Open

Original Article

The judgement process in evidence-based medicine and health technology assessment

Michael P. Kelly[a,b,*] and Tessa A. Moore[c]

[a]The National Institute for Health and Clinical Excellence, (NICE), Mid City Place, 71, High Holborn, London, WC1V 6NA, UK.
[b]General Practice and Primary Care Research Unit, Institute of Public Health, University of Cambridge.
E-mail: mike.kelly@nice.org.uk
[c]Head of School Effectiveness, Education, Children's and Cultural Services, London Borough of Richmond upon Thames.

*Corresponding author.

Abstract This article describes the judgements used to interpret evidence in evidence-based medicine (EBM) and health technology assessment (HTA). It outlines the methods and processes of EBM and HTA. Respectively, EBM and HTA are approaches to medical clinical decision making and efficient allocation of scarce health resources. At the heart of both is a concern to review and synthesise evidence, especially evidence derived from randomised controlled trials (RCTs) of clinical effectiveness. The driver of the approach of both is a desire to eliminate, or at least reduce, bias. The hierarchy of evidence, which is used as an indicator of the likelihood of bias, features heavily in the process and methods of EBM and HTA. The epistemological underpinnings of EBM and HTA are explored with particular reference to the distinction between rationalism and empiricism, developed by the philosopher David Hume and elaborated by Immanuel Kant in the *Critique of Pure Reason*. The importance of Humian and Kantian principles for understanding the projects of EBM and HTA is considered and the ways in which decisions are made in both, within a judgemental framework originally outlined by Kant, are explored.
Social Theory & Health (2012) **10,** 1–19. doi:10.1057/sth.2011.21;
published online 14 December 2011

Keywords: evidence-based medicine; health technology assessment; hierarchy of evidence; randomised controlled trials; internal validity; external validity

Introduction

This article considers the way in which judgements are used to interpret evidence and make decisions in evidence-based medicine (EBM) and health technology assessment (HTA). Specifically, it will be shown that the judgements in both EBM and HTA may be understood with reference to the ideas of rationalism and empiricism, as outlined in the writings of Hume (1748) and subsequently elaborated by Kant (1781). This essay develops this theme by describing the principles of EBM and HTA. It goes on to examine the problem of uncertainty associated with bias that EBM and HTA seek to minimise and explores the ways in which the ideas of empiricism and rationalism are integral to the way that EBM and HTA operate. The significance of the concepts of internal and external validity is examined and the role of deduction and induction described. The article goes on to suggest that the ideas of Hume and Kant can help explain the interpretation and decision-making process in both EBM and HTA, as well as illuminating some of the tensions that arise in doing EBM and HTA.

An Outline of EBM and HTA

EBM is conventionally defined as '… the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients' (Sackett *et al*, 1996, p. 71). The underlying principle of EBM is to use the evidence that is the least likely to be biased as the basis for clinical decision making. The rationale behind this is that medical interventions are inherently risky to patients and very costly for whoever has to pay for them. EBM therefore starts from the premise that the platform for the practice of medicine should be the best available (meaning least biased and therefore most trustworthy) evidence. This, it is argued, offers the patient protection from medical incompetence, from other risks and from the inflated claims of drug effectiveness made by a profit-driven pharmaceutical industry (Greenhalgh, 2001).

HTA is the first cousin of EBM. Its origins lie in the escalating costs of pharmaceuticals across the globe and the need to find transparent, fair and scientifically robust ways of determining whether new drugs are effective and cost effective (Sorensen *et al*, 2010). It has its own Annual International Conference. There are many agencies around the world that sit between the pharmaceutical industry and the payers, be they governments or insurance companies. These agencies judge the clinical and cost effectiveness of these new technologies. HTA essentially uses the same approach as EBM, but with a very strong health economics component, usually cost utility analysis, as part of the decision-making process. The most well-known EBM/HTA agency in the UK

is The National Institute for Health and Clinical Excellence (NICE) (Kelly *et al*, 2010).

Over the years, there have been some important contributors to the ideas of EBM and HTA. One of the most interesting and influential is Archie Cochrane, a chest physician, fighter against Franco's forces in the Spanish Civil War, prisoner of the Germans after the Battle of Crete in 1941, supporter in the early days of the British Sociological Association's Medical Sociology Group, first President of the Faculty of Public Health and all round iconoclast. In his book *Effectiveness and Efficiency: Random Reflections on Health Services*, he asked a set of simple questions (Cochrane, 1972). Do we know whether intervention *x* for condition *y* is effective? How do we know it is effective? How do we know whether it is more or less effective than intervention *z*? On what basis do we make that judgement of effectiveness? Do we know what it costs? Is it cost effective? If it is not cost effective, why is it still being used? What are the dangers posed to patients of treatments about which we are scientifically uncertain? Are the treatments dangerous? Why are we using potentially dangerous or worthless treatments?

His suggestion that the randomised controlled trial (RCT) should be the starting point for answering these questions, although now part of the conventional evidence-based wisdom, did not immediately find favour. In 1972, when *Effectiveness and Efficiency* was published, the RCT was still then the exception rather than the rule, as the basis for medical decision making and health economics did not form any component of resource allocation in British health services. Cochrane's argument that some kind of cost effectiveness consideration should be a part of clinical decision making was initially taken seriously by few commentators. He was highly critical of the waste he observed in the use of unproven technologies and, although the kind of economics he advocated in *Effectiveness and Efficiency* now seems rather crude, he had in essence asked the question that few had had the good sense to ask before. He made common cause with Alan Williams, professor of economics at the University of York, who did much to develop the principles of cost utility analysis as a means of understanding and doing resource allocation in health care better. Williams' work and that of his many colleagues and students at the University of York laid the foundations of modern British health economics and the York group still play a very important role in HTA internationally (Culyer *et al*, 1972; Drummond *et al*, 2005, 2007; Lomas *et al*, 2005; Culyer, 2006).

