

Do Long Radiology Workdays Affect Nodule Detection in Dynamic CT Interpretation?

Elizabeth A. Krupinski, PhD^a, Kevin S. Berbaum, PhD^b, Robert T. Caldwell, MFA^b,
Kevin M. Scharz, PhD^b, Mark T. Madsen, PhD^b, David J. Kramer, BS^b

Purpose: A previous study demonstrated decreased diagnostic accuracy for finding fractures and decreased ability to focus on skeletal radiographs after a long working day. Skeletal radiographic examinations commonly have images that are displayed statically. The aim of this study was to investigate whether diagnostic accuracy for detecting pulmonary nodules on CT of the chest displayed dynamically would be similarly affected by fatigue.

Methods: Twenty-two radiologists and 22 residents were given 2 tests searching CT chest sequences for a solitary pulmonary nodule before and after a day of clinical reading. To measure search time, 10 lung CT sequences, each containing 20 consecutive sections and a single nodule, were inspected using free search and navigation. To measure diagnostic accuracy, 100 CT sequences, each with 20 sections and half with nodules, were displayed at preset scrolling speed and duration. Accuracy was measured using receiver operating characteristic curve analysis. Visual strain was measured via dark vergence, an indicator of the ability to keep the eyes focused on the display.

Results: Diagnostic accuracy was reduced after a day of clinical reading ($P = .0246$), but search time was not affected ($P > .05$). After a day of reading, dark vergence was significantly larger and more variable ($P = .0098$), reflecting higher levels of visual strain, and subjective ratings of fatigue were also higher.

Conclusions: After their usual workday, radiologists experience increased fatigue and decreased diagnostic accuracy for detecting pulmonary nodules on CT. Effects of fatigue may be mitigated by active interaction with the display.

Key Words: Reader fatigue, observer performance, dark vergence, dynamic CT

J Am Coll Radiol 2012;9:191-198. Copyright © 2012 American College of Radiology

INTRODUCTION

Today's advanced imaging modalities are acquiring more and more images that must be interpreted in less and less time [1-8]. Concerns have been raised that radiologists' workloads are becoming so demanding that fatigue and reduced time for interpretation are negatively affecting diagnostic accuracy [9-14]. In court, a plaintiff's attorney has argued that a radiologist missed a breast lesion because he was overworked [9].

Although radiologist fatigue has been a concern for years, only recently have dedicated studies been conducted. Some early studies did not examine fatigue or viewing times directly. For example, Oestmann et al [12]

demonstrated that detection accuracy for lung nodules decreased as viewing time decreased, but fatigue was not examined. Bechtold et al [11] found that error rates in the interpretation of abdominal CT more than doubled when radiologists read out >20 studies in a day. This retrospective review and classification of errors in clinical cases was not a controlled examination of fatigue.

More recently, studies have examined reader accuracy at different times during the day, with mixed results. Taylor-Phillips et al [15] examined data from the UK Breast Screening Programme for nearly 200,000 cases in an attempt to relate accuracy to time of day and reading time. They found that recall rates varied with time of day but not in the same way for the individual readers. Some readers had lower recall rates in the afternoon, while others did not. Recall rates tended to decline with increased reading time (ie, recall rates were lower around lunch and the end of the day), but again it varied considerably among readers. The sample was too noisy to document anything significant beyond a possible trend. This study did not directly examine fatigue or conduct a con-

^aDepartment of Radiology, University of Arizona, Tucson, Arizona.

^bDepartment of Radiology, University of Iowa, Iowa City, Iowa.

Corresponding author and reprints: Elizabeth A. Krupinski, PhD, University of Arizona, Department of Radiology, 1609 N Warren, Building 211, Room 112, Tucson, AZ 85724; e-mail: krupinski@radiology.arizona.edu.

This work was supported in part by grant R01 EB004987 from the National Institute of Biomedical Imaging and Bioengineering (Bethesda, Md).

trolled study in which readers read a dedicated set of cases before and after a day of clinical reading.

Al-s'adi et al [16] also found that breast lesion detection varied with time of day but that no particular time of day had a significant effect. Radiologists at a national meeting were recruited to read a set of mammograms during 1 of 4 reading times (7 to 10 AM, 10 AM to 1 PM, 1 to 4 PM, and 4 to 8 PM). There were no statistically significant differences in sensitivity, specificity, or area under the receiver operating characteristic (ROC) curve (AUC) as a function of time of day. Limitations of this study included that readers participate in only a single session and that they could choose the time of their participation, possibly choosing times of higher performance or motivation.

