Published in final edited form as:
J Am Coll Radiol. 2012 March ; 9(3): 191–198. doi:10.1016/j.jacr.2011.11.013.

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

Elizabeth A. Krupinski, PhD[1], **Kevin S. Berbaum, PhD**[2], **Robert T. Caldwell, MFA**[2], **Kevin M. Schartz, PhD**[2], **Mark T. Madsen, PhD**[2], and **David J. Kramer**[2]

[1]Department of Radiology, The University of Arizona

[2]Department of Radiology, The University of Iowa

Abstract

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. This study investigated whether diagnostic accuracy for detecting pulmonary nodules in computed tomography (CT) of the chest displayed dynamically would be similarly affected by fatigue.

Methods—Twenty-two radiologists and 22 residents were given two tests searching CT chest sequences for a solitary pulmonary nodule before and after a day of clinical reading. To measure search time, ten lung CT sequences, each containing 20 consecutive sections and a single nodule, were inspected using free search and navigation. To measure diagnostic accuracy, one hundred CT sequences, each with 20 sections and half with nodules, were displayed at preset scrolling speed and duration. Accuracy was measured using ROC 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 ($p = 0.0246$), but search time was not affected ($p > 0.05$). After a day of reading, dark vergence was significantly larger and more variable ($p = 0.0098$), 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.

Keywords

reader fatigue; observer performance; dark vergence; dynamic CT

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 radiologist workloads

Corresponding author: Elizabeth A. Krupinski, PhD, Department of Radiology, University of Arizona, 1609 N. Warren, Bldg 211, Rm 112, Tucson, AZ 85724, 520-626-4498 (ph), 520-626-4376 (fax), krupinski@radiology.arizona.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

are becoming so demanding that fatigue and reduced time for interpretation are negatively impacting 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 [11] found that error rates in the interpretation of abdominal CT more than doubled when radiologists read out more than 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 (i.e., 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 controlled 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 varies 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 one of 4 reading times (7:00 – 10:00; 10:00 – 13:00; 13:00 – 16:00; 16:00 – 20:00). 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. Limitation of this study include readers only participating in a single session and that they could choose the time of their participation, possibly choosing a time 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 (a.k.a. dark focus) was measured as an indicator of visual workload on oculomotor equilibrium. Subjective measures of physical and visual strain and/or fatigue were collected. The results indicated that diagnostic accuracy was reduced significantly from before to after the day of clinical reading ($p < 0.05$) and the radiologists and residents became more myopic. Subjective ratings indicating increased lack of energy, physical discomfort, sleepiness, physical exertion, lack of motivation, and eyestrain. In general, the residents exhibited greater effects of fatigue on all measures compared to 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, and 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, magnetic resonance imaging (MRI) and digital breast tomosynthesis are viewed in fundamentally different way than radiographs.

The sequences of tomographic sections are typically viewed in ciné-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 characteristic reflect the sensitivities of the two parts of the visual system handling perception of these stimuli. As the 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 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 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 Health Insurance Portability and Accountability Act standards. We used 110 chest CT examinations selected from existing databases [21-24], approximately half (60) with a solitary pulmonary nodule and half (50) nodule free. Approximately half of the nodules were moderately subtle and the other half subtle as determined in the previous studies. In order 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 two end slices. This insured that the entire nodule would be visible without getting cut off at the boundaries. Standard lung window/level setting was used and observers were not allowed to adjust settings during testing. Additional examinations were used in a demonstration prior to 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 (AZ) and the University of Iowa (IA), with 11 attending radiologists and 11 radiology residents at each institution. Table 1 provides the gender, average age, percent wearing corrective lenses, and type of lenses worn for the observers at both institutions. Table 2 indicates how long on average they had been reading cases prior to 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 (e.g., cine first, right then left etc.). The preferences are shown in Figures 1 and 2.

Procedure

Cases were displayed using customized WorkstationJ software developed at the (copyright 2011, the University of Iowa) [25]. Data were collected at two points in time for each observer: once prior to 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 since the Early session for some readers was in the afternoon before starting a night shift and the Late session was in the morning after coming off call.

