Evidence-based radiology for diagnostic imaging: What it is and how to practice it

C. García Villar

Unidad Clínica de Radiodiagnóstico, Hospital Universitario Puerta del Mar, Cádiz, Spain

Received 14 November 2010; accepted 4 February 2011

Abstract Evidence-based radiology is defined as the decision that results from integrating clinical information to select the most appropriate imaging test on the basis of the best available evidence, the physician’s experience, and the patient’s expectations. The practice of evidence-based radiology consists of five steps: formulating the question, performing an efficient search of the literature, critically evaluating the literature, applying the results of the search and evaluation while taking into account our experience and the patient’s values, and evaluating the results obtained within our own practice. In diagnostic imaging, the number of resources available for evidence-based radiology is increasing: apart from books, articles, and web pages on this subject, evidence-based radiology is receiving more attention at diagnostic imaging conferences. The principles of evidence-based radiology will help promote the appropriate use of resources, greatly benefiting patients (decreasing the use of examinations that use ionizing radiation), professionals (less overload), and managers (more efficient use of resources).

© 2010 SERAM. Published by Elsevier España, S.L. All rights reserved.
Levels of evidence and grades of recommendation

The levels of evidence were set with the aim to help professionals assess the strength or robustness of the results obtained in a research. It is a hierarchical classification according to the scientific rigor of the design of studies. There are five levels of evidence that range from level 1 (best evidence) to level 5 (least solid evidence). From this classification, levels of recommendation are established concerning a specific health care procedure or intervention: A (highly recommendable), B (recommendable), C (not very recommendable) and D (not recommendable). If from one given disease, different types of questions can be raised that can relate to its etiology and risk factors (what causes this disease?), to its frequency (how common is this disease?), to its diagnosis (has this patient this disease? or what is the best test to confirm or rule out the diagnosis of suspicion?), to its prognosis (which one of these patients will develop this disease?), or to its treatment (what is the best treatment?). Different studies will be designed depending on the type of question to be answered.5,117

Therefore, Oxford’s CEBM5 sets the levels of evidence and grades of recommendation depending on whether the questions to be formulated are regarding treatment, prognosis, diagnosis or economic analysis. Table 1 shows the classification of levels of evidence and grades of recommendation for diagnostic tests.

According to the design of the studies, they can be classified as observational (the researcher is prospectively or retrospectively a spectator of what is happening) and experimental (the researcher controls the factor under study).19

Within the observational studies there are the cohort studies, the case–control studies and the transversal or prevalence studies.10

Normally, an observational study with an outcome variable (disease determined by a reference test or gold standard) and a predictive variable (test under study) is brought up in order to evaluate diagnostic tests. Therefore, in CEBM’s classification (Table 1), the design of study considered the most appropriate in order to compare two diagnostic tests is the cohort study (level of evidence 1b). Although even better than a cohort study is a systematic review (SR) of various cohort studies. A SR performs a systematic search of all cohort studies on a subject, appraises them critically and summarises the outcome according to a set of predetermined criteria.21 A meta-analysis always includes a statistical treatment of data, whereas a SR may not.

Case–control studies can be applied in radiology although their use is not very extended. Cost-effectiveness studies are increasingly common in our field.22
**Table 1** Classification of levels of evidence and grades of recommendation for diagnostic studies according to Oxford’s CEBM.

<table>
<thead>
<tr>
<th>Grade of recommendation</th>
<th>Level of evidence</th>
<th>Type of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (highly recommendable)</td>
<td>1a</td>
<td>Systematic reviews or meta-analyses from level 1 studies, which fit the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>homogeneity criteria&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>A (highly recommendable)</td>
<td>1b</td>
<td>Cohort studies that compare blindly and independently an appropriate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>group of consecutive patients. The diagnostic test under study and the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reference standard are applied to all patients</td>
</tr>
<tr>
<td>A (highly recommendable)</td>
<td>1c</td>
<td>Diagnostic studies with high sensitivity and specificity</td>
</tr>
<tr>
<td>B (recommendable)</td>
<td>2a</td>
<td>Systematic reviews of level 2 studies that fit the homogeneity criteria</td>
</tr>
<tr>
<td>B (recommendable)</td>
<td>2b</td>
<td>Cohort studies that compare blindly and independently a group of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nonconsecutive patients or reduced to a narrow group of individual studies,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to whom the diagnostic test and the reference standard is applied</td>
</tr>
<tr>
<td>B (recommendable)</td>
<td>3a</td>
<td>Systematic reviews that match the homogeneity criteria for level 3 studies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or higher</td>
</tr>
<tr>
<td>B (recommendable)</td>
<td>3b</td>
<td>Blind and independent comparison of an adequate group of nonconsecutive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>patients, not applying the reference standard to all patients</td>
</tr>
<tr>
<td>C (not very recommendable)</td>
<td>4</td>
<td>Case–control studies or reference standard studies not applied</td>
</tr>
<tr>
<td></td>
<td></td>
<td>independently or blindly</td>
</tr>
<tr>
<td>D (not recommendable)</td>
<td>5</td>
<td>Expert’s opinion without critical appraisal of the literature</td>
</tr>
</tbody>
</table>

