Peer Review in Cardiothoracic Radiology

Jeffrey P. Kanne, MD

Abstract: A variety of peer review methods can be used as part of quality assurance and quality improvement in cardiothoracic radiology. Traditionally, peer review in radiology is a retrospective process relying primarily on review of previously interpreted studies at the time of follow-up or additional imaging. However, peer review can be enhanced with other methods such as double reads, focus practice review, practice audit, and correlation with operative and pathologic findings. Furthermore, feedback from referring physicians can be extremely useful in improving the quality of a radiology practice. This article discusses peer review in radiology with a focus on cardiothoracic imaging. Types of peer review, advantages and shortcomings, and future challenges are addressed.

Key Words: peer review, quality improvement, cardiothoracic imaging

(J Thorac Imaging 2014;29:271–280)

LEARNING OBJECTIVES

After completing this SA-CME activity, physicians should be better able to:
1. Compare the advantages and disadvantages of various methods of peer review in radiology
2. Assess how participation in peer review can be incorporated into an American Board of Radiology Maintenance of Certification Program
3. Identify the necessary components to establish a focused practice review

Peer review in its various manifestations has become a nearly universal practice among physicians and is primarily the result of increased public and regulatory scrutiny of physician professional performance that followed publication of the Institute of Medicine 2000 report To Err Is Human.1 Further increasing peer review activities are requirements published by The Joint Commission (TJC), the primary accreditation body for hospitals. In 2004, TJC published new guidelines that mandated collection and use of provider-specific data in the credentialing process, including clinical judgment, technical skills, professionalism, communication, and continued self-improvement and education.2 These guidelines were revised in 2007 with emphasis on aligning provider-specific data with the 6 core competencies developed jointly by the ACGME and the American Board of Medical Subspecialties.3 These core competencies include: patient care, medical and clinical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and system-based practice.

The primary driving force behind peer review in radiology is the American College of Radiology’s RADPEER program, which was developed in direct response to the Institute of Medicine’s To Err Is Human report and was made available to radiologists in 2005.4 Through RADPEER, participating radiologists can view aggregate peer review data stratified by modality and practice site and can compare their own data with those of other radiologists in their respective practice and with national aggregate data. According to the American College of Radiology, > 17,000 radiologists are currently participating in the program.5 Although RADPEER is one of the most popular forms of peer review in radiology, it is only a single tool in the peer review process. Other forms of peer review include double reads, focus practice review, practice audit, and correlation with operative and pathologic findings (Table 1).

This article discusses peer review in radiology with special emphasis on cardiothoracic imaging. Types of peer review, advantages and shortcomings, and future challenges are addressed. Most importantly, the reader will recognize that no one peer review program is perfect and that instituting a more comprehensive evaluation of professional practice that includes a variety of peer review methods is likely the best method to characterize radiologist performance.

MEASURING PERFORMANCE IN RADIOLOGY

Measurements of radiologist performance need to be meaningful to individual radiologists, practice leadership, credentialing bodies, and society as a whole. Metrics should be easily reproducible and selected on the basis of published evidence or agreed-upon standards or guidelines and should be relevant to each radiologist’s respective individual practice.3 When measuring performance in any manner, an adequate number of data points should be collected to ensure that data are meaningful. For example, some groups recommend that 3% to 5% of cases undergo peer review6 with the argument that review of 0.1% of a radiologist’s interpretive reports provides a much lower level of insight into a radiologist’s practice than does a sample of 3%. However, Hussain et al7 correctly point out that this recommendation is arbitrary and is not supported by scientific evidence.

Diagnostic accuracy is the most logical and presumably the most important performance metric in diagnostic radiology because of its direct link to patient outcome.5 However, one should keep in mind that factors affecting outcomes in health care are typically multifactorial and that diagnostic imaging is usually only 1 component of diagnostic evaluation and therapy. Furthermore, errors in diagnosis on medical imaging studies can have variable impact on patient management or ultimate outcome. For
example, failing to detect a 2 cm lung carcinoma on a chest radiograph likely will significantly adversely affect patient outcome, whereas inaccurately characterizing the pattern of severe fibroblastic diffuse lung disease on high-resolution computed tomography (HRCT) may have little impact on patient outcome.

