WHY EVIDENCE BASED MEDICINE MAY BE BAD FOR YOU AND YOUR PATIENTS

Geoffrey M. Shaw[1] MBChB, FANCZA, FJFICM

J. Geoffrey Chase[2] PhD

[1] Clinical Sr. Lecturer Christchurch School of Medicine and Health Sciences, University of Otago, c/o Department of Intensive Care, Christchurch Hospital, Christchurch, New Zealand.
Email: geoff.shaw@cdhb.govt.nz

[2] Professor, University of Canterbury, Centre for Bio-Engineering, Department of Mechanical Engineering, Private Bag 4800, Christchurch, New Zealand.
Email: geoff.chase@canterbury.ac.nz

Word count: 4901

ABSTRACT

Evidence based medicine (EBM) has inveigled itself into almost every aspect of medical practice in the last decade (or more) with promises of better delivery of health care and a more rational basis by which to deliver that care. Thus, an appropriate question at this time might be: How does the evidence for the so-called "evidence" of EBM stack up?

This review argues that the basis of EBM is so deeply flawed that in many cases it cannot usefully inform clinical practice, reflected in fact by the current majority outcome of most trials as "no-blood," or no result. The illusion that knowledge can be created using only empiric data from large randomised controlled trials (RCTs) that test single hypotheses and draw conclusions from multi-variate regression correlations is examined. Particular attention is paid to how this idea undermines the advancement of in-depth, fact-based medical knowledge and, as a result, patient care.

The flawed logic behind the EBM rationale of treating *risk,* as determined by probabilistic outcome measurements, means EBM will never be equipped to identify individual benefit or harm. Thus, it can never usefully inform the attending clinician about titrating therapy to the individual case. Hence, the results of EBM can never improve a *single* patient's care, only the averaged odds faced by that patient being treated by a statistical outcome correlation based therapeutic choice, at best.

Finally, despite its emphasis on large trials utilising inherently flawed logic, EBM is ruling the "mindshare" of medical and administrative thinking, precluding other lines of thought and stifling scientific debate. Clinicians need to re-think how they gather, interpret and act on evidence. and, in fact, what defines "evidence" outside the definition of trial size, randomisation and blinding. Through its implicit emphasis on treating risk, and not disease, EBM paradoxically risks much more.

The monotonous march of performing large RCTs, crunching the statistics and presenting the results for the latest journal issue, has come to define medical research. As a result, it has stifled the idea that reasoned evidence, based on the complementary use of both hypothesis-based trials and logic-based reasoning has any contribution to make in this brave new world. Forgetting of course that it was just such complementary forms of research engagement that drove the last several hundred, if not thousand, years of scientific and medical research – until now.

So, enter, if you will, this editorial on the state of the evidence, or lack thereof, in medical research today. The authors claim no competing interests outside the biases that they entered the room with.

INTRODUCTION

One of the first reported clinical trials designed to test a hypothesis was conducted by Archie Cochrane at a POW camp in Salonica, Greece, during the Second World War.[1] He had noticed an increasing number of prisoners complaining of heavily swollen ankles that seemed to affect the British more seriously, but apparently did not affect the cooks. He reasonably concluded that the cause was most likely nutritional and diagnosed hypoproteinaemic oedema. However, the Germans ignored his concerns attributing the oedema to too much sun and refused to measure the blood protein.

Soon he was seeing more than 20 cases per day. The numbers kept rising and the situation was getting more desperate. To avoid panic he disguised the figures by only including those who had pitting oedema above the knee. To convince the Germans Archie needed another angle, since the Germans were arguably less interested in Allied welfare at the time, and soon convinced himself he was witnessing an epidemic of *wet beriberi.*

He decided to do an experiment, and randomly selected a cohort of 20 young males, all in their early 20's, all emaciated with oedema above the knee. He assigned 10 in each of two small wards. All the prisoners were given the same food rations, however those in one ward, the "yeast room," received a supplement of yeast three times a day, (purchased on the black market, supplier unnamed), while those in the other received vitamin C.

