Please read these carefully before beginning to score the abstracts assigned to you, even if you have used this system before.
# CONTENTS

<table>
<thead>
<tr>
<th>Overview</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3-7</td>
</tr>
</tbody>
</table>

1) Grading Diagnostic Studies Abstracts | 8-12 |

1.1. ‘Technical Performance’ abstracts | 8,9 |
1.2. ‘Diagnostic Performance’ abstracts | 10,11 |
1.3. ‘Diagnostic Thinking Impact’ abstracts | 12 |
1.4. ‘Therapeutic Impact’ abstracts | 12 |
1.5. ‘Impact on Health’ abstracts | 12 |

2) Grading Radiologic-Pathologic Correlation studies | 13,14 |

3) Grading Case Reports | 15 |

4) Grading Pictorial Essays, Reviews, Molecular & Genetic imaging studies, ‘Other’ studies. | 16 |

5) Grading Interventional Radiology Abstracts | 17,18 |

**Appendix**

*Diagnostic Studies:*
- Background Information & Discussion | 19-22 |

*Secondary Evidence*
- Evidence-Based Reviews | 22 |

*Interventional Radiology* | 23 |
Overview:

These grading categories are being used for three reasons:

1) To provide the reviewers with a set of objective criteria to use as well as their expert opinion when judging ‘scientific’, as opposed to ‘descriptive’ radiology research.
2) To provide researchers with information that they can use to write better abstracts and to make their research programmes stronger and more original.
3) To provide ESGAR with a ‘research profile’ that the Programme Committee can use to make policy decisions about programme content.

GRADING CATEGORIES

1. *Diagnostic Studies:

In order for a test to be fully evaluated, it must be assessed sequentially at all levels. Satisfactory performance at a lower level does not mean the test will perform well at a higher level. Satisfactory performance at all levels establishes the effectiveness of a test for clinical practice.

1.1. **Technical Performance** – These studies deal with how diagnostic tests are done best. They concentrate on innovation and refinement of imaging techniques and assess whether imaging methods reliably result in good quality images that are anatomically representative. Results typically contain ‘p values’. Studies on contrast media and complications should also be included in this section as they typically address basic feasibility and safety issues.

1.2. **Diagnostic Performance** – These studies deal with how well diagnostic tests discriminate between patients who have, and who have not, got the target disorder. The methods typically involve comparison of the diagnostic test in question with a ‘gold standard’ and the results typically contain test properties such as sensitivity, specificity, predictive values, ROC curves etc.

1.3. **Diagnostic Thinking Impact** - These studies look at diagnostic tests in two ways. First, they may specifically assess the extent to which the application of an imaging test displaces alternative technologies (including other imaging tests). Second, they may specifically evaluate the effect of a test on clinical management pathways - for example, by assessing the clinician’s diagnostic thinking before and after receiving the test results. They do not report accuracy, sensitivity, specificity etc (If you have these results in the abstract it is a Diagnostic Performance study).

1.4. **Therapeutic Impact** – Diagnosis is only important in as much as it leads to a change of management or allows a firm, helpful, prognosis. These studies typically evaluate the effect of a test on the management of a target disorder. For example, the management strategy might be decided without the test, the clinician then given the test result and any changes in management recorded / measured.

1.5. **Impact on Health** – These studies measure the effects of diagnostic tests on patient outcomes (including but not restricted to mortality and morbidity from disease).

The cost-effectiveness of an imaging test can be measured at any level.


2. **Pictorial Essays**

These studies describe what the authors consider to be the important, characteristic and / or unusual findings in certain pathological conditions. The authors have typically gathered a series of patients with a particular disease / condition, imaged them with one or more methods and described the findings in their patients.
3. Radiologic-Pathologic Correlation
These studies focus on the correlation between radiology images and autopsy or surgical specimens. There is some potential for overlap with technical performance studies and pictorial essays. However, technical performance studies concentrate more on the technique and methods of test performance than on the pathology in question and pictorial essays do not deal predominately with surgical / autopsy specimens.

