World J. Surg. 23, 789–793, 1999

WORLD
Journal of
SURGERY

Evidence-based Surgery: A Passing Fad?

Nick Black, M.D.

Health Services Research Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Abstract. Recent years have witnessed the development of a new movement within health care: the promotion of "evidence-based medicine" (EBM). EBM is about integrating individual clinical expertise and the best external evidence derived from scientific research. Advocates claim that much medical practice is based too much on opinion and experience and insufficiently on research evidence. Their approach would increase the quality of care and its efficiency. This paper describes the principal steps in the evidence-based approach—systematic reviews of the literature and meta-analyses—and its shortcomings in surgery. These include the reliance of EBM on randomized trials, the lack of generalizability of scientific evidence to individual patients, the lack of attention to third party interests, the threat to the "art" of medicine, and the dangers of an oversimplistic approach. Although EBM clearly has a place, it does not have all the answers.

The management of health services presents enormous challenges. First, health care is highly complex: There are dozens of occupational groups employed in providing health care, often competing with one another. Complexity also arises because no two patients are identical, which restricts the extent to which mass, standardized processing can be used. For many patients, management of their care requires multiple activities to be coordinated. Second, all this complexity is ever changing. Third, unlike many other complex, dynamic organizations, some of the employees, most particularly the doctors, not only exercise enormous influence on how resources are used but also have considerable autonomy.

As if these challenges were not enough, health services are subject to many external pressures described as "environmental turbulence" [1]. Four principal sources of turbulence can be identified: those who pay for health services; public or consumer opinion; health care staff; and the medical–industrial complex (backed up by much of the biomedical research community) [2]. All are legitimate stakeholders, though their relative importance varies by country.

Given the existence of such diverse influences, it is not surprising that resources may not be used in the most effective way. This is seen when data on geographic variations in surgical rates are considered. Whereas the rates of some operations (e.g., groin hernia repair) vary little, for some they vary considerably (prostatectomy, hysterectomy) and for others the variation is enormous (hemorrhoidectomy, tonsillectomy) [3]. The extent of variation depends on the operation and not on the country studied or the overall level of funding of a health system. In other words, hernia repair shows little variation in all countries studied, whereas tonsillectomy is highly variable everywhere. These differences cannot be explained by different levels of need or the availability of services. Instead, it has been shown that most of the variation is due to differences in the judgment of clinicians.

This has long been recognized. A study during the 1920s in New York City on tonsillectomy demonstrated the vagaries of clinical judgment [4]. A series of one thousand 11-year-olds had their throats examined. It was found that 61% had already had their tonsils removed. Of the other 39%, the examining doctor thought that about half of them needed tonsil surgery. Those children deemed to have healthy tonsils were taken and examined by another doctor who thought that half of them required surgery. Again the healthy children were reexamined by yet another doctor who declared that half of them needed their tonsils removed. After four examinations, only 65 of 1000 children would have escaped with their tonsils intact.

Such variation in clinical opinion is not a thing of the past. More recently a group of physicians in the United States and a group in the United Kingdom were asked to consider several hundred clinical situations or patient scenarios and for each decide whether they thought coronary surgery was appropriate, inappropriate, or equivocal [5]. Despite doctors in the two countries sharing many of the same textbooks and journals, there was a considerable difference between the countries: the U.S. physicians considered surgery was appropriate in many more situations than the British physicians. Similar differences have been found for other operations (Table 1).

These sorts of differences have led investigators to suggest that there is a need to establish a much stronger base of scientific research evidence as to effectiveness and appropriateness of treatments. One group of clinicians and epidemiologists based at McMaster University, Canada have adopted the application of research evidence to clinical practice as their mission and termed their approach "evidence-based medicine." What is it, what can it contribute, and to what extent is it just a passing fad?

Evidence-based Movement

The principal plank of the evidence-based (EB) approach is the view that without the current best scientific evidence, clinical practice risks rapidly becoming out of date to the detriment of patients [6]. It does, however, also recognize that without clinical

Table 1. British and US clinicians' views of the appropriateness of coronary artery bypass grafting.