By the third to fourth day differences became apparent with 8 out of 10 in the yeast room feeling better, while none in the vitamin C room improved. Archie persuaded his captors that the thiamine in the yeast extract had cured the beriberi and ankle oedema. Based on this clear evidence the Germans changed their stance, and the next morning a large amount of yeast arrived. Within a few days the food rations were increased from 600 to 800 calories per day.

However, there was just one problem. The hypothesis was wrong (ankle oedema due to beriberi) and the treatment (yeast extract) accidentally provided the "cure" to this wrong problem. In actual fact, the ankle oedema was due to hypoprotienaemia, and *not* beriberi.

Archie was later "horrified" that he had deceived himself, and the Germans, believing deep down that he was always dealing with hypoproteinaemic oedema. In his words[1]:

"As regards the trial I always felt rather emotional about it and ashamed of it, I have seldom referred to it since. It was a poor attempt. I was testing the wrong hypothesis, the numbers were too small, and they were not randomized. The outcome measure was pitiful, and the trial did not go on for long enough……my first, worst, and most successful trial."

Thus, was born the first origins of the tsunami that is today evidence based medicine (EBM), which governs the very financial and medical decisions that dictate how, when, and where treatment for many serious diseases is given, … or withheld.

In particular, there is a growing demand for "evidence" or proof through randomised clinical trials (RCTs) as the gold standard. However, despite the clear warning of Archie Cochrane's actual result, there is no equivalent demand for evidence that the underlying hypothesis or treatment being tested is correct or likely. Nor is there an equivalent demand that the large, costly RCTs that rarely deliver statistically significant outcomes provide the cost efficiency that their backers often claim will come from using the results.

WHAT'S WRONG WITH THE EVIDENCE

Recently, we have seen a variety of very large, very costly clinical RCTs, all designed to better inform practice. One such area of intense effort relevant to critical care has concerned the treatment of acute sepsis. The Surviving Sepsis Campaign, generously supported by Eli-Lilly, graded 52 recommendations for treatment.[2] The grades were scaled from A (supported by at least two large RCTs with clear cut results and low risk of type I and II errors) to E (supported by level IV or V evidence; non-randomised studies, historical controls, expert opinion, case series etc).

The results? Of the 52 recommendations, 36 were "C" or lower. Of these only 9 were supported by small RCTs or trials using non-randomised controls with uncertain results and moderate to high risk of type I or II errors. The remaining 27 (over half of the 52 recommendations) were graded "E". Only 5 recommendations were graded "A".

Even more interestingly, of the "A" or "B" evidence just under half (7) recommended things that should *not* be done, and thus recommend no specific treatment(s). Hence, they excluded just one practice (of many), leaving every other treatment option completely open. This solution by gradual elimination is much like telling small child what they cannot do and not telling them about the expected or necessary behaviour(s) you, as the parent, should expect. In contrast, of the grade "E" evidence the vast majority (24 of 27) was for specific treatments we should do, but lacking in convincing, statistical proof.

The outcome? Where convincing evidence does exist EBM tells us more about what not to do than what specifically will work! Interestingly, there was no new science, physiology or pathology reported and our understanding of the disease and how it works was little changed. Thus, outside of "don't do this" and a few broad recommendations, the field itself stood has largely still, at the cost of many millions of dollars not spent on treating patients.

The above exemplifies how "evidence" is becoming defined only in terms of empirically derived data, with the lack of such data translating into the mantra *"there is no evidence…"*. Of great concern is that no consideration is given for evidence obtained through deductive or fact-based logically reasoned evidence.

The flaws and the inequity of the EBM approach was recently lampooned by Smith et al[3] who, writing in the British Medical Journal, noted that while parachutes were well known to prevent injury, there was no randomised evidence to "prove" their efficacy for "gravity challenges" and somewhat tongue-in-cheek suggested that the

protagonists of EBM should therefore have equipoise and volunteer for such a randomised trial!

