

Francesco Sardanelli  
Myriam G. Hunink  
Fiona J. Gilbert  
Giovanni Di Leo  
Gabriel P. Krestin

## Evidence-based radiology: why and how?

Received: 5 May 2009  
Revised: 21 June 2009  
Accepted: 1 July 2009  
Published online: 4 September 2009  
© European Society of Radiology 2009

M. G. Hunink  
Program for Health Decision Science,  
Harvard School of Public Health,  
Boston, MA, USA

F. J. Gilbert  
Aberdeen Biomedical Imaging Centre,  
University of Aberdeen,  
Aberdeen, AB25 2ZD, UK

F. Sardanelli (✉) · G. Di Leo  
Dipartimento di Scienze Medico-  
Chirurgiche, Unità di Radiologia,  
IRCCS Policlinico San Donato,  
Università degli Studi di Milano,  
Via Morandi 30,  
20097, San Donato Milanese,  
Milan, Italy  
e-mail: f.sardanelli@grupposandonato.it  
e-mail: francesco.sardanelli@unimi.it  
Tel.: +39-02-52774468  
Fax: +39-02-5279695

M. G. Hunink · G. P. Krestin  
Department of Radiology,  
Erasmus University Medical Center,  
Rotterdam, The Netherlands

M. G. Hunink  
Department of Epidemiology,  
Erasmus University Medical Center,  
Rotterdam, The Netherlands

**Abstract** *Purpose:* To provide an overview of evidence-based medicine (EBM) in relation to radiology and to define a policy for adoption of this principle in the European radiological community. *Results:* Starting from Sackett's definition of EBM we illustrate the top-down and bottom-up approaches to EBM as well as EBM's limitations. Delayed diffusion and peculiar features of evidence-based radiology (EBR) are defined with emphasis on the need to shift from the demonstration of the increasing ability to see more and better, to the demonstration of a significant change in treatment planning or, at best, of a significant gain in patient outcome. The "as low as reasonably achievable" (ALARA) principle is thought as a dimension of

EBR while EBR is proposed as part of the core curriculum of radiology residency. Moreover, we describe the process of health technology assessment in radiology with reference to the six-level scale of hierarchy of studies on diagnostic tests, the main sources of bias in studies on diagnostic performance, and levels of evidence and degrees of recommendations according to the Centre for Evidence-Based Medicine (Oxford, UK) as well as the approach proposed by the GRADE working group. Problems and opportunities offered by evidence-based guidelines in radiology are considered. Finally, we suggest nine points to be actioned by the ESR in order to promote EBR. *Conclusion:* Radiology will benefit greatly from the improvement in practice that will result from adopting this more rigorous approach to all aspects of our work.

**Keywords** Evidence-based medicine · Evidence-based radiology · ALARA · Health technology assessment · Degrees of recommendations

### Introduction

Over the past three decades, the medical community has increasingly supported the principle that clinical practice should be based on the critical evaluation of the results

obtained from medical scientific research. Today this evaluation is facilitated by the Internet which provides instantaneous online access to the most recent publications even before they appear in print form. More and more information is solely accessible through the Internet and

through quality- and relevance-filtered secondary publications (meta-analyses, systematic reviews and guidelines). This principle—a clinical practice based on the results (the evidence) given by the research—has engendered a discipline, *evidence-based medicine* (EBM), which is increasingly expanding into healthcare and bringing a striking change in teaching, learning, clinical practice and decision making by physicians, administrators and policy makers. EBM has entered radiology with a relative delay, but a substantial impact of this approach is expected in the near future.

The aim of this article is to provide an overview of EBM in relation to radiology and to define a policy for this principle in the European radiological community.

### What is EBM?

Evidence-based medicine, also referred to as *evidence-based healthcare* or *evidence-based practice* [1], has been defined as “the systematic application of the best evidence to evaluate the available options and decision making in clinical management and policy settings”, i.e. “integrating clinical expertise with the best available external clinical evidence from research” [2].

This concept is not new. The basis for this way of thinking was developed in the nineteenth century (Pierre C. A. Luis) and during the twentieth century (Ronald A. Fisher, Austin Bradford Hill, Richard Doll and Archie Cochrane). However, it was not until the second half of the last century that the Canadian School led by Gordon Guyatt and Dave L. Sackett at McMaster University (Hamilton, Ontario, Canada) promoted the tendency to guide clinical practice using the best results—the evidence—produced by scientific research [2–4]. This approach was subsequently refined also by the Centre for Evidence-Based Medicine (CEBM) at University of Oxford, England [1, 5].

Dave L. Sackett said that:

Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external evidence from systematic research [6].

A highly attractive alternative but more technical definition, explicitly including diagnosis and investigation, has been proposed by Anna Donald and Trisha Greenhalgh:

Evidence-based medicine is 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 [4].

However, EBM is not only the combination of current best available external evidence and individual clinical expertise. A third factor must be included in EBM: the patient’s values and choice [6]. “It cannot result in slavish, cookbook approaches to individual patient care” [6]. Thus, EBM needs to be the integration of: (i) research evidence, (ii) clinical expertise and (iii) patient’s values and preferences [6–8]. Clinical expertise “decides whether the external evidence applies to the individual patient”, evaluating “how it matches the patient’s clinical state, predicament, and preferences” [6]. A synopsis of this process is given in Fig. 1.

Two general approaches are usually proposed for applying EBM [8–10] (Fig. 2):

- The *top-down* approach, when academic centres, special groups of experts on behalf of medical bodies, or specialized organizations (e.g. the Cochrane collaboration; <http://www.cochrane.org>) provide high-quality primary studies (original research), systematic reviews and meta-analyses, applications of decision analysis, or issue evidence-based guidelines and make efforts for their integration into practice
- The *bottom-up* approach, when practitioners or other physicians working in a day-by-day practice are able “to ask a question, search and appraise the literature, and then apply best current evidence in a local setting”

Both approaches can open a so-called *audit cycle*, when one physician takes a standard and measure her/his own practice against it. However, the top-down approach involves a small number of people considered as experts and does not involve physicians acting at the local level. There is a difference between the production of systematic reviews and meta-analyses (that are welcome as an important source of information by local physicians who want to practice the bottom-up model) and the production

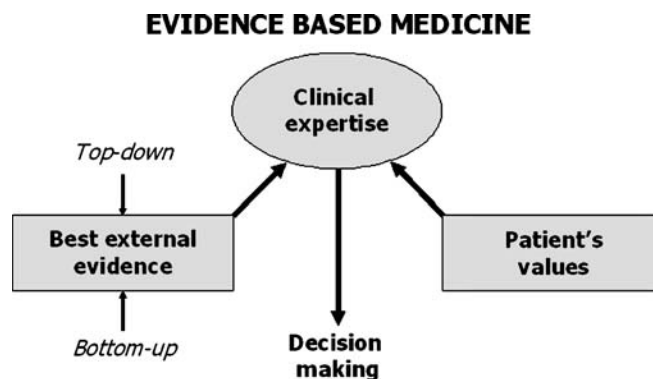
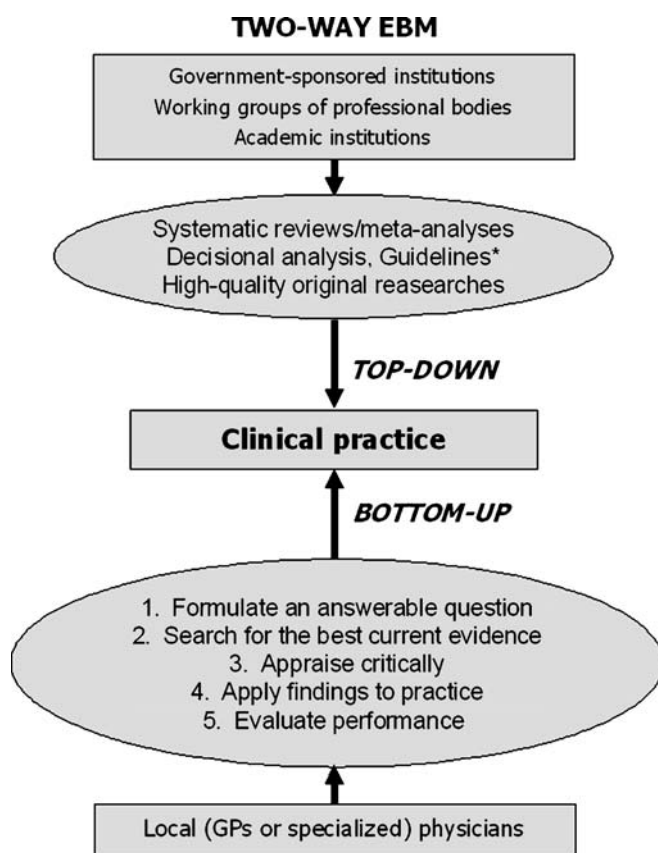


