Title
The Effect of Teaching Search Strategies on Perceptual Performance

Journal
Academic Radiology, 24(6)

ISSN
1076-6332

Authors
van der Gijp, Anouk
Vincken, Koen L
Boscardin, Christy
et al.

Publication Date
2017-06-01

DOI
10.1016/j.acra.2017.01.007

Peer reviewed

The Effect of Teaching Search Strategies on Perceptual Performance

Anouk van der Gijp, MD, Koen L. Vincken, PhD, Christy Boscardin, PhD, Emily M. Webb, MD,
Olle Th.J. ten Cate, PhD, David M. Naeger, MD

Rationale and Objectives: Radiology expertise is dependent on the use of efficient search strategies. The aim of this study is to investigate the effect of teaching search strategies on trainee's accuracy in detecting lung nodules at computed tomography.

Materials and Methods: Two search strategies, "scanning" and "drilling," were tested with a randomized crossover design. Nineteen junior radiology residents were randomized into two groups. Both groups first completed a baseline lung nodule detection test allowing a free search strategy, followed by a test after scanning instruction and drilling instruction or vice versa. True positive (TP) and false positive (FP) scores and scroll behavior were registered. A mixed-design analysis of variance was applied to compare the three search conditions.

Results: Search strategy instruction had a significant effect on scroll behavior, $F(1.3) = 54.2$, $P < 0.001$; TP score, $F(2) = 16.1$, $P < 0.001$; and FP score, $F(1.3) = 15.3$, $P < 0.001$. Scanning instruction resulted in significantly lower TP scores than drilling instruction (M = 10.7, SD = 5.0 versus M = 16.3, SD = 5.3), $t(18) = 4.78$, $P < 0.001$; or free search (M = 15.3, SD = 4.6), $t(18) = 4.44$, $P < 0.001$. TP scores for drilling did not significantly differ from free search. FP scores for drilling (M = 7.3, SD = 5.6) were significantly lower than for free search (M = 12.5, SD = 7.8), $t(18) = 4.86$, $P < 0.001$.

Conclusions: Teaching a drilling strategy is preferable to teaching a scanning strategy for finding lung nodules.

Key Words: Radiology education; pulmonary nodules; medical image perception; search strategies.

INTRODUCTION

Perceptual errors account for a substantial part of misdiagnoses in radiology (1) and can be related to the search behavior of the observer (2). For educational purposes, it is important to identify which visual search patterns are most effective and to investigate if teaching search strategies improves perception.

Visual search characteristics that are related to expertise and high performance have been identified in various radiology perception tasks (3). For example experts tend to fixate on abnormalities faster (4–6) and need less time and a smaller number of eye fixations to inspect the image (7,8). These characteristics derive from experience, and they lack an underlying structure that can be taught to novices.

Some specific visual search patterns are found to be related to high performance (4,9–11). Most patterns apply to visual search in X-rays, such as chest X-rays or mammography. Two visual search types are distinguished for searching chest computed tomography (CT) images: "scanners" and "drillers" (11). Scanners tend to visually search a single slice, before scrolling further through the stack, whereas drillers focus their eyes on one quadrant of the lung fields and quickly scroll through the stack in depth before moving to another quadrant. Drillers outperformed scanners with respect to higher true positive rates and a larger lung coverage (11). One interesting finding was that, when given the option to search freely, more experienced readers tend to select "drilling" as a search pattern (the more effective pattern), suggesting it might be a pattern that has, consciously or unconsciously, evolved through instruction or practice. The relationship between search patterns and experience has been noted in several other studies (4,9–11), although it is unknown if experts unconsciously adopt these patterns or deliberately chose or had acquired one, as a *strategy*.

Teaching junior trainees to use expert search strategies may not necessarily be effective. First, learning the strategy may not be easy, particularly given that most experts acquire their behaviors after years of practice. Second, the improvement in perceptual performance that comes with experience is probably due to multiple factors. Knowledge gained and feedback received are known to be critical factors in developing visual

