to universal prominence. Educators, legislators, insurance providers, professional associations, healthcare practitioners of various modalities, lawyers, journalists and other commentators – all appear to be driven by a mania for 'evidence' in medicine. However, EBM is a deeply flawed system and under the pervasive influence of the pharmaceutical industry these flaws have become more entrenched rather than identified and corrected. Moreover, the motive of 'medicine for profit' has given rise to a situation that not only resists correction, but also tends to amplify and expand the scope of bad science and poor evidence in medicine. This paper does not argue against the idea that medical practice should be based on the best available evidence. The examination here pertains to the kinds of evidence that have become acceptable, and those which are now devalued, many of which are just as - if not more - relevant to a holistic clinical practice. The main tenet here is that, in terms of medical epistemology, the pendulum has swung too far in the direction of empiricism - an untenable position with respect to clinical knowledge - and that it is time for it to swing back to the sensible middle ground between quantitative and qualitative research. Additionally, it is crucial that we understand the multifarious ways in which this newly developing system may be perverted to serve other agendas. These considerations are all the more important for the complementary medicine profession, which appears to be in the process of wholeheartedly embracing the current paradigm of EBM, 'warts and all'.

By: Tony Reid

Keywords: Evidence-based medicine, EBM, randomised controlled clinical trial, RCT, epidemiology, acupuncture, Chinese medicine.

*'A little learning is a dangerous thing;
drink deep, or taste not the Pierian spring:
there shallow draughts intoxicate the brain,
and drinking largely sobers us again.'*

– Alexander Pope (*An Essay on Criticism*, 1709)

Introduction

The regulatory changes that followed in the wake of the thalidomide crisis mandated the evaluation of drug efficacy and safety in controlled clinical trials. Beginning with the 1962 Kefauver-Harris Drug Amendments to the Federal Food Drug & Cosmetic Act, these standards for the premarket evaluation of new drugs rapidly became the international standard (USFDA, 2009; Healy, 2012; Gøtsche, 2013). This gave impetus to the further development, refinement and elaboration of randomised controlled clinical trial (RCT) methodology, which aimed to maximise objectivity and minimise bias. During the 1990s, together with the extensive use of modern information technology, this blossomed into the evidence-based medicine (EBM) movement, with Ian Chalmers setting up the Cochrane Centre in 1992 and David Sackett becoming its chief spokesperson (Claridge & Fabian, 2005; Healy, 2012).

Bolstered by early successes in refuting the

effectiveness of some established medical services and procedures, and the introduction of new evidence-based guidelines for the treatment of diseases such as asthma, EBM has become the 'new paradigm' for teaching and practising clinical medicine. However, as an Australian academic has recently pointed out, the arguments in favour of EBM are so strong and essentially irrefutable that 'we run the risk of no longer seeing EBM for what it is. It has achieved cult status. To question it is treasonous, politically incorrect, antediluvian, paternalistic, to be condemned and marginalized' (Little, 2003). Moreover, in Australia the undergraduate medical course structure (and also the postgraduate TCM curriculum) does not encourage critical evaluation of the limitations and abuses of 'the evidence' (Office of Medical Education, University of Sydney, 2014; University of Western Sydney, 2014). In general, for both orthodox and complementary medicine only the basics of EBM are taught within an already crowded tertiary level curriculum, and this tends to lead to a reliance on evidence-based summaries (e.g. evidence-based practice guidelines), where somebody else has done the assessment (Abbot et al., 2014).

Another factor in the unquestioning acceptance of EBM lies in the authoritative tone of academic papers on the subject. This is a universal feature of

academic writing, which has the requirement to be terse (i.e. observe a very strict word limit), and to have all major issues and facts backed up with suitable references, to which readers with time constraints of their own may rarely refer, and which may in some cases be of poor quality. A succinctly worded argument, followed by the names of eminent researchers and academics together with a publication date, creates an apparently rigorous impression that may not bear up under closer scrutiny. As this is an unavoidable feature of academic papers, it is best for readers to maintain a sceptical attitude and as much as possible evaluate both the arguments together with the citations: *caveat emptor, caveat lector*.¹

A corollary of the above is the awe-inspiring nature of the systematic review and meta-analysis: the rigorous inclusion criteria, the complex statistical analyses and the substantial number of individual trials involved. All of these taken together, with the impressive list of experts who command our respect for having undertaken such work, tends to generate in readers a sense of self-effacing deference. Who are we to question this gold standard of clinical truth? We read the abstract, paying particular attention to the main results and the authors' conclusions, and then return to our practice, fortified with a renewed feeling of certainty. However, as we shall see below, all that glitters may in fact not be gold.

Finally, there is the inescapable human aspect that surrounds, informs and filters all of these issues. We crave certainty, we wish to help our patients, we want our cherished theories and concepts to be confirmed. Anything that offers or appears to have the potential to offer these things tends to be perceived with a halo of authenticity that may obscure the facts or context (Healy, 2012; Healy, 2003; Gøtzsche, 2013). We all bring these intrinsic biases to our scientific endeavours, and they creep undetected into even the most rigorous scientific research. Although the sources of bias in clinical research are well documented (Higgins & Green, 2011; Simundic, 2013), human frailties may sometimes prevail. This underlines the need for constant vigilance and speaks to the overarching requirement that all clinical trials and all of the data be freely available and easily accessible for scrutiny by the scientific community.

No less important is the influence of the English language upon the way in which we are accustomed to think and express ideas. The verb 'to be' has a lot to answer for in this regard. Scientific truth is highly nuanced and we are rightly warned to avoid the use of 'always' and 'never' in medical discussions. However, our language encourages us to think in terms of absolutes, not in terms of degrees or possibilities. The appropriate way to express the positive findings (i.e. outcomes) of a study that establishes correlation, rather than causation (i.e. a randomised controlled clinical trial), is to say that the findings 'suggest' that the outcomes 'may be' due to the intervention. Even where there is a very high degree of

correlation, we need to acknowledge that the conclusions based upon these findings may be altered or refuted once we uncover some new piece of evidence. Moreover, until a causal mechanism or pathway has been established we may rightly question the currently available best evidence.

There are many aspects of EBM itself that are problematic, from its philosophical underpinnings to its application in clinical practice. These items have been discussed and debated at length in the literature, particularly over the first decade or so since the introduction of EBM, and several attempts have been made to summarise and evaluate them. Perhaps the best of these appeared in 2004, categorising the criticisms and limitations of EBM as: 'reliance on empiricism, narrow definition of evidence, lack of evidence of efficacy, limited usefulness for individual patients, and threats to the autonomy of the doctor/patient relationship' (Cohen et al., 2004). Carefully laid out and explored, these points are clearly elucidated and easy to understand. However, after having given the foundations of EBM a good shaking, the message appears to have gone largely unheeded (Greehalgh, 2014). Perhaps this is because, while Cohen et al. provide an excellent academic perspective on these matters, the psychological, social, political and economic factors that have been driving this radical change in orthodox medicine remained largely unacknowledged. With the benefit of hindsight and another decade on the clock, it is apparent that the rapid and wholesale adoption of our current - severely distorted - version of EBM has less to do with the efforts of an enthusiastic and articulate group of Canadian epidemiologists than with these other factors. Perhaps chief amongst them is the dominant role of the large multinational pharmaceutical companies in clinical research. This dominance leads to a pervasive conflict of interest in the generation, synthesis and dissemination of the evidence that is meant to guide contemporary medical practice. As we shall see, such industry dominance has found little, if any, resistance.

What is EBM?

Evidence based medicine has been defined broadly as: 'the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research' (Sackett et al., 1996). In addition, the practice of EBM requires that the healthcare practitioner integrates 'clinical expertise and patient's values with the evidence in a way that leads to a rational, acceptable management strategy' (Straus & Sackett, 1998).

In clinical practice, EBM has been defined more specifically in terms of a four-step process that begins and ends with the patient encounter:

1. Formulating a focused clinical question
 2. Searching for evidence
 3. Appraising the evidence
 4. Drawing conclusions that impact on practice
- (Lang, 2004)

This process is what is generally meant by the term 'evidence based practice' (EBP), although the two terms - EBM and EBP- are frequently used synonymously. It involves the cumulative gathering and assessing of information from the practitioner's encounter with the patient (case history taking and eliciting the patient's expectations), through the above four steps, and then back to the patient for discussion and implementation of treatment options. Although implicit, the practitioner's clinical expertise (together with that of his/her peers) as well as the patient's values, are notably downplayed in the four-step paradigm.

Central to this whole process is the concept of a hierarchy of evidence, where, moving from strongest and most reliable to weakest and least reliable, we have:

1. Meta-analysis (MA) of several homogenous RCTs and systematic review (SR) of the same (or a large well-conducted RCT)
 2. Individual RCT
 3. Observational studies (patient-important outcomes)
 4. Basic research (test tube, animal, human physiology)
 5. Clinical experience and expert opinion
- (Lang, 2004)

Not long ago, the clinical judgement of a senior doctor was universally respected, just as the seasoned judgement of professionals working in other fields is generally placed in high regard. Since the advent of EBP, and the paradigm shift that is embodied in the above hierarchy, this has been turned on its head, to the extent that unless a treatment has been evaluated in an RCT (or preferably in an SR or MA of several RCTs) it is viewed with deep mistrust (Tonelli, 1998; Little, 2003; Healy, 2012). Indeed, this may even apply to practices and procedures that have a long history of satisfactory clinical use. This was the subject of a mock SR published recently in the *British Medical Journal* regarding the lack of evidence for the use of a parachute when jumping from a plane (Smith & Pell, 2003).

Under the influence of EBM all other types of medical knowledge are regarded as having less validity than the various types of scientific studies. Moreover, evidence-based clinical guidelines may be used as the basis for legal action in cases where a non-standard treatment has led to an adverse outcome, in spite of having been agreed to by both patient and doctor (Little, 2003; Healy, 2012). It is therefore understandable that the priorities of the medical profession lie in elaborating and refining the procedures as well as collecting and analysing the data obtained from

clinical research. Noticeably lacking, however, is guidance for the appropriate use and prioritisation of other forms of medical knowledge, such as individual clinical experience, expert opinion, patient values and basic physiological principles (Tonelli, 1998).