Observers at one site (AZ) 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 five 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 0 – 10 point scale where 10 indicates that they are 10 times as fatigued/stressed/unmotivated etc. than if they were reporting a 1 (i.e., interval scale data). An average score for each of the five 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. SOFI does not measure visual fatigue so it was complemented with the oculomotor strain sub-scale from the Simulator Sickness Questionnaire (SSQ) [28-29]. Subjects report their ratings on a set of seven dimensions (i.e., general discomfort, fatigue, headache, eyestrain, difficulty focusing, difficulty concentrating, blurred vision) using a 1 – 4 point scale ranging from none to 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 impacts 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, PA). The device measures dark vergence and generates two metrics called V or angle (deg) and meter-angle. 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 a NEC MultiSync LCD 2490WUXi color display (maximum luminance 400 cd/m^2; contrast ratio 800:1; resolution 1920 × 1200; screen size 24.1”) that was calibrated to the DICOM (Digital Imaging and Communications in Medicine) Grayscale Standard Display Function (GSDF) [34].

The test session was divided into two parts. In Part I (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 a ROC analysis. Total time spent viewing each sequence was recorded.

In Part II (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 sec/slice for a total of 14.18 sec total viewing time. After each sequence was displayed, the software guided them 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 it asked them 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 one hour to complete.

Statistical Tests

Diagnostic accuracy was derived from the confidence data and was measured using area under the Receiver Operating Characteristic (ROC) curve (AUC). 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), institution (Arizona, Iowa), and a 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) so were analyzed using an ANOVA with Early vs Late and location (AZ vs IA) as independent variables. The visual strain (dark vergence) measures (continuous ratio data) were also analyzed with an ANOVA with Early and Late pre and post-session recordings as the independent variables. Post-hoc F-tests were used to examine individual variable differences and interactions.

The SOFI survey uses interval scales for reporting so an ANOVA was used to analyze these data. The SSQ survey uses a 1-4 ordinal scale for reporting and thus a Wilcoxon Signed Rank test was used to analyze these data.

Results

Diagnostic Accuracy & Viewing Time

In Part I (free scrolling), number of nodules detected (of 10) and time to indicate the nodule (which ended the trial) were analyzed in two separate ANOVAs. There was no significant difference between Early and Late in the number of nodules detected (F = 1.42, p = 0.24). 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. Average of the median viewing time in the Early session was 26.83 sec and 26.85 sec during the Late session (F = 0.00, p = 0.99).

In Part II (fixed search) area under the ROC curve (AUC) was used to measure accuracy for detecting nodules. For the ANOVA and PROPROC AUC measures, the only significant effect was the training level by time-of-day interaction (F = 5.45, p = 0.0246). This effect is illustrated in Figure 3.

Follow-up F-tests indicate that for the attending radiologists the Early to Late change in PROPROC AUC (0.873 to 0.882) was not significant (F = 0.86, p = 0.37) and for the residents the Early to Late change in PROPROC AUC (0.906 to 0.863) was marginally significant (F = 3.93, p = 0.063).

For the ANOVA for LROC AUC measures, the only significant effect was the training level by time-of-day interaction (F = 6.40, p = 0.0154). This effect is illustrated in Figure 4.

Follow-up F-tests indicate that for the attending radiologists the Early to Late change in LROC AUC (0.706 to 0.755) was marginally significant (F = 4.13, p = 0.057) and for the residents the Early to Late change in PROPROC AUC (0.789 to 0.742) was not significant (F = 2.13, p = 0.162).

Visual Strain Results

Both of the dark vergence measures (V and MA) showed increased variability for the Late versus Early reading sessions (box plots Figure 5). The MA metric revealed a statistically significant increase for Late compared to Early sessions (F = 6.793, p = 0.0098). The V metric also showed an increase for the Late session, but it did not reach statistical significance (F = 1.507, p = 0.2210).

Fatigue Survey Results

The scores for each of the five SOFI factors (AZ readers) were analyzed with an ANOVA with session (Early vs. Late) and experience (Attending vs. Resident) as independent variables. Average rating values for each factor are shown in Table 3. It can be seen for all measures ratings were higher (more severe) for the Late compared to the Early sessions. For all of the measures the residents gave higher ratings than the attending radiologists.

