The methodology of evidence-based medicine

Evidence-based medicine (EBM) (sometimes called evidence-based health care or EBHC to broaden its application to allied health care professionals) has been defined as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients." Trisha Greenhalgh and Anna Donald define it more specifically as "the use of mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients."

EBM seeks to assess the strength of the evidence of risks and benefits of treatments (including lack of treatment) and diagnostic tests. This helps clinicians predict whether a treatment will do more good than harm.

Evidence quality can be assessed based on the source type (from meta-analyses and systematic reviews of triple-blind randomized clinical trials with concealment of allocation and no attrition at the top end, down to conventional wisdom at the bottom), as well as other factors including statistical validity, clinical relevance, currency, and peer-review acceptance. EBM recognizes that many aspects of health care depend on individual factors such as quality- and value-of-life judgments, which are only partially subject to quantitative scientific methods. Application of EBM data therefore depends on patient circumstances and preferences, and medical treatment remains subject to input from personal, political, philosophical, ethical, economic, and esthetic values.

Because EBM is used in allied fields, including dentistry, nursing and psychology, evidence-based practice (EBP) is a more encompassing term.

Background and definition

Evidence based medicine (EBM) has evolved from clinical epidemiology, a discipline promoted by the creation of the Journal of Clinical Epidemiology in 1988.
Clinical epidemiology aims to bridge the gap between clinical practice and public health using population health sciences to inform clinical practice. Thus, the methodology that underpins EBM applies methods used in the field of epidemiology to the clinical context (i.e. clinical epidemiology). In essence, EBM incorporates this quantitative (as well as qualitative) methodology in the “art” of clinical practice, so as to make the framework for clinical decisions more objective by better reflecting the evidence from research. By introducing scientific methods – particularly the methods of the population sciences – in clinical decision making, EBM has driven a transformation of clinical practice in medicine.

In 1996 David Sackett wrote that "Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients." This definition, put forward by one of the original proponents of evidence-based medicine, has since been adopted by major organizations, including the Cochrane Collaboration and the Centre for Evidence Based Medicine.

**Evidence-based health service**

An evidence-based health service tends to generate an increase in the competence of health service decision makers and is the practice of evidence-based medicine at the organizational or institutional level. It strengthens the motivation of any health service decision-maker to use scientific methods when making a decision and details of this approach to health services and public health has been discussed in a book titled *evidence-based healthcare & public health*.

**Evidence-based decision making**

The results of population-based research form the foundation of evidence-based medicine. It aims to use the experience of a population of patients reported in the research literature to guide decision making in practice. This practice of evidence-based medicine (EBM), which later evolved to evidence-based health care (EBHC), requires the application of population-based data to the care of an individual
In the past, we have relied on the experience of physicians or other health care workers to make decisions about therapy. In the current information era, this approach would be suboptimal as health care workers rapidly find themselves unable to cope with the influx of a huge variety of new information, from the irrelevant to the very important. Therefore, Evidence-based decision making gradually emerged as a solution to integrate the best research evidence with clinical expertise and patient values and expectations as practiced by the individual health care provider. The concepts and ideas attributed to and labeled collectively as EBM/EBHC have now become a part of daily clinical lives, and health care professionals increasingly hear about evidence-based guidelines, care paths, questions and solutions. The controversy has shifted from whether to implement the new concepts to how to do so sensibly and efficiently, while avoiding potential problems associated with a number of misconceptions about what EBM/EBHC is and what it is not. The EBM/EBHC-related concepts of hierarchy of evidence, meta-analyses, confidence intervals, study design, and so forth are so widespread, that health care professionals have no choice but to become familiar with EBM/EBHC principles and methodologies.

**Process and progress**

The five steps of EBM in practice were first described in 1992 and the experience of delegates attending the 2003 Conference of Evidence-Based Health Care Teachers and Developers was summarized into five steps and published in 2005. This five step process can broadly be categorized as:

1. Translation of uncertainty to an answerable question and includes critical questioning, study design and levels of evidence
2. Systematic retrieval of best evidence available
3. Critical appraisal of evidence for internal validity that can be broken down into aspects regarding:
   - Systematic errors as a result of selection bias, information bias and confounding
4. Application of results in practice
5. Evaluation of performance

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

Ex cathedra statements by the "medical expert" are considered to be the least valid form of evidence. All "experts" are now expected to reference their pronouncements to scientific studies.

The systematic review of published research studies is a major method used for evaluating particular treatments. The Cochrane Collaboration is one of the best-known, respected examples of systematic reviews. Like other collections of systematic reviews, it requires authors to provide a detailed and repeatable plan of their literature search and evaluations of the evidence. Once all the best evidence is assessed, treatment is categorized as "likely to be beneficial", "likely to be harmful", or "evidence did not support either benefit or harm".

