



Published in final edited form as:

Acad Radiol. 2017 September ; 24(9): 1058–1063. doi:10.1016/j.acra.2017.03.021.

The Impact of Fatigue on Satisfaction of Search in Chest Radiography

Elizabeth A. Krupinski, PhD¹, Kevin S. Berbaum, PhD², Kevin M. Schartz, PhD², Robert T. Caldwell, MFA², and Mark T. Madsen, PhD²

¹Department of Radiology & Imaging Sciences Emory University; Atlanta, GA

²Department of Radiology University of Iowa; Iowa City, IA

Abstract

Purpose—To assess the nature of the satisfaction of search (SOS) effect in chest radiography when observers are fatigued; determine if we could replicate recent findings that have documented the nature of the SOS effect to be due to a threshold shift rather than a change in diagnostic accuracy as in earlier film-based studies.

Materials and Methods—Nearing or at the end of a clinical workday, 20 radiologists read 64 chest images twice, once with and once without the addition of a simulated pulmonary nodule. Half of the images had different types of “test” abnormalities. Decision thresholds were analyzed using the center of the range of false positive (FP) and true positive (TP) fractions associated with each Receiver Operating Characteristic (ROC) point for reporting test abnormalities. Detection accuracy was assessed with ROC technique and inspection time was recorded.

Results—The SOS effect was confirmed to be a reduction in willingness to respond (threshold shift). The center of the FP range was significantly reduced (FP = 0.10 without added nodules, FP = 0.05 with added nodules, $F(1,18) = 19.85$, $p = 0.0003$). The center of the TP range was significantly reduced (TP = 0.39 without added nodules, TP = 0.33 with added nodules, $F(1,18) = 10.81$, $p = 0.004$).

Conclusion—This study suggests that fatigue does not change the nature of the SOS effect, but rather may be additive with the SOS effect. SOS reduces both true and false positive responses, whereas fatigue reduces true positives more than false positives.

Introduction

In recent years there has been concern over the impact of errors (1, 2) and fatigue on performance in medicine in general (3, 4) and for radiology in particular (5–7). Increased volumes of imaging studies that are more complex in nature have led to increased workloads

Corresponding Author: Elizabeth A. Krupinski, PhD, Department of Radiology & Imaging Sciences, Emory University, 1364 Clifton Rd NE Room D107, Atlanta, GA 30322, 404-712-3868 (ph), 404-712-7387 (fax), ekrupin@emory.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.

None of the authors have any conflict of interests to report

and longer hours interpreting cases (6). Fatigue has both mental and physical consequences, can occur with or without feelings of sleepiness, and can be measured in a variety of subjective and objective ways (see 8 for a recent review). The effects of fatigue can vary as a function of task, level of experience, environmental conditions, and health status. For radiologists, visual fatigue and eye strain may be of particular interest (9, 10) given the visual nature of medical image interpretation. Cognitive fatigue may be important as well but has not been studied systematically in radiology.

Physical measures of visual strain (accommodation and dark vergence) indicate that fatigue can reduce a radiologist's ability to focus, especially at close viewing distances (9, 10). It has also been found that subjective feelings of physical, mental and emotional strain (11, 12) increase in radiologists who have spent as little as 8 hours interpreting clinical images. Both physical and subjective measures indicate that residents may actually be more vulnerable to effects of fatigue than attending radiologists.

Extended workdays have also been associated with reduced diagnostic accuracy performance in a number of clinical reading tasks (10, 12, 13). Studies generally demonstrated a statistically significant average decrease in diagnostic accuracy of approximately 4% for single lesion detection tasks (fractures in bone images, nodules in CR and CT chest). More complex tasks have yet to be investigated.

The presence of multiple abnormalities in radiology examinations yields more reading errors than single abnormalities (see 14 for a historical review of the SOS phenomenon). In chest radiography, classic SOS studies showed a reduction in accuracy (as measured by Receiver Operating Characteristic (ROC) analyses) for the detection of native lesions or abnormalities in the presence of simulated pulmonary nodules (15, 16). A recent studies (17, 18) reexamined the SOS effect in chest radiography with modern imaging techniques (i.e., computed radiography (CR) images viewed on a workstation monitor) instead of film images on a view box. In contrast to the previous film-based studies, they observed that diagnostic accuracy for native lesions was *not* affected by the addition of nodules. Instead, they observed a *threshold shift* (reduction in both true and false positives), reflecting a reluctance to report the additional abnormalities. The reasons for this change in the nature of the SOS effect are not clear, but the authors suggest it may be due to greater utilization of advanced imaging techniques such as CT and MR changing the way radiologists approach the interpretation of radiographic images; or perhaps better training methods have increased awareness of SOS.