---

Acad Radiol 2017; 24:762–767

From the Department of Radiology and Biomedical Imaging, University of California, San Francisco, 505 Parnassus Ave., M-391, San Francisco, CA 94143-0628 (A.v.d.G., E.M.W., D.M.N.); Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands (K.L.V.); Office of Medical Education, University of California, San Francisco, San Francisco, California (C.B.); Center for Research and Development of Education, University Medical Center Utrecht, Utrecht, The Netherlands (O.TJ.t.C.). Received November 14, 2016; revised January 16, 2017; accepted January 17, 2017. **Address correspondence to:** D.M.N. e-mail: david.naeger@ucsf.edu

http://dx.doi.org/10.1016/j.acra.2017.01.007

expertise (12–15). Therefore, it is not evident that learners' perceptual performance will improve simply by using the search strategies of experts. However, some perceptual tasks, such as finding lung nodules on chest CT scans, do not depend on a large knowledge base, and therefore teaching a search strategy may improve detection. Experimental studies may be beneficial to determine if search patterns can be taught to junior observers, and if this can improve perceptual performance.

The aim of this research study is twofold: (1) to investigate if drilling and scanning search strategies can be taught to junior radiology trainees, and (2) to compare the effect of teaching each search strategy on trainee's perception accuracy of lung nodule detection. We hypothesized that junior radiology trainees could adopt a new search strategy after instruction and that the use of a drilling strategy would improve the trainees' perceptual performance compared to a scanning strategy.

MATERIALS AND METHODS

Design

An experimental study was conducted to compare the effect of two teaching methods on perceptual performance. A randomized crossover design was chosen to adjust for individual variation in performance, differences in search strategies prior to any search strategy instruction, and possible differences in search behavior due to the sequence of the search strategy instructions. The design is illustrated in Figure 1.

Study Population and Procedure

Over a 3-month period, 19 (70%) first and second-year radiology trainees of a US academic medical center's radiology residency program enrolled in the study. Participants were randomly divided into two groups: both groups first watched an instructional video that provided lung nodule definitions. For the purpose of the study, a pulmonary nodule was defined as a solid opacity with a diameter greater than or equal to 3.0 mm. Ground glass or calcified nodules were not included. The instructional video showed examples of true nodules and also addressed other pulmonary abnormalities that were not considered nodules, such as consolidations, linear densities, pleural irregularities, and apical scarring, all illustrated by examples. Study participants then completed a pretest using free search (Test 1).

After the free search test (Test 1), group A started with the drilling instruction video, followed by Test 2, the scanning instruction video, and Test 3. Group B received the scanning and drilling instruction in opposite order. The drilling instructions explained the drilling search strategy: mentally dividing each lung into three regions (anterior, middle and posterior) and scrolling through each region individually, while keeping the eyes fixated in that region. The scanning instruction explained the scanning search strategy: reviewing all visible lung parenchyma at once (both sides), while slowly scrolling down, image by image.

In all three tests, participants were asked to mark as many true pulmonary nodules as possible, while avoiding marking any foci not meeting the study's definition of a true nodule. There was a time limit of 4 minutes per case.

The digital assessment program VQuest (www.vquest.eu) was used for the viewing and marking of lung nodules. This program is designed to deliver tests containing volumetric images, and allows for registering all scroll movements and mouse clicks. During the tests, participants could scroll through the stack of images, zoom in or out, adjust contrast settings, and measure findings. All stacks were viewed in axial plane. To select a lung nodule, participants were instructed to place a marker by clicking in the center of the nodule.

Tests

Tests 1, 2, and 3 were unique tests, each containing seven volumetric pulmonary CT scans. In total, each test consisted of 31 true nodules spread out over the seven CT scans. Nodules were 3 to 6 mm, with an average of 4 mm. The scans were retrieved from the picture archiving and communication system of the institution and were reviewed by two experienced radiologists (with 10 and 5 years of experience). Disagreement was resolved in consensus format. The selected chest CT scans had, on average, 349 slices, and slice thickness was 1.25 mm in all cases. The tests were made as equivalent as possible, by means of a test blueprint (Table 1). Each test was similar with regard to total number of nodules, the size of the nodules, the distribution of cases with fewer and more nodules, and the distribution of easy versus difficult cases. Nodules attached to vessels, bronchi, mediastinal structures, or diaphragm were considered difficult, whereas all other locations were considered easy.

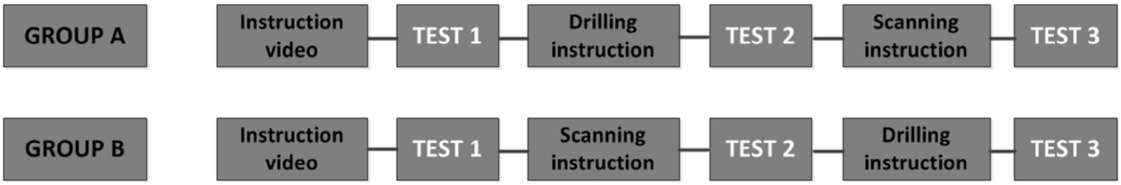

Figure 1. Study design.