Cochrane was right about the RCT. It does offer the best assessment of the question of clinical effectiveness. To determine the effectiveness of a drug therapy and some other types of intervention, the RCT provides more certainty than any other method because it maximises internal validity by reducing bias (Campbell and Stanley, 1966). It is a method that is designed to allow the

reasonable conclusion that the effect that is being witnessed is the consequence of the intervention. By controlling out of the process factors that could contaminate the relationship between the independent and dependent variable, the observer has as much certainty as possible that the relationship is real rather than an artefact of the research process or some other variables confounding the relationship. It also allows degrees of effect or effect size to be estimated. Trials of effectiveness form the platform for the subsequent economic analysis of cost effectiveness. The results of trials are never absolutely certain, however, because even the best trials are flawed in various ways. In addition, the size and time scale of most trials mean that they cannot detect all possible adverse reactions, and as more and more observations accumulate original observations may be modified (Rawlins, 2008). Moreover, trials are not always the appropriate method to investigate the question (Black, 1996; Glasziou *et al*, 2007; Vandenbroucke, 2008). However, for the purposes of assessing *clinical* effectiveness, the double-blind RCT sets a threshold that no other method has yet surpassed.

Therefore, the RCT is the starting point methodologically for EBM and HTA. However, EBM and HTA are much more than just RCTs. As the evidence-based approach has evolved, a number of practices have developed. In schematic terms, these involve a sequence of activities. The sequence depends on the pre-existence of a body of primary research, some of which will be RCTs. Then, and the first stage in the sequence, a clinical problem is expressed as a question, for example – what is the comparative effectiveness of different treatments for a particular disease? It is usual to use the so-called PICO framework to do this. This involves specifying the *population*, the *intervention*, the *comparator* and the *outcome*.

Second, the potential answer to the question is sought from the peer-reviewed scientific literature – not opinion or current practice, but empirical evidence that has been subject to the detailed scrutiny of peer review. Modern search techniques and search engines, as well as comprehensive if not exhaustive databases, allow the world literature to be searched for the evidence and also help to preclude the possibility that findings will be out of date or US or UK centric. The literature searching is done systematically using highly sophisticated computer-based search protocols according to clear and explicit inclusion and exclusion criteria. The process is replicable, transparent and auditable (Egger *et al*, 2001; Greenhalgh, 2001; Centre for Reviews and Dissemination, 2008).

Third, the evidence that is found in this way is then subject to critical appraisal using well-defined methods designed to weed out bias. The focus is on internal validity. Studies are ranked in the hierarchy of evidence. The hierarchy of evidence represents levels of types of evidence where internal validity is improved at each succeeding step up the hierarchy. With each step up the

hierarchy, the chances of bias are lessened. RCTs score highly because their *raison d'êtr*e is the controlling out of factors, which can cause bias.

Fourth, the studies are then assessed for quality within the levels of the hierarchy (Egger *et al*, 2001; Greenhalgh, 2001). Sometimes, the data extracted from the studies ranked highest (good RCTs) are synthesised in a procedure called meta-analysis. Meta-analysis involves summing the results of trials together in order to determine the overall result of the accumulated evidence. Even without formal meta-analysis, there are a number of techniques that allow comprehensive accounts of the accumulated evidence to be provided (NICE, 2009). Following this, it is a relatively straightforward matter, at least in principle, to determine the answer to the original comparative effectiveness question. Finally, all of this can be assessed for its cost effectiveness using the principles of cost utility analysis.

The practitioners of the evidence-based approach usually take it as self-evident that EBM is superior to other forms of medical decision making and other forms of resource allocation because of the scientific and rigorous nature of their practice. The practitioners of EBM tend to start with a very practical and praiseworthy precept, which is that the recipients of medical treatments deserve the best and safest treatments that are available. These can be determined using scientific methods to search, synthesise, appraise and analyse the cumulative evidence. Further, they acknowledge that until relatively recently the degree of variation in medical practice and wide differences in the drug prescribing behaviour of doctors were not grounded in science. Occupational practices were embedded in learning in medical school, the dominance of influential medical professors and other medical leaders and the influence of the pharmaceutical industry. In addition, as the growth in the volume of scientific medical information has accelerated exponentially since the 1960s, doctors simply could not keep up with the amount of new information. Even the most conscientious and assiduous physician or decision maker could never hope to keep pace with the growing amount of evidence. The systematic review and synthesis of the best available evidence offers a means of looking at accumulated findings from multiple investigations and, in part at least, solves the problem of information overload.

HTA is based on the principle that markets alone are not the optimal means of resource allocation in health care and that the principles of fairness and transparency should be paramount. Cost utility analysis provides a method that embraces costs through the concept of the Quality Adjusted Life Year. This combines a measure of quality of life with an estimate of what society is willing to pay for treatments, to produce an assessment of how to allocate resources fairly. This idea goes beyond a simple measure of costs and allows comparisons to be made between different treatments. The process acknowledges that there

is always a limited budget to pay for health care and that demand for scarce health care resources outstrips the ability of individuals or society to pay for all possible treatments. It also embraces the idea that the improvements whether measured in the direct effect of the drug or the patient's quality of life is not uniform and that, therefore, some means of comparison between different treatments is required. The cost utility measure provides a mechanism to weave a way through this complexity (Drummond *et al*, 2005).

