Medicine, Health Care and Philosophy (2005) 8:255–260
DOI 10.1007/s11019-004-7345-8

Scientific Contribution

The challenges of evidence-based medicine: A philosophical perspective

Abhaya V. Kulkarni
Departments of Surgery and Population Health Sciences, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada M5G 1X8 (E-mail: abhaya.kulkarni@sickkids.ca)

Abstract. Although evidence-based medicine (EBM) has gained prominence in current medical practice and research, it has also had to deal with a number of problems and inconsistencies. For example, how do clinicians reconcile discordant results of randomized trials or how do they apply results of randomized trials to individual patients? In an attempt to examine such problems in a structured way, this essay describes EBM within a philosophical framework of science. Using this approach, some of the problems and challenges faced by EBM can be explained at a more fundamental level. As well as by employing a similar description of the competing *alternative research tradition of clinical medicine*, this essay not only highlights the philosophical differences between these two modes of medical practice, but suggests that they, in fact, play a *de facto* complementary role in current clinical medicine.

Key words: evidence-based medicine, philosophy of science, research tradition

Introduction

In recent years, evidence-based medicine (EBM) has evolved as the "new paradigm" for medical practice and research (Evidence-based Medicine Working Group, 1992). The rise of EBM has been described as replacing the "*former paradigm*" of clinical medicine, which was built on the foundation of clinical experience and the application of basic pathophysiological mechanisms to clinical decision-making (Evidence-based Medicine Working Group, 1992). Evidence-based medicine, on the other hand, is supposed to rely on systematic, reproducible, unbiased observation and acknowledges that basic pathophysiological principles and expert opinion are insufficient, and, at times, inaccurate, in guiding clinical decisions (Evidence-based Medicine Working Group, 1992). The EBM Working Group singled out the two vanguards of EBM: randomized controlled trials (RCT) and meta-analyses (MA) (Evidence-based Medicine Working Group, 1992). At the outset, it is important to acknowledge that while the term EBM is relatively new, some of its fundamental concepts took root centuries ago. For example, in 1835, Pierre Louis' *numerical method* (la méthode numérique) was based on assessing outcomes in many patients and subjecting them to statistical analysis, rather than just describing individual cases (Freedman, 1999). This was a radical concept at the time, but now one that has become quite familiar under the aegis of EBM.

However, questions and problems with EBM have been raised (Charlton and Miles, 1998; Editorial, 1995). For example, how do we apply the results of a RCT to an individual patient? How do we reconcile the results of discordant RCT's? Is the *former paradigm* (which we will henceforth term the *alternative research tradition of clinical medicine* (ARTCM)) completely incompatible with the principles of EBM? These are difficult questions and it is perhaps useful to begin by thinking about EBM and the ARTCM in a more fundamental, philosophical context.

For example, the term EBM seems to imply that all other forms of medicine are, in fact, not based on evidence. However, the true distinction between EBM and the ARTCM is based on a fundamental difference in their philosophical assumptions about what things in clinical medicine are able to be studied (*ontology*) and how clinical medicine can and should be studied (*epistemology* and *methodology*). Recognizing this, it is fairer to say that the distinction between EBM and the ARTCM lies in what each considers *acceptable* evidence.

This essay will examine EBM and the ARTCM within Laudan's philosophical model of science (Laudan, 1977). The major theories and the

empirical and conceptual problems associated with EBM will be discussed within this philosophical context. The essay will conclude with what seems to be the *de facto* complementary role of EBM and the ARTCM in current medical practice.