4. Case Reports
These studies describe the findings in a patient with a specific condition. This grading category may be used for single-patient reports on diagnostic or interventional topics.

5. Reviews
This includes reviews using ‘Evidence-Based Practice’ or traditional narrative review methods.

6. Interventional Studies:
These studies describe the results from original research about an interventional procedure.

7. Molecular & Genetic Imaging:
These studies examine a molecular or genetic aspect of imaging (e.g. correlation of DNA microarray testing and traditional imaging to look for differences in phenotypes of a disease) or intervention (e.g. image-guided injection of gene therapy).

8. Other
This category includes Educational studies describing teaching / learning / evaluation methods and any diagnostic or interventional study that cannot be included in one of the alternative categories.

The Online Abstract Submission System

Submitters will first choose the abstract type and topic (e.g. Diagnostic, GI – Colorectal). This is to aid scientific session and the e-Poster organisation of accepted abstracts. To finalise the submission of an abstract the submitter will choose a grading category. The ESGAR ‘Research Profile’, graphed below, shows the abstracts submitted to ESGAR 2015 by grading category.
The commonest diagnostic submissions in 2016 are expected to be technical performance, diagnostic performance, pictorial essays and case reports. To simplify entry, the grading system online interface has been designed so that these will be the first 4 choices for submitters. Choices for abstract submission will be labelled D for Diagnostic Radiology, IR for Interventional Radiology and R for reviews (Table).

**TABLE**: Comparison of online submission and grading categories.

<table>
<thead>
<tr>
<th>Online Submission Category</th>
<th>Grading Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic categories are arranged in order of frequency at ESGAR – see graph above</td>
<td>This defines the grading method used for each submission category and provides information for the ESGAR ‘Research Profile’</td>
</tr>
<tr>
<td>D1</td>
<td>Technical Performance</td>
</tr>
<tr>
<td>D2</td>
<td>Diagnostic Performance</td>
</tr>
<tr>
<td>D3</td>
<td>Pictorial Essay</td>
</tr>
<tr>
<td>D4</td>
<td>Case Report</td>
</tr>
<tr>
<td>D5</td>
<td>Diagnostic Thinking Impact</td>
</tr>
<tr>
<td>D6</td>
<td>Therapeutic Impact</td>
</tr>
<tr>
<td>D7</td>
<td>Impact on Health</td>
</tr>
<tr>
<td>D8</td>
<td>Radiologic-Pathologic Correlation</td>
</tr>
<tr>
<td>D9</td>
<td>Molecular and Genetic Imaging</td>
</tr>
<tr>
<td>IR1</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>IR2</td>
<td>Non-randomized controlled trial</td>
</tr>
<tr>
<td>IR3</td>
<td>Case series</td>
</tr>
<tr>
<td>IR4</td>
<td>Case report</td>
</tr>
<tr>
<td>IR5</td>
<td>Laboratory / animal study (interventional)</td>
</tr>
<tr>
<td>R1</td>
<td>Review using 'Evidence-Based’ Methods</td>
</tr>
<tr>
<td>R2</td>
<td>Review using traditional methods</td>
</tr>
<tr>
<td>Other</td>
<td>Other studies (diagnostic &amp; interventional)</td>
</tr>
</tbody>
</table>
Overview Continued:

Issues to be considered when scoring abstracts

1) The Diagnostic Studies sub-categories of Technical Performance, Diagnostic Performance and the grading categories of Radiologic-Pathologic Correlation and Interventional Radiology are categories where objective assessment of Methods and Results is possible. Each abstract will receive a maximum of 10 points from the assessment of:

A. Methods
B. Results
C. Expert opinion of the abstract quality
D. Priority score

2) For Diagnostic Studies sub-categories of Pictorial Essays, Diagnostic Thinking Impact, Therapeutic Impact, Impact on Health and Molecular & Genetic Imaging, objective assessment of Methods and Results is not yet possible. Each abstract will receive a maximum of 10 points from the assessment of:

