Examining the Impact of the 2016 ACR RADPEER® Scoring Update in Practice

Tanya Hioe, DO
Andrew Kent Moriarity, MD

RSNA 2020

Advanced Radiology Services PC
College of Human Medicine
MICHIGAN STATE UNIVERSITY
Spectrum Health

Background

- ACR RADPEER® is currently the leading method for radiology peer review in the United States and uses a numeric score-based approach to assess diagnostic accuracy of imaging reports
- The initial 4-point scale was revised in 2009 with terminology that was considered more widely applicable and outcomes based with a focus on if a diagnostic discrepancy was “clinically significant”
- In 2016, the scale was condensed to a 3-point scale with the goal of facilitating non-punitive learning as recommended by the International Board of Medicine with an expanded focus on discrepancy classification

Background

- The purpose of this study was to determine how the 2016 change impacted radiologist-assigned scores before and after implementation of the 2016 changes in a private practice setting

2002 RADPEER® Scoring

SCORE	INTERPRETATION
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

Background

2009 RADPEER® Scoring

SCORE	INTERPRETATION	MODIFIER
1	Concur with interpretation	N/A
2	Discrepancy in interpretation/not ordinarily expected to 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

2016 RADPEER® Scoring

SCORE	INTERPRETATION	MODIFIER
1	Concur with interpretation	N/A
2	Discrepancy in interpretation/not ordinarily expected to 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

Methods

- Peer review scores for 6-months **before** and **after** implementation of the 2016 RADPEER® score consolidation
- To monitor performance of the scoring during each period, all discrepancy scores >2 during the pre-period and >1 during the post-period were adjudicated by either the subspecialty section chief or the quality committee chair
- Scores of each category were compared to measure significant change
  - As the revision merged the “3” & “4” categories into a singular “3” category, the pre-revision “3” & “4” categories were totaled and compared against the post-revision “3” category

Results

Total number of peer reviews: **21,003**

Pre-Revision (July-Dec 2019)

Reviews: **11,498** (55% of total)

Score	# Cases Reviewed	Percentage
1	11,186	97.29%
2	284	2.47%
3	24	0.21%
4	4	0.03%

Post-Revision (Jan-Jun 2020)

Reviews: **9,505** (45% of total)

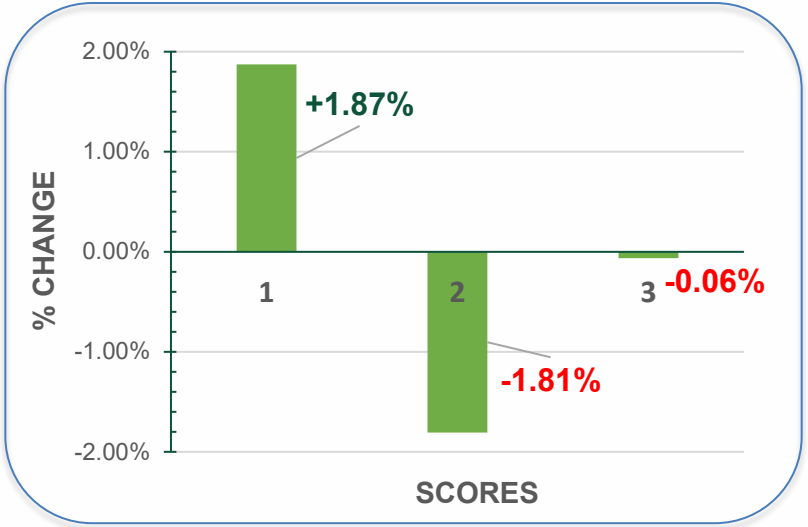
Score	# Cases Reviewed	Percentage
1	9425	99.16%
2	63	0.66%
3	17	0.18%

Results

- Score category consolidation resulted in an increased proportion of “1” scores and decreased proportions of “2” and “3” scores

Score	Pre-Revision %	Post-Revision %	% Change
1	97.29%	99.16%	+1.87%
2	2.47%	0.66%	-1.81%
3	0.24%*	0.18%	-0.06%

*Total of “3” & “4” scores in the pre-revision period



Discussion

- Revision shifted more scores to “1” and away from “2”in the post-revision period
- Although the “3” and “4” scores in the pre-revision were already low, the equivalent scoring was further decreased post-revision
- Comparison across the revised sub-categories of “a” (unlikely to be clinically significant) and “b” (likely to be clinically significant) was not significantly different between the periods
- There were twice as many cases regraded by the subspecialty during the post-revision, 0.5% (48) compared with 0.2% (23) in the pre-revision which may reflect unfamiliarity with the new scoring system

Conclusion

- The updated 2016 RADPEER® scoring sought to shift reviewer focus from determining the severity of an error towards nonpunitive peer learning
- In clinical practice, this change resulted in non-significant changes to score classification and did not increase the number of peer-learning opportunities when applied to a randomly selected population of cases
- Future revisions are needed to refine the process in support of peer learning

References

1. Goldberg-Stein S, Frigini LA, Long S, Metwalli Z, Nguyen XV, Parker M, Abujudeh H (2017) ACR RADPEER committee white paper with 2016 updates: revised scoring system, new classifications, self-review, and subspecialized reports. J Am Coll Radiol 14(8):1080–1086