Philosophical models of science

Over the years, many academics have attempted to construct a philosophical framework within which to describe science and its progress. Some of the more eminent thinkers who have participated in this ongoing dialogue include Karl Popper, Imre Lakatos, Thomas Kuhn, and Larry Laudan. (Chalmers, 1999; Jarvie, 1998; Kuhn, 1970; Lakatos, 1978; Laudan, 1977; Maher, 1998) For many centuries scientific knowledge was thought to accumulate through induction: many observations leading to theory (Chalmers, 1999; Maher, 1998). Karl Popper rejected this, citing Hume's criticism that induction is logically invalid. (Chalmers, 1999; Jarvie, 1998) He instead defined science and its progress through the *falsification* of individual theories that could never be proved to be true, but could, with a single observation, be shown to be false. Thomas Kuhn recognized that this was an inadequate model for scientific progress since many theories throughout the history of science had initially been apparently falsified, yet were not rejected outright (Kuhn, 1970). Kuhn described science in terms of *scientific paradigms* within which scientists work and solve problems. However, once a paradigm meets with unavoidable and substantial anomalies (*crises*), it is abandoned for another paradigm (*paradigm shift*). Imre Lakatos, a pupil of Popper and a contemporary of Kuhn, rejected both philosophers in favor of his own model of science in which scientists worked within *research programmes* (Lakatos, 1978). These consisted of hard core central theories and several more minor theories that constituted the protective belt. Attempts to discredit the central theory were usually absorbed by the protective belt theories, which could be revised.

Aside from these different philosophical models of science, the one that is perhaps most suited for this discussion is Laudan's model of the *research tradition*. A *research tradition* is "a set of general assumptions about the entities and processes in a domain of study, and about the appropriate methods to be used for investigating the problems and constructing the theories in that domain" (Laudan, 1977). Research traditions have numerous specific theories and they hold certain philosophical assumptions about the nature of reality (*ontology*) and the ways in which that reality can be studied (*epistemology* and *methodology*). Research traditions also face both *empirical problems* (questions about the objects of study) and *conceptual problems* (inconsistencies between theories in a research tradition or incompatibility with a widely held worldview) (Laudan, 1977).

The remainder of this essay will focus on a discussion of research traditions. It is for this reason that the term *former paradigm* was abandoned at the beginning of this essay since it is misleading in the context of the current discussion: the term makes explicit reference to the Kuhnian model of science (which will not be appealed to in this work) and implies that a paradigm shift as occurred (which is not the thesis of this paper). Therefore, this entity will be referred to in this essay as the ARTCM.

The alternative research tradition of clinical medicine

The ARTCM assumes an objective, knowable reality. The acceptable ways in which to study this reality include vast clinical experience and basic science experimentation. In this research tradition, the unit of observation is the individual patient, attempting to take into account all of his/her unique features. It is also assumed that the basic mechanisms of disease can be known, such that the results of non-human experimentation can be applied confidently to clinical situations. The highest source of knowledge in this research tradition is experienced, scientific authority, manifested as the internationally recognized expert.

Problem solving within the ARTCM research tradition is accomplished by extrapolating from several particular cases to the general case. This process of using several, finite observations in drawing conclusions regarding that which has not yet been observed relies on *inductive logic* (Murphy, 1997). The more similar observations that are made, i.e., the greater the clinical experience, the more confidence one has that it will be similarly repeated in the next case.

The research tradition of EBM

Evidence-based medicine assumes an objective reality that is approximately knowable and is based on statistical probability. The acceptable methods of studying this reality are through systematic, unbiased observation, exemplified by the RCT. Bias and the

potential for interaction between observer and subject are well recognized and attempts are made to avoid this. The unit of observation is not the individual patient (except in the case of n-of-1 trials – to be discussed later), but, rather, groups of patients.

When solving problems within the EBM research tradition, the basic mode of investigation involves hypothesis testing. It is tempting to describe this as a form of *deductive logic* (Murphy, 1997), since one begins with a null hypothesis which, if true, would lead to certain expected observations. If those observations are not found, one would conclude, at a given level of probability, that the null hypothesis is not true. There are two problems with this interpretation. First, the logic of hypothesis testing in RCT's has its basis in *inductive logic* since the hypotheses one begins with are not axioms (as strictly required for true deductive logic), but are themselves based on previous experience and expectation. Second, the statistical *p*-values that are calculated at the end of a RCT actually represent the *probability of the data* given the (null) hypothesis, whereas our real interest lies in knowing the *probability of the hypothesis* given the data (Diamond and Forrester, 1983; Murphy, 1997). From the perspective of *Bayesian logic*, this is a very important, albeit subtle, difference.

