Evidence-based medicine

Contents

1 Overview
2 Classification
   2.1 Evidence-based guidelines
   2.2 Evidence-based individual decision making
3 History
4 Qualification of evidence
   4.1 Categories of recommendations
5 Statistical measures in evidence-based medicine
6 Quality of clinical trial publications
   6.1 Limitations of available evidence
7 Effectiveness
8 Criticism of evidence-based medicine
9 Evidence-Based Policy
10 See also
11 References
12 External links

Evidence-based medicine (EBM) aims to apply evidence gained from the scientific method to certain parts of medical practice. It seeks to assess the quality of evidence relevant to the risks and benefits of treatments (including lack of treatment). According to the Centre for Evidence-Based Medicine, "Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients."[2]

EBM recognizes that many aspects of medical care depend on individual factors such as quality and value-of-life judgments, which are only partially subject to scientific methods. EBM, however, seeks to clarify those parts of medical practice that are in principle subject to scientific methods and to apply these methods to ensure the best
prediction of outcomes in medical treatment, even as debate about which outcomes are desirable continues.

The foundation of evidence-based medicine is the systematic review of evidence for particular treatments, mainly randomized controlled trials. The Cochrane Collaboration leads this effort. A 2001 review of 160 Cochrane systematic reviews in the 1998 database revealed that, according to two readers, 41.3% concluded positive or possibly positive effect, 20% concluded evidence of no effect, 8.1% concluded net harmful effects, and 21.3% of the reviews concluded insufficient evidence.[3] A review of 145 alternative medicine Cochrane reviews using the more up-to-date 2004 database revealed that 38.4% concluded positive effect or possibly positive (12.4%) effect, 4.8% concluded no effect, 0.69% concluded harmful effect, and 56.6% concluded insufficient evidence.[4]

1 Overview

Using techniques from science, engineering, and statistics, such as meta-analysis of medical literature, risk-benefit analysis, and randomized controlled trials, EBM aims for the ideal that healthcare professionals should make "conscientious, explicit, and judicious use of current best evidence" in their everyday practice.

Generally, there are three distinct, but interdependent, areas of EBM. The first is to treat individual patients with acute or chronic pathologies by treatments supported in the most scientifically valid medical literature. Thus, medical practitioners would select treatment options for specific cases based on the best research for each patient they treat. The second area is the systematic review of medical literature to evaluate the best studies on specific topics. This process can be very human-centered, as in a journal club, or highly technical, using computer programs and information techniques such as data mining. Increased use of information technology turns large volumes of information into practical guides. Finally, evidence-based medicine can be understood as a medical "movement" in which advocates work to popularize the method and usefulness of the practice in the public, patient communities, educational institutions, and continuing education of practicing professionals.

Evidence-based medicine has demoted ex cathedra statements of the "medical expert" to the least valid form of evidence. All "experts" are now expected to reference their pronouncements to scientific studies.

2 Classification

Two types of evidence-based medicine have been proposed.[5]

2.1 Evidence-based guidelines

Evidence-based guidelines (EBG) is the practice of evidence-based medicine at the organizational or institutional level. This includes the production of guidelines, policy, and regulations. This approach has also been called evidence based healthcare.[6]

2.2 Evidence-based individual decision making
Evidence-based individual decision (EBID) making is evidence-based medicine as practiced by the individual health care provider. There is concern that current evidence-based medicine focuses excessively on EBID.\(^5\)

3 History

Although testing medical interventions for efficacy has existed since the time of Avicenna's *The Canon of Medicine* in the 11th century, it was only in the 20th century that this effort evolved to impact almost all fields of health care and policy. Professor Archie Cochrane, a Scottish epidemiologist, through his book *Effectiveness and Efficiency: Random Reflections on Health Services* (1972) and subsequent advocacy, caused increasing acceptance of the concepts behind evidence-based practice. Cochrane's work was honoured through the naming of centres of evidence-based medical research — *Cochrane Centres* — and an international organization, the *Cochrane Collaboration*. The explicit methodologies used to determine "best evidence" were largely established by the McMaster University research group led by David Sackett and Gordon Guyatt. The term "evidence based" was first used in 1990 by David Eddy.\(^5\) The term "evidence-based medicine" first appeared in the medical literature in 1992 in a paper by Guyatt *et al.*\(^10\)

4 Qualification of evidence

Evidence-based medicine categorizes different types of clinical evidence and ranks them according to the strength of their freedom from the various biases that beset medical research. For example, the strongest evidence for therapeutic interventions is provided by systematic review of randomized, double-blind, placebo-controlled trials involving a homogeneous patient population and medical condition. In contrast, patient testimonials, case reports, and even expert opinion have little value as proof because of the placebo effect, the biases inherent in observation and reporting of cases, difficulties in ascertaining who is an expert, and more.

