Effect of Shift, Schedule, and Volume on Interpretive Accuracy: A Retrospective Analysis of 2.9 Million Radiologic Examinations[1]

Tarek N. Hanna, MD*
Christine Lamoureux, MD*
Elizabeth A. Krupinski, PhD
Scott Weber, BA
Jamlik-Omari Johnson, MD

Purpose: To determine whether there is an association between radiologist shift length, schedule, or examination volume and interpretive accuracy.

Materials and Methods: This study was institutional review board approved and HIPAA compliant. A retrospective analysis of all major discrepancies from a 2015 quality assurance database of a teleradiology practice was performed. Board-certified radiologists provided initial preliminary interpretations. Discrepancies were identified during a secondary review by a practicing radiologist or through an internal quality assurance process and were vetted through a consensus radiology quality assurance committee. Unique anonymous radiologist identifiers were used to link the discrepancies to radiologists' shifts and schedules. Data were analyzed by using analysis of variance, *t* test, or χ^2 test.

Results: A total of 4294 major discrepancies resulted from 2922377 examinations (0.15%). There was a significant difference for shift length ($P < .0001$) and volume ($P < .0001$) for shifts with versus those without discrepancies. On average, errors occurred a mean (± standard deviation) of 8.97 hours ± 2.28 into the shift (median, 10 hours; interquartile range, 2.0 hours). Significantly more errors occurred late in shifts than early ($P < .0001$), peaking between 10 and 12 hours. The number of major discrepancies in a single shift ranged from one to four, with a significant difference in the number of discrepancies as a function of study volume (volume for all shifts, 67.60 ± 60.24; volume for shifts with major discrepancies, 118.96 ± 66.89; $P < .001$). Despite a trend for more discrepancies after more consecutive days worked, the difference was not significant ($P = .0893$).

Conclusion: Longer shifts and higher diagnostic examination volumes are associated with increased major interpretive discrepancies. These are more likely to occur later in a shift, peaking after the 10th hour of work.

[1] From the Division of Emergency Radiology, Department of Radiology and Imaging Sciences, Emory University Midtown Hospital, Emory University, 550 Peachtree Rd, Atlanta, GA 30308 (T.N.H., E.A.K, J.O.J.) and Virtual Radiologic, Eden Prairie, Minn. (C.L., S.W.). Received March 9, 2017; revision requested April 24; revision received June 6; accepted July 11; final version accepted September 14. **Address correspondence to** T.N.H. (e-mail: *tarek.hanna@emory.edu*).

*T.N.H. and C.L. contributed equally to this work.

For diagnostic radiologists, accuracy of interpretation is a fundamental determinant of quality and competency, and peer review remains the central method for routinely surveying diagnostic accuracy in the clinical practice setting (1,2). Medical error is an important cause of death (3). Following a task force created in response to the seminal Institute of Medicine report, "To Err Is Human" (4), the American College of Radiology launched RADPEER to members in 2002 (5). The RADPEER database generates reports that provide summary statistics and comparisons for each radiologist by modality, summary data for each facility, and summative data for all participating facilities (5). According to cumulative RADPEER data through 2007, 0.39% of all studies received a score of 3 or 4, indicating a potentially important diagnostic misinterpretation (5).

There is substantial literature conceptualizing and categorizing radiologic errors (6–9); however, apart from the RADPEER data, few large studies have been performed to examine the discrepancy rates detected in double readings by practicing radiologists (10,11). A 2014 meta-analysis uncovered 58 prior articles and an aggregate staff major discrepancy rate of 2.4% for interpretations of findings in computed tomographic (CT) examinations in adults (11). Soffa et al (10) included all modalities in their study and found a major disagreement rate of 3.48%. Few previous studies in the radiology literature included time of interpretation, radiologist shift, radiologist schedule, or interpretative volume, which would have allowed insight into how these factors affect diagnostic errors. Most prior work to examine these factors was compiled through resident data (12,13). The primary aim of our study was to determine whether radiology shift length, schedule, and diagnostic volume for practicing board-certified radiologists are associated with major discrepancies. The results could be useful to practices to minimize error through evidence-based schedule and shift optimization.

Advances in Knowledge

- Longer shifts were more likely to be associated with errors than were shorter shifts (shifts without discrepancies versus with discrepancies had a mean length of 7.98 hours ± 2.63 vs 8.97 hours ± 2.28; $P < .0001$); on average, discrepancies occurred 9 hours into a shift and peaked between 10 and 12 hours into a shift.
- A significantly higher number of major discrepancies occurred during the latter portions of shifts ($P < .0001$).
- Shifts with major discrepancies had significantly higher absolute study volume ($P < .0001$) and higher study volume per hour ($P < .0001$); for all shifts, the total volume was 67.60 ± 60.24, with mean number of studies per hour of 10.95 ± 6.8.
- For shifts in which a major discrepancy occurred, the total volume was 118.96 ± 66.89, and the mean number of studies per hour was 13.06 ± 6.10; in shifts with one major discrepancy or more, the number of errors increased as a function of study volume ($P < .001$).
- There was no significant difference in number of errors as a function of consecutive days worked ($P = .0893$).