Fig. 1 The general scheme of evidence-based medicine (from ref. [68], p. 3). See Fig. 2 for the top-down and bottom-up approaches to the best external evidence



**Fig. 2** Top-down and bottom-up processes for evidence-based medicine (EBM) (from ref. [68], p. 3). \*Appropriateness criteria are not included in the top-down EBM approach because they are produced on the basis of experts' opinion, even though formalized procedures (such as the modified Delphi protocol) are frequently used and experts commonly base their opinion on systematic reviews and meta-analyses [39]

of guidelines which could be considered as an external cookbook (confused as mandatory standard of practice) by physicians who feel themselves removed from the decision process [10]. On the other hand, the bottom-up approach (which was thought of as the EBM method before the top-down approach [11]) implies a higher level of knowledge of medical research methodology and EBM techniques by local physicians than that demanded by the top-down approach. In either case, a qualitative improvement in patient care is expected. At any rate, clinical expertise must play the pivotal role of integrator of external evidence and patient's values and choice. When decision analyses, meta-analyses and guidelines provide only part of the external evidence found by the local physicians, the two models act together, as hopefully should happen in practice. Moreover, a particular aim of the top-down approach is the identification of knowledge gaps to be filled in by future research. In this way, EBM becomes a method to redirect medical research towards purposes for an improved medical

practice [11]. In fact, one outcome of the production of guidelines should be the identification of the questions still to be answered.

However, EBM is burdened by limitations and beset by criticisms. It has been judged as unproven, very time-consuming (and therefore expensive), narrowing the research agenda and patients' options, facilitating cost cutting, threatening professional autonomy and clinical freedom [6, 8, 12]. On objective evaluation, these criticisms seem to be substantially weak due to the pivotal role attributed to the "individual clinical expertise" by EBM and to the general EBM aim "to maximize the quality and quantity of life for individual patients" which "may raise rather than lower the cost of their care" as pointed out by Sackett in 1996 [6].

Other limitations seem to be more relevant. On the one hand, large clinical areas—radiology being one of them—have not been sufficiently explored by studies according to EBM criteria. On the other hand, real patients can be totally different from those described in the literature, especially due to the presence of comorbidities, making the conclusions of clinical trials not directly applicable. This event is the day-by-day reality in geriatric medicine. The ageing population in Western countries has created a hard benchmark for EBM. These EBM limitations could be related to a general criticism which considers that in the EBM perspective the central feature would be the patient population and not the individual patient [13, 14]. Finally, we should avoid an unbridled enthusiasm for clinical guidelines, especially if they are issued without clarity as to how they were reached or if questionable methods were used [15].

However, all these limitations are due to a still limited EBM development and application rather than as a result of intrinsic problems with EBM. Basically, the value of EBM should be borne in mind, as EBM aims to provide the best choice for the individual patient with the use of probabilistic reasoning. The proponents of EBM are investing significant effort in improving contemporary medicine.

The application of EBM presents a fundamental difficulty. Not only producing scientific evidence but also reading and correctly understanding the medical literature, in particular, syntheses of the best results such as systematic reviews and meta-analyses, requires a basic knowledge of and confidence with the principles and techniques of *descriptive and inferential statistics applied in medical research*. In fact, this is the only way to quantify the uncertainty associated with biological variability and the changes brought about by the patient's disease. It also allows one to manage with the indices and parameters involved in these studies and it is the only mean to judge their quality level. This theoretical background is now emerging as a very important expertise required by any physician of the new millennium.

## Delayed diffusion of EBM in radiology and peculiar features of evidence-based radiology

Radiology is not outside of EBM, as stated by Sackett in 1996: “EBM is not restricted to randomised trials and meta-analyses.[...] To find out about the accuracy of a diagnostic test, we need to find proper cross sectional studies of patients clinically suspected of harbouring the relevant disorder, not a randomised trial” [6]. *Evidence-based radiology* (EBR), also called *evidence-based imaging*, first appeared in the literature only in recent years. We decided to adopt here the terminology *evidence-based radiology* not to restrict the field of interest but to highlight that radiologists are the main addressees of this article. Radiologists are the interpreters of the images and are required to understand the implications of their findings and reports in the context of the available evidence from the literature.

Until 2000, few papers on EBR were published in non-radiological journals [16–20] and in one journal specialized in dentomaxillofacial radiology [21]. From 2001 to 2005, several papers introduced the EBM approach in radiology [2, 22–37]. The first edition of the book *Evidence-Based Imaging* by L. Santiago Medina and C. Craig Blackmore was only published in 2006 [38]. The diffusion of EBM in radiology was therefore delayed. From this viewpoint, radiology is “behind other specialties” [39]. According to Medina and Blackmore: “only around 30% of what constitutes ‘imaging knowledge’ is substantiated by reliable scientific inquiry” [38]. Other authors estimate that less than 10% of standard imaging procedures is supported by sufficient randomized controlled trials, meta-analyses or systematic reviews” [19, 26, 40].

The ‘EBR delay’ is also due to several particular traits of our discipline. The comparison between two diagnostic imaging modalities is vastly different from the well-known comparison between two treatments, typically between a new drug and a placebo or standard care. Thus, the classical design of randomized controlled trials is not the standard for radiological studies. What are the peculiar features of radiology to be considered?

First, the evaluation of the diagnostic performance of imaging modalities must be based on knowledge of the technologies used for image generation and postprocessing. Technical expertise has to be combined with clinical expertise in judging when and how the best available external evidence can be applied in clinical practice. This aspect is as important as the “clinical expertise” (knowledge of indications for an imaging procedure, imaging interpretation and reporting, etc.). Dodd et al. [33] showed the consequences of ignoring a technical detail such as the slice thickness in evaluating the diagnostic performance of magnetic resonance (MR) cholangiopancreatography. Using a 5-mm instead of a 3-mm thickness, the diagnostic performance for the detection of choledocholithiasis changed from 0.57 sensitivity and 1.0 specificity to 0.92

sensitivity and 0.97 specificity [33]. If the results of technically inadequate imaging protocols are included in a meta-analysis, the consequence will be underestimation of the diagnostic performance. Technical expertise is crucial for EBR.

At times progress in clinical imaging is essentially driven by the development of new technology, as was the case for MR imaging at the beginning of the 1980s. However, more frequently, an important gain in spatial or temporal resolution, in signal-to-noise or contrast-to-noise ratio are attained through hardware and/or software innovations in pre-existing technology. This new step broadens the clinical applicability of the technology, as was the case for computed tomography (CT) which evolved from helical single-slice to multidetector row scanners, thus opening the way to cardiac CT and CT angiography of the coronary arteries. To be updated with technological development is a hard task for radiologists and a relevant part of the time not spent with imaging interpretation should be dedicated to the study of new imaging modalities or techniques. In radiological research, each new technology appearing on the market should be tested with studies on its technical performance (image resolution, etc.).

Second, the increasing availability of multiple options in diagnostic imaging should be taken into consideration along with their continuous and sometimes unexpected technological development and sophistication. Thus, the high speed of technological evolution created not only the need to study theory and practical applications of new tools, but also to start again and again with studies on technical performance, reproducibility and diagnostic performance. The faster the advances in technical development, the more difficult it is to do the job in time. This development is often much more rapid than the time required for performing clinical studies for the basic evaluation of diagnostic performance. From this viewpoint, *we are often too late with our assessment studies*.

However, the most important problem to be considered with a new diagnostic technology is that “a balance must be struck between apparent (e.g. diagnostic) benefit and real benefit to the patient” [19]. In fact, a qualitative leap in radiologic research is now expected: from the demonstration of the increasing ability to see more and better, to the demonstration of a significant change in treatment planning or, at best, a significant gain in patient health and/or quality of life—the patient outcome.

Third, we need to perform studies on the reproducibility of the results of imaging modalities (intraobserver, interobserver and interstudy variability), an emergent research area which requires dedicated study design and statistical methods (e.g. Cohen’s kappa statistics, Bland–Altman plots and intraclass correlation coefficients). In fact, if a test shows poor reproducibility, it will never provide good diagnostic performance. Good reproducibility is a necessary (but not sufficient) condition for a test to be useful.

Lastly, we should specifically integrate a new aspect into EBR, the need to avoid unnecessary exposure to ionizing radiation, according to the *as low as reasonably achievable* (ALARA) principle [41–43] and government regulations [44–46]. The ALARA principle might be considered as embedded in radiological technical and clinical expertise. However, in our opinion, it should be regarded as a *fourth dimension of EBR*, due to the increasing relevance of radioprotection issues in radiological thinking. The best external evidence (first dimension) has to be integrated with patient’s values (second dimension) by the radiologist’s technical and clinical expertise (third dimension) taking into the highest consideration the ALARA principle (fourth dimension). A graphical representation of the EBR process, including the ALARA principle, is provided in Fig. 3.