Physicians	Appropriate (%)	Equivocal (%)	Inappropriate (%)
USA	62	25	13
UK	41	24	35

From Brook et al. [5], with permission.

expertise practice risks becoming tyrannized by evidence, as even excellent external evidence may be inapplicable to or inappropriate for an individual patient. In the words of one of the founders of the EB movement: "It's about integrating individual clinical expertise and the best external evidence" [6], though explanations of how this is to be carried out leaves many clinicians unconvinced about its practicality.

How is evidence-based medicine (EBM) carried out? It is basically about ways of accessing, appraising, and incorporating scientific evidence about the appropriate clinical management of patients. It involves four steps [7]:

1. Formulate a clear clinical question from a patient's problem.
2. Search the literature for relevant clinical articles.
3. Evaluate (critically appraise) the evidence for its validity and usefulness.
4. Implement useful findings in clinical practice.

Before considering the practicalities, the theoretic dilemma of defining what is "appropriate" must be addressed. The idea that some treatments are effective and others ineffective is an oversimplification. Instead, patients extend over a spectrum from healthy to severely ill. The challenge is to define where the threshold for an intervention should lie; that is, when do the benefits or advantages outweigh the costs/disadvantages by a sufficiently wide margin for the intervention to be deemed "appropriate"?

Systematic Reviews and Meta-analysis

The cornerstone of the EB approach is consideration of the results of what they define as the best evidence—that from randomized controlled trials (RCTs). To do this it is necessary to review all the relevant literature systematically. Such a review may already exist. There are now electronic databases of systematic reviews such as the Cochrane Database and DARE. Of what does a systematic review consist? It involves four main steps.

1. *Identify topic, define inclusion and exclusion criteria.* There is a need to clarify exactly what question(s) requires answering. For example, in a recent review of laparoscopic cholecystectomy [8] the questions were as follows: Is laparoscopic cholecystectomy as effective as open and minilaparotomy cholecystectomy for treatment of patients with symptomatic gallstones? Is laparoscopic cholecystectomy as safe or safer than open and minicholecystectomy? Is postoperative recovery from laparoscopic cholecystectomy faster than from open and minicholecystectomy?

2. *Literature search using multiple sources.* Several approaches can and should be adopted: electronic databases such as MEDLINE and EMBASE; citations in those papers found; contacting relevant governmental organizations and members of professional organizations. An attempt should be made to discover any unpublished studies, as they are more likely to have reached so-called "negative" results. In other words, the published literature may be biased toward positive results.

Table 2. Methodologic quality of randomized trials conducted in two areas of surgery.

Parameter	Groin hernia repair	Surgery for stress incontinence
Bias (20)	14 (3–20)	12 (3–14)
Confounding (20)	13 (0–20)	10 (7–15)
External validity (20)	0 (0–13)	0 (0)
Power (20)	0 (0–4)	0 (0–16)
Reporting (20)	13 (3–20)	12 (6–19)
Total (100)	41 (9–71)	32 (23–48)

Results are medians (range).

3. *Assess the quality of each study.* The methodologic quality of each study then needs to be assessed. Various checklists exist for assessing randomized trials [9]. Such checklists assess quality on several dimensions. The results of two reviews demonstrate that the quality of much research is poor (Table 2).

4. *Synthesize findings: from qualitative to statistical meta-analysis.* Having decided which studies are the least biased and most reliable, the findings of various studies must be combined to obtain an overview. Studies in surgery tend to be heterogeneous as regards the way cases were defined, the inclusion criteria used, the way cases were clinically managed, and the outcomes measured. Statistical meta-analysis is therefore rarely justified. Instead, a more qualitative overview is needed.

Limitations of EBM

Reliance on RCTs Too Limited in Surgery

The widely held view that experimental methods (RCTs) are the "gold standard" for evaluation has led to the denigration of nonexperimental methods to the extent that research funding bodies and journal editors automatically reject them. This stems from a misguided notion that everything can be investigated using an RCT. In practice, other types of study design are frequently used in surgery (Table 3).

The limitations of randomized trials can be seen as deriving from either the inherent nature of the method (a limitation in principle) or from the way trials are conducted (a limitation in procedure). There are four main reasons: experimentation may be unnecessary, inappropriate, impossible, or inadequate.