The goal of the present study was to determine the nature of the SOS effect in chest radiography after observers have been interpreting cases for approximately 8 hours. Eight hours was the average amount of prior reading time in previous studies of reader fatigue that demonstrated oculomotor and subjective fatigue in readers. Specifically, we wanted to find out whether SOS under conditions of fatigue would lead to an SOS effect on detection accuracy as documented in the original SOS studies (15, 16) or an SOS effect on decision thresholds as found in the more recent ones (17, 18).

Materials & Methods

Cases

The same set of 64 computed radiography (CR) chest cases from the most recent chest SOS study (17) were used. All cases were acquired from clinical studies and anonymized in accordance with local institutional review board approval. As with the previous study, 33 had a native abnormality verified by information from the patient records including follow-up studies, clinical course, surgery, lab results, and autopsy reports; and 31 had no native abnormalities. The native abnormalities included: 3 aneurysms (chest), 1 aortic calcification, 1 asbestosis, 1 cardiomegaly, 2 cervical ribs, 1 clavicle fracture, 1 dilated esophagus, 2 free air hemi-diaphragm, 1 gallstones, 1 gastric air shadow compressed, 2 hiatal hernia, 1 middle lobe collapse (see Figure 1), 1 Morgagni hernia, 1 pneumonia, 2 pneumothorax, 1 renal stone, 2 rib fractures, 2 right-sided aortic arch, 1 scapula fracture, 4 tracheal deviation (neck mass), 1 tuberculosis, and 1 Zenker's diverticulum. In order to test the SOS effect for detection accuracy of the native abnormalities, a solitary simulated pulmonary nodule was added to each case, ensuring that all image characteristics (i.e., patient, native abnormality, window/level setting) were exactly the same for the SOS (with nodule) and non-SOS (no nodule) conditions. Care was taken not to superimpose nodules on the native lesions. The previous study (17) verified the realism of the simulated nodules, level of detectability and placement in the images.

Readers & Image Presentation

An SOS study requires readers to participate in two separate sessions, viewing the same set of images twice - once with and once without the added distractor nodules. Based on our previous fatigue studies measuring changes (~ 4% on average) in diagnostic accuracy, 20 radiologists (10 faculty and 10 residents from the University of Arizona and 10 faculty and 10 residents from the University of Iowa) from the University of Arizona and the University of Iowa served as readers in the study in order to obtain adequate power. All readers reviewed the cases after a workday of interpreting clinical cases (i.e., not on an academic day), comparable to the amount of time readers in our previous studies observing a significant drop in accuracy had spent reading cases before participating in the experiment. The study was approved by the institutional review boards of both universities and all readers signed informed consent documents. Each participant received \$200 per session for their participation.

A counterbalanced design was used, in which each reader saw each case ($n = 64$) twice, once with (SOS) and once without (non-SOS) the added simulated pulmonary nodule (i.e., $n = 128$ total case stimuli). In each viewing session, half of the cases ($n = 16$ with native abnormalities and $n = 16$ without a native abnormality in one session; $n = 17$ with native abnormality and $n = 15$ without a native abnormality in the other session) contained a pulmonary nodule and half did not. In the subsequent session, the cases were seen in the opposite condition. Presentation order was randomized for each session. At least 3 weeks passed between the two viewing sessions to promote forgetting of the images.

The images were displayed on a NEC MultiSync color LCD monitor (2490WUXi; NEC; Tokyo, Japan). The display resolution was 1920×1200 ; maximum luminance 400 cd/m^2 ; contrast ratio 800:1; screen size 24.1 inches; and it was calibrated to the DICOM (Digital Imaging and Communications in Medicine) Grayscale Standard Display Function (GSDF) (20). Viewing software designed by the University of Iowa to simulate clinical viewing conditions and record responses (e.g., time to make a response, total time spent viewing an image, response location, decision confidence, use of window/level) was used for the viewing sessions (21). Room lights were set to 30 lux. Reader distance to the display was not set so readers could choose their own comfortable viewing distance.