The reality is that most successful and typical treatments lack evidence – like the parachute. For example, we regularly ventilate sedated patients, yet no RCT has *proven* it to be more effective in any outcome than not ventilating. Similarly so for even sedating these patients in the first place, or treating pneumonia with penicillin or any other antibiotic (versus not treating them with antibiotics).

We use these hopefully obvious examples to point out that many of the common "foundation" therapies we take for granted today, such as sedation, antibiotics, ventilation, and many others, were *never proven* to the satisfaction or standard demanded by today's EBM protagonists. By today's EBM debate, we should stop these treatments until they are truly proven. Yet, to do so would cause needless suffering and death. This course of action is about as logical as banning parachutes for skydivers until they are proven more safe than any other randomly selected alternative therapy for gravity challenges.

More specifically, there is not one truly new and novel therapeutic approach in use today that was developed and based on EBM alone. The foundation of modern medicine, from which effectively all modern therapies have their origin, is based on hypothesis-based deductive reasoning whose origins lie in "first principles" application of physiology, bio-chemistry and clinical experience. Or, in some cases, just plain serendipity!

REASONING

So, how was it we got so far in the first place? Let us first consider how we acquire knowledge. The process we use to link knowledge to the observed world is called *inferential* reasoning of which there are 2 types [4]:

- Deductive inference (objective)
- Inductive inference

With deductive inference, we start with a hypothesis that we assume to be true. We then predict what we should expect to see. This reasoning is objective because what we predict will always be true, so long as the hypothesis is true. The problem is that this method of reasoning does not allow us to predict behaviours or develop our knowledge beyond the original hypothesis. This is problematic because no new knowledge can be created.

Inductive inference goes in the opposite direction. Hence, our observations can be used to develop new hypotheses. For example, deductive reasoning tells us all patients with asthma will wheeze (or have airflow limitation), but Chevalier Jackson over 100 years ago recognized through inductive reasoning "*all that wheezes is not asthma*"[5] In other words, wheeze is not just limited to the diagnosis (hypothesis) of than asthma. Hence, the clinical history and examination is used to gather as much data as possible, from which we can infer several other tenable hypotheses - the differential diagnosis. If for example our differential diagnosis includes inhalation of

a foreign body, we can look for it with a bronchoscope. If we find a peanut, we can then deduce this caused the wheeze. Deductive and inductive inferential reasoning are so truly inextricably entwined in the practice (and research) of clinical medicine that it is sometimes difficult to determine what is a fact from what is assumed or possibly fiction. More specifically, the practice of medicine uses both forms of reasoning repeatedly in even the simplest situations, often without conscious recognition of such.

In the same way we reason and formulate diagnoses for our patients, scientists using the same reasoning develop new paradigms and hypotheses. *Rationalism* emphasises new knowledge that is acquired through reason.[6] Essentially, the rationalist acquires knowledge by carrying out simple sequential experiments of "n=1". Each experiment builds on knowledge and understanding gained from all the previous work so that in time a *coherent body of evidence* is developed. As long as we continue to ensure that our paradigms (i.e. the internal models of the physical world on which we base our understanding) and our observations of the physical world are coherent, the closer we get to the truth.

However, knowledge can also be derived through the senses alone. For example, observations noting the time the sun rises every morning do not require reasoned understanding of celestial mechanics, yet these observations can be used to predict the exact time the sun will rise every day throughout the year. This is *empiricism*, which emphasises new knowledge acquired by observation.[6] Hence, this knowledge is not acquired by reason, and never has to be.

Claude Bernard, one of the first rationalists in modern medicine used physiological "first principles" and reasoning to develop new knowledge and treatments[6]. However, his contemporary Pierre Charles Alexander Louis, an empiricist and founder of *Médecine d'Observation*, believed empiric data could (and preferably should) be used inform the clinician[6]; with no biologically plausible mechanism being required (e.g. the use of leeches in pneumonia). Our limited understanding of the complex physical processes occurring in the natural world are usually based on this form of empiric evidence and much of our medical knowledge is necessarily derived in this manner.