For Lack of Energy ($F = 9.13$, $p = 0.0044$) and Lack of Motivation ($F = 8.23$, $p = 0.0066$) the differences were statistically significant. For Physical Exertion, Physical Discomfort and Sleepiness, the Early to Late differences were not statistically significant. For the SSQ survey the residents again had higher ratings overall than the attending and the ratings for Early were significantly lower for both groups (i.e., less severe) than for Late reporting (Z = -3.509, p = 0.0004).

Discussion

Diagnostic Accuracy

Our study revealed some decreases in diagnostic accuracy as a function of the work of interpreting clinical images. Part II used automated scrolling to collect 100 ROC trials in under 50 minutes. We had judged that collecting our data in under an hour per session was necessary in order 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 radiologist 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 impact observer performance for 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 version 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-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 I 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 I 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 confident 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 & Reading Time

Dark vergence was a fairly effective measure of visual strain or fatigue at least using the MA metric. After a long day/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 compared to before as well as more myopic overall Late compared to 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/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 visual acuity of 23 radiologists between 7:50 – 10:30, 12:00 – 15:30 and after 15:30 on several workdays. They found no significant differences in acuity as a function of time of day. We would expect decreased acuity based on decrements in ability to keep the eyes focused on the display screen. Safdar et al. did not however report on exactly how much clinical reading the observers had been engaged in prior to each measurement.

Unno et al. [42] compared visual acuity, convergence, and pupil diameter of younger and older subjects before and after reading 2D vs 3D (stereoscopic) radiographs. They observed some possible trends in each of these measures with the 3D reading impacting 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/night of clinical reading. In both studies, the residents had higher ratings on all of the measures compared to the attending radiologists. It is interesting to note that even though 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 impacted 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 an attending radiologist the learning process itself is quite fatiguing and stressful, thus impacting 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 are limitations associated with this study. Although we did include a free search condition (Part I), the main study (Part II) 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 Phase I (and eliminated phase II) 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” their performance was not affected by fatigue. The results however indicate otherwise. Even if they were trying to combat their fatigue and maintain accuracy, at least for the residents and some of the attending 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 were overall less impacted by fatigue than the residents in that their diagnostic accuracy in the main test (Phase II) was not impacted 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 etc. Less cognitive and physical energy was needed compared to 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 are simply much more experienced than the residents and over many years of clinical reading have developed ways to compensate for fatigue.

Summary

After a day/night of clinical reading, radiologists have increased symptoms of fatigue, and increased oculomotor strain as evidence 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.

This work was supported in part by grant R01 EB004987 from the National Institute of Biomedical Imaging and Bioengineering

References

1. Bhargavan M, Sunshine JH. Utilization of radiology services in the United States: levels and trends in modalities, regions, and populations. Radiology. 2005; 234:824–832. [PubMed: 15681686]
2. 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; 9:1430–1433. [PubMed: 12553355]
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:W103–112. [PubMed: 17242214]
4. Sunshine JH, Maynard CD. Update on the diagnostic radiology employment market: findings through 2007-2008. J Am Coll Radiol. 2008; 5:827–833. [PubMed: 18585660]
5. Lu Y, Zhao S, Chu PW, Arenson RL. An update survey of academic radiologists’ clinical productivity. J Am Coll Radiol. 2008; 5:817–826. [PubMed: 18585659]
6. Nakajima Y, Yamada K, Imamura K, Kobayashi K. Radiologist supply and workload: international comparison – Working Group of Japanese College of Radiology. Radiat Med. 2008; 26:455–465. [PubMed: 18975046]
7. 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.
8. Meghea C, Sunshine JH. Determinants of radiologists’ desired workloads. J Am Coll Radiol. 2007; 4:143–144. [PubMed: 17412250]
9. Berlin L. Liability of interpreting too many radiographs. Am J Roent. 2000; 175:17–22.
10. Fitzgerald R. Error in radiology. Clin Radiol. 2001; 56:938–946. [PubMed: 11795921]
11. Bechtold RE, Chen MYM, Ott DJ, Zagoria RJ, Scharling ES, Wolfman NT, Vining DJ. Interpretation of abdominal CT: analysis of errors and their causes. J Comp Assist Tomogr. 1997; 21:681–685.
12. Oestmann JW, Greene R, Kushner DC, Bourgouin PM, Linetsky L, Llewellyn HJ. Lung lesions: correlation between viewing time and detection. Radiol. 1988; 166:451–453.
13. The Royal College of Radiologists. Workload and Manpower in Clinical Radiology. BFCR. 1999; 99(5)
14. European Society of Radiology. Risk management in radiology in Europe IV. ESR/EAR Office Vienna; Austria: 2004.
15. 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.
16. 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.
17. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long radiology workdays reduce detection accommodation accuracy. J Am Coll Radiol. 2010; 7:698–704. [PubMed: 20816631]
18. Kulikowski JJ, Tolhurst DJ. Psychophysical evidence for sustained and transient detectors in human vision. J Physiol. 1973; 232:149–162. [PubMed: 4733479]
19. 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. [PubMed: 766038]
20. Yantis S, Jonides J. Abrupt visual onsets and selective attention: evidence from visual search. J Exptl Psych: Hum Percep Perf. 1984; 10:601–621.

