
Volume 11 - Issue 4, 2011 - Management Matters

Auditing Radiology Reports

Author

Dr. Howard Galloway

Chief Radiologist Imaging Partners

UK

galloway.howard@gmail.com

The advent of teleradiology and the outsourcing of reporting have increased focus on the audit of radiology reports. However, thus far there has been little examination of the evidence that could underpin a rational approach to audit. Audit data as well as performance monitoring provide information that may be used to systematically improve the quality of radiology reporting.

Four major questions arise when considering how this might be achieved:

1. What are the reported rates of discrepancy and are they comparable, appropriate or meaningful?
2. Can acceptable performance levels be determined on the basis of what is reported in the literature?
3. What are the current methods of audit and what are the current classification systems?
4. What would the features of a system that could potentially inform evidence-based interventions be?

Studies Report on Discrepancy Rates

A large number of reported studies compare the performance of residents with specialists or specialists with sub-specialists, which do not reflect most common situations of peer review. In addition, studies from large teleradiology services rely on client sites to report discrepancies, which may not provide a true indication of discrepancy rates. The largest series of 124,870 cases reported by 10 radiologists from a U.S. teleradiology provider reports an average discordance rate of 1.0 percent, with body CT having the highest rate of 2.1 percent (Wong, 2005). This is the lowest rate reported in the literature but relies on client sites for overreading.

Branstetter (2007) demonstrated a significant discordance rate between preliminary and final reports for senior radiologists in an emergency situation of 88 / 3,587 (2.5 percent) for CT and 3 / 253 (0.5 percent) for CR in a major trauma centre. This is similar to the ACR RadPeer self-reported rate from institutions and practices across the U.S. of 2.1 percent for major discrepancies across all modalities. Stevens (2008) summarised previous studies with a variety of selection criteria showing a major discrepancy rate from 0.5 to five percent with a mean of 1.79 percent. However, even among staff radiologists, there is substantial inter observer variation (between 2.1 percent to 23 percent) in the interpretation of cross-sectional imaging studies

In terms of the level of discrepancy that might require remedial action, there is very little in the literature. Anecdotally, concerns about reporting performance are raised on the basis of individual cases rather than review of overall performance. One study that may provide an indication is that of Siegle (1998) who reported a review of 1,100 studies by 35 community radiologists over seven years showing a 4.4 percent mean rate of disagreement with three percent felt to be below standard of care.

From the reported rates, it would seem that a significant discrepancy rate of two - three percent for body CT would be an acceptable level of performance. What constitutes an unacceptable level of performance is less clear, although anecdotally, in some organisations levels of or above four percent on random sampling prompt a more systematic performance review.

Current Audit Process Insufficient

The current audit process identifies discrepancies but does not provide sufficient information for specific process improvement. Any current positive effect of audit is likely to be due to a non-specific manifestation of the Hawthorne effect where process improves only because it is being watched. General studies demonstrate only small changes in doctor behaviour with the provision of various forms of feedback. Uniformly negative feedback has been found in many circumstances to produce a paradoxical effect with decreased motivation, disengagement and a resulting decrease in performance.

What systems may be used to classify discrepancies and what potential do they have to inform systematic quality improvement? There are three main ways of classifying discrepancy data. These are:

1. Clinical Impact

This is the system most commonly used and is advised by the NHS. This system has potential utility particularly when liaising with clinicians and client departments and determining clinical follow-up action. Problems, however, include inconsistency and the difficulty of assessing the clinical impact from the request and radiological data alone. There are issues with peer review radiologists having sufficient expertise to reliably and consistently classify clinical impact. The most important problem is that it does not provide information about the possible cause of the discrepancy.

2. Peer Assessment

The American College of Radiology RadPeer project uses a system based on the reviewing radiologists' opinion of the difficulty of the diagnosis. This shows that on random review radiologists concur with the original interpretation ~97 percent of cases. This has the value of incorporating a peer assessment of the difficulty of the case but does not provide more specific information to inform causal analysis of discrepancies.

3. Causal Classifications

Unlike classification systems based on putative clinical impact, a classification based on proposed cause offers the opportunity to undertake targeted action for quality improvement.

Alternative Classification Systems

Renfrew (1992) proposed a classification system that tries to assign possible causes based on decision analysis methodology. This has consistently shown false negative errors and perceptual / cognitive factors to be the major component of discrepancies. More sophisticated variations on this theme have been used e.g. the Australian AIMS system, and are forming the basis of the proposed WHO classification. However, personal experience of attempting to apply these systems to radiology has shown that they are too complicated for routine use. A simplified system based on Renfrew's paper classifying discrepancies into three broad categories as false positive, false negative or misattribution with further simple sub-classification is proposed.

Errors due to misattribution would be expected to be amenable to focused education or feedback from clinical outcomes. False positive errors likewise may represent failure to appreciate normal variations, which can be dealt with by education.

Mammography screening has demonstrated that feedback of operative results can inform the refinement of callback rates. False negative discrepancies are the most common. It is tempting to ascribe such errors to lack of care and attention. This, however, is not productive, as errors have been consistently reported to occur at a relatively constant rate over time and across different individuals. The majority of false negative discrepancies are apparently perceptual errors often in the setting of multiple abnormalities either related or unrelated.

Attempts to reduce false negative perceptual errors could focus on both systematic and individual factors. The work environment may contribute with poor lighting, cramped conditions and multiple distractions playing a role. Individual factors such as fatigue or eyesight illness may also play a role. Workload itself may be a major factor and there are surprisingly little data to correlate workload with discrepancy rates. One study showed that the discrepancy rate for body CT studies doubled faculty radiologists reported more than 20 studies a day but this was not statistically significant

Concluding Thoughts

The need for and inevitability of radiology reporting audit is now generally accepted. The challenge is to perform audit in a statistically valid manner and collect data that provides information that can potentially inform rational process improvement. Human factors are central in this and the effects of various forms of feedback and the need for radiologists to have professional and individual ownership of the process is vital if audit is to achieve the key goal of improving patient care.

Published on : Fri, 30 Dec 2011