THE EBM MODEL: THE END JUSTIFYING THE MEANS

This might all sound familiar and many might ask: How is the EBM model different? EBM uses the same data and methods as these two models, which have (in partnership) delivered all we have today, but … in reverse! Rather than state a hypothesis and prove it, or try something and see if it works (leading to further refinement on the evident data), EBM uses data and statistics to inferentially manufacture the knowledge. Much like Archie Cochrane found his cure for beriberi.

Let's examine further the trial of Archie Cochrane. Archie assumed the hypothesis that the yeast extract (with vitamin B) would reduce the ankle oedema. His results showed this basic hypothesis to be true. So far, so good, in terms of the Rationalist Model. He next induced from these results that yeast extract cured beriberi through

the action of vitamin B, his (assumed) or inferred causal condition. However, this inference or induction was wrong.

More specifically, the EBM approach (hypothesis testing) ***relies on the assumption of truth in the null hypothesis***, when calculating a *P* value. It thus creates an illusion of evidence in the significance (or lack thereof) of that *P* value that may not be accurate or valid. It is this error that infects EBM today.

Regardless whether or not one accepts the rationalist or the empiricist models, the emphasis on empirically derived evidence in EBM puts the cart before the horse. In addition, the EBM model does not, and cannot, provide any *useful* guidance on how to interpret reasoned data. In fact, in the EBM paradigm expert opinion based on case series (rationalism) is relegated to the lowest grade, Grade V, of evidence.

In the 1920's and 1930's scientists pushed for alternative methods of statistical inference that used only deductive probabilities, hoping to avoid Bayesian statistics. Fisher proposed the *P* value as measure of the strength of evidence if an experiment is repeated many times - a *frequentist* approach that is not truly viable for today's large RCT. Concomitantly, Neyman and Pearson developed the methods of hypothesis testing.

However, this approach was problematic because it could not give the probability of an underlying truth. Such a method would require inductive elements that would inevitably lead back to Bayes theorem, the very thing they were trying to avoid[4]. This significant short-coming was rationalised by Neyman and Pearson who stated:

"…no test based upon a theory of probability can by itself provide any valuable evidence of the truth or falsehood of a hypothesis… Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not often be wrong" [7]

However, R.A. Fisher was perturbed that science could be advanced along these lines[8]:

"*The idea that this responsibility can be delegated to a giant computer programmed with decision functions belongs to a phantasy of circles rather remote from scientific research*"

Doctors and medical students will often state that when the *P* value is <0.05 there is a 95% or greater chance the null hypothesis is incorrect. However, the *P* value is calculated on the assumption the null hypothesis is correct and therefore can never be a measure of the probability the null hypothesis is false[4]:

"The P value is defined as the probability under the assumption of no effect or no difference (the null hypothesis) of obtaining a result equal to or more extreme than what was actually observed"

Knowledge acquired through hypothesis testing uses deduction (*P* values) to test or compare treatments, and uses induction from the outcomes of the study to infer

evidence (accepting or rejecting the null hypothesis). Steven Goodman, an epidemiologist from Johns Hopkins University School of Medicine argues[9]:

"*The approaches of each side have been improperly combined, creating a new procedure with such a strong illusion of coherence that even when it produces problems, the combination is not recognised as their source*"

In which he was perhaps trying to tell us that evidence should be based on measured, observed, scientific fact and first principles, rather than inferred from a potentially flawed hypothesis created from a statistical inference.

In spite of these problems, there is an almost slavish reliance on the *P* value of the results, the statistical goodness if you will, as well as the average difference or value it is differentiating. There is thus no means of using these results to treat or manage the variability in conditions and presentation and symptoms that define any typical or specific patient – and thus no means of individualising care.

