Theme 3. Introduction to evidence-based medicine

Evidence-based medicine (EBM) or **evidence-based practice** (EBP) aims to apply the best available evidence gained from the scientific method to clinical decision making. It seeks to assess the strength of evidence of the risks and benefits of treatments (including lack of treatment) and diagnostic tests. This helps clinicians to learn whether or not any treatment will do more good than harm.

The essential components of EBM are:

1. ***The evidence,***
2. ***Clinical experience***
3. ***The patient.***

EBP is "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of the individual patient (Fig. 3.1***). It means integrating individual clinical expertise with the best available external clinical evidence from systematic research."* (Dr. David Sackett, 1996).**

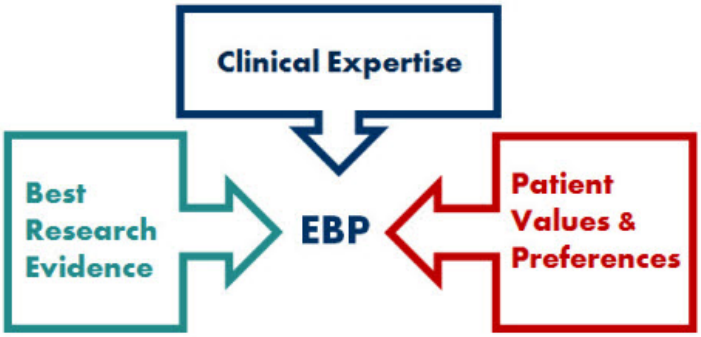

Fig. 3.1 The EBP Process

EBM focuses on the individual patient. The term "evidence-based healthcare" is sometimes used to describe the application of evidence-based approaches at the population level. The decision about groups of patients or populations is based on a combination of three factors:

1. ***Evidence***
2. ***Values***
3. ***Resources***

Evidence quality can be assessed based on the source type (***from meta-analyses and systematic reviews of double-blind, placebo-controlled clinical trials 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/EBP 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 scientific methods. EBP, 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 continues about which outcomes are desirable.

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***.

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.

The Cochrane Library contains high-quality, independent evidence to inform healthcare decision-making. It includes the databases.

Databases included in The Cochrane Library

The Cochrane Database of Systematic Reviews

Full-text systematic reviews and meta-analyses carried out to a common protocol and quality standard by Cochrane Collaboration researchers throughout the world. There are currently 3,625 completed reviews on the database (November 2008), with a further 1,921 protocols for works in process.

The Database of Abstracts of Reviews of Effects (DARE)

This database contains 9,025 quality-appraised abstracts of systematic reviews carried out by other researchers. Protocols and quality standards may vary but this provides a useful starting point if no appropriate Cochrane review has been carried out.

The Cochrane Central Register of Controlled Trials

All Cochrane reviews start with a comprehensive review of the literature. This database provides abstracts of all controlled studies identified by the research groups, as well as further results trawled from electronic databases. There are currently around 550,000 studies included in the register.

The Health Technology Assessment Database

Produced by the Centre for Reviews and Dissemination (CRD) at York University, this database brings together details of 7,528 health technology assessments from around the world. These reviews tend to focus on efficient use of healthcare resources and often include epidemiological and economic assessments.

The NHS Economic Evaluation Database

This also originates from the CRD and focuses purely on those reviews that evaluate the economics of healthcare interventions. It currently contains details on 24,451 such appraisals.

Generally, there are three distinct, but interdependent, areas of evidence based medicine.

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 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.

Ranking 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 (Fig. 3.2). 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.

Fig. 3.2. Evidence pyramid or the hierarchy of study design

Clinical Trials Types

A clinical trial is any form of planned experimental study designed, in general, to evaluate the effect of a new treatment on a clinical outcome in humans.

Clinical trials can be categorised according to their objectives and also the way they are organised.

Clinical trials may be described as ***randomised trials*, *blind trials*, *add-on studies, open label trials* or *randomised double blind placebo controlled trials***. These terms provide information about the way that the trials is organised which may have consequences for the subjects (mostly with regard to the inclusion of a placebo group which is fairly common).

Clinical Trials Types: (In order from strongest to weakest empirical evidence inherent to the design when properly executed.)

1. **Experimental Studies:**
   a. **Randomized Controlled Clinical Trial (RCT).**
   b. **Randomized Cross-Over Clinical Trial.**
   c. **Randomized Controlled Laboratory Study.**
2. **Observational Studies:**
   a. **Cohort (Incidence, Longitudinal) Study.**
   b. **Case-Control Study.**
   c. **Ecologic (Aggregate) Study.**
   d. **Cross-Sectional Study (Prevalence Study).**
   e. **Case Series.**
   f. **Case Report.**

Levels of evidence

Systems to stratify evidence by quality have been developed. The value of evidence can be ranked according to its potential for bias. Table 3.1 shows the classification used by the Scottish Intercollegiate Guidelines Network grading evidence for its clinical guidelines.

Table 3.1.