A. Expert opinion of the abstract quality
B. Priority score

3) For Case Reports, objective assessment is not yet possible. Case reports will not normally be accepted as oral presentations, but will be accepted as posters or rejected. Each abstract will receive a maximum of 8 points from the assessment of:

A. Expert opinion of the abstract quality
B. Whether Eurorad submission is recommended

4) For the grading category 'Other', objective assessment is not yet possible. Each abstract will receive a maximum of 10 points from the assessment of:

A. Expert opinion of the abstract quality
B. Priority score

Scoring
The reviewers will choose one score for each category (A-D or A-B), using the guidelines below. If a reviewer cannot, from the abstract, establish whether one of the objective criteria is met, they should regard it as ‘not present.’ The final score will be the sum of the scores.

The Priority Score: Please do not use the priority score frequently or lightly. It is intended to allow you to highlight early, preliminary research that is in your opinion of significant importance, but which would not have scored highly on the 'Methods and Results’ section or to ‘boost’ an abstract deemed important.

Disagreement with submission category: If you disagree with the submitter’s choice of category this should not cause a problem as the 2016 system only requires you to insert your FINAL or SUMMARY score for each abstract.
**Recommend reject; other comments:**

If you recommend an abstract should be rejected, you will be asked to specify why, using the following letter codes.

**Reject:**

- **V** Inappropriate material for ESGAR (uroradiology, obstetric / gynaecologic imaging)
- **W** No data or data is insufficient to support the conclusions
- **X** Grossly inappropriate methods/ design
- **Y** Gross neglect of the requested abstract format / English language

Other comments – please add to the end of the line allotted to that abstract.

Please note that if a reviewer decides to reject an abstract for reasons **V** or **Y** above, detailed appraisal is not mandatory. The reviewer may enter ‘0’ points for any grading criterion.

The following pages describe the system in detail. You may find it useful to print them off and read them before and during your abstract scoring.

Many thanks again for your assistance with this important task.
**Grading Category 1: Diagnostic Studies**

1.1. **Technical Performance Abstracts**  
(refer to grading catalogue D1)

This is the level that deals with Technical Efficacy [1, 2]. The question is ‘Can we see what we are looking for?’

Each abstract will receive a maximum of 10 points from the assessment of:
A. Methods  
B. Results  
C. Expert opinion of the abstract quality  
D. Priority score

The reviewers will choose one score for each category (A-D), using the guidelines below. ‘Don’t know = not present.’ The final score will be the sum of the scores.

**A) Methods:**

<table>
<thead>
<tr>
<th>Description</th>
<th>Points awarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not enough information to replicate the study technique</td>
<td>0.1</td>
</tr>
<tr>
<td>Almost enough information to replicate the study technique</td>
<td>0.5</td>
</tr>
<tr>
<td>The technique is well described and could be replicated, but a small number of subjects / samples have been presented.</td>
<td>1</td>
</tr>
<tr>
<td>The technique is well described and could be replicated. Subject numbers appear satisfactory.</td>
<td>2</td>
</tr>
</tbody>
</table>

**B) Results:**

<table>
<thead>
<tr>
<th>Description</th>
<th>Points awarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data from a small number of subjects / samples.</td>
<td>0.1</td>
</tr>
<tr>
<td>Raw data from an apparently adequate number of subjects / samples.</td>
<td>0.5</td>
</tr>
<tr>
<td>Statistical analyses have been applied to the data or the abstract states that a statistician was consulted about data analysis. The results are inconclusive.</td>
<td>1</td>
</tr>
<tr>
<td>Statistical analyses have been applied to the data or the abstract states that a statistician was consulted about data analysis. The results are conclusive.</td>
<td>2</td>
</tr>
</tbody>
</table>
Grading Category 1: Diagnostic Studies

1.1 Technical Performance Abstracts (scoring instructions continued)

C) Expert Opinion

Extremely high quality and relevant. Must be presented.  4  
High quality, should be presented.  3  
Acceptable quality, may be presented.  2  
Below average quality.  1  
Reject.  0  

D) Priority Score

Originality  1  
Importance to ESGAR  1  

Important: Please DO NOT use the priority score frequently or lightly. It is intended to allow you to highlight early, preliminary research of significant importance, which would not have scored highly on the ‘Methods and Results’ section or to ‘boost’ an abstract at your discretion.
Grading Category 1: Diagnostic Studies

1.2. Diagnostic Performance Abstracts
(refer to grading catalogue D2)

This is the level that deals with Diagnostic Accuracy Efficacy [1, 2]. The question is ‘How well can we see what we are looking for?’