Materials and Methods

The institutional review board of Emory University in Atlanta, Ga, approved this retrospective, Health Insurance Portability and Accountability Act–compliant study. The requirement for informed consent was waived. All patient and reporting radiologist identifiers were removed from the records before they were received for analysis.

Implication for Patient Care

- The effect of shift length and diagnostic volume on major discrepancy rates should be considered when schedules are crafted and error patterns in individual practices are examined.

Participants

Data for our study originated from an international teleradiology practice consisting of 370 American Board of Radiology–certified radiologists who provide interpretations 24 hours per day, 7 days per week, 365 days per year. All preliminary interpretations are subsequently read by a second practicing radiologist. This second interpreting radiologist is then able to submit discrepancies, which are logged in a database.

All 4294 major discrepancies in preliminary examinations occurring in calendar year 2015 were extracted from the database (n = 2 922 377 total examinations) of this large teleradiology organization. Studies that were not read by a second radiologist made up approximately 0.054% of total preliminary study volume (1590 of 2 922 377). Included major discrepancies occurred from January 1, 2015, at 1:55 hours through December 31, 2015, at 22:39 hours. The quality assurance process by which retrospective discrepancies could be reported tracked the 2015 studies for 8 months, such that a discrepancy could be discovered and reported up to September 1, 2016, at which point the data were extracted for analysis.

https://doi.org/10.1148/radiol.2017170555

Content codes: HP SQ

Radiology 2018; 287:205–212

Author contributions:
Guarantors of integrity of entire study, T.N.H., C.L.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, T.N.H., C.L., J.O.J.; clinical studies, C.L.; statistical analysis, T.N.H., E.A.K., S.W., J.O.J.; and manuscript editing, T.N.H., C.L., E.A.K., J.O.J.

Conflicts of interest are listed at the end of this article.

See also the editorial by Arenson in this issue.

Table 1

Experience of Interpreting Radiologists

Time Since ABR Certification*	No. of Radiologists
0–4 years	74 (20)
5–9 years	94 (25.4)
10–14 years	89 (24.1)
15–19 years	41 (11.1)
20–24 years	34 (9.2)
25–29 years	25 (6.8)
30–34 years	10 (2.7)
35–39 years	2 (0.5)
≥45 years	1 (0.3)
Total no. of radiologists	370 (100)

Note.—Data in parentheses are percentages. ABR = American Board of Radiology.

* If time since ABR certification was unavailable, time since residency completion was used.

Table 2

Fellowship Training of Interpreting Radiologists

Fellowship Description	No. of Radiologists
No fellowship*	143 (38.6)
Abdominal imaging	68 (18.4)
Neuroradiology	39 (10.5)
Musculoskeletal	25 (6.8)
Other†	24 (6.5)
Vascular and interventional	21 (5.7)
MR imaging	18 (4.9)
Pediatric	9 (2.4)
Mammography/breast imaging	9 (2.4)
Nuclear medicine	5 (1.4)
Thoracic imaging	5 (1.4)
Emergency radiology	4 (1.1)
Total	370 (100)

Note.—Data in parentheses are percentages.

* This refers to formal fellowship training. Many of these radiologists practice as subspecialists based on experience.

† Includes ultrasonography (US), combination fellowships, and cross-sectional imaging.

Quality Assurance Process and Definition of a Major Discrepancy

All diagnostic radiologic examinations included in this study were preliminarily interpreted by one of 370 board-certified radiologists affiliated with the teleradiology practice. Experience of these radiologists is reported in Table 1 and fellowship training of interpreting radiologists is reported in Table 2. These examinations were subsequently reinterpreted by a separate practicing radiologist, who generated a final radiology report. These second radiologist readers are affiliated with groups or hospital systems that contracted with the teleradiology organization. If the second radiologist reader disagreed with the preliminary interpretation, he or she generated a quality assurance discrepancy submission to the teleradiology organization. A rare exception to this process occurred when the teleradiology client requested to have the same radiologist finalize his or her own preliminary interpretation; this occurred in a maximum of 0.054% (1590 of 2 922 377) of examinations. Less commonly, internal quality assurances were generated by internal board-certified radiologists who came across a study in their daily practice and reported it to the quality assurance committee.

