Diagnostic errors: the effect of shift length, schedule and volume

By Dr C Lamouruex, Dr TN Hanna & Dr JO Johnson

With the increased utilization of medical imaging over the past several decades, efforts to optimize patient care through improving diagnostic imaging accuracy are more relevant than ever.

More than 85 million CT examinations were performed in the United States in 2011 alone [1], meaning that even a 0.01% improvement in diagnostic accuracy could affect 8500 patients. The diagnostic radiologist is thus faced with not only the challenge of interpreting increased numbers of imaging exams per hour and per day but also of meeting the demand for coverage 24 hours a day, seven days a week. The reality is that increased interpretive volumes per radiologist are needed [2,3] as well as variable schedules and shift lengths.

This article summarizes the results of a recent study of the effect of exam volumes, shift length and schedule on radiologist performance and diagnostic accuracy.

Studies evaluating the effect of interpretive volume on diagnostic accuracy of residents and attending physicians have shown a relationship between diagnostic errors and increased volume. In one such study by Fitzgerald [4] there were increased errors in CT interpretation when more than 20 were read per day. Increased speed of interpretation was also linked to a 26.6% increase in errors in 80% of radiologists [5]. Shift length has previously been shown to affect diagnostic errors toward the end of a shift. For example, Ruutiainen *et al.* [6] showed a relationship between resident diagnostic errors in the last 2 hours of an overnight shift. Studies of interns, residents and nurses have demonstrated a relationship between errors and shift schedule, with increased errors associated with overnight shifts [7], and with shifts longer than 12-12.5 hours [8,9]. Most prior work to examine these factors in radiology was compiled through resident data [10]. However, data regarding the prevalence of diagnostic errors by attending-level radiologists in general is limited. Notable exceptions include the American College of Radiology (ACR) RADPEER program launched in 2002 which reports important diagnostic errors in 0.39% of exams interpreted by participating radiologists through 2007 [11].

" ... the most errors occurred in the last 10% of the shift..."

DESIGN AND METHODOLOGY

The major goal of our retrospective study was to determine whether there was an association between attending level radiologist performance and effects of volume, shift length and schedule in a large teleradiology practice, Virtual Radiologic Professionals, LLC (vRad) [12] At the time of the study (calendar year 2015), there were 370 ABR-certified radiologists in the practice providing interpretations 24 hours a day/7 days a week. Teleradiologists provided preliminary interpretations. A second radiologist, from a client practice, subsequently rendered a final interpretation. Through a rigorous quality assurance process between the client and teleradiology practices, 4294 major discrepancies out of a total pool of 2,922,377 imaging exams were identified. Major discrepancies are defined as a significant diagnostic interpretive error with implication or possible implication on patient care. Each major discrepancy was vetted through a formal vRad QA Committee and the severity of the discrepancy was determined. Examinations with major discrepancies were categorized as acute or chronic with an urgency description. Data collection included shift start and end time and date of shift, shift type (regularly scheduled, holiday, weekend, extra, backup) and shift volume. Extra shifts

The Authors

Dr C Lamoureux[1], Dr TN Hanna[2] & Dr JO Johnson[2]

1. Virtual Radiologic, Eden Prairie, Minn. U.S.A.

2. Division of Emergency Radiology, Department of Radiology and Imaging Sciences,

Emory University Midtown Hospital, Emory University,

550 Peachtree Rd, Atlanta, GA, USA

Communicating Author : Dr Christine Lamoureux.

email: *Christine.Lamoureux@vrad.com*

were defined as shifts of variable lengths worked before or after a regular shift or at times on days when a regular shift was not scheduled. A backup shift was defined as a shift for which the radiologist was scheduled to interpret a variable number of exams only if needed during the shift period,.

RESULTS

Overall Error Rate:
The overall rate of major diagnostic errors was 0.15%. Out of a total of 2,922,377 examinations interpreted, there were 4294 major diagnostic discrepancies were identified (0.15%). Given the differences in study and practice mix, this 0.15% error rate is comparable to the reported RADPEER rate of 0.39%. CT examinations made up 79.9% of these exams and had the highest number of discrepancies (86.6%).

Effects of Shift Length on Error Rate:
The mean length of a radiologist shift where a major discrepancy occurred was 8.97 +/-2.28 hours, and on average, errors occurred 9 hours into a shift, with the peak incidence between 10-12 hours. Significantly more errors occurred late in shifts than occurred early (P<.0001). When shifts were divided into tenths, the fewest errors occurred in the first 10% of the shift, and the most errors occurred in the last 10% of the shift.