Another performance metric that can be assessed for each radiologist is adherence to agreed-upon practice guidelines. For example, a radiology practice can agree to use the Fleischner Society’s published guidelines for management of incidentally detected small lung nodules,\(^5\) and adherence to and appropriate use of these guidelines can be measured for each radiologist as a component of peer review. Wide variability in management recommendations can be confusing to referring physicians and patients alike.

Other facets of radiologist performance such as interpersonal communication can be determined by feedback from colleagues, trainees, staff, and patients. A combination of personal evaluations and review of patient or staff complaints or commendations can be incorporated into any professional practice review, giving a more comprehensive overview of radiologists’ professional practice.

### METHODS OF PEER REVIEW IN RADIOLOGY

Traditionally, peer review in radiology has been and continues to primarily be a retrospective process. Archiving of diagnostic imaging tests and accompanying interpretative reports makes retrospective peer review easy to perform and widely available. With this method, the reviewing radiologist reviews a previously interpreted study with its accompanying report and assigns a score that reflects agreement or various levels of disagreement. Some scoring systems subdivide scores into those that are clinically relevant and those that are not (Table 2).\(^4\) Studies can be flagged for peer review through random assignment or can be selected when they are rereviewed for consultation or multidisciplinary conferences. When the reviewer disagrees with the initial interpreting radiologist’s report, feedback is provided to that radiologist through a variety of mechanisms.

Double reading is another peer review method used in diagnostic imaging.\(^10,11\) The goal of this prospective approach to peer review is to promptly identify interpretative errors that could have an adverse effect on patient outcome. Most commonly with this method, a second radiologist is assigned to review a recently performed imaging study, typically one performed the same day. Ideally, the second reviewer is privy to the same clinical information and comparison studies and is blinded to the interpretation of the first radiologist. After the second interpretation is rendered, the 2 involved radiologists can reconcile any discrepancies. A third radiologist can be solicited for consultation if disagreement persists. This model is often used in screening mammography and can easily be used for chest radiography given the relatively brief amount of time required for interpretation and reporting of these studies.

Professional auditing is another retrospective form of radiologic peer review in which image interpretation is compared against a reference standard such as pathologic data, operative findings, and clinical follow-up. With

### TABLE 1. Types of Peer Review With Respective Advantages and Disadvantages

<table>
<thead>
<tr>
<th>Method of Peer Review</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrospective</td>
<td>Data easily accessible, can be performed-on-the-fly</td>
<td>Unblinded, potential for bias, significant errors may not be detected in a timely manner</td>
</tr>
<tr>
<td>Double read</td>
<td>Real-time peer review, potential to promptly identify clinically significant errors</td>
<td>Labor and time intensive, requires large number of studies to be meaningful</td>
</tr>
<tr>
<td>Professional audit</td>
<td>Uses reference standards, encourages use of electronic health record and outcomes</td>
<td>Labor intensive, limited types of studies that can be reviewed, validity of reference standard may be in question</td>
</tr>
<tr>
<td>Structured feedback</td>
<td>Looks at entire radiology report, encourages focus on communication, solicits input from referring physicians</td>
<td>Labor intensive, may miss clinically significant findings when focus is too much on report verbiage</td>
</tr>
<tr>
<td>FPR</td>
<td>Process highly structured and defined upfront with robust appeals process, involves departmental leadership</td>
<td>Labor intensive, may incite animosity from potential “punitive” nature</td>
</tr>
<tr>
<td>Comprehensive professional review</td>
<td>Evaluates all facets of professional practice in radiology, aligned with regulatory requirements</td>
<td>Labor intensive, obtaining data may be difficult, may prove to be more challenging for smaller practices</td>
</tr>
</tbody>
</table>