Clients of the teleradiology practice are aware of the quality assurance system and are provided with information on how to categorize discrepancies. All studies submitted through the quality assurance process included a categorization as follows: *(a)* slight discrepancy with no effect on patient care; *(b)* minor discrepancy with subclassification: *(i)* N = no effect on patient care, *(ii)* Q = questionable effect on patient care, *(iii)* C = chronic (with effect on patient care), and *(iv)* A = acute (with effect on patient care); and *(c)* major discrepancy, with effect on patient care.

The discrepancy between preliminary and final reports (suggesting a potential interpretive error) was then reviewed by a panel of practicing board-certified radiologists in the teleradiology practice, and a consensus discrepancy category was determined. A major discrepancy included findings that practicing radiologists should make almost all the time and that affect patient care. In our study, we only examined studies categorized as having a major discrepancy after review by the panel.

Data Mining

For every major discrepancy, the original examination type (modality and study description), date of interpretation, interpretation time to the nearest second, and location of interpretation were obtained. The major discrepancy was categorized as acute or chronic. Acute findings are those that need to be acted on during the current health care visit and may have implications for patient care in minutes, hours, or days. Studies were also categorized according to an urgency description: nonemergent, emergent, highly emergent, stroke protocol, and trauma protocol. Every major discrepancy had an associated unique anonymous identifier, representing the original interpreting radiologist. A database containing radiologist shift and schedule information was created by using these anonymous identifiers.

This database contained the following: start date and time of shift, end date and time of shift, shift type (regularly scheduled, holiday, weekend, extra, backup), and shift volume. Shifts vary in length because radiologists work different regular shifts and can work extra shifts to cope with study volume. Backup shifts occur when another radiologist needs coverage for health or family reasons. Shift volume is defined as the total number of diagnostic examinations a radiologist interpreted during their shift. Extra shifts are defined as adding shift time before and/or after a regularly scheduled shift. These extra shifts can be of any length and include shifts of less than 1 hour. Backup shifts are added work hours during days off but usually occur during the shift hours usually worked by the radiologist in question. The individual radiologist's clinical calendar allowed the authors to calculate how many consecutive clinical shifts were worked before the shifts containing the major discrepancy.

Data Analysis

Data were analyzed by using analysis of variance or *t* tests for continuous data or with χ^2 tests for categorical data. Statistical software (StatView; SAS Institute, Cary, NC) was used, and a *P*

Table 3

Breakdown by Modality and Body Part for Major Discrepancies

Discrepancies per Modality	Value
Overall	
No. of examinations	2 922 377
No. of major discrepancies	4294
Discrepancy rate (%)	0.15
CT	
Body part imaged	
Abdomen and pelvis	1585
Head	729
Chest	626
Spine	337
Other	274
Face	169
No. of examinations	2 305 238
No. of major discrepancies	3720
Discrepancy rate (%)	0.16
Total discrepancies (%)	86.6
Radiography	
Body part imaged	NC
No. of examinations	234 661
No. of major discrepancies	245
Discrepancy rate (%)	0.10
Total discrepancies (%)	5.7
Magnetic resonance (MR) imaging	
Body part imaged	
Head	118
Spine	68
Other	42
No. of examinations	81 937
No. of major discrepancies	76
Discrepancy rate (%)	0.03
Total discrepancies (%)	1.7
US	
Body part imaged	
Pelvis	35
Vascular or other	32
Abdomen	9
No. of examinations	287 403
No. of major discrepancies	76
Discrepancy rate (%)	0.03
Total discrepancies (%)	1.7
Nuclear medicine	
Body part imaged	NC
No. of examinations	12 745
No. of major discrepancies	22
Discrepancy rate (%)	0.17
Total discrepancies (%)	0.5
Modality NC	
No. of examinations	393
No. of major discrepancies	3
Discrepancy rate (%)	0.76
Total discrepancies (%)	0.1

Note.—NC = not categorized.

value of less than .05 was considered to indicate a significant difference.

Results

There were 4294 major discrepancies from a total preliminary study volume of 2 922 377, resulting in an aggregate major discrepancy rate of 0.15%. Major discrepancies occurred across multiple modalities and study types; 79.9% of examinations were CT, and 86.6% of major discrepancies occurred during CT interpretations (Table 3).

Shift Length

Diagnostic radiologists worked 86 754 shifts, consisting of regularly scheduled and extra shifts. Regular shifts varied and had a mean (± standard deviation) shift length of 7.98 hours ± 2.63 (range, 0.016–16.0 hours), and extra shifts had a mean length of 1.85 hours ± 2.57 (range, 0.016–14.92 hours). Shifts in which major discrepancies occurred had a mean length of 8.97 hours ± 2.28 (range, 0.13–15.42 hours). Mean shift length significantly differed ($P < .0001$) for shifts with major discrepancies versus those without major discrepancies. On average, discrepancies occurred 8.97 hours ± 2.28 into the shift (median, 10 hours; interquartile range, 2.0 hours; minimum, 0.1 hours; maximum, 15.4 hours). When we converted the data into time categories (eg, ≤1 hour, 1.1–2 hours), significantly more errors occurred late in shifts than occurred early (χ^2 = 6313.42; $P < .0001$) (Fig 1).