21. Schartz KM, Berbaum KS, Madsen M, Thompson BH, Mullan BF, Caldwell RT, Hammett B, Ellingson AN, Franken EA Jr. Multiple diagnostic task performance in computed tomography examination of the chest. In preparation.

22. Krupinski EA, Siddiqui K, Siegel E, Shrestha R, Grant E, Roehrig H, Fan J. Influence of 8-bit vs 11-bit digital displays on observer performance and visual search: a multi-center evaluation. J Soc Inform Display. 2007; 15:385–390.

23. 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; 13:305–312. [PubMed: 16488842]

24. Madsen MT, Berbaum KS, Schartz K, Caldwell RT. Improved implementation of the abnormality manipulation software tools. Proc SPIE Med Imag. 2011; 7966:796612-1–7.

25. Schartz 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.

26. Ahsberg E. Dimensions of fatigue in different workplace populations. Scandinavian J Psych. 2000; 41:231–241.

27. Ahsberg E, Gamberale F, Gustafsson K. Perceived fatigue after mental work: an experimental evaluation of a fatigue inventory. Ergonomics. 2000; 43:252–268. [PubMed: 10675062]

28. 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.

29. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. Intl J Aviation Psych. 1993; 3:203–220.

30. Rosenfield M. Tonic vergence and vergence adaptation. Optom Vis Sci. 1997; 74:303–328. [PubMed: 9219290]

31. Jaschinski W, Jainta S, Hoormann J, Walper N. Objective and subjective measurements of dark vergence. Ophthal Physiol Opt. 2007; 27:85–92.

32. Tyrell RA, Leibowitz HW, Herschel W. The relation of vergence effort to reports of visual fatigue following prolonged near work. Hum Factors. 1990; 32:341–357. [PubMed: 2258180]

33. Owens DA, Wolf-Kelly K. Near work, visual fatigue, and variations in oculomotor tonus. Invest Ophthalmol Vis Sci. 1987; 28:743–749. [PubMed: 3557879]

34. [August 3, 2011] DICOM (Digital Imaging and Communications in Medicine). http://medical.nema.org/

35. Metz CE, Pan X. “Proper” binormal ROC curves: theory and maximum-likelihood estimation. J Math Psych. 1999; 43:1–33.

36. Pan X, Metz CE. The “proper” binormal model: parametric ROC curve estimation with degenerate data. Acad Radiol. 1997; 4:380–389. [PubMed: 9156236]

37. Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of “proper” binormal ROC curves. Academic Radiology. 2007; 14:814–829. [PubMed: 17574132]

38. Swensson, RG. Measuring detection and localization performance. In: Barrett, HH.; Gmitro, AF., editors. Proc IMPI ’93. London: Springer-Verlag, London; 1993. p. 525-541.