Major theories of the EBM research tradition

The research tradition of EBM contains many theories, but there are a few that stand out as the most prominent:

The RCT provides the best evidence for making clinical decisions

The RCT represents the current state-of-the-art application of the guiding methodological principles of EBM: systematic, unbiased observation in a hypothesis-testing format. The process of randomization allows, hopefully, for a balance in the known and *unknown* prognostic factors and also validates the use of certain statistical techniques (Armitage, 1982; Jadad, 1998).

Biological rationale and expert opinion are an insufficient basis for making clinical decisions

Within the EBM research tradition, biological rationale and expert opinion can only provide the starting point for clinical inquiry as to what may, potentially, be efficacious. The final arbiter of the true efficacy is the application of EBM methodology in the form of RCT's and MA's. Biological rationale and expert opinion might prove useful when proper evidence (RCT or MA) is lacking or when it cannot be feasibly obtained for a particular medical problem (Evidence-based Medicine Working Group, 1992, Jadad, 1998).

The application of EBM principles results in better patient care

It is perhaps a little ironic that the grandest theory in the EBM research tradition is also the one with the least evidence to back it up (Evidence-based Medicine Working Group, 1992; Sackett et al., 2000). Regardless, this is a strongly held belief amongst the EBM community and one of the reasons that this research tradition has gained such prominence.

Problems faced by the EBM research tradition

As with any research tradition, there are several problems, both empirical and conceptual, that are faced by EBM. The degree to which EBM has provided solutions to some problems has resulted in its early success and its current position of dominance in medical research. However, the degree to which it will be able to provide solutions to the unsolved problems will determine its progress, if any. What follows is a partial list of solved and unsolved problems.

Empirical problem: How does one resolve the problems of bias and chance error?

Bias is "any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth" (Last, 1995). The *epistemology* of the EBM research tradition explicitly recognizes bias in clinical research and provides the methodology required to minimize it. One example of this is the emphasis on randomization techniques with concealment of allocation and observer blinding.

Similarly, for chance error, while the research tradition has not found a way of eliminating it, it has provided the means of quantifying and minimizing it. This is best exemplified by the use of statistical testing with explicit *p*-values (to examine false positive error) and sample size and power estimations (to examine false negative error).

Empirical problem: How does one resolve the issues upon which clinicians differ in their opinion?

The EBM research tradition has provided the methodology and "rules" that can be used, in theory, to resolve issues of dispute among clinicians. By performing systematic, unbiased observation in the hypothesis-testing format of the RCT an "even playing field", free of the opinions of rival clinicians, can be created. While this has been successful in many areas of medicine (McIntosh, 1991), the solution is not always so clear. This has led to some of the currently unsolved problems that will be discussed.

Empirical problem: How does one resolve the issues upon which several RCT's and MA's differ?

Randomized controlled trials of the same intervention may produce results that differ in the degree or direction of treatment effect. As part of the solution, EBM spawned an area of research in which the object of study was not patients, but, rather, the RCT's themselves (Jadad, 1998). This led to the recognition of varying degrees of bias in the conduct and reporting of RCT's and the association between lower quality RCT's and exaggerated treatment effects (Assendelft et al., 1995; Jadad, 1998; Jadad and Rennie, 1998; Khan et al., 1996; Kunz and Oxman, 1998; Schulz et al., 1995). The methods used to investigate RCT bias itself have relied on the very same methodology provided by EBM in the form of systematic, unbiased observation. Here we see the EBM research tradition in its heuristic role, attempting to provide clues to solve its own empirical problems.