Study Volume per Shift

Mean study volume for all shifts throughout the year was 67.60 ± 60.24, with mean studies per hour of 10.95 ± 6.81. Mean study volume for shifts in which major discrepancies occurred was 118.96 ± 66.89 examinations, with the mean number of studies per hour of 13.06 ± 6.10. There was a significant difference (t = 50.36; $P < .0001$) in the mean volume of studies read for those with and those without major discrepancies. 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 (t = 642.27; $P < .0001$). When we examined only shifts with major discrepancies, there was a significant difference in the number of errors as a function of study volume (F = 20.99, $P < .001$), with volumes significantly lower for one error versus two, three versus four, and two versus three (Fig 2).

Consecutive Days Worked

The mean number of prior consecutive days in which a radiology shift was worked was 3.38 days ± 3.23 (range, 0–14 days). The number of errors did not differ as a function of the number of days worked consecutively (F = 2.171; P = .0893) (Fig 3).

Shift Type, Modality, and Urgency

Of the 4294 major discrepancies, 4203 occurred on regularly scheduled shifts or extra shifts. The remaining 91 major discrepancies occurred on backup shifts (Table 4). There was a significant difference in error rate as a function of type of shift (χ^2 = 29.28; $P < .0001$), with regular shifts having more errors than did extra or backup shifts. However, extra shifts were significantly shorter than regular shifts.

When we examined shifts with at least one error, there was no significant difference in errors as a function of modality (χ^2 = 11.01; P = .5283). There was no significant difference in errors as a function of whether the discrepancy was acute or chronic (χ^2 = 3.794; P = .2845). There was no significant difference in errors as a function of critical finding (χ^2 = 0.391; P = .9420). There was a significant difference in errors as a function of urgency description (χ^2 = 22.37; P = .0336), with trauma protocols having fewer single-error shifts and more multierror shifts than the other types of studies (Table 5).

Monthly Analysis

By using the entire set of studies read, the monthly major discrepancy rate as a function of shift type (regular vs extra) was assessed. There were no significant differences in miss rates as a function of shift type (t = 0.101; P =