39. Swensson RG. Unified measurement of observer perrformance in detecting and localizing target objects on images. Med Phys. 1996; 23:1709–1725. [PubMed: 8946368]

40. Krupinski EA, Berbaum KS. Measurement of visual strain in radiologists. Acad Radiol. 2009; 16:947–950. [PubMed: 19406673]

41. Safdar N, Mai J, Siddiqui K, Janjua R, Siegel E. Evaluation of the visual acuity of radiologists. Paper presented at the 2005 Radiological Society of North America Meeting.

42. Unno YY, Tajima T, Kuwabara T, Hasegawa A, Natsui N, Ishikawa K, Hatada T. Analysis of physiological impact while reading stereoscopic radiographs. Proc SPIE Med Imag. 2011; 7966:79660C-1–15.

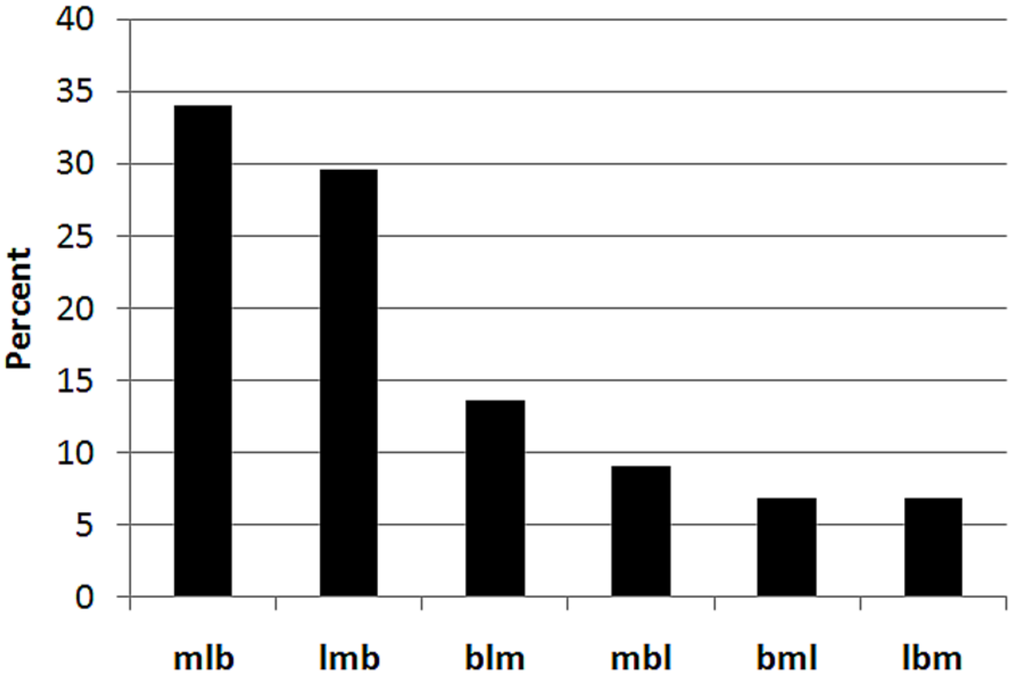

Figure 1.