Procedure

Each session was preceded by a set of instructions and a practice case to become familiar with the software. The instructions noted that each case could contain no, one or several abnormalities, and that the types of abnormalities were varied. They were not informed about abnormality frequency. The readers' task was to indicate with a mouse cursor, the location of any abnormality detected. If an abnormality was indicated, a pop-up box appeared and readers indicated their decision confidence using a subjective probability scale (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%), and in a text box describe the abnormality. If the reader thought the case contained no abnormality, they proceeded to the next case and the case was scored as definitely normal (0% confidence).

Statistical Analyses

Full details of how the older SOS studies (15, 16) and the more recent (17) SOS chest study were scored can be found in (17). In the current SOS study we were concerned with whether or not the addition of nodules impacts the detection of native lesions. Therefore, true positives and false negatives (TP, FN) were determined based on the reports associated with the native abnormalities. The nodules serve as the "distractors" to establish the SOS context. We are not concerned with the detection of these added nodules, but with their impact on detection of the native abnormalities which is the way that the SOS effect has been traditionally defined. Reader performance measures were not rewarded or penalized for the nodule decisions reported. False positives (FP) were scored only when occurring on the cases without native abnormalities.

Detection Accuracy

As in previous studies, we used the empirical ROC method with detection accuracy measured using true-positive fraction (TP fraction = sensitivity) at the FP fraction of 0.1 ($1 - \text{specificity} = 0.9$) because this index focuses on a part of the ROC curve well-supported with empirical ROC points (22). For generality, we repeated the analysis using the contaminated binormal model (CBM) (23). The multi-reader multi-case (MRMC) ROC methodology developed by Dorfman, Berbaum, and Metz (DBM) (24, 25) has recently been extended in software (OR-DBM MRMC 2.4, available from <http://perception.radiology.uiowa.edu>). Because SOS affects readers rather than patients, generalization to the population of readers is fundamental. Therefore, reader was treated as a random factor and patient and treatment as fixed factors. To accomplish this we selected

DBM Analysis 3 (Obuchowski-Rockette Analysis): Random Readers and Fixed Cases (results apply to the population of readers).

Decision Thresholds

Decision thresholds are defined by parameters of FP rates at different thresholds of certainty to gauge willingness to report an abnormality (at each level of certainty) when no abnormality is present. A decision threshold shift implies that both false positive fractions (FPFs) and true positive fractions (TPFs) change together (though non-linearly), causing the Receiver Operating Characteristic (ROC) points to move up or down the ROC curve without it changing its height. A difference in detection accuracy in the absence of a decision threshold shift would change the height of the ROC curve because the FPFs and TPFs at each threshold of certainty do not change together. ROC curves increase in height because either the TPFs are increasing while the FPFs stay the same, or the FPFs are decreasing while the TPFs stay the same. Attributing differences in FPFs and TPFs between experimental conditions to either a difference in accuracy (curve height) or to a difference in willingness to report (position of points along the curve) is the basic goal of ROC analysis.

Decision Threshold Shifts

Recently decision threshold shifts have been found in chest radiography SOS (17, 18) supplanting the reduction in detection accuracy found in the past (15, 16). Decision threshold shifts reflect reductions in TPFs accompanied by reductions in FPFs: the ROC points move downward along the ROC curve. We studied decision thresholds using the FPF associated with each ROC point (26). In our scoring, a report of abnormality at a location other than the simulated nodule was considered a FP if it occurred on a case with no native abnormality and it referred to an abnormality other than a nodule (results were similar when nodule FP reports were included). For each reader-treatment combination, the midpoint of the FP range of ROC points was computed as the average of most and least conservative ROC points. As such, we get an array of midpoint values consisting of the two SOS conditions X 20 readers (40 values). A similar array was obtained for TP responses. Analyses of Variance (ANOVA) with a within-subject factor for SOS treatment condition and a between-subject factor for resident vs. faculty were used to analyze midpoints of FP and TP ranges separately (BMDP2V, release: 8.0.; 1993; BMDP Statistical Software, Inc. Statistical Solutions Ltd., Cork, Ireland (<http://www.statsol.ie>)).