Figure 1

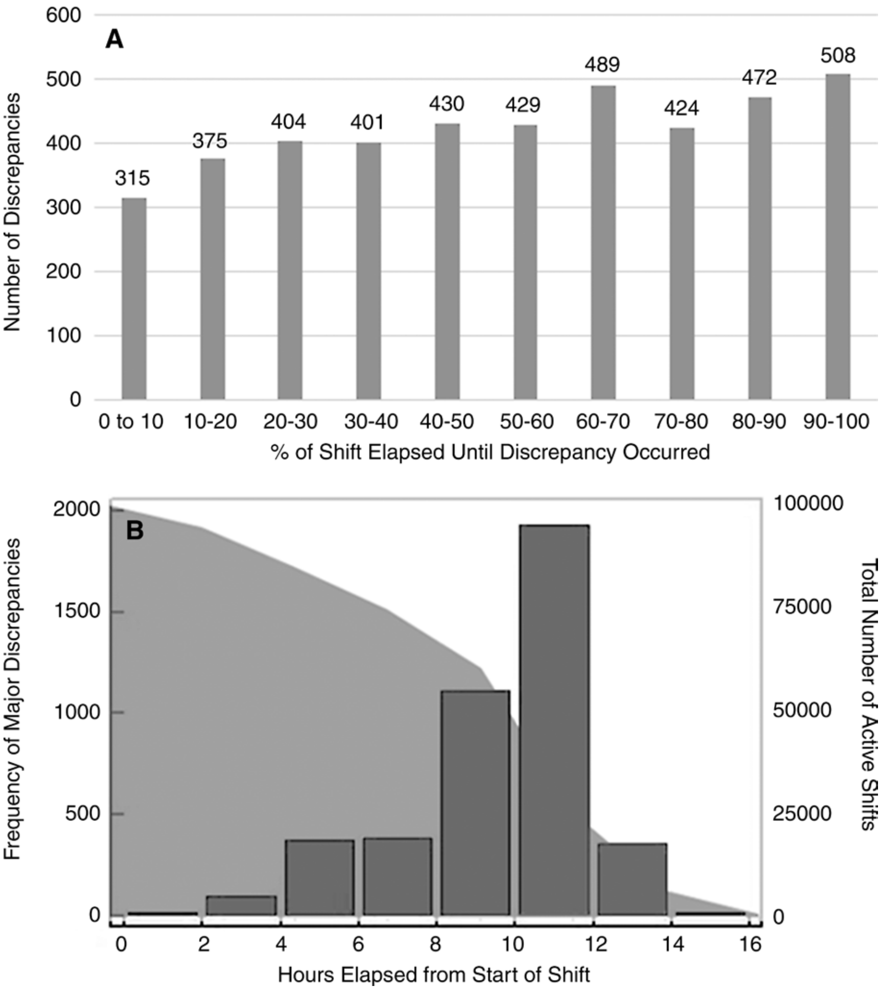

Figure 1: Graphs show analysis of the association of shift length and occurrence of major discrepancies. *A*, Bar graph shows that discrepancies trend upward as the percentage of time elapsed since the start of the shift increases. *B*, Graph shows frequency of major discrepancies displayed on the basis of hours elapsed since the start of the shift (bars). Time since the start of the shift is grouped into 2-hour intervals. The total number of active shifts at each time interval is displayed as the gray area behind the bars (right axis).

Figure 2

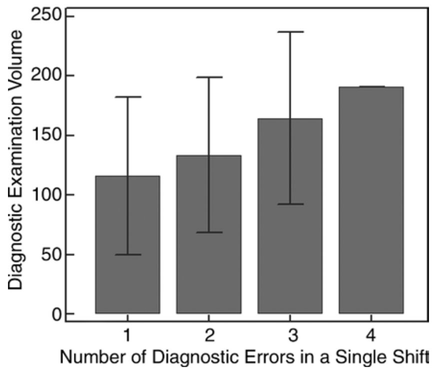

Figure 2: Graph shows association of diagnostic volume and major discrepancies. All radiologist shifts containing at least one diagnostic error were grouped according to error frequency (range, one to four), and diagnostic examination volume for each shift group is displayed. Bars represent mean volume as a function of number of errors, and the lines are the standard deviations.

Figure 3

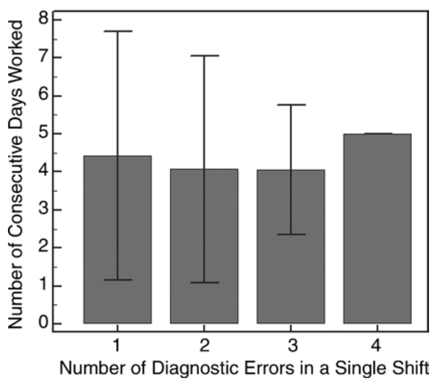

Figure 3: Graph shows association of consecutive days worked and error frequency. All radiologist reading shifts containing at least one diagnostic error were grouped by error frequency (range, one to four); for each group, the number of consecutive days worked prior to current shift is displayed. The lines are the standard deviations.

.921) or month (F = 0.965; P = .520) (Fig 4). Monthly, there was no correlation (r = 0.006) between miss rate and mean number of studies per shift or shift length (r = 0.015).

Discussion

In this retrospective analysis of 2.92 million examinations, we demonstrate that major diagnostic radiologic interpretive discrepancies are associated with shift length and higher diagnostic examination volumes and are more likely to occur later in a shift. More consecutive days worked was not statistically significantly associated with major discrepancy rates. In a 2014 meta-analysis that included 388 124 adult CT examinations, Wu et al (11) reported an aggregate major discrepancy rate for staff radiologists of 2.4%; however, only CT examinations were analyzed, and the studies involved were heterogeneous with respect to population, practice setting versus simulated environment, resident involvement, and major discrepancy definition. Discrepancy rates across the 58 prior studies pooled by Wu et al (11) ranged from 0.3% to 57.8%. The 2009 RADPEER white paper reported a RADPEER level 3 rate of 0.32% and level 4 rate of 0.07% (5). Although different language is used for categorizing major discrepancies in our study, our aggregate discrepancy rate of 0.15% is similar to prior RADPEER data. It is important to note that only

Table 4

Distribution of Error Occurrence and Frequency by Shift Type

Shift Type	Zero-Error Shifts	One-Error Shifts	Two-Error Shifts	Three-Error Shifts	Four-Error Shifts	Total Errors	Total Shifts with Errors
Regular	52 001 (93.8)	3174 (5.7)	268 (0.5)	16 (0.03)	1 (0.002)	3762	3459/55 460 (6.2)
Extra	30 868 (98.6)	412 (1.3)	13 (0.04)	1 (0.003)	0 (0)	441	426/31 294 (1.4)
Backup	NA*	83	4	0	0	91	87/4802 (1.8)*
Total	82 869	3669	285	17	1	4294	. . .