TABLE 1. Test Blueprints with Number, Size, and Location of Nodules

	Number Per Case			Size			Location*	
	0–2	3–6	>6	3–4 mm	5–6 mm	Average Size (mm)	Easy	Difficult
Test 1	2	3	2	23	8	4.0	25	6
Test 2	2	3	2	23	8	4.1	25	6
Test 3	2	3	2	23	8	4.0	25	6

* Difficult: attached to vessels, bronchi, mediastinal structures, or diaphragm; Easy: anywhere else.

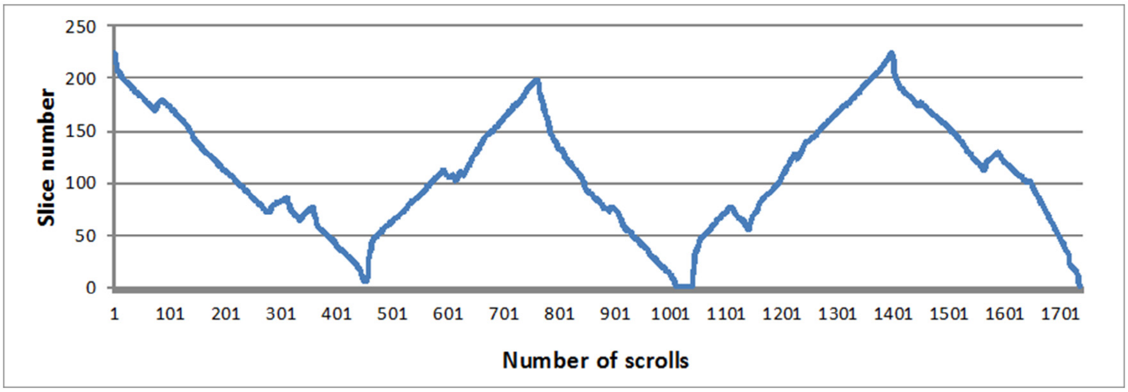

Figure 2. Example of a scroll pattern with five runs.

Questionnaire

Participants completed a short questionnaire to evaluate the study setup and to determine how participants searched for nodules in case of free search and how they perceived the two instructed search strategies. Response formats were a five-point Likert scale (5 = strongly agree, 4 = agree, 3 = undecided, 2 = disagree, 1 = strongly disagree), dichotomous (yes or no), or open. The survey questions are listed in Appendix S1.

Analysis

Reliability of the true positive (TP) scores was estimated with Cronbach's alpha, using the true nodules as items. Average *P* values of the items were calculated to estimate test difficulty.

Two main outcomes were measured, related to the two research questions:

1. Scroll behavior: for each case, the number of runs was determined. A run is defined by Venjakob et al. as a scroll movement in one direction that covers at least 50% of the total scan length, ignoring interruptions by any smaller scroll movements in the other direction (16). Figure 2 presents an example of a scroll pattern with five runs, despite some interrupting smaller movements. The total scrolling time was also calculated for each case.

2. Perceptual performance: the average numbers of TP and false positives (FPs) of Tests 1, 2, and 3 were calculated for both groups. We were primarily interested in differences in TPs. FP numbers were also measured, given that an increase in TPs is sometimes achieved at the cost of an increase in FPs.

For both outcome measures, a mixed-design analysis of variance was used for the statistical analysis. The instructed search strategy (no scroll instruction, drilling instruction, or scanning instruction) was the within-subject factor. Between-subject factors were study group (A or B) and year of training (first or second year of residency). Paired *t* tests were used for post hoc testing in normal distributed data and Wilcoxon signed-rank tests in non-normal distributed data.

Institutional Review Board Approval

The study was approved by the Institutional Review Board of the institution. Participation was voluntary and all participants gave written informed consent. This study did not involve protected health information.

RESULTS

Participants

After randomization, 9 participants were assigned to group A and 10 to group B, of which 4 in each group were second-year residents.

Test Performance

Reliabilities of the tests were acceptable: Cronbach's α were 0.74, 0.82, and 0.84 for Tests 1, 2, and 3, respectively. Average *P* values were 0.49, 0.39 and 0.48, respectively.