The impact of fatigue was directly studied by Krupinski et al [17] using skeletal radiographs with fractures as the detection task. Forty radiologists and residents interpreted a set of 60 patient examinations before and after a day in the reading room. Resting state of accommodation (also known as dark focus) was measured as an indicator of visual workload on oculomotor equilibrium. Subjective measures of physical and visual strain and fatigue were collected. The results indicated that diagnostic accuracy was reduced significantly from before to after the day of clinical reading ($P < .05$), and the radiologists and residents became more myopic. Subjective ratings indicated increased lack of energy, physical discomfort, sleepiness, physical exertion, lack of motivation, and eye-strain. In general, the residents exhibited greater effects of fatigue on all measures compared with the attending radiologists. The conclusion was that after a day of clinical reading, radiologists have reduced ability to focus and a reduced ability to detect fractures, as well as increased symptoms of fatigue and visual strain.

The results of this study probably generalize well to most radiographic modalities. However, there are usually few radiographs per patient, and the images are static. Tomographic modalities such as CT, MRI, and digital breast tomosynthesis are viewed in fundamentally different way than are radiographs.

The sequences of tomographic sections are typically viewed in cine animation mode, with successive sections presented one after another under the radiologist's control. The difference between static and dynamic displays places different demands on the visual system. A very basic, yet critical, distinction in the human visual system is between channels processing static stimuli and channels processing moving or changing stimuli [18-20]. Briefly, the transient visual channel, with high temporal resolution but poor spatial resolution, serves as an "early warning system" for the sustained visual channel, which has poor temporal resolution and high spatial resolution. Things that move or change attract attention and eye movements. That is why people wave when they want to attract attention and why warning signals flash off and on. It is why things that move seem blurry and things that

do not move seem to be sharp. These characteristics reflect the sensitivities of the two parts of the visual system handling perception of these stimuli. As a radiologist cycles dynamically through a sequence of CT sections, the sudden onset and offset of a pulmonary nodule captures the viewer's attention and directs it to the location of change [20]. With dynamic images, the motion channel of visual processing which directly affects attention comes into play, and the task of guiding the eyes around the changing image in search of a lesion becomes more complex. Thus, the impact of fatigue may differ for dynamic and static image interpretation.

The goal of the present study was to measure diagnostic accuracy for pulmonary nodule detection in dynamic CT chest sequences before and after a day (or night) of diagnostic image interpretation. We also investigated a measure of visual strain, the resting static of convergence, often referred to as dark vergence.

METHODS

This study was approved by the institutional review boards at both the University of Arizona and the University of Iowa.

Images

All images were stripped of patient identifiers to comply with HIPAA standards. We used 110 chest CT examinations selected from existing databases [21-24], approximately half ($n = 60$) with a solitary pulmonary nodule and half ($n = 50$) nodule free. Approximately half of the nodules were moderately subtle and the other half subtle as determined in the previous studies. To standardize the viewing conditions for all observers, we restricted each case to 20-slice sequences. For the nodule cases, the slices (3 mm) were selected such that the nodule did not appear in the 2 end slices. This ensured that the entire nodule would be visible without getting cut off at the boundaries. Standard lung window and level settings were used, and observers were not allowed to adjust settings during testing. Additional examinations were used in a demonstration before testing to familiarize observers with the task, reporting procedure, and presentation software.

Observers

Observers were attending radiologists and radiology residents at the University of Arizona and the University of Iowa, with 11 attending radiologists and 11 radiology residents at each institution. Table 1 provides the genders, average ages, percentage wearing corrective lenses, and types of lenses worn for the observers at both institutions. Table 2 indicates how long on average they had been reading cases before the test sessions.

The participants were also asked to indicate whether they had a preferred order in which they viewed CT chest image areas (bone, mediastinum, lung) and in what manner they preferred to view them (cine first, right then left, etc). Preferences are shown in Figures 1 and 2.

Table 1. Characteristics of participating University of Arizona and University of Iowa attending radiologists and residents

	Attending Radiologists		Residents	
	Arizona	Iowa	Arizona	Iowa
Gender				
Male	8	9	8	8
Female	3	2	3	3
Age (y)				
Men	42.63 ± 13.12 (33-71)	48.78 ± 14.26 (33-74)	30.75 ± 2.05 (28-33)	30.25 ± 1.61 (28-32)
Women	42.33 ± 10.75 (30-54)	45.50 ± 1.73 (44-47)	28.67 ± 0.52 (28-29)	30.67 ± 2.58 (29-34)
Wear corrective lenses	63.64%	90.91%	72.73%	72.73%
Type of lenses				
Glasses/contact lenses full-time	85.71%	100%	100%	100%
Reading glasses	14.29%			

Note: Data are expressed as numbers, mean ± SD (range), or percentages.

Procedure

Cases were displayed using customized Workstation J software developed at the University of Iowa [25]. Data were collected at two points in time for each observer: once before any diagnostic reading activity (early) and once after a day of diagnostic reading (late). It should be noted that we use the terms *early* and *late*, rather than *morning* and *afternoon*, because the early sessions for some readers were in the afternoon before starting a night shift, and the late sessions were in the morning after coming off call.

