Journal of Cancer Education (2019) 34:1213–1218
https://doi.org/10.1007/s13187-018-1431-8

Measuring the Effects of Education in Detecting Lung Cancer on Chest Radiographs: Utilization of a New Assessment Tool

Junghyun Kim[1] · Kwan Hyoung Kim[2]

Published online: 25 September 2018

Abstract

This study was designed to evaluate the effect of group and individualized educational lectures to accurately interpret chest radiographs of lung cancer patients and to introduce a new educational tool in evaluating skills for reading chest radiographs. Utilizing "hotspot" technology will be instrumental in measuring the effect of education in interpreting chest radiographs. There were 48 participants in the study. Chest radiographs of 100 lung cancer patients and 11 healthy patients taken at various time points were used for evaluation. Using "hotspot" technology, lesions on each radiograph were outlined. Values were taken at baseline, after which the group received lectures. Several days later, they underwent exam 2. Exam 3 was conducted after individualized lectures. A final exam was taken after the participants underwent individualized training within 2 months. Scores significantly improved after the individual lessons ($p < 0.001$). This improvement in performance decreased in the final examination. Statistically significant differences were observed between exam 2 vs. exam 3 and exam 3 vs. the final exam ($p < 0.001$, $p < 0.001$). Participants demonstrated more improvement in detecting lesions in abnormal chest radiographs than in identifying normal ones. Although there was significant improvement in detecting abnormal radiographs by the end of the study ($p < 0.001$), no improvement was observed in detecting normal ones. We measured lung cancer detection rate using a new "hotspot" detection tool for chest radiographs. With the proposed scoring system, this tool could be objectively used in evaluating the educational effects.

Keywords Lung · Education · Radiograph · Hotspot

Introduction

Chest radiographs are among the most frequently used forms of medical imaging worldwide. In the UK alone, 7.7 million chest radiographs were recorded in 2012 and 7.4 million in 2013 [1]. Chest radiographs are used as a screening tool for a plethora of chest abnormalities. Although it is difficult to determine how many chest radiographs were ordered for the purpose of lung cancer diagnosis, it is indisputable that chest radiographs present an abundant source of information on lung cancers.

✉ Kwan Hyoung Kim
kwan-kim@catholic.ac.kr

1 Veterans Health Service Medical Center, Seoul, Republic of Korea

2 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

However, because chest radiographs have a low sensitivity, it is important to develop an adequate, organized approach to reading chest radiographs that ensures optimal accuracy. To this end, there have been efforts to utilize computers to increase the accuracy of detection of chest abnormalities [2–6]. However, even with the development of computational tools, the ability of the medical staff to interpret chest radiographs is crucial. The latest guidelines from the US Preventive Services Task Force recommend annual screening with low-dose computed tomography in adults aged 55–80 years, if they have a 30 pack-year smoking history within the past 15 years [7]. Although computed tomography is used as the mainstay method for lung cancer screening in symptomatic patients, it is accompanied by large costs for manpower, infrastructure, and resources. In addition, there is the limitation of eligible patient populations to consider. With evidence suggesting that chest radiographs may allow early detection of lung cancer [8–11] and the high volume of chest radiographs being obtained at present, it is important to examine whether physicians can be better educated to detect malignancies when reading chest radiographs. There is a wide variability in chest radiograph reading capabilities across specialties [12–14]. As many clinical decisions are affected by information gained from chest radiographs, improper interpretation could have detrimental effects on patient care [15]. Multiple studies have demonstrated that inaccurate interpretations can cause misdiagnoses and change disease management. Therefore, training and education are important for all resident physicians regardless of their specialties.

However, to date, there is no well-established educational method for reading chest radiographs. Most of the education currently relies on on-site experience. In addition, much of radiology has shifted its focus toward other imaging modalities.

To improve competency in the interpretation of a chest X-ray of lung cancer, we developed a new evaluation tool for measuring performance in malignancy detection utilizing hotspot technology. We hypothesized that combined with a novel scoring system, this tool will allow objective assessment of the effects of various educational methods for chest radiograph interpretation.