IGN classification for grading evidence

1++	**High-quality meta-analyses, systematic reviews of RCT*s or RCTs with a very low risk of bias**

1+	Well-conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1–	Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High-quality systematic reviews of case-control or cohort studies High-quality case-control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2+	Well-conducted case-control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal
2–	Case-control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies; for example, case reports, case series
4	Expert opinion

* RCT: randomised controlled trial; SIGN: Scottish Intercollegiate Guidelines Network

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.

Critical appraisal

For any clinician, the real key to assessing the usefulness of a clinical study and interpreting the results to an area of work is through the process of ***critical appraisal***. ***This is a method of assessing and interpreting the evidence by systematically considering its validity, results and relevance to the area of work considered***.

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. Such recommendations use:

- **Level A**: Good scientific evidence suggests that the benefits of the clinical service substantially outweigh the potential risks.
- **Level B**: At least fair scientific evidence suggests that the benefits of the clinical service outweigh the potential risks.
- **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.
- **Level D**: At least fair scientific evidence suggests that the risks of the clinical service outweigh potential benefits.
- **Level E**: Scientific evidence is lacking, of poor quality, or conflicting, such that the risk versus benefit balance cannot be assessed.

The practice of EBM involves a five-step approach

1. Define the problem
2. Find the information you need
3. Critically appraise the information
4. Apply the evidence to the patient
5. Evaluate how much this evidence is useful to the practice

The practice of evidence-based medicine is not an easy thing. It demands an investment in time and resources that some may feel is untenable, given the busy work lives one leads today. Suppose the same clinical question is asked by several doctors. It is unpractical to have each doctor

on his own to work through the five-step approach. This is the reason why various evidence-based methods and resources have been developed to help the busy practitioner.

Evidence-based medicine attempts to express clinical benefits of tests and treatments using mathematical methods.

Clinical decision-making is complex and based upon accurate evaluation of clinical findings using diagnostic tests and reference standard data.

A diagnostic test is used to determine the presence or absence of a disease when a subject shows signs or symptoms of the disease

A screening test identifies asymptomatic individuals who may have the disease The diagnostic test is performed after a positive screening test to establish a definitive diagnosis.

One very important step in the assessment of a patient is the decision of what diagnostic test should be used in determining a patient's needs. In most cases, that decision is based on the physician's experience and knowledge. However, sometimes it is difficult to decide which test is the best for a patient in terms of costs, accuracy, reliability, safety, and efficacy. In this case, there are some simple methods the physician can use to obtain some evidence to support his decision-making.

The term gold standard refers to a benchmark that is the available under reasonable conditions. Indeed, is not the perfect test, but merely the best available one that has a standard with known results.

Examples of some common screening tests:

— Pap smear for cervical dysplasia or cervical cancer
— Fasting blood cholesterol for heart disease
— Fasting blood sugar for diabetes
— Blood pressure for hypertension
— Mammography for breast cancer
— PSA test for prostate cancer
— Fecal occult blood for colon cancer
— Ocular pressure for glaucoma

The basic idea of diagnostic test interpretation is to calculate the probability a patient has a disease under consideration given a certain test result. A 2 by 2 table is used as a mnemonic device. Be sure to label the table with the test results on the left side and the disease status on top as shown here:

	Disease present	**Disease absent**
Test positive	**A True Positive**	**B False Positive**
Test negative	**C False Negative**	**D True Negative**

Sensitivity is the proportion of patients with disease who test positive.

To answer question: How good a test is at detecting people with disease.

In probability notation: $P(T^+|D^+) = A / (A+C)$.

Specificity is the proportion of patients without disease who test negative.

Reflects how good the test is at IDing people without disease.

In probability notation: $P(T^-|D^-) = D / (B + D)$.

Pretest Probability is the estimated likelihood of disease before the test is done. It is the same thing as **prior probability** and is often estimated. If a defined population of patients is being

evaluated, the pretest probability is equal to the **prevalence** of disease in the population. It is the proportion of total patients who have the disease.

In probability notation: $P(D^+) = (A+C) / (A+B+D+C)$.

Pretest probability is the likelihood that a patient has a disorder before the performance of a diagnostic test. If the pretest probability of a specific disorder is *very high*, then a diagnostic test should not be needed to *confirm* the diagnosis. If the pretest probability of a specific disorder is *very low*, then a diagnostic test is not needed to *rule out* a diagnosis.

Sensitivity and specificity describe how well the test discriminates between patients with and without disease. They address a different question than we want answered when evaluating a patient, however. What we usually want to know is: given a certain test result, what is the probability of disease? This is the **predictive value** of the test.

Predictive value of a positive test is the proportion of patients with *positive* tests who have disease.

In probability notation: $P(D^+|T^+) = A / (A+B)$.

This is the same thing as **post-test probability** of disease given a positive test. It measures how well the test rules in disease.

Predictive value of a negative test is the proportion of patients with *negative* tests who *do not* have disease.

In probability notation: $P(D^-|T^-) = D / (D+C)$. It measures how well the test rules out disease. Notice that this is *not* the same as post-test probability of disease given a negative test which is one minus the predicitive value of a negative test.