Under the influence of EBM all other types of medical knowledge are regarded as having less validity than the various types of scientific studies.

The randomised-controlled clinical trial

If the crown of EBM is the systematic review or meta-analysis of data, the jewel is the well-conducted clinical trial, which is generally regarded as the 'gold standard for evaluating the effectiveness of interventions' (Akobeng, 2005). An RCT is 'an experiment in which two or more interventions, possibly including a control intervention or no intervention, are compared by being randomly allocated to participants. In most trials one intervention is assigned to each individual ...' (Cochrane Collaboration, 2014). In order to effectively evaluate a particular healthcare intervention, a trial must be designed to incorporate various means that minimise possible sources of bias, and hence confounding factors that may influence the results. Principal amongst these are: randomisation, allocation concealment, blinding and intention to treat analysis (Akobeng, 2005; Higgins & Green, 2011). For a new drug or device to be approved by regulatory authorities, it must be shown to be significantly more effective than placebo.

Studies showing that a new treatment is as good as or marginally better than an older treatment (which has not been the subject of an RCT, as is the case with most older treatments) are not acceptable, as it is possible that neither drug may be effective (Healy, 2003; Healy, 2012; Gøtzsche, 2013). However, the use of a placebo control is questionable in many instances, particularly where the outcomes are highly subjective in nature (e.g. mood or pain). In these situations the placebo effect may actually be very weak, as the natural course of the disorder may lead to clinical improvement, remission or fluctuations in severity such that a significant number of subjects would be expected to improve over the duration of the trial without any treatment at all. Thus, in these instances the same logic would apply, negating the apparent effectiveness of the treatment in question (Gøtzsche, 2013). For this reason, one should not accept at face value any therapeutic claims arising out of clinical trials that require a large sample size in order to detect a marginal effect.

As noted above, a well conducted RCT is designed

to minimise possible sources of bias. 'The reliability of the results of a randomized trial depends on the extent to which potential sources of bias have been avoided' (Higgins & Green, 2011). This point is well worth some consideration. Bias may be minimised, but it is not possible to eliminate it entirely. Therefore a small amount of bias is an inevitable accompaniment to even the best RCTs (Simundic, 2013; Gøtzsche, 2013; Sackett, 2000). Almost by definition, we are not able to measure the extent of such residual bias; we only know that it must exist, and have faith that it is minimal. There are two important consequences of such residual bias. One is that in trials where the treatment produces only a small effect, it would take only a very modest fluctuation in the degree of bias (e.g. in the completion of a rating scale for pain or depressed mood) to completely change the trial outcome, especially if we are dealing with statistically significant changes, which are generally of a much smaller order than clinically significant changes (Gøtzsche, 2013). The other point is that when combining several RCTs together in a systematic review or meta-analysis, while there is a chance that the various small biases will negate each other, there is also the chance that they may add together, thus skewing the results.

The great strength of RCTs is that they are able to provide potentially meaningful data to support specific actions (i.e. the administration of a potentially effective treatment, the avoidance of a potentially useless or dangerous treatment or the avoidance of certain risk factors) in the absence of identified causes. There may be no theoretical underpinnings to support the outcome - indeed the outcome may go against the currently accepted model - and yet a well conducted RCT is able to tell us that a particular intervention is strongly associated with a particular outcome in a specified patient group, or that a particular factor is strongly associated with a certain harm in a certain population. We are thus able to take curative or preventative action, without having to wait for the scientific breakthrough that explains why this should be so. In an early example of the use of statistical studies in medicine, preventative action was taken during the 1854 cholera epidemic in Soho based on the findings of Dr. John Snow, who linked the disease with contamination of the water supply by showing a significant increase in cases amongst people using water from the Broad Street pump. Based on this compelling evidence, the handle of the pump was eventually removed in the absence of precise knowledge of the causative factors (Hempel, 2009 in Healey, 2012). It was not until several decades later that these factors were elucidated through the work of Dr. Robert Koch (Altman, 1987 [in Healy, 2012]).

On the other hand, being able to access statistically significant data in the absence of a precise knowledge of causes has a significant downside. This is best illustrated by an extreme example. Suppose we are orthopaedic

surgeons investigating treatments for fractures of the right forearm (unaware of the precise cause of the patients' distress and disability). As active treatment we choose to apply a plaster cast to either of the upper limbs, and as placebo we immobilise the neck. Inevitably our RCT will show that patients with a cast on either limb do significantly better than those whose necks were treated. In this way we have demonstrated that applying a cast to either limb 'works' in cases with a fracture on the right side.² Real-world examples of comparable practices are not uncommon, as we shall see in the discussion of industry involvement below.

It bears repeating that statistical significance means, on the one hand, that there is a considerably higher probability that the results under consideration did not occur due to the play of chance. However, it also means that the results could, in fact, have occurred by chance (this is the significance of the 'p value'³), and indeed will actually occur at some time, even though the odds are very low. This is discussed in more detail below.

The take-home message is that statistical studies are best utilised in the assessment of risk, as well as to show that a particular treatment is of no significant benefit, or that the risk to benefit ratio is unfavourable. In such cases we can promptly avoid the risk or abandon the useless treatment, while we continue our investigation into specific causes and mechanisms. If we believe that a favourable outcome in an RCT actually proves that a treatment works, we are in the same position as the optimistic punter who believes that very favourable odds on a horse at the race track 'proves' that the horse has already won.⁴

Statistical tests, such as RCTs, are capable only of demonstrating relationships, not providing proof. A well-designed RCT with a strong result (i.e. a very low p value), may point to probable causes. But they remain just that: probable. The test has only confirmed a hypothesis. There may be other competing hypotheses, which should also be examined and tested if sufficiently plausible, following which causal mechanisms should be identified that explain the favoured hypothesis. In this way the meaningful correlations revealed by an RCT may generate further investigations aimed at elucidating causes. This is how the science of medicine progresses, otherwise we are opening ourselves up to error.

Another critical weakness of RCTs is that the rigidity of the clinical trial structure must inevitably fail to account for the complex nature of clinical problems. By their very design, RCTs neglect important contextual factors that have a marked impact on disease prognosis, such as the presence of a loving family environment, a support network of friends and other carers (e.g. religious affiliations), job satisfaction, ability to engage in recreational activities and physical fitness (Cohen et al., 2004; Dossey, 1995). These are largely ignored when allocating patients to active treatment or placebo.

However, there is a growing body of research that tends to confirm their importance, particularly in cardiovascular disease (Khayyam-Nekouei et al., 2013; Everson-Rose & Lewis, 2005; Cohen & Herbert, 1996).

Additionally, it is very rare for two different studies on the same topic to be exactly the same, with differences in study design, definition of the disease, populations admitted, ancillary therapies, unanticipated consequences and definition of outcome. Therefore, conflicting or contradictory results are often to be found where there are several different trials dealing with the same topic. While two studies that reach the same conclusion in regards to therapy may strengthen this conclusion, studies that disagree (not due to bias or too small a sample size) may provide important clinical insights that would otherwise be lost when pooling the data (Horowitz, 1987). For these reasons, the inherent reductionism of RCTs tends to 'dumb down' clinical practice, which of necessity must deal with real-life patients – their individual health problems and individual responses to treatment. These are obscured in favour of statistical data related to the mean, standard deviation and, of course, statistical significance. In this way an epidemiological perspective is favoured over the more relevant clinical perspective (Horowitz, 1995). As we shall see below, there is a parallel dumbing down effect that is an intrinsic concomitant of SRs and MAs.

Systematic review and meta-analysis

'A systematic review is a critical assessment and evaluation of all research studies that address a particular clinical issue. The researchers use an organized method of locating, assembling, and evaluating a body of literature on a particular topic using a set of specific criteria. A systematic review typically includes a description of the findings of the collection of research studies. The systematic review may also include a quantitative pooling of data, called a meta-analysis.' (US Department of Health and Human Services, 2014). In contrast to a journalistic review, the primary studies that are included are selected according to explicit and reproducible methods. Thus, any bias associated with selection and rejection of studies is limited.

In the best-case scenario there is a minimal heterogeneity amongst studies included in an SR, and all are of a uniformly high quality, regardless of their results. However, the emphasis on uniform high quality of studies fails to adequately address the issue of heterogeneity. In spite of adherence to a standard methodology, the complexities of clinical practice will inevitably generate significant differences amongst studies on a particular topic. In practice, studies dealing with the same topic are rarely homogenous, as noted above. In terms of methodology, some may score more or less highly according to the selection criteria, and the cut-off point below which a given study is to be rejected may be the

result of an arbitrary decision by one or two reviewers. These factors inevitably lead to a degree of heterogeneity amongst included studies that may be sufficient to skew the resultant conclusions.

The inherent reductionism of RCTs tends to 'dumb down' clinical practice ...

Additionally, the selection criteria may be flawed in such a way that inadequate or irrelevant studies are included. A recent example of this potential may be seen in an SR in the Cochrane Library to assess the effects of acupuncture for treating peripheral joint osteoarthritis. As there is currently no consensus within the Chinese medicine community regarding what constitutes an acupuncture treatment nor what constitutes a course of acupuncture⁵ (White et al., 2008), the 'adequacy of acupuncture' selection criteria are of doubtful validity. Moreover, studies in which one or, in some instances, two out of the four criteria were not met were included in the review: 'Only two of the trials ... were judged adequate in terms of the acupuncturist's experience ... For five of the trials ... the number of acupuncture sessions was judged inadequate.' It was noted under the heading 'Adequacy of Acupuncture assessments': 'Two acupuncturists ... who have a combined acupuncture clinical experience of nearly forty years in treating knee OA, and who have both previously worked on RCTs and systematic reviews of acupuncture, independently assessed the adequacy of the acupuncture administered in the trials. Consensus was achieved by discussion ... The assessors had previously used this adequacy assessment instrument for the earlier systematic review ... of which this is an update.' (Manheimer et al., 2010). This amounts to deference to expert opinion, which, although it may seem out of place in the 'pinnacle of the evidence hierarchy', is nevertheless a ubiquitous, though rarely acknowledged, element at all levels of evidence.