Each abstract will receive a maximum of 10 points from the assessment of:
C. Methods
D. Results
E. Expert opinion of the abstract quality
F. Priority score

The reviewers will choose one score for each category (A-D), using the guidelines below. ‘Don’t know = not present.’ The final score will be the sum of the scores.

A) Methods:

<table>
<thead>
<tr>
<th>Description</th>
<th>Points awarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent, blinded comparison of an appropriate spectrum of consecutive patients; method well described and all underwent the study test and the same, reliable, gold standard</td>
<td>2</td>
</tr>
<tr>
<td>Independent, blinded comparison of an appropriate spectrum of consecutive patients; method well described and all underwent the study test and more than one gold standard (surgery, histology or imaging and adequate clinical follow-up).</td>
<td>1.6</td>
</tr>
<tr>
<td>As above, but a narrow spectrum of study individuals or non-consecutive patients were studied.</td>
<td>1.2</td>
</tr>
<tr>
<td>As above, but one or more imaging tests were used as a gold standard or the gold standard was not applied to all the patients or the gold standard was not blindly or independently applied.</td>
<td>0.8</td>
</tr>
<tr>
<td>The methods for doing the test are not satisfactorily described</td>
<td>0.4</td>
</tr>
<tr>
<td>I can not establish the above points well enough from the Methods section</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Grading Category 1: Diagnostic Studies

1.2. Diagnostic Performance Abstracts (scoring instructions continued)

B) Results:

Detailed analysis (ROC curves, 'p' values, most properties below plus Likelihood Ratios, another analysis based on the advice of a statistician). 2

Sensitivity, Specificity and Predictive Values are all reported. 95% Confidence Intervals are reported for Sensitivity and Specificity. Disease prevalence in the study population is also reported. 1.6

Sensitivity, Specificity and Predictive Values are reported. 1.2

Sensitivity, Specificity and Accuracy are reported. 0.8

Sensitivity or Specificity only / other weak analysis 0.4

Raw data only 0.2

C) Expert Opinion

Extremely high quality and relevant. Must be presented. 4

High quality, should be presented. 3

Acceptable quality, may be presented. 2

Below average quality. 1

Reject. 0

D) Priority Score

Originality 1

Importance to ESGAR 1

Important: Please DO NOT use the priority score frequently or lightly. It is intended to allow you to highlight early, preliminary research of significant importance, which would not have scored highly on the ‘Methods and Results’ section or to ‘boost’ an abstract at your discretion.
Diagnostic Studies: Abstract Scoring for

1.3. Diagnostic Thinking Impact (refer to grading catalogue D5)
1.4. Therapeutic Impact (refer to grading catalogue D6)
1.5. Impact on Health (refer to grading catalogue D7)

Checklists for more transparent evaluation of the Methods and Results sections of these types of study are not available. These categories of abstract will be scored in the traditional way (expert opinion alone!)

Each abstract will receive a maximum of 10 points from the assessment of:
A. Expert opinion of the abstract quality
B. Priority score

The reviewers will choose one score for each category (A-B), using the guidelines below. The final score will be the sum of the scores.

