

ACR RADPEER Committee White Paper with 2016 Updates: Revised Scoring System, New Classifications, Self-Review, and Subspecialized Reports

Shlomit Goldberg-Stein, MD^a, L. Alexandre Frigini, MD^b, Scott Long, MD^c, Zeyad Metwalli, MD^{b,d}, Xuan V. Nguyen, MD, PhD^e, Mark Parker, MD^f, Hani Abujudeh, MD, MBA^g

Abstract

The ACR's RADPEER program is currently the leading method for peer review in the United States. To date, more than 18,000 radiologists and more than 1,100 groups participate in the program. The ABR accepted RADPEER as a practice quality improvement in 2009, which can be applied toward maintenance of certification; there are currently over 2,200 practice quality improvement participants. There have been ongoing deliberations regarding the utility of RADPEER, its goals, and its scoring system since the preceding 2009 white paper. This white paper reviews the history and evolution of RADPEER and eRADPEER, the 2016 ACR Peer Review Committee's discussions, the updated recommended scoring system and lexicon for RADPEER, and updates to eRADPEER including the study type, age, and discrepancy classifications. The central goal of RADPEER to aid in nonpunitive peer learning is discussed.

Key Words: RADPEER learning, peer review, peer learning, RADPEER, nonpunitive

J Am Coll Radiol 2017;14:1080-1086. Copyright © 2017 American College of Radiology

INTRODUCTION

In 1999, the Institute of Medicine (IOM) reported that medical errors accounted for nearly 100,000 preventable deaths each year in the United States alone [1]. Patient safety proved elusive, and in September 2015, a follow-up IOM report focusing on diagnostic errors cited that these errors contribute to an alarming 10% of patient deaths [2]. One of the many stated goals of that report was to “develop and deploy approaches to

identify, learn from, and reduce diagnostic errors and near misses in clinical practice” [2].

In addition to the unnecessary, irreplaceable loss of life, preventable medical errors prolong the course and duration of patient hospitalization, increase patient morbidity and suffering, and further accentuate the cost of health care delivery on the order of tens of billions of dollars per year. Unfortunately, as the original IOM report highlighted, to err is indeed human, but it was also understood that human errors may occur in a predictable pattern and frequency [1,2]. It is within this context that we provide a history of the evolution of the ACR's RADPEER™ and eRADPEER programs, and then describe the most recent updates of 2016, which are a deliberate attempt to facilitate peer learning, as valued by the IOM.

The original IOM report did not cite medical imaging in particular as an area of medicine fraught with high error rates [1]. Nevertheless, in response to that IOM report, and in the interest of public safety and the health care community, the ACR task force established several committees to specifically examine this issue. One such

^aMontefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York.

^bBaylor College of Medicine, Houston, Texas.

^cSouthern Illinois University School of Medicine, Springfield, Illinois.

^dMichael E. DeBakey VA Medical Center, Houston, Texas.

^eOhio State University Wexner Medical Center, Columbus, Ohio.

^fVirginia Commonwealth University Medical Center, Richmond, Virginia.

^gCooper University Hospital of Rowan University, Camden, New Jersey.

Corresponding author and reprints: Dr Shlomit Goldberg-Stein, MD, 111 East 210 Street Bronx, NY 10463; e-mail: sgoldberg@montefiore.org.

The authors have no conflicts of interest related to the material discussed in this article.

committee examined professional self-evaluation and peer review and subsequently developed the ACR's first RADPEER program. After its successful piloting at 14 sites in 2001 and 2002, this RADPEER program was made available to ACR members in 2002 [3].

RADPEER was built on the premise that whenever a radiologist interprets a new imaging study in conjunction with a preceding comparative examination of a similar nature, there is a professional opportunity to assess the accuracy of the original interpretation and reader's (ie, radiologist's) performance with respect to that examination [3,4]. When utilized appropriately, RADPEER is a simple, cost-effective performance tool allowing previous exams and their interpretations to be reviewed for accuracy. The resulting data can then be collected for systematic analysis of an individual radiologist or radiology group's performance and to identify possible deficiencies or areas for improvement [3,4]. Numerous radiologist performance audits have reported daily practice error rates on the order of 3% to 5%; a well-designed and well-conducted peer review system could provide physicians and health care providers with an opportunity to identify patterns of error, intervene, and implement strategies to reduce specific medical errors, which in turn reduce patient harm [3-11].