Observers in Arizona completed the Swedish Occupational Fatigue Inventory (SOFI), which was developed and validated to measure perceived fatigue in work environments [26,27]. The instrument consists of 20 expressions distributed on 5 latent factors: lack of energy, physical exertion, physical discomfort, lack of motivation, and sleepiness. Subjects report their ratings for each of the 20 questions using a scale ranging from 0 to 10, where 10 indicates that they are 10 times as fatigued, stressed, unmotivated, and so on, than if they were reporting a score of 1 (ie, interval scale data). An average score for each of the 5 latent variables is derived from the individual questions within the set of 20 that contribute to the latent factors [26,27]. Physical exertion and physical discomfort are considered physical dimensions of fatigue, while lack of motivation and sleepiness are considered primarily mental factors. Lack of energy is a general factor reflecting both physical and mental aspects of fatigue. Lower scores indicate lower levels of perceived fatigue than higher scores. The SOFI does not measure

visual fatigue, so it was complemented with the oculomotor strain subscale from the Simulator Sickness Questionnaire (SSQ) [28,29]. Subjects report their ratings on a set of 7 dimensions (ie, general discomfort, fatigue, headache, eyestrain, difficulty focusing, difficulty concentrating, and blurred vision) using a scale ranging from 1 (none) to 4 (severe) (ordinal scale data).

Visual strain was assessed by measuring dark vergence, the resting state of convergence of the eyes [30-33], measured in the absence of stimuli (including light). There is evidence that prolonged near work affects dark vergence (as it does accommodation), resulting in inducement of temporary myopia. In this study, we measured dark vergence using the Vergamatic II USB (manufactured by Steven Spadafore, Franklin and Marshall College, Lancaster, Pennsylvania). The device measures dark vergence and generates two metrics called V or angle (in degrees) and meter angle (MA). Angle (V) is approximately equal to the angle between the lines from the optical center of the eyes to the point of fixation and the parallel rays that would define the gaze direction if the eyes were fixated at infinity. Meter angle is the linear equivalent of V. Measures were made before and after each reading session.

After an introduction and review of the practice cases, the observers viewed the CT sequences on an NEC MultiSync LCD 2490WUXi color display (maximum luminance, 400 cd/m²; contrast ratio, 800:1; resolution, 1,920 × 1,200; screen size, 24.1 inches) that was calibrated to the Digital Imaging and Communications in Medicine grayscale standard display function [34].

Table 2. Data for attending radiologists and residents for the early and late sessions regarding sleep, case reading, and eye conditions on the days of the study

Variable	Attending Radiologists		Residents	
	Arizona	Iowa	Arizona	Iowa
Hours reading early	0.24 ± 0.33 (0-1)	0 ± 0 (0)	0.18 ± 0.34 (0-1)	0 ± 0 (0)
Hours reading late	7.00 ± 1.07 (5-8.5)	8.05 ± 1.37 (7-10)	9.77 ± 5.65 (5-25)	8.28 ± 0.65 (7.5-10)

Note: Data are expressed as mean ± SD (range).

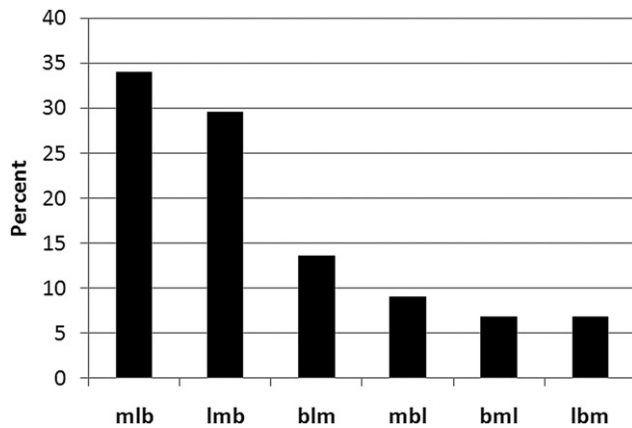


Fig 1. Distribution of responses as to whether readers had a preferred order in which they viewed CT chest image areas (bone, mediastinum, lung).

The test session was divided into 2 parts. In part 1 (free scrolling), the readers were presented with 10 of the cases, each containing a nodule. In this part, they used the mouse to scroll back and forth through the CT sequences at their own pace. Their task was to determine if a nodule was present, locate it with a cursor, and provide a rating of their decision confidence both in adjectival form (definite, probable, possible, suspicious) and subjective probability (10-100 in 10-point intervals) to be used in ROC curve analysis. Total time spent viewing each sequence was recorded.

In part 2 (fixed scrolling), 100 CT sequences were shown to the readers but at a fixed rate and number of passes through each sequence. Each sequence went through 4 passes (sections 1 to 20, 20 to 1, 1 to 20, and 20 to 1) at a rate of 0.18 seconds per slice, for a total of 14.18 seconds total viewing time. After each sequence was displayed, the software guided readers through a series of responses to indicate whether a nodule was present or absent and, if present, to indicate its location (right or left lobe and anterior, central, or posterior portion of the lung). Finally, the software asked readers to indicate their confidence in the decision as a subjective probability (10%-100% in 10% intervals), before prompting them to go to the next sequence. Each session took approximately 1 hour to complete.