Finally, Archie Cochrane felt ashamed he had deluded himself into thinking the ankle oedema was due to an epidemic of beriberi even though he deep down believed the problem was related to low dietary protein. But how else could he convince his captors to do anything? This raises an interesting additional question: to what extent do commercial or political interests, or necessity, over-ride our reasoning in favour of manufacturing knowledge that justifies the end? The recent situation in South Korea comes to mind in this instance.

More succinctly perhaps, statistical tests should be used as tools to confirm evidence, or *the means*, rather than being relied upon to generate new knowledge, *or the end*. In EBM today it seems as if the ends justify the means.

THE COST AND HARM OF EBM

Medicine was for many years based on various reasoning models that are well founded today in modern engineering. In particular, engineers, who are somewhat notorious in many places for their frugality, would rarely contemplate a major, costly, large trial without some sure knowledge of the likely outcome. For example, one would not imagine manufacturing an "all-new" jet fighter without testing prototypes first of the entire plane, not to mention individual critical elements like engines. Interestingly, the EBM approach assumes that any such foreknowledge destroys equipoise, and is thus to be avoided.

The end result is that engineers would tackle the problem as shown below in Figure 1. In contrast, today's EBM approach reverses this process with much of the research effort and expenditure being used for testing hypotheses that lack foundation- the ends justifying the means.

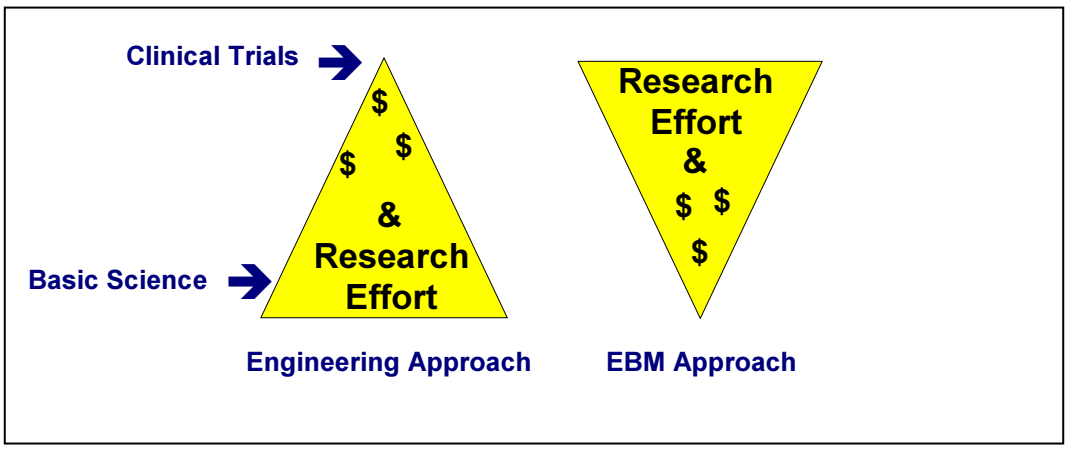

Figure 1. Two approaches to medical research. The "engineering approach" has a strong a foundation of basic science culminating in a successful RCTs, whereas the "EBM approach" has little foundation with many, costly and often unsuccessful, RCTs

The end result is wasted time, effort and cost. In contrast, if only successful science leads to RCTs the outcome is not only more efficient research (and many smaller trials yielding insight), but also likely fewer large RCTs, each of which requires potentially far fewer patients and far less cost. It is this same approach of following "winners" and culling "losers" that drives successful business and venture investors in all fields. In the same way, engineering research and science, and once upon a time medical science, presented several "horses" and followed the "winners" over several incremental steps to the foundation treatments we have today.