Effect of Diagnostic Volume on Error Rate:
The mean volume of studies was higher in those shifts with major discrepancies then those without major discrepancies (P <0001). The mean study volume for all shifts for the year was 67.60 +/- 60.20, with a mean studies per hour of 10.95+/- 6.81. The mean study volume for shifts in which major discrepancies occurred was 118.96 +/- 66.89 exams, with a mean studies per hour of 13.06 +/-6.10. Moreover, the mean number of studies per hour significantly differed for shifts with a major discrepancy compared with the mean per-hour read rate for the year (P<.0001), and the incidence of multiple errors in given shift increased with study volume in those shifts where an error occurred (P<.001).

Effect of Number of Days Worked and Shift Type:
The mean number of prior consecutive days in which a radiologist worked was 3.38 days +/-3.23 days (range 0-14 days). The number of errors did not differ as a function of the number of days worked consecutively (P<.0893), however, there was a significant difference in error rate as a function of shift type, with more errors occurring during regular shifts (which were often longer than the other types of shifts). For shifts with at least one error, there was no significant difference in errors as a function of modality, whether the discrepancy was acute or chronic or had a critical finding. Regarding urgency description, exams with trauma protocols resulted in more multi-error shifts than other types of studies.

LIMITATIONS

The practice setting for our study introduced some limitations. Because the data was from a teleradiology practice, there was an over-representation of night shifts compared with conventional practices impacting the modality mix (fewer x-rays and MRIs at night). An increase in the acuteness of the exam may also be more prevalent at night, altering the distribution and types of interpreted pathologies. Fewer distractions from phone calls, consultations and lack of performance of procedures likely increased the concentrated reading time and increased volumes read by the teleradiologist. The nature of the preliminary interpretation versus a final interpretation may also be a limitation of this study, as the second radiologist generating the final read had access to but may or may not have referenced the preliminary interpretation which could lead to decreased discrepancy rates because of alliterative cognitive bias. Other potential contributing factors include more clinical data available to the second radiologist and the benefit of time.

STUDY IMPLICATIONS

Interpretive errors are a rare but real occurrence in diagnostic radiology. As radiologists we are continuously seeking ways to improve performance. This study gives us some tools to optimize shift, schedule, and volume parameters in our own practices:

PRACTICAL ADVICE

- Read slower: closer to 11 studies an hour then 13 (particularly when there is a high proportion of CT and trauma protocols) and closer to 70 studies a shift then 120.
- Shorten shifts: diagnostic errors peak between 10-12 hours into a shift.
- Be vigilant late in shifts: the last 10% of the shift had more errors then any other 10% shift period.
- Take a break: although not statistically significant, the study shows that consecutive days work trended toward worsening interpretive performance.

REFERENCES:

1. IMV Medical Information Division. IMV CT Market Outlook Report. IMV Medical Information Division; Des Plaines, IL;2012.
2. Sunshine JH, Burkhardt JH. Radiology groups' workload in relative value units and factors affecting it. Radiology 2000; 214(3): 815-822.
3. McDonald RJ, Schwartz KM, Eckel LJ, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Acad Radiol 2015 22(9): 1191-1198.
4. Folkard S, Lombardi DA. Modeling the impact of the components of long work hours on injuries and "accidents." Am J Ind Med 2006: 49(11): 953-963.
5. Fitzgerald R. Error in radiology. Clin Radiol 2001;561(12)938-946
6. Sokolovskaya E, Shinde T, Ruchman RB, et al. The effect of faster reporting speed for imaging studies on the number of misses and interpretation errors: a pilot study. J Am Coll Radiol 2015; 12(7): 683-688.
7. Ruutiainen AT, Durand DJ, Scanlon MH, Itri JN. Increased error rates in preliminary reports issued by radiology residents working more than 10 consecutive hours overnight. Acad radiol 2013; 20(3): 305-311.
8. Olds DM & Clarke SP. The effect of work hours on adverse events and errors in health care. J Safety Res 2010; 41(2): 153-162.
9. Scott LD, Rogers AE, Hwang WT, Zhang Y. Effects of critical care nurses' work hours on vigilance and patient's safety. Am J Crit Care 2006; 15(1): 30-37
10. Hanna TN, Loehfelm T*, Khosa F, Rohatgi S, Johnson JO. Overnight shift work: factors contributing to diagnostic discrepancies. Emer Radio. 2015; 23(1): 41-47.
11. Jackson VP, Cushing T. Abujudeh HH, *et al*. RADPEER scoring white paper. J Am Coll Radiol 2009; 6(1): 21-25.
12. Hanna TN, Lamoureux C, Krupinski EA, Weber S, Johnson JO. Effect of shift, schedule, and volume on interpretive accuracy: a retrospective analysis of 2.9 million radiologic examinations. Radiology 2018; 287(1): 205-212.