Note.—Data in parentheses are percentages of shift type. NA = not applicable.

* Total number of backup shifts was 4802. However, backup-shift radiologists interpret examinations only if needed on the basis of practice volume. Because these are not interpretive shifts, they were excluded from the total shift numbers.

Table 5

Number of Errors Occurring per Shift as a Function of Study Type

No. of Errors	Emergent	Nonemergent	Highly Emergent	Stroke Protocol	Trauma Protocol
1	86 (2957/3439)	100 (2/2)	84 (554/660)	83 (65/78)	74 (85/115)
2	12 (413/3439)	0 (0/2)	15 (99/660)	17 (13/78)	25 (29/115)
3	1.5 (52/3439)	0 (0/2)	1 (7/660)	0 (0/78)	1 (1/115)
4	0.5 (17/3439)	0 (0/2)	0 (0/660)	0 (0/78)	0 (0/115)

Note.—Values are expressed as percentage, with numerators and denominators in parentheses. There was a significant difference, with trauma protocols having fewer single error shifts than other study types.

a subset of radiologic errors cause a change in patient treatment (14,15).

The effect of study volume on diagnostic radiology interpretive errors by residents and attending physicians has been studied across multiple radiology subspecialties (7,13,16). Although the volume of studies read per hour or shift varies markedly across different practice settings, volumes per radiologist have increased substantially over those from prior decades (17,18). Fitzgerald (7) proposed that greater workload increased the likelihood of diagnostic errors by radiologists. The results of that study showed increased error rates in the interpretation of abnormal CT findings when the radiologist read more than 20 studies per day. Conversely, a separate study demonstrated no statistical significance in the relationship of increased neuroradiology study volume and major discrepancies between attending physicians and residents (13). Sokolovskaya et al (16) studied whether the speed at which a radiologist reads affects major errors by attending radiologists reading at two different speeds and found that four of the five radiologists made more errors (26.6%) at the faster reading speed than at the slower speed (10%). Interestingly, in mammography, an increasing volume of total annual examinations interpreted (up to 2000–3000 studies) improved radiologist performance (19).

Our data show that major diagnostic discrepancies are more likely to occur in longer shifts and disproportionately occur later in a shift, peaking between 10 and 12 hours. This result is substantiated by prior literature both within and outside of radiology (20–22). In a single institutional study by Ruutiainen et al (21), major discrepancy rates for radiology residents increased significantly in the final 2 hours of a 12-hour overnight call shift (odds ratio, 1.94; confidence interval: 1.18, 3.21). However, the authors of this study exclusively studied overnight shifts and may have been influenced by the physiologic effects of fatigue and sleep deprivation. Krupinski et al (23) have shown that after a clinical workday, radiologists demonstrate oculomotor strain, reduced ability to focus, and decreased diagnostic accuracy. In nonradiologic literature, authors of a systematic review detected statistically significant declines in error rates with shorter shifts (20). Notably, much literature involves trainees, and experience may affect the ability to deal with fatigue-related errors.

In our study, there was a trend toward more errors after more consecutive days worked, and regular shifts had more major discrepancies than did extra or backup shifts (which were shorter and may have occurred during a more rested state). Authors of prior studies have examined the effects of shift and schedule on error rates of nurses, interns, and residents. Compared with day shifts, errors and adverse incidents increased: 15% during evening shifts, 28% during night shifts, 17% with the third consecutive night shift, and 36% with the fourth consecutive night shift (24). Increased cumulative weekly hours have been associated with increased rate of injury to workers (25). There was increased risk for needle-stick injury, illness, and work missed among nurses with mandatory overtime (26). Nursing errors increase substantially among nurses working more than 40 hours per week with shift lengths of greater than 12 hours (27) or 12.5 hours (28). A systematic literature review in 2009 uncovered 12 studies showing clinical (mortality, morbidity, complications, medical error, readmissions, and resource use) and 15 showing laboratory (cognitive and fine motor skills, use of surgical simulation devices) fatigue-related errors that adversely affected patient safety (29). Numerous studies have shown that medical trainees working