By introducing statistical techniques for pooling results and incorporating factors such as the quality of the primary RCT's, MA looked to be a promising means of reconciling discordant RCT results (Cook et al., 1995; Dickersin and Berlin, 1992). However, another problem developed: how do you reconcile conflicting results from large RCT's versus MA of smaller studies? This is a source of continuing debate with no clear resolution in sight (Cappelleri et al., 1996; Chalmers, 1991; LeLorier et al., 1997; Villar et al., 1995).

Empirical problem: How does one apply the results from a group of patients to a single individual?

The RCT considers and analyzes patients in groups: these are the units of observation. It is, however, difficult to apply these RCT results to clinical practice, in which the unit of observation and the unit of treatment is the individual patient (Horwitz, 1995). The EBM research tradition suggests that clinicians reduce that patient to a set of prognostic factors in order to make a comparison with the groups of patients treated in RCT's. This is, essentially, a means to harmonize the different methodologies. However, one still cannot know that a given treatment will be beneficial for *that particular* patient (Guyatt et al., 1986).

The EBM research tradition, again in its heuristic role, has attempted to come up with a solution, albeit a partial one, to this problem: the n-of-1 trial. This applies all the fundamental principles of EBM into the systematic, unbiased study of a single patient to determine the best treatment for that individual (Guyatt et al., 1986, 1990). Unfortunately, the actual clinical scenarios to which this approach can be applied are limited, leaving a vast area of clinical medicine without a solution from the EBM research tradition (Jadad, 1998).

Conceptual problem: How does one determine when further evidence should be sought?

For a given hypothetical medical intervention, there may be a number of RCT's that have demonstrated varying results. If, at this point, a minority of physicians are convinced of the efficacy of the intervention, can the medical problem be considered solved? Or, must the research continue until a larger majority of physicians are convinced of its efficacy? Or, perhaps, until a group of internationally recognized experts are convinced? The problem is that, despite the high premium given to systematic, unbiased observation, a medical problem within the EBM research tradition still relies on majority, expert opinion to be considered, finally, "solved". Furthermore, the logic involved in this process is inductive: the greater the accumulated evidence, the greater the confidence in the conclusion. The results of the most recent evidence need to be interpreted in light of the prior probability (Bayesian logic). These elements – the reliance on expert opinion and the use of inductive logic – represent some of the basic characteristics of the ARTCM. This, then, is the conceptual problem and internal conflict for EBM: a research tradition that is built on the premise that expert opinion is *not* the best source for solving medical problems relies on this very same expert opinion in the final stages of problem-solving. This conceptual problem is partly tied in

to the unsolved empirical problems of reconciling inconsistent results from RCT's and MA's. Until those empirical problems are solved, the EBM research tradition will still need to rely on expert opinion. The presence of publications such *ACP Journal Club* and *Evidence-Based Medicine* further suggests that the "evidence" in EBM is not so "self-evident" and does frequently require expert interpretation (Jadad, 1998).

Integration of the research traditions

Some have taken the current prominence of EBM to mean this research tradition has simply replaced one form of authority for another: the clinical epidemiologist now dictates how a patient must be managed, ignoring the individual patient's special circumstances. Sackett et al., however, argue that in practising EBM "external clinical evidence can inform, but never replace, individual clinical expertise" (Sackett, 1997). Straus and Sackett have more formally defined this complementary role of EBM and the ARTCM in an article entitled "Applying evidence to the individual patient" (Straus and Sackett, 1999). In deciding whether to implement a given therapeutic intervention for a certain patient, their approach involves adjusting the results of RCT's and MA's by an "adjustment factor". The adjustment factor is the physician's best estimate of how increased or decreased the patient's absolute risk reduction is compared to those studied in the RCT's and MA's. A similar adjustment factor is used to adjust the risk of adverse events. These adjustments can be viewed as an application of Bayesian logic: the statistical results of the studies are being interpreted in light of certain prior probabilities regarding the expected outcome for this particular patient. These are then combined with the patient's own relative preferences. While the RCT's and MA's clearly come from the realm of EBM, it must be assumed that the adjustment factors come from the physician's own clinical expertise. And, to be strict, the use of clinical expertise falls within the realm of the ARTCM, since it depends on the physician's personal experience and their interaction with the patient. Personal experience and patient interaction are by no means systematic, reproducible, or unbiased. As well, it is important to recognize the individual patient's autonomy and his/her right to choose. Their prior knowledge and experiences, their interaction with their physicians, and their own unique set of values and preferences will all play a large role in the decision the patient ultimately makes. One must also recognize that medical decisions are not made in a vacuum and factors outside of the physician and the patient may frequently have impact. For example, the utility of a given intervention, including its cost and convenience can be relevant.