Source: Centre for Evidence Based Medicine at the University of Oxford website<sup>2</sup>.  
<sup>a</sup> Homogeneity: the results from the different studies must be treated similarly, deleting the possible variations that might exist between individual studies.

**How is the EBR practiced?**

The creation and evolution of Internet have allowed the development of the practice of EBR in a way that any radiologist who has a question can perform an efficient search of the relevant literature, select the studies that provide a higher level of evidence, critically appraise them, apply the conclusions of the study to their daily practice and evaluate the impact of that specific implementation.<sup>23</sup>

The practice of EBR establishes five steps:

**Step 1: formulate a question**

EBR is expected to offer useful solutions to specific clinical problems by achieving valid and current information in order to take decisions on our patients.<sup>24,25</sup>

Formulating a question is the most important step within the process. It requires thorough thinking on it and making it, since it will be the starting point. Normally, within diagnostic radiology the majority of questions relate to the superiority of an imaging modality over another regarding a specific pathology.<sup>3</sup>

When formulating a question it must be divided into little pieces to facilitate the subsequent search of an answer within the literature. A well-structured question consists of four parts<sup>26</sup>:

- Define the patient, group of patients or the problem of interest.
- Define the intervention (in our case the diagnostic test) to be evaluated.
- Compare the test to be evaluated with the one considered the standard reference (gold standard) (if any).
- Define the outcome or result to be evaluated.

Thus, following the acronym PICO (''P'' patient; ''I'' intervention; ''C'' comparison; ''O'' outcome), the question will be ready for the search.

Let’s imagine that we are on call and we get a call from the ER about a 35-year-old patient who after a thoracic trauma of high impact presents central thoracic pain and hypotension. The chest radiograph is normal. Should any other tests be done? Is the chest radiograph enough to diagnose or rule out an aortic rupture or would it be better to perform a CT?

In this example, the four parts of the question would be: ''P'' thoracic trauma; ''I'' chest radiograph; ''C'' computed tomography; ''O'' aortic rupture diagnosis. This clinical setting could also be solved by formulating a multiple-comparison question that would be as follows: in patients with a suspected aortic rupture trauma, are the chest radiograph, the computed tomography and the aortography equivalent for the diagnosis of the presence, severity and level of rupture?

**Step 2: find the best possible evidence**

Once the question has been formulated, we must know where to search for the most relevant literature and how to do it in a fast and efficient way.<sup>27</sup>

In the search for information, we radiologists are faced with an enormous volume of literature on diagnosis that is published not only in journals specialized in radiology but also in journals of other specialties.<sup>28</sup> Where should we start from?

In order to classify the different types of information, Haynes proposed a few years back the model of ''the pyramid of evidence'' in which a hierarchy of all literature available would be established. At the beginning, it had four levels called ''4S''<sup>29</sup>: the foundation was made of the
primary sources (original studies) and in higher levels were the secondary sources (synthesis, synopsis and information systems). Subsequently, this pyramid was redefined into five steps and it now has six levels (model "6S" of the pyramid of evidence) (Fig. 1).

How can this pyramid guide the professionals who must take a decision so that they can find the evidence needed in a fast and safe way?

Normally, secondary sources are better than primary; therefore, the literature that appears in higher steps is considered scientifically better than that of lower levels. The search for evidence must start at the highest possible level of the pyramid.

At the vertex are the support systems for clinical decisions, computerised decision support systems, which are computerised information systems used to integrate clinical and patient information with the aim to take decisions regarding their care. They summarise all the relevant and important evidence on a clinical problem and generate specific recommendations for a given patient after having introduced the details in the program. This system is for example being used in the United Kingdom to manage oral anticoagulation.

In radiodiagnosis there is not at present a clinical decision support system, although there are already some studies that evaluate the impact that its development would have.

The summaries are on the next level. They integrate the information based on evidence regarding a specific problem and are updated regularly. ClinicalEvidence and UpToDate are examples of these summaries. In this group there are also the clinical practice guidelines based on evidence, such as the ones found in The National Guidelines Clearinghouse.