RADPEER scoring was originally performed using machine-readable cards. This was updated to a web-based program, eRADPEER, in 2005. The diagnostic accuracy of the previous interpreting radiologist's report was scored by the contemporary interpreting radiologist, using a 4-point rating system (Table 1) [3,4,9]. In general, scores of 1 and 2 did not require any additional action or intervention, but scores of 3 and 4 required an internal review by the local peer review committee to either substantiate or modify the assigned score. Although this scoring system initially worked well, questions with its application soon arose [9]. In particular, score 3 ("Diagnosis should be made most of the time") did not distinguish a diagnostic misinterpretation from a possible disagreement in interpretation. Furthermore, a

score of 3 could also potentially be applied in scenarios in which the original examination was correctly interpreted, but the current interpreting radiologist simply "felt" this was an easy diagnosis to make. It was likewise unclear if a score of 2 ("Difficult diagnosis, not ordinarily expected to be made") represented an actual disagreement in diagnostic interpretation or was simply being used in cases of a great diagnostic pickup ("a good call") [9].

In 2009, these issues were addressed in the published release of a second RADPEER Committee white paper [9]. Updated RADPEER lexicon was adopted for the 4-point scoring system. The new language clarified that a score of 2 was not a "good call" but rather a discrepancy in interpretation for a difficult finding (ie, an understandable miss). Additionally, terminology for scores 3 and 4 was updated, with the new terminology considered more widely applicable and outcomes based. To aid the radiologist in assigning the score, the 2009 white paper provided applicable clinical examples of each score. Also, the updated RADPEER system added optional "a" and "b" categories addressing whether or not the diagnostic discrepancy was clinically significant or not (Table 2) [9].

In revising the scoring terminology, one concern raised by task force members was that a major shift in the definitions of the scores would cause previously accumulated data to be lost. According to the 2009 white paper, the ACR's Research Department at the time felt that changes preserving the 4-point system, and, in particular, preserving the distinction between a score of 1 versus any other score, allowed for reliable cumulative historical data to remain. The Research Department did, however, suggest that the task force revisit the terminology and consider changes about every 5 years [9].

Table 1. Original 2002 RADPEER Scoring System

| Score | Meaning |
|-------|--|
| 1 | Concur with interpretation |
| 2 | Difficult diagnosis, not ordinarily expected to be made |
| 3 | Diagnosis should be made most of the time |
| 4 | Diagnosis should be made almost every time—misinterpretation of findings |

Table 2. Revised 2009 RADPEER Scoring System

| Score | Meaning | Optional |
|-------|--|--|
| 1 | Concur with interpretation | |
| 2 | Discrepancy in interpretation/ not ordinarily expected to be made (understandable miss) | a. Unlikely to be clinically significant b. Likely to be clinically significant |
| 3 | Discrepancy in interpretation/ should be made most of the time | a. Unlikely to be clinically significant b. Likely to be clinically significant |
| 4 | Discrepancy in interpretation/ should be made almost every time—misinterpretation of finding | a. Unlikely to be clinically significant b. Likely to be clinically significant |

RADPEER is currently the leading method for peer review in the United States. As of January 31, 2017, RADPEER had a total of 1,118 groups participating in the program, with a total of 18,305 physicians, and 2,241 of those being practice quality improvement participants. To date, the total number of records received in RADPEER is over 30 million. The ABR accepted RADPEER as a practice quality improvement in 2009, which could be applied toward the participating radiologist's maintenance of certification. RADPEER can also satisfy both hospital and Joint Commission credentialing requirements as part of the accreditation process, which mandates departmental data regarding the performance of their radiologist(s) [12].