A) Expert Opinion

Extremely high quality and relevant. Must be presented. 8
High quality, should be presented. 6
Acceptable quality, may be presented. 4
Below average quality. 2
Reject. 0

B) Priority Score

Originality 1
Importance to ESGAR 1

Important: Please DO NOT use the priority score frequently or lightly. It is intended to allow you to highlight early, preliminary research of significant importance, which would not have scored highly on the ‘Methods and Results’ section or to ‘boost’ an abstract at your discretion.
Grading Category 3: Radiologic-Pathologic Correlation Abstracts
(refer to grading catalogue D8)

This topic separates out studies which attempt to correlate abnormal radiology images with gross pathology as defined by surgical resection or autopsy. The focus is on the understanding of the pathologic basis of the abnormalities rather than on the imaging technology. These abstracts will receive a maximum of 10 points from the assessment of:

A) Methods
B) Results
C) Expert opinion of the abstract quality
D) Priority score

The reviewers will choose one score for each category (A-D), using the guidelines below. ‘Don’t know = not present.’ The final score will be the sum of the scores.

A) Methods:

<table>
<thead>
<tr>
<th>Points awarded</th>
<th>Major assumptions made without pathologic correlation.</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Little or no information on rad-path technique; or major assumptions made on the basis of correlation only with other imaging methods. Author’s assumptions may not be correct, as they are not based on pathologic proof.</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Some detail given about rad-path correlation technique AND / OR good technique but low subject numbers</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Meticulous rad-path correlation described in Methods. Satisfactory subject numbers.</td>
<td>2</td>
</tr>
</tbody>
</table>

B) Results:

<table>
<thead>
<tr>
<th>Points awarded</th>
<th>This report is anecdotal and not generalizable to other patient populations</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Principles of radiologic-pathologic correlation are applied to the interpretation of radiologic studies</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Satisfies above criterion and applies an understanding of the clinical and pathologic implications of the radiologic appearances to image interpretation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Satisfies above criteria AND refines the differential diagnosis in various entities and organ systems, based on specific, cross-related features</td>
<td>2</td>
</tr>
</tbody>
</table>

Grading Category 2 – Radiologic-Pathologic Correlation – Page 1 of 2
Grading Category 3: Radiologic-Pathologic Correlation Abstracts

Abstract scoring continued

C) Expert Opinion

Extremely high quality and relevant. Must be presented. 4
High quality, should be presented. 3
Acceptable quality, may be presented. 2
Below average quality. 1
Reject. 0

D) Priority Score

Originality 1
Importance to ESGAR 1

Important: Please DO NOT use the priority score frequently or lightly. It is intended to allow you to highlight early, preliminary research of significant importance, which would not have scored highly on the ‘Methods and Results’ section or to ‘boost’ an abstract at your discretion.
Grading Category 4: CASE REPORTS
(refer to grading catalogue D4)

These are abstracts with 1 case representing a particular pathology. It is hoped that this section of the ESGAR meeting can provide cases for the Eurorad database (see below). As they are anecdotal and descriptive, their maximum potential score is < scientific or educational submissions.

Please score case reports as follows:

Each abstract will receive a maximum of 8 points from the assessment of:

A. Expert opinion of the abstract quality
   B. Priority score

The reviewers will choose one score for each category (A-B), using the guidelines below. The final score will be the sum of the scores.

A) Expert Opinion

Extremely high quality and relevant. Must be presented. 7
High quality, should be presented. 5
Acceptable quality, may be presented. 3
Below average quality. 2
Reject. 0

B) Eurorad Score

The authors should be asked to submit abstracts to Eurorad B

Abstracts recommended for Eurorad should have one or more of the following characteristics:

– Contain descriptions of unusual cases.
– Describe imaging features of pathological conditions especially well.
– Describe imaging techniques optimally demonstrating pathological conditions.

