Published in final edited form as:

Acad Radiol. 2007 September ; 14(9): 1069–1076. doi:10.1016/j.acra.2007.06.001.

Does computer-aided diagnosis for lung tumors change satisfaction of search in chest radiography?

Kevin S. Berbaum, PhD, Robert T. Caldwell, MFA, Kevin M. Schartz, PhD, Brad H. Thompson, MD, and E.A. Franken Jr., MD

Departments of Radiology (K.S.B., R.T.C., K.M.S.B.H.T., E.A.F.) and Psychology (K.S.B.), University of Iowa, Iowa City

Abstract

Rationale and Objectives—Computer-aided diagnosis (CAD) has been developed to ensure that the radiologist considers suspect focal opacities that may represent cancer in chest radiography. Although CAD was not developed to counteract the satisfaction of search (SOS) effect, it may be an effective intervention to do so. The objective of this study is to determine whether an idealized CAD can reduce SOS effects in chest radiography.

Materials and Methods—Fifty-seven chest radiographs, half of which demonstrated diverse, native abnormalities were read twice by 16 observers, once with and once without the addition of a simulated pulmonary nodule. Simulated CAD prompts were provided during the interpretation, which unerringly pointed to the *added* simulated nodule. Area under the ROC curve for detecting the native abnormalities was estimated for each observer in each treatment condition. In addition to testing for the SOS effect in the presence of CAD prompts, results were compared to those of a previous SOS study.

Results—Significantly more nodules were reported in the SOS with CAD experiment than in the original SOS experiment (49 vs. 43, $p < 0.01$). An SOS effect was found even when CAD prompts were provided; ROC areas for detecting native abnormalities were reduced with added nodules (0.68 vs. 0.65, $p(\text{one-tailed}) < 0.05$). Comparison of the current experiment with CAD and the previous SOS experiments failed to show a significant difference of the magnitude of the SOS effect ($p = 0.52$). The threshold for reporting was more conservative with CAD prompts than in SOS studies ($p = 0.052$).

Conclusion—Our results indicate that the CAD prompts, even those that always point to their target lesion without false-positive error, fail to counteract SOS in chest radiography. The stricter decision thresholds with CAD prompts may indicate less visual search for native abnormalities.

Keywords

Diagnostic radiology; observer performance; images; interpretation; quality assurance

Address correspondence to: KSB, Department of Radiology, 3170 Medical Laboratories, The University of Iowa, Iowa City, IA 52242.

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.

INTRODUCTION

Satisfaction of search (SOS) occurs when a lesion is “missed” after detecting another lesion in the same image. An SOS effect in chest radiology, defined operationally as a reduced accuracy in detecting native abnormalities on chest radiographs in the presence of simulated pulmonary nodules, has been demonstrated [1–3].

A promising approach to prevent SOS errors in medical imaging diagnosis is computer-aided diagnosis (CAD). CAD has been used to find tumors on chest radiographs, chest computed tomography, and mammograms. The goal of computer-aided diagnosis for nodule detection is to ensure that the interpreting physician considers regions that have a high probability of disease. An obvious radiology examination in which to study the effects of CAD on SOS is chest radiography. Not only have SOS effects on the detection of lung nodules and other abnormalities been quantified [1–4], CAD systems for nodule detection in chest imaging are also well developed and commercially available [5–7].

In the current study, we attempt to discover whether computer-aided diagnosis can counteract satisfaction of search effects in chest radiology, or alternatively, whether there are penalties with reduced detection of unprompted abnormalities. Current CAD algorithms find nodules in chest radiographs missed by unaided readers but also report false-positive locations [8–11]. It is not possible to predict the level of performance CAD will ultimately achieve as algorithms continue to improve [12–15]. To find out whether a future, perfectly performing CAD can prevent SOS in chest radiography, we simulated CAD prompts that pointed only to simulated nodules we placed in the lung fields. We reasoned that if such a CAD system could not prevent SOS, it would be unlikely that a less accurate CAD could do so.