The guideline movement has sprung up in parallel with EBM and HTA. A guideline is an evidence-based protocol, providing doctors and others involved in the care of patients with descriptions of optimal algorithms for pathways of care and interventions. A guideline is based on cumulative findings from systematic reviews. A guideline will in theory encapsulate the most up-to-date and best evidence and help eliminate the variation in medical practice, which characterised earlier generations. The Cochrane Collaboration, a worldwide network of systematic reviewers producing up-to-date reviews of many hundreds of different aspects of medical care and interventions, is a major and easily accessible resource to underpin those doing EBM and HTA. In the UK, the Centre for Reviews and Dissemination at the University of York is a major UK resource underpinning EBM, HTA and guidelines.

EBM has undoubtedly brought improvement to a range of patients' experiences and reduced some of the risks associated with medical care. It has also put the funding of new technologies on a fairer and more transparent basis where HTA is used properly. For the most part, the proponents of the EBM approach, the vast majority of whom are trained initially in biomedicine, statistics or economics, do not tarry with the theoretical niceties with which this article is concerned. Instead, the refinement of the methodology of trial design, processes of literature searching and development of clinical guidelines have been central to the methodological concerns of EBM. The logic of the scientific methods used is assumed to be optimal, to be linear and to be progressive. There is no sense at all of science as a social or historical product and of ways of knowing and understanding as sociologically or historically grounded (Pickstone, 2007). Therefore, underlying sociological, epistemological and philosophical questions remain little discussed. However, there are some extremely interesting and important issues to which the whole business gives rise. This article explores one of these – the basis of the interpretations and judgements involved.

Uncertainty

EBM and HTA depend on the principle of evidence accumulation. In other words, they make a virtue out of the exponential growth in medical knowledge.

The idea is that the greater the number of observations, the greater the degree of accuracy about that which is being observed and the greater the chance of the elimination of uncertainty. Further, if multiple studies with multiple results are pooled, then there is an even better chance of the results being averaged out in an optimally accurate way. This is a way of dealing with uncertainty. It recognises explicitly that single observations may be unreliable and that multiple observations offer protection against outliers. It also presumes that repeated observations will in the long run be more likely to be a truer representation of the outcomes of interest, and those initial and single observations are a poor basis for inductive reasoning.

The question of uncertainty has several dimensions. Underlying the reason why observations vary in drug trials is individual human biological and genetic variation. Each time a drug is administered to an individual patient, it produces biochemical changes in the human body. Some of these may be efficacious in terms of the desired outcome – some will not be. The human body is a dynamic biological system, and different human bodies vary in the way in which they respond to the administration of any pharmaceutical agent. This is by virtue of metabolism, nutritional status, age, genetics, intercurrent disease, other drugs being used and the effects of the lifecourse, as well as the biochemical interaction. There will, therefore, always be a range of ways in which the human organism responds. This is why a drug that has a therapeutic benefit in one patient will not necessarily always work in the same way in all patients. The purpose of the clinical trial is to determine whether, on average, it will work for most or for some subset in the population. The idea of biological variation is intrinsic to the business of doing medical research and indeed medicine more generally. The concepts of efficacy and effectiveness capture this idea. Efficacy deals with the question of whether or not a particular intervention works under highly controlled laboratory or laboratory-like conditions. Effectiveness is concerned with the degree to which it works on average in real-world settings in clinical practice.

A trial is just one aggregated average overall observation and looking at the results of one trial may mislead. The phenomena that early trials tend to achieve better results than later ones, that positive findings tend to get reported more readily than negative ones in the scientific literature and that all sorts of Hawthorne, placebo and reactivity effects will be at work (Blalock, 1970; Orne, 1970) mean that it is now considered much wiser, if possible, to look to the multiple trials of the same agent to see what on average the result is of all the trials, the *sine qua non* of EBM. This narrows down the possibility that one early trial result will be an outlier. And, obviously, within the method, great pains are taken to avoid placebo and Hawthorne effects by blinding the intervention and control subjects and their doctors as to which are taking the active drug.

The Problem of Observation

However, the question of uncertainty goes still further than biological variation, to the process of observation itself. This problem was formulated more than 250 years ago in Enlightenment philosophy most clearly by David Hume (Hume, 1748[1]) in Scotland, and later by Immanuel Kant (Kant, 1787[2]) in Prussia. Both were influenced by the then emerging science of optics. Hume and Kant argued that there is a distinction to be made between the observation and the thing that is being observed. What we observe – the observation – is a representation in our mind, which comes to us via our senses and any scientific instrument we are using, of the real thing that is being observed. Applied to any process of scientific observation, the distinction made by Hume and Kant indicates that there is an underlying reality, but our ability to perceive or observe it is always partial. All scientific data are but representations of reality. Our observations must not be confused with reality. Further, this observational process produces distortion and bias because we are fallible observers and the tools we use to do the observation, whether our own senses or some scientific instrument, will be limited by our human capacity to observe and interpret and the technology we are using. This applies equally well to a doctor listening to a heartbeat through a stethoscope and inferring heart disease, to the results of a trial, to the observation of a specimen under a microscope or to perceiving a landscape in the distance. In the case of a clinical trial of a drug, the drug has a real biochemical reaction and there will be average overall effects in population groups in the trial, but the data from the trial are ways of representing that reality in two-dimensional space – on screen or on paper. That representation will in turn require interpretation: a process of inferring what the underlying biological reality is on the basis of the two-dimensional representation of it – on paper or on screen.