Results

All readers (14 males, 6 females; mean age residents 32.2 years, mean age attendings 44.5) reviewed the cases after an average of 8.1 hours (SE = 0.81) interpreting an average of 44.9 cases (SE = 7.00) clinical cases: 55% reported reading a single modality (plain film, CT, MRFI, nuclear medicine, interventional, fluoroscopy), while 28% reported reading a mix of 2 modalities and 17% 3 modalities. All readers had 20–20 or corrected vision.

Diagnostic Performance

Average ROC points associated with non-nodule reports for faculty and residents in the non-SOS and SOS conditions (each point is the average of 10 readers) are shown in Figure 2.

According to the OR-DBM MPMC analysis, there was no significant difference between the SOS (mean empirical AUC = 0.667) and non-SOS conditions (mean empirical AUC = 0.679), ($F(1, 19) = 1.04, p = 0.3204$). Other analytic choices such as other ROC models (contaminated binormal model) and other accuracy parameters (true positive fraction at a fixed false positive fraction = 0.1) yielded similar results. We note that when this analysis was extended to include a between-subject factor for level of training, there remained no significant SOS difference in detection accuracy and there was no significant interaction of level of training with SOS condition, indicating no SOS effect on detection accuracy for either group.

Decision Thresholds

From Figure 2, it seems clear that the FPFs associated with the averaged operating points were reduced in the SOS condition (i.e., the solid symbols shift lower to and to the left) relative to the non-SOS condition (open symbols) for both faculty and residents. The statistical analyses confirmed the shifts in decision thresholds.

The TPF at midpoint of the TP range of the operating points was significantly less for residents than faculty (0.294 vs. 0.423, $F(1, 18) = 5.30, P < 0.05$). Similarly, the midpoint of the FP range of the operating points was significantly less in residents than faculty (0.058 vs. 0.093, $F(1, 18) = 4.72, P < 0.05$). Taken together, these findings suggest that residents were less willing to report abnormality than faculty, reducing both true and false positive responses. Thus, there was overall more conservative reporting in the SOS condition (i.e., a threshold shift not a change in accuracy).

The TPF at midpoint of the TP range of the operating points was significantly reduced in the SOS condition (TP = 0.385 for the non-SOS condition vs. 0.333 for SOS condition, $F(1, 18) = 10.81, P < 0.01$). The midpoint of the FP range of the operating points was also significantly reduced in the SOS condition (0.102 for the non-SOS condition vs. 0.049 for SOS condition, $F(1, 18) = 19.85, P < 0.001$). Taken together, these findings suggest that readers were less willing to report abnormality in the SOS condition, reducing both true and false positive responses.

Inspection Time

A two-way ANOVA with within-subjects factors for SOS manipulation (addition of nodule) and presence of test abnormality (naïve abnormality) was conducted. The median of each reader's inspection times for the four types of trials was computed: (1) normal cases without nodules, (2) normal cases with added nodules, (3) abnormal cases without added nodules and (4) abnormal cases with added nodules. The addition of the nodule *did not* significantly affect inspection time ($F(1, 19) = 0.96, p = 0.3400$), with non-SOS readings requiring 48.3 seconds on average and SOS readings requiring 46.5 seconds. Presence of a native abnormality required great inspection time ($F(1, 19) = 62.23, p < 0.0001$) with 40.2 seconds required without native abnormalities and 54.5 seconds with native abnormalities. Of course, these inspection times include reporting time so it is not surprising that more time is required when there is something to report.

Discussion

The present results support the findings from the most recent SOS chest radiography studies (17, 18). Instead of observing a change in detection accuracy for SOS vs non-SOS as in earlier chest SOS studies (15, 16), we observed a *shift in reporting criteria* (17), reflecting a reluctance to report (i.e., lower TP and FP rates) or a more conservative reporting strategy. Why the nature of the SOS effect has changed is not clear. The inspection time data address a potential cause of the threshold shift: reduced inspection (search) time. However, there was no significant difference in inspection time between SOS and non-SOS conditions. Subjects did not respond less, making fewer true and false positive responses, because they looked at the images less. The threshold shifts were not the result of changed search behavior, but of reluctance to report abnormalities (native) after another abnormality (nodule) had already been reported. This does mean that all nodules were detected or that every nodule detection resulted in missing the native abnormality. Rather it means that overall when a nodule was present there is a reduced willingness to report the native lesion.