Systems to stratify evidence by quality have been developed, such as this one by the U.S. Preventive Services Task Force for ranking evidence about the effectiveness of treatments or screening:

- Level I: Evidence obtained from at least one properly designed randomized controlled trial.

- Level II-1: Evidence obtained from well-designed controlled trials without randomization.

- Level II-2: Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group.

- Level II-3: Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled trials might also be regarded as this type of evidence.
Level III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

The UK National Health Service uses a similar system with categories labeled A, B, C, and D. The above Levels are only appropriate for treatment or interventions; different types of research are required for assessing diagnostic accuracy or natural history and prognosis, and hence different "levels" are required. For example, the Oxford Centre for Evidence-based Medicine suggests levels of evidence (LOE) according to the study designs and critical appraisal of prevention, diagnosis, prognosis, therapy, and harm studies.[11]

Level A: Consistent Randomised Controlled Clinical Trial, cohort study, all or none (see note below), clinical decision rule validated in different populations.

Level B: Consistent Retrospective Cohort, Exploratory Cohort, Ecological Study, Outcomes Research, case-control study; or extrapolations from level A studies.

Level C: Case-series study or extrapolations from level B studies.

Level D: Expert opinion without explicit critical appraisal, or based on physiology, bench research or first principles.

A newer system is by the Grade Working Group and takes in account more dimensions that just the quality of medical evidence.[12] “Extrapolations” are where data is used in a situation which has potentially clinically important differences than the original study situation. Thus, the quality of evidence to support a clinical decision is a combination of the quality of research data and the clinical 'directness' of the data.[13]

Despite the differences between systems, the purposes are the same: to guide users of clinical research information about which studies are likely to be most valid. However, the individual studies still require careful critical appraisal.

Note: The all or none principle is met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it.

4.1 Categories of recommendations

In guidelines and other publications, recommendation for a clinical service is classified by the balance of risk versus benefit of the service and the level of evidence on which this information is based. The U.S. Preventive Services Task Force uses:[14]

Level A: Good scientific evidence suggests that the benefits of the clinical service substantially outweighs the potential risks. Clinicians should discuss the service with eligible patients.

Level B: At least fair scientific evidence suggests that the benefits of the clinical service outweighs the potential risks. Clinicians should discuss the service with eligible patients.
Level C: At least fair scientific evidence suggests that there are benefits provided by
the clinical service, but the balance between benefits and risks are too close for
making general recommendations. Clinicians need not offer it unless there are
individual considerations.

Level D: At least fair scientific evidence suggests that the risks of the clinical service
outweighs potential benefits. Clinicians should not routinely offer the service to
asymptomatic patients.

Level I: Scientific evidence is lacking, of poor quality, or conflicting, such that the
risk versus benefit balance cannot be assessed. Clinicians should help patients
understand the uncertainty surrounding the clinical service.

5 Statistical measures in evidence-based medicine

Evidence-based medicine attempts to express clinical benefits of tests and treatments
using mathematical methods. Tools used by practitioners of evidence-based medicine
include:

Likelihood ratios. The pretest probability of a particular diagnosis, multiplied by the
likelihood ratio, determines the posttest probability. This reflects Bayes' theorem. The
differences in likelihood ratio between clinical tests can be used to prioritize clinical
tests according to their usefulness in a given clinical situation.

The area under the receiver operator characteristic curve (AUC-ROC) reflects the
relationship between sensitivity and specificity for a given test. High-quality tests will
have an AUC-ROC approaching 1, and high-quality publications about clinical tests
will provide information about the AUC-ROC. Cutoff values for positive and
negative tests can influence specificity and sensitivity, but they do not affect AUC-
ROC.