Figure 4

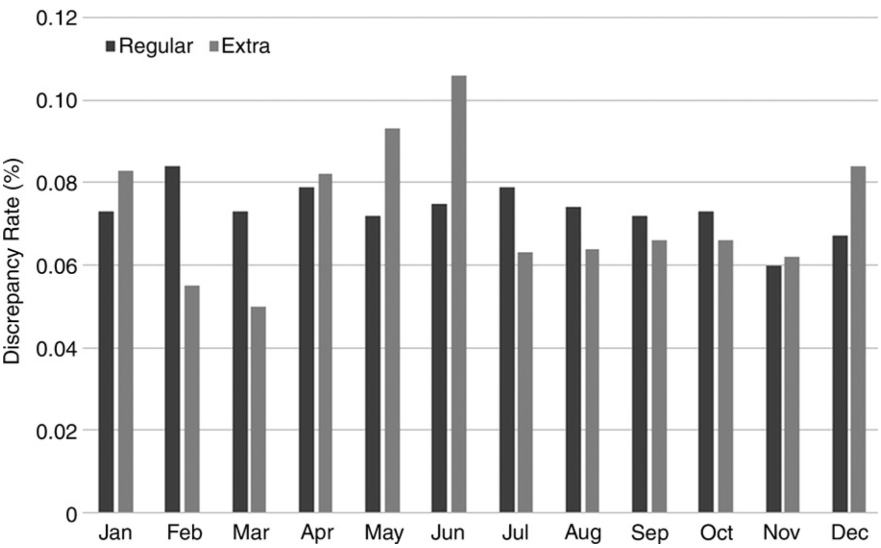

Figure 4: Graph shows major discrepancy rate by month displayed by shift type.

fewer hours per week and shorter shifts make fewer serious errors (30,31), and this can reduce patient mortality (32). Conversely, in the Flexibility in Duty Hours Regulations for Surgical Trainees trial, no harm to patients was demonstrated with removal of restrictions on shift length and time off between shifts, but there was increased dissatisfaction regarding time off (33).

The practice of diagnostic radiology is nearly entirely computer-based, leading to an enormous volume of data that can be harnessed for analysis and error minimization. The mining of big data in health care in general and radiology specifically has been slow but is very promising as a quality assurance tool (34,35). These data show the utility of aggregate tracking and mining of discrepancies at a practice level; similar reports could be generated for individual radiologists who may have unique situations causing interpretive errors.

Our study had limitations. These data arise from a teleradiology practice, and the following differences between teleradiology and other practice types could influence the results: *(a)* There is likely a greater representation of overnight interpretations in this study compared with that in conventional practices; *(b)* the overall modality mix in these data may be different, in particular regarding a lower radiography fraction and lower rates of MR imaging; *(c)* the practice environment, with fewer interruptions by phone calls, consultations with peers and referring clinicians, and fewer procedures, resulting in more concentrated periods of reading time, may increase interpreted volumes; and *(d)* a greater fraction of acute studies may alter the pathology distribution over that of other practices. These factors may affect the aggregate major discrepancy rate; however, we believe that it would have a less significant effect on the distribution of errors across shifts and schedules. Nonetheless, this should be considered in interpreting these results.

The second radiologist reader had probably already viewed the initial interpretation in our study (ie, they were not blinded to the preliminary interpretation), which can lower discrepancy rates (11). This may result from alliterative cognitive bias, essentially the tendency to repeat prior interpretations because of the (possibly unconscious) influence of knowing that a board-certified radiologist has already rendered a diagnosis (9). However, there may be heterogeneity in how individual radiologists practice, with some not looking at the initial interpretation until after viewing the study.

Additional discrepancies discovered after September 1, 2016, were not included; theoretically, major discrepancies of chronic findings (eg, neoplasms) may be recognized later and could increase the cumulative discrepancy rate. However, on the basis of 2015 data (including both minor and major discrepancies), reported discrepancies after 9 months are very rare (0.1%); as such, the authors believe that major discrepancies reported after this date are unlikely to alter the findings of our study.

Finally, up to 0.054% (1590 of 2 922 377) of preliminary interpretations in our study were subsequently given final interpretations by the same radiologist. Statistically, the lack of separate radiologist readers for these examinations may have resulted in up to three major discrepancies being excluded from our analysis, but this should not have affected the results.

In conclusion, the results of this study advance our understanding of how radiologist shift length, diagnostic volume, and consecutive days worked influence major discrepancy rates. In an ongoing effort to minimize radiologic errors, such information may be useful in evidence-based optimization of shifts and schedules.

Disclosures of Conflicts of Interest: **T.N.H.** disclosed no relevant relationships. **C.L.** disclosed no relevant relationships. **E.A.K.** disclosed no relevant relationships. **S.W.** disclosed no relevant relationships. **J.O.J.** disclosed no relevant relationships.

References

1. Bender LC, Linnau KF, Meier EN, Anzai Y, Gunn ML. Interrater agreement in the evaluation of discrepant imaging findings with the Radpeer system. AJR Am J Roentgenol 2012;199(6):1320–1327.
2. Mahgerefteh S, Kruskal JB, Yam CS, Blachar A, Sosna J. Peer review in diagnostic radiology: current state and a vision for the future. RadioGraphics 2009;29(5):1221–1231.
3. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ 2016;353:i2139.