Accepted case reports will be presented at ESGAR as e-Posters. Authors will be informed that submission to Eurorad has been recommended by the reviewers.
Grading Categories 2, 5, 7, 8:

Pictorial Essays (refer to grading catalogue D3)
Reviews (refer to grading catalogues R1 + R2)
Molecular & Genetic (refer to grading catalogue D9)
‘Other’ (refer to grading catalogue ‘Other’)

These categories of study will be graded using expert opinion. They include pictorial essays, in which authors identify a group of patients with a particular disease and describe the imaging findings in that disease using one or more imaging methods; reviews using Evidence-Based methodology; Molecular & Genetic Imaging or Interventional research and any diagnostic and interventional study that cannot be included in one of the alternative categories (e.g. educational studies describing teaching / learning / evaluation methods). Each abstract will receive a maximum of 10 points from the assessment of:

A. Expert opinion of the abstract quality
B. Priority score

The reviewers will choose one score for each category (A-B), using the guidelines below. The final score will be the sum of the scores.

A) Expert Opinion

Extremely high quality and relevant. Must be presented.  8
High quality, should be presented.                  6
Acceptable quality, may be presented.             4
Below average quality.                           2
Reject.                                         0

B) Priority Score

Originality                                     1
Importance to ESGAR                             1

Important: Please DO NOT use the priority score frequently or lightly. It is intended to allow you to highlight work of significant importance or to ‘boost’ an abstract at your discretion.
Grading Category 6: Interventional Radiology Abstracts
(refer to grading catalogues IR 1 – IR 5)

Studies focusing on image-guided therapy will be included in this category. Each abstract will receive a maximum of 10 points from the assessment of:

A. Methods
B. Results
C. Expert opinion of the abstract quality
D. Priority score

The reviewers will choose one score for each category (A-D), using the guidelines below. ‘Don’t know = not present.’ The final score will be the sum of the scores.

A) Methods:

<table>
<thead>
<tr>
<th>Points awarded</th>
<th>A. Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Randomized Controlled Trial</td>
</tr>
<tr>
<td>1.6</td>
<td>Prospective Controlled Trial</td>
</tr>
<tr>
<td>1.2</td>
<td>Retrospective Controlled Trial or it is not clear from the abstract if a trial is prospective or retrospective</td>
</tr>
<tr>
<td>0.8</td>
<td>Case Series</td>
</tr>
<tr>
<td>0.4</td>
<td>Laboratory Research (including animal studies)</td>
</tr>
</tbody>
</table>

B) Results:

Procedural Benefit (Choose one score between 0.4 and 1.6)

A controlled study with either: a) Detailed statistical analysis (e.g. hazard ratio or odds ratio with 95% confidence intervals, Kaplan-Meier life table analysis) or b) Relative Risk and Absolute Risk with Numbers Needed to Treat or Harm and 95% Confidence Intervals calculated from comparison of Experimental and control group event rates. 1.6

Hypothesis testing used to calculate ‘p’ values from comparison of experimental and control group event rates. 1.2

Control Group event rate compared with Experimental Group event rate or statistical analysis used to extract significant data from case series. 0.8

Raw data from case series 0.4
Grading Category 6: Interventional Radiology Abstracts

Abstract scoring continued

Procedural Harm (Choose one score between – 0.2 and + 0.4)

Technique claimed to be safe with > 300 patients (at most, a 1% level of complications could be missed). 0.4

Technique claimed to be safe with > 100 < 300 patients (at most, a 3% level of major complications could be missed). 0.2

Technique claimed to be safe with 50 – 100 patients OR no comment about safety. 0

Technique claimed to be safe with < 50 patients (a 6% level of major complications could be missed) -0.2

C) Expert Opinion

Extremely high quality and relevant. Must be presented. 4

High quality, should be presented. 3

Acceptable quality, may be presented. 2

Below average quality. 1

Reject. 0

D) Priority Score

Originality 1

Importance to ESGAR 1

Important: Please DO NOT use the priority score frequently or lightly. It is intended to allow you to highlight early, preliminary research of significant importance, which would not have scored highly on the ‘Methods and Results’ section or to ‘boost’ an abstract at your discretion.
This section contains some background information and references which should clarify newer aspects of the grading system.