Hume (1748, p. 79), despite being an avowed empiricist, argued that experience and observation were fallible guides to the real world. In other words, empirical observation is subject to distortion. We needed, he suggested, to proportion our belief in the evidence before us – our observations – on the basis of past observations, the weight of evidence, probability (what he rather confusingly called moral reasoning) and the similarity of observations to previous observations (1748: 80). For Hume, there was always some process going on of what today we would call social construction. This does not mean that the world around us is illusory. It means that our ability to observe real things is always partial. Kant embraced the same idea. In Kant's terms, we can only perceive things as they appear to be – what he called *phenomenon* – rather than as they really are – what he called *noumenon*.

This fundamental position is simultaneously central to EBM and ignored by it. It is central because EBM acknowledges that single observations are

potentially unreliable and that the best (but not necessarily final) account of the relationship between the drug and the outcome is obtained by multiple observations. This is firmly in the Humian/Kantian tradition. However, EBM seldom goes the stage further and questions the reality of the evidence itself. It tends to treat the evidence not as a proxy for reality, but as if it were reality. This is to confuse empirical observation with reality itself in what appears to be a crude form of positivism. All science can ever realistically do empirically is describe the observations made as best and as accurately as possible. The data from a trial are not the real underlying biological process – they are the best available approximation of it. Consequently, not only can bias never be eliminated completely, but it is also a feature intrinsic to all scientific observation. EBM strives to reduce bias *as if* its complete elimination were possible. All observation involves distortion because by its very nature the act of observation changes that which we observe. This, EBM never acknowledges. Instead, the evidence is given an ontological status as something real, not as a representation of the underlying reality. Further, the principles of the hierarchy of evidence designed to eliminate bias are premised on the idea that if only bias can be eliminated, then the truth will emerge. This is, at least in Humian and Kantian terms, impossible.

Interpreting Uncertain Observations

The issues that the Enlightenment philosophers can assist us with, however, go further and actually help to get EBM off the hook of the accusation of naïve positivism and confusing facts and observations. Hume (1748, pp. 17–25) divides reasoning into (i) that concerned with the relations between ideas called demonstrative reasoning or rationalism, and (ii) that concerned with matters of fact called factual reasoning or empiricism. The first is *a priori*, that is, precedes observation. This form of reasoning for Hume includes geometry, arithmetic and algebra. Matters of fact on the other hand are *a posteriori* and derived from evidential observations (1748: 18) usually involving some assessment of cause and effect (1748: 19). These are *a posteriori* – after the fact or after the observation. Demonstrative reasoning is about the relations between ideas (1748: 25). Demonstrative reasoning is deductive. It proceeds with absolute certainty, based as it is on the logical relations between ideas. Factual reasoning is inductive and involves drawing apparently reasonable but not logically certain conclusions based on the available and incomplete evidence (Millican, 2007, p. xxxvii).

The rhetoric of EBM is located firmly in the empiricist, matters of fact camp. Most of the proponents of EBM would see themselves as dealing with matters

of fact or real things. However, there are very strong elements of rationalist thought at the heart of EBM. The principles of the elimination of the possibility of bias in the hierarchy of evidence, of the rule-driven principles of guideline development and appraisal are based on an ideal version of the scientific method, which owe more to the logical precepts of the *a priori* relations of ideas than they do to messy empirical observation. Hume's warnings that induction was probabilistic and that, therefore, interpretative processes were required to understand the data are the underdeveloped component in EBM. Every time an empirical statement is made, a set of rationalist judgements have been used to make sense of the data and to interpret the evidence. Although EBM has been particularly effective at elucidating the empirical end of things, it has been much less reflective about the rationalist processes it uses.

Although Hume and Kant agree on the object–observation distinction and are agreed that knowledge is either rationalist or empiricist, after that they part company. Hume, in the end, argued for the superiority of the empiricist method. The *Enquiry* ends with his famous words. 'If we take in our hand any volume of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number*? No. *Does it contain any experimental reasoning concerning matters of fact and existence*? No. Cast it then into flames. For it can be nothing but sophistry and illusion'.

Kant is less convinced. An admirer of Hume, but he was much less certain that empirical knowledge on its own could be so easily divorced from rationalism or what he called pure reason (Kant, 1787, p. 41). Kant argues that there must be knowledge that is independent of experience, which is not *a posteriori* (Kant, 1787, pp. 42–43) because, without some *a priori* knowledge, we would be unable to make sense of what we observe (Kant, 1787, p. 42). *A priori* knowledge is independent of experience, whereas empirical *a posteriori* knowledge is not (Kant, 1787, p. 43; Weldon, 1958, p. 101). For Kant, experience is neither the product of pure reason nor of empirical induction; it is both. All knowledge begins with experience, but not all knowledge is derived from experience, the bridge is the interpreting observer.

The Types of Judgements Made About Evidence

Kant argues that we need some kind of criteria to distinguish between pure and empirical knowledge (Kant, 1787, p. 43). To explore the relationship between the two forms of knowledge, one based on pure *a priori* reasoning and one based on observation, Kant develops an argument about what he calls judgements or rules (Korner, 1955, p. 104), meaning the way we interpret the objects of our perception. Judgements can be analytic or synthetic (Korner, 1955). An

analytic judgement is one in which the meaning of the predicate is included in the meaning of the subject (Kant, 1787, pp. 45–49; Ward, 2006, p. 16). A simple example of this is 'all bachelors are men'. The same meaning is intrinsic in each term – men and bachelor – because the definition of bachelor means an unmarried man. Analytic judgements like this are *a priori*, and to deny them is self-contradictory (Korner, 1955; Ward, 2006, p. 18).