### TABLE 2. ACR RADPEER Scoring System\(^4\)

<table>
<thead>
<tr>
<th>Score</th>
<th>Meaning</th>
<th>Optional</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Concur with interpretation</td>
<td>(1) Unlikely to be clinically significant</td>
</tr>
<tr>
<td>2</td>
<td>Discrepancy in interpretation/not ordinarily expected to be made (understandable miss)</td>
<td>(2) Likely to be clinically significant</td>
</tr>
<tr>
<td>3</td>
<td>Discrepancy in interpretation/should be made most of the time</td>
<td>(1) Unlikely to be clinically significant</td>
</tr>
<tr>
<td>4</td>
<td>Discrepancy in interpretation/should be made almost every time—misinterpretation of finding</td>
<td>(2) Likely to be clinically significant</td>
</tr>
</tbody>
</table>
shortcomings and solutions

The primary intent of peer review in radiology is to improve patient care through better outcomes by reducing interpretative and procedural errors. However, along with advantages, each model of peer review has its shortcomings. Factors that limit the effectiveness and acceptance of peer review include conscious and unconscious bias among reviewers’ unclear policies, lack or perceived lack of transparency, absence of evidence-based reference standards for many cases, additional time pressures on already busy radiologists, beliefs that the process will not result in meaningful improvement in patient care, and even legal concerns.

The retrospective case review model has received the most attention for its inherent limitations.6,13,14 For example, personal and professional biases compounded with increasing pressures on radiologists’ time can lead reviewers to select less complex studies or apparently normal studies for review. Increasing the randomness of reviewed cases can improve the quality of a peer review program. One method to improve the randomness of peer review is to assign the first case encountered with a relevant comparison on any given day. Expanding on this model, peer review can be performed in this manner to include a specific number or type of modality or body part. More sophisticated systems can randomly select cases for review from a radiology information system or from a voice recognition software database, assign the case for review, and track the progress of that review.

Anonymity of reviewer and reviewee is frequently missing with retrospective peer review. Reviewers can easily identify the interpreting radiologist of the case under review, the knowledge of whom may consciously or unconsciously bias the choice to review as well as bias the score assigned. Similarly, reviewees can typically determine the person performing peer review by ascertaining who interpreted the first case encountered with a relevant comparison. Additionally, there is the fear of potential retribution.

Another major criticism of retrospective peer review is that a significant time gap may exist between the original interpretation and retrospective review of that case, potentially missing the opportunity to avoid significant adverse impact on patient care. One solution to this issue is to define a maximum time period between the current examination and the comparison examination being used for peer review,15 although this may narrow the pool of eligible studies and likely overemphasize inpatient examinations and radiography. Establishing different maximum time frames for each modality depending on its respective relative frequency across a practice may overcome this shortcoming.

Although double reading provides the advantage of near-real-time peer review, major drawbacks still exist. Most importantly, double reading requires additional radiologist time. Furthermore, to make this type of peer review meaningful, a reasonable number of cases need to be double read to achieve statistical usefulness. This increased burden of work has the potential to incite animosity toward the peer review process and reduce compliance.

Practice audits are advantageous because of direct comparison with objective (or mostly objective) data such as operative findings. However, this approach limits the scope and types of cases that can be reviewed. For example, the increasing implementation of electronic health records, access to comparative data becomes more readily available for this type of case review. Advantages to this method include the use of some objective data, the potential to correlate imaging interpretation with outcomes or operative findings, and encouragement to radiologists to regularly use the electronic health record as part of their practice. To institute this method of peer review, however, practices need to agree upfront upon what types of imaging studies, disease categories, and subspecialties to target, and robust clinical data such as pathology reports and operative notes need to be readily accessible.

A relatively novel approach to peer review in radiology is one that relies on direct input from referring physicians.12 Structured feedback is solicited from referring physicians and focuses on other aspects of diagnostic imaging reporting aside from accurate identification and interpretation of findings. Referring physicians can provide feedback on report features such as language, internal consistencies, length, whether or not the clinical question is addressed, and inclusion of appropriate or relevant recommendations. For example, retrospective peer review of a complex cardiac magnetic resonance imaging (MRI) examination may result in agreement with a highly detailed and accurate yet verbose and disorganized report but can easily miss the fact that the referring physician failed to understand the clinical implications of the examination results. Key advantages of this type of peer review include focusing the radiologist on the needs of the referring physician, aiming for improved communication between radiologist and referring physician, and striving for improved patient care through more accurate communication among health care team members.