MATERIALS AND METHODS

Experimental Conditions

Laboratory studies use an operational definition of SOS not found in accounts of clinical errors. According to this operational definition, *the lesion that is missed because of SOS is shown to be detected in the absence of other lesions*. To accomplish this, a known abnormality is defined as the *test* abnormality because detection of that abnormality is measured. The test abnormalities are always presented twice to observers: once alone and once with another abnormality added to that same study. SOS occurs when the test abnormality is missed in the presence of the added abnormality, but not in its absence. Although detection of any abnormality may affect detection of any other abnormality, SOS can only be measured rigorously on the test abnormality because it is the only abnormality that is presented by itself as a critical control condition.

The detection accuracy for native, subtle lesions was compared with that for those same lesions when a simulated pulmonary nodule was added photographically to the radiograph. Thus, the background anatomy and actual lesions were perfectly matched for the two conditions. Simulated and native lesions were not spatially superimposed, and the native abnormalities were physically identical with and without the nodules.

To test whether an idealized CAD prompt could alter the satisfaction of search effect, we used the same two conditions that were used in SOS demonstrations: presentation of each chest radiograph with and without a simulated pulmonary nodule. The experiment consisted of two parts, separated in time by two months. Half the cases presented during each part contained a simulated pulmonary nodule, and half the cases did not. Thus, during the course of the two parts, each radiograph appeared twice (once with and once without the simulated nodule). Simultaneously with the presentation of examination, observers were provided with a CAD

prompt on the report form that indicated the location of a potential pulmonary nodule. CAD prompts pointed unerringly to the simulated nodules we placed in the lung fields and were only present when a nodule appeared in the radiograph. Thus, we repeated the SOS experiment, but added CAD prompts that unerringly pointed at the nodules we placed there. Because the nodules only appeared in the SOS condition, the CAD prompts could only appear in that condition.

Figure 1 shows a response form for a patient in the experiment that includes photographs of the patient's images with a CAD prompt superimposed. A white circular disk was overlaid on a digitally acquired photograph of the radiograph and corresponded to location of a simulated pulmonary nodule. The response form for each patient presented in the experiment included photographs that represented the actual radiographs as displayed on the film alternator.

Case Sample

This experiment was performed using 57 radiographic examinations of the chest that had been used in earlier SOS studies [1,3,16,17]. Thirty cases had diverse, subtle, and clinically important native abnormalities, and 27 had no native abnormalities. As in other SOS experiments, the detection accuracy for native abnormalities when they appeared without simulated nodules was compared with accuracy for those same abnormalities when a simulated pulmonary nodule was introduced. However, in the current experiment, the simulated pulmonary nodule was always accompanied by a CAD prompt.

The native abnormalities of our sample included 6 abnormalities of the lungs, 4 abnormalities of the heart and/or great vessels, 4 abnormalities of the mediastinum and/or trachea, 3 abnormalities of the chest wall, ribs or sternum, no abnormalities of the pleural cavity, 7 abnormalities of the humeri, scapulae or clavicles, no abnormalities of the neck and/or cervical airway, 2 abnormalities of the vertebrae, 2 abnormalities of the abdomen, and 2 abnormalities in other areas of the film.

Procedure

Sixteen volunteers from the residents and faculty of our Department of Radiology served as observers. Their informed consent was obtained after the nature of their task was explained. Observers were told that this was not an examination of individual performance and that their responses would be coded to ensure confidentiality. The frequency of abnormalities was not revealed. Before actual data collection, a practice case was used to ensure proper understanding of the procedure.

A booklet was provided to observers for recording their responses. Figure 1 presents a facsimile of one page of this booklet. As each case was presented, observers were asked to inspect the radiograph until they felt certain that all abnormal features had been reported. The following is an excerpt from the instructions:

“Some of the cases you will read will be completely normal and others will contain one or more clinically significant abnormalities. We are interested only in those features that represent significant abnormality. Avoid “over calling”. A list of features to generally ignore is posted next to the alternator and is also included in your scoring booklet.

There is a separate response sheet for each case presented. The top of the response sheet lists the alternator position, the patient's age and sex, and sometimes a comment about some attribute of the film we would like you to ignore. At the bottom of the response sheet there is a photostat representing a PA view of the case (and sometimes a lateral view). On some PA views there will be a CAD indication of the location of

a potential pulmonary nodule (white dot). Please review this area on the film and reject or accept this indication by checking the appropriate box. If you accept the CAD indication, circle your confidence that the feature is abnormal. Otherwise, leave the CAD section blank.