When there is no summary the next step is to search for the synopses of synthesis, which summarise and group the SR data. The synopses of synthesis are, in other words, a systematic review of systematic reviews that meet inclusion and exclusion criteria. They consist of a summary (synopses) of the corresponding SR and are accompanied with comments on the methodological quality of the SR and their applicability in daily practice. These synopses of synthesis can be found in the ACP Journal Club and Evidence-Based Medicine. Another source can be found in the Center for Reviews and Dissemination (CRD) at the University of York, which is a database that in itself contains three databases. One of them is the Database of Abstracts of Reviews of Effects (DARE) that contains structured summaries of RS that meet quality criteria.

If these synopses of synthesis do not exist or are insufficient, then we should turn to the basics of SR, which can be available in EvidenceUpdates and Cochrane Library, and contain synthesis on the effectiveness of health care interventions and some diagnostic tests.

If we cannot find what we are searching for, the next step is the synopses of the original studies. The advantages of a synopsis of an original study over just an original study are that they are briefer, have an added comment and have passed a quality filter and clinical relevance filter.

Finally, if we cannot find the answer in the secondary literature, we must search within the original studies of databases or primary sources such as Pubmed.

Step 3: critically appraise the literature

Once having defined the question to be answered and having identified the relevant literature, we must consider the design of the studies to be critically appraised since we will establish our levels of evidence and grades of recommendation around it (Table 1).

Imagine that we found a SR of cohort studies that concluded that aortic CT is not superior to chest radiography for the diagnosis of an aortic rupture. Since this design represents the highest level of evidence, should we just believe it? Apart from the design, we should also raise other questions to establish if the results and conclusions of the research are valid and applicable. We must therefore critically appraise it.

In an article on diagnosis, the three key questions to be formulated are to determine whether the results of the study are valid, what those results are and if they are applicable to our setting. We must therefore read with careful attention the materials and methods section, and the results section.

Are the results on the study valid?: materials and methods section

Was there a comparison between the test being evaluated and the one considered reference standard? The most correct procedure is to apply the reference standard test on all patients, regardless of the result of the test being evaluated. It is also important to find out whether it existed a blind comparison between both tests, that is, if those who interpreted the results of the test under study were aware of the results of the reference standard test (and vice versa).

Did the test include a proper spectrum of patients? The article must explain how the subjects were recruited and define the inclusion or exclusion criteria followed.

Is the test clearly described? It must be clearly defined as what are a positive result and a negative result. Furthermore, it is especially important in imaging studies to describe the technical aspects in order that the test can be reproduced in another department. Additional aspects to be taken into account are for example exposure to radiation. The
Table 2  Formulae necessary to interpret a diagnostic test.

<table>
<thead>
<tr>
<th>Test to be evaluated</th>
<th>Reference standard</th>
<th>Without the disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Negative</td>
<td>FN</td>
<td>TN</td>
</tr>
<tr>
<td>Sensitivity = TP/(TP + FN)</td>
<td>Specificity = TN/(FP + TN)</td>
<td></td>
</tr>
<tr>
<td>Positive predictive value = TP/(TP + FP)</td>
<td>Negative predictive value = TN/(FN + TN)</td>
<td></td>
</tr>
<tr>
<td>Positive likelihood ratio = S/(1 − E)</td>
<td>Negative likelihood ratio = (1 − S)/E</td>
<td></td>
</tr>
</tbody>
</table>

FN: false negatives; FP: false positives; TN: true negatives; TP: true positives.

justification/optimization concept is important for the radiation protection of patients.47

**What are the results?: results section**

Statistical analysis represents a problem to the majority of clinicians. Although this is a general article, some basic concepts that will be useful when interpreting a study on diagnostic tests must be defined.

Can likelihood ratios be calculated? The studies on diagnostic tests, set out an outcome variable (disease determined by an adequate reference test) and a predictive variable (test under study).18 The aim is to measure the strength of association between both tests using the sensitivity (percentage of people with the disease who have a positive test) and specificity (proportion of healthy people with negative tests) so that the ability of a test to correctly classify a person according to the presence of a disease can be quantified (Table 2).

The positive predictive value (likelihood of disease presence with a positive test) and the negative predictive value (likelihood of healthiness with a negative test) can be calculated from the sensitivity and the specificity.18

Likelihood ratios, which unlike predictive values do not vary depending on the prevalence of the disease, can also be calculated. They can also be positive or negative:

- Positive likelihood ratio: indicates how more probable a positive result is in sick patients than in healthy ones. It should desirably be higher than 1.
- Negative likelihood ratio: indicates how more probable a negative result is in sick patients that in healthy ones. It should desirably be lower than 1.