Increasingly, RADPEER is electronically implemented into the radiologists' routine digital work flow, prompting radiologists in 5% (or more) of cases to switch their roles from that of a primary reader to a peer reviewer, ideally in an anonymous and spontaneous or random manner. Used appropriately, RADPEER produces scores reflecting levels of agreement or disagreement with regards to an original diagnostic interpretation of imaging findings between radiologists at different points in time [12,13]. The reviewing radiologist has the prerogative of providing comments along with any discrepant scores or findings. If a discrepant reading is reported (score 2b or greater), an internal peer review committee or practice manager reviews the case before ACR submission. As part of this review process, the original interpreting radiologist should be made aware of the finding and allowed to explain his or her view and, where relevant, appeal the reported discrepant reading. This process also enables the involved radiologist to address any educational gaps necessary to avoid repeating the same or similar interpretive errors [14]. There must also be a mechanism in place to notify the affected patient's clinical provider or the patient themselves of the discrepant reading, if the discrepancy could impact the patient's medical care [14].

Each medical practice, institution, or radiology group is assigned a unique identifier number by the ACR. Facilities also assign each physician user a numeric identifier to maintain confidentiality when performance data are submitted to the ACR [9]. The actual radiologist's name is not provided to the ACR. The ACR utilizes the submitted data from RADPEER to create online comprehensive data summary reports. These reports are then made available to the radiology group's chair or medical director summarizing statistics

and comparisons for each radiologist by imaging modality. Summary data for each facility by imaging modality compared with data from across all participating facilities is also made available. This information should then ideally be used to guide the focus of continuing medical education, improving interpretive skills, and improving patient care [9,12-14].

ACR RADPEER COMMITTEE

In August 2015, the ACR RADPEER Committee convened to re-evaluate RADPEER and its scoring terminology. This was in keeping with the suggestion of the ACR's Research Department, mentioned previously and documented in the 2009 white paper, that "the terminology be reevaluated and change considered about every 5 years" [9]. In closed committee meetings, members reflected upon the scoring categories until committee consensus was reached. The scoring updates are detailed in the following section. The committee also set out to evaluate the eRADPEER product, the web-based program first offered to ACR members in 2005. In evaluating the eRADPEER program in quarterly meetings, committee members were asked to suggest improvements to the software itself and to the reports obtained from RADPEER. Consensus emerged within the committee surrounding the updates to eRADPEER, comprehensively outlined in the following text.

ACR RADPEER AND eRADPEER UPDATES

RADPEER Scoring System

It is accepted practice that for discrepant findings with scores 2b or greater, local institution or practice-based peer review committees convene to uphold, upgrade, or downgrade the final RADPEER score [14]. A common theme among committee members' written responses to questions regarding the scoring system was the sentiment that excessive time was unnecessarily spent in peer review committee meetings deliberating between a case's final score of 3 versus 4. Many members felt that because no added learning benefit resulted from that deliberation, it was an unnecessary and unhelpful burden on the local committees to make that determination. Because the primary goal of peer review is to enhance learning, the first action of the ACR RADPEER Committee was to hold a vote on whether to remove the score of 4. This would effectively merge the previous scores 3 and 4 into only one "discrepant score." The vote passed unanimously in favor of dropping to a 3-point scoring system. At the time of

committee vote, two major reasons were recorded as motivations for this change. First, members felt that distinguishing between interpretations that “should be made most of the time” and those that “should be made almost every time” was both subjective and nonreproducible. Second, all committee members agreed that time wasted debating and voting over a score of 3 versus 4 has no real value and served only to distract from the goals of peer learning. The sentiment that brought the issue to vote was carried by the entire committee membership. The committee concluded that eliminating the score of 4 would simplify the peer review process without compromising committee goals.

The revised 2016 RADPEER scoring system replaces the previous 4-point scale with a 3-point scale (Table 3). As in the previous scoring system, a peer review score of 1 will continue to indicate agreement with the previous reader (“concur with interpretation”). Discrepancies for which the correct interpretation is “not ordinarily expected to be made” (ie, understandable miss) will continue as a score of 2. Similarly, any discrepancy in interpretation that “should be made most of the time” will continue as a score of 3. The major update is that the score of 4 previously used for “discrepancy in interpretation/should be made almost every time—misinterpretation of finding” has been eliminated, and such discrepancies will, under the new scoring system, be joined with the cases assigned a score of 3. The option to designate discrepancies as “likely to be clinically significant” or not (options a and b) will continue to be available in the revised scoring system.