EBR should be considered as part of the core curriculum of radiology residency. Efforts in this direction were made in the USA by the Radiology Residency Review Committee, the American Board of Radiology and the Association of Program Directors in Radiology [39].

### Health technology assessment in radiology and hierarchy of studies on diagnostic tests

In the framework described above, EBM and EBR are based on the possibility of getting the best external evidence for a specific clinical question. Now the problem is: how is this evidence produced? In other words, which methods should be used to demonstrate the value of a diagnostic imaging technology? This field is what we name *health technology assessment* (HTA) and particular features of HTA are important in radiology. Thus, EBR

may exist only if a good radiological HTA is available. As said by William Hollingworth and Jeffery J. Jarvik, “the tricky part, as with boring a tunnel through a mountain, is making sure that the two ends meet in the middle” [11].

According to the UK HTA programme, HTA should answer four fundamental questions on a given technology [11, 47]:

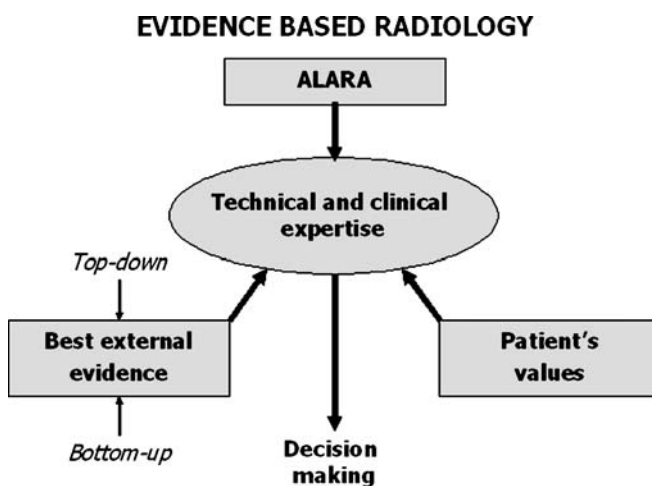
1. Does it work?
2. For whom?
3. At what cost?
4. How does it compare with alternatives?

In this context, an increasing importance has been gained by the use of three different terms. While *efficacy* reflects the performance of medical technology under ideal conditions, *effectiveness* evaluates the same performance under ordinary conditions and *efficiency* measures the cost-effectiveness [48]. In this way the development of a procedure in specialized or academic centres is distinguished by its application to routine clinical practice and from the inevitable role played by the economic costs associated with implementation of a procedure.

To evaluate the impact of the results of studies, i.e. the level at which the HTA was performed, we need a hierarchy of values. Such a hierarchy has been proposed for diagnostic tests and also accepted for diagnostic imaging investigations. During the 1970s, the first classification proposed five levels for the analysis of the diagnostic and therapeutic impact of cranial CT [49]. By the 1990s [50], this classification had evolved into a six-level scale, thanks to the addition of a top level called *societal impact* [51–53]. A description of this scale was more recently presented in the radiologic literature [2, 54].

This six-level scale (Table 1) is currently widely accepted as a foundation for HTA of diagnostic tools. This framework provides an opportunity to assess a technology from differing viewpoints. Studies on technical performance (level 1) are of key importance to the imaging community, and the evaluation of diagnostic performance and reproducibility (level 2) are the basis for adopting a new technique by the radiologists and clinicians. However, radiologists and clinicians are also interested in how an imaging technique impacts patient management (levels 3 and 4) and patient outcomes (level 5), while healthcare providers wish to ascertain the costs and benefits of reimbursing a new technique, from a societal perspective (level 6). Governments are mainly concerned about the societal impact of new technology in comparison with that of other initiatives they may be considering.

Note that this hierarchical order is a one-way logical chain. A positive effect at any level generally implies a positive effect at lower levels but not vice versa [11]. In fact, while a new diagnostic technology with a positive impact on patient’s outcome probably has better technical performance, higher diagnostic accuracy, etc. compared with the standard technology, there is no certainty that a



**Fig. 3** The process of evidence-based radiology (from ref. [68], p. 7). *ALARA* “as low as reasonably achievable”, refers to ionizing radiation exposure

**Table 1** Hierarchy of studies on diagnostic tests

Level	Parameters under investigation
6. Societal impact	Benefit–cost and cost-effectiveness analysis from a social perspective
5. Patient outcomes	Fraction of patients improved with the test compared with that of those improved without the test; difference in morbidity between the patients with the test and those without the test; gain in quality-adjusted life years (QALYs) obtained by the patients with the test compared with those without the test
4. Therapeutic impact	Fraction of patients for whom the test is judged useful for treatment planning or for whom the treatment planning is modified on the basis of the information supplied by the test
3. Diagnostic impact	Fraction of patients for whom the test is judged useful for rendering the diagnosis or for whom the diagnosis is substantially modified after the test; positive and negative likelihood ratios
2. Diagnostic performance	Sensitivity, specificity, accuracy, positive predictive value, negative predictive value and receiver operator characteristic (ROC) analysis; intraobserver, interobserver and interstudy reproducibility
1. Technical performance	Grey-scale range; modulation transfer function; sharpness; spatial resolution, in-plane (line pairs per mm, pixel size) and through-the-plane (slice thickness), integrated in voxel size; signal-to-noise ratio; contrast resolution (contrast-to-noise ratio); time resolution (images/s) etc.

From ref. [68], p. 9

radiologic test with a higher diagnostic accuracy results in better patient outcomes. If we have demonstrated an effective diagnostic performance of a new test (level 2), the impact on a higher level depends on the clinical setting and frequently on conditions external to radiology. It must be demonstrated with specifically designed studies. We might have a very accurate test for the early diagnosis of the disease X. However, if no therapy exists for the disease X, no impact on patient outcomes can be obtained. Alternatively, we may have a new test for the diagnosis of disease Y, but if there is uncertainty on the effectiveness of different treatments of disease Y it may be difficult to prove that the new test is better than the old one. HTA should examine the link between each level and the next in the chain of this hierarchy to establish the clinical value of a radiological test.

Cost-effectiveness can be included in HTA at any level of the hierarchic scale, as cost per examination (level 1), per correct diagnosis (level 2), per invasive test avoided (level 3), per changed therapeutic plan (level 4) and per gained quality-adjusted life expectancy or per saved life (levels 5 and 6) [11]. Recommendations for the performance of cost-effectiveness analyses, however, advocate calculating incremental costs per quality-adjusted life year gained and doing this from the healthcare or societal perspective. Only then are the results comparable and meaningful in setting priorities.

New equipment or a new imaging procedure should have extensive HTA assessment before it is adopted in day-to-day practice. Thereafter follows a period of clinical evaluation where diagnostic accuracy is assessed against a known gold standard. Indeed, the radiological literature is mainly composed of level 1 (*technical performance*) and level 2 (*diagnostic performance*) studies. This is partly inevitable. The evaluation of the technical and diagnostic performance of medical imaging is a typical function of

radiologic research. However, radiologists less frequently study the *diagnostic impact* (level 3) or *therapeutic impact* (level 4) of medical imaging, while *outcome* (level 5) and *societal impact* (level 6) analysis is positively rare in radiologic research. There is a “shortage of coherent and consistent scientific evidence in the radiology literature” to be used for a wide application of EBR [2]. Several papers have recently appeared exploring levels higher than those concerning technical and diagnostic performance, such as the Scottish Low Back Pain Trial, the DAMASK study and others [35, 55–57].

This lack of evidence on patient outcomes is a void also for well-established technologies. This is the case for cranial CT for head injuries, even though in this case the diagnostic information yielded by CT was “obviously so much better than that of alternative strategies that equipoise (genuine uncertainty about the efficacy of a new medical technology) was never present” and “there was an effective treatment for patients with subdural or epidural haematomas—i.e. neurosurgical evacuation” [11]. However, cases like this are very rare, and “in general, new imaging modalities and interventional procedures should be viewed with a degree of healthy skepticism to preserve equipoise until evidence dictates otherwise” [11].

This urgent problem has been recently highlighted by Kuhl et al. for the clinical value of 3.0-T MR imaging. They say: “Although for most neurologic and angiographic applications 3.0 T yields technical advantages compared to 1.5 T, the evidence regarding the added clinical value of high-field strength MR is very limited. There is no paucity of articles that focus on the technical evaluation of neurologic and angiographic applications at 3.0 T. This technology-driven science absorbs a lot of time and energy—energy that is not available for research on the actual clinical utility of high-field MR imaging” [58]. The same can be said for MR spectroscopy of brain tumours

[11, 59], with only one of 96 reviewed articles evaluating the additional value of this technology compared with MR imaging alone [60].