Continue describing other abnormal features on the film. As you do, number them and draw their locations on the photostat, give a likely diagnosis if possible and rate your confidence (one to four) that the feature is abnormal. The rating categories are, 1) suspicious, but probably normal; 2) possibly abnormal; 3) probably abnormal; and 4) definitely abnormal. If you find no abnormal features, check the box on the response sheet indicating, "normal features, nothing to report" and proceed to the next case."

Inspection of the posteroanterior radiograph and, when available, the lateral radiograph continued until the observer believed that all abnormal features had been reported.

Scoring

Because adding a simulated nodule to each examination was an experimental manipulation, ratings associated with the simulated nodules were not scored for ROC analysis. The primary criterion for determining whether a rating applied to the simulated rather than to the native lesion was the location of the drawn lesion. The diagnosis given to the drawn feature served as a secondary criterion. When no native abnormality was present on the radiograph, ratings with a location or category description that could not be assigned to the simulated lesion were scored as a false-positive result; that is, they were scored toward normal examinations. When a native abnormality was present, ratings not attributed to a simulated lesion were scored as a true-positive result; that is, they were scored toward abnormal examinations. If these ratings were at the wrong location, however, they were treated as false-negative responses by assigning to them the lowest confidence rating.

For our scoring, the box labeled "normal, nothing to report" represented the lowest confidence that an abnormality was present and was translated into a rating of 1 on a 1 to 5 scale. In other words, whenever only the nodule was reported and otherwise only these boxes were checked, then a rating of one was scored for that case. When an abnormal feature was detected, the observer was asked to report a confidence rating in abnormality using four categories: suspicious, but probably normal; possibly abnormal; probably abnormal; and definitely abnormal. These responses were translated into 2, 3, 4, and 5, respectively, on the 1 to 5 scale.

Comparison with Earlier Experiments

Data from previous SOS experiments in chest radiology [1,3] were reanalyzed for comparison with data from our current experiment. Data analyzed were limited to the same 57 cases that were used in the current study. Nineteen observers (different from the current group) had participated in the initial SOS experiment unaltered by any intervention [3]. Like the current experiment, the original SOS experiments consisted of two parts, separated in time by several months. Thirty of the examinations contained a subtle native abnormality, and twenty-seven examinations were normal chest radiographs. Half the cases presented during each part contained a simulated pulmonary nodule, and half the cases did not. Thus, during the course of the two parts, each radiograph appeared twice (once with and once without the simulated nodule).

Data Analysis

The contaminated binormal model developed by Dorfman and Berbaum [18] gives proper ROC curves. The area under the contaminated binormal ROC curve was used to measure detection accuracy for native abnormalities. Areas under the ROC curves for the detection of native abnormalities were estimated for each observer in each experimental condition using the

contaminated binormal model, and the average areas were compared by using a *t*-test for paired observations [19] with degrees of freedom determined by the number of readers. This was the fundamental test of whether SOS occurred in the current experiment.

To compare the ROC areas for detecting native abnormalities and the number of our simulated nodules reported in the current experiment and the SOS study [1,3], we compared ROC areas in the control conditions of the two experiments, conditions which did not differ in any way. In this condition, no nodules were added and because the CAD prompts only pointed to placed nodules, there were no CAD prompts in either condition. The only difference that could result would be related to differences in reader accuracy for our set of radiographs. Because of evidence of unequal variances between groups, we performed a non-parametric Wilcoxon-Mann-Whitney rank-sum test [20] instead of a *t*-test for independent groups [19].

Of course, we knew from the outset that the level of experience of the readers in the two experiments would differ. In the prior SOS study [1,3], there were 1 second-year resident, 1 third-year resident, 5 fellows, and 12 faculty radiologists (4 of whom were chest radiologists). In the current experiment, there were 3 second-year residents, 5 third-year residents, 6 fourth-year residents, and 2 fellows (neither were chest radiologists). Of course, by the end of their first year, radiology residents have traditionally taken call, reading radiographic studies independently. Although different levels of training and image interpretation experience could lead to differences in detection accuracy between the experiments, there were other differences in the circumstances of reading. Between the time when the SOS experiments were conducted, in 1989 [1] and 2000 [3], and when the current experiment was conducted, in 2006 and 2007, the practice of radiology changed. Today's readers have less experience with the look of film and the computed radiographs that they are familiar with differ in a number of respects from the film radiographs that we used in both the current and previous studies. The difference in familiarity with the images, as well as other changes in the practice of radiology between the current and previous experiments, could account for any difference in overall performance as easily as differences in experience.