Of course, when one considers the 90% failure rate (to deliver a *P* value) in many modern RCTs, there is no discussion of what was learned. The reason is that with no *P* value there is no "inferred knowledge". However, what is also not discussed is the potential harm done to patients in not achieving an outcome. Thus, a limited number of RCTs based on very strong foundation knowledge and results, and thus a foreknowledge of a likely successful outcome, might well be less harmful to patients.

Consider a RCT of treatment A versus control B for condition X. Consider the outcome where 5% who received treatment A, who would have died, end up surviving. However, another 5% who receive treatment A, who would have survived, end up dying. The net outcome in terms of RCT and EBM statistics is no effect. However, the reality is that one (sub) group of patients had definite benefit, while another group was harmed.

Consider also the rise and subsequent fall of postmenopausal hormone replacement therapy (HRT). In 2002 researchers published principal results from the Women's Health Initiative (WHI) Randomized Controlled trial, of 16608 women aged 50-79 years who received either oestrogen and progesterone or placebo.[10] In May 2002 the data and safety monitoring board recommended stopping the trial after 5.2 years (planned duration 8.5 years) because the test statistic for invasive breast cancer exceeded the stopping boundary and the global index statistic supported risks exceeding benefits. Absolute excess risks per 10000 person years attributable to oestrogen plus progesterone were 7 more coronary heart disease events (non-fatal and fatal myocardial infarction) 8 more strokes, 8 more pulmonary embolisms, and 8 more invasive breast cancers. However, there were risk reductions of 6 fewer colorectal cancers, and 5 fewer hip fractures per 10000 patient years. This concern spurred other

researchers to examine the WHI findings along with three other earlier trials of HRT in over 20000 women. The results were much the same: an excess of stroke, pulmonary embolism and breast cancers, with a smaller reduction in colorectal cancer and hip fractures.[11]

Empiric evidence from hypothesis testing in randomised clinical trials is at the heart of EBM. However, this knowledge is probabilistic and implicit in this is the concept of risk. By treating *risk* as a surrogate of disease EBM fails to identify *individual* harm or benefit.

The problem of treating risk and subjecting large heterogeneous populations to a single intervention is that some patients are benefited, we don't know whom, and some are harmed, we also don't know whom. In the end, we have learned very little, but we have spent a great deal of time, labour and cost doing so. And of course there is a journal publication or two.

Treatment of risk is also attractive to health administrators, because it usually takes the form of an easily understood simple treatment that can be applied across an entire population. However, there is a catch. Specifically, probabilistic outcome measurements (risk or surrogate disease) cannot be tailored to individual responses and therefore treatment may continue indefinitely, as shown Figure 2.

In addition the benefits are likely to be time limited; there is more bang for your buck if you start early. However, when does the benefit become less than the risk? How long should we keep treating a patient to avoid the risk of disease? Or by analogy, how long is it useful to keep paying life-insurance? At some stage disease and/or death catches up with all of us.

In answer, the empiric data, so much beloved of EBM protagonists, cannot *usefully* inform the clinician when a treatment should be started, how it should be titrated, or when it should be stopped. Hence, guidance about the use of expensive individual treatments with significant risks, such as activated protein C in intensive care, remains loose[12].

For example, consider the small change to medical history if you will, if the PROWESS[13] study had titrated activated protein C to surrogates of sepsis (patient responses), rather than just body weight (not a surrogate of sepsis). Or perhaps tried several such titrations over a series of several smaller incremental pilot trials, and then built on this knowledge and developed *reasoned and rational* methods for the use of activated protein C. We might then not (still) be stuck debating the one-size-fits-all method used in this large costly RCT. Ultimately, and more importantly, we might have been better informed on how to best use the drug for less cost. Oh, and we would also have also done more to improve patient outcome and safety, as well.