This study has two main objectives: first, to introduce a new evaluation tool and examine its application by teaching resident physicians across various specialties how to better interpret chest radiographs and, second, to examine the effects of education in detecting lung cancers by utilizing chest radiographs.

Methods

Participants

After obtaining approval from the XXX Hospital institutional review board (Study No. UC16EISI0076), 52 resident physicians who were part of the residency training program at the YYY University of XXX Hospital, a 761-bed facility, between March 2016 and February 2017, were randomly selected and consented to be recruited for the study. Physicians were gathered from across 11 specialties and had varying levels of training. This was a convenience sample, and no formal sample size analysis was conducted.

Study Set

One hundred and eleven plain digital chest radiographs from 100 lung cancer patients confirmed by pathology and 11 non-cancer healthy controls were selected from 1500 radiographs taken at XXX Hospital between 2005 and 2015. Two researchers selected radiographs with various lesion sites and sizes to allow for comprehensive assessment of performance. All radiographs were obtained before or at diagnosis.

We used a modified hotspot method using Adobe Captivate 9 (Adobe Systems Inc., San Jose, CA, USA). By default, a hotspot is visible and is rectangular. The shapes were then changed to free form and turned into custom hotspots that outlined the lung lesions. We adjusted the borders of the hotspots so that they were not visible by setting the line width to 0 and changing the opacity of the outlined area to 0%. The radiographs were then compiled into a 111-question examination, in which malignant areas were marked with transparent outlines. Two board-certified radiologists and one board-certified pulmonologist, each with 33 years of experience, allocated a score to each radiograph, depending on the lesion size, anatomical location, and contrast to adjacent structures (Table 1). A maximum of 10 points was assigned to each question, providing a total of 761 points. To assure reliable point allocation, points were independently assigned to 10 sample radiographs by each investigator and then compared, with an inter-rater reliability of 0.93. The hotspot program recorded their answers on spreadsheets.

Examinations

During the examinations, participants were asked to examine the radiographs and use the mouse cursor to click the region that they believed indicated a malignant lesion. Participants were informed that some of the chest radiographs were normal. Participants were awarded a full score for each abnormal lesion correctly identified or for correctly identifying normal radiographs. If participants clicked on a region outside the outlined area, a score of 0 was given for the question. No partial scores were awarded.

Education Program

Exam 1 was provided to establish a baseline. Within 1 month, participants received a group lecture on how to read and detect malignancies on chest radiographs. Several days later, the participants took exam 2. Each participant was then provided with an individualized lecture within 1 month of the group lecture. Exam 3 was then taken several days after the individualized lectures. Participants were provided the final exams within 2 months after their individualized lectures. Although the same 111 radiographs were used in all exams, the order of presentation was scrambled so that participants could not memorize the answers from previous examinations.

In the group lectures, the lecturer, a board-certified pulmonologist specializing in critical care with 33 years of experience, trained groups of 6-10 resident physicians in a classroom setting. The lecturer used Microsoft PowerPoint slides to present material on the basics of chest radiography, chest radiological anatomy, signs, and tips on chest radiograph interpretation. Exercise questions were provided with chest radiographs of patients with various cell types and stages of lung

Table 1 Point allocation criteria

Points allocated	0	1	2	3	4
Size	N/A	> 5 cm	3–5 cm	< 3 cm	N/A
Location	N/A	Open area	Located on the border of structure	Hidden by adjacent structure	N/A
Contrast to adjacent structure	Very strong contrast or prominent change of contour	Moderate contrast	Mild contrast	Minimal contrast	Faint contrast

cancer. The lecture and the exercise were conducted in a total of two 1-hr sessions. The lecturer allowed the students to briefly study the chest radiographs before providing them with the answer. Although resident physicians have various levels of proficiency, this aspect was not considered due to time constraints and because some resident physicians were losing focus or falling asleep.

In individualized lectures, each resident physician was provided with the same lecture content using computer-based material but was provided more time to independently find the lesions on the chest radiographs. The lecturer provided tips and exercises on how to improve their reading skills. These lectures were provided in two 30–60-min sessions. Participants in this method were noted to have better attention spans.

The group and individualized lectures were presented by one identical lecturer, a medical pulmonologist with 33 years of clinical experience in the field and 31 years of experience in educating resident physicians.