On the other hand, judgements may be synthetic. In this case, the meaning of the predicate is not contained in the subject (Ward, 2006, p. 16). Two things that are not intrinsic to each other are brought together (Korner, 1955). An example of this would be 'the book is blue'. We are bringing together two ideas, book and blue. They are not intrinsic to each other because books can be of any colour and the colour blue could apply to many different objects (Guyer, 2006). The blue book on my table is one specific object located in space and time, observed empirically. The book is blue is a synthetic judgement because the idea blue is not intrinsic to the idea book (Guyer, 2006, p. 47). Synthetic judgements are *a posteriori* because they are established by recourse to experience (Ward, 2006, p. 18), in this instance of books and of colours. Simple *a priori* knowledge is analytic, and empirical *a posteriori* knowledge is synthetic.

However, Kant suggested that it is more complex than this because some *a priori* judgements are synthetic not analytic (Kant, 1787, p. 51). For Kant, these synthetic *a priori* judgements are the fundamental judgements or precepts of geometry, mathematics, natural science and metaphysics, that is, they are more than just contained in the subject and predicate and cannot be determined on the basis of experience alone (Guyer, 2006, p. 47; Ward, 2006, p. 19). For Kant, the most important ideas of this type relate to space and time (Korner, 1955). *A priori* rationalism provides the conceptual architecture in the form of synthetic *a priori* judgements and is the basis for empirical science.

The apparently simple distinction made by Hume between rationalism and empiricism therefore turns out for Kant to be more complex. For Kant, there were three elements involved in knowing and understanding and used in the process of interpreting sense data or scientific evidence – to recap – (i) analytic *a priori* judgements, (ii) synthetic *a posteriori* judgements and (iii) synthetic *a priori* judgements. These judgements or rules are quite helpful for considering the processes and practices of EBM.[3]

Therefore, following Kant's argument, first there are the analytic *a priori* judgements. These are a very significant part of the EBM armoury. They are the logical and methodological givens. They are self-referential in the sense that by definition they are true and to deny them would be self-contradictory. The most important analytic *a priori* propositions in EBM are those that give rise to the hierarchy of evidence. The hierarchy is the operationalisation of the *a priori* analytic principle, and fundamentally rationalist idea, that there is a true and

real relationship between phenomena and that extraneous or confounding factors mask that true relationship. By reducing bias, we can get closer to that real relationship. The problems described in the previous sections of this article relating to the biases that arise as a consequence of the process of observation, or the difficulties surrounding uncertainty or the fallibility of the human observer, are thought of, in this view, as masking our ability to see the true nature of things. The methodological task following from this is to try to limit the impact of such things in order to reduce uncertainty and bias and to see things as they really are. The concept of efficacy is premised on this precept.

The idea that there are true and real relationships between things, usually defined as the independent and dependent variable, is the central *a priori* analytic judgement in EBM. A number of other analytic *a priori* tools follow from this principle, including the judgements that: RCTs control extraneous or confounding factors in an assessment of an intervention; confidence intervals help distinguish between true effects and chance ones; and that summing results in meta-analyses produces a truer result than a single observation. The hierarchy of evidence is not about the *possibility* of the elimination of bias, but about the actual or real elimination of bias to reveal a pure relationship uncluttered by other things.

Internal validity is intrinsic to the hierarchy. The precept of internal validity is that it is possible to be certain that the action of the independent variable is the reaction in the dependent variable, that the measure of the reaction is true and that if repeated under the same conditions it will produce the same degree of change in the dependent variable. This is an *a priori* position because empirically and after the fact, this can never be demonstrated; it is an ideal position. And even empirically, it can only be demonstrated by controlling for all possible confounders, something that is never attained in real life. Moreover, in medical practice, the way disease presents is not as a single entity but frequently as multiple morbidities, that is, the patient has more than one thing wrong with them. Therefore, the idea of simple causes and effects as implied by the focus on internal validity in EBM is seldom the reality of medical presentations. Nevertheless, methods that are strong on control and on internal validity sit at the top of the evidence hierarchy in EBM. The analytic *a priori* judgement that follows from this is that it is possible to distinguish between types of method on the basis of their ability to eliminate bias. This is an entirely rationalist position that owes little to empirical science because the idea of a real pure relationship between the independent and dependent variable is only possible in the realm of pure reason and is about the relationship between ideas. It is further based ontologically on the idea that single outcomes could have single causes and that other things being equal these could be found and measured precisely. All this reasoning operates in the world of ideas and pure

reason, not empirical science. These principles are the basis for deductive reasoning in EBM.

What the hierarchy of evidence attempts to do is describe a science in which the elimination of bias is a real possibility and to privilege those methods that control the extraneous factors out of the equation. They are givens; they are true by reference to their meaning and to deny them would be self-contradictory. None of this is to deny that bias introduced by the act of observation or fallible human observers should not be minimised, or that science should not strive for accuracy and objectivity. In this regard, the hierarchy serves well. However, its fundamental principle is not about eliminating bias deriving from observation, but bias meaning eliminating distortions in the true relations between phenomena.

Kant's second set of judgements or rules are synthetic *a posteriori* judgements. These are judgements that are made on the basis of observation, can only be made after the fact and are in essence pure evidence, or more precisely the representations of the reality that the evidence describes, or as Kant would put it as they appear to be. Therefore, for example, the statement that comparative effectiveness of compound x over compound y in cases of disease z has n difference in effect size, is judged on the basis of its synthetic *a posteriori* quality. It involves three after-the-fact observations, which are brought together. The observation of what x does, the observation of what y does and the difference between them. This is the core synthetic *a posteriori* matter of fact at the heart of EBM, and indeed is often thought of as the very quintessence of EBM and HTA. These matters of fact can then be the basis of induction. The concept of effectiveness as against efficacy is synthetic and *a posteriori*.