There are genuine reasons for rarely attaining the highest impact levels of efficacy by radiological research. On the one hand, increasingly rapid technologic development forces an endless return to low impact levels. Radiology was judged as the most rapidly evolving specialty in medicine [19]. On the other hand, level 5 and 6 studies entail long performance times, huge economic costs, a high degree of organization and management for longitudinal data gathering on patient outcomes, and often require a randomized study design (the average time for 59 studies in radiation oncology was about 11 years [61]). In this setting, there are two essential needs: full cooperation with clinicians who manage the patient before and after a diagnostic examination and methodological/statistical expertise regarding randomized controlled trials. Radiologists should not be afraid of this, as it is not unfamiliar territory for radiology. More than three decades ago, mammographic screening created a scenario in which the early diagnosis by imaging contributed to a worldwide reduction in mortality from breast cancer, with a high societal impact.

Lastly, alternatives to clinical trials and meta-analyses exist. They are the so-called pragmatic or quasi-experimental studies and decision analysis.

A *pragmatic study* proposes the concurrent development, assessment and implementation of new diagnostic technologies [62]. An empirically based study, preferably using controlled randomization, integrates research aims into clinical practice, using outcome measures reflecting the clinical decision-making process and acceptance of the new test. Outcome measures include: additional imaging studies requested; costs of diagnostic work-up and treatments; confidence in therapeutic decision making; recruitment rate; and patient's outcome measures. Importantly, time is used as the fundamental dimension, e.g. as an explanatory variable in data analysis to model the learning curve, technical developments and interpretation skill. Limitations of this approach can be the need for dedicated and specifically trained personnel and the related economic costs to be covered presumably by governmental agencies [63]. However, this proposal seems to show the potential to answer the dual demand from the faster and faster technology evolution of radiology and the need to attain higher levels of radiological studies, obtaining in a unique approach data on diagnostic confidence, effect on therapy planning, patient outcome measures and cost-effectiveness analysis.

*Decision analysis* integrates the best available evidence and patient values into a mathematical model of possible strategies, their consequences and the associated outcomes. Through analysis of the sensitivity of model results to varying assumptions it can explore the effect of the limited external validity associated with clinical trials [7, 64]. It is a particularly useful tool for evaluating diagnostic tests by

combining intermediate outcome measures such as sensitivity and specificity obtained from published studies and meta-analyses with long-term consequences of true and false, positive and negative outcomes. Different diagnostic or therapeutic alternatives are visually represented by means of a decision tree and dedicated statistical methods are used (e.g. Markov model, Monte Carlo simulation) [7, 65]. This method is typically used for cost-effectiveness analysis.

This approach has been evaluated over a 20-year period from 1985, when the first article concerning cost-effectiveness analysis in medical imaging was published and included 111 radiology-related articles [66]. The average number of studies increased from 1.6 per year (1985–1995) to 9.4 per year (1996–2005). Eighty-six studies were performed to evaluate diagnostic imaging technologies and 25 were performed to evaluate interventional imaging technologies. Ultrasonography (35%), angiography (32%), MR imaging (23%) and CT (20%) were evaluated most frequently. Using a seven-point scale, from 1=low to 7=high, the mean quality score was  $4.2 \pm 1.1$  (mean  $\pm$  standard deviation), without significant improvement over time. Note that quality was measured according to US recommendations for cost-effectiveness analyses, which are not identical to European standards, and the power to demonstrate an improvement was limited [67]. The authors concluded that “improvement in the quality of analyses is needed” [66].

A simple way to appraise the intrinsic difficulty in HTA of radiological procedures is to compare radiological with pharmacological research. After chemical discovery of an active molecule, development, cell and animal testing, phase I and phase II studies are carried out by the industry and very few cooperating clinicians (for phase I and II studies). In this long phase (commonly about 10 years), the majority of academic institutions and large hospitals are not involved. When clinicians are involved in phase III studies, i.e. large randomized trials for registration, the aims are already at level 5 (outcome impact). *Radiologists have to climb 4 levels of impact before reaching the outcome level.* We can imagine a world in which new radiologic procedures are also tested for cost-effectiveness or patient outcome endpoints before entering routine clinical practice, but the real world is different and we have much more technology-driven research from radiologists than radiologist-driven research on technology.

Several countries have well-developed strategies for HTA. In the UK the government funds a HTA programme where topics are prioritised and work is commissioned in relevant areas. In Italy, the Section of Economics in Radiology of the Italian Society of Medical Radiology has connections with the Italian Society of HTA for dedicated research projects. Research groups competitively bid to undertake this work and close monitoring is undertaken to ensure value for money. Radiologists in the USA have formed the American College of Radiologists Imaging Network (ACRIN) ([www.ACRIN.org](http://www.ACRIN.org)) to perform such studies. In Europe, EIBIR (<http://www.EIBIR.org>)

[eibir.org](http://eibir.org)) has formed EuroAIM to undertake such studies. Since 2005, the Royal Australian and New Zealand College of Radiologists developed a program focusing on implementing evidence into practice in radiology: the Quality Use of Diagnostic Imaging (QUDI) program (<http://www.ranzcr.edu.au/qualityprograms/qudi/index.cfm>). The program is fully funded by the Australian federal government and managed by the College.

It is important that new technologies are appropriately assessed before being adopted into practice. However, with a new technology the problem of when to undertake a formal HTA is difficult. Often the technology is still being developed and refined. An early assessment which can take several years might not be relevant if the technology is still undergoing continuing improvement. However, if we wait until a technology is mature then it may already have been widely adopted into practice and so clinicians and radiologists are very reluctant to randomize patients into a study which might deprive them of the new imaging test.

With increasingly expensive technology, new funding mechanisms may be required to allow partnership between industry, the research community and the healthcare system to allow timely, planned introduction of these techniques into practice so the benefit to patients and society can be fully explored before widespread adoption in the healthcare system takes place.

### Sources of bias in studies on diagnostic performance

The quality of HTA studies is determined by the quality of the information provided by the original primary studies on

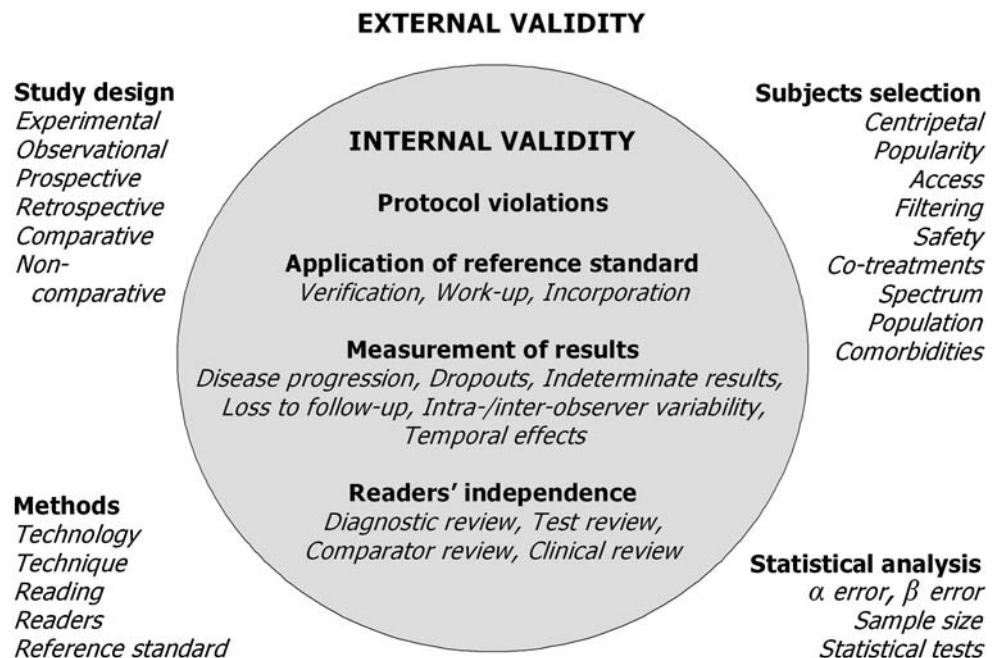
which it is based. Thus, the quality of the original studies is the key point for implementing EBR.

Which are the most important sources of bias for the studies on diagnostic performance? We should distinguish between biases influencing the *external validity* of a study, that is the applicability of its results to clinical practice, and biases influencing the *internal validity* of a study, that is its inherent coherence. Biases influencing the external validity are mainly due to *selection of subjects and choice of techniques leading to lack of generalizability*. Biases influencing the internal validity are due to errors in the methods used in the study (Fig. 4). External and internal validity are related concepts: *the internal validity is a necessary but not sufficient condition in order that a study has external validity* [68].