Statistical Analysis

Performances in each examination were compared to identify any statistically significant improvements in lesion detection. Continuous data were reported as means ± standard deviations and categorical data as a number and percentage. The scores from each test were compared according to participant specialties. A p value of < 0.05 was considered statistically significant. A t test was conducted. The software IBM SPSS Statistics 24.0 was used for statistical analysis [16].

Results

Four participants were excluded from the study since they were not able to attend all the lectures and/or examinations.

Improvements in Lesion Detection After Education

Forty-eight residents were recruited for the study. Participants scored 35.76 ± 10.25% at baseline. There were no statistically significant improvements between the baseline exam and the exam taken after the group lecture. Exam scores improved significantly to 79.86 ± 10.12% after the individualized lectures ($p < 0.001$). However, this improvement in performance decreased in the final exam (54.55 ± 9.39%). Statistically significant differences were also observed when comparing exam 2 and exam 3 ($p < 0.001$), exam 2 and the final exam ($p < 0.001$), and exam 3 and the final exam ($p < 0.001$).

Comparison of Lesion Detection in Abnormal vs. Normal Chest Radiographs

Participants showed more improvements in detecting lesions in abnormal chest radiographs than in detecting normal ones (Fig. 1). At baseline, participants achieved a score of 32.12 ± 10.18%. This gradually improved to a peak of 81.25 ± 10.22% by exam 3 and 54.35 ± 9.61% by the final exam. Detection of normal radiographs was 57.33 ± 29.94% at baseline. However, performance significantly decreased to 23.48 ± 28.6% in exam 2 ($p < 0.001$). This then improved to 71.56 ± 21.75% by exam 3 and 55.72 ± 22.31% by the final exam. Although there was a statistically significant improvement in detecting abnormal chest radiographs by the end of the study ($p < 0.001$), there was no significant improvement in detecting normal chest radiographs.

Comparison Between Departments

Resident physicians from the internal medicine department did not show a significantly different performance in reading normal or abnormal chest radiographs when compared to residents from other departments (Table 2).

Comparison Between Levels of Training

There were no significant differences in performance between residents depending on the levels of training.

Discussion

We observed an overall significant improvement in performance after the individual lectures. The decline in performance after 2 months, despite continued clinical practice in the intervention period, suggests the need for periodic

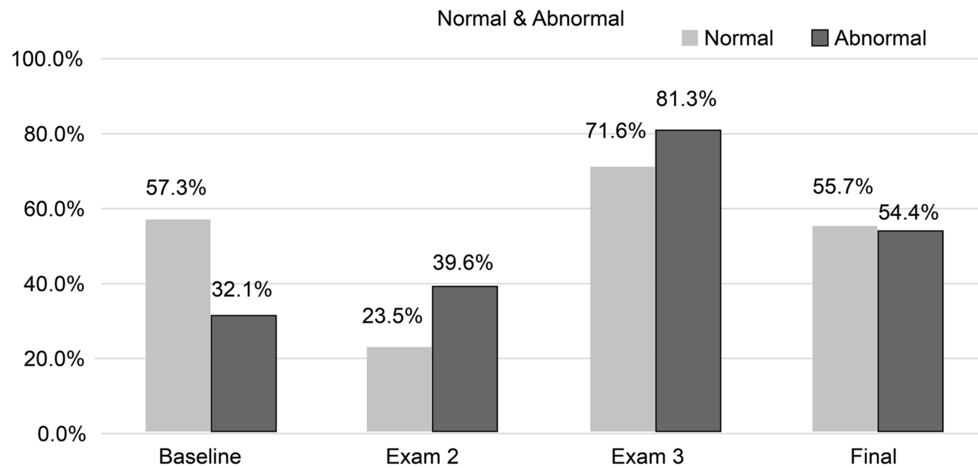

Fig. 1 Examination score percentages (normal and abnormal)

education. Although the magnitude of the benefit is expected to differ depending on the education method and target, Keijzers et al. report that a single lecture did not prove to be effective [17]. Further research is needed to determine whether frequent repeated education will be required to achieve a certain level of performance.