We know also that there are common situations in which the EBM research tradition has no choice but to rely on the ARTCM. This occurs when several RCT's or MA's for a given topic do not agree or when no proper RCT's or MA's exist for a given problem.

Some may have seen EBM as the panacea for medical uncertainty and, to the extent that it has failed to do so for many problems in medicine, this may have led to "paralytic indecisiveness" among clinicians (Naylor, 1995). This has not been the case, however. Elements of the ARTCM remain strong, despite the rise of EBM, and they serve a useful role in filling in the gaps. Some may argue with the philosophical descriptions or the conclusions presented in this essay. It seems, however, that the *de facto* complementary role of these two research traditions represents the current state and foreseeable future of medicine. Comfortingly, this is not a new concept: in paraphrasing William Osler, Naylor wrote, "let us agree that good clinical medicine will always blend the art of uncertainty with the science of probability" (Naylor, 1995). It is clear that this is how EBM must currently be conducted because of the problems outlined in this essay. What is not clear is how these problems will be solved and, if they are, what EBM will look like in the future.

Acknowledgements

Dr. Kulkarni was supported by the Duncan L. Gordon Fellowship from the Hospital for Sick Children Foundation and the Wilder Penfield Clinical Investigation Fellowship from the Congress of Neurological Surgeons.

References

Armitage, P.: 1982, 'The Role of Randomization in Clinical Trials', *Statistics in Medicine* 1, 345–352.

Assendelft, W.J.J., B.W. Koes, P.G. Knipschild and L.M. Bouter: 1995, 'The Relationship Between Methodological Quality and Conclusions in Reviews of Spinal Manipulation', *Journal of the American Medical Association* 274, 1942–1948.

Cappelleri, J.C., J.P.A. Ioannidis, C.H. Schmid, S.D. deFerranti, M. Aubert, T.C. Chalmers and J. Lau: 1996, 'Large Trials vs Meta-Analysis of Smaller Trials. How do their Results Compare?', *Journal of the American Medical Association* 276, 1332–1338.

Chalmers, A.F.: 1999, *What Is This Thing Called Science?* Indianapolis, Indiana: Hackett Publishing Co., Inc.

Chalmers, T.C.: 1991, 'Problems Induced by Meta-Analyses', *Statistics in Medicine* 10, 971–980.

Charlton, B.G. and A. Miles: 1998, 'The Rise and Fall of EBM', *Quarterly Journal of Medicine* 91(5), 371–374.

Cook, D.J., D.L. Sackett and W.O. Spitzer: 1995, 'Methodologic Guidelines for Systematic Reviews of Randomized Control Trials in Health Care from the Potsdam Consultation on Meta-Analysis', *Journal of Clinical Epidemiology* 48, 167–171.

Diamond, G.A. and J.S. Forrester: 1983, 'Clinical Trials and Statistical Verdicts: Probable Grounds for Appeal', *Annals of Internal Medicine* 98, 385–394.

Dickersin, K. and J.A. Berlin: 1992, 'Meta-Analysis: State-Of-The-Science', *Epidemiological Review* 14, 154–176.

Editorial: 1995. 'Evidence-Based Medicine, in Its Place', *Lancet* 346, 785.

Evidenced-Based Medicine Working Group: 1992, 'Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine', *Journal of the American Medical Association* 268, 2420–2425.

Freedman, D: 1999, 'From Association to Causation: Some Remarks on the History of Statistics', *Statistical Science* 14, 243–258.