A 2007 analysis of 1016 systematic reviews from all 50 Cochrane Collaboration Review Groups found that 44% of the reviews concluded that the intervention was "likely to be beneficial", 7% concluded that the intervention was "likely to be harmful", and 49% concluded that evidence "did not support either benefit or harm". 96% recommended further research. A 2001 review of 160 Cochrane systematic reviews (excluding complementary treatments) 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. A review of 145 alternative medicine Cochrane reviews using the 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.

Generally, there are three distinct, but interdependent, areas of evidence-based medicine. The first is to treat individual patients with acute or chronic pathologies with 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 human-centered, as in a journal club, or 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.

**Assessing the quality of evidence**

Evidence-based medicine categorizes different types of clinical evidence and rates or grades 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, triple-blind, placebo-controlled trials with allocation concealment and complete follow-up involving a homogeneous patient population and medical condition. In contrast, patient testimonials, case reports, and even expert opinion (however some critics have argued that expert opinion "does not belong in the rankings of the quality of empirical evidence because it does not represent a form of empirical evidence" and continue that "expert opinion would seem to be a separate, complex type of knowledge that would not fit into hierarchies otherwise limited to empirical evidence alone.") 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.

**US Preventive Services Task Force (USPSTF)**

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.

**UK National Health Service**

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:

- **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.

**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:

- Level A: Good scientific evidence suggests that the benefits of the clinical service substantially outweigh 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.
GRADE working group

A newer system was developed by the GRADE working group and takes into account more dimensions than just the quality of medical research. It requires users of GRADE who are performing an assessment of the quality of evidence, usually as part of a systematic review, to consider the impact of different factors on their confidence in the results. Authors of GRADE tables, grade the quality of evidence into four levels, on the basis of their confidence in the observed effect (a numerical value) being close to what the true effect is. The confidence value is based on judgements assigned in five different domains in a structured manner. The GRADE working group defines 'quality of evidence' and 'strength of recommendations' based on the quality as two different concepts which are commonly confused with each other.

Systematic reviews may include Randomized Controlled trials that have low risk of bias, or, observational studies that have high risk of bias. In the case of Randomized controlled trials, the quality of evidence is high, but can be downgraded in five different domains.

- Risk of bias: Is a judgement made on the basis of the chance that bias in included studies have influenced the estimate of effect.
- Imprecision: Is a judgement made on the basis of the chance that the observed estimate of effect could change completely.
- Indirectness: Is a judgement made on the basis of the differences in characteristics of how the study was conducted and how the results are actually going to be applied.
- Inconsistency: Is a judgement made on the basis of the variability of results across the included studies.
- Publication bias: Is a judgement made on the basis of the question whether all the research evidence has been taken to account.

In the case of observational studies, the quality of evidence starts of lower and may be upgraded in three domains in addition to being subject to downgrading.
• Large effect: This is when methodologically strong studies show that the observed effect is so large that the probability of it changing completely is less likely.

• Plausible confounding would change the effect: This is when despite the presence of a possible confounding factor which is expected to reduce the observed effect, the effect estimate still shows significant effect.

• Dose response gradient: This is when the intervention used becomes more effective with increasing dose. This suggests that a further increase will likely bring about more effect.

Meaning of the levels of Quality of evidence as per GRADE

• High Quality Evidence: The authors are very confident that the estimate that is presented lies very close to the true value. One could interpret it as: there is very low probability of further research completely changing the presented conclusions.

• Moderate Quality Evidence: The authors are confident that the presented estimate lies close to the true value, but it is also possible that it may be substantially different. One could also interpret it as: further research may completely change the conclusions.

• Low Quality Evidence: The authors are not confident in the effect estimate and the true value may be substantially different. One could interpret it as: further research is likely to change the presented conclusions completely.

• Very low quality Evidence: The authors do not have any confidence in the estimate and it is likely that the true value is substantially different from it. One could interpret it as: New research will most probably change the presented conclusions completely.

Guideline panelists may make **Strong** or **Weak recommendations** on the basis of further criteria. Some of the important criteria are:

• Balance between desirable and undesirable effects (not considering cost)
• Quality of the evidence
• Values and preferences
• Costs (resource utilization)
Despite the differences between systems, the purposes are the same: to guide users of clinical research information on 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.

**Statistical measures**

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 ratio**

Main article: Likelihood ratios in diagnostic testing

The pre-test odds of a particular diagnosis, multiplied by the likelihood ratio, determines the post-test odds. (Odds can be calculated from, and then converted to, the [more familiar] 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.

**AUC-ROC**

The area under the receiver operating 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 / harm**

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 \( p_A \) and \( p_B \) of this endpoint under treatments A and B, respectively, are known, then the NNT is computed as \( 1/(p_B-p_A) \). The NNT for breast mammography is 285; that is, 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.