In Manheimer et al. (2010) there were significant differences amongst the studies: most involved OA of the knee joint, while the remainder dealt with OA of the hip. Amongst these there was considerable variation in the severity and duration of the disorder, the acupuncture points selected and the frequency of treatments. However, the major portion of the SR was devoted to description and discussion of methodology, with only a single small paragraph outlining some of the clinical differences between the studies and a table showing the adequacy of the acupuncture assessment instrument (which was applied after the studies had been selected for inclusion). This illustrates Horwitz's contention that 'scientific rigidity

creates only the illusion of homogeneity for clinical trials' (Horowitz, 1987). By convention homogeneity is judged according to the overlapping of the confidence intervals⁶ of the various trials selected for the meta-analysis.⁷ In other words, homogeneity is defined statistically, not clinically,⁸ thus favouring a mathematical or statistical perspective over a clinical one. The richness and complexity of acupuncture in clinical practice is therefore obscured in favour of a statistical analysis to provide a standardised result.

SRs routinely ignore the intrinsic differences amongst the trials from which the pooled data are derived, tending to regard the included trials in much the same light as a single multicentre trial (Horowitz, 1995). It should therefore come as no surprise to find that in many instances the conclusions reached by a single large RCT are at odds with prior SRs and MAs (of smaller trials) on the same topic. A review of 12 large RCTs, to which 19 MAs corresponded, with a total of 40 relevant outcomes, noted that 'if there had been no subsequent randomized, controlled trial, the meta-analysis would have led to the adoption of an ineffective treatment in 32 percent of cases ... and to the rejection of a useful treatment in 33 percent of cases'. As the authors concluded, 'Pooled results incorporate the biases of individual studies and embody new sources of bias, mostly because of the selection of studies and the inevitable heterogeneity among them' (Leloir et al., 1997).

When confronted with an SR, it is in a clinician's best interests to appraise each trial separately. 'Rather than pooling the data and blurring the distinctions that are such a dominant feature of clinical trials, the results of our analysis would encourage readers to assess each trial individually. Such a strategy would take advantage of the diversity in patients, therapies and trial designs by allowing pragmatic clinicians to distinguish the effects of treatment among distinctive studies.' (Horowitz, 1987).

Philosophical conundrums: randomness, significance, certainty, objectivity

As will be discussed below, objectivity may be achieved by degrees, but there will always be at least some element of subjectivity present in every observation, calculation and conclusion. Rather than shy away from or deny the ubiquity of subjectivity, it is proposed that it should be acknowledged and accorded a central place in the generation of medical knowledge. Therefore this section presents some personal opinions in relation to some of the philosophical underpinnings of EBM.

Randomness and 'the play of chance' are nebulous concepts, as lampooned by Tom Stoppard in the opening scene of his play, *Rosencrantz and Guildenstern are Dead* (Stoppard, 1967, pp.11-18). Even if the subjects in a trial are assigned to its various arms 'randomly', there is always the likelihood that randomness may work against the aims

of allocation - just as at some point in time it is possible for one hundred tosses of the coin to all come up heads. Of course, the various patient groups are analysed afterwards in order to detect any discrepancies. However, the need for this type of allocation review points to a conundrum in the structure of a 'controlled' trial and brings into question the desirability of including randomisation. There seems to be an inherent contradiction in having randomisation that needs to be monitored. This issue has been highlighted within epidemiology and discussed elsewhere (Gordis, 2005, in Maier & Shibles, 2011).

Furthermore, the occurrence of seemingly impossible events or sequences of events (i.e. those with only a very small possibility) highlights the 'hidden' side of randomness or chance. To our usual ways of thinking these concepts imply the 'law of averages' and the 'fifty-fifty split', and indeed most of the time this understanding serves us well. However, there are times when we are thrust from under the shelter of the central portion of the bell shaped curve,⁹ to be confronted with unexpected, atypical - if not downright weird - phenomena. These sorts of occurrences are also part of randomness or the play of chance and lead to the idea that whatever can happen, will happen.¹⁰ Even with the most sophisticated statistical computations, no one is able to predict exactly when these outlying phenomena will occur. Our own existence on this planet is testament to this. According to the theory of evolution, human life (indeed, all life) is based on the repeated occurrence of events with only the minutest possibility (Darwin, 1909, Morris, 2003).

The fact of our existence has profound implications on the current application of statistical models to populations. It appears reasonable to expect that the odds for the occurrence of a particular phenomenon would be different depending on the time scale involved (millions of years versus thousands of years versus several human generations versus a single lifetime versus several years versus a 12-week trial) and the numbers of subjects involved (all life on the planet versus a single species versus all members of that species with a particular disease versus a limited number of subjects with a particular disease). It would therefore be reasonable to expect that different statistical models should apply at various orders of magnitude. This is graphically illustrated by the fact that a drug with only weak therapeutic effects requires a large number of subjects in a clinical trial in order to demonstrate statistical significance; and that, conversely, adverse reactions to a drug may be undetected (i.e. not reach statistical significance) in a trial with relatively small numbers. Also in this connection, most healthcare practitioners are aware of the disproportionate numbers of rare and unusual cases seen in the early stages of clinical practice when patient numbers are small; once the practice has become established and larger numbers of patients are seen annually, the common disorders are seen commonly

and the rarer disorders seen more rarely (allowing, of course, for the location and nature of the practice). While it is beyond the scope of this paper to explore the validity of statistical methodology with respect to medical research, it appears that the above considerations, even if they have been elaborated within the science of statistics, have not percolated down into the field of clinical medicine.

The concept of statistical significance is another contentious area. The results of a clinical trial are expressed in terms of their statistical significance, i.e. the likelihood (preferably very small) that they could have occurred by chance. There are two issues here. One is that by their very nature RCTs are designed to prove or disprove the null hypothesis, i.e. whether or not the treatment under consideration does not work. In other words, an RCT is designed to answer the question: 'Is this treatment doing nothing or not doing nothing?' A positive result tells us that the treatment in question is 'not doing nothing', and that this 'not doing nothing' is closely or loosely associated with various clinical effects, some or most of the time. In this way, an RCT may demonstrate correlation, but it does not reveal causation. Such a procedure appears to fall a little short of what one would expect of a 'gold standard' for clinical efficacy. The other aspect of statistical significance is that the application of ' $p < 0.05$ ' is too rigid. As elaborated by Fisher (1926), the man who promulgated this approach, the standard for statistical significance should be flexible. Some instances should require only very small odds in order to reach significance, e.g. for very severe or life-threatening side effects.

The notion of objectivity and the removal of bias requires clarification. In so far as all perception arises in the mind as a result of neuronal activity, perception is inherently subjective. Human beings are incapable of absolute objectivity. Certain consensus observations may approach objectivity more or less closely. However, pure objectivity can never be reached. The impossibility of objective observation has been clearly and extensively discussed elsewhere (Cohen et al., 2004; Harari, 2001). The world-view of the observer determines what questions are asked, what information is deemed important and what information is deemed background noise. According to quantum theory in contemporary physics, the observer affects the observed reality in such a way as to change what is observed to align with the cognitive bias of the observer. In other scientific disciplines the effect of observer influence is acknowledged, but not in medicine (Greenwood, 2002; Dossey, 1995). Thus, research may be more objective or less objective, but never completely objective. In empirical research objectivity is a goal, not a fact (Cohen et al., 2004). Especially dangerous is the presumption of objectivity and the belief that RCTs actually prove something. Correlation does not prove causation (Aldrich, 1995). Indeed there are many steps to traverse from an 'observed *association*' to a verdict of

causation' (Hill, 1965). In addition, there are the inevitable limitations of inductive evidence. We are generally unable to be certain that we have all of the relevant facts in order to draw a valid conclusion. It is always possible that some new facts may come to light that serve to invalidate our previous conclusions. This appears to be the case with the results of EBM thus far, e.g. recommendations regarding hormone replacement therapy in menopausal women and, more recently, dietary fat intake (Little, 2003; Harcombe et al., 2015).

Research may be more objective or less objective, but never completely objective.

The above arguments have been put forward in order to stimulate discussion and debate over these issues. Most of them are not new. A.B. Hill, regarded as the 'world's leading medical statistician' (Armitage, 1991) – the man who introduced RCTs into clinical medicine – warned against the overemphasis of statistical significance and the inadequacies resulting from undetected systematic errors such as measurement error, confounding and selection bias (Hill, 1965; Phillips & Goodman, 2004). The various anomalies discussed in this paper, as well as the numerous cases in which incorrect or inappropriate conclusions have been drawn from the data under consideration should give ample reason to question whether the fault lies with the correct or incorrect application of statistical and epidemiological methodology. A fine example of the conundrums that may be generated in this connection may be seen in the following paper, the title of which says it all: 'Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics' (Heres et al., 2006). So much for 'scientifically proven' clinical efficacy. Although the authors of this study were investigating sources of bias in company-sponsored trials, I believe that results such as these also call into question the validity of the methodology, as there will always be a certain amount of unavoidable bias, which is either not detected or not detectable, and this may significantly influence the outcomes of trials.

Other important sources of clinical knowledge

Empirical research, such as the RCT that aims for maximum objectivity, is best applied to relatively simple and clear-cut clinical scenarios, such as the clinical effects of a drug intervention. Such an approach is less useful for the more complex and multifactorial clinical scenarios (Greenwood,

2002; Horowitz, 1987; Horowitz, 1995). Indeed, illness is a predominantly subjective experience, doctor-patient interactions are subjective on both sides, and the effects of treatment have a strong subjective component. Thus, because of their nature and principally through the application of double blinding and their intrinsic clinical reductionism, objective studies cannot be applied to interactive therapies (e.g. acupuncture, psychotherapy, etc.) without fundamentally changing them. This leads to the conundrum that in order to obtain a valid result, the therapy itself needs to be invalidated. In as much as all therapeutic modalities are interactive to some extent, these considerations apply across the board. When seen in this light, arguments concerning the reliability and accuracy of RCTs, intriguing as they are, become somewhat of a red herring, and we return to the definition of EBM quoted above (Sackett et al., 1996) and the deficiencies pointed out by Tonelli (1998) concerning the lack of guidance for the appropriate use and prioritisation of other forms of medical knowledge.

