Redefining the Practice of Peer Review Through Intelligent Automation Part 2: Data-Driven Peer Review Selection and Assignment

Bruce I. Reiner¹

Published online: 27 July 2017
© Society for Imaging Informatics in Medicine 2017

Abstract In conventional radiology peer review practice, a small number of exams (routinely 5% of the total volume) is randomly selected, which may significantly underestimate the true error rate within a given radiology practice. An alternative and preferable approach would be to create a data-driven model which mathematically quantifies a *peer review risk score* for each individual exam and uses this data to identify high risk exams and readers, and selectively target these exams for peer review. An analogous model can also be created to assist in the assignment of these peer review cases in keeping with specific priorities of the service provider. An additional option to enhance the peer review process would be to assign the peer review cases in a truly blinded fashion. In addition to eliminating traditional peer review bias, this approach has the potential to better define exam-specific standard of care, particularly when multiple readers participate in the peer review process.

Keywords Peer review · Data mining · Report analysis · Workflow distribution

Introduction

In conventional radiology practice, peer review is routinely performed on 5% of all imaging exams, as required by the Joint Commission [1]. This results in an exceedingly large number of cases (up to 95%) being excluded from formal peer

review, with the potential for clinically significant diagnostic errors going undetected. In addition to this small sample size, conventional peer review is further limited by the customary practice of randomized selection, which is theoretically designed to represent the full spectrum of radiology practice [2]. In reality, however, certain radiology exams have a greater potential for diagnostic error based upon a variety of factors including the interpreting radiologist, exam type, accessible data, clinical indication, technical exam attributes, and patient profile. While isolated attempts have been made to address this limitation in peer review selection through mathematical modeling [3], no comprehensive strategy has been described to date which takes into account the myriad of factors which can be used to define the potential diagnostic error rate in radiology reporting (i.e., peer review risk). Creation of such a data-driven model provides an opportunity for radiology providers to create statistical peer review risk scores for individual exams a priori, which in turn can be used for targeted and proactive peer review at the point of care.

Defining Diagnostic Error Risk Factors

The primary goal of “targeted” peer review is to identify specific factors which may result in a higher than normal likelihood of report diagnostic error and use this information to create a data-driven *Peer Review Risk Score*, which can serve as an objective trigger for peer review selection and assignment. Figure 1 lists a number of data categories which can be used in the quantification of peer review risk, each of which may in turn contain a number of sub-variables. As an example, for the category Patient Profile variables used to quantify peer review risk may include patient age, current clinical status, comorbidities, prior surgery, body habitus, and compliance.

✉ Bruce I. Reiner
breiner1@comcast.net

¹ Department of Radiology, Veterans Affairs Maryland Healthcare System, 10 North Greene Street, Baltimore, MD 21201, USA

1. Patient Profile
2. Clinical Indication
3. Exam Type
4. Radiologist of Record
5. Report Uncertainty
6. Historical Report Data
7. Follow-up Recommendations
8. Technical Limitations
9. Medicolegal Risk
10. Workflow Outliers

Fig. 1 High priority peer review indicators

The data elements associated with each category could be extracted from a variety of sources including the computerized physician order entry system (CPOE), radiology information system (RIS), picture archival and communication system (PACS), electronic medical record (EMR), and radiology reports. Both historical and current radiology reports would provide important information in quantifying peer review risk by providing data related to current and past imaging findings, clinical data, follow-up recommendations, technical limitations, and radiologist uncertainty. Examples of “report triggers” or indicators for high peer review risk may include critical results findings (e.g., appendicitis, cancer), follow-up recommendations (e.g., repeat CT in 3 months, biopsy), technical limitations (e.g., motion artifact, suboptimal contrast opacification), and radiologist uncertainty (e.g., “cannot exclude”, “appears to be”). In addition to these text-based “report triggers”, exam-specific workflow data may provide additional high risk triggers; such as unexpectedly shortened exam interpretation time, failure to review portions of the imaging dataset, and non-utilization of decision support technology.

A variety of statistical methods [3–5] and data mining technologies [6–8] can be used for automated data extraction and risk quantification, which can be iteratively refined through continuous analysis of the peer review database. The peer review database could be created by longitudinal and continuous collection of peer review and outcomes data [9], which provides the ability to search and analyze data based upon the 10 categories of data defined in Fig. 1. In addition to using this data for generalizable peer review risk scores, comparable data specific to each individual radiologist can be used to create *radiologist-specific* peer review risk scores, taking into account each individual radiologist’s exam-specific peer review analytics.