Hussain et al.17 implemented a more vigorous method of radiology peer review by developing a focused practice review (FPR) to supplement routine retrospective peer review. In addition to randomly detected errors, FPRs included clinically reported errors, which were misdiagnoses reported by referring physicians. The FPR method developed by Hussain and colleagues17 has a clearly defined sequence of review and, most importantly, a robust appeals process. The department quality officer, the division director, and the department chair review cases with significant errors, and, at the conclusion of the FPR process, corrective action is dictated by the department chair. This type of peer review has the advantages of being highly structured and directly involving senior departmental leadership.

A truly comprehensive professional peer review program extends beyond image interpretation to include all aspects of professional practice. Donnelly and colleagues1 at Cincinnati Children’s Hospital Medical Center worked with their medical staff office to develop a comprehensive radiologist performance assessment modeled on the 2007 TJC standards. In addition to using peer review of case interpretation, Donnelly expanded performance review to include other facets of professional practice such as evaluation of professional interactions with patients and staff, participation in continuing medical education, maintaining appropriate skills certification such as cardiopulmonary resuscitation, adherence to departmental communication policies, participation in quality and safety improvement projects, and completion of institutionally mandated training. This approach to professional peer review focuses on the entire professional practice of radiologists and is not just limited to image interpretation.
vast majority of small lung nodules are benign, so failure to detect a small nodule on chest CT may go unnoticed if the patient does not undergo any further evaluation. In addition, an overlooked small pulmonary embolism could result in no adverse clinical outcome and thus would not be captured by peer review. A second shortcoming of practice auditing is that it assumes that what is defined as the “reference standard” is accurate. Comparing coronary CT angiography results with conventional coronary angiography would seem to many to be a useful benchmark for a practice audit. However, subtle abnormalities can be overlooked on conventional angiography, and because of the 2-dimensional nature of conventional angiographic images, lesion severity can be overscored or underscored. Another example is distinguishing between restrictive and constrictive cardiac physiology. Echocardiography, angiographic, or clinical evaluation may establish a diagnosis of restrictive cardiomyopathy, whereas cardiac MRI clearly shows constrictive pericarditis. Problems also arise when the pathologic diagnosis is used as the reference standard. For example, a radiologist with experience in diffuse lung disease may give a diagnosis of nonspecific interstitial pneumonia on an HRCT of a patient with basal-predominant ground-glass opacity, reticulation, and traction bronchectasis with subtle subpleural sparing. However, the expert pulmonary pathologist renders a final pathologic diagnosis of usual interstitial pneumonia for the surgical biopsy specimen. An experienced thoracic radiologist will immediately recognize that this type of discordance between HRCT and histopathologic findings is not at all uncommon in the setting of diffuse lung disease and does not necessarily reflect an interpretative error on either part but rather illustrates the significant overlap of HRCT and pathologic features of diffuse lung disease. In contrast, a general radiologist with little exposure to diffuse lung disease may consider this, in the context of peer review, to be an interpretative error on the part of the radiologist, because the surgical biopsy is believed to be the reference standard. To reconcile discrepancies like this, the individual performing peer review must comprehend the accuracy of each examination being reviewed and the accuracy of the test or observation being used as the reference standard, both of which may be difficult for those whose practice covers a broad spectrum of modalities and diseases.

Feedback from referring physicians regarding the quality and utility of a diagnostic imaging report can be very informative for a radiologist. However, this approach to peer review can fail to reveal deficiencies in actual image interpretation. For example, a referring physician may be pleased with the quality of one radiologist’s report of a chest radiograph but does not convey the fact that subtle left lower lobe pneumonia was overlooked. Another scenario to consider is one in which HRCT shows scattered foci of patchy ground-glass opacity, and the differential diagnosis is appropriately broad given limited clinical information, yet the referring physician is frustrated because a specific diagnosis or more limited differential diagnosis is not provided. In order for this type of peer review to be successful, an easy mechanism should be in place for referring physicians to submit comments and requests for review of clinically reported diagnostic errors.