It should be noted that decision thresholds are defined by responses to normal cases, without reference to responses to abnormal cases. As such, TPFs are not needed to study threshold shifts. This is particularly true when there is a change in FPFs associated with ROC points in one experimental condition relative to another, but no AUC difference. The TPFs have to change with the FPFs for the ROC points to remain on the same ROC curve so that the AUCs do not differ. The results show corresponding shifts in both FPFs and TPFs. This is what one expects, and so confirming it is helpful. However, even if there were no TPF changes, we would still conclude that there were decision threshold shifts based on the changes in FPFs and lack of difference in AUC. One might not statistically detect a change in TPFs associated with points because while FPFs and TPFs vary together in a threshold shift, they can vary together non-linearly (thus the need to fit ROC curves). Owing to the fact that smaller effect sizes are harder to detect than larger ones (requiring more data), one may have less statistical power to detect a TPF change because it could be smaller than the FPF change.

These inspection time results were actually quite similar to those found in earlier studies that demonstrated a change in accuracy (i.e., not a threshold shift). For example, in a 1991 study Berbaum et al. studied the time course of SOS using the same chest image with native abnormalities and added nodule paradigm (27). They found that time to detect the nodules did not depend on the presence of native abnormalities (and vice versa), and total viewing times were the same for zero, one or two lesions. Therefore it seems that radiologists have not fundamentally changed the way they visually search and process image data, but rather may have changed their thresholds for deciding to report abnormalities.

The presence of fatigue did not alter the type of the SOS effect demonstrated in recent studies (17, 18): there was reduced willingness to report abnormalities rather than reduced detection accuracy. We now have 3 studies suggesting that the nature of the SOS effect has changed since the original film-based studies were conducted. Unfortunately, a direct statistical comparison of the current and previous data (17, 18) would not be meaningful because of differences in reader sampling involving many more faculty readers in the current

study. We can contrast the results and conclusions. For both detection accuracy and for decision thresholds, the direction, magnitude and statistical significance were similar for the fatigued and non-fatigued readers (17). This suggests an account of how fatigue and satisfaction of search combine to affect radiologist performance. Previous research on fatigue suggests that fatigue decreases detection accuracy (10, 12). The current study and previous research on fatigue (17, 18) suggests that SOS involves a reluctance to report abnormalities when other abnormalities have already been reported. The current study indicates that fatigue does not alter the SOS effect. Therefore, the effects of fatigue (reduced accuracy leading to missed abnormalities) and of multiple abnormalities (reduced willingness to report leading to missed abnormalities and fewer false positives) may be *additive*. Although this parsimonious account must be treated as a working hypothesis that needs to be tested in further experiments, radiologists should be aware of the likely increase in error rates due to combined effects of fatigue and SOS. Although more research is needed, there have been a number of articles with recommendations on how to reduce or possibly even avoid fatigue in the radiology reading room (4–8). The more obvious solutions include such things as taking scheduled breaks and maintaining regular sleep habits, but some of the more technical include the use of tools such as checklists, structured reporting and computer-aided detection/decision systems (5).

Conclusions

The current research provides a better understanding of the combined effects of fatigue and satisfaction of search effects. The effect of fatigue is reduced detection accuracy indicated by a lowered ROC curve (leading to missed abnormalities). The effect of multiple abnormalities is reduced willingness to report indicated by lowered points along the ROC curve (leading to missed abnormalities and fewer false positives). The study also supports the conclusion of previous studies that the nature of SOS in chest radiography has changed. Further research is needed. A series of experiments that allow within subject comparisons of SOS and non-SOS and fatigued and non-fatigued conditions may permit further advances in our understanding of how fatigue and multi-trauma interact.

Acknowledgments

Support: This work was supported in part by grant 5R01EB004987-07 from the NIBIB

References

1. Kohn, LT, Corrigan, JM., Donaldson, MS., editors. To Err is Human: Building a Safer Health System. National Academy Press; Washington, DC: 1999.
2. Bleich, S. Medical Errors: Five years after the IOM report. Jul. 2005 Commonwealth Fund Publication #830 http://www.commonwealthfund.org/usr_doc/830_Bleich_errors.pdf Last accessed July 13, 2016
3. O'Donnell EP, Humeniuk KM, West CP, Tiburt JC. The effects of fatigue and dissatisfaction on how physicians perceive their social responsibilities. Mayo Clin Proc. 2015; 90:194–201. [PubMed: 25659237]
4. Chan AOM, Chan YH, Chuang KP, Ng JSC, Neo PSH. Addressing physician quality of life: understanding the relationship between burnout, work engagement, compassion fatigue and satisfaction. J Hosp Admin. 2015; 4:46–55.

5. Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *RadioGraphics*. 2015; 35:1668–1676. [PubMed: 26466178]
6. Rohatgi S, Hanna TN, Sliker CW, Abbott RM, Nicola R. After-hours radiology: challenges and strategies for the radiologist. *Am J Roentgen*. 2015; 205:956–961.
7. Hoffmann JC, Mittal S, Hoffmann CH, Fadl A, Baadh A, Katz DS, Flug J. Combating the health risks of sedentary behavior in the contemporary radiology reading room. *Am J Roentgen*. 2016; 206:1135–1140.
8. Waite S, Kolla S, Jeudy J, Legasto A, Macknik SL, Martinez-Condo S, Krupinski EA, Reede DL. Tired in the reading room: the influence of fatigue in radiology. *J Am Coll Radiol*. 2016 In Press.
9. Krupinski EA, Berbaum KS. Measurement of visual strain in radiologists. *Acad Radiol*. 2009; 16:947–950. [PubMed: 19406673]
10. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. *J Am Coll Radiol*. 2010; 7:698–704. [PubMed: 20816631]
11. Krupinski EA, MacKinnon L, Reiner BI. Feasibility of using a biowatch to monitor GSR as a measure of radiologists' stress and fatigue. *Proc SPIE Med Imag*. 2015; 9416:941613.
12. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Madsen MT, Kramer DJ. Do long radiology workdays affect nodule detection in dynamic CT interpretation? *J Am Coll Radiol*. 2012; 9:191–198. [PubMed: 22386166]
13. Taylor-Phillips S, Elze MC, Krupinski EA, Dennick K, Gale AG, Clarke A, Mello-Thoms C. Retrospective review of the drop in observer detection performance over time in lesion-enriched experimental studies. *J Digital Imaging*. 2015; 28:32–40.
14. Berbaum, K., Franken, E., Caldwell, R., Schartz, K. Satisfaction of search in traditional radiographic imaging. In: Samei, E., Krupinski, E., editors. *The Handbook of Medical Image Perception and Techniques*. New York (NY): Cambridge University Press; 2010. p. 107-138.
15. Berbaum KS, Franken EA, Dorfman DD, et al. Satisfaction of search in diagnostic radiology. *Invest Radiol*. 1990; 25:133–140. [PubMed: 2312249]
16. Berbaum KS, Dorfman DD, Franken EA, et al. Proper ROC analysis and joint ROC analysis of the satisfaction of search effect in chest radiography. *Acad Radiol*. 2000; 7:945–958. [PubMed: 11089697]
17. Berbaum KS, Krupinski EA, Schartz KM, et al. Satisfaction of search in chest radiography 2015. *Acad Radiol*. 2015; 22:1457–1465. [PubMed: 26363824]
18. Berbaum KS, Krupinski EA, Schartz KM, et al. The influence of a vocalized checklist on detection of multiple abnormalities in chest radiography. *Acad Radiol*. 2016; 23:413–420. [PubMed: 26916249]
19. Berbaum KS, Franken EA, Caldwell RT, et al. Can a checklist reduce SOS errors in chest radiography? *Acad Radiol*. 2006; 13:296–304. [PubMed: 16488841]
20. DICOM PS3.14 2016c – Grayscale Standard Display Function. <http://dicom.nema.org/medical/dicom/current/output/pdf/part14.pdf> Last accessed July 25, 2016
21. Schartz KM, Berbaum KS, Caldwell RT, Madsen MT. WorkstationJ: workstation emulation software for medical image perception and technology evaluation research. *Proc SPIE Med Imag*. 2007; 6515:65151I–65151I1.
22. Jiang Y, Metz CE, Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology*. 1996; 201:745–750. [PubMed: 8939225]
23. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data - Part II. A formal model. *Academic Radiology*. 2000; 7:427–437. [PubMed: 10845402]
24. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method. *Invest Radiol*. 1992; 27:723–731. [PubMed: 1399456]
25. Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Acad Radiol*. 2008; 15:647–61. [PubMed: 18423323]
26. Swets, JA., Pickett, RM. *Evaluation of diagnostic systems: methods from signal detection theory*. New York: Academic Press; 1982. p. 39