Objective studies cannot be applied to interactive therapies (e.g. acupuncture, psychotherapy, etc.) without fundamentally changing them.

The rise in popularity of objective knowledge and its elevation to the apogee in orthodox medicine has gone hand in hand with the increasing reliance by Western medical practitioners on clinical measurements, e.g. blood pressure, serum lipids, glucose tolerance and liver function, as well as pseudo-measurements such as patient questionnaires. This has often been to the detriment of the clinical encounter, eroding the observational and interpersonal communication skills that had been the cornerstone of good medical practice up until the latter part of the 20th century (Spence, 2013; Healy, 2012; Wen & Kosowsky, 2012). However, there are many important aspects of a patient's presentation that, while detectable, are not in fact measurable – measurable in a way that provides meaningful comparisons to other similar patients. Patient experiences of pain and depressed mood are prime examples of this. Moreover, in many complementary medical systems (e.g. traditional Chinese medicine) the importance of recognisable - but non-measurable – aspects of a disease in an individual are a central feature (Greenwood, 2002; Tonelli & Callahan, 2001). If we try to investigate or understand these factors by utilising an unsuitable enquiry system we will inevitably generate incorrect or irrelevant data, including finding solutions to the wrong problems, solving irrelevant or unimportant problems, rejecting a correct hypothesis and accepting an incorrect hypothesis (Oschman & Oschman, 1997;

Greenwood, 2002). In other words, as has been observed, the evidence-based tail ends up wagging the clinical dog (Greenhalgh et al., 2014).

Some potentially fruitful sources of medical knowledge in relation to the non-measurable factors within a clinical encounter have been explored by various authors, many of which centre on the individual patient (Healy, 2012; Greenwood, 2002; Tonelli & Callahan, 2001; Dossey, 1995). These include:

- Expert opinion or consensus of experts, based upon clinical experience
- Application of anatomical, physiological, biochemical or pathophysiological principles
- Reasoning based on theory (e.g. according to the principles of one of the complementary medicine modalities)
- N-of-1 trial (Lillie et al., 2011; Kravitz & Duan, 2014)¹¹
- Single case causality assessment
- Case studies

Note that the clinical judgement of the individual healthcare practitioner is central to most of the above items. As mentioned above, the mature judgement of a professional is highly valued in every other sphere except, until only very recently, in medicine. A crucial component of professional judgement in medicine (and also in general) involves the development of pattern recognition based on the progressive accumulation of a repertoire of illness scripts (Matsui & Kawaguchi, 2014; Pelaccia et al., 2011; Coderre et al., 2003). Experienced doctors diagnose more accurately than final year medical students for this reason – their ability to recognise essentially similar clinical scenarios from prior experience. In large part this unique ability depends upon mental processes occurring below the threshold of conscious awareness; it is mostly intuitive (Coderre et al., 2003).

Historically, all the major advances in medicine up until the mid-20th century have followed from case study reports (CSRs). That is to say, major advances have occurred in the absence of RCTs: new syndromes were differentiated and described, and patient responses to treatment, both beneficial and deleterious, were brought to light. Although undervalued today, the usefulness of CSRs is still recognised, in spite of their obvious and well-known limitations (Nissen & Wynn, 2014). However, discussions on the merits of CSRs would be more realistic and clinically relevant if they were informed by a more accurate assessment of the context. There are two critical ideas that pertain here. One is that RCTs (to which CSRs are inevitably compared) are by no means free of limitations. The other is the crucial role that CSRs have played in the past. Moreover, we should be wary of the tendency, perhaps the hallmark of immature thinking,¹² to undervalue history and overvalue current knowledge.

Although it is beyond the scope of this paper to examine them in detail, some of the forces that have driven this shift in emphasis regarding the acceptability and credibility of various forms of medical knowledge, include the following:

- The analytical nature of Western science.
- The human fear of uncertainty together with the craving for certainty.
- Peer group pressure and the influence of opinion leaders.
- The static world view underlying a belief in objectivity.
- The professional as well as human desire to help alleviate sickness in the best possible way.
- The fact that EBM is still only a nascent theoretical system in need of refinement and development under real-world conditions.

Having noted the above, perhaps the major driving force behind the development and clinical application of EBM at present is the pharmaceutical industry, utilising what has been referred to as 'the most sophisticated marketing system on the planet' (Healy, 2012; Goldacre, 2012; Kassirer, 2005) to exploit all of the above factors, largely for its own benefit and largely to the detriment of medical science and patients worldwide. With hindsight, all of the developments described in the following section are predictable - though inconvenient - consequences of placing the levers of control into the hands of vested interests.

Involvement of Industry

A system with this many holes is ripe for exploitation, and indeed, this is what has happened. Moving from a position of self-effacing deference to the medical profession to one of almost complete dominance in the space of less than 40 years, the multinational pharmaceutical giants have magnified the inherent distortions within EBM, and added quite a few of their own (Greenhalgh et al., 2014; Gøtzsche, 2013; Healy, 2012; Goldacre, 2012; Kassirer, 2005; Healy, 2008; Healy, 2004). Since the 1990s governments in the developed nations have pursued policies which involved cutting funding for medical research and encouraging industry to partner with universities and the medical community in order to conduct research. After all, they are involved in clinical research anyway, and are unable to bring a new drug to market without the support of RCTs. However, this is a bit like asking a group of heroin addicts to be in charge of the production and distribution of medical narcotics, using the argument that their experience with procuring and using the drug as well as their continued need for it uniquely qualifies them for this task. The reason for the over-the-top analogy here is to drive home the fact that these are all predictable outcomes.

As a result of these changes in policy, the big pharmaceutical companies now dominate clinical

research. Not only that, but their influence has extended to all aspects of Western medical clinical practice. This has been extensively documented elsewhere (Gøtzsche, 2013; Healy, 2012; Goldacre, 2012; Kassirer, 2005; Healy, 2008; Healy, 2004). Multinational corporations of all kinds have one major, over-riding purpose: to increase profits for their shareholders – not the advancement of science nor the increased welfare of humankind. We cannot condemn the big pharmaceutical companies for pursuing their businesses in the most efficient and profitable manner. We and our governments have given them an open door to subvert Western medical science and clinical practice for their own purposes – and they have been highly successful in doing this. These kinds of behaviour on the part of big business, who place the generation of profits above all else, are not new; they have been a notable characteristic of our current economic system since its inception (Patel, 2009; Munz, 1986; Klein, 2008). However, this does not make such activities any less heinous, nor should we allow ourselves to become inured to them.

The errors that may be committed with regard to RCTs may be divided into two categories: mistakes (both unintentional and deliberate) that could be made by anybody, and outright fraud (generally requiring a considerable amount of money and influence). We could be protected against both types of error if pharmaceutical companies made their data freely available for scrutiny by the scientific community. Unfortunately, this is not the case, as companies sequester their trial data and only tend to publish that which is favourable for their products (Gøtzsche, 2013; Healy, 2012; Goldacre, 2012; Healy, 2008). The former category includes the common mistake that may be made when writing up the results of statistical tests. The following list represents an anatomy of bias, providing a summary of the common ways in which errors, unintentioned, well intentioned or otherwise, may find their way into the results of clinical research.

- Misuse of findings with statistical significance, e.g. giving undue importance to trivial effects.
- Failure to adjust for baseline differences between groups, particularly when they favour the intervention group.
- Failure to test the data (i.e. the population from which the data are drawn) to see if they are normally distributed and to use non-parametric tests of correlation if they are not.
- Ignoring all withdrawals (drop-outs) and non-responders, so the analysis only concerns subjects who fully complied with treatment.
- Assuming that one set of data can be plotted against another and an 'r value' (Pearson correlation coefficient) can be calculated, and assuming that a 'significant' r value proves causation.
- Inappropriate handling of outliers (points which lie a

long way from the others on the graph).

- Omission of confidence intervals in the discussion of results, particularly if the confidence intervals overlap zero difference between the groups.
- Stopping a trial early and writing (and publishing) the report if the difference between two groups becomes significant before the end of the trial.
- Extending a trial until the results reach statistical significance if at the end of a trial the results are approaching, but not quite reaching statistical significance.
- Analysing according to subgroups (not in the initial trial design) after a trial has been completed if the trial results are not favourable, in order to find a subgroup with a statistically significant response.
- Using a different analysis of the data (not in the original trial design) to find some sort of favourable outcome if the original trial does not give the desired result.

(Greenhalgh, 1997)

The following group of deliberate misrepresentations have been brought to light when legal action has been taken against a particular company following disastrous outcomes due to a product, revealing that cheating on statistical tests is commonly behind the fraudulent results that have led to the catastrophe under investigation. These include:

- Conducting trials on populations of patients that do not represent the real-world patients in whom that drug will be used.
- Head to head trials where a new drug is compared to an older drug, which is administered incorrectly (i.e. too low a dose or wrong timing).
- Use of composite outcomes to overstate the benefits of a drug and understate the adverse effects.
- Selectively publishing trials with positive findings for a drug and withholding those with negative results.
- Miscoding of adverse events, e.g. suicidal events on active treatment coded as 'miscellaneous effects', 'emotional volatility' or 'overdose'.
- Re-allocation of subjects with severe adverse effects on the drug into the placebo group.
- Removing subjects who have responded favourably from the placebo group.
- The practice of 'ghost-writing', where a company commissions an agency to write up a study with conclusions in the abstract (that are favourable to the drug being tested) which contradict the findings of the study – and then have eminent academics put their names to it.