Peer Review Selection and Prioritization

The resulting data from peer review risk quantification (i.e., peer review risk scores) can in turn be used as one important component in the peer review selection process, with the intended goal of proactively identifying and intervention in

high risk exams. If effectively accomplished and promptly acted upon, one could theoretically use the knowledge gained in the peer review process to *identify and correct* diagnostic errors *before* clinical management decisions and actions are taken. In essence, this would become a data-driven format for selectively identifying high risk exams for “second reads”, in which both the primary reader and designated peer reviewer independently interpret the same exam, with the corresponding reports reviewed and analyzed before the *corrected* final report is issued.

A number of logistical and workflow constraints exist in implementing this type of system, the foremost of which include time and workload. Since current practice requires prompt report turnaround measured in minutes, the implementation of such a system would require a priori or immediate identification of the high risk exam, followed by rapid selection and assignment of the appropriate peer reviewer, and an automated technology for comparative report review and reconciliation (which will be discussed in article 3 of this series). Equally important are workload and financial limitations, which mandate that high levels of productivity are maintained in the face of low reimbursements and the additional peer review requirements are minimally taxing to already stress radiologist resources.

One strategy for addressing these challenges is to create a standardized method for hierarchical grading of “high risk” exams, which take into account not only the degree (or magnitude) of peer review risk but also the potential clinical ramifications. As an example, in the interpretation of a “high peer review risk” chest CT exam, the finding of pulmonary nodule is clinically important and non-emergent, whereas the finding of aortic dissection is both emergent and of high clinical significance. As a result, the report turnaround requirements for the “high peer review risk” pulmonary nodule case may be categorized as *routine*, while the aortic dissection may be categorized as *urgent*. A representative schema for classifying these report turnaround requirements is listed in Fig. 2 and provides a method for limiting workload and time constraints for designated “high risk” peer review exams.

In addition to computerized (i.e., automated) peer review selection through quantification of peer review risk scores, peer review selection can be performed manually. Manual selection of peer review cases can be initiated by a variety of authorized sources including the radiologist of record (i.e., primary reader), referring clinician, or radiologist colleague. As an example, during the course of the *report lifetime*, a peer

- A. Highly Emergent (15 minutes)
- B. Emergent (1 hour)
- C. Urgent (12 hours)
- D. Short Term (24 hours)
- E. Routine (72 hours)

Fig. 2 Representative schema for “High Risk” peer review turnaround time requirements

review request may be generated during primary interpretation (by the radiologist of record), during the time the report is read (by the referring clinician), or at the time of exam follow-up (by the radiologist colleague). These manual peer review requests would be triggered by uncertainty and/or disagreement of report findings. As a method of guarding against excessive use, these manual peer review requests are recorded in the peer review database and analyzed for peer review outcomes, in order to ensure that utilization is appropriate.

The third method for peer review selection is *randomized* selection, which is intended to maintain a broad base of selected cases. In the absence of random peer review selection, exam type and reader distribution of peer review cases may become skewed in favor of more complex exam types and readers with relatively poor peer review performance. Randomized selection is maintained to ensure that the cases selected are reflective of the case and reader mix as a whole. The final determination of relative frequency of automated, manual, and random peer review case selection should be determined by practice requirements and refined through outcomes analysis of the Peer Review Database. The ultimate goal is to create a balance between optimizing clinical outcomes, workflow, and available resources.

Peer Review Assignment

Peer review assignment can be random and/or targeted based upon the specific circumstances of the case in question and the preferences of the institutional and/or individual provider of record. Randomized peer review assignment allows for automated selection of a peer reviewer irrespective of their individual attributes, institutional profile, or performance record. Targeted peer review assignment provides a mechanism for selecting peer reviewers based upon specific search criteria of interest (Fig. 3). These search criteria may be determined by a variety of individuals including the primary reader (i.e., radiologist of record), institutional provider, referring clinician, or hospital administrator. The theoretical advantage of targeted peer review assignment is that it provides a data-driven tool for selecting peer reviewers based upon specific search criteria. As an example, if a radiology department administrator receives a complaint from a referring clinician about a

1. Clinical Experience
2. Institutional Demographics
3. Geographic Location
4. Technology in Use
5. Education and Training
6. Exam Volume
7. Historical Peer Review Performance
8. Peer Reviewer Availability

Fig. 3 Variables for targeted peer review assignment

possible missed finding on an imaging report and wants a neutral and unbiased second opinion, the case could be directed for peer review. In an attempt to ensure that the second reader practices in a “comparable” practice setting, the administrator may request primary search criteria of “Institutional Demographics” and secondary search criteria of “Technology in Use”. By doing so, the peer review database can be searched to identify potential reviewers whose institutional demographics and technology match those of the host institution.