Experimentation May Be Unnecessary. When the effect of an intervention is dramatic, the likelihood of unknown confounding factors being important is so small they can be ignored. There are many well known examples of such interventions, including anesthesia for surgical operations and immobilization of fractured bones.

Experimentation May Be Inappropriate. There are three situations in which randomized trials may be inappropriate. The first is that they are rarely large enough to quantify accurately the infrequent adverse outcomes. This limitation has been recognized with the establishment in many countries of postmarketing surveillance schemes to detect rare adverse effects of drugs. The need for

Table 3. Study designs used to evaluate the effectiveness of three common surgical procedures.

Study design	Laparoscopic cholecystectomy (no.)	Surgery for stress incontinence (no.)	Groin hernia repair (no.)
Randomized trials	15	11	33
Nonrandomized trials/prospective cohort studies	21	20	18
Retrospective cohort studies	19	45	30
Cohort with nonparallel comparison	21	0	11

similar surveillance schemes for surgical interventions has recently been recognized in the United Kingdom with the establishment of a scheme for surgeons to register innovations. At present it is done only on a voluntary basis. The lack of such schemes means that there is still uncertainty as to whether laparoscopic techniques are associated with an increased risk of injuries, such as bile duct damage during cholecystectomy [8]. Huge observational datasets are the only practical means of acquiring such vital information.

A second limitation of trials is when the outcomes of interest are far in the future. For example, the loosening of artificial hip joints for which a 10- to 15-year follow-up is needed. The practical difficulties of maintaining such prolonged prospective studies (whether experimental or observational) are considerable, as are their costs. As a practical alternative to doing nothing, retrospective observational studies can be used to obtain some information on long-term outcomes [10].

Third, a randomized trial may be inappropriate because the very act of random allocation may reduce the effectiveness of the intervention. For example, if a surgeon is experienced in open cholecystectomy, being asked to randomize patients to open or laparoscopic procedures may well achieve better results with the open operations simply because of his lack of familiarity with the laparoscopic approach. As a consequence, randomized trials may find less benefit for the new procedure than observational studies in which surgeons use their preferred option. The same may be true for many interventions in which clinicians and patients have a preference (despite agreeing to random allocation) [11].

Experimentation May Be Impossible. The first, and most familiar, is the reluctance and refusal of clinicians to participate. Just because clinical uncertainty, manifested as a variation in practice, may exist does not mean that each individual clinician lacks certainty as to how to practice. In other words, although "collective equipoise" existed, "individual equipoise" was absent [12]. Even when clinicians purport to participate, randomization may be subverted by clinicians deciphering the assignment sequence [13].

A second potential obstacle is ethical objections. It is most unlikely that any ethics committee would sanction the random allocation of patients to cardiac transplantation versus medical management.

Third, researchers may also meet legal obstacles to performing a randomized trial. The classic example is the attempt to subject radial keratotomy (an operation to correct short sightedness) to a randomized trial in the United States [14]. The researchers were blocked by private ophthalmologists who faced a major loss of income if the procedure was declared "experimental" because it would mean health insurance companies would no longer reimburse them. As a result of legal action, the academic ophthalmologists were forced to declare the operation safe and effective and abandon any attempt at evaluation.

The final reason it is not always possible to conduct randomized trials is simply the scale of the task confronting the research community. There is an immense number of health care interventions in use. In addition, most interventions have many components. Consider a simple surgical operation: preoperative tests, anesthesia, the surgical approach, wound management, postoperative nursing, discharge practice—and these are just the principal components.

Experimentation May Be Inadequate. The external validity, or generalizability, of the results of randomized trials is often low [15]. The extent to which the results of a trial are generalizable depends on the extent to which the outcome of the intervention is determined by the particular person providing the care. The outcome of surgery may be highly dependent on the characteristics of the provider, the setting, and the patients.

There are three reasons randomized trials may have low external validity. The first is that the surgeons who participate may be unrepresentative. They may have a particular interest in the topic or be enthusiasts and innovators. In addition, the setting may be atypical, such as a teaching hospital. Second, the patients who participate may be atypical. All trials exclude certain categories of patients from their study. Often the exclusion criteria are so restrictive the patients who are eligible for inclusion represent only a small proportion of the patients being treated in normal practice. Only 4% of patients currently undergoing coronary revascularization in the United States would have been eligible for inclusion in the trials conducted during the 1970s [16].