(Gøtzsche, 2013; Goldacre, 2012; Healy, 2012; Healy, 2008; Healy, 2004)

In the words of Gøtzsche: 'There are very little high-quality published data. Neither the drug industry nor

publicly employed researchers are particularly willing to share their data with others, which essentially means that science ceases to exist. Scrutiny of data by others is a fundamental aspect of science that moves science forward, but that is not how it works in healthcare. Most doctors are willing to add their names to articles produced by drug companies, although they are being denied access to the data they and their patients have produced and without which the articles cannot be written. This is corruption of academic integrity and betrayal of the trust patients have in the research enterprise. No self-respecting scientists should publish findings based on data to which they do not have free and full access.' (ANH-Intl, 2013).

The consequences of this withholding of data by the pharmaceutical industry, some of which may not be strictly relevant to the topic of this paper, graphically illustrate the range and gravity of problematic outcomes generated by the flaws in EBM. Bad science and bad data are one thing, but the effects of industry control are pervasive. Some of these effects that relate directly to EBM have been succinctly summarised by Greenhalgh et al. (2014): '... the drug and medical devices industries increasingly set the research agenda. They define what counts as disease (for example, female sexual arousal disorder, treatable with sildenafil and male baldness, treatable with finasteride) and predisease "risk states" (such as low bone density, treatable with alendronate). They also decide which tests and treatments will be compared in empirical studies and choose (often surrogate) outcome measures for establishing "efficacy". Furthermore, by overpowering trials to ensure that small differences will be statistically significant, setting inclusion criteria to select those most likely to respond to treatment, manipulating the dose of both intervention and control drugs, using surrogate endpoints, and selectively publishing positive studies, industry may manage to publish its outputs as "unbiased" studies in leading peer reviewed journals ... Evidence based medicine's quality checklists and risk of bias tools may be unable to detect the increasingly subtle biases in industry sponsored studies.'

The main consequences of industry domination are summarised below. While only a few representative examples are provided to illustrate what are widespread phenomena within the industry, I strongly urge the reader to review some of the references in order to gain an understanding of the full extent and gravity of the situation.

- Unnecessary harms: Obviously, if you exaggerate the benefits and hide the hazards, people are going to get hurt. Although many authors point to the harms caused due to these practices (Goldacre, 2012; Healy, 2012; Healy, 2003), Gøtzsche (2013) places pharmaceutical drugs as the third leading cause of death after heart disease and cancer.

- **Control of medical practice:** Evidence-based guidelines, which are derived to a large extent on the published results of industry-funded drug trials, often recommending on-patent drugs which are no better than older off-patent (and hence much cheaper) drugs, or recommending ineffective or harmful treatments (Healey, 2012; Gøtzsche, 2013). Invention and promotion of new 'diseases' or exaggeration of the importance of specific risk factors (to both doctors and patients) so that new drugs may be marketed to doctors (Healy, 2012; Gøtzsche, 2013; Greenhalgh et al., 2014). Removal or undermining of avenues for dissent by promoting a culture of 'high level scientific evidence' which allows no room for individual practitioners' observations and professional opinions in peer-reviewed literature (Tonelli, 1998; Little, 2003; Healy, 2012).

Some representative examples of such industry involvement are described below:

1. Intimidation of the profession by statistical analysis: Publishing false conclusions of MA findings.

Although this example comes from the years just prior to the enunciation and elaboration of the principles of EBM, regulatory authorities had for some time required evidence from RCTs in order to allow a drug onto the market, and the scientific community recognised the value of high-level evidence in determining the risk-benefit ratio of new treatments. In 1991 the *British Medical Journal* published a paper (Beasley et al., 1991) by a group of company employees – tertiary qualifications notwithstanding – pooling 'data from 17 double blind clinical trials in patients with major depressive disorder comparing fluoxetine (n=1765) with a tricyclic antidepressant (n=731) or placebo (n=569), or both' in response to growing concerns over a possible increased suicide risk for patients on fluoxetine. The abstract, loaded as it is with complex statistical jargon relating to the incidence of suicidal acts as well as suicidal ideation (assessed according to the Hamilton rating scale), concludes that: 'Data from these trials do not show that fluoxetine is associated with an increased risk of suicidal acts or emergence of substantial suicidal thoughts among depressed patients.' However, if one reads the entire paper, neglecting the dubious findings derived from a rating scale,¹³ one finds a total of six suicidal acts in the Prozac group (n=1765) and one in the placebo group (n=569). These figures show there were just under two times as many suicidal acts on fluoxetine compared to placebo. Analysing the data presented, the relative risk for suicidal acts is 1.9 times greater on fluoxetine than placebo, with a 95 per cent confidence interval from 0.2 to 16, i.e. while there may be no significant increase in risk, there may just as likely be a 16 fold increase in risk, and the most

likely increase in risk would be 1.9 fold. At the time of publication there were either no letters of protest against the study's false conclusions, or the journal decided not to publish them. Although the general level of sophistication regarding the pitfalls of EBM in the medical and other healthcare professions has increased considerably since that time, unfortunately the sophistication of vested interests has consistently outpaced them, with the result that only little has changed with respect to what medical journals will publish and what they will not (Gøtzsche, 2013; Goldacre, 2012; Healy, 2012).

The undue influence of pharmaceutical companies through exploiting the weaknesses within EBM has transformed Western medicine over the past fifty years ...

2. Hiding harms and overstating the benefits of a drug: making extensive use of ghost-writing

The drug olanzapine (Zyprexa) has been surrounded by considerable controversy, and came to market via some highly suspect trial data. Although not associated with the tardive dyskinesia of older 'antipsychotic' drugs, olanzapine has not demonstrated advantages over many of the older first generation antipsychotics, such as haloperidol. Olanzapine was originally synthesised in 1982 and was one of a series of compounds that the manufacturer developed in its search for a safer 'second-generation' (or 'atypical') antipsychotic. They were patented but never marketed, due to problems with toxicity. However, the company decided to select what was apparently the least toxic compound and take it to market. Renewal of patent (US 5,229,382) was granted in 1993 on the basis of the unique benefit that it caused less increase in serum cholesterol in beagle dogs compared with one of its sister compounds (ethyl-olanzapine, which had never been marketed). Based on this fact alone, the compound should never have received the patent. However, with the renewed patent the company ran four clinical trials on olanzapine, which were not published in full. Nevertheless, these four trials gave rise to 234 publications (Healy, 2012, p.142), with one trial having been published 142 times in papers and conference abstracts (Gøtzsche, 2013, p. 231). In many instances these papers were ghost-written, fraudulently promoting 'off label' uses for Olanzapine, which led to unnecessary deaths and disability in children and the elderly (Gøtzsche, 2013, pp.31-32, 230-32, 260; Goldacre, 2012, p.293; Healy, 2012, pp.31-2). The SR conducted by the Cochrane Collaboration proved inconclusive: 'The large proportion of participants leaving studies early in these trials makes it difficult to draw firm conclusions on olanzapine's

clinical effects. For people with schizophrenia it may offer antipsychotic efficacy with fewer extrapyramidal adverse effects than typical drugs, but more weight gain. There is a need for further large, long-term randomised trials with more comprehensive data.' (Duggan et al., 2005). But this was based only on data that had been made available by the manufacturer. Some of the sequestered data, accessed through subsequent legal proceedings or leaked by company employees, have revealed marked increases in suicides, diabetes, cholesterol elevation and weight gain, making it possibly the most harmful of all the available antipsychotic drugs (Götzsche, 2013, pp. 231-2; Healy, 2012, p.142-3).

The warning 'caveat emptor; caveat lector' still stands until one simple, essential solution has been implemented: free public access to all clinical trials and to all of the raw data from trials.

3. Abuse of findings of statistical significance; lobbying by industry and industry backed patient groups to influence guidelines and government policies

One of the prime examples of this category is the plethora of virtually incomprehensible studies on drugs for the treatment of Alzheimer's disease. Supported by complex and sophisticated statistical analyses (mostly of results from rating scales), these studies all somehow confirm the 'efficacy' of various cholinesterase inhibitors and Memantine, either singly or in various combinations. However, on closer examination, one finds that most patients do not respond to treatment and, in those that do, the effects are equivocal, modest and temporary, with small and short-lived improvements in the rate of decline of cognitive function and activities of daily living. Unfortunately these 'modest improvements' only refer to statistically significant improvements. It is doubtful that they are, in fact, clinically significant. What these trials are really saying is this: while it may be possible to measure a small amount of improvement, which represents a retardation in the expected rate of decline in the patient's cognitive abilities, functional ability, behaviour and psychological state, this occurs in only one out of every ten to twelve patients that receive treatment, and will generally not make an appreciable difference to the course of the illness and the burdens on the caregivers (Lancôt et al., 2003; Tampi & van Dyck, 2007; Lancôt et al., 2009). In spite of these findings, which were also supported in a 2010 observational study involving 938 patients (Santoro et al., 2010), intensive lobbying by the manufacturers (including legal proceedings against the National Institute for Health and Care Excellence [NICE]) and by patient

groups financed by these companies, influenced NICE to reverse its initial guidance in relation to these drugs, endorsing what are essentially useless treatments, and allowing their availability to UK patients on NHS subsidy (NICE, 2011).

The undue influence of pharmaceutical companies through exploiting the weaknesses within EBM has transformed Western medicine over the past fifty years, both in terms of how new therapeutic knowledge is disseminated to doctors, as well as how medicine is now practised. In terms of the original goals of protecting the public from dangerous treatments and improving patient care by making the most effective and economical treatments available, the system has been turned on its head (Greenhalgh et al., 2014, Goldacre, 2012; Healy, 2012). There have been tragic consequences for both medical practice and scientific research. In terms of the former, there are now some very harmful and yet minimally effective drugs in widespread use, some of which should never been allowed onto the market. Moreover, buried amongst the data, most of which has not been published, there may be a few highly effective drugs that are very specific in terms of the subgroup of patients in which they provide their maximum benefits, and possibly also quite specific in terms of the particular groups of patients in which the risk to benefit ratio would preclude their use. Unfortunately, given the present state of affairs, we will never know. In terms of scientific research, looking at the host of deleterious consequences, even allowing for the degree of cheating that has come to light, the methodology itself, even if strictly and carefully followed, must also be called into question.