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

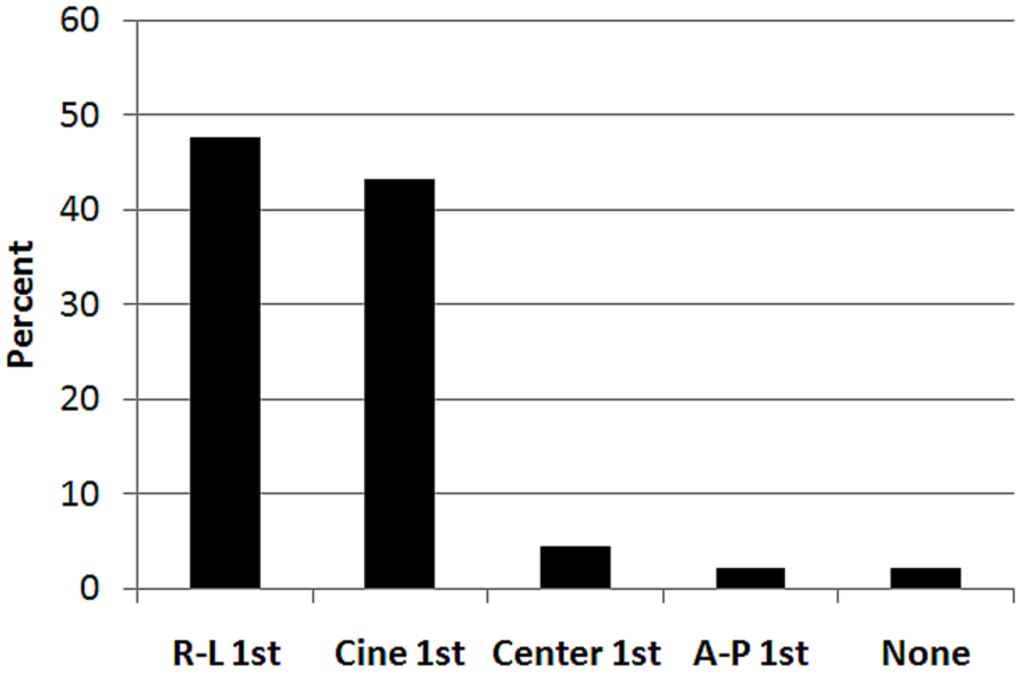

Figure 2.

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

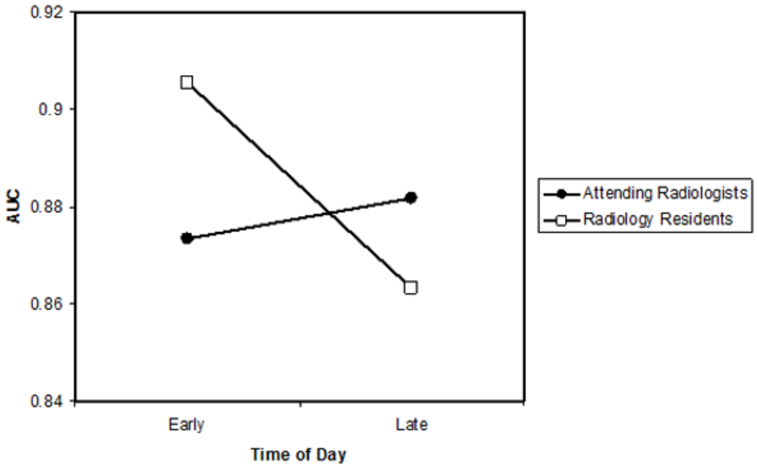

Figure 3.
For the ANOVA and PROPROC AUC measures, figure showing the significant effect of training level by time-of-day interaction.

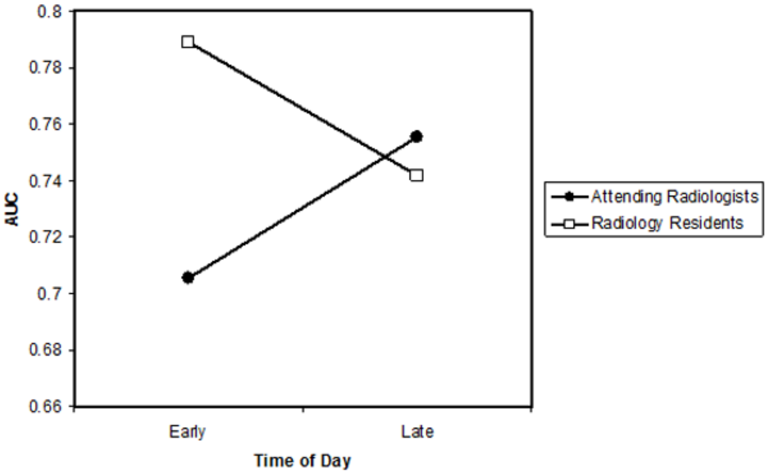

Figure 4.

For the ANOVA for LROC AUC measures, figure showing the significant effect of training level by time-of-day interaction.