In certain unique circumstances, peer review may be requested by non-medical parties in an attempt to better define the standard of care. Examples may include a third party payer seeking verification as to the necessity of recommended follow-up tests or a plaintiff attorney in a medical malpractice case. In the latter example of an alleged missed finding, the plaintiff attorney may request blinded peer review by a variety of reviewers from different practice backgrounds and experience levels in an attempt to better define the standard of care from a diverse number of practicing radiologists.

As previously mentioned, emergent and “high peer review risk” cases may be selected for real-time peer review as a means to provide second opinions to assist the primary reader at the point of care. When time is of the essence, the search criteria can include both peer reviewer performance data as well as availability for expedited peer review. In order to minimize peer review workload for individual peer reviewers, the workflow engine integrated with the Peer Review Database can continuously track peer review assignments for individual peer reviewers, with the goal of evenly distributing peer review cases across the large spectrum of potential peer reviewers. Since all peer review reports are independently created without pre-existing knowledge of the primary reader report, both the primary and peer reviewer reports are subjected to “agnostic” finding-specific peer review analysis. This ensures that the peer review process is equal and fair to all participants, while simultaneously recording peer review report data for both the primary reader and peer reviewer. As a result, all assigned peer reviewers are in effect being subjected to the same peer review process as the primary readers they are being tasked to evaluate.

Conclusion

The goal of the proposed peer review model is the creation of a dynamic system which provides for a variety of selection techniques including automated peer review selection based upon quantifiable peer review risk, manual selection based on end-user identification, and randomized selection from the broad case pool. Incorporating institutional and individual provider workflow data provides a method of ensuring that peer review case distribution is synergistic with regular exam

workload. At the same time, exams determined to be both high risk for peer review discrepancy and of emergent clinical status can automatically trigger real-time targeted peer review, thereby providing an expert “second read” with the goal of expediting accurate diagnosis at the point of care so that effective clinical management is not delayed.

Peer review assignment is also dynamic in nature, with a variety of methods and criteria used for analyzing peer reviewer status and availability. In addition to utilizing finding-specific performance data derived from the peer review database, an integrated peer review workflow engine can assist in determining individual peer reviewer availability, exam backlog, and expected report turnaround time. The ultimate goal is to provide balance and data-driven intelligence to both peer review selection and assignment, while maximizing clinical outcomes through early and effective identification and intervention of report discrepancies.

References

1. McEnery KW, Sutor CT, Hildebrand S et al.: Integration of radiologist peer review into clinical review workstation. *J Digit Imaging* 13:101–104, 2000
2. Sheu YR, Feder E, Balsim I et al.: Optimizing peer review: a mathematical model for selecting future cases based on prior errors. *J Am Coll Radiol* 7:431–438, 2010
3. Zhou KH, O'Malley AJ, Mauri L: Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115:654–657, 2007
4. Glas AS, Lijmer JG, Prins MH et al.: The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiology* 56:1129–1135, 2003
5. Lasko TA, Bhagwat JG, Zou KH et al.: The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform* 38: 404–415, 2005
6. Friedman C, Alderson PO, Austin JHM et al.: A general natural language text processor for clinical radiology. *J Am Med Inform Assoc* 1:161–174, 1994
7. Gong T, Tan CL, Leong TY, et al: Text mining in radiology reports. In: 2008 Eighth IEEE International Conference on Data Mining; Dec 15–19. Pisa: Data Mining, pp 815–820, 2008
8. Reiner B: Uncovering and improving upon the inherent deficiencies of radiology reporting through data mining. *J Digit Imaging* 23:109–118, 2010
9. Reiner B: Redefining the practice of peer review through intelligent automation. Part 1: Creation of a Standardized Methodology and Referenceable Database. *J Digit Imaging* 2017 (in press)