Government intervention is required in order to wrest control of both the research agenda as well as the collection, collation and dissemination of the data away from the pharmaceutical industry. In view of the industry's enormous wealth and political influence, this concern is part of the larger issue of government regulation of the 'free market', versus a laissez-faire approach. Since Alan Greenspan's admissions at the House Oversight Committee hearings in the wake of the Global Financial Crisis (Waxman-Greenspan, 2008; Patel, 2009), the debate about whether or not our market economy requires stringent regulation by governments is much like a debate about whether or not we need traffic lights at intersections and a speed limit. Government intervention is necessary, and it is essential if we are to have a research agenda driven by the welfare of patients rather than the profits of large multinational companies. The possible solutions to this impasse and proposals for future development along these lines have been discussed in detail elsewhere (Healy 2012; Götzsche, 2013; Goldacre, 2012; Kassirer, 2005). However, in view of the unremitting power and influence of the vested interests, it appears that little will change

in the short term. Therefore the warning ‘caveat emptor; caveat lector’ still stands until one simple, essential solution has been implemented: free public access to all clinical trials and to all of the raw data from trials.

Concluding remarks

In his 1965 address to the Heberden Society, Hill pointed out many of the weaknesses and difficulties within the burgeoning science of controlled clinical trials, noting that the pendulum of medical fashion had already begun to swing in the direction of controlled trials.¹⁴ It is a record of issues unresolved and warnings unheeded: undue reliance on controlled trials as a means to gain clinical knowledge; the risks of extrapolating from trial results (i.e. generalising from the results in the particular group of patients); the use of complex statistical techniques to legitimise useless or trivial data; the need to account for biological variability and perhaps identify a subgroup of patients in whom an intervention may work well, in spite of this intervention not having been shown to be of benefit in the trial group as a whole; how to make statistical data relevant for the individual doctor in a particular clinical encounter; the limitations of double-blinding; and the tensions between subjectivity and objectivity. Hill also noted the emerging trend for industry to use RCT data in their sales patter, and also the trend within the medical profession to downplay the importance of clinical observations and clinical reasoning in favour of trial data that may just as equally ‘mislead as well as to lead’ (Hill, 1966).¹⁵

It is evident that EBM as it stands is neither working in the interests of doctors nor their patients. A response to ‘Evidence based medicine: a movement in crisis?’ (Greenhalgh et al., 2014) sums up the dilemma for doctors:

‘The basic issue for a practising doctor is that most of the evidence based literature is by researchers and pharma companies. The so called evidence is ambiguous [sic] contradictory most of the time with no take home messages or real answers. The literature in EBM appears to give more importance to statistical validity and correctness and after the reading the articles leaves us more confused than ever. I tried to learn statistics and invested considerable amount of time in attending workshops on Biostatistics but I still fail to get practical information which I can apply to the patient and explain to the patient in simple language.’ (Kumar, 2014)

Epidemiological methods, which are used in order to quantify the risks posed by various hazards, including environmental factors and drug treatments, have been applied and developed within medicine to include the quantitative assessment of beneficial effects of drug treatments. This methodology has always been strong on correlation and weak on causation, and this was quite

suitable for its original purpose: quantified risk was to be considered in reference to the qualitative variables of severity of risk, inconvenience or suffering that may arise from avoiding the putative risk factor/s, perceived benefits provided by the risk factor, etc. Thus, if we only have very slight evidence that an anti-emetic drug may cause abortion, we would be justified in restricting its use in pregnant women with morning sickness, regardless of the discomfort that may result from this decision¹⁶ (Hill, 1965). In other words ‘causal decision-making in epidemiology cannot be a statistical or quantitative procedure, but rather the critical thinking (speaking) and creative reasoning of the physician in cooperation with the patient, who preferably has similar qualities’ (Maier & Shibles, 2011). Unfortunately, this great strength of epidemiological methodology – to quantify risk factors – has been turned on its head by the pharmaceutical industry, who are using the appearance of this methodology to hide important risks. Moreover, their pervasive influence appears to be driving the move to downplay qualitative considerations within medical research.

It is evident that EBM as it stands is neither working in the interests of doctors nor their patients.

Provided that unrealistic expectations are not placed upon them, and that they are seen for what they are, the tools of EBM have the potential to enhance clinical practice. ‘EBM is not a new philosophy of medicine, but is instead a useful, imperfect tool available as an aid in making individual and group health care decisions, and in discussing care with patients’ (Cohen et al., 2004). However, the evidence hierarchy does not serve practising clinicians well; under the banner of EBM, the richness and complexity of ‘clinical evidence’ has devolved to signify the results of a properly conducted RCT or equivalent, and, as we have seen above, such trials may not in many instances have been properly conducted. Moreover, such a narrow scope of ‘evidence’ may have only a tangential relationship to clinical practice in the real world. Indeed, we also need to stop conflating ‘evidence’ with ‘proof’ and to refrain from using the term ‘evidence’ in this way. A more accurate description of the evidence hierarchy, taking into consideration twenty years of clinical application, would be to say that the items towards the top may potentially yield less biased outcomes, while those towards the bottom may potentially yield more biased outcomes. In other words, the results of an apparently well conducted RCT may in fact be heavily influenced by various biases, to the point where they are detrimental when applied clinically, while the seasoned judgement of a mature clinician who takes seriously the life-long commitment to

learning, may in fact provide the most effective solutions for a multifactorial clinical encounter.

Many of the commentators cited above have been quick to point out that proper evaluation of new drug treatments can only occur when all clinical trial data are made freely available to the medical community as well as the general public, as this is the best way to detect errors, both unintentional and intentional. While this is a self-evident truth, the underlying idea is that 'scrutiny of data by others is a fundamental aspect of science that moves science forward' (ANH-Intl, 2013).

In place of the linear 'evidence hierarchy', a circular arrangement of a broad variety of sources of clinical knowledge is proposed...

The implication is that medical science as a whole (not simply the aspect of drug evaluation) can only progress through properly conducted, reported and disseminated RCTs. There is a deep philosophical conundrum here. To put it bluntly, this is decidedly not how medical science or any other science moves forward. As noted elsewhere, EBM presumes that observations can be made by a naïve, completely objective observer, and that medical knowledge should be based on these supposedly unbiased observations. In this way, EBM seeks to remove medicine from its theoretical underpinnings. However it is impossible for anyone to make an observation that is objective and free from the theories and biases that are held in the mind of the observer (Harari, 2001; Cohen et al., 2004). Science progresses through the interplay of theory and observation: theory is challenged by progressively clearer observations, which in turn lead to new and better theories, which in turn permit more specific, more detailed and hence more useful observations. Progress in medical knowledge has the twin arms of improved understanding of physiological and pathophysiological processes,¹⁷ together with the associated improvement in medical observations. EBM has led to many disastrous consequences and the expectation that they can be remedied or prevented by more of the same is simply misguided.

Clinicians need a much broader scope of sources for medical knowledge than simply the undue reliance on RCTs. We need to preserve, renew and invent useful and valid means of acquiring clinical knowledge as well as a flexible system for prioritising various forms of clinical knowledge in different scenarios. Instead of denying or shying away from the essential subjectivity inherent in all aspects of clinical medicine, it should be acknowledged and accorded a central place within a new schema of clinical evidence. In place of the linear 'evidence hierarchy', a circular arrangement of a broad variety of sources of clinical knowledge is proposed - a circle around

the circumference of which are arranged different types of medical knowledge to be drawn upon, as appropriate, by the clinician placed at the centre, exercising his or her clinical judgement, informed by past experience and the present data from the patient. We honour Sackett's use of the words 'judicious' and 'integrating' when we place the clinical judgement and expertise of the practitioner at the centre of the evidence wheel: 'the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.' (Sackett et al., 1996).

However, this model also neglects the patient, who needs to be accorded a rightful place within the evidence continuum. The patient may also be viewed as the centre of another circle around which are arranged the events of his/her life, past and present, including the set of interrelated problems of which the current clinical presentation is a part. Perhaps the area of overlap between these two circles, always including the two central areas, may best represent the clinical encounter.

With the current trend towards increasing research within complementary modalities, including the development of suitable models for such research, it is crucial that the research agenda begins and remains in the hands of the profession, and that we learn from the history of EBM in its Western medical context so that we may remain vigilant over the course of the continuing development of the complementary medicine evidence base.

Logged in journal subscribers can comment on this article at www.jcm.co.uk

Endnotes

- 1 Caveatemptor: 'buyer beware'; caveat lector: 'reader beware'.
- 2 This example is adapted from Healy, 2012, p. 81-82, who links research of this nature to what has actually been occurring in regard to the promotion of SSRI's in the treatment of depression by industry.
- 3 The p value expresses the probability that the results could have occurred due to the play of chance. By convention, trial results are said to be statistically significant when p is less than five per cent. However, as long as p is greater than zero, there is a finite possibility that the results could, in fact, have occurred by chance. In other words whatever can happen will happen. Strictly speaking, in order to properly validate statistical significance, a trial should be repeated 20 times and should show the same result 19 times.
- 4 Although the rudiments of statistical knowledge and enquiry go back to ancient Greek mathematicians and Roman philosophers, the science of statistics has its origins in the perennially popular pastime of gambling. The church, which dominated scholarship up until fairly recently, frowned upon discussions of randomness or chance events, for obvious reasons (Cowles & Davis, 1982).
- 5 The standard for a course of acupuncture in Chinese hospitals includes a high level of training and experience of the acupuncturist administering treatment, prolonged needle retention and ten 'daily' treatments, delivered over 12 days (to allow for the