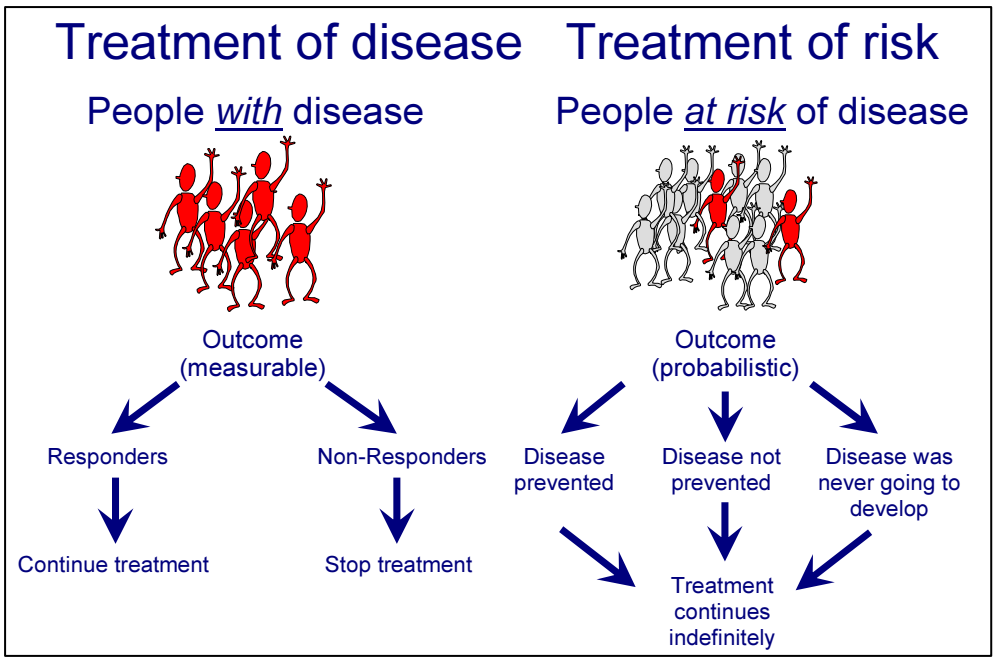

Figure 2. (Modified, with permission from the authors) [6] Treatment of risk might continue indefinitely without individual patient responses to guide treatment.

THE PROBLEM OF VARIABILITY

The cause of all this debate, conjecture and conundrum? Variability. No two patients respond the same for any or all treatments. Developing treatments based on averages and net outcomes over large cohorts totally ignores this fact. Yet it is the ongoing foundation for developing and "proving" new treatments.

EBM is the ultimate utilitarian approach to medicine, the greatest good for the greatest number. It is thus appears entirely ethically defensible to save a few by killing slightly fewer. However, it would be ethically more responsible to identify and save individuals who would die, while not harming those who would live. EBM does not and *cannot* provide the knowledge or means to do this.

Today, when it suits, many decisions to withhold payments for treatments, drugs, or clinical staff, are made using an EBM approach with no proof that the evidence, like Archie Cochrane's initial foray, is valid. This approach is variably, and at worst cynically, used to deny high cost individual treatments or pharmaceuticals to individuals.

Consider the plight of Ruby a 32 year old woman with treatment-resistant chronic arthritis who has tried a variety of different disease modifying drugs, all of which failed to control the inflammation and/or were associated with unacceptable side effects[6]. A trial of anti tumour necrosis factor (anti-TNF) therapy resulted in dramatic improvement, yet treatment had to be withdrawn through lack of funding. Anti-TNF therapy costs approximately $NZ30 000 per year and attracts no tax payer subsidy. However, an estimated 300 patients in New Zealand would benefit. In this case, the variable response of these patients and the "evidence" of efficacy don't match, and the losers are the individuals.

CONCLUSIONS

The knowledge gained in small trials and empiric testing yields valuable insights into the "how" of treating specific patients. This approach also yields insight into the variability in patient cohorts and responses to treatments that statistical correlations and averages cannot. As a result, this approach can provide a better knowledge base for individualising therapy and treatment, as well as for judging the appropriateness of a given therapy for a specific individual. Particularly, prior to, or in the process of, developing the foundation for a large RCT.