Thus, all kinds of bias influence the external validity of a study. However, while lack of generalizability has a negative effect on the external validity but the study can retain its internal validity, errors in performing the study have a negative effect primarily on internal validity and secondarily on external validity. The lack of internal validity makes the results themselves not reliable. In this case the question about the external validity (i.e. the application of the results to clinical practice) makes no sense. As a consequence, only the results of a study not flawed by errors in planning and performance can be applied to clinical practice [69].

Several items are present in both planning and performing a study. Consider the reference standard: an error in planning is to choose an inadequate reference standard (imperfect reference standard bias); an error in performing the study is an incorrect use of the planned reference standard. We can go the wrong way either choosing incorrect rules or applying

**Fig. 4** Synopsis of the sources of bias in the studies on diagnostic performance (from ref. [68], p. 166). To apply the results of a study to clinical practice, it must have internal validity (i.e. absence of substantial errors in the methods used in the study) and external validity (i.e. generalizability to other settings). For more details on each of the sources of bias, see refs. [68–70]





right rules incorrectly (but also adding errors in the application of already incorrect rules). There is probably only one right way to do a correct study but infinite ways to introduce errors that make a study useless.

A bias in performing the study can be due to:

1. Defects in protocol application
2. Unforeseen events or events due to insufficient protocol specification
3. Methods defined in the study protocol which implied errors in performing the study

For items 2 and 3, the defects in performing the study depend in some way on error in planning. This does not seem to be the case for item 1. However, if in a study we have many protocol violations, the study protocol was probably theoretically correct but only partially applicable. In other words, biases in performing the study frequently have their ultimate origin in planning error(s).

More details on each of the sources of bias can be found in the articles by Kelly et al. [69] and Sica et al. [70].

## The STARD initiative

The need for an improved quality of studies on diagnostic performance has been present for many years. In 1995 Reid et al. [71] published the results of their analysis on 112 articles regarding diagnostic tests published from 1978 to

1993 in four important medical journals. Overall, over 80% of the studies had relevant biases flawing their estimates of diagnostic performance. In particular: only 27% of the studies reported the disease spectrum of the patients; only 46% of the studies had no work-up bias; only 38% had no review bias; only 11% reported the confidence intervals associated with the point estimates of sensitivity, specificity, predictive values etc.; only 22% reported the frequency of indeterminate results and how they were managed; only 23% of the studies reported a reproducibility of the results.

In this context, a detailed presentation of the rules to be respected for a good-quality original article on diagnostic performance was outlined in an important paper [72], published in 2003 in *Radiology* and also in *Annals of Internal Medicine*, *British Medical Journal*, *Clinical Chemistry*, *Journal of Clinical Microbiology*, *The Lancet* and *Nederlands Tijdschrift voor Geneeskunde*. It is a practical short manual to check the quality of a manuscript or published paper. An extremely useful checklist is provided for authors in order to avoid omitting important information. The paper is entitled "Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative"; STARD is an acronym for standards for reporting of diagnostic accuracy. The authors evaluated 33 papers which proposed a checklist for studies on diagnostic performance. From a list of 75 recommendations, 25 were judged important. The gap to be filled in was testified by Smidt et al. in a study published in 2005 [73]. They evaluated 124 articles on diagnostic performance published

**Table 2** Levels of evidence of studies on diagnostic performance

Level	Study type
1a	Systematic reviews with homogeneous meta-analyses of level-1 studies Multicentre studies, in consecutive patients with a reliable and systematically applied reference standard, of diagnostic criteria previously established by explorative studies
1b	Single-centre studies, in consecutive patients with a reliable and systematically applied reference standard, of diagnostic criteria previously established by explorative studies
1c	Studies of diagnostic examinations with very high sensitivity (absolute <i>snout</i> ) and of diagnostic examinations with very high specificity (absolute <i>spin</i> )*
2a	Systematic reviews with homogeneous meta-analyses of level-2 or higher studies
2b	Explorative studies of diagnostic criteria in cohorts of patients with a reliable and systematically applied reference standard; definition of diagnostic criteria on parts of cohorts or on databases
3a	Systematic reviews with homogeneous meta-analyses of level-3 or higher studies
3b	Studies of non-consecutive patients and/or without systematic application of the reference standard
4	Case-control studies Studies with inadequate or non-independent reference standard
5	Experts' opinions without critical evaluation of the literature

Source: Centre for Evidence-Based Medicine, Oxford, UK (<http://www.cebm.net/index.aspx?o=1025>; accessed 24 Feb 2008); Dodd et al. [33]; with modifications

\*An examination is **snout** when its negative result excludes the possibility of the presence of the disease (when a test has a very high sensitivity, a **negative** result rules **out** the diagnosis); it is instead **spin** when its positive result definitely confirms the presence of the disease (when a test has a very high **specificity**, a positive result rules **in** the diagnosis) [33]

in 12 journals with impact factor of 4 or higher using the 25-item STARD checklist. Only 41% of articles reported more than 50% of STARD items, while no articles reported more than 80%. A flow chart of the study was presented in only two articles. The mean number of reported STARD items was 11.9. Smidt et al. concluded: “Quality of reporting in diagnostic accuracy articles published in 2000 is less than optimal, even in journals with high impact factor” [73].

The relatively low quality of studies on diagnostic performance is a relevant threat to the successful implementation of EBR. Hopefully, the adoption of the STARD requisites will improve the quality of radiological studies but the process seems to be very slow [11], as demonstrated also by the recent study by Wilczynski [74].

Other shared rules are available for articles reporting the results of randomized controlled trials, the CONSORT statement [75], recently extended to trials assessing non-pharmacological treatments [76] or of meta-analyses, the QUOROM statement [77].

In particular, systematic reviews and meta-analyses in radiology should evaluate the study validity for specific issues, as pointed out by Dodd et al. [33]: detailed imaging methods; level of excellence of both imaging and reference standard; adequacy of technology generation; level of ionizing radiation; viewing conditions (hard versus soft copy).

### Levels of evidence

The need to evaluate the relevance of the various studies in relation to the reported level of evidence generated a hierarchy of the levels of evidence based on study type and design.

According to the Centre for Evidence-Based Medicine (Oxford, UK), studies on diagnostic performance can be ranked on a five-level scale, from 1 to 5 (Table 2). Resting on similar scales, four degrees of recommendations, from A to D, can be distinguished (Table 3).

However, we should consider that we have today multiple different classifications of the levels of evidence and of degrees of recommendation. The same degree of recommendation can be represented in different systems using capital letters, Roman or Arabic numerals, etc., generating confusion and possible errors in clinical practice.

A new approach to evidence classification has been recently proposed by the GRADE working group [78] with special attention paid to the definition of standardized criteria for releasing and applying clinical guidelines. The GRADE system states the need for an explicit declaration of the methodological core of a guideline, with particular regard to: quality of evidence, relative importance, risk–benefit balance and value of the incremental benefit for each outcome. This method, apparently complex, finally provides four simple levels of evidence: *high*, when further research is thought unlikely to modify the level of confidence of the estimated effect; *moderate*, when further research is thought likely to modify the level of confidence

**Table 3** Degrees of recommendation

Degree of recommendation	Study type
A	Level 1 studies
B	Consistent level 2 or 3 studies or extrapolations* from level 1 studies
C	Consistent level 4 studies or extrapolations* from level 2 or 3 studies
D	Level 5 studies or low-quality or inconclusive studies of any level

Adapted from results reported by the Centre for Evidence-based Medicine, Oxford, UK (<http://www.cebm.net/index.aspx?o=1025>; accessed 24 Feb 2008)

\**Extrapolation* is the translation of a study to clinical situations different from those of the original study

of the estimated effect and the estimate itself of the effect; *low*, when further research is thought very likely to modify the level of confidence of the estimated effect and the estimate itself of the effect; *very low*, when the estimate of the effect is highly uncertain. Similarly, the risk–benefit ratio is classified as follows: *net benefit*, when the treatment clearly provides more benefits than risks; *moderate*, when, even though the treatment provides important benefits, there is a trade-off in terms of risks; *uncertain*, when we do not know whether the treatment provides more benefits than risks; *lack of net benefit*, when the treatment clearly provides more risks than benefits. The procedure gives four possible recommendations: *do it* or *don't do it*, when we think that the large majority of well-informed people would make this decision; *probably do it* or *probably don't do it*, when we think that the majority of well-informed people would make this decision but a substantial minority would have an opposite opinion. The GRADE system finally differentiates between *strong recommendations* and *weak recommendations*, making the guidelines application to clinical practice easier. Methods for applying the GRADE system to diagnostic tests were recently issued [79]

### Development of evidence-based guidelines in radiology

Clinical guidelines are defined by the Institute of Medicine (Washington DC, USA) as “systematically developed statements to assist practitioners and patient decisions about appropriate health care for specific clinical circumstances” [15, 80, 81]. This purpose is reached by seeking “to make the strengths, weaknesses, and relevance of research findings transparent to clinicians” [82]. Guidelines have potential benefits and harms [15], also from the legal viewpoint [82], and only rigorously developed evidence-based guidelines minimize the potential harms [15, 26]. However, rigorously developed evidence-based guidelines are also not a pure objective product. They imply a decision

process in which opinion is gathered and used, at least because “conclusive evidence exists for relatively few healthcare procedures” and “deriving recommendations only in areas of strong evidence would lead to a guideline of limited scope and applicability” [83]. Thus, a guideline is a sum of evidence and experts’ opinion, taking into account “resource implications and feasibility of interventions” [83]. As a matter of fact, “strong evidence does not always produce a strong recommendation” [83].