TABLE 2. Scroll Behavior and Perceptual Performance Measures Per Search Strategy Condition

	Free Search	With Drilling Instruction	With Scanning Instruction
Scroll behavior			
Number of runs per case, Mdn (IQR)	10 (4)	10 (4)	2 (3)
Scrolling time per case in seconds, Mdn (IQR)	219.6 (35.9)	208.0 (39.3)	167.0 (96.6)
Perceptual performance			
True positives, M (SD)	15.3 (4.6)	16.3 (5.3)	10.7 (5.0)
False positives, M (SD)	12.5 (7.8)	7.3 (5.6)	5.6 (4.8)

IQR, interquartile range; M, mean; Mdn, median; SD, standard deviation.

TABLE 3. Mixed-Design ANOVA for Scroll Behavior Outcomes

	Number of Runs		Time	
	F	*P*	*F*	*P*
Between subjects				
Year of residency	1.0	0.33	0.4	0.55
Study group	3.5	0.08	0.7	0.41
Within subjects				
Search strategy	54.2	<0.001	10.5	<0.001
Search strategy × Year of residency	0.1	0.88	0.7	0.49
Search strategy × Study group	28.3	0.74	0.9	0.43
Search strategy × Year of residency × Study group	0.4	0.37	1.8	0.18

ANOVA, analysis of variance.
Study group: intervention group A or B; search strategy: free search, drilling instruction, or scanning instruction; year of residency: first or second year.

TABLE 4. Mixed-Design ANOVA for Perceptual Performance Outcomes

	TP		FP	
	F	*P*	*F*	*P*
Between subjects				
Year of residency	9.0	<0.01	1.4	0.26
Study group	0.12	0.74	0.04	0.85
Within subjects				
Search strategy	16.1	<0.001	15.3	<0.001
Search strategy × Year of residency	1.5	0.85	0.3	0.66
Search strategy × Study group	28.3	0.07	0.9	0.84
Search strategy × Year of residency × Study group	0.4	0.96	0.9	0.38

ANOVA, analysis of variance; FP, false positives; TP, true positives.
Study group: intervention group A or B; search strategy: free search, drilling instruction, or scanning instruction; year of residency: first or second year.

Scroll Behavior

Median number of runs and scrolling time for each search strategy are provided in Table 2.

There was no significant effect of study group or year of residency on the number of runs or time. There was a significant within-subject effect of search strategy on number of runs, $F(1.3) = 54.2$, $P < 0.001$, and on time, $F(2) = 10.5$, $P < 0.001$, indicating that the instructed search strategy significantly affected the amount of drilling and the time required to complete the task. There were no significant interaction effects for both outcomes (Table 3).

The number of runs was significantly lower after participants were instructed to scan (Mdn = 2, interquartile range [IQR] = 3), than after drilling instruction (Mdn = 10, IQR = 4), $z = -3.8$, $P < 0.0$, or when allowed to search as desired (Mdn = 10, IQR = 4), $z = -3.8$, $P < 0.001$. The participants' free search and drilling instructed search had similar amounts of drilling.

The scanning search required significantly less time compared to the free search (Mdn = 167.0, IQR = 96.6 vs Mdn = 219.6, IQR = 35.9), $z = -3.6$, $P < 0.001$, although the drilling search (Mdn = 208.0, IQR = 39.3) was not found to be significantly different from scanning, $z = -1.4$, $P = 0.147$, or free search, $z = -1.9$, $P = 0.053$.

Perceptual Performance

Mean numbers of TPs and FPs for each search strategy condition are given in Table 2. There was a significant effect of year of residency on TP score, $F(1) = 9.0$, $P < 0.01$, although no significant effect on FP score (Table 4). Within-subject analysis shows a significant effect of instructed search strategy on the number of TPs, $F(2) = 16.1$, $P < 0.001$, and FPs, $F(1.3) = 15.3$, $P < 0.001$. Search strategy accounted for 51.8% of the variability in the number of TPs and 50.1% of the variability in the number of FPs. There were no significant interaction effects with the two between-subject factors, study group and year of residency, for both performance outcomes.

Pairwise analysis showed that drilling (M = 16.3, standard deviation [SD] = 5.3) and free search (M = 15.3, SD = 4.6) both resulted in significantly higher TP scores than scanning (M = 10.7, SD = 5.0); $t(18) = 4.78$, $P < 0.001$, and $t(18) = 4.44$, $P < 0.001$, respectively. TP scores of free search and drilling did not significantly differ.