It might well be said that EBM provides only one side of the evidence supporting an underlying truth, and does not ever really examine that truth directly – because it is not equipped to do so. However, as EBM methods and approaches, with their large clinical trials, currently rule the "mindshare" of medical thinking, we might wish to raise one last question. Specifically, what is the harm being caused to the general advancement of medical knowledge by censoring knowledge with large trials utilising flawed methods of investigation that preclude the inputs of other lines of thought?

Is not the basis of the best science the open debate of ideas based on evidence, yet today's EBM approaches would say that no debate is needed as long as you have your p-value. Similar views have been ardently expressed by Charlton and Miles[14]:

"Tremendous advances have been made in establishing the EBM brand name, obtaining massive government funding, manoeuvring to a position of unchallenged authority within the NHS managerial and political hierarchy (and its priesthood among BMJ editorial staff) and promoting EBM as a marketing slogan for lucrative conferences, courses books, journals, people and organisations"

Equally importantly, we might also ask whether we are satisfied with how EBM is being used to determine treatment options and how it affects individuals in our care?

The answer? We need to rethink about how we gather evidence in clinical medicine. We need to revisit our roots in utilising our powers of deduction to develop the next generation of significant medical advances. Or perhaps we should take Max Planck's sage advice and just wait for EBM to pass like most fads whose foundations are flawed?

"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up familiar with it…" [15]

Max Planck, German Physicist, 1858-1947

REFERENCES

[1] Cochrane AL, Blythe M (2004). Salonica. In: One Man's Medicine: an autobiography of Professor Archie Cochrane. London: BMJ, 1989, pp 61-72. Reproduced with permission in The James Lind Library (www.jameslindlibrary.org). Accessed Thursday 28 December 2006.
© Max Blythe, 2004.

[2] Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J,Gea-Banacloche J, Keh D, Marshall JC, Parker MM, Ramsay G, Zimmerman JL, Vincent JL, Levy MM; Surviving Sepsis Campaign Management Guidelines Committee. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. Crit Care Med 2004 Mar;32(3):858-73.

[3] Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ 2003 Dec 20;327(7429):1459-61.

[4] Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med 1999 Jun 15;130(12):995-1004.

[5] Krieger BP. When wheezing may not mean asthma. Postgrad Med 2002;112(2):101-11

[6] O'Donnell JL, Smyth D, Frampton C. Prioritizing health-care funding. Intern Med J. 2005 Jul;35(7):409-12.

[7] Neyman J Pearson E. On the problem of the most efficient tests of statistical hypotheses. Philos Trans R Soc Lond A 1933;231:289-337

[8] Statistical Methods and Scientific Inference 3rd Ed NY Macmillian 1973

[9] Steven Goodman. P values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. Am J Epidemiology 1993:137:485-495

[10] Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women:principal results From the Women's Health Initiative randomized controlled trial. JAMA 2002 Jul 17;288(3):321-33.

[11] Beral V, Banks E, Reeves G. Evidence from randomised trials on the long-term effects of hormone replacement therapy. Lancet 2002 Sep 21;360(9337):942-4.

[12] Finfer S, Felton T, Blundell A, Lipman J; ANZICS Clinical Trials Group Sepsis Investigators. Estimate of the number of patients eligible for treatment with drotrecogin alfa (activated) based on differing international indications: post-hoc analysis of an inception cohort study in Australia and New Zealand. Anaesth Intensive Care. 2006 Apr;34(2):184-90

[13] Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, Steingrub JS, Garber GE, Helterbrand JD, Ely EW, Fisher CJ Jr. Recombinant human protein C Worldwide Evaluation in Severe Sepsis (PROWESS) study group. Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 2001 Mar 8;344(10):699-709

[14] Charlton BG, Miles A. The rise and fall of EBM. QJM 1998;91:371-374

[15] http://thinkexist.com/quotation/a_new_scientific_truth_does_not_triumph_by/158371.html accessed 31 December 2006