FPR as described by Hussain et al provides a more structured approach to peer review that involves senior departmental leadership, a robust appeals process, and defined management options. Although this process may enhance the effectiveness of peer review, it has the potential of creating negative feelings toward the process. Specifically, according to Hussain and colleagues, “completed FPR cases formed the basis of morbidity and mortality presentations and punitive management decisions.” In my opinion, the mere use of the word “punitive” can have unintended consequences in the realm of peer review.

Finally, the comprehensive professional evaluation described by Donnelly likely provides the broadest form of peer review and considers the whole spectrum of professional radiologic practice. To establish and maintain a program like this can be extremely resource intense. Implementing a robust professional assessment as such probably can be accomplished in a large community practice or academic department with sufficient administrative support, but acquiring and tracking the requisite data can be more challenging for those in smaller practices.

**MY PRACTICE IS COLLECTING DATA, BUT NOW WHAT?**

Once peer review data collection begins, participants will start to expect to see feedback in some form or another (Fig. 1). As described by Hussain et al, FPR uses a predefined structured, tiered method of review including chief quality officer, division chief, and department chair and includes an appeals process. At the end of the FPR, the department chair makes a decision as to what further action is needed. In contrast, Butler and Forghani make the case for a shift away from metrics and error identification to improving the performance of the entire practice, in a sense shifting the bell curve to the right rather than chopping off the lower tail. The premise of their proposal is similar to that of the aviation industry, which is built around learning from past mistakes of others to prevent future errors as well as establishing a culture in which all parties involved are empowered to speak up when they feel an error has been made or, more importantly, is about to be made.

No matter which approach to peer review a practice selects, it is critical to use collected data in an educational format. Some practices may use the more traditional morbidity and mortality conference, in which case specifics are discussed in an open forum with no anonymity, and the “owner” of the case must defend or justify actions taken or reasons for an interpretation. Other groups may opt for the other end of the spectrum in which difficult or “missed” cases are presented in a nonpunitive, anonymous forum with the sole aim to educate attendees. In addition to the cases themselves, participants may be assigned to present on a specific topic. As a result of either of these approaches to group evaluation of reviewed cases, practices may adopt a uniform image acquisition process (eg, single HRCT acquisition protocol), agree to use a defined lexicon of terminology (eg, Fleischner Society’s glossary of terms for thoracic imaging), or adopt practice guidelines (eg, Fleischner Society’s recommendations for management of subsolid pulmonary nodules).

Once a peer review program is underway, continued support from participating radiologists is key to maintaining success. Documenting practice improvement or standardization may convince skeptics of the value of the peer review process. Furthermore, peer review activities can also be used to satisfy part IV (Practice Quality Improvement) of the American Board of Radiology’s Maintenance of Certification program. Diplomats enrolled in Maintenance of Certification can use their own peer review data to design and implement a quality improvement project. Ultimately,
a robust peer review program will facilitate a change from a culture of circumspection and paranoia to one of collaboration and openness, with the goal of improving patient care.

**SPECIFIC CHALLENGES IN CARDIOTHORACIC RADIOLOGY**

One of the most difficult challenges facing peer review in cardiothoracic imaging is the relative paucity of peers.