Application of a clinical guideline involves interpretation, as is the case for the EBM principle where the best external evidence from research has to be combined with clinical expertise on each specific case and patient. As stated also by the World Health Organization: “Guidelines should provide extensive, critical, and well balanced information on benefits and limitations of the various diagnostic and therapeutic interventions so that the physician may exert the most careful judgment in individual cases” [84].

For over 10 years, the UK Royal College of Radiologists has produced guidance on making the best use of the radiology department and most recently has published MBUR6 [85]. In 2001 there was a fundamental shift in the way these guidelines were developed by the adoption of a more formal approach to the process of gathering and synthesizing evidence. A template was provided to individual radiologists tasked with providing an imaging recommendation so that there was transparency as to how literature was collected and distilled before a guideline was produced. A detailed example of how this was done was published on imaging recommendations on osteomyelitis [36]. The more formal process of gathering evidence by information scientists highlighted the deficiencies of the imaging literature—there were relatively few outcome studies on the impact on patient management or health outcome, a number of studies where the reference standard had been suboptimal and many others where the study methodology had been inadequately described. The requirement for more high-quality imaging studies became apparent. Often new technology becomes part of routine clinical practice prior to extensive evaluation, making outcome studies impossible to perform: e.g. although there is no good evidence of benefit of CT in lung pathology it is inconceivable to attempt a randomized controlled trial comparing CT versus no CT. Neither clinicians nor patients would tolerate being randomized to a no-CT arm. This is the situation for many commonly used imaging procedures where the guidelines have been written by consensus of a panel of experts rather than by evidence from a randomized controlled trials or meta-analyses.

A number of bodies have produced guidelines for imaging—examples include the American College of Radiology [86], the Canadian Association of Radiologists [87], the European Society of Radiology and the European radiological subspecialty societies [88], as well as the radiological societies of individual European countries. While some of these guidelines are based on strong evidence resulting from systematic reviews and meta-analyses, others

were formulated on the sole basis of consensus of expert opinion. Where consensus is used the guidance can be conflicting even when this has been developed in the same country. While there may be cogent reasons why a particular guideline varies from one country to another it is somewhat surprising that there is so much variation when these are supposedly based on evidence. There is a requirement for international cooperation on the gathering and distillation of information to give the imaging community improved understanding of the basis of imaging recommendations. Similarly, given the relative paucity of evidence in certain areas, an international effort to identify and prioritize research requires to be undertaken. This would provide funding bodies the opportunity to collaborate and ensure that a broad range of topics could be addressed across Europe and North America.

We should remember that, as was recently highlighted by Kainberger et al. [26], guidelines are issued but they are commonly accepted by very few clinicians [89] or radiologists [90]. In fact, in the paper by Tigges et al., USA musculoskeletal radiologists, including those of the Society of Skeletal Radiology, were surveyed in 1998 regarding their use of the musculoskeletal appropriateness criteria issued by the American College of Radiology. The response rate was 298/465 (64%) and only 30% of respondents reported using the appropriateness criteria, without difference among organizations or for private practice compared with academic radiologists [90].

---

### **Methods to promote EBR in the European radiological community**

The relative delay in the introduction of EBM in radiology underlines the need for actions aimed at promoting EBR in Europe. This is not an easy task because a cultural change is required. Probably only the new generations of radiologists will fully adopt the new viewpoint in which the patient(s) and the population take the centre stage rather than the images and their quality. The introduction of EBR in the day-by-day practice cannot be solved by a simple series of instructions. A lot of education in research methodology, EBM and HTA in radiology must be done. We suggest several possible lines of action in this direction.

#### **EBR European group**

We propose creating a permanent group of European radiologists dedicated to EBR under the control of the Research Committee of the ESR, in order to coordinate all the lines of action described below. This group could be basically composed of radiologists expert in EBR nominated by ESR subspecialty societies (one for each society), members of the ESR Research Committee and other experts, radiologists and non-radiologists, nominated by the Chairman of the ESR Research Committee.

## Promotion of EBR teaching in postgraduate education in radiology at European universities

The current status of courses in biostatistics and methods for EBR in teaching programs in postgraduate education in diagnostic radiology in European universities should be evaluated.

EBR should be introduced as part of the core curriculum of the residency teaching programs, including the basics of biostatistics applied to radiology, possibly organized as follows:

- First year: sensitivity, specificity, predictive values, overall accuracy and receiver operator characteristic (ROC) analysis; pre-test and post-test probability, Bayes theorem, likelihood ratios and graphs of conditional probability; variables and scales of measurement; normal distribution and confidence intervals; null hypothesis and statistical significance, alpha and beta errors, concept of study power; EBM and EBR principles; self-directed learning: each resident should perform one case-based self-directed bottom-up EBR research project used as a problem-solving approach for decision making related to clinical practice.
- Second year: parametric and non-parametric statistical tests; association and regression; intra- and interobserver reproducibility; study design with particular reference to randomization and randomized controlled trials; study power and sample size calculation; sources of bias in radiologic studies; systematic reviews/meta-analyses; decision analysis and cost-effectiveness analysis for radiological studies; levels of evidence provided by radiological studies; hierarchy of efficacy of radiologic studies; two case-based self-directed bottom-up EBR assignments per resident.
- Third year: two case-based self-directed bottom-up EBR assignments per resident.
- Fourth year: two case-based self-directed bottom-up EBR assignments per resident.
- Introduction of a specific evaluation of the research work, including EBR research and authorship of radiological papers, in the resident's curriculum and for annual and final grading.
- Systematic evaluation of the level of involvement of residents in radiology in radiological research (e.g. number of papers published with one or more residents as authors).

## Intra- and interdepartmental EBR groups in the EuroAIM context

We propose creating intra- and interdepartmental EBR groups, starting with the departments of radiology of academic institutions, teaching and research hospitals. These radiological institutions should be connected in the

EuroAIM network [91] dedicated to EBR in the context of the EIBIR. These groups should be dedicated to:

- Promotion of the adoption of local or international guidelines concerning the use of imaging technology
- Monitoring the effect of the implementation of guidelines in clinical practice
- Day-to-day data collection (clinical data, imaging data, follow-up data) in hospital information systems to analyze the value of imaging technology

Subspecialty groups within EuroAIM and ESR subspecialty societies should collaborate on writing evidence-based guidelines for the appropriate use of imaging technology.

## Redirection of European radiological research

We propose elaborating a strategy to redirect the European radiological research of primary studies (original research) towards purposes defined on the basis of EBR methodology. In particular:

- To change the priority interest of academic radiologists from single-centre studies mainly aimed at estimating diagnostic performance (frequently on relatively small samples) to large multicentre studies, possibly pan-European, including patient randomization and measurement of patient outcomes
- To promote secondary studies (systematic reviews and meta-analyses, decision analyses) on relevant topics regarding the use of diagnostic imaging and interventional radiology
- To collaborate within the context of EuroAIM in performing such studies

## EBR at the ECR

We propose to implement a more detailed grid for ECR abstract rating with an enhanced role of methodology. A model for this could be the process of abstract evaluation adopted by the European Society of Gastrointestinal and Abdominal Radiology in recent years. An explicit mandatory declaration of the study design for each scientific abstract submitted to the ECR could be considered in this context.

We suggest the organization of focused EBR courses during future ECR congresses. Subspecialty committees could propose sessions on particular topics, such as “Evidence-based coronary CT”, “Evidence-based breast MR imaging” and “Evidence-based CT-PET imaging”.

We propose to plan specific ECR sessions dedicated to the presentation and discussion of European guidelines worked out by the ESR subspecialty societies or EuroAIM groups. These sessions could be included in the “professional challenges”.