**Quality of clinical trials**

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 other clinical contexts.
- Follow-up. 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.
Limitations and criticism

Although evidence-based medicine is regarded as the gold standard of conventional clinical practice, limitations or criticisms of its use exist, such as:

- EBM produces quantitative research, especially from randomized controlled trials (RCTs). Accordingly, results may not be relevant for all treatment situations. [http://en.wikipedia.org/wiki/Evidence-based_medicine - cite_note-34](http://en.wikipedia.org/wiki/Evidence-based_medicine - cite_note-34)
- RCTs are expensive, influencing research topics according to the sponsor's interests.
- There is a lag between when the RCT is conducted and when its results are published.
- Certain population segments have been historically under-researched (racial minorities and people with co-morbid diseases), and thus the RCT restricts generalizing.
- Not all evidence from an RCT is made accessible. Treatment effectiveness reported from RCTs may be different than that achieved in routine clinical practice.
- Published studies may not be representative of all studies completed on a given topic (published and unpublished) or may be unreliable due to the different study conditions and variables.
- EBM applies to groups of people but this does not preclude clinicians from using their personal experience in deciding how to treat each patient. One author advises that "the knowledge gained from clinical research does not directly answer the primary clinical question of what is best for the patient at hand" and suggests that evidence-based medicine should not discount the value of clinical experience. Another author stated that "the practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research".
- In areas where frames (contextual and presentational influences on perceptions of reality) obscure facts, hypocognition has been blamed for preventing the practical application of EBM.
Assessing the teaching of evidence-based medicine

Two instruments, the Berlin questionnaire are the most validated. [http://en.wikipedia.org/wiki/Evidence-based_medicine](http://en.wikipedia.org/wiki/Evidence-based_medicine) These questionnaires have been used in diverse settings.

In psychiatry

Standard descriptions about mental illnesses, such as the Diagnostic and Statistical Manual of Mental Disorders, have been criticized as incompletely justified by evidence. In many cases, it is unknown whether a particular "disease" has one, several, or no underlying biological causes, with controversy arising over whether some diseases are merely an artifact of the attempt to construct a unified classification scheme, rather than a "real" disease.

While some experts point to statistics in support of the idea that a lack of adoption of research findings results in suboptimal treatment for many patients, others emphasize the importance of the skill of the practitioner and the customization of the treatment to fit individual needs. There is some controversy over whether mental illnesses are too complex for broad population studies to be helpful.

History

Traces of evidence-based medicine's origin can be found in ancient Greece. 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.

In 1967, the American physician and mathematician Alvan R. Feinstein published his groundbreaking work *Clinical Judgment*, which together with Archie Cochrane’s famous book *Effectiveness and Efficiency* (1972) led to an increasing acceptance of clinical epidemiology and controlled studies during the 70s and 80s and paved the way for the institutional development of EBM in the 90s. Cochrane's efforts were recognized by the fact that an international network for efficacy assessment in medicine - the Cochrane Collaboration – was posthumously named
after him. However, Cochrane himself did not live to see the foundation and institutionalization of the EBM movement and Feinstein became later in his life one of its sharpest methodological critics.

The explicit methodologies used to determine "best evidence" were largely established by the McMaster University research group led by David Sackett and Gordon Guyatt. Guyatt later coined the term “evidence-based” in 1990. The term "evidence-based medicine" first appeared in the medical literature in 1992 in a paper by Guyatt et al. Relevant journals include the British Medical Journal's *Clinical Evidence*, the *Journal Of Evidence-Based Healthcare* and *Evidence Based Health Policy*. All of these were co-founded by Anna Donald, an Australian pioneer in the discipline.

**EBM and ethics of experimental or risky treatments**

Insurance companies in the United States and public insurers in other countries usually wait for drug use approval based on evidence-based guidelines before funding a treatment. Where approval for a drug has been given, and subsequent evidence-based findings indicating that a drug may be less safe than originally anticipated, some insurers in the U.S. have reacted very cautiously and withdrawn funding. For example, an older generic statin drug had been shown to reduce mortality, but a newer and much more expensive statin drug was found to lower cholesterol more effectively. However, evidence came to light about safety concerns with the new drug which caused some insurers to stop funding it even though marketing approval was not withdrawn.

Some people are willing to take their chances to gamble their health on the success of new drugs or old drugs in new situations which may not yet have been fully tested in clinical trials. However insurance companies are reluctant to take on the job of funding such treatments, preferring instead to take the safer route of awaiting the results of clinical testing and leaving the funding of such trials to the manufacturer seeking a license.
Sometimes caution errs in the other direction. 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 in vitro fertilization 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.

**Application of the evidence based model on other public policy matters**

There has been discussion of applying what has been learned from EBM to public 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.


Doi 2012


Evidence-based medicine. Wikipedia the free encyclopedia.