4. Kohn LT, Corrigan J, Donaldson MS. To err is human: building a safer health system. Washington, DC: National Academies Press, 2000.
5. Jackson VP, Cushing T, Abujudeh HH, et al. RADPEER scoring white paper. J Am Coll Radiol 2009;6(1):21–25.
6. Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. RadioGraphics 2015;35(6):1668–1676.
7. Fitzgerald R. Error in radiology. Clin Radiol 2001;56(12):938–946.
8. FitzGerald R. Radiological error: analysis, standard setting, targeted instruction and teamworking. Eur Radiol 2005;15(8):1760–1767.
9. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive error in radiology. AJR Am J Roentgenol 2017;208(4):739–749.
10. Soffa DJ, Lewis RS, Sunshine JH, Bhargavan M. Disagreement in interpretation: a method for the development of benchmarks for quality assurance in imaging. J Am Coll Radiol 2004;1(3):212–217.
11. Wu MZ, McInnes MD, Macdonald DB, Kielar AZ, Duigenan S. CT in adults: systematic review and meta-analysis of interpretation discrepancy rates. Radiology 2014;270(3):717–735.
12. Hanna TN, Loehfelm T, Khosa F, Rohatgi S, Johnson JO. Overnight shift work: factors contributing to diagnostic discrepancies. Emerg Radiol 2016;23(1):41–47.
13. Verdoorn JT, Hunt CH, Luetmer MT, et al. Increasing neuroradiology exam volumes on-call do not result in increased major discrepancies in primary reads performed by residents. Open Neuroimaging J 2015;8:11–15.
14. Carney E, Kempf J, DeCarvalho V, Yudd A, Nosher J. Preliminary interpretations of after-hours CT and sonography by radiology residents versus final interpretations by body imaging radiologists at a level 1 trauma center. AJR Am J Roentgenol 2003;181(2):367–373.
15. Yoon LS, Haims AH, Brink JA, Rabinovici R, Forman HP. Evaluation of an emergency radiology quality assurance program at a level I trauma center: abdominal and pelvic CT studies. Radiology 2002;224(1):42–46.
16. 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.
17. Sunshine JH, Burkhardt JH. Radiology groups' workload in relative value units and factors affecting it. Radiology 2000;214(3):815–822.
18. 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.
19. Coldman AJ, Major D, Doyle GP, et al. Organized breast screening programs in Canada: effect of radiologist reading volumes on outcomes. Radiology 2006;238(3):809–815.
20. Reed DA, Fletcher KE, Arora VM. Systematic review: association of shift length, protected sleep time, and night float with patient care, residents' health, and education. Ann Intern Med 2010;153(12):829–842.
21. 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.
22. Mann FA, Danz PL. The night stalker effect: quality improvements with a dedicated night-call rotation. Invest Radiol 1993;28(1):92–96.
23. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. J Am Coll Radiol 2010;7(9):698–704.
24. 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.
25. Dembe AE, Erickson JB, Delbos RG, Banks SM. The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. Occup Environ Med 2005;62(9):588–597.
26. Gershon RR, Pearson JM, Sherman MF, Samar SM, Canton AN, Stone PW. The prevalence and risk factors for percutaneous injuries in registered nurses in the home health care sector. Am J Infect Control 2009;37(7):525–533.
27. 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.
28. Scott LD, Rogers AE, Hwang WT, Zhang Y. Effects of critical care nurses' work hours on vigilance and patients' safety. Am J Crit Care 2006;15(1):30–37.
29. Caruso JW, Veloski J, Grasberger M, et al. Appendix C: Systematic review of the literature on the impact of variation in residents' duty hour schedules on patient safety. In: Philibert I, Amis S Jr, eds. The ACGME 2011 Duty Hour Standards: Promoting Quality of Care, Supervision, and Resident Professional Development: Enhancing Quality of Care, Supervision, and Resident Professional Development. Chicago, Ill: ACGME, 2011; 101–106.
30. Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 2004;351(18):1838–1848.
31. Rothschild JM, Landrigan CP, Cronin JW, et al. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. Crit Care Med 2005;33(8):1694–1700.
32. Bhavsar J, Montgomery D, Li J, et al. Impact of duty hours restrictions on quality of care and clinical outcomes. Am J Med 2007;120(11):968–974.
33. Bilimoria KY, Chung JW, Hedges LV, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. N Engl J Med 2016;374(8):713–727.
34. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. BioMed Res Int 2015;2015:370194.
35. Murphy DR, Meyer AN, Bhise V, et al. Computerized triggers of big data to detect delays in follow-up of chest imaging results. Chest 2016;150(3):613–620.