---

**FIGURE 1.** Screen capture of peer review reporting system. Names and examination information have been obscured to protect confidentiality.
that may exist in one practice. This becomes a particular issue in subspecialty practices wherein only a handful of radiologists interpret complex cardiothoracic imaging studies. In this situation, maintaining anonymity may be near impossible and using double reads becomes unfeasible. Interpretation of cardiac MRI and HRCT of the chest requires a certain level of expertise, and the available pool of radiologists who provide useful peer review may be extremely limited or nonexistent. Furthermore, agreement in interpretation of these complex studies can vary even among “experts.”22,23 By combining a variety of peer review methods such as feedback from cardiologists and pulmonologists as well as using results from biopsy, bronchoscopy, other cardiac imaging studies, and operative findings, the value of peer review can be increased. Because of the relative lack of dedicated cardiothoracic radiologists, practice structures, or historic modality-driven workflows, chest imaging studies are often interpreted by general radiologists or radiologists with other subspecialty training. Defining the standard by which peer review should be performed when radiologists vary in their level of expertise and training becomes a challenge. For a general practice or a completely subspecialized practice, it is safe to say that the standard should be that of a generalist or subspecialist, respectively. However, when cardiothoracic imaging studies are interpreted by cardiothoracic radiologists and other radiologists, how does one define the standard? In my opinion, no simple answer exists. If an emergency radiologist interprets a pulmonary CT angiogram correctly as “no pulmonary embolism” but incorrectly misinterprets the diffuse lung cysts of lymphangioleiomyomatosis as “diffuse emphysema,” should that radiologist be considered performing below the standard of care? This question is not easy to answer. However, what is clear is that standards and expectations need to be defined upfront by each practice and that the goal of the peer review program is to improve the quality of care delivered by the practice as a whole.

Another challenge in cardiothoracic radiology peer review is ensuring that an adequate number of cases are reviewed. This can particularly be a problem with chest radiography, wherein volumes are usually high but the rate of significant pathology is typically quite low. For example, if a cardiothoracic radiologist routinely interprets 100 chest radiographs per day, sampling 2 or 3 examinations may be insufficient to detect significant overlooked abnormalities as often is the case, chest radiographs are not infrequently obtained for dubious indications such as “gastrointestinal bleed,” “preop,” or “evaluate.”

Although complex imaging studies such as cardiac MRI may seem like a good choice for peer review, the relative complexity of these cases can make performing peer review a significant chore for the reviewer, especially when a double reading technique is used. Questions arise as to how much effort the reviewer should expend on the case review: Does every sequence need to be reviewed in detail? Does every sequence need to be reviewed in detail? Do much effort the reviewer should expend on the case review: review a significant chore for the reviewer, especially when a cardiothoracic radiologist routinely interprets 100 chest radiography, wherein volumes are usually high but the rate of significance of care delivered by the practice as a whole.

Another challenge in cardiothoracic radiology peer review is ensuring that an adequate number of cases are reviewed. This can particularly be a problem with chest radiography, wherein volumes are usually high but the rate of significant pathology is typically quite low. For example, if a cardiothoracic radiologist routinely interprets 100 chest radiographs per day, sampling 2 or 3 examinations may be insufficient to detect significant overlooked abnormalities as often is the case, chest radiographs are not infrequently obtained for dubious indications such as “gastrointestinal bleed,” “preop,” or “evaluate.”

Although complex imaging studies such as cardiac MRI may seem like a good choice for peer review, the relative complexity of these cases can make performing peer review a significant chore for the reviewer, especially when a double reading technique is used. Questions arise as to how much effort the reviewer should expend on the case review: Does every sequence need to be reviewed in detail? Does every sequence need to be reviewed in detail? Do much effort the reviewer should expend on the case review: review a significant chore for the reviewer, especially when a cardiothoracic radiologist routinely interprets 100 chest radiography, wherein volumes are usually high but the rate of significance of care delivered by the practice as a whole.

Another challenge in cardiothoracic radiology peer review is ensuring that an adequate number of cases are reviewed. This can particularly be a problem with chest radiography, wherein volumes are usually high but the rate of significant pathology is typically quite low. For example, if a cardiothoracic radiologist routinely interprets 100 chest radiographs per day, sampling 2 or 3 examinations may be insufficient to detect significant overlooked abnormalities as often is the case, chest radiographs are not infrequently obtained for dubious indications such as “gastrointestinal bleed,” “preop,” or “evaluate.”


Questions marked with an asterisk are ABR Self-Assessment Module (SAM) questions. Participants can claim credit for the SAM regardless of the test outcome. Notify the ABR of the SAM completion, or visit the ABR website at www.theabr.org to set up or login to your personal database to record the number of SAMs you completed. The SAM ID number will be printed on the CME certificate for your records. If you wish to include the ID number in your ABR database, contact a MOC Specialist at the ABR office for instruction by calling 520-519-2152.