**Grading Category 1 – Diagnostic Studies**

Technology assessment is defined as any process of examining and reporting properties of a medical technology used in health care, such as safety, feasibility, efficacy and the indications for use, cost-effectiveness and consequences of the technology. This broad range has been sub-divided by several authors in hierarchical models that are based on a global view of radiology [1-3]. For an imaging examination to be successful at a higher level, it must first be successful at lower levels. The reverse is not necessarily true. The levels suggested by Mackenzie and Dixon [3] are listed below. If you go to the ESGAR ‘Instructions to Authors’ link in the abstract submission system you will find all this information online. You can link to the examples from this page by copying the URL, and pasting it into your browser.

1.1. **Technical Performance:**
(How can we see what we are looking for?)
This level includes all studies that assess the ability of imaging technology to identify pathology. There is some conceptual overlap with rad-path correlation studies, but the difference is that the focus is on the understanding of the imaging technology rather than the pathologic basis of the abnormalities.

- **Example** A study evaluating automated polyp measurement in CT colonography.

This level also includes studies that assess real or potential morbidity of diagnostic imaging procedures (e.g. radiation dose, contrast toxicity etc).

- **Example** A study evaluating the cardiovascular effects of CT colonography.

1.2. **Diagnostic Performance:**
(How well can we see what we are looking for?).
This level includes all studies that compare an imaging technology (e.g. MR Cholangiography) with a gold standard (e.g. contrast cholangiography) to assess the performance of the new technology. Studies done for these purposes should derive test “properties” (sensitivity, specificity, predictive values) or use statistical techniques such as ROC curves in the results.

- **Example** A study evaluating the diagnostic performance of MRCP

1.3. **Diagnostic Thinking Impact:**
(Does the test have an impact on diagnostic thinking?)
For example, a study that measures the clinician’s subjectively estimated diagnosis before and after test information is received.

- **Example** A study in critically ill ICU patients undergoing CT scanning, evaluating the impact of CT on clinical diagnosis.

1.4. **Therapeutic Impact:**
(Will the test results change the treatment plan?)
These studies might assess the number (percentage) of times a medical or surgical procedure was avoided or a management plan was changed due to image information. **Example** A study at this level might assess the pre-operative plan for colorectal liver metastasis resection based on ultrasound, CT and MRI results, and evaluate the effects of a new test (e.g. PET) on management.


1.5. **Impact on Health:**
(Will the test results change patient outcomes?)

In the broadest context, Adrian Dixon has theorised that this really means ‘Will the results have an impact on the patient’s health?’ [3]

**Example** A study randomising acute surgical patients to have abdominal CT routinely on admission from the emergency room or only if clinically indicated, and assess the impact of early CT scanning on patient outcome.

http://bmj.com/cgi/content/abstract/325/7377/1387

Studies that assess whether the benefit of the test is worth the cost can be carried out for any of the above levels. These are cost-benefit and cost-effectiveness analyses, which look at diagnostic imaging from a societal viewpoint. This level applies to therapeutic studies and therefore can apply to Interventional Radiology studies also.

**Example** A study evaluating the cost-effectiveness of different imaging strategies for rectal cancer.


**DISCUSSION**

Technology Assessment levels form a ‘Conceptual Continuum of Efficacy’ [1, 2]. This is important for two reasons.

1. Thornbury and Fryback, in considering this model of efficacy for diagnostic imaging, stated that a localised view of the goal of diagnostic radiology would be to provide the best images and the most accurate diagnoses possible. Diagnosis is not, however, an end in itself. It is only helpful in so far as it leads to a change of management or a firm, helpful, prognosis [3]. Diagnostic radiology as part of a larger system whose goal is to treat patients effectively and efficiently. The key point is that the global "systems" view of efficacy forces one to consider standards that go beyond the performance / quality of the test or procedure to examine the ultimate value or benefit that is derived from those examinations [2].