Statistical Tests

Diagnostic accuracy was derived from the confidence data and was measured using the AUC. The AUC was estimated for each observer in each experimental condition, and the average areas were compared using an analysis of variance (ANOVA). Between-subject variables were level of training (attending, resident) and institution (Arizona, Iowa), and the within-subject (or repeated-measures) variable was the reading session time of day (early, late). Two ROC methods were used. The first was PROPROC [35-37], which does not take into account

lesion location, and the second was LROC [38,39], which does take location into account. Post hoc *F* tests were used to examine individual variable differences and interactions.

The viewing times were measured in seconds (continuous ratio data) and were analyzed using an ANOVA with early vs late and location (Arizona vs Iowa) as independent variables. The visual strain (dark vergence) measures (continuous ratio data) were also analyzed with an ANOVA with early and late pre-session and post-session recordings as the independent variables. Post hoc *F* tests were used to examine individual variable differences and interactions.

The SOFI uses interval scales for reporting, so an ANOVA was used to analyze these data. The SSQ uses an ordinal scale ranging from 1 to 4 for reporting, and thus Wilcoxon's signed-rank test was used to analyze these data.

RESULTS

Diagnostic Accuracy and Viewing Time

In part 1 (free scrolling), the number of nodules detected (of 10) and the time to indicate the nodule (which ended the trial) were analyzed in 2 separate ANOVAs. There was no significant difference between early and late sessions in the number of nodules detected ($F = 1.42$, $P = .24$). The attending radiologists detected 81% of the nodules on average in the early session and 80% during the late session. The residents detected 79% on average during the early session and 75% during the late session. There was also no significant difference in viewing time per image. The median viewing time for the 10 trials was computed for each reader in each treatment. The averages of the median viewing time were 26.83 seconds in the early session and 26.85 seconds during the late session ($F = 0.00$, $P = .99$).

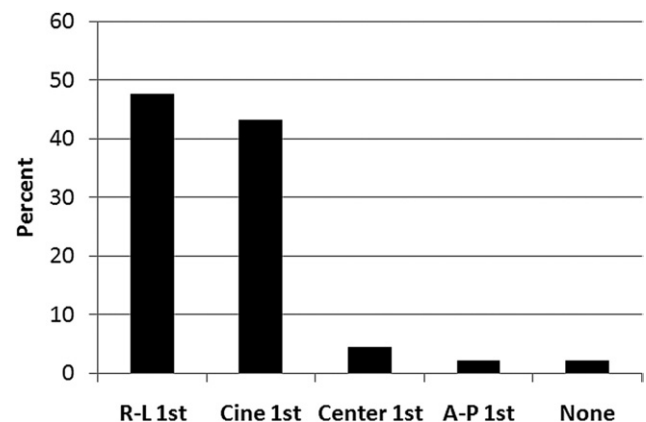


Fig 2. Distribution of responses as to whether and in what manner readers preferred to view CT images (cine first, right then left, etc).

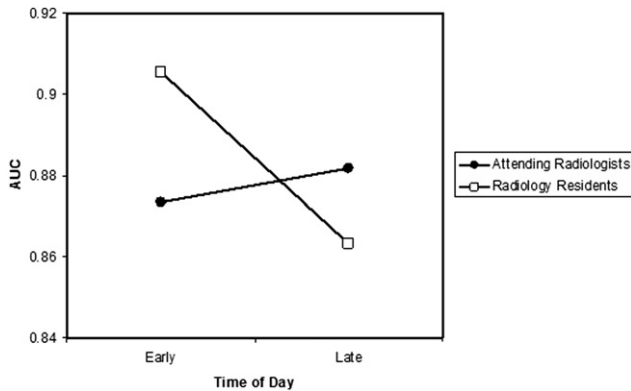


Fig 3. For the analysis of variance and PROPROC area under the curve measures, the significant effect of the interaction between training level and time of day.

In part 2 (fixed search), AUCs were used to measure accuracy for detecting nodules. For the ANOVA and PROPROC AUC measures, the only significant effect was the interaction between training level and time of day ($F = 5.45, P = .0246$). This effect is illustrated in Figure 3.

Follow-up F tests indicated that for the attending radiologists, the change in PROPROC AUC between the early and late sessions (0.873 to 0.882) was not significant ($F = 0.86, P = .37$), and for the residents, change in PROPROC AUC between the early and late sessions (0.906 to 0.863) was marginally significant ($F = 3.93, P = .063$).

For the ANOVA for LROC AUC measures, the only significant effect was the interaction between training level and time of day ($F = 6.40, P = .0154$). This effect is illustrated in Figure 4.

Follow-up F tests indicated that for the attending radiologists, the change in LROC AUC between the early and late sessions (0.706 to 0.755) was marginally significant ($F = 4.13, P = .057$), and for the residents, the change in LROC AUC between the early and late sessions (0.789 to 0.742) was not significant ($F = 2.13, P = .162$).