Some concern was raised over the potential impact of changes in the scoring scale on comparability of new RADPEER data with prior data. To address this issue, the committee agreed that existing, archived peer review scores of 4 will be merged into score category 3, for purposes of statistics and comparison. In this way, all discrepant cases of 3 or 4 in the archive would be

compared with future cases scored as 3, which is in reality a merged score of categories 3 and 4. The language describing the remaining three score categories remains unchanged. In this way, comparability with previously accumulated data is preserved.

Study Type Classification

Several updates were proposed by the 2016 ACR RADPEER Committee for the ACR RADPEER data summary reports. The current reports provide comparative data across all participating practices and facilities, with comparisons made by imaging modality alone (ie, x-ray, CT, ultrasound, MRI). The committee felt that to facilitate peer learning, more granular data reports would be helpful. Therefore, classification by body system was introduced (Table 4).

Age Classification

The committee similarly felt it would be useful to separate out pediatric cases from adult cases to provide more specific information in its reports. The updated version of eRADPEER allows users to identify cases by age, with a new “pediatric only” option. This is considered an improvement because pediatric radiology is a recognized subspecialty and cases falling into the pediatrics category may want to be reviewed separately to determine patterns of weakness or performance gaps that would be lost in aggregated data including both adults and children.

Discrepancy Classifications

To provide more information about identified discrepancies in interpretation, the committee proposed that discrepancies be categorized. The proposed categories are error in perception, error in interpretation, and error in communication. The classification according to discrepancy type can occur either at the time of case submission by the reviewing radiologist or at the time of committee review. The result is a more specific RADPEER report, which may drive practice quality improvements. Notably, these categories have been the primary contributors to medical malpractice suits involving radiologists [15-17].

An error in *perception* is a cognitive error, which occurs when the interpreting radiologist fails to identify or describe an abnormality on a diagnostic image. A perceptive error can be retrospectively identified on the image once the error is identified. Perceptual errors account for the majority (up to 80% in some series) of clinically relevant discrepancies in radiologic reporting [18-21].

Table 3. RADPEER Scoring System (Effective May 2016)

| Score | Meaning | Optional |
|-------|---|--|
| 1 | Concur with interpretation | |
| 2 | Discrepancy in interpretation/ not ordinarily expected to be made (understandable miss) | a. Unlikely to be clinically significant b. Likely to be clinically significant |
| 3 | Discrepancy in interpretation/ should be made most of the time | a. Unlikely to be clinically significant b. Likely to be clinically significant |

Table 4. Classification of Examination Type

| Original Classification by Modality | New Classification | | |
|-------------------------------------|--------------------|---------------------------|-------------------------|
| | By Modality | By Body System (Optional) | By Age Group (Optional) |
| Radiography | Radiography | All | Adult only |
| US | US | Abdomen | Pediatric |
| CT | CT | Breast | |
| MRI | MRI | Cardiovascular | |
| Nuclear medicine | Nuclear medicine | Chest | |
| Mammography | Mammography | GU | |
| IR | IR | MSK | |
| PET | PET | Neuro | |
| Fluoroscopy | Fluoroscopy | OB/GYN | |
| | | Vascular | |

GU = genitourinary radiology; IR = interventional radiology; MSK = musculoskeletal radiology ; Neuro = neuroradiology; OB/GYN = obstetrics and gynecology; US = ultrasonography.

An error in *interpretation* is also a cognitive error that occurs when the interpreting radiologist correctly identifies and describes an imaging abnormality but fails to identify the significance or clinical impact of the abnormality. Interpretive errors are rarely due to a lack of knowledge [19,20]. Rather, interpretive errors are more likely due to incorrect judgment or misclassification by the radiologist, accounting for 26% of errors in a review of 182 errors by Renfrew et al [19].

An error in *communication* is a failure to communicate the results of an imaging study to the appropriate provider caring for the patient. This can be due to an unclear or misleading report or a report that does not effectively communicate the radiologist’s recommendations. It can also be due to failure to properly communicate urgent or unexpected findings effectively or within an appropriate time frame. It can be due to simple failure to transmit the report to the referring physician. This is a common problem in radiology. Whang et al reported that 4% of malpractice claims involving radiologists, where a primary error could be elucidated, involved a communication error [21]. However, in the series by Renfrew et al, communication errors were present in up to 10% of cases reviewed [19]. Brenner et al reported communication errors in 20% of cases reviewed in the Physicians’ Insurers Association of America/American College of Radiology report. In 10% of claims reviewed, a written radiology report was not forwarded to the ordering physician or the patient; issuance of the report was delayed in an additional 10% of reviewed claims [22].