2. For many years, radiologists have relied on the rapid progression of imaging technology for research material. This has led to a situation where early access to new technology and completion of descriptive studies have given radiologists all the
access to presentations and publications and prestige that they wanted within radiology [4]. For this reason, most radiology research is at Levels 1 and 2 of this hierarchy (Technical and Diagnostic Performance in Mackenzie and Dixon’s model [3]). These are important steps. Without them, we could not confidently progress in our imaging strategies. The problems began when radiologists brought their level 1 and 2 research to public health and private sector paymasters etc. seeking funding to apply the technologies more widely. What these groups want is evidence that the patient’s health care can be improved by using the tests. Radiologists were not well-prepared for this [4]. Radiology needs more research at Levels 3 – 6 (Diagnostic Impact, Therapeutic Impact and Impact on Health in Mackenzie and Dixon’s model [3]). The key point to note about these models of technology assessment is that for a technology to be fully evaluated at a higher level, it must already be fully evaluated at the lower levels. No technology has been fully evaluated until it’s performance at all levels has been assessed [1].

We can use this in two ways:

1. To understand what the most important features of study design and data analysis are at each level. This will focus and hopefully improve our critical appraisal and study design skills.

2. To give researchers ideas about new research projects at the next highest level, once lower level research has been satisfactorily completed.

Both of these should help radiologists to improve the impact of their research. That will, we hope, give researchers all the presentations, publications and prestige that they want – and increase the evidence base from which we negotiate with public and private sector health care paymasters. Our long-term goal is to help ESGAR members increase the impact of their research and of the society in abdominal imaging and intervention, internationally.

Efficacy and Effectiveness research are both important. Efficacy can be defined as “the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use”. Effectiveness “reflects performance of a medical technology under ordinary, rather than ideal, conditions” [5]. Differences in local expertise or local disease profiles may mean that a technology does not always perform as well at the ‘effectiveness’ level as it did under ideal circumstances. Radiologists need to know about this and ESGAR wants to provide a platform for exchange of information on both efficacy and effectiveness of imaging techniques in abdominal disease, so don’t be put off if you are not working in a major academic centre. Your local data could be very important!

References:

Grading Categories 2 & 3 – Pictorial Essays and Radiologic-Pathologic Correlation

These are two of the key areas of radiological activity. Expert opinion grading is used for Pictorial Essays. The objective grading guidelines for rad-path correlation studies aim to identify factors that predict the reliability of the study conclusions. The best rad-path correlation studies will have high quality correlation with pathology from an adequate number of subjects, and will contribute to our knowledge of the pathologic features that assist in differential diagnosis of findings in abdominal radiology.

Grading Category 5 – Reviews

The concept of ‘evidence-based’ reviews and ‘secondary evidence’ are relatively new.

The ‘Evidence-Based Medicine’ system can be applied to single papers or to reviews. Systematic reviews, which use explicit methodology and often pool data with sophisticated statistical techniques (e.g. meta-analysis), are common in the medical and surgical literature. They are replacing ‘expert’ and ‘consensus’ reviews as they are less subject to bias. They are becoming more common in the radiology literature. In the ‘evidence-based’ paradigm, all original research studies are considered ‘primary’ evidence. Reviews performed according to explicit methodology that present relevant, filtered, best current evidence about common dilemmas and emerging technologies are considered ‘secondary’ evidence. If well done, these reviews have more impact than single studies and are a potentially interesting addition to the ESGAR meeting and, subsequently, a valuable online resource for ESGAR members. They will, as in previous years, be submitted with the primary scientific submissions and, if accepted, presented in the scientific sessions.

You can link to the examples from here by copying the URL into your browser.

Diagnostic Radiology Example: A review and meta-analysis of CT, MRI and PET imaging of colorectal liver metastases.


References

Grading Category 6: Interventional Radiology

Interventional radiologists are becoming increasingly aware that the strength of scientific arguments about the relative value of interventional procedures versus traditional surgical alternatives lies in the quality of the evidence rather than in the quantity of cases done (1). This section of the grading system is built around that concept, drawing on previous publications (2-4) to assess the quality of the study design, the strength of the results and the validity of any statements regarding the safety of the procedure in question (5).

References:

2. Levels of Evidence. NHS Centre for Evidence Based Medicine, Oxford University. In: http://www.cebm.net/?o=1025