Our main concern, therefore, was not to test performance differences between the current and previous experiments, but to test whether there was a difference in the magnitude of SOS effects between these experiments. To do so, we subtracted the ROC area of the SOS condition from the ROC area of the non-SOS condition to get a difference score for SOS. We tested whether there was a difference in the magnitude of the SOS effect between the current experiment and the previous SOS experiments with a Wilcoxon-Mann-Whitney rank-sum test, preferring the non-parametric test because of the difference in variance between the experiments noted above.

ROC points are defined by sensitivities and specificities associated with boundaries between rating categories. Decision thresholds are revealed by fitting an ROC model to the response probabilities on either side of boundaries between the categorical ratings. For example, "definitely abnormal" versus "probably abnormal" defines the most conservative decision threshold in our study. The next most conservative decision threshold is defined by the boundary between "probably abnormal" and "possibly abnormal", the next by the boundary between "possibly abnormal" and "suspicious, but probably normal." The most lenient decision threshold in our study was defined by the boundary between "suspicious, but probably normal" and not making any response. Clearly, not all of the decision thresholds are equally consequential in patient care. The difference between "probably abnormal" and "definitely abnormal" would often lead to the same treatment, but the difference between "possibly abnormal" and "suspicious, but probably normal" may lead to very different treatments. Furthermore, the different thresholds may be more or less affected by factors such as SOS or CAD prompts. Thresholds close to not reporting may be more affected by factors that change the focus of attention than thresholds differentiating between high levels of certainty.

To determine whether there are any shifts in response readiness associated with added nodules or CAD prompts, we analyzed the decision threshold between “possibly abnormal” and “suspicious, but probably normal.” There are a variety of ways of measuring decision thresholds, but the most intuitive of these is simply the false-positive fraction associated with the ROC point. We measure the decision threshold by the false-positive rate associated with the fitted ROC points. In our experience, these predicted ROC points differ little from the observed ROC points, which are defined by sensitivities and specificities that can be calculated directly from the data. By limiting ourselves to one ROC point, we can use the same statistical methods as for ROC areas, except that, having no predicted difference direction, we use two-tailed tests (p values reported are two-tailed unless otherwise noted).

RESULTS

Effect of Simulated CAD on Reporting Nodules

We counted the number of simulated nodules that were reported by each reader in the SOS condition of each experiment and tested for a difference in the rate of report using a t -test for independent groups. Significantly more simulated nodules were reported in the SOS with CAD experiment than in the original SOS experiment (49.3 vs. 43.4 nodule reports, $t(33) = 3.31$, $p < 0.01$).

Contaminated Binormal Areas for Detecting Native Abnormalities

In the SOS experiment with CAD prompts of nodules provided, the paired t -test on contaminated binormal areas for detecting native abnormalities estimated for each reader in each condition found a significant difference without and with added nodules (ROC area of 0.681 without nodules vs. 0.653 with nodules and prompts, $t(15) = 1.76$, $p(\text{one-tailed}) < 0.05$); thus CAD failed to eliminate the SOS effect. A similar analysis of the data from the previous SOS study without CAD prompts [1,3] also demonstrated an SOS effect. A paired t -test on contaminated binormal areas for detecting native abnormalities estimated for each reader in each condition found a significant difference without and with added nodules (ROC area of 0.745 without nodules vs. 0.701 with nodules, $t(18) = 5.11$, $p(\text{one-tailed}) < 0.001$).

Preliminary to testing for a difference in detection accuracy in *control* conditions of the two experiments using a t -test for independent groups, a test on variances showed them to differ (Levene $F(1,33)$ for variability = 5.04, $p < 0.05$), indicating a violation of the heterogeneity of variance assumption of the t -test. A non-parametric Wilcoxon-Mann-Whitney rank-sum test showed non-SOS condition detection accuracy for *test* abnormalities to be greater in the original SOS experiment than in the current experiment ($p < 0.05$, median ROC area = 0.748 vs. 0.673). We conclude that for whatever reason, accuracy was higher for the original readers in the SOS experiments than for the readers of the current experiment.