All these association measures will allow us to interpret the clinical applicability of a test under study.

What is the accuracy of the results? In order to achieve this accuracy we must calculate the confidence intervals, among which the estimate that we are searching for can be found (the exact value cannot be known) with a defined degree of certainty (95%, 99%).

**Can the results be applied to our setting?**

Will the reproducibility of the test and its interpretation be satisfactory in our setting?

We must consider whether the scope of the test is too different from our setting.

- Is the test acceptable in this case?
- We must consider the availability of the test, its risks/discomfort and the costs.
- Will the results of the test change our management?
- From the clinical setting, if the approach is not going to change, the test will not be useful.
- We must consider a treatment threshold and a pre-test probability and post-test probability of disease.

**Step 4: apply**

Once major evidence for the clinical question has been found, the next step is to use our own clinical experience and apply it to the values and preferences of the patient.

Before we decide whether to apply the results of our study to our patient, we must assess48:

- If the diagnostic test can be reproduced in our unit.
- Consider the available alternatives.
- Calculate the pretest probability of our patients, that is, the probability that the patient has the disease (or condition) before performing the test or diagnostic test.
- Check if the patient or the group of patients is similar to the subjects of the study. The main features that can affect our decision include the stage or severity of our patient’s disease. Other factors such as age, sex and comorbidity are also important.
- Weigh up the pros and cons of the diagnostic test for every patient.

In some occasions, the application of the evidence to patients is called “external validity”, in other words, the generalization of the results obtained from our research.

**Step 5: evaluate**

The last step is to evaluate the results within our own clinical practice.49 This can be achieved by evaluating effectiveness and efficiency. This is very important because the results obtained in specialized centres can differ from those obtained locally and therefore need to be evaluated locally.
Evidence-based radiology for diagnostic imaging: What it is and how to practice it

Specific resources of evidence-based Radiology

Books

The authors Medina and Blackmore have written two books on EBR. The first one is then called Evidence-Based Imaging: Optimizing Imaging in Patient Care, which include thirty chapters that evaluate the diagnostic options for different diseases. On the same line, these authors published recently a book about EBR in pediatrics. It can also be useful when writing or critically appraising a study the book Biostatistics for Radiologist, written by Sardanelli and Di Leo, which includes basic definitions on statistics, design of studies and statistical calculations necessary to design and interpret an article on radiology.

Journal articles

More and more journals include articles about different aspects of EBR. The journals Radiology, Seminars in Roentgenology and Academic Radiology, among others, have published a series on the different steps in EBR. SR and meta-analyses on diagnostic tests can be found not only in journals on other specialities but also in those specialized in radiology.

Furthermore, there are more articles that follow the service model question-answer, in other words, they are written like a structured answer to a specific clinical question; Critical Appraisal Topic (CAT). After formulating the clinical question, there is an explanation of the searching strategy that has been used and the articles selected that can best answer it through a summary of results. Finally, a comment on the design of the study and its applicability is included. Thus, they follow an EBM methodology. Although in their preparation they are not as complex as a SR or a meta-analysis, they are a useful tool.

Some examples of CAT can be found in journals such as The Canadian Association of Radiologist Journal, Seminars in Roentgenology and Abdominal Imaging.

Websites

http://www.evidencebasedradiology.net is a site developed by radiologists from Ireland, which provides an up to date practice of EBM. It has a free access section (where the EBR steps are explained) and a private part containing numerous links to articles and other electronic resources.

http://radiologiaevidencia.org is a Spanish site still under construction that is aiming to have more than 1000 references all coming from secondary literature classified by organs and systems. It has a general part and links to other articles and resources. It could be accessed free and will be available in Spanish and English from June 2011.

http://www.evidencebasedradiology.net is a site developed by radiologists from Ireland, which provides an up to date practice of EBM. It has a free access section (where the EBR steps are explained) and a private part containing numerous links to articles and other electronic resources.

http://www.evidencebasedradiology.net is a site developed by radiologists from Ireland, which provides an up to date practice of EBM. It has a free access section (where the EBR steps are explained) and a private part containing numerous links to articles and other electronic resources.