### Shared rules for developing EBR-based ESR guidelines (a guideline for guidelines)

We propose the adoption of new shared rules for issuing guidelines based on EBR. Examples of these rules can be found at the website of the AGREE Collaboration [92]. In this perspective, a guideline should include:

- Selection and description of the objectives
- Methods for literature searching
- Methods for classification of the evidence extracted from the literature
- Summary of the evidence extracted from the literature
- Practical recommendations, each of them validated by one or more citations and tagged with the level of evidence upon which it is based
- Instructions for application in clinical practice

Before the final release of a guideline, external reviewers should validate its validity (experts in clinical content), clarity (experts in systematic reviews or guidelines development) and applicability (potential users) [83]. Moreover, a date for updating the systematic review which underpins the guideline should be specified [83].

Thus, the usual method consisting of experts' opinions combined with a non-systematic (narrative) review should be overcome. Guidelines officially issued by ESR subspecialty societies should be worked out according EBR-based formal steps defined in a specific ESR document drafted by the ESR-EBR group, discussed with the boards of the subspecialty societies and finally approved by the ESR board.

### Educational programs on EBR-based guidelines by ESR subspecialty societies

We propose organizing courses, seminars and meetings aimed at the diffusion of EBR-based guidelines by the ESR subspecialty societies, as already done by the European Society of Gastrointestinal Radiology, in order also to get feedback on the degree of the theoretical acceptance (first round) and practical acceptance (second round).

### European meetings on EBR-based guidelines with non-radiologists

We propose organizing, as with ESR subspecialty societies, European meetings with other societies of specialists involved in specific clinical fields to present EBR-based rules for the correct request and use of imaging modalities and interventional procedures. The documents offered as a basis for discussion should be the guidelines described in the preceding section.

### Periodical control of the adoption of EBR-based guidelines issued by ESR subspecialty societies

We propose to periodically check the level of adoption of the EBR-based guidelines issued by ESR subspecialty societies by means of surveys which could be conducted in full cooperation with the national societies or their sections or with subspecialty national societies.

## Conclusions

European radiologists need to embrace EBM. Our specialty will benefit greatly from the improvement in practice that will result from this more rigorous approach to all aspects of our work. Wherever radiologists are involved in producing guidelines, refereeing manuscripts, publishing work or undertaking research, cognizance of EBR principles should be maintained. If we can make this step-by-step change in our approach, we will improve radiology for future generations and our patients. EBR should be promoted by ESR and all the European subspecialty societies.

**Acknowledgement** We sincerely thank Professor Yves Menu (Department of Radiology, Saint Antoine Hospital, Paris) for his suggestions regarding the subsection "EBR at the ECR".

## References

1. Malone DE (2007) Evidence-based practice in radiology: an introduction to the series. *Radiology* 242:12–14
2. Evidence-Based Radiology Working Group (2001) Evidence-based radiology: a new approach to the practice of radiology. *Radiology* 220:566–575
3. Greenhalgh T (2006) How to read a paper. The basics of evidence-based medicine, 3rd edn. Blackwell, Oxford, pp ix–xii
4. Greenhalgh T (2006) How to read a paper. The basics of evidence-based medicine, 3rd edn. Blackwell, Oxford, pp 1–3
5. Centre for Evidence-Based Medicine (2008) <http://cebm.net>. Accessed 24 Feb 2008
6. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312:71–72
7. Hunink MGM, Glasziou PP, Siegel JE, Weeks JC, Pliskin JS, Elstein AS, Weinstein MC (2001) Decision making in health and medicine: integrating evidence and values. Cambridge University Press, Cambridge, UK

8. Malone DE, Staunton M (2007) Evidence-based practice in radiology: step 5 (evaluate)—caveats and common questions. *Radiology* 243:319–328
9. Dodd JD (2007) Evidence-based practice in radiology: steps 3 and 4—appraise and apply diagnostic radiology literature. *Radiology* 242:342–354
10. van Beek EJ, Malone DE (2007) Evidence-based practice in radiology education: why and how should we teach it? *Radiology* 243:633–640
11. Hollingworth W, Jarvik JG (2007) Technology assessment in radiology: putting the evidence in evidence-based radiology. *Radiology* 244:31–38
12. Trinder L (2000) A critical appraisal of evidence-based practice. In: Trinder L, Reynolds S (eds) *Evidence-based practice: a critical appraisal*. Blackwell Science, Oxford, pp 212–214
13. Tonelli MR (1998) The philosophical limits of evidence-based medicine. *Acad Med* 73:1234–1240
14. Raymond J, Trop I (2007) The practice of ethics in the era of evidence-based radiology. *Radiology* 244:643–649
15. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J (1999) Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ* 318:527–530
16. Acheson L, Mitchell L (1993) The routine antenatal diagnostic imaging with ultrasound study. The challenge to practice evidence-based obstetrics. *Arch Fam Med* 2:1229–1231
17. No authors listed (1997) Routine ultrasound imaging in pregnancy: how evidence-based are the guidelines? *Int J Technol Assess Health Care* 13:475–477
18. No authors listed (1997) Reports from the British Columbia Office of Health Technology Assessment (BCOHTA). Routine ultrasound imaging in pregnancy: how evidence-based are the guidelines? *Int J Technol Assess Health Care* 13:633–637
19. Dixon AK (1997) Evidence-based diagnostic radiology. *Lancet* 350:509–512
20. Mukerjee A (1999) Towards evidence based emergency medicine: best BETs from the Manchester Royal Infirmary. Magnetic resonance imaging in acute knee haemarthrosis. *J Accid Emerg Med* 16:216–217
21. Liedberg J, Panmekiate S, Petersson A, Rohlin M (1996) Evidence-based evaluation of three imaging methods for the temporomandibular disc. *Dentomaxillofac Radiol* 25:234–241
22. Taïeb S, Vennin P (2001) Evidence-based medicine: towards evidence-based radiology. *J Radiol* 82:887–890
23. Arrivé L, Tubiana JM (2002) “Evidence-based” radiology. *J Radiol* 83:661
24. Bui AA, Taira RK, Dionisio JD et al (2002) Evidence-based radiology: requirements for electronic access. *Acad Radiol* 9:662–669
25. Guillerman RP, Brody AS, Kraus SJ (2002) Evidence-based guidelines for pediatric imaging: the example of the child with possible appendicitis. *Pediatr Ann* 31:629–640
26. Kainberger F, Czemberek H, Frühwald F, Pokieser P, Imhof H (2002) Guidelines and algorithms: strategies for standardization of referral criteria in diagnostic radiology. *Eur Radiol* 12:673–679
27. Bennett JD (2003) Evidence-based radiology problems. Covered stent treatment of an axillary artery pseudoaneurysm: June 2003–June 2004. *Can Assoc Radiol J* 54:140–143
28. Blackmore CC (2003) Evidence-based imaging evaluation of the cervical spine in trauma. *Neuroimaging Clin N Am* 13:283–291
29. Cohen WA, Giauque AP, Hallam DK, Linnau KF, Mann FA (2003) Evidence-based approach to use of MR imaging in acute spinal trauma. *Eur J Radiol* 48:49–60
30. Goergen SK, Fong C, Dalziel K, Fennessy G (2003) Development of an evidence-based guideline for imaging in cervical spine trauma. *Australas Radiol* 47:240–246
31. Medina LS, Aguirre E, Zurakowski D (2003) Introduction to evidence-based imaging. *Neuroimaging Clin N Am* 13:157–165
32. Blackmore CC (2004) Critically assessing the radiology literature. *Acad Radiol* 11:134–140
33. Dodd JD, MacEneaney PM, Malone DE (2004) Evidence-based radiology: how to quickly assess the validity and strength of publications in the diagnostic radiology literature. *Eur Radiol* 14:915–922
34. Erden A (2004) Evidence based radiology. *Tani Girisim Radyol* 10:89–91
35. Gilbert FJ, Grant AM, Gillan MGC (2004) Low back pain: influence of early MR imaging or CT on treatment and outcome—multicenter randomized trial. *Radiology* 231:343–351
36. Matowe L, Gilbert FJ (2004) How to synthesize evidence for imaging guidelines. *Clin Radiol* 59:63–68
37. Giovagnoni A, Ottaviani L, Mensà A et al (2005) Evidence based medicine (EBM) and evidence based radiology (EBR) in the follow-up of the patients after surgery for lung and colon-rectal carcinoma. *Radiol Med* 109:345–357
38. Medina LS, Blackmore CC (2006) *Evidence-based imaging*, 1st edn. Springer, New York
39. Medina LS, Blackmore CC (2007) Evidence-based radiology: review and dissemination. *Radiology* 244:331–336
40. Royal College of Radiologists Working Party (1998) *Making the best use of a department of clinical radiology: guidelines for doctors*, 4th edn. The Royal College of Radiologists, London
41. No authors listed (2004) Proceedings of the second ALARA conference. February 28, 2004. Houston, Texas, USA. *Pediatr Radiol* 34(Suppl 3):S162–S246
42. Prasad KN, Cole WC, Haase GM (2004) Radiation protection in humans: extending the concept of as low as reasonably achievable (ALARA) from dose to biological damage. *Br J Radiol* 77:97–99
43. Semelka RC, Armao DM, Elias J Jr, Huda W (2007) Imaging strategies to reduce the risk of radiation in CT studies, including selective substitution with MRI. *J Magn Reson Imaging* 25:900–909
44. Council of the European Union (1997) Council Directive 97/43/Euratom of 30 June 1997 on health protection of individuals against the dangers of ionizing radiation in relation with medical exposure, and repealing Directive 84/466/Euratom. *J Eur Commun L* 180:22–27 ([http://europa.eu.int/eurlex/en/dat/1997/en\\_397L0043.html](http://europa.eu.int/eurlex/en/dat/1997/en_397L0043.html))
45. Barr HJ, Ohlhaber T, FINDER C (2006) Focusing in on dose reduction: the FDA perspective. *AJR Am J Roentgenol* 186:1716–1717
46. FDA Radiological Health Program (2008) Available via: <http://www.fda.gov/cdrh/radhealth/index.html>. Accessed 24 Feb 2008
47. White SJ, Ashby D, Brown PJ (2000) An introduction to statistical methods for health technology assessment. *Health Technol Assess* 4(i–iv):1–59
48. Hillman BJ, Gatsonis CA (2008) When is the right time to conduct a clinical trial of a diagnostic imaging technology? *Radiology* 248:12–15
49. Fineberg HV, Bauman R, Sosman M (1977) Computerized cranial tomography. Effect on diagnostic and therapeutic plans. *JAMA* 238:224–227
50. Fryback DG, Thornbury JR (1991) The efficacy of diagnostic imaging. *Med Decis Making* 11:88–94
51. Thornbury JR (1994) Clinical efficacy of diagnostic imaging: love it or leave it. *AJR Am J Roentgenol* 162:1–8
52. Mackenzie R, Dixon AK (1995) Measuring the effects of imaging: an evaluative framework. *Clin Radiol* 50:513–518