The problem of eligibility may be exacerbated by a poor recruitment rate. Although most trials fail to report their recruitment rate [8], those that do suggest rates are often low. As little is yet known about the sort of people who are prepared to have their treatment allocated on a random basis, it seems wise to assume that they may differ in important ways from those who decline to take part. The third and final problem of generalizing the results of randomized trials is that treatment may be atypical. Patients who participate may receive better care, regardless of which arm of the trial they are in [17].

As a result of these problems, randomized trials generally offer an indication of the efficacy of an intervention rather than its effectiveness in everyday practice. Although the latter can be achieved through "pragmatic" trials that evaluate normal clinical practice, they are rarely undertaken [18]. Most randomized trials are "explanatory"; that is, they provide evidence of what can be achieved in the most favorable circumstances.

Generalizability of Evidence to Individual Patients

Evidence from RCTs provide overall probabilities of outcomes, not precise, certain predictions for any individual patient. By definition, they provide an overview for all patients included in the trial. It is then assumed that this applies to every patient presenting with the condition. As Rothwell has pointed out [19]:

Patients included in a clinical trial are heterogeneous and may, for example, differ in the severity of illness and consequently in the absolute risk of a poor outcome. Therefore, a treatment that produces an overall relative risk reduction, but has significant morbidity or mortality, may be ineffective or even harmful in patients at low risk.

He went on to reanalyze the results of an RCT (European Carotid Surgery Trial) in which carotid endarterectomy was compared with no surgery in patients who had recent carotid-distribution transient ischemic attacks or nondisabling strokes and who had severe stenosis of the relevant carotid artery (70–99%).

Overall, patients experienced a 50% reduction in relative risk with surgery. Whereas those with a high (> 15%) or moderate (10–15%) baseline risk of a poor outcome benefited from relative risk reductions, those with a low baseline risk (< 10%) gained no benefit or even suffered some harm. Operations in 7 high risk patients or 14 moderate risk patients would result in one stroke prevented, whereas for every 71 operations in low risk patients one additional stroke was likely.

Ignores Societal or Third Party Interests

Some critics have pointed out that inherent to the EBM approach is a disregard for the cost of treatments, whether to the patient or the society [20]. EBM does not help decide the threshold at which patients should be accepted for treatment or assist in deciding between the competing priorities of different patients. EBM is focused on individual patients and does not acknowledge the opportunity cost to other patients.

Endangers the "Art" of Medicine

Other commentators have expressed concern about the threat EBM poses to the personal relationship doctors have with patients [21]. Although recognizing the potential benefits of a more scientific approach to clinical practice, McCormick pointed out the dangers [21]:

Guidelines derive from population studies and are not always applicable to the unique person who decides to consult. All might be well if guidelines were securely based and if they were perceived as giving advice rather than mandatory instruction. Unfortunately many guidelines are insecurely based, and doctors, sometimes fearing medico-legal consequences are motivated towards slavish adherence to them.

Danger of Overly Simplistic Application

Related to the just-stated problem, others have warned of the danger of EBM focusing attention on the measurable, usually medical treatment aspects of health services at the expense of the caring aspects [22]:

Most health service managers do not wish to downgrade caring any more than clinicians, but they are compelled to seek efficiency in terms that effectively ignore what is difficult to measure.

In a similar vein, Flood [23] and Klein [24] have warned about the adoption of scientism based as it is on "over-hopeful assumptions that are unlikely to be fully realized." The danger Klein envisaged is that a vulgarized form of the new scientism will be taken up by payers, purchasers, and managers, which will result in eventual disillusionment. This is not to argue against the application of science to improve health services but, rather, "to be as aware of the limitations . . . as of the benefits."

Others have taken a less compromising line, perceiving a serious threat to individual freedom of thought and action [25].

Clinicians should not be bashful about putting "evidence-based" mania in its place. Overweening and unjustifiable ambition should be opposed; just as we ought to oppose the imposition of any system of rationalistic dictatorship based upon simplistic and incomplete analysis. Indeed, in the mad world of health reform, a totalitarian system of EBM might be easier to establish than to dislodge.