27. Berbaum KS, Franken EA, Dorfman DD, Rooholamini SA, Coffman CE, Cornell SH, Cragg AH, Galvin JR, Honda H, Kao SC, et al. Time course of satisfaction of search. *Invest Radiol.* 1991; 26:640–648. [PubMed: 1885270]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

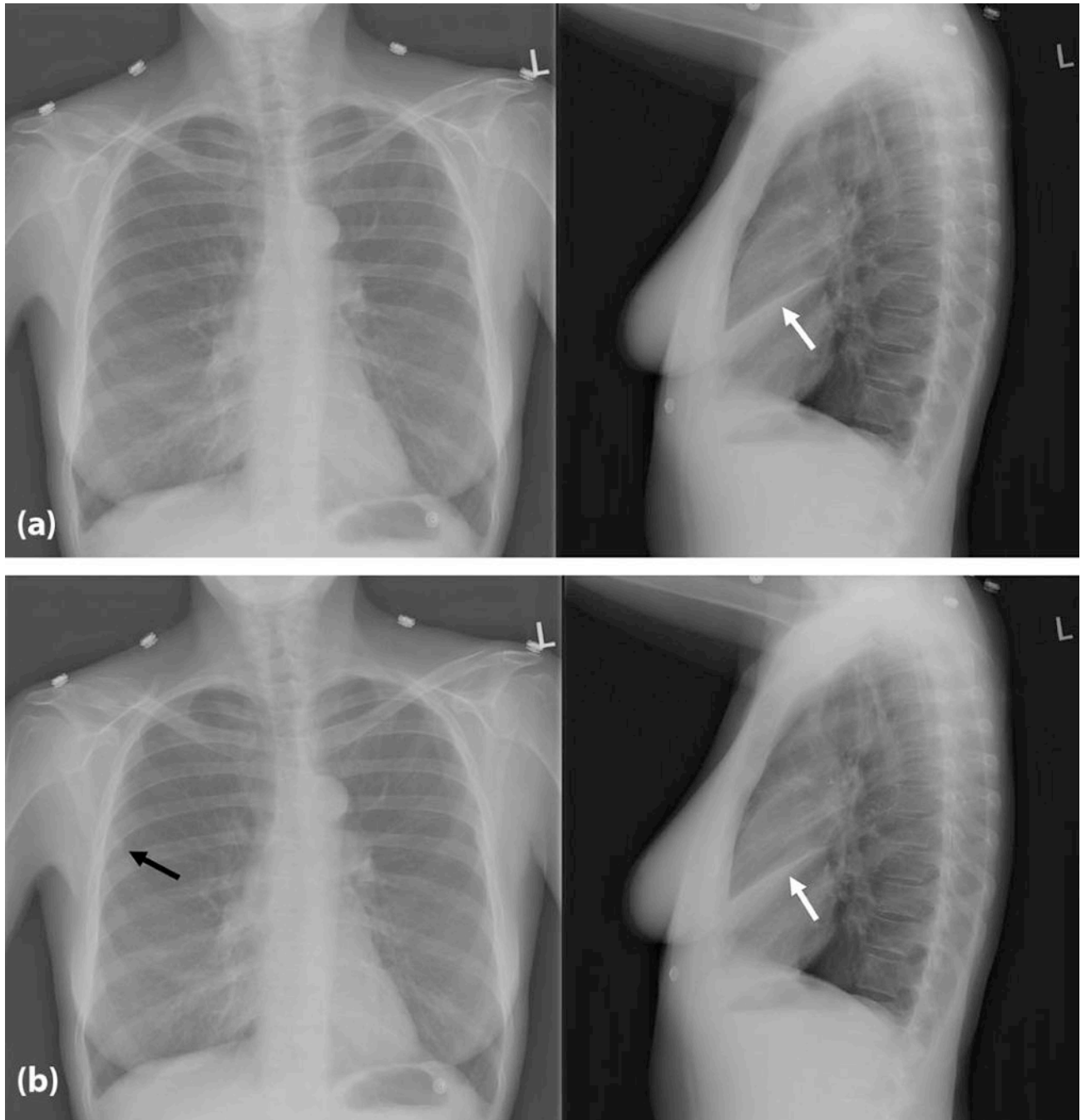


Figure 1. Example of one of the cases used in the study, showing as the native abnormality a right middle lung collapse (white arrow) in the non-SOS (top without added nodule) and SOS (bottom with added nodule indicated by the black arrow) conditions.