Visual Strain Results

Both of the dark vergence measures (V and MA) showed increased variability for the late vs early reading sessions (see box plots in Figure 5). The MA metric revealed a statistically significant increase for late compared with early sessions ($F = 6.793, P = .0098$). The V metric also showed an increase for the late session, but it did not reach statistical significance ($F = 1.507, P = .2210$).

Fatigue Survey Results

The scores for each of the 5 SOFI factors (Arizona readers) were analyzed using an ANOVA with session (early vs late) and experience (attending radiologist vs resident) as independent variables. Average rating values for each factor are shown in Table 3. It can be seen for all measures, that ratings were higher (more severe) for the late compared with

the early sessions. For all of the measures, the residents gave higher ratings than the attending radiologists.

For lack of energy ($F = 9.13, P = .0044$) and lack of motivation ($F = 8.23, P = .0066$), the differences were statistically significant. For physical exertion, physical discomfort, and sleepiness, the early-to-late differences were not statistically significant. For the SSQ, the residents again had higher ratings overall than the attending radiologists, and the ratings for the early session were significantly lower for both groups (ie, less severe) than for the late session ($Z = -3.509, P = .0004$).

DISCUSSION

Diagnostic Accuracy

Our study revealed some decreases in diagnostic accuracy as a function of the work of interpreting clinical images. Part 2 used automated scrolling to collect 100 ROC trials in <50 minutes. We had judged that collecting our data in <1 hour per session was necessary to limit adding to the fatigue levels. Both proper ROC and location-specific ROC methods demonstrated a statistically significant training-by-workload interaction, with attending radiologists tending to increase in accuracy with work and residents tending to decrease accuracy with work. Attending radiologists either improved after working (LROC) or stayed the same (ROC). Residents either decreased in accuracy (ROC) or stayed the same (LROC). These significant interactions mirror our findings with fracture detection used to measure the effects of fatigue [17]. Long reading days do affect observer performance for the interpretation of dynamic CT sequences, much as they do with static image interpretation.

An interesting question is why the proper ROC analysis and the LROC analysis gave differing versions of the statistical interaction between training and fatigue: LROC showed increasing attending performance, while proper ROC showed decreasing resident performance. Of course, it should be noted that the direction of non-

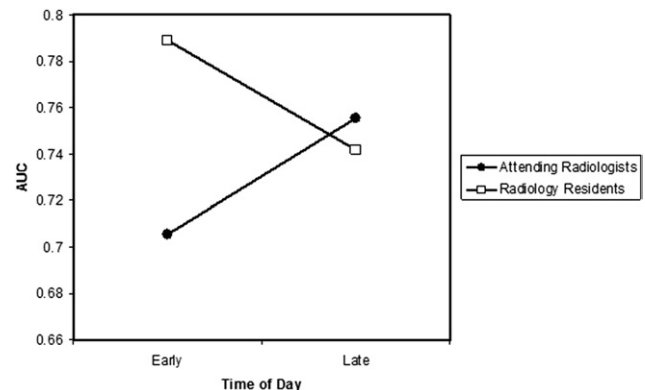


Fig 4. For the analysis of variance for LROC area under the curve measures, the significant effect of the interaction between training level and time of day.