- weekend). In Western countries, the acupuncturist generally has considerably less training and experience, needles may be retained for as little as 15 to 20 minutes, the number of treatments varies, and the frequency of treatments may range from twice weekly to once monthly. There appears to be a growing body of evidence supporting frequency of treatments at more than once a week and a course of no less than twelve treatments (McDonald, 2012).
- 6 The confidence interval (CI) is the range of values within which we are fairly certain that the true result lies. It is another way of expressing the statistical significance of the trial results. Generally a CI of 95 per cent is applied in RCTs, corresponding to the accepted cut-off for statistical significance of 5 per cent. This generally yields a small range of values, which the researchers are 95 per cent certain represent the actual trial result.
 - 7 This is generally displayed graphically, where all of the results are lined up in a column and the range of values for each trial is denoted by a horizontal line, making it readily apparent whether the requisite overlapping is present.
 - 8 If the 'lower confidence limit of every trial is below the upper confidence limit of all the others ... statistically speaking, the trials are homogenous' (Greenhalgh, 1997).
 - 9 The bell-shaped curve is the graphic representation of normal (or Gaussian) distribution, in which continuous data (e.g. height, weight, IQ of a human population) are represented. The average value (the mean) is generally at the peak of the bell, and as you move away from the mean value in either direction, the number of subjects decreases; 95 per cent of subjects have data that are within two standard deviations of the mean, i.e. 95 per cent of the population studied are represented in the bulk of the bell shape. The outliers under the outer lip of the bell, representing the other 5 per cent generally have the more extreme qualities.
 - 10 If you combine this concept with the observation that there are always more ways for something to be done incorrectly rather than correctly, or for things to go wrong rather than right, you have the basis of 'Murphy's Law'.
 - 11 Essentially this involves the application of scientifically rigorous methodology to the study of a single patient. As an example let us take a patient who has experienced a possible adverse reaction to treatment. The application of challenge, dechallenge and rechallenge gives us the best and most reliable information concerning this particular patient's response to the treatment. 'Challenge' is when the patient has received the treatment and shortly afterwards develops the adverse symptoms. Dechallenge is when the treatment is withdrawn. If the symptoms disappear, the treatment may have caused them. Rechallenge is when, after resolution of the symptoms, the treatment is given again. If the patient again exhibits the adverse symptoms, we can be sure that the treatment has caused these symptoms in this patient.
 - 12 Perhaps most succinctly epitomised in the Arabic proverb: 'They asked Abboud of Omdurman: "Which is better, to be young or to be old?" He said: "To be old is to have less time before you and more mistakes behind. I leave you to decide whether this is better than the reverse."' (Idries Shah. *Thinkers of the East*. London: The Octagon Press, 2002)
 - 13 A rating scale is not much better than a check-list, and requires interpretation both by the patient and the administrator (i.e. it is very far from being an objective measuring device, and it allows much scope for bias). At best it may be used as a guide for professional interrogation. However it is rarely administered by a qualified medical professional, who would know when to probe a little deeper and when to clarify a particular point. These issues are discussed in Healy (2012), pp. 177-181.
 - 14 The analogy being that 'any belief that the controlled trial is the only way [in which we can study therapeutic efficacy] would mean not that the pendulum had swung too far but that it had come right off its hook.'
 - 15 It is remarkably refreshing to read this fine example of medical discourse from the days before industry and other vested interests had placed their stranglehold around it. The paper is available for free download at: <http://ard.bmj.com/content/25/2/107.full.pdf+html>.
 - 16 This is the reason why regulatory authorities in Western countries have been able to ban or severely restrict the use of herbal medicines, for which there may be only a very small degree of risk, detected using very low level evidence: in the absence of acceptable empirical evidence of benefit, a minute risk is judged against no benefit at all. Thus, avoidance of risk trumps availability every time.
 - 17 And, in the context of psychiatry, the improved understanding of psychological, psychosocial and psychopathological processes.

References

- Abbot, P., Reath, J., Rosenkranz, S., Usherwood, T. & Hu, W. (2014). "Increasing GP Supervisor Research Skills – Enhancing Clinical Practice and Teaching". *AFP* (43) 5, 327-330
- Akobeng, A. (2005). "Understanding Randomised Controlled Trials", *Arch Dis Child*; 90, 840–844
- Akobeng, A. (2005). "Understanding Systematic Reviews and Meta-analysis", *Arch Dis Child*; 90, 845–848.
- Aldrich, J. (1995). "Correlations Genuine and Spurious in Pearson and Yule", *Statist. Sci.*, 10(4), 364-376
- Alliance for Natural Health International (ANH-Intl) (2013). "Interview with Dr Peter Götzsche, author of 'Deadly Medicines and Organised Crime'", available at: <<http://www.anh-europe.org/ANH-Intl+Exclusive+Interview+with+Dr+Peter+Götsche+author+of+Deadly+Medicines+and+Organised+Crime>> [accessed 18/05/15]
- Armitage, P. (1991). "Obituary: Sir Austin Bradford Hill, 1897-1991", *Journal of the Royal Statistical Society*, 154(3), 482–484
- Beasley, C., Dornseif, B., Bosomworth, J., Saylor, M., Rampey, A., Heiligenstein, J., Thompson, V., Murphy, D. & Masica, D. (1991). "Fluoxetine and Suicide: a Meta-Analysis of Controlled Trials of Treatment for Depression", *BMJ*, 303, 685-92
- Begaud, B. (1984). "Standardized Assessment of Adverse Drug Reactions: The Method Used in France. Special Workshop—Clinical", *Drug Inf J*, 18(3–4), 275–81
- Choudhry, N., Stelfox, H. & Detsky, A. (2002). "Relationships Between Authors of Clinical Practice Guidelines and the Pharmaceutical Industry", *JAMA*, 287(5), 612-617
- Claridge, J. & Fabian, T. (2005). "History and Development of Evidence-Based Medicine", *World J Surg*, 29(5), 547-53
- Coderre, S., Mandin, H., Harasym, P. & Fick, G. (2003). "Diagnostic Reasoning Strategies and Diagnostic Success", *Med Educ*, 37, 695–703
- Cohen, A., Stavri, P. & Hersh, W. (2004). "A Categorization and Analysis of the Criticisms of Evidence-Based Medicine", *Int J Med Inform*, 73(1), 35-43
- Cohen, S. & Herbert, T. (1996). "Health Psychology: Psychological Factors and Physical Disease from the Perspective of Human Psychoneuroimmunology", *Annu Rev Psychol*, 47, 113-142
- Cowles, M. & Davis, C. (1982). "On the Origins of the .05 Level of Statistical Significance", *Am Psychol*, 37, 553-558
- Darwin, C. (1909). *The Origin of Species*. P.F. Collier & Son Co.: New York
- Dossey, L. (1995). "How Should Alternative Therapies Be Evaluated? An Examination of Fundamentals", *Altern Ther Health Med*, 1(2), 6-10, 79-85
- Duggan, L., Fenton, M., Rathbone, J., Dardennes, R., El-Dosoky, A. & Indran, S. (2005). "Olanzapine for schizophrenia", *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD001359, available at <<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001359.pub2/pdf>> [Accessed 13/02/15]
- Everson-Rose, S. & Lewis, T. (2005). "Psychosocial Factors and Cardiovascular Diseases, *Annu Rev Public Health*, 26, 469-500
- Evidence-Based Medicine Working Group (1992). "Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine", *JAMA*, 268(17), 2420-5
- Fisher, R. (1926). "The Arrangement of Field Experiments", *Journal*