Because of the significant test for variability reported above, a nonparametric approach is indicated for further tests between groups. Having subtracted the ROC area of the SOS condition from the ROC area of the non-SOS condition for each reader to get a difference score for SOS, we tested whether there was a difference in the magnitude of the SOS effect between the current experiment and the previous SOS experiments using a non-parametric Wilcoxon-Mann-Whitney rank-sum test. This test failed to find a significant difference ($p = 0.52$, with a median SOS difference of 0.044 for the original SOS experiment and of 0.028 for the current experiment). Therefore, there was no statistically significant difference in the magnitude of the SOS effect between the current and previous experiments. Given this result, it was reasonable to test the SOS effect across both experiments. A t -test for paired observations on ROC areas for detecting native abnormalities gave a significant difference (ROC area of 0.716 without nodules vs. 0.679 with nodules, $t(34) = 4.28$, $p < 0.001$).

Thresholds for Detecting Native Abnormalities

In the current experiment with CAD prompts of nodules provided, the paired t -test on the false-positive fraction associated with the “suspicious, but probably normal”/“possibly abnormal” decision threshold for detecting native abnormalities estimated for each reader in each condition found a significant difference without and with added nodules (false-positive fraction of 0.151 without nodules vs. 0.103 with nodules and prompts, $t(15) = 2.33$, $p = 0.0342$), suggesting a threshold shift toward stricter reporting with the addition of nodules. By contrast, a similar analysis of the data from the previous SOS study without CAD prompts [1,3] demonstrated no threshold shift. A paired t -test on decision threshold for detecting native abnormalities estimated for each reader in each condition found no significant difference without and with added nodules (false-positive fraction of 0.227 without nodules vs. 0.192 with nodules and prompts, $t(18) = 1.58$, $p = 0.132$).

A t -test for independent samples was used to test for a difference in decision threshold in control conditions of the two experiments. This test demonstrated a marginally significant difference in the decision thresholds (false-positive rate of 0.227 without CAD vs. 0.151 with CAD, $t(33) = 2.02$, $p = 0.052$). Therefore, there was a shift toward a more conservative threshold for reporting with the CAD prompt experiment, even in the condition without nodules or CAD prompts. We performed a similar set of tests for a difference in decision threshold in experimental conditions (with added nodules) of the two experiments. This test demonstrated a significant difference in the decision thresholds (false-positive rate of 0.188 without CAD vs. 0.103 with CAD, $t(33) = 2.60$, $p < 0.05$). Therefore, the same shift was evident in the experimental conditions as well.

For each reader, we subtracted the false-positive rate of the SOS condition from the false-positive rate of the non-SOS condition to get a difference score. We tested whether there was a difference in the magnitude of the threshold shift between the current experiment and the previous SOS experiments using a t -test for independent groups. The difference in false-positive rates between the experiments was not significantly different (false-positive rate of 0.048 for the current experiment and of 0.035 for the previous SOS experiment, $t(33) = 0.43$, $p = 0.67$). Therefore, there was no statistically significant difference between the current and previous experiments in the magnitude of the threshold shift from the conditions without and with nodules. Given this result, it was reasonable to test for the threshold shift with the addition of nodules across both experiments. A t -test for independent groups gave a statistically significant difference (false-positive rate of 0.192 without nodules vs. 0.151 with nodules, $t(34) = 2.71$, $p < 0.05$). Thus, there was a shift toward more conservative thresholds for reporting native abnormalities with the addition of the nodules.

DISCUSSION

The rationale for CAD is that computer algorithms and human interpreters may have different, but complementary strengths. The hope is that by providing the human observer with the computer's findings, detection accuracy will be improved. Computer aids to help find cancer in mammograms, chest radiographs, and chest CT have been developed, where the observer's task is difficult but focused on a single type of disease. The goal of CAD for nodule detection is to ensure that the radiologist considers regions that have a high probability of potential cancer. We hoped that computer diagnostic aids might alleviate SOS effects, reducing misses that affect patient morbidity and mortality. This hope was not realized in the results of the experiment.