Subspecialized Reports

Stratifying cases not only by modality but also by examination type, patient age, and by discrepancy type in

the RADPEER system can allow participants to make targeted practice improvement decisions based on areas of identified weakness. Discrepancy classification in combination with the other modifiers may further aid in focusing a radiology practice’s effort at reducing discrepancies. Similarly, if errors in communication are found to be prevalent, interventions by a radiology practice can be focused on improving communication between radiologists and the referring physicians. The committee felt strongly that the addition of these classifications to the ACR report would increase the utility of the report.

Self-Review

The final eRADPEER proposal, passed unanimously by the ACR 2016 RADPEER Committee, is the new ability to perform a self-review. What this entails technically is that the logged-in user ID will be available for selection in the “reviewed physician” menu. This self-reporting feature opens the review process to self-criticism, which encourages self-directed learning and which also bolsters the openness of the learning environment. This feature may be especially useful in environments where peer review cases are gathered and shown anonymously in a conference format for learning purposes alone. This easy mechanism by which to share a “case you missed” with the group serves the goals of peer learning.

NONPUNITIVE PEER LEARNING

The committee believes that the updates to RADPEER and eRADPEER outlined here are a step in the right direction, within the confines of the existing ACR RADPEER construct, toward shifting from a traditional model of scoring-based peer review to an eventual model

where a collaborative peer learning environment is truly achievable [23].

Learning is the most important outcome that can be achieved using RADPEER data. RADPEER learning cannot be effectively achieved if there is a perception of punishment. Fear of punishment is a hindrance to full, authentic, and sincere participation. Attention should be paid to the construction of an environment around RADPEER to guarantee the elimination of the punitive perception.

Despite the existence of a standard scoring system, the committee believes that if RADPEER is applied in a nonpunitive manner, it can serve as an effective learning tool. The committee recommends that, if not already in place, an important step toward a positive, cultivated learning environment for peer review is to establish a local culture of comfort and confidence in your system. This emerges from the knowledge that participation will not result in any punitive measures or judgment or result in being chastised. To that end, the explicit committee recommendation is that RADPEER data should never be the basis for punitive action or remediation of an individual practitioner.

The committee hopes that the updated, more granular reports and the error classifications will serve to aid in peer learning. RADPEER can be used as a platform, with complementary learning systems attached. The learning systems can be comprised of a committee or a conference. The categories can aid in the preparation of theme-based or subspecialty-type learning conferences that allow for all radiologists and trainees to learn from their recognized errors.

RADPEER is a quality improvement tool, a platform of sorts. How it is used determines what outcome is to be gained. Education sessions are at the heart of serious peer learning processes. The RADPEER Committee recommends that local peer review committees or their designees vet and prepare cases anonymously for regularly scheduled learning conferences with educational case content for review to aid the explicit learning benefit of all radiologists and trainees. These recommendations are in keeping with the advice of the most recent IOM report, which advises developing and deploying approaches to reduce diagnostic errors and near-misses [2].

CONCLUSIONS

To facilitate peer learning, the ACR RADPEER Committee enacted several changes to RADPEER and eRADPEER in May 2016. The update to a 3-point

scoring system was designed to unburden local peer review committees from the unhelpful task of stratifying discrepant cases. The newly categorized RADPEER reports are intended to facilitate identification of error patterns, to aid in the development of focused subspecialty peer learning conferences, and to bolster strategies to reduce specific errors, and in turn, to improve patient care. In a nonpunitive environment, RADPEER learning can flourish as a powerful improvement tool.

TAKE-HOME POINTS

- Learning is the most important outcome that can be achieved using RADPEER data; education sessions are at the heart of serious peer learning processes.
- The revised RADPEER scoring system replaces the previous 4-point scale with a 3-point scale, effectively merging previous scores of 3 and 4.
- The scoring system was updated to unburden reviewers from the process of determining the severity of an error and instead refocus efforts toward nonpunitive peer learning.
- eRADPEER updates include classifications by age, body system, and discrepancy type (perception, interpretation, communication) to facilitate learning conferences and focused improvements.