Kant's third set of judgements are the synthetic *a priori* judgements. These are required because, as both Hume and Kant noted, induction is a process involving judgement and interpretation. Inductive reasoning needs more than matters of fact. For Kant, it needs synthetic *a priori* judgements. These are plentiful in EBM. Synthetic *a priori* judgements transcend the empiricist world of synthetic *a posteriori* evidence of effectiveness and the rationalist analytic *a priori* world of the hierarchy of evidence. The synthetic *a priori* judgements bridge the two domains and without them EBM would be impossible.

The first example of synthetic *a priori* approach to illustrate the point is medical diagnosis. Critical to the process are clinical judgement, disease labels and observation. Diagnostic categories or disease labels are essential to EBM and HTA, as a diagnostic category is normally the focus of the efforts to determine clinical and cost effectiveness. Simplistically, it might be imagined that diagnosis involves fitting the observed symptoms to the agreed disease taxonomy. Taxonomies are familiar enough from any medical textbook, based

on a variety of ways of arranging medical knowledge. Attached to disease taxonomies are descriptions of the epidemiology, aetiology (if known), therapeutics and prognostics. Taxonomies change as medical knowledge advances; they are not fixed and immutable (Bell, 2010). The taxonomies are typifications that can only exist in an ideal rationalist sense. This is because the way that pathology manifests itself empirically in the human or in any animal or plant does not follow the strict limits of the disease as described in the taxonomy. The taxonomy is a social product that changes, not a fixed underlying reality.

In any event, there is always variation. In part, this variation is associated with the individual biological and genetic difference referred to above. It also arises because biological and genetic variations interact with different aetiological agents. Empirically, disease is therefore a spectrum of pathology and many patient presentations are a cluster of multiple pathologies and morbidities. These will more or less approximate to the taxonomic ideal (Bell, 2010). Therefore, taxonomies not only change through time, they always have to be used flexibly in diagnosis. Doctoring is not simply about fitting sets of observed symptoms to taxonomies. Clinical judgement involves being flexible with the taxonomies. If all that medicine involved was applying the categories described in the textbook, then anyone who could read and understand the textbook could be a doctor. Clinical training involves learning that the categories are the necessary, but not sufficient condition for diagnosis – clinical judgement is also involved. It is what is *not* in the textbook that is the basis of the practice of clinical medicine. Therefore, the very concept of the identification via diagnosis of a particular disease is synthetic and *a priori* as it rests, and is only possible, on the basis of empirical observation *and* clinical judgement.

The synthetic *a priori* idea of diagnosis involving clinical judgement is a very different type of judgement to applying the mechanics of the hierarchy of evidence or a statistical test. The two ways of thinking sit decidedly uncomfortably together in EBM. The certainties of the analytic *a priori* concepts of confidence intervals, hierarchies of evidence and the elimination of bias are a long way removed from the kinds of subjectivities involved in clinical judgement. Therefore, many of the tensions associated with doing EBM can be seen to revolve around the rationalist–empiricist divide and the contrast between the processes of inductive reasoning associated with clinical activity and the deductive reasoning associated with the mechanics and techniques of EBM and HTA (Barnett *et al*, 2009). Many of the critiques of EBM that arise precisely because of these two ways of thinking – the synthetic *a priori* clinical judgement versus the rule-driven certainties of EBM grounded in analytic *a priori* judgements – are different (Egger *et al*, 2001). Both sides interestingly appeal to the

evidence – the synthetic *a posteriori* concepts – as a rhetorical device to justify their position (see Russell *et al*, 2008). Kant's eighteenth-century epistemology offers an explanation of this very modern problem.

The second key synthetic *a priori* idea is that of modelling, which is also at the heart of the EBM/HTA enterprise and an area where the principles are hotly contested. Economic modelling, which is such a central part of HTA, is founded on the juxtaposition of different concepts; say the amount of health improvement or quality of life gained as a consequence of the administration of such and such a degree of medical intervention. This is the synthetic element. The *a priori* element comes from the association that is known empirically to exist in general terms between these two elements and the ability to predict *a priori* that these elements will be conjoined in the future. Economic modelling is classically synthetic and *a priori*.

The third example of the synthetic *a priori* group of concepts is external validity (Campbell and Stanley, 1966). Unlike internal validity, which is an *a priori* analytic concept, built entirely out of rationalist principles, external validity is classically *a priori* and synthetic. External validity is conventionally defined as dealing with the question of whether the results obtained in one setting would apply in another (Campbell and Stanley, 1966). In EBM and HTA classically, it is about determining whether the findings of one trial are transferable or generalisable more broadly. Statistical generalisability is usually corralled in order to help reach the decision. Applied in EBM, it is about dealing with the question of whether the results of study A will help patient B, and can team C in hospital or primary care setting D implement it in such a way that it will work as in study A. In areas where the judgements involve a long causal chain from the intervention to the outcome as in public health or other social care or educational interventions (Kelly *et al*, 2010; Kelly and Moore, 2010), the problem of external validity is still more vexing (See also Pawson, 2002, 2006). In these cases, we are trying to determine whether the results of study A, which describes a rather loosely defined and often poorly specified intervention B, applied in manner C, in context D, by team E, in organisation F, to sub population G, will produce the same result as it did when it was originally done two decades ago in study A (Moore and Rutherford, 2012). These judgements are quintessentially and inevitably difficult. It is not going to be possible to derive the judgements from better more highly powered studies, covering all sub populations, nor to find all the details of the fidelity of the intervention or the process of implementation (Davidson *et al*, 2003). Therefore, we have to use synthetic *a priori* judgements if we are to work scientifically. In short, external validity is empirical evidence conjoined with theory and is about probabilistic statements in the face of real-world uncertainty; internal validity is about rationalist certainty in the world of ideas.