Guyatt, G, D. Sackett, W. Taylor, J. Chong, R. Roberts and S. Pugsley: 1986, 'Determining Optimal Therapy - Randomized Trials in Individual Patients', *New England Journal of Medicine* 314, 889–892.

Guyatt, G.H., J.L. Keller, R. Jaeschke, D. Rosenbloom, J.D. Adachi and M.T. Newhouse: 1990. 'The n-of-1 Randomized Controlled Trial: Clinical Usefulness', *Annals of Internal Medicine* 112, 293–299.

Horwitz, R.I.: 1995. ''Large-Scale Randomized Evidence: Large, Simple Trials and Overviews of Trials': Discussion. A Clinician's Perspective on Meta-Analyses', *Journal of Clinical Epidemiology* 48, 41–44.

Jadad, A.R.: 1998, *Randomised Controlled Trials*. London: BMJ Books.

Jadad, A.R. and D. Rennie: 1998, 'The Randomized Controlled Trial Gets a Middle-Aged Checkup', *Journal of the American Medical Association* 279, 319–320.

Jarvie, I.C.: 1998, 'Karl Raimund Popper', in: E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. New York: Routledge.

Khan, K.S., S. Daya and A.R. Jadad: 1996, 'The Importance of Quality of Primary Studies in Producing Unbiased Systematic Reviews', *Archives of Internal Medicine* 156, 661–666.

Kuhn, T.S.: 1970, *The Structure of Scientific Revolutions,* (2nd ed). Chicago: University of Chicago Press.

Kunz, R. and A.D. Oxman: 1998, 'The Unpredictability Paradox: Review of Empirical Comparisons of Randomised and Non-Randomised Clinical Trials', *British Medical Journal* 317, 1185–1190.

Lakatos, I. 1978, *The Methodology of Scientific Research Programmes. Philosophical Papers*. New York: Cambridge University Press.

Last, J.M.: 1995, *A Dictionary of Epidemiology*. New York: Oxford University Press.

Laudan, L.: 1977, *Progress and Its Problems. Toward a Theory of Scientific Growth*. London: University of California Press, Ltd.

LeLorier, J., G. Gregoire, A. Benhaddad, J. Lapierre and F. Derderian: 1997, 'Discrepancies Between Meta-Analyses and Subsequent Large Randomized, Controlled Trials', *New England Journal of Medicine* 337, 536–542.

Maher, P.: 1998, 'Inductive Inference,' in: E. Craig (ed)., *Routledge Encyclopedia of Philosophy*. New York: Routledge.

McIntosh, H.: 1991, '1971–1991: Clinical Trials Evolve to Become Rigorous Scientific Testing Ground', *Journal of the National Cancer Institute* 83 (10), 668–670.

Murphy, E.A.: 1997, *The Logic of Medicine* (2nd ed.). Baltimore: The Johns Hopkins University Press.

Naylor, C.D.: 1995, 'Grey Zones of Clinical Practice: Some Limits to Evidence-Based Medicine', *Lancet* 345, 840–842.

Sackett, D.L.: 1997, *Evidence-Based Medicine: How to practice and teach EBM*. New York: Churchill Livingstone.

Sackett, D.L., S.E. Straus, W.S. Richardson, W. Rosenberg and R.B. Haynes: 2000, *Evidence-Based Medicine: How to Practice and Teach EBM*. New York: Churchill Livingstone.

Schulz, K.F., I. Chalmers, R.J. Hayes and D.G. Altman: 1995. 'Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effect in Controlled Clinical Trials', *Journal of the American Medical Association* 273, 408–412.

Straus, S.E. and D.L. Sackett: 1999, 'Applying Evidence to the Individual Patient', *Annals of Oncology* 10(1), 29–32.

Villar, J., G. Carroli and J.M. Belizian: 1995, 'Predictive Ability of Meta-Analyses of Randomised Controlled Trials', *Lancet* 345, 772–776.