REFERENCES

1. Kohn LT, Corrigan JM, Donaldson MS, eds. *To err is human: building a safer health system*. Washington, DC: National Academy Press; 1999.
2. Balogh EP, Miller BT, Ball JR, eds. *Board on Health Care Services, Institute of Medicine. Improving diagnosis in health care*. Washington, DC: The National Academy of Sciences, The National Academy Press; 2015.
3. Halstead MJ. Radiology PEER review as an opportunity to reduce errors and improve patient care. *J Am Coll Radiol* 2004;1:984-7.
4. Borgstede J, Lewis R, Bhargavan M, Sunshine J. RADPEER quality assurance program: a multifacility study of interpretive disagreement rates. *J Am Coll Radiol* 2004;1:59-65.
5. Radiology Quality Institute. *Diagnostic accuracy in radiology: defining a literature-based benchmark*. Beachwood, OH: Radiology Quality Institute; 2012.
6. Berlin L. Radiologic errors: acceptable practice or malpractice. *In Pract* 2007;1:5-7.
7. Siegle R, Baram E, Reuter S, Clarke E, Lancaster J, McMahan A. Rates of disagreement in imaging interpretation in a group of community hospitals. *Acad Radiol* 1998;5:148-54.
8. Edwards MT. Peer review: a new tool for quality improvement. *Physician Exec* 2009;35:54-9.
9. Jackson VP, Cushing T, Abujudeh HH, Borgstede JP, Chin KW, Grimes CK, et al. RADPEER™ scoring white paper. *J Am Coll Radiol* 2009;6:21-5.
10. Altman D, Clancy C, Blendon R. Improving patient safety-five years after the IOM report. *N Engl J Med* 2004;351:2041-3.

11. Loreto M, Kahn D, Glanc P. Survey of radiologist attitudes and perceptions regarding the incorporation of a departmental PEER review system. *J Am Coll Radiol* 2014;11:1034-7.
12. Abujudeh H, Pyatt RS Jr, Bruno MA, et al. RADPEER PEER review: relevance, use, concerns, challenges, and direction forward. *J Am Coll Radiol* 2014;11:899-904.
13. Fotenos A, Nagy P. What are your goals for peer review? A framework for understanding different methods. *J Am Coll Radiol* 2012;9:929-30.
14. Larson PA, Pyatt RS Jr, Grimers CK, Abudujeh HH, Chin KW, Roth CJ. Getting the most out of RADPEERTM. *J Am Coll Radiol* 2011;8:543-8.
15. Berlin L. Malpractice and radiologists. *AJR Am J Roentgenol* 1980;135:587-91.
16. Berlin L. Malpractice and radiologists, update 1986: an 11.5-year perspective. *AJR Am J Roentgenol* 1986;147:1291-8.
17. Berlin L, Berlin JW. Malpractice and radiologists in Cook County, IL: trends in 20 years of litigation. *AJR Am J Roentgenol* 1995;165:781-8.
18. Berlin L, Hendrix RW. Perceptual errors and negligence. *AJR Am J Roentgenol* 1998;170:863-7.
19. Renfrew DL, Franken EA Jr, Berbaum KS, Weigelt FH, Abu-Yousef MM. Error in radiology: classification and lessons in 182 cases presented at a problem case conference. *Radiology* 1992;183:145-50.
20. Fitzgerald R. Error in radiology. *Clin Radiol* 2001;56:938-46.
21. Whang JS, Baker SR, Patel R, Luk L, Castro A 3rd. The causes of medical malpractice suits against radiologists in the United States. *Radiology* 2013;266:548-54.
22. Brenner RJ, Lucey LL, Smith JJ, Saunders R. Radiology and medical malpractice claims: a report of the practice standards claims survey of the Physician Insurers Association of America and the American College of Radiology. *AJR Am J Roentgenol* 1998;171:19-22.
23. Larson DB, Donnelly LF, Podberesky DJ, Merrow AC, Sharpe RE Jr, Kruskal JB. Peer feedback, learning, and improvement: answering the call of the institute of medicine report on diagnostic error. *Radiology* 2017;282:1-11.