For the FP scores, post hoc analysis revealed that free search ($M = 12.5$, $SD = 7.8$) resulted in significantly higher FP rates than drilling ($M = 7.3$, $SD = 5.6$), $t(18) = 4.86$, $P < 0.001$, and scanning ($M = 5.6$, SD 4.8), $t(18) = 4.44$, $P < 0.001$.

Questionnaire

The vast majority of participants (89%) agreed or totally agreed that reviewing chest CT scans in this research study using VQuest was substantially similar to clinical practice. Both instructional videos were found to provide sufficient instruction on using the search strategies by 95% of the participants.

After completing the free search test, all participants reported that they were aware of their search strategy when reading chest CT scans. Seventeen participants specified their search strategy and all were variants of drilling: 53% reported dividing the lungs into two, three, or four regions, and searched each region individually; the remainder searched lobe by lobe. In addition, seven participants (37%) built in an extra check of the edges of the lungs at the pleura, fissures, mediastinum, and hilar structures.

The drilling search strategy was not new for most participants—74% had heard about it whereas only 11% was familiar with the scanning search strategy. Participants felt that drilling led to a higher detection rate than scanning ($M = 3.7$ vs 1.8, $z = 3.8$, $P < 0.001$), and found drilling to be more time efficient ($M = 4.2$ vs 1.8, $z = 3.6$, $P < 0.001$) and easier to use continuously ($M = 4.3$ vs 1.9, $z = 3.9$, $P < 0.001$). Of all participants, 58% planned to use drilling as their new search strategy, which reportedly was different or somewhat different from the strategy they used before the study. The remainder preferred to keep using their own search strategy.

DISCUSSION

Search strategy instruction had a significant effect on both scroll behavior and perceptual performance. The scanning instruction decreased the number of long scroll movements and scrolling time. The drilling instruction did not alter scroll behavior significantly. However, the majority of participants reported already using some kind of drilling strategy at their free search. Perceptual performance following drilling search instructions outperformed performance following scanning search instructions in terms of TPs. Compared to a free search, the use of a drilling search strategy did not result in more TP findings. The study confirms our hypothesis that drilling outperforms scanning for detecting lung nodules, although drilling did not improve the baseline performance, probably explained by an a priori search strategy similar to drilling.

The benefit after drilling instruction may have been caused by initial habits of drilling and not by the instruction. Still, there was an improvement after instruction: it did reduce the number of FP findings, on top of an initial habit of a more productive drilling strategy at baseline. This improvement did not coincide with a decline in TP rate, whereas the scanning instruction reduced FP rate at the cost of TPs.

The decline in FP findings after instruction was an unexpected result, as the strategies were anticipated to improve actual detection rather than avoiding overcalls of non-nodule structures. The search strategy instruction may have induced a more focused attention and kept participants from being distracted by non-nodule structures. Another explanation may be that participants were more vigilant to mark any possible abnormality they found at the first test and became less attentive and less willing to spend extra time as the experiment proceeded, resulting in a drop of FPs in the second and third tests. FP findings have been associated with an increase in reading time, as previously found in breast cancer detection (17).

None of the participants reported the use of scanning in their free search, nor did anyone show a scroll pattern that indicated scanning behavior. This differs from the study of Drew et al., reporting both "scanners" and "drillers" among radiologists in a lung nodule detection task, although scanners were in the minority (11). Interestingly, in that study the scanners were less experienced readers than the drillers, which would raise the expectation of finding even more scanners among the junior residents in our study. Possibly the use of thin section scans intrinsically evokes a drilling search, as scanning simply takes too much time. Another way to deal with large numbers of thin slices is the use of maximum intensity projections, which have been shown to improve lung nodule detection accuracy (18).