Conclusion

In some respects, the evidence-based approach may seem at first sight in the critical sociological gaze to be little more than abstracted empiricism, and in many ways it is its own worst enemy being locked into a particular view of itself as the highest form of empirical medical science. Some proponents of EBM do seem to have a very narrow conception of the enterprise in which evidence based means empirically based, with theory and opinion being cast as the lower orders of the hierarchy of evidence. However, thinking about EBM in a framework of some of the ideas of Hume and Kant casts it into a different light. Therefore, intriguingly, the evidence-based approach as applied in EBM and HTA offers an illustration of Kant's fundamental epistemological concepts in action.

Arguably, this observation is of little interest to practitioners of the arts of EBM and HTA, as it is perfectly possible to make an assessment of cost effectiveness of a new pharmacological agent or to quality appraise a series of systematic reviews without ever knowing anything of Kant or Hume. On the other hand, articulating the judgement processes is interesting in itself. However, more importantly, the concepts outlined here can help illuminate some of the real tensions that arise in trying to reach judgements in EBM and HTA, where real uncertainty often surrounds the processes. The usual common sense reaction to uncertainty by the practitioners of the arts of EBM and HTA is to try to work harder and harder, like Boxer the horse in Orwell's *Animal Farm*, to perfect the methods that will eliminate the bias. They seldom recognise the fact that they are straddling a very old intellectual divide, well known to Enlightenment philosophers.

Some agencies with HTA responsibilities have explicitly acknowledged the tensions and sought to move beyond methodological to philosophical solutions. NICE, for example, began quite early on to try to articulate tensions and ways of resolving them in its Social Value Judgements Papers (NICE, 2008). The need for these arose amidst the realisation that evidence alone did not speak for itself, that interpretative processes were involved in making sense of the evidence and that the interpretive processes did not themselves or could not themselves be derived from science *per se*. The current article adds to that thinking by suggesting that certain philosophical ideas can help to articulate or describe the judgemental and interpretive processes involved. What this article has not considered, and which remain important areas for consideration, are the discursive and rhetorical devices used by the actors involved, the group dynamics and the way they overlay the decision-making processes. However, those will be the subject of another article.

Acknowledgement

Thanks to Trisha Greenhalgh for comments on an earlier version of this article.

About the Authors

Mike P. Kelly is Director of the Centre of Public Health Excellence at NICE. He is an Honorary Senior Visiting Fellow in the General Practice and Primary Care Research Unit, Institute of Public Health, University of Cambridge. He also has honorary appointments at the Universities of Manchester and Sheffield. His research interests include evidence-based approaches to health improvement, and methodological problems in public health research. In 2010, he was awarded the Alwyn Smith Prize of the Faculty of Public Health.

Tessa A. Moore has a career spanning over 30 years in education. She taught for Local Authorities in Essex, Shropshire and Telford & Wrekin. Following 10 years of headship, she was seconded to the National College for School Leadership as an Assistant Director of the Primary Leadership Programme. She also worked for 4 years as a Senior Education Adviser and is now Head of School Effectiveness in the London Borough of Richmond upon Thames.

Notes

1 All references to Hume are taken from the 2007 Oxford University Press edition, Hume (1748/2007) *An Enquiry Concerning Human Understanding,* Oxford: Oxford University Press. ed and intro P. Millican. First published 1748.

2 All references to Kant are taken from the Kemp Smith translation in the Palgrave Macmillan edition, Kant (1781, 1787/2007) *The Critique of Pure Reason*, trans Norman Kemp Smith, introduction Howard Caygill, bibliography Gary Banham, Basingstoke: Palgrave/Macmillan.

3 The distinctions made by Kant were developed subsequently by John Stuart Mill (1843). Mill distinguished between verbal and real propositions and between merely apparent and real inferences. The distinction, Mill himself noted, corresponds to Kant's analytic and synthetic judgements. Mill differed from Kant by asserting that apparent inferences have no genuine cognitive content. Pure mathematics and logic do, according to Mill, contain real positions and have genuine cognitive content. In the final analysis, Mill held that all reasoning is empirical through a process he called enumerative induction or simple generalisation from experience. This he recognised was a fallible process (Skorupski, 2005). There are some striking similarities between Kant and Mill, and it would be possible to use Mill as a basis to deconstruct EBM. However, we propose that the original Kantian scheme of judgements offers the more helpful way of describing the judgement process in EBM rather than subsequent empiricist or indeed positivist arguments.

References

Barnett, D., Stevens, A., Boysen, M. and Longson, C. (2009) Developing evidence based guidance for health technologies: The nice experience. In: A. Killoran and M.P. Kelly (eds.) *Evidence Based Public Health: Effectiveness and Efficiency*. Oxford: Oxford University Press.

Bell, J. (2010) *Redefining Disease. The Harveian Oration of 2010*. London: Royal College of Physicians.

Black, N. (1996) Why we need observational studies to evaluate the effectiveness of healthcare. *British Medical Journal* 312(May): 1215–1218.

Blalock, H.M. (1970) *An Introduction to Social Research*. Engelwood Cliffs, NJ: Prentice Hall.

Campbell, D.T. and Stanley, J.C. (1966) *Experimental and Quasi Experimental Designs for Research*. Chicago: Rand McNally.