http://www.aur.org is the American College of Radiology, started to develop in 1990 the Appropriateness Criteria, which are guides on clinical practice. In order to prepare these guides, a committee of experts meets to create various clinical settings, searches for relevant literature to answer them, critically appraises them and they finally applies them to that setting. Thus, a table with a list of recommendations is drawn up prioritizing among the different diagnostic tests within each clinical setting. The range runs from 1 (the least recommendable) until 9 (the most recommendable). Apart from the table there is also a summary of the literature consulted from which this decision has been made and the most relevant bibliography. Although methodologically they follow the five steps of EBR, there are some limitations to these criteria. One of them is that there is no explanation about the searching strategy used: the inclusion and exclusion criteria used by the authors to establish a recommendation are not mentioned. Another important limitation is that a critical appraisal of the articles is not carried out, therefore it cannot be ascertained whether the chosen articles represent a good methodology. There is much variability regarding the design of the studies: they may include from a meta-analysis or a SR well developed, to an opinion article from an expert. For all that, although they are good tools we must use them with precaution since they may be seriously biased. Although they have been established for over 20 years, they have not been well disseminated among the medical community and therefore are not known or applied by the majority of clinicians.

Workshops and presentations in congresses

- European Society of Radiology: in the Congress held in Vienna in 2009, an EBR European Working Group was established, subject to The European Network for the Assessment of Imaging in Medicine (EuroAIM), which in turn is part of the European Institute for Biomedical Research (EIBIR). This working group is lead by Professor Francesco Sardanelli (Milan, Italy) and has 42 members from 12 different countries. During this year, they have been analyzing which subjects in radiology have been sufficiently studied by a SR or a meta-analysis and which have not, assessing the quality of these studies. Another one of their objectives is to create a young group with educational purposes.

- European Society of Gastrointestinal and Abdominal Radiologist: they are including EBR workshops in their annual congresses.

Discussion

The EBR presents limitations although at the same time it provides great benefits.

It has been pointed out that the practice of EBR takes much time and energy: it is easier when we are asked to perform a test, with no discussion, instead of performing a search and argue the pros and cons of carrying it out. It may be so the first few times, although if this systematic work is incorporated to our usual practice there can be more material and available resources that might help more colleagues. In some occasions, it can also seem to threaten
the autonomy and freedom of physicians since they must follow "strict" action guidelines far from it. Guidelines are set as recommendations, but above those recommendations is our "individual clinical experience". In other words, if a guideline revising a specific pathology in a defined group of patients establishes a recommendation and our experience tells us the contrary, we are under no obligation to follow it.

A real limitation present is that the field of the imaging tests has not been sufficiently explored by studies that follow EBR principles. However, there is a great effort to identify which are the areas with more deficiencies within radiology in order to avoid this problem. This is a slow and expensive process but it must be pursued little by little without giving up on the way.

All these aspects start from a common premise: training is required in order to learn how to raise questions correctly, carry out efficient search strategies and critically appraise the literature in order to decide whether to apply it or not.

The fast incorporation of new technologies, the increase in the demand of services and the absence of quality scientific evidence have led to an increase in the variability of criteria for the use of specific diagnostic procedures. This variability can cause an overuse in some places and an underuse of the mentioned procedures in other places. All these facts generate doubts about the quality of the care given to patients and cause the need to look for strategies and methods to develop agreed criteria that will help in decision taking on the use of specific procedures in clinical practice.

One of the most used tools in which the EBM methodology is based on is the one created by the Research and Development Corporation together with the University of California Los Angeles, who established the RAND/UCLA Appropriateness Method. This method is based on the synthesis of evidence and on the experts’ opinion and is used to establish whether the execution of a procedure on a specific patient can be appropriate, inappropriate or doubtful within certain clinical circumstances. Not only RAND but also the use of other evaluation methods pretends, among other things, to provide the tools that can be applied in medical practice and have been used in treatment and diagnostic aspects. As for diagnosis, this method has been used to analyse the appropriate use of certain techniques such as endoscopy and colonoscopy, but there have not been found other studies on other types of diagnostic procedures. Although the American College of Radiology uses the methodology of the criteria of appropriate use they do not evaluate a technique but a specific clinical setting in which the different tests that could be performed are evaluated. The diagnostic tests could therefore constitute an area of development for the adequate use of studies.

Conclusions

The principles of EBR can be applied to all aspects of radiology and will help promote the appropriate use of imaging procedures.

A practice based on the principles of EBR contributes with enormous benefits not only to patients (less examinations using ionizing radiations) but also to professionals (less medical overload) and to managers (more efficient use of resources). This is a change of mentality and practice that concerns the whole radiological community, not just one person.

Conflict of interests

The author declares no conflict of interests.

Acknowledgement

I would like to thank Dr. Antonio Martín Mateos, Director of the Unidad Clínica de ORL from the Hospital Universitario Puerta del Mar for having transmitted his enthusiasm for this subject over the years and for his continuous contribution to my work.

References