Our study has several limitations. First, the sample size of 19 participants is relatively small. However, each participant was asked to review multiple scans with multiple nodules, and indeed the study was powered to detect a difference as evidenced by our results. We do not expect that the main results would substantially change with a larger sample size, although subtle differences might have emerged in comparisons that were not statistically significantly different. Second, although we tried to equalize difficulty across the three tests by means of a test blueprint, Test 2 proved to be more difficult than the other two tests. The nonsignificant effect of the search strategy and study group interaction is probably due to this inequality in test difficulty that may have diminished the degree to which a participant could profit from the search strategy that was applied in the second test. This did not jeopardize our scanning versus drilling comparison, because the scanning and the drilling conditions were equally affected. However, it may have caused an underestimation of the effect of instruction on perceptual performance. Third, our study was specifically targeted to the detection of lung nodules (selected for a number of reasons); this means that we cannot generalize the results to other detection tasks. Finally, we only used scroll behavior and self-reports to estimate participants' search strategies, and did not apply eye tracking to further characterize visual attention. A higher drilling fraction does indicate more drilling, although we cannot further specify how participants divided the lung into regions and in which order.

Future research can be directed toward comparing scanning and drilling search strategies in other detection tasks, and also to exploring which type of drilling strategy would be optimal. For example, it would be helpful to compare drilling strategies dividing the lung into different numbers of regions, or to investigate if building in checks of the lung edges is beneficial.

CONCLUSION

Search strategy instruction can influence scroll behavior and perceptual performance of junior radiology residents completing a lung nodule detection task. In junior trainees, a drilling strategy yields a better perceptual performance than a scanning strategy. Teaching a scanning strategy further decreases the perceptual performance of junior radiology residents below their baseline performance. Teaching a drilling strategy for detecting lung nodules in chest CT scans is therefore preferable.

REFERENCES

1. Donald JJ, Barnard SA. Common patterns in 558 diagnostic radiology errors. J Med Imaging Radiat Oncol 2012; 56:173–178.
2. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978; 13:175–181.
3. van der Gijp A, Ravesloot CJ, Jarodzka H, et al. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ Theory Pract 2016; doi:10.1007/s10459-016-9698-1. Epub ahead of print.
4. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. Acad Radiol 1996; 3:137–144.
5. Cooper L, Gale A, Saada J, et al. The assessment of stroke multidimensional CT and MR imaging using eye movement analysis: does modality preference enhance observer performance? In: Manning DJ, Abbey CK, eds. Medical imaging 2010: image perception, observer performance, and technology assessment. 2010. doi:10.1117/12.843680.
6. Wood G, Knapp KM, Rock B, et al. Visual expertise in detecting and diagnosing skeletal fractures. Skeletal Radiol 2013; 42:165–172.
7. Alzubaidi M, Black JA, Jr, Patel A, et al., eds. Conscious vs. subconscious perception, as a function of radiological expertise. Proceedings—IEEE Symposium on Computer-Based Medical Systems; 2009.
8. Manning D, Ethell S, Donovan T, et al. How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography 2006; 12:134–142.
9. Kok EM, de Bruin ABH, Robben SGF, et al. Looking in the same manner but seeing it differently: bottom-up and expertise effects in radiology. Appl Cogn Psychol 2012; 26:854–862.
10. Hu CH, Kundel HL, Nodine CF, et al. Searching for bone fractures: a comparison with pulmonary nodule search. Acad Radiol 1994; 1:25–32.
11. Drew T, Le-Hoa Vo M, Olwal A, et al. Scanners and drillers: characterizing expert visual search through volumetric images. J Vis 2013; 13.
12. Baghdady MT, Pharoah MJ, Regehr G, et al. The role of basic sciences in diagnostic oral radiology. J Dent Educ 2009; 73:1187–1193.
13. Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. Acad Radiol 1999; 6:575–585.
14. Nodine CF, Kundel HL, Lauver SC, et al. Nature of expertise in searching mammograms for breast masses. Acad Radiol 1996; 3:1000–1006.
15. Elmore JG, Wells CK, Howard DH. Does diagnostic accuracy in mammography depend on radiologists' experience? J Womens Health 1998; 7:443–449.
16. Venjakob A, Marnitz T, Mahler J, et al. Radiologists' eye gaze when reading cranial CT images. Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment. 2012;8318.
17. Nodine CF, Mello-Thoms C, Kundel HL, et al. Time course of perception and decision making during mammographic interpretation. AJR Am J Roentgenol 2002; 179:917–923.
18. Valencia R, Denecke T, Lehmkuhl L, et al. Value of axial and coronal maximum intensity projection (MIP) images in the detection of pulmonary nodules by multislice spiral CT: comparison with axial 1-mm and 5-mm slices. Eur Radiol 2006; 16:325–332.

APPENDIX. SUPPLEMENTARY DATA

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.acra.2017.01.007.