Conclusions

Given these challenges, will EBM prove to be no more than a passing fad? Although advocates of EBS adhere to an oversimplified notion of clinical practice and health care policy, the key part of their message is of great importance. It is clearly in everyone's interest that health services be of proven effectiveness. To base decisions for individual patients or for populations only on evidence of clinical effectiveness is inadequate. Two other interests must be taken into account: the cost of services, regardless of who is paying, and the preferences of patients. EBM seems to shy away from such considerations because of the complexity they introduce into the conduct of evidence-based practice. While such complexity represents a serious challenge, there is a solution in the form of decision analysis [26]:

EBM, as so far expounded, reflects a problem-solving attitude that results in a heavy concentration on RCTs and meta-analyses, rather than a broad decision making focus that concentrates on meeting all the requirements of a good clinical decision. The latter include: ensuring that inferences from RCTs and meta-analyses to individual patients are made explicitly; paying equally serious attention to evidence on values and costs as to clinical evidence; and accepting the inadequacy of "taking into account and bearing in mind" as a way of integrating the multiple and distinct elements of a decision.

The potential for decision analysis to resolve some of the dilemmas has even been recognized by some of EBM's strongest advocates [27]. If such recognition of the limitations of a purer EBM approach were to spread among its advocates, there is a chance that it will prove to be more than a passing fad. The only problem is that the modified EBM that emerges looks rather like something we could call GMP—good medical practice—and something that we could all support. As Carr-Hill has noted [28]:

No one denies the importance of evidence: it is a sine qua non of professional practice. But often, there are no simple answers to apparently simple questions: there is a role for judgement in decision-making—just as in criminal law trials both prosecution and defence try to build a convincing picture to place before the jury. But this does not mean that lawyers or juries ignore the evidence: indeed, would it not have been seen as rather silly to promote "evidence-based law"?

Résumé

Ces dernières années, on a vu se développer un nouvelle approche dans notre système de santé, la promotion de la médecine «factuelle» («evidence-based medicine»). La médecine factuelle concerne l'intégration de l'expertise clinique individuelle et la meilleure preuve externe dérivée de la recherche scientifique. Ces défenseurs clament que trop de pratiques médicales sont basées sur l'opinion et l'expérience personnelles de certains, sans s'appuyer suffisamment sur des preuves provenant de la recherche. Cette nouvelle approche augmenterait la qualité des soins et leur efficience. Ce chapitre décrit les étapes principales de l'approche de la médicine factuelle- les revues systématique de la

littérature et les méta-analyses- ainsi que ses défauts en ce qui concerne la chirurgie. Ceux-ci comprennent : la dépendance de la médicine factuelle sur les études randomisées, l'insuffisance de la généralisation des preuves scientifiques aux patients individuels, le manque d'attention apportée aux intérêts des autres partenaires dans le système de soins, le menace de voir disparaître «l'art» de la médecine, et les dangers d'une approche trop simpliste. La médecine factuelle a certainement sa place, mais elle ne donne pas toutes les réponses à toutes les questions.

Resumen

En los últimos años se ha evidenciado el desarrollo de un nuevo movimiento en el campo de la atención de la salud: la promoción de la "medicina basada en la evidencia" (MBE). La MBE se refiere a la integración de la habilidad clínica individual con la mejor evid*encia externa derivada de la investigación científica. Sus defensores proclaman que mucha de la práctica médica se fundamenta excesivamente en la opinión y la experiencia, y poco en la evidencia que provee la investigación. Dicen que la adopción de este enfoque vendría a incrementar la calidad y eficiencia de la atención. El presente artículo describe los aspectos principales del enfoque "basado en la evidencia"-revisiones sistemáticas y meta-análisis- y sus limitaciones en la cirugía. Tales aspectos incluyen la dependencia de la MBE de los estudios randomizados, la falla en cuanto a la generalización de la evidencia científica a los pacientes individuales, la falta de atención a los intereses de terceros, la amenaza contra el "arte" de la medicina y los peligros de un aproche excesivamente simplista. Aunque es claro que la MBE tiene un lugar definido, no es la dueña de todas las respuestas.