Number needed to treat or Number needed to harm are ways of expressing the
effectiveness and safety of an intervention in a way that is clinically meaningful. In
general, NNT is always computed with respect to two treatments A and B, with A
typically a drug and B a placebo (in our example above, A is a 5-year treatment with
the hypothetical drug, and B is no treatment). A defined endpoint has to be specified
(in our example: the appearance of colon cancer in the 5 year period). If the
probabilities pA and pB of this endpoint under treatments A and B, respectively, are
known, then the NNT is computed as 1/(pB-pA). The NNT for breast mammography
is 1/285, so 285 mammograms need to be performed to diagnose one breast cancer.
As another example, an NNT of 4 means if 4 patients are treated, only one would
respond.

An NNT of 1 is the most effective and means each patient treated responds, e.g., in
comparing antibiotics with placebo in the eradication of Helicobacter pylori. An NNT of
2 or 3 indicates that a treatment is quite effective (with one patient in 2 or 3 responding to
the treatment). An NNT of 20 to 40 can still be considered clinically effective.[15]
6 Quality of clinical trial publications

Evidence-based medicine attempts to objectively evaluate the quality of clinical research by critically assessing techniques reported by researchers in their publications.

- Trial design considerations. High-quality studies have clearly-defined eligibility criteria, and have minimal missing data.

- Generalizability considerations. Studies may only be applicable to narrowly-defined patient populations, and may not be generalizable to clinical practice.

- Followup. Sufficient time for defined outcomes to occur can influence the study outcomes and the statistical power of a study to detect differences between a treatment and control arm.

- Power. A mathematical calculation can determine if the number of patients is sufficient to detect a difference between treatment arms. A negative study may reflect a lack of benefit, or simply a lack of sufficient quantities of patients to detect a difference.

6.1 Limitations of available evidence

It is recognised that not all evidence is made accessible, that this can limit the effectiveness of any approach, and that effort to reduce various publication and retrieval biases is required.

Failure to publish negative trials is the most obvious gap, and moves to register all trials at the outset, and then to pursue their results, are underway. Changes in publication methods, particularly related to the Web, should reduce the difficulty of obtaining publication for a paper on a trial that concludes it did not prove anything new, including its starting hypothesis.

Treatment effectiveness reported from clinical studies may be higher than that achieved in later routine clinical practice due to the closer patient monitoring during trials that leads to much higher compliance rates.[16]

7 Effectiveness

There are mixed reports about whether evidence-based medicine is effective. Using the classification scheme above --dividing evidence-based medicine into evidence-based guidelines (EBG) and evidence-based individual decision (EBID)-- may explain the conflict. It is difficult to find evidence that EBID improves health care,[citation needed] whereas there is growing evidence of improvements in the efficacy of health care when evidence-based medicine is practiced at the organizational level.[17] One of the virtues of healthcare accreditation is that it offers an opportunity to assess the overall functioning of a hospital or healthcare organisation against the best of the currently-available evidence and to assist the hospital or healthcare organisation to move towards a more effective application of evidence-based medical.
8 Criticism of evidence-based medicine

The neutrality of this section is disputed. Please see the discussion on the talk page. (December 2007)
Please do not remove this message until the dispute is resolved.

Critics of EBM say lack of evidence and lack of benefit are not the same, and that the more data are pooled and aggregated, the more difficult it is to compare the patients in the studies with the patient in front of the doctor — that is, EBM applies to populations, not necessarily to individuals. In *The limits of evidence-based medicine*, Tonelli argues that "the knowledge gained from clinical research does not directly answer the primary clinical question of what is best for the patient at hand." Tonelli suggests that proponents of evidence-based medicine discount the value of clinical experience.

However, many proponents of EBM argue that the best practice of EBM does not discount clinicians' own experience. For instance David Sackett writes that "the practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research".

Although evidence-based medicine is becoming regarded as the "gold standard" for clinical practice and treatment guidelines, there are a number of reasons why most current medical and surgical practices do not have a strong literature base supporting them.

- In some cases, such as in open-heart surgery, conducting randomized, placebo-controlled trials would be unethical, although observational studies may address these problems to some degree.

- Certain groups have been historically under-researched (racial minorities and people with many co-morbid diseases), and thus the literature is sparse in areas that do not allow for generalizing.