Centre for Reviews and Dissemination. (2008) *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare*. York: Centre for Reviews and Dissemination, University of York.

Cochrane, A.L. (1972) *Effectiveness and Efficiency: Random Reflections on Health Services*. London: British Medical Journal/Nuffield Provincial Hospitals Trust.

Culyer, A.J. (2006) The bogus conflict between efficiency and equity. *Health Economics* 15(11): 1155–1158.

Culyer, A.J., Lavers, R.J. and Williams, A. (1972) Health indicators. In: A. Shonfield and S. Shaw (eds.) *Social Indicators and Social Policy*. London: Heinemann.

Davidson, K. *et al* (2003) Evidence – based behavioral medicine: What it is and how do we achieve it? *Annals of Behavioral Medicine* 26(3): 161–171.

Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J. and Stoddart, G.L. (2005) *Methods for the Evaluation of Healthcare Programmes*, 3rd edn. Oxford: Oxford University Press.

Drummond, M. *et al* (2007) Assessing the Challenges of Applying Standard Methods of Economic Evaluation to Public Health Interventions, Public Health Research Consortium, Final Report York: University of York.

Egger, M., Davey Smith, G. and Altman, B.G. (2001) *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books.

Glasziou, P., Chalmers, I., Rawlins, M. and McCulloch, P. (2007) When are randomised trials unnecessary? Picking signal from noise. *British Medical Journal* 334(February): 349–351.

Greenhalgh, T. (2001) *How to Read a Paper: The Basics of Evidence Based Medicine*. London: BMJ Books.

Guyer, P. (2006) *Kant*. London: Routledge.

Hume, D. (1748/2007) *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press, P. Millican (ed. and intro.). First published 1748.

Kant, I. (1781, 1787/2007) *The Critique of Pure Reason*, Translated by Norman Kemp Smith, Introduction Howard Caygill, bibliography Gary Banham. Basingstoke, UK: Palgrave/Macmillan.

Kelly, M.P. and Moore, T.A. (2010) *Making a Difference: Using the NICE Public Health Guidance and Embedding Evaluation*. London: IDEA.

Kelly, M.P., Morgan, A., Ellis, S., Younger, T., Huntley, J. and Swann, C. (2010) Evidence based public health: A review of the experience of the national institute of health and clinical excellence (NICE) of developing public health guidance in England. *Social Science and Medicine* 71: 1056–1062.

Korner, S. (1955) *Kant*. Harmondsworth, UK: Penguin.

Lomas, J., Culyer, T., McCutcheon, C., McCAuley, L. and Law, S. (2005) *Conceptualizing and Combining Evidence for Health System Guidance*. Ottawa, Canada: Canadian Health Services Research Foundation.

Mill, J.S. (1843/1947) *A System of Logic*. London: New impression, Longmans, Green, 1947.

Millican, P. (2007) *Introduction Hume's Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.

Moore, T.A and Rutherford, D. (2012) Primary strategy learning networks: A local study of a national initiative. *Educational Management Administration and Leadership*, published online before print, 15 November 2011, doi: 10.1177/1741143211420612.

NICE. (2008) *Social Value Judgements: Principles for the Development of NICE Guidance*, 2nd edn. London: NICE, http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf, accessed 17 November 2011.

NICE. (2009) *Methods for the Development of NICE Public Health Guidance*, 2nd edn. London: NICE, http://www.nice.org.uk/aboutnice/howwework/developingnicepublichealthguidance/publichealthguidanceprocessandmethodguides/public_health_guidance_process_and_method_guides.jsp?domedia=1&mid=F6A97CF4-19B9-E0B5-D42B4018AE84DD51, accessed 17 November 2011.

Orne, M.T. (1970) On the social psychology of the psychological experiment: with particular reference to the demand characteristics and their implications. In: Denzin, N.K. (ed.) *Sociological Methods: A Sourcebook*. Chicago: Aldine.

Pawson, R. (2002) Evidence based policy: In search of a method. *Evaluation* 8(2): 157–181.

Pawson, R. (2006) *Evidence Based Policy: A Realist Perspective*. London: Sage.

Pickstone, J.V. (2007) Working knowledges before and after circa 1800: Practices and disciplines in the history of science, technology and medicine. *Isis* 98(3): 489–516.

Rawlins, M. (2008) *De Testimonio: On the Evidence for Decisions About the Use of Therapeutic Interventions, The Harveian Oration of 2008*. London: Royal College of Physicians.

Russell, J., Greenhalgh, T., Byrne, E. and McDonnell, J. (2008) Recognizing rhetoric in health care policy analysis. *Journal of Health Services Research and Policy* 13(1): 40–46.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B and Richardson, W.S. (1996) Evidence based medicine: What it is and what it isn't. *British Medical Journal* 312: 71–72.

Sorensen, C., Drummond, M., Børlum Kristensen, F. and Busse, R. (2010) *How can the impact of health technology assessments be enhanced? Policy brief*, World Health Organization on behalf of the European Observatory on Health Systems and Policies, http://bit.ly/a1NxTR, accessed 17 November 2011.

Skorupski, J. (2005) *The Oxford Companion to Philosophy*, In: J.S. Mill and T. Honderich (eds.) 2nd edn. Oxford: Oxford University Press, pp. 599–602.

Vandenbroucke, J.P. (2008) Observational research, randomised trials, and two views of medical science. *PLoS Med* 5(3): e67.

Ward, A. (2006) *Kant: The Three Critiques*. Cambridge: Polity.

Weldon, T.D. (1958) *Kant's Critique of Pure Reason*, 2nd edn. Oxford: Clarendon.