Although CAD increased the recognition of the simulated nodules, it did not eliminate the SOS effect of nodules on the detection of native abnormalities. The current experiment cannot address whether CAD itself could affect detection of the native abnormalities because of

differences from earlier SOS studies in the level of training of observers, their familiarity with our type of images, and changes in the practice of radiology in the intervening period. We hope eventually to repeat the SOS with CAD experiment with a companion SOS experiment using matched observers and computed radiography images.

The decision threshold we studied was chosen to reflect the decision to report or not report abnormality. Our findings for this threshold level may not necessarily be reproduced at more conservative thresholds that differentiate between high levels of certainty in reporting. This decision threshold shifted toward more conservative reporting in the current study relative to those in the previous SOS study [3]. Shifts in decision thresholds have not been previously noted in SOS studies with chest radiography, only the simple reduction in detection accuracy. This might happen if all thresholds were studied together because more conservative thresholds may not be affected. Shifts in decision thresholds are more typical of SOS effects in contrast studies of the abdomen [21–23]. In research using gaze-dwell time recording, reduced thresholds have been linked to a global change in search behavior—a visual neglect of the regions containing the abnormalities. To make a similar determination for CAD prompts and SOS effects in chest radiography, we need measurements of visual search. Although our initial investigation of the influence of CAD on SOS gave a negative result with regard to the question of whether SOS could be prevented using CAD, it suggests further experiments that could lead to a better understanding of both SOS and the psychophysical effects of CAD.

Our study is not the first to demonstrate limitations in computer diagnostic aids. Alberdi, Povyakalo, Strigini, et al. [24] studied the influence of failure to prompt cancers by a CAD for mammography that had previously shown high sensitivity, though low specificity on the sensitivity of human readers. Sensitivity of human readers was significantly reduced when CAD failed to mark cancers. They explained this finding by suggesting that readers, based on past experience with the CAD tool that ordinarily finds most cancers but also gives many false-positive indications, assumed that the absence of a CAD prompt was a reliable indication of a normal examination, and so paid less attention to cases with no prompts or discounted their own detection. The surprising outcome was that the absence of CAD prompts seemed to be used by readers more than their presence. There have been other reports of poorer detection of non-prompted cancers when CAD prompts are provided [25]. Our own results are less startling or worrisome. They only suggest that CAD does not reduce SOS error for other types of abnormality by easing the radiologist's burden of detecting nodules, an activity that seems to produce the SOS errors.

Acknowledgments

Supported by USPHS Grants R01 EB/CA00145 and R01 EB/CA00863 from the National Cancer Institute, Bethesda, Maryland.

References

1. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Satisfaction of search in diagnostic radiology. *Invest Radiol* 1990;25:133–140. [PubMed: 2312249]
2. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Time course of satisfaction of search. *Invest Radiol* 1991;26:640–648. [PubMed: 1885270]
3. Berbaum KS, Dorfman DD, Franken EA Jr, Caldwell RT. Proper ROC analysis and joint ROC analysis of the satisfaction of search effect in chest radiography. *Acad Radiol* 2000;7:945–958. [PubMed: 11089697]
4. Samuel S, Kundel HL, Nodine CF, Toto LC. Mechanism of satisfaction of search: Eye position recordings in the reading of chest radiographs. *Radiology* 1995;194:895–902. [PubMed: 7862998]
5. Abe H, Macmahon H, Shiraishi J, Li Q, Engelmann R, Doi K. Computer-aided diagnosis in chest radiology. *Semin Ultrasound CT MR* 2004;25:432–7. [PubMed: 15559126]