Likelihood ratio

The likelihood ratio incorporates both the sensitivity and specificity of the test and provides a direct estimate of how much a test result will change the odds of having a disease. The **likelihood ratio for a positive result** (LR+) tells you **how much the odds of the disease increase when a test is positive**. The **likelihood ratio for a negative result** (LR-) tells you **how much the odds of the disease decrease when a test is negative**.

You combine the likelihood ratio with information about

1. the prevalence of the disease,
2. characteristics of your patient pool,
3. information about this particular patient

to determine the post-test odds of disease.

If you want to quantify the effect of a diagnostic test, you have to first provide information about the patient. **You need to specify the pre-test odds: the likelihood that the patient would have a specific disease prior to testing**. The pre-test odds are usually related to the prevalence of the disease, though you might adjust it upwards or downwards depending on characteristics of your overall patient pool or of the individual patient.

The likelihood ratio of a positive test result (LR+) is sensitivity divided by 1- specificity.

$$LR^+ = \frac{sensitivity}{1 - specificity}$$

The likelihood ratio of a negative test result (LR-) is 1- sensitivity divided by specificity.

$$LR^- = \frac{1 - sensitivity}{specificity}$$

LRs >10 or <0.1 cause large changes in likelihood.
LRs 5-10 or 0.1-0.2 cause moderate changes.

LRs 2-5 or 0.2-0.5 cause small changes.
LRs between <2 and 0.5 cause little or no change.

To calculate **post-test odds**, you multiply pre-test odds by the likelihood ratio.

$$odds_{post} = odds_{pre} \times likelihood\ ratio$$

The post-test odds represent the chances that your patient has a disease. It incorporates information about the disease prevalence, the patient pool, and specific patient risk factors (pre-test odds) and information about the diagnostic test itself (the likelihood ratio).

Example

An early test for developmental dysplasia of the hip. The test has *92% sensitivity* and *86% specificity* in boys . *The likelihood ratio for a positive result* from this test is *0.92 / (1-0.86) = 6.6* for boys. *The likelihood ratio for a negative result* from this test is *(1-0.92) / 0.86 = 0.09* (or roughly 1/11).

Suppose one of our patients is a boy with no special risk factors. The diagnostic test is positive. What can we say about the chances that this boy will develop hip dysplasia? *The prevalence of this condition* is *1.5%* in boys. This corresponds to an odds of one to 66. Multiply the odds by the likelihood ratio, you get 6.6 to 66 or roughly 1 to 10. The post test odds of having the disease is 1 to 10 which corresponds to a probability of 9%.

Suppose we had a negative result, but it was with a boy who had a family history of hip dysplasia. Suppose the *family history* would change *the pre-test probability to 25%*. How likely is hip dysplasia, factoring in both the family history and the negative test result? A probability of 25% corresponds to an odds of 1 to 3. The likelihood ratio for a negative result is 0.09 or 1/11. So the post-test odds would be roughly 1 to 33, which corresponds to a probability of 3%.

Notice *that a negative test seems to change things more than a positive test. There are two factors at work here. First, a positive result multiplies the pre-test odds by a factor of only seven whereas a negative result divides the pre-test odds by 11. This means that the test is better at ruling out a condition than ruling it in.*

Second, *the impact of a test is usually greatest for mid-sized probabilities. If a condition is either very rare, or very common, then only a very definitive test is likely to change things much. But mid-sized probabilities (say between 20% and 80%) will change greatly on the basis of even a moderately precise test.*

***EBM resources on the web*:**

- Glossary of EBM Terms: http://www.cebm.utoronto.ca/glossary/
- SUNY Downstate Medical Center EBM Course: http://library.downstate.edu/EBM2/contents.htm
- Users' Guides to Evidence-Based Practice (Centre for Health Evidence) http://www.cche.net/usersguides/main.asp
- McMaster University Health Information Research Unit http://hiru.mcmaster.ca/
- AHRQ – Evidence-Based Practice http://www.ahrq.gov/clinic/epcix.htm

Control questions

1. What is the aim of evidence-based medicine?
2. How did evidence-based medicine appear?
3. What are main areas of evidence based medicine?
4. What clinical trial types are used in evidence based medicine?
5. List levels of evidence.
6. Which statistical characteristics are used in diagnostics tests?

7. What is likelihood ratio?
8. How does evidence-based medicine attempt to evaluate objectively the quality of clinical research?

References

1. Evidence-Based Medicine Working Group Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. JAMA 1992 Nov 4;268(17):2420-5. http://www.cche.net/text/usersguides/ebm.asp
2. Sackett, D. Evidence-based Medicine - What it is and what it isn't. BMJ 1996; 312:71-72. http://www.bmj.com/cgi/content/full/312/7023/71
3. Lang, Eddy. The why and the how of evidence-based medicine. McGill Journal of Medicine 2004 8:90-94. http://www.mclibrary.duke.edu/training/courses/ebm600/lang.pdf.