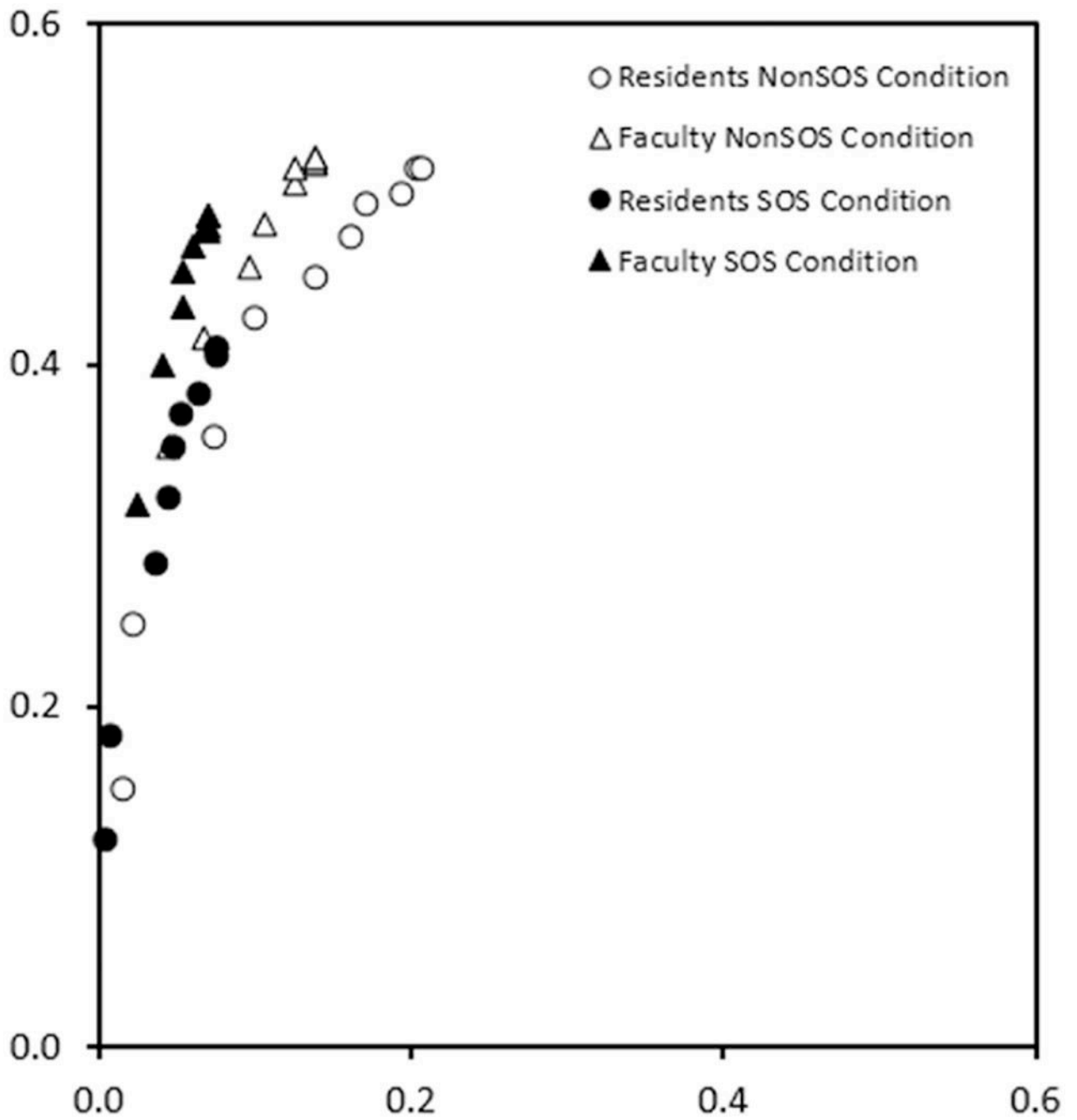


Figure 2. Average ROC points associated with non-nodule reports for faculty and residents in the non-SOS and SOS conditions (each point is the average of 10 readers).