- The types of trials considered "gold standard" (i.e. randomized double-blind placebo-controlled trials) may be expensive, so that funding sources play a role in what gets investigated. For example, public authorities may tend to fund preventive medicine studies to improve public health as a whole, while pharmaceutical companies fund studies intended to demonstrate the efficacy and safety of particular drugs.

- The studies that are published in medical journals may not be representative of all the studies that are completed on a given topic (published and unpublished) or may be misleading due to conflicts of interest (i.e. publication bias). Thus the array of evidence available on particular therapies may not be well-represented in the literature. A 2004 statement by the International Committee of Medical Journal Editors that they will refuse to publish clinical trial results if the trial was not recorded publicly at its outset, may help with this, although this has to date still not been actioned.
• The quality of studies performed varies, making it difficult to generalize about the results.

An additional problem is that large randomized controlled trials are useful for examining discrete interventions for carefully defined medical conditions. The more complex the patient population (e.g. severity of condition, co-morbid conditions, etc) in the study, the more difficult it is to assess the treatment effect (i.e., treatment mean - control group mean), relative to the random variation (within group variation of both the treatment and control groups). Because of this, a number of studies obtain non-significant results, either because there is insufficient power to show a difference, or because the groups are not well-enough "controlled". Ironically, the fewer restrictions there are on who can participate in a study (i.e., the greater the generalizability of the results to the type of patient being seen in a real world setting) the less able the study to detect real differences between groups for a given sample size.

Furthermore, evidence-based guidelines do not remove the problem of extrapolation to different populations or longer timeframes. Even if several top-quality studies are available, questions always remain about how far, and to which populations, their results are "generalizable". Furthermore, skepticism about results may always be extended to areas not explicitly covered: for example a drug may influence a "secondary endpoint" such as test result (blood pressure, glucose, or cholesterol levels) without having the power to show that it decreases overall mortality or morbidity in a population.

In managed healthcare systems, evidence-based guideline have been used as a basis for denying insurance coverage for some treatments which are held by the physicians involved to be effective, but of which randomized controlled trials have not yet been published. In some cases, these denials were based upon questions of induction and efficacy as discussed above. For example, if an older generic statin drug has been shown to reduce mortality, is this enough evidence for use of a much more expensive newer statin drug which lowers cholesterol more effectively, but for which mortality reductions have not had time enough to be shown? If a new, costly therapy that works on tumor blood vessels causes two kinds of cancer to go into remission, is it justified as an expense in a third kind of cancer, before this has specifically been proven? Kaiser Permanente did not change its methods of evaluating whether or not new therapies were too "experimental" to be covered, until it was successfully sued twice: once for delaying IVF treatments for two years after the courts determined that scientific evidence of efficacy and safety had reached the "reasonable" stage, and in another case where Kaiser refused to pay for liver transplantation in infants when it had already been shown to be effective in adults, on the basis that use in infants was still "experimental." Here again the problem of induction plays a key role in arguments.

9 Evidence-Based Policy

In his 1996 inaugural speech as President of the Royal Statistical Society, Adrian Smith held out evidence-based medicine as an exemplar for all public policy. He proposed that Evidence-Based Policy should be established for education, prisons and policing policy and all areas of government.
The Users' Guides to the Medical Literature are a series of journal articles[^25], and more recently a comprehensive textbook[^26], that provide invaluable tips for clinicians wishing to incorporate evidence-based medicine into their practices.

10 See also

- Adverse drug reaction
- Adverse effect (medicine)
- Clinical trials with surprising outcomes
- Consensus (medical)
- Epidemiology
- Evidence-based design
- Evidence-based management
- Evidence-based pharmacy in developing countries
- Evidence based policy
- Evidence based practice
- Guideline (medical)
- History of medicine
- Hospital accreditation
- Medical algorithm
- Medical research
- Medicine
- Nocebo
- Placebo (origins of technical term)
- Policy-based evidence making
- Quality control
- Source criticism
- Systematic review
11 References


6. Amazon.com: Evidence-Based Healthcare: J. A. Muir Gray: Books


11. Oxford Centre for Evidence-based Medicine *Levels of Evidence and Grades of Recommendation*


22. http://sfgate.com/cgi-bin/article.cgi?f=/c/a/2006/02/12/MNGD0H7AGT1.DTL&hw=stanford&sn=013&sc=110