53. Thornbury JR (1999) Intermediate outcomes: diagnostic and therapeutic impact. *Acad Radiol* 6(suppl 1):S58–S65
54. Sunshine JH, Applegate KE (2004) Technology assessment for radiologists. *Radiology* 230:309–314
55. Brealey SD, DAMASK (Direct Access to Magnetic Resonance Imaging: Assessment for Suspect Knees) Trial Team (2007) Influence of magnetic resonance of the knee on GPs' decisions: a randomised trial. *Br J Gen Pract* 57:622–629
56. Oei EH, Nikken JJ, Ginai AZ, From the Program for the Assessment of Radiological Technology (ART Program) et al (2009) Costs and effectiveness of a brief MRI examination of patients with acute knee injury. *Eur Radiol* 19 (2):409–418
57. Ouwendijk R, de Vries M, Stijnen T, from the Program for the Assessment of Radiological Technology et al (2008) Multicenter randomized controlled trial of the costs and effects of noninvasive diagnostic imaging in patients with peripheral arterial disease: the DIPAD trial. *AJR Am J Roentgenol* 190:1349–1357
58. Kuhl CK, Träber F, Schild HH (2008) Whole-body high-field-strength (3.0-T) MR imaging in clinical practice. Part I. Technical considerations and clinical applications. *Radiology* 246:675–696
59. Jordan HS, Bert RB, Chew P, Kupelnick B, Lau J (2003) Magnetic resonance spectroscopy for brain tumors. Agency for Healthcare Research and Quality, Rockville, p 109
60. Möller-Hartmann W, Herminghaus S, Krings T et al (2002) Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. *Neuroradiology* 44:371–381
61. Soares HP, Kumar A, Daniels S et al (2005) Evaluation of new treatments in radiation oncology: are they better than standard treatments? *JAMA* 293:970–978
62. Hunink MG, Krestin GP (2002) Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* 222:604–614
63. Jarvik JG (2002) Study design for the new millennium: changing how we perform research and practice medicine. *Radiology* 222:593–594
64. Launois R (2003) Economic assessment, a field between clinical research and observational studies. *Bull Cancer* 90:97–104
65. Plevritis SK (2005) Decision analysis and simulation modeling for evaluating diagnostic tests on the basis of patient outcomes. *AJR Am J Roentgenol* 185:581–590
66. Otero HJ, Rybicki FJ, Greenberg D, Neumann PJ (2008) Twenty years of cost-effectiveness analysis in medical imaging: are we improving? *Radiology* 249:917–925
67. Hunink MG (2008) Cost-effectiveness analysis: some clarifications. *Radiology* 249:753–755
68. Sardanelli F, Di Leo G (2008) Biostatistics for radiologists. Springer, Milan, pp 165–179
69. Kelly S, Berry E, Roderick P et al (1997) The identification of bias in studies of the diagnostic performance of imaging modalities. *Br J Radiol* 70:1028–1035
70. Sica GT (2006) Bias in research studies. *Radiology* 238:780–789
71. Reid MC, Lachs MS, Feinstein AR (1995) Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 274:645–651
72. Bossuyt PM, Reitsma JB, Bruns DE et al (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Radiology* 226:24–28
73. Smidt N, Rutjes AW, van der Windt DA et al (2005) Quality of reporting of diagnostic accuracy studies. *Radiology* 235:347–353
74. Wilczynski NL (2008) Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication—before-and-after study. *Radiology* 248:817–823
75. Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357:1191–1194
76. Boutron I, Moher D, Altman DG, Schulz KF, Ravaut P, CONSORT Group (2008) Methods and processes of the CONSORT Group: example of an extension for trials assessing non-pharmacologic treatments. *Ann Intern Med* 148:W60–W66
77. Moher D, Cook DJ, Eastwood S et al (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet* 354:1896–1900
78. Atkins D, Best D, Briss PA, for the GRADE working group et al (2004) Grading quality of evidence and strength of recommendations. *BMJ* 328:1490 (<http://www.bmj.com/cgi/content/full/328/7454/1490>)
79. Schünemann HJ, Oxman AD, Brozek J, for the GRADE Working Group et al (2008) Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 336:1106–1110
80. Field MJ, Lohr KN, eds (1992) Guidelines for clinical practice: from development to use. National Academy, Washington DC
81. Lohr KN (1992) Reasonable expectations: from the Institute of Medicine. Interview by Paul M Schyve. *QRB Qual Rev Bull* 18:393–396
82. Hurwitz B (1999) Legal and political considerations of clinical practice guidelines. *BMJ* 318:661–664
83. Shekelle PG, Woolf SH, Eccles M, Grimshaw J (1999) Clinical guidelines: developing guidelines. *BMJ* 318:593–596
84. Schmidt HG, van der Arend A, Moust JH, Kokx I, Boon L (1993) Influence of tutors' subject-matter expertise on student effort and achievement in problem-based learning. *Acad Med* 68:784–791
85. Royal College of Radiologists (2007) Making the best use of clinical radiology services (MBUR), 6th edn. <http://www.rcr.ac.uk/content.aspx?PageID=995>. Accessed 21 June 2009
86. American College of Radiologists (2009) Guidelines available at: [http://www.acr.org/Secondary/MainMenu/Categories/quality\\_safety/guidelines.aspx](http://www.acr.org/Secondary/MainMenu/Categories/quality_safety/guidelines.aspx). Accessed 21 June 2009
87. Canadian Association of Radiologists (2009) Guidelines available at: <http://www.car.ca/content.aspx?page=Guidelines&spg=home&lang=E&IID=>. Accessed 21 June 2009
88. European Society of Radiology (2009) Guidelines available at: <http://www.myesr.org/cms/website.php?id=%2Fen%2Fsearchresults.htm&cx=014135113606645554273%3A1jgwz0kdufju&cof=FORID%3A11&sa=Search&q=guidelines#1545>. Accessed 21 June 2009
89. Cabana MD, Rand CS, Powe NR et al (1999) Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 282:1458–1465
90. Tigges S, Sutherland D, Manaster BJ (2000) Do radiologists use the American College of Radiology musculoskeletal appropriateness criteria? *AJR Am J Roentgenol* 175:545–547
91. The European Network for the Assessment of Imaging in Medicine (EuroAIM) (2009) <http://www.eibir.org/cms/website.php?id=/de/index/newfilename/newfilename.htm>. Accessed 21 June 2009
92. Appraisal of Guidelines Research and Evaluation (AGREE) <http://www.agreecollaboration.org/instrument/>. Accessed 21 June 2009