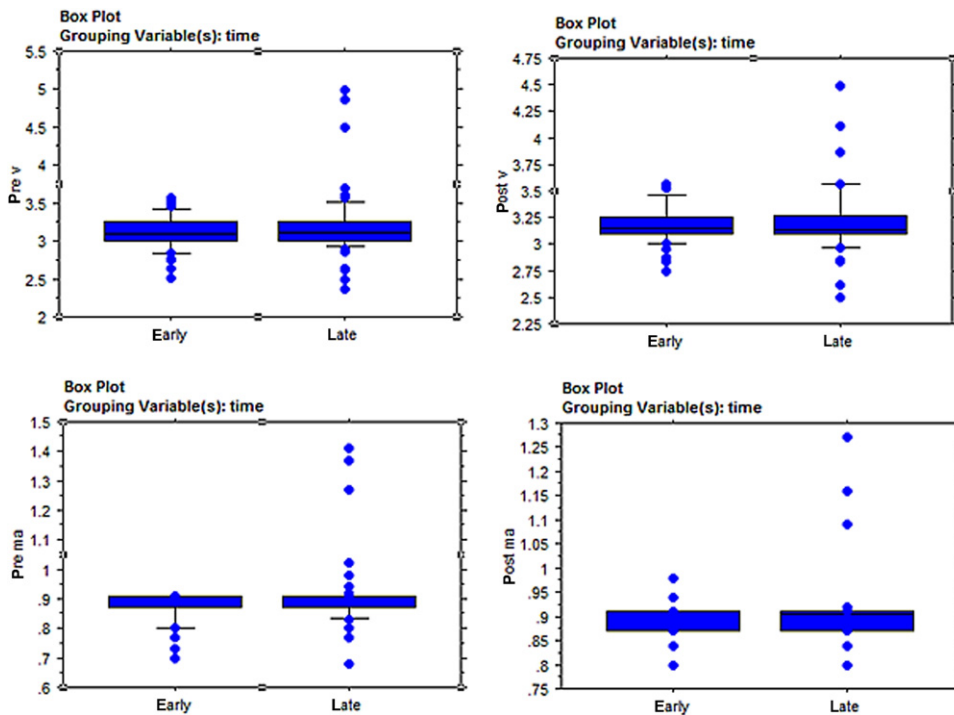


Fig 5. Box plots of the dark vergence measures (angle [V] and meter angle [MA]) showing increased variability for the late versus early reading sessions. $V (^{\circ}) = \{ATAN[(-CM + IPD)/vergence\ distance]\} \times 57.295$. $MA = \text{inverse of vergence distance (m)} = V\{0.01 \times [vergence\ distance + (vergence\ distance + CM)/(IPD - CM)]\}$.

significant effects was consistent with the significant effects. The difference between ROC scoring and LROC scoring is that the former may give credit for a false-positive response identifying a nonexistent nodule combined with a false-negative response failing to identify a real nodule. This (LROC) should provide a more accurate scoring of responses.

Part 1 of our experiment used free scrolling to focus on fatigue effects on visual search time, closely resembling actual clinical reading. Neither response time nor hit rate detecting nodules depended on interpretive work. Part 1, with only 10 target nodules and no trials without nodules, was not designed to measure diagnostic accuracy. The instructions were designed to encourage observers to search until they were confi-

dent that they had located a pulmonary nodule. The purpose was to determine where visual search became less efficient. It did not. Perhaps active interaction with the workstation provides a measure of physical activity sufficient to ward off the effects of fatigue.

Visual Strain and Reading Time

Dark vergence was a fairly effective measure of visual strain or fatigue, at least using the MA metric. After a long day or night of clinical reading, there was much more variability in both the V and MA metrics, and for MA, the values increased significantly. The results are supportive of those observed with the accommodation measure used in the bone fracture study [17,40]: readers were essentially more myopic after each reading session

Table 3. SOFI and SSQ ratings for University of Arizona radiologists early and late in the day

Variable	Attending Radiologists		Residents	
	Early	Late	Early	Late
SOFI				
Lack of energy	1.07 ± 1.42	3.23 ± 2.50	2.05 ± 1.96	4.41 ± 3.54
Physical discomfort	0.80 ± 1.21	0.84 ± 1.25	0.98 ± 0.95	2.02 ± 2.28
Sleepiness	0.96 ± 1.58	1.98 ± 2.31	1.64 ± 1.89	3.32 ± 3.22
Physical exertion	0.21 ± 0.42	0.27 ± 0.75	0.34 ± 0.49	0.86 ± 1.65
Lack of motivation	0.80 ± 1.07	1.98 ± 1.96	1.46 ± 1.31	3.48 ± 2.65
SSQ Eye Strain	1.14 (0.25)	1.43 (0.68)	1.15 (0.50)	1.71 (1.11)

Note: Data are expressed as mean ± SD or as median (interquartile range). SOFI = Swedish Occupational Fatigue Inventory; SSQ = Simulator Sickness Questionnaire.

compared with before as well as more myopic overall late compared with early.

As noted earlier, induced myopia is a common finding in observers engaged in prolonged near-vision work, which is exactly what radiologists are engaged in as they sit in front of computer displays for hours on end interpreting image cases. We have now established in two separate studies with two separate measures of the set point of accommodation and convergence that radiologists experience induced myopia after a long day or night of reading. However, we cannot yet establish a causal relationship between visual changes and reduced diagnostic accuracy. Evidence from other studies is mixed. For example, Safdar et al [41] tested the visual acuity of 23 radiologists between 7:50 and 10:30 AM, between 12 and 3:30 PM, and after 3:30 PM on several workdays. They found no significant differences in acuity as a function of time of day. We would expect decreased acuity on the basis of decrements in ability to keep the eyes focused on the display screen. However, Safdar et al did not report on exactly how much clinical reading the observers had been engaged in before each measurement.

Unno et al [42] compared visual acuity, convergence, and pupil diameter of younger and older subjects before and after reading 2-D vs 3-D (stereoscopic) radiographs. They observed some possible trends in each of these measures, with the 3-D reading affecting the measures more, but no statistically significant differences were observed. The limitations of this study were that they did not use radiologists as observers, and the study was focused on fatigue associated with a relatively brief reading of stereo pairs.

Subjective Ratings of Fatigue

The SOFI and SSQ ratings are very similar to those observed in the fracture study [17]. Both the attending radiologists and the residents subjectively felt more fatigued after a day or night of clinical reading. In both studies, the residents had higher ratings on all of the measures compared with the attending radiologists. It is interesting to note that although the attending radiologists felt fatigued and experienced induced myopia, as evidenced by the dark vergence measurements, they did not have an associated decrease in diagnostic accuracy. In the fracture study, they did exhibit a decrease in performance, but as in this study, the residents were clearly more affected by fatigue than the more experienced attending radiologists. In the present study, it was the residents' drop in diagnostic accuracy that contributed more to the statistical significance than the attending radiologists'.