- of the Ministry of Agriculture, 33, 503-13
- Goldacre, B. (2012). *Bad Pharma. How Drug Companies Mislead Doctors and Harm Patients*. Faber and Faber Inc: New York
- Gordis, L. (2005). *Epidemiology*. 3rd ed. Elsevier: Philadelphia. In Maier & Shibles, 2011, Ch.4
- Gotzsche, P. (2013). *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*. Radcliffe Publishing: London
- Greenhalgh, T. (1997). "How to Read a Paper: Statistics for the Non-Statistician. II: "Significant" Relations and Their Pitfalls", *BMJ*, 315(7105), 422-425
- Greenhalgh, T. (1997). "How to Read a Paper: Papers that Summarise Other Papers (Systematic Reviews and Meta-Analyses)", *BMJ*, 315(7109), 672-5
- Greenhalgh, T., Howick, J., Maskrey, N. & Evidence Based Medicine Renaissance Group (2014). "Evidence Based Medicine: a Movement in Crisis?", *BMJ*, 348, g3725
- Greenwood, M. (2002). "Acupuncture and Evidence-Based Medicine: A Philosophical Critique", *Med Acupunct*, 13(2), page number missing
- Harari, E. (2001). "Whose Evidence? Lessons from the Philosophy of Science and the Epistemology of Medicine", *Aust NZ J Psychiatry*, 35(6), 724-30
- Harcombe, Z., Baker, J., Cooper, S., Davies, B., Sculthorpe, N., DiNicolantonio, J. & Grace, F. (2015). "Evidence from Randomised Controlled Trials Did Not Support the Introduction of Dietary Fat Guidelines in 1977 and 1983: a Systematic Review and Metaanalysis." *Open Heart*, 2(1): doi:10.1136/openhrt-2014-000196
- Healy D. (2003). *Let Them Eat Prozac*. James Lorimer and Company: London
- Healy, D. (2008). *Mania. A Short History of Bipolar Disorder*. The Johns Hopkins University Press: Baltimore
- Healy, D. (2012). *Pharmageddon*. University of California Press: Berkeley & Los Angeles
- Hempel, S. (2009). *The Strange Case of the Broad Street Pump: John Snow and the Mystery of Cholera*. University of California Press: Berkeley
- Heres, S., Davis, J., Maino, K., Jetzinger, E., Kissling, W. & Leucht, S. (2006). "Why Olanzapine Beats Risperidone, Risperidone Beats Quetiapine, and Quetiapine Beats Olanzapine: an Exploratory Analysis of Head-to-Head Comparison Studies of Second-Generation Antipsychotics", *Am J Psychiatry*, 163(2), 185-94
- Higgins, J. & Green, S. (Ed.) (2011). "Assessing Risk of Bias in Included Studies in Cochrane Handbook for Systematic Reviews of Interventions, Part 2: General Methods for Cochrane Reviews, Assessing Risk of Bias in Clinical Studies", available at: <http://handbook.cochrane.org/chapter_8/8_4_introduction_to_sources_of_bias_in_clinical_trials.htm> [Accessed 21/11/14]
- Hill, A. (1965). "The Environment and Disease. Association of Causation?", *Proceedings of the Royal Society of Medicine*, 58, 295-300
- Hill, A. (1966). "Reflections on the Controlled Trial", *Ann. Rheum. Dis.*, 25, 107-113
- Horwitz, R. (1987). "Complexity and Contradiction in Clinical Trial Research", *Am J Med*, 82(3), 498-510
- Horwitz, R. (1995). "Large-Scale Randomized Evidence: Large, Simple Trials and Overviews of Trials: Discussion. A Clinician's Perspective on Meta-Analyses", *J Clin Epidemiol*, 48(1), 41-4
- Kassirer, J. (2005). *On the Take: How Medicine's Complicity with Big Business Can Endanger Your Health*. Oxford University Press: New York
- Kaul, S. & Diamond, G. (2010). "Trial and Error. How to Avoid Commonly Encountered Limitations of Published Clinical Trials" *J Am Coll Cardiol*, 55(5), 415-427
- Khayyam-Nekouei, Z., Neshatdoost, H., Yousefy, A., Sadeghi, M., & Manshaee, G. (2013). "Psychological Factors and Coronary Heart Disease", *ARYA Atherosclerosis*, 9(1), 102-111
- Klein, N. (2008). *The Shock Doctrine: The Rise of Disaster Capitalism*. Picador: New York
- Kravitz, R. & Duan, N. (Ed) (2014). "Design and Implementation of N-of-1 Trials: A User's Guide", AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality, available at <www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm> [Accessed 09/02/15]
- Kumar, A. (2014). "BMJ Rapid Responses to Greenhalgh et al. (2014). Evidence based medicine: a movement in crisis?", available at: <http://www.bmj.com/content/348/bmj.g3725/rr/811572> [Accessed 09/02/15]
- Lancôt, K., Herrmann, N., Yau, K., Khan, L., Liu, B., LouLou, M. & Einarson, T. (2003). "Efficacy and Safety of Cholinesterase Inhibitors in Alzheimer's Disease: a Metaanalysis", *CMAJ*, 169, 557-64
- Lancôt, K., Rajaram, R. & Herrmann, N. (2009). "Review: Therapy for Alzheimer's Disease: How Effective Are Current Treatments?", *Ther Adv Neurol Disorders*, 2(3), 163-180
- Lang, E. (2004). "The Why and the How of Evidence Based Medicine", *MJM*, 8, 90-94
- Lelorie, J., Gregoire, G., Benhaddad, A., Lapiere, J. & Derderian, F. (1997). "Discrepancies between Meta-Analyses and Subsequent Large Randomized, Controlled Trials", *N Engl J Med*, 337(8), 536-542
- Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). "The N-of-1 Clinical Trial: the Ultimate Strategy for Individualizing Medicine?", *Personalized Medicine*, 8(2), 161-173
- Little, M. (2003). "'Better than Numbers ...' A Gentle Critique of Evidence Based Medicine", *ANZ J Surg*, 73(4), 177-82
- Maier, B. & Shibles, W. (2011). *The Philosophy and Practice of Medicine and Bioethics. A Naturalistic-Humanistic Approach*. Springer Science + Business Media: Dordrecht
- Manheimer, E., Cheng, K., Linde, K., Lao, L., Yoo, J., Wieland, S., van der Windt, D., Berman, B. & Bouter, L. (2010). "Acupuncture for Peripheral Joint Osteoarthritis", *Cochrane Database of Systematic Reviews*, Issue 1. Art. No.: CD001977. Available at: <http://www.thecochranelibrary.com/userfiles/ccoach/file/Acupuncture_ancient_traditions/CD001977.pdf> [Accessed 23/12/14]
- Matsui, K. & Kawaguchi, K. (2014). "The Development of Pattern Recognition via Clinical Experience: A Preliminary Study", *Educ in Med J*, 6(4), 57-65
- McDonald, J.L. (2012). "Acupuncture Research: Current Status and Future Directions". International Scientific Acupuncture and Meridian Symposium (iSAMS), University of Technology Sydney (UTS), Sydney, Australia. From as yet unpublished PhD paper: "The Effects of Acupuncture on Mucosal Immunity in the Upper Respiratory Tract" submitted to Griffith University in September 2014
- Morris, H., (2003). "The Mathematical Impossibility of Evolution." Institute of Creation Research Website. Available at <www.icr.org/article/493/> [Accessed 19/05/15]
- Muntz, S. (1986). *Sweetness and Power: The Place of Sugar in Modern History*. Penguin Books: New York
- National Institute for Health and Care Excellence (NICE) (2011). *Donepezil, Galantamine, Rivastigmine and Memantine for the Treatment of Alzheimer's Disease*. Guidance for Dementia, available at <http://www.nice.org.uk/guidance/ta217> [Accessed 29/01/15]
- Nissen, T. & Wynn, R. (2014). "The Clinical Case Report: a Review of Its Merits and Limitations", *BMC Res Notes*, 7, 264
- Office of Medical Education, University of Sydney (2014). Evidence Based Medicine, from Office of Medical Education, About Us, Educational Theory, available at <http://sydney.edu.au/medicine/ome/about/educational-theory.php#ebm2> [Accessed 14/11/14]
- Oschman, J. & Oschman, N. (1997). *Readings on the Scientific Basis of Bodywork, Energetic, and Movement Therapies. A Collection of 21 Articles*. N.O.R.A. Press: Dover
- Patel, R. (2009). *The Value of Nothing: How to Reshape Market Society and Redefine Democracy*. Picador: New York
- Pelaccia, T., Tardif, J., Tribi, E. & Charlin, B. (2011). "An Analysis of Clinical Reasoning Through a Recent and Comprehensive Approach: the Dual-Process Theory", *Medical Education Online*, 16, 10.3402/meo.v16i0.5890. doi:10.3402/meo.v16i0.5890
- Phillips, C. & Goodman, K. (2004). "The Missed Lessons of Sir Austin Bradford Hill", *Epidemiol Perspect Innov*, 1(1), 3
- Sackett, D. (2000). "Why Randomized Controlled Trials Fail but Needn't: 1. Failure to Gain 'Coal-Face' Commitment and to Use the Uncertainty Principle", *CMAJ*, 162(9), 1311-1314
- Sackett, D., Rosenberg, W., Gray, J., Haynes, R. & Richardson, W. (1996). "Evidence Based Medicine: What It Is and What It Isn't", *BMJ*, 312(7023), 71-72
- Santoro, A., Siviero, P., Minicuci, N., Bellavista, E., Mishto, M., Olivieri, F., Marchegiani, F., Chiamenti, A., Benussi,

- L., Ghidoni, R., Nacmias, B., Bagnoli, S., Ginestroni, A., Scarpino, O., Feraco, E., Gianni, W., Cruciani, G., Paganelli, R., Di Iorio, A., Scognamiglio, M. & Grimaldi, L. (2010). "Effects of Donepezil, Galantamine and Rivastigmine in 938 Italian Patients with Alzheimer's Disease: a Prospective Observational Study", *CNS Drugs*, 24(2), 163-76
- Simundic, A. (2013). "Bias in Research", *Biochemia Medica* 2013, 23(1), 12-15
- Smith, G. & Pell, J. (2003). "Parachute Use to Prevent Death and Major Trauma Related to Gravitational Challenge: Systematic Review of Randomised Controlled Trials", *BMJ*, 327(7429), 1459-1461
- Spence, D. (2013). "What is the Point of Doctors", *BMJ*, 347, f7380. doi: 10.1136/bmj.f7380.
- Stoppard, T. (1967). *Rosencrantz and Guildenstern are Dead*. Faber and Faber: London
- Straus, S. & Sackett, D. (1998). "Using Research Findings in Clinical Practice", *BMJ*, 317, :339-42
- Tampi, R. & van Dyck, C. (2007). "Memantine: Efficacy and Safety in Mild-to-Severe Alzheimer's Disease", *Neuropsychiatr Dis Treat*, 3(2), 245-258
- The Cochrane Collaboration Website (2014). Glossary, available at <<http://www.cochrane.org/glossary/5#term372>> [Accessed 05/12/14]
- Tonelli, M. (1998). "The Philosophical Limits of Evidence-Based Medicine", *Acad Med*, 73(12), 1234-40
- Tonelli, M. & Callahan, T. (2001). "Why Alternative Medicine Cannot Be Evidence-Based", *Acad Med*, 76(12), 1213-20
- U.S. Food and Drug Administration (2009), Promoting Safe and Effective Drugs for 100 Years, available at <<http://www.fda.gov/AboutFDA/WhatWeDo/History/ProductRegulation/PromotingSafeandEffectiveDrugsfor100Years/>> [Accessed 05/01/15]
- Uman, L. (2011). "Systematic Reviews and Meta-Analyses", *J Can Acad Child Adolesc Psychiatry*, 20(1), 57-59
- University of Western Sydney (2014), Postgraduate courses, Health Science Courses, Master of Chinese Medicine, available at <http://www.uws.edu.au/future-students/postgraduate/postgraduate/postgraduate_courses/health_science_courses/master_chinese_medicine> [Accessed 14/11/14]
- US Department of Health and Human Services website, (2014) Agency for Healthcare Research and Quality, from Glossary of Terms, available at <<http://effectivehealthcare.ahrq.gov/index.cfm/glossary-of-terms/?pageaction=showterm&termid=70>> [Accessed 14/11/14]
- Waxman-Greenspan Testimony (2008). "Waxman questions Greenspan on ideology, October 23, 2008." From C-Span Video Library, available at: <<http://www.c-span.org/video/?c3342718/waxman-greenspan-testimony>> [Accessed 14/11/14]
- Wen, L. & Kosowsky, J., (2012). *When Doctors Don't Listen: How to Avoid Misdiagnoses and Unnecessary Tests*. St Martin's Press: New York
- White, A., Cummings, M., Barlas, P., Cardini, F., Filshie, J., Foster, N., Lundberg, T., Stener-Victorin, E. & Witt, C. (2008). "Defining an Adequate Dose of Acupuncture Using a Neurophysiological Approach--A Narrative Review of the Literature", *Acupunct Med*, 26(2), 111-20