6. Kakeda S, Moriya J, Sato H, Aoki T, Watanabe H, Nakata H, Oda N, Katsuragawa S, Yamamoto K, Doi K. Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. *AJR Am J Roentgenol* 2004;182:505–10. [PubMed: 14736690]
7. Freedman M. State-of-the-art screening for lung cancer (part 1): the chest radiograph. *Thorac Surg Clin* 2004;14:43–52. [PubMed: 15382307]
8. Das M, Mühlenbruch G, Mahnken AH, Flohr TG, Gündel L, Stanzel S, Kraus T, Günther RW, Wildberger JE. Small pulmonary nodules: Effect of two computer-aided detection systems on radiologist performance. *Radiology* 2006;241:564–571. [PubMed: 17057074]
9. Sakai S, Soeda H, Takahashi N, Okafuji T, Yoshitake T, Yabuuchi H, Yoshino I, Yamamoto K, Honda H, Doi K. Computer-aided nodule detection on digital chest radiography: validation test on consecutive t1 cases of resectable lung cancer. *J Digit Imaging* 2006;19:376–82. [PubMed: 16763934]
10. Shiraishi J, Abe H, Li F, Engelmann R, MacMahon H, Doi K. Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs ROC analysis of radiologists' performance. *Acad Radiol* 2006;13:995–1003. [PubMed: 16843852]
11. Qiang, Li. Recent progress in computer-aided diagnosis of lung nodules on thin-section CT. 1: Comput Med Imaging Graph 2007. in press; [Epub ahead of print] (<http://www.sciencedirect.com/science/article/B6T5K-4N9MYKV-3/2/094de96a6b678e72973822b412b31922>)
12. Shiraishi J, Li F, Doi K. Computer-aided diagnosis for improved detection of lung nodules by use of posterior-anterior and lateral chest radiographs. *Acad Radiol* 2007;14:28–37. [PubMed: 17178363]
13. Suzuki K, Shiraishi J, Abe H, MacMahon H, Doi K. False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network. *Acad Radiol* 2005;12:191–201. [PubMed: 15721596]
14. Marten K, Engelke C, Seyfarth T, Grillhosi A, Obenauer S, Rummeny EJ. Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. *Clin Radiol* 2005;60:196–206. [PubMed: 15664574]
15. Shiraishi J, Li Q, Suzuki K, Engelmann R, Doi K. Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: localized search method based on anatomical classification. *Med Phys* 2006;33:2642–53. [PubMed: 16898468]
16. Berbaum KS, Franken EA Jr, Dorfman DD, et al. The role of faulty decision making in the satisfaction of search effect in chest radiology. *Acad Radiol* 2000;7:1098–1106. [PubMed: 11131054]
17. Berbaum KS, Franken EA Jr, Caldwell RT, Scharzt KM. Can a checklist reduce SOS errors in chest radiography? *Academic Radiology* 2006;13:296–304. [PubMed: 16488841]
18. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data - Part II. A formal model *Acad Radiol* 2000;7:427–437.
19. McNemar, Q. Psychological statistics. Vol. 4. New York, NY: John Wiley and Sons; 1969. p. 110-121.
20. Siegel, S.; Castellan, NJ, Jr. Nonparametric statistics for the behavioral sciences. Vol. 2. New York: McGraw-Hill Book Company; 1988. p. 73-167.
21. Franken EA Jr, Berbaum KS, Lu CH, et al. Satisfaction of search in detection of plain film abnormalities in abdominal contrast examinations. *Invest Radiol* 1994;29:403–409. [PubMed: 8034444]
22. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Cause of satisfaction of search effects in contrast studies of the abdomen. *Acad Radiol* 1996;3:815–826. [PubMed: 8923900]
23. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Can order of report prevent satisfaction of search in abdominal contrast studies? *Acad Radiol* 2005;2:74–84. [PubMed: 15691728]
24. Alberdi E, Povyakalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Acad Radiol* 2004;11:909–18. [PubMed: 15354301]
25. Poller, WR.; Zheng, B.; Sumkin, JH.; Gur, D. Detecting abnormalities in non-cued areas of digitized mammograms: An observer experience. *Medical Imaging 2002: Image Perception, Observer Performance, and Technology Assessment*. In: Chakraborty Dev, P.; Krupinski Elizabeth, A., editors. *Proceedings of SPIE*. Vol. 4686. 2002. p. 84-88.

Panel 50, right: Practice Case. This is a 35-year old male.

If you find no abnormal features, check the box "normal features, nothing to report" and proceed to the next case. Please reject or accept any CAD indication. Otherwise, please describe all abnormal features, draw a line to their locations on the image below, give a likely diagnosis, and rate your confidence that the feature is abnormal using the following scale:

☐ normal features, nothing to report

1 = suspicious, but probably normal
 2 = possibly abnormal
 3 = probably abnormal
 4 = definitely abnormal

☐ Reject CAD	☐ Accept CAD	Circle Confidence			
		1	2	3	4

Abnormal feature	Specific diagnosis	Confidence

Figure 1.

Facsimile of a page from the scoring booklet. White dot on the radiograph indicates the location of a potential pulmonary nodule.