Further study is needed to determine why this difference between residents and attending radiologists exists, but two possibilities come to mind immediately. The first is that the residents are still in a learning phase during their routine workdays, and although they clearly do not read as many images as attending radiologists, the learning process itself is quite fatiguing and stressful, thus affecting them more at the end of the day. The second

possibility is that the attending radiologists are quite fatigued as well at the end of their shifts but through experience have learned to compensate for their fatigue better perhaps by being more careful during reading and pacing themselves better than the residents.

Limitations

There were limitations associated with this study. Although we did include a free search condition (part 1), the main study (part 2) was restricted to 20 contiguous sections that were scrolled through automatically by the computer for a set amount of time. This is quite unlike clinical reading but was necessary for this study, as we wanted all readers to complete the study within about 1 hour and to have read the same number of cases. Although this could have made the task less fatiguing than in true clinical reading, we still observed a statistically significant drop in diagnostic accuracy after a long day of clinical reading. If we had actually replicated clinical reading with free search of 100 cases in part 1 (and eliminated part 2) that included all of the slices, it seems likely that we would have observed an even greater decrement in performance. A future study is warranted to follow up on this possibility.

A second limitation is that the readers knew that the study was about fatigue. However, intuitively, one would think that knowing the study was about fatigue would have led to readers trying to compensate for or overcome their fatigue in the late session, just to "prove" that their performance was not affected by fatigue. However, the results indicate otherwise. Even if they were trying to combat their fatigue and maintain accuracy, at least for the residents and some of the attending radiologists, this did not happen; accuracy was degraded after a long day of clinical reading. They did not rise to the occasion.

It is interesting that the attending radiologists were overall less affected by fatigue than the residents, in that their diagnostic accuracy in the main test (part 2) was not affected greatly late in the day. There are two possible contributing factors. The first was noted above: the automatic scrolling and set viewing time may somehow lessen the impact of fatigue, perhaps by reducing the need to interact with the computer, decide how fast to scroll, when to stop, and so on. Less cognitive and physical energy was needed compared with traditional "active" reading, so more attentional and cognitive resources could be devoted to the detection task. This is one avenue for potential future investigation. The second possibility is that the attending radiologists are simply much more experienced than the residents and over many years of clinical reading have developed ways to compensate for fatigue.

CONCLUSIONS

After a day or night of clinical reading, radiologists have increased symptoms of fatigue and increased oculomotor strain, as evidenced by more variability in dark vergence.

Residents have reduced detection accuracy for lesion targets in dynamic CT sequences, although paradoxically, attending radiologists do not. These results parallel those for accuracy in detecting fractures in static bone images. Radiologists need to be aware of the effects of fatigue on diagnostic accuracy and take steps to mitigate these effects.

ACKNOWLEDGMENTS

We would like to thank the 44 attending radiologists and radiology residents in the departments of radiology at the University of Arizona and the University of Iowa for their time and participation in this study.