References

1. Evans, T.: Strategic response to environmental turbulence. In: Managers as Strategists, G. Parston, editor. London, King Edward's Hospital Fund for London, 1986
2. Relman, A.S.: The new medical-industrial complex. N. Engl. J. Med. *303*:963, 1980
3. McPherson, K., Wennberg, J.E., Hovind, O.B., Clifford, P.: Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. N. Engl. J. Med. *307*:1310, 1982
4. American Child Health Association: Physical Defects: The Pathway to Correction. New York, American Child Health Association, 1934
5. Brook, R.H., Kosecoff, J.B., Park, R.E., Chassin, M.R., Winslow, C.M., Hampton, J.R.: Diagnosis and treatment of coronary disease: comparison of doctors' attitudes in the USA and the UK. Lancet *1*:750, 1988
6. Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., Richardson, W.S.: Evidence based medicine: what it is and what it isn't. B.M.J. *312*:71, 1996
7. Rosenberg, W., Donald, A.: Evidence based medicine: an approach to clinical problem-solving. B.M.J. *310*:1122, 1995
8. Downs, S.H., Black, N.A., Devlin, H.B., Royston, C.M.S., Russell, R.C.G.: Systematic review of the effectiveness and safety of laparoscopic cholecystectomy. Ann. R. Coll. Surg. Engl. *78*:241, 1996
9. Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P., Walsh, S.: Assessing the quality of randomised controlled trials: an annotated bibliography of checklists. Controlled Clin. Trials *16*:62, 1995
10. Stauffer, R.N.: Ten-year follow-up study of total hip replacement. J. Bone Joint Surg. Am. *64*:983, 1982
11. Brewin, C.R., Bradley, C.: Patient preferences and randomised clinical trials. B.M.J. *299*:313, 1989
12. Lilford, R., Jackson, J.: Equipoise and the ethics of randomization. J. R. Soc. Med. *88*:552, 1995
13. Schulz, K.F.: Subverting randomization in controlled trials. J.A.M.A. *274*:1456, 1995
14. Chalmers, I.: Minimizing harm and maximizing benefit during innovation in health care: controlled or uncontrolled experimentation? Birth *13*:155, 1986
15. Editorial: Cross design synthesis: a new strategy for studying medical outcomes? Lancet *340*:944, 1992
16. Hlatky, M.A., Califf, R.M., Harrell, F.E., Lee, K.L., Mark, D.B., Pryor, D.: Comparison of predictions based on observational data with the results of randomized controlled trials of coronary artery bypass surgery. J. Am. Coll. Cardiol. *11*:237, 1988
17. Stiller, C.A.: Centralised treatment, entry to trials, and survival. Br. J. Cancer *70*:352, 1994
18. Schwartz, D., Lellouch, J.: Explanatory and pragmatic attitudes in clinical trials. J. Chronic Dis. *20*:637, 1967
19. Rothwell, P.M.: Can overall results of clinical trials be applied to all patients? Lancet *345*:1616, 1995
20. Maynard, A.: Evidence-based medicine: an incomplete method for informing treatment choices. Lancet *349*:126, 1997
21. McCormick, J.: Death of the personal doctor. Lancet *348*:667, 1996
22. Hart, J.T., Dieppe, P.: Caring effects. Lancet *347*:1606, 1996
23. Flood, A.B.: Scientific bases versus scientism in health services. J. Health Serv. Res. Policy *1*:63, 1996
24. Klein, R.: The NHS and the new scientism: solution or delusion? Q. J. Med. *89*:85, 1996
25. Charlton, B.G.: Restoring the balance: evidence-based medicine put in its place. J. Eval. Clin. Pract. *3*:87, 1997
26. Dowie, J.: Evidence-based, cost-effective and preference-driven medicine: decision analysis based medical decision making is the prerequisite. J. Health Serv. Res. Policy *1*:104, 1996
27. Greenhalgh, T., Worrall, J.G.: From EBM to CSM: the evolution of context-specific medicine. J. Eval. Clin. Pract. *3*:105, 1997
28. Carr-Hill, R.: Welcome? To the brave new world of evidence based medicine. Soc. Sci. Med. *41*:1467, 1995