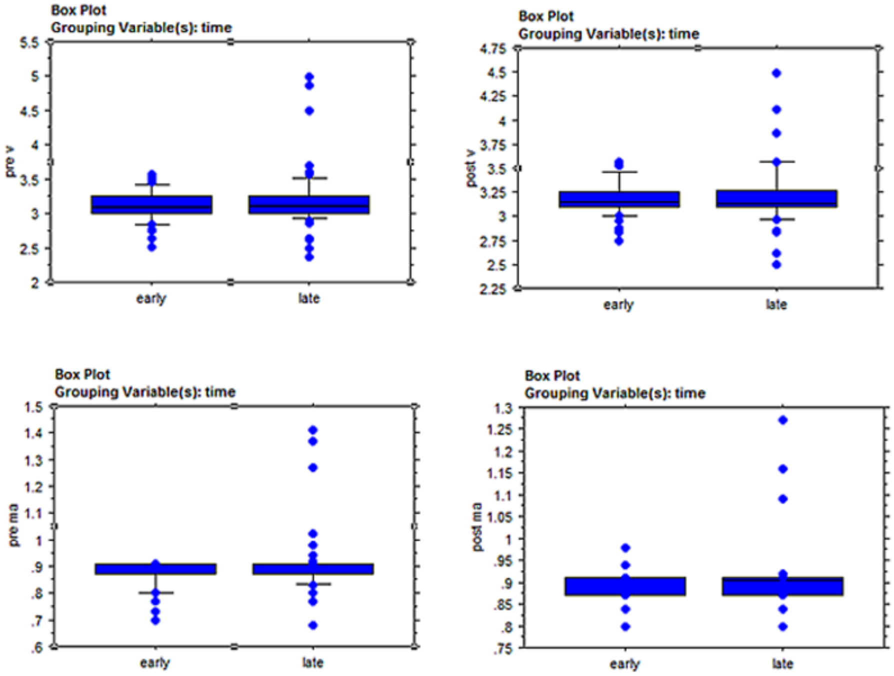

Figure 5.

Box plots of the dark vergence measures (V and MA) showing increased variability for the Late versus Early reading sessions.

V = angle (deg) = (ATAN((-CM+IPD)/vergence distance))*57.295.

MA = meter angle = inverse of vergence distance (m) = V(0.01*(vergence distance+ (vergence distance + CM)/(IPD-CM)))

Table 1

Characteristics of participating Arizona (AZ) and Iowa (IA) Attendings and Residents.

	AZ Attendings	IA Attendings	AZ Residents	IA Residents
Gender	8 male, 3 female	9 male, 2 female	8 male, 3 female	8 male, 3 female
Average male age	42.63 (sd = 13.12; range = 33-71)	48.78 (sd = 14.26, range = 33-74)	30.75 (sd = 2.05, range = 28-33)	30.25 (sd = 1.61, range = 28-32)
Average female age	42.33 (sd = 10.75, range = 30-54)	45.50 (sd = 1.73, range = 44-47)	28.67 (sd = 0.52, range = 28-29)	30.67 (sd = 2.58, range = 29-34)
Wear corrective lenses	63.64%	90.91%	72.73%	72.73%
Type of lenses	85.71% glasses/contacts full-time; 14.29% readers	100% glasses/contacts full-time	100% glasses/contacts full-time	100% glasses/contacts full-time

Table 2

Data for Attendings and Residents for the Early and Late sessions regarding sleep, case reading and eye conditions on the days of the study.

	AZ Attendings	IA Attendings	AZ Residents	IA Residents
Hours reading early	0.24 (sd = 0.33; range = 0 – 1)	0 (sd = 0; range = 0)	0.18 (sd = 0.34; range = 0 – 1)	0 (sd = 0; range = 0)
Hours reading late	7.00 (sd = 1.07; range = 5 - 8.5)	8.05 (sd = 1.37; range = 7 - 10)	9.77 (sd = 5.65; range = 5 - 25)	8.28 (sd = 0.65; range = 7.5 - 10)

Table 3

Mean and standard deviations (in parentheses) of the SOFI and median and IQR for the SSQ survey ratings for AZ Attendings and Residents Early and Late in the day.

	Attendings Early	Attendings Late	Residents Early	Residents Late
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)