REFERENCES

- Bhargavan M, Sunshine JH. Utilization of radiology services in the United States: levels and trends in modalities, regions, and populations. *Radiology* 2005;234:824-32.
- DiPiro PJ, vanSonnenberg E, Tumeh SS, Ros PR. Volume and impact of second-opinion consultations by radiologists at a tertiary care cancer center: data. *Acad Radiol* 2002;9:1430-3.
- Ebbert TL, Meghea C, Iturbe S, Forman HP, Bhargavan M, Sunshine JH. The state of teleradiology in 2003 and changes since 1999. *AJR Am J Roentgenol* 2007;188:W103-12.
- Sunshine JH, Maynard CD. Update on the diagnostic radiology employment market: findings through 2007-2008. *J Am Coll Radiol* 2008;5:827-33.
- Lu Y, Zhao S, Chu PW, Arenson RL. An update survey of academic radiologists' clinical productivity. *J Am Coll Radiol* 2008;5:817-26.
- Nakajima Y, Yamada K, Imamura K, Kobayashi K. Radiologist supply and workload: international comparison—Working Group of Japanese College of Radiology. *Radiat Med* 2008;26:455-65.
- Mukerji N, Wallace D, Mitra D. Audit of the change in the on-call practices in neuroradiology and factors affecting it. *BMC Med Imag* 2006;6:13.
- Meghea C, Sunshine JH. Determinants of radiologists' desired workloads. *J Am Coll Radiol* 2007;4:143-4.
- Berlin L. Liability of interpreting too many radiographs. *AJR Am J Roentgenol* 2000;175:17-22.
- Fitzgerald R. Error in radiology. *Clin Radiol* 2001;56:938-46.
- Bechtold RE, Chen MYM, Ott DJ, et al. Interpretation of abdominal CT: analysis of errors and their causes. *J Comp Assist Tomogr* 1997;21:681-5.
- Oestmann JW, Greene R, Kushner DC, Bourgooin PM, Linetsky L, Llewellyn HJ. Lung lesions: correlation between viewing time and detection. *Radiol* 1988;166:451-3.
- Royal College of Radiologists. Workload and manpower in clinical radiology. BFCR(99)5. London: Royal College of Radiologists.
- European Society of Radiology. Risk management in radiology in Europe IV. Vienna, Austria: ESR/EAR Office; 2004.
- Taylor-Phillips S, Clarke A, Wallis M, Wheaton M, Duncan A, Gale AG. The time course of cancer detection performance. *Proc SPIE Med Imag* 2011;7966:796605-1-8.
- Al-s'adi M, McEntee MF, Ryan E. Time of day does not affect radiologists' accuracy in breast lesion detection. *Proc SPIE Med Imag* 2011;7966:796608-1-7.
- Krupinski EA, Berbaum KS, Caldwell RT, Scharz KM, Kim J. Long radiology workdays reduce detection accommodation accuracy. *J Am Coll Radiol* 2010;7:698-704.
- Kulikowski JJ, Tolhurst DJ. Psychophysical evidence for sustained and transient detectors in human vision. *J Physiol* 1973;232:149-62.
- Breitmeyer BG, Ganz L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol Rev* 1976;83:1-3.
- Yantis S, Jonides J. Abrupt visual onsets and selective attention: evidence from visual search. *J Exp Psychol Hum Percept Perform* 1984;10:601-21.
- Scharz KM, Berbaum KS, Madsen M, et al. Multiple diagnostic task performance in computed tomography examination of the chest. In preparation.
- Krupinski EA, Siddiqui K, Siegel E, et al. Influence of 8-bit vs 11-bit digital displays on observer performance and visual search: a multi-center evaluation. *J Soc Inform Display* 2007;15:385-90.
- Madsen MT, Berbaum KS, Caldwell RT. A new software tool for removing, storing and adding abnormalities to medical images for perception research studies. *Acad Radiol* 2006;13:305-12.
- Madsen MT, Berbaum KS, Scharz K, Caldwell RT. Improved implementation of the abnormality manipulation software tools. *Proc SPIE Med Imag* 2011;7966:796612-1-7.
- Scharz KM, Berbaum KS, Caldwell RT, Madsen MT. Workstation J. workstation emulation software for medical image perception and technology evaluation research. *Proc SPIE Med Imag* 2007;6515:651511-1-11.
- Ahsberg E. Dimensions of fatigue in different workplace populations. *Scand J Psychol* 2000;41:231-41.
- Ahsberg E, Gamberale F, Gustafsson K. Perceived fatigue after mental work: an experimental evaluation of a fatigue inventory. *Ergonomics* 2000;43:252-68.
- Kennedy RS, Lane NE, Lilienthal MG, Berbaum KS, Hettinger LJ. Profile analysis of simulator sickness symptoms: application to virtual environment systems. *Presence* 1992;1:295-301.
- Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. *Intl J Aviation Psych* 1993;3:203-20.
- Rosenfield M. Tonic vergence and vergence adaptation. *Optom Vis Sci* 1997;74:303-28.
- Jaschinski W, Jainta S, Hoormann J, Walper N. Objective and subjective measurements of dark vergence. *Ophthalm Physiol Opt* 2007;27:85-92.
- Tyrell RA, Leibowitz HW, Herschel W. The relation of vergence effort to reports of visual fatigue following prolonged near work. *Hum Factors* 1990;32:341-57.
- Owens DA, Wolf-Kelly K. Near work, visual fatigue, and variations in oculomotor tonus. *Invest Ophthalmol Vis Sci* 1987;28:743-9.
- Digital Imaging and Communications in Medicine. Available at: <http://medical.nema.org>. Accessed August 3, 2011.
- Metz CE, Pan X. "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *J Math Psychol* 1999;43:1-33.
- Pan X, Metz CE. The "proper" binormal model: parametric ROC curve estimation with degenerate data. *Acad Radiol* 1997;4:380-9.
- Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves. *Acad Radiol* 2007;14:814-29.
- Swenson RG. Measuring detection and localization performance. In: Barrett HH, Gmitro AF, eds. *Proceedings IMPI '93*. London: Springer-Verlag; 1993:525-41.
- Swenson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys* 1996;23:1709-25.
- Krupinski EA, Berbaum KS. Measurement of visual strain in radiologists. *Acad Radiol* 2009;16:947-50.
- Safdar N, Mai J, Siddiqui K, Janjua R, Siegel E. Evaluation of the visual acuity of radiologists. Presented at: Annual meeting of the Radiological Society of North America; 2005.
- Unno YY, Tajima T, Kuwabara T, et al. Analysis of physiological impact while reading stereoscopic radiographs. *Proc SPIE Med Imag* 2011;7966:79660C-1-15.