The lack of comparable improvement after the group lectures may be attributable to several factors. The structure of the group lectures limits the ability of the lecturer to cater to each student's needs. In addition, it creates a more passive learning environment, which may have contributed to the relative lack of attention from the resident physicians.

It is worth noting that there was no significant difference in improvement between internal medicine (IM) residents and non-IM residents. This suggests that with adequate education, physicians will be able to effectively detect lung lesions, regardless of their specialty.

The significance of this study is that it introduces a new evaluation tool for the education of chest radiograph interpretation, with promising possibilities for application in multiple medical specialties. Study participants scored an average of 36.72% at baseline, which showed peak improvement at 82.01% after individual lessons, and ultimately improved to 56.02% by the end of the study when compared with the baseline performance. This demonstrated a significant positive impact on the detection of abnormalities using chest radiographs. This study suggests that with adequate physician training, lung malignancy detection using chest radiographs can substantially improve.

Education on reading chest radiographs is conducted in the emergency department, in radiology, or in case presentations and lectures that involve chest radiographs. The differences in baseline performance observed between residents of different specialties are likely attributable to different educational methods used in different departments [18]. Van Geel et al. suggested that systematic evaluation of chest radiographs should be prioritized over non-systematic evaluation [19]. However, Chen et al. suggested that education resembling game play is the most effective [20]. This study used both systematic and non-systematic evaluation approaches.

We did not perform a direct comparison between the effects of group lectures vs. individualized lectures. However, since there was no significant improvement after the group lecture, it is likely the educational effect observed was due mainly to individual lectures. It is interesting to note that Rozenshtein et al. reported that a massed educational

Table 2 Comparison of exam scores between internal medicine and other specialties

Internal medicine ($n = 12$)					
Baseline score (mean ± SD)	Exam 2 score (mean ± SD)		Exam 3 score (mean ± SD)	Final score (mean ± SD)	
300.67 ± 87.16	267.67 ± 87.05		608.09 ± 85.15	417.60 ± 66.90	
Baseline vs. exam 2	Baseline vs. exam 3	Baseline vs. final exam	Exam 2 vs. exam 3	Exam 2 vs. final exam	Exam 3 vs. final exam
$t = 1.588$ $p = 0.151$	$t = 9.348$***	$t = 5.660$***	$t = 13.038$***	$t = 7.804$***	$t = 6.652$***
Others ($n = 36$)					
Baseline score (mean ± SD)	Exam 2 score (mean ± SD)		Exam 3 score (mean ± SD)	Final score (mean ± SD)	
262.63 ± 74.04	289.11 ± 62.30		607.58 ± 75.70	414.40 ± 73.65	
Baseline vs. exam 2	Baseline vs. exam 3	Baseline vs. final exam	Exam 2 vs. exam 3	Exam 2 vs. final exam	Exam 3 vs. final exam
$t = 2.464$*	$t = 18.653$***	$t = 7.964$***	$t = 17.923$***	$t = 7.618$***	$t = 12.469$***

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

method was more effective and longer lasting than an interleaved method [21]. It remains to be seen whether the improvement observed in the current study was due to the synergetic effects of both group and individualized lectures or whether it was mainly attributable to individualized lectures. Future comparative studies will help to identify the most efficient education method.

Furthermore, to the best of our knowledge, this study is the first to use the hotspot technology from Adobe Captivate as an educational tool. Utilization of the hotspot technology allows education to focus on teaching learners to pinpoint specific regions with anomalies. Unlike past educational methods used to detect nodes, this study outlined lesions of different shapes, comparing radiographs with computed tomography scans to outline areas hidden by overlapping structures. This technique has shown promise as a resource that can be used by medical professionals to improve and measure their image analysis skills. Although this study examined the utility of the technique for interpreting chest radiographs, its application can potentially extend to any imaging modality, including computed tomography, magnetic resonance imaging, and ultrasound.

There are several limitations to this study. First, the study was limited to resident physicians from a single hospital in the Republic of Korea. However, it is a pilot study and remains the largest of its kind. Second, due to scheduling issues, some residents had a larger time gap between the different stages. Although this discrepancy in the timings between participants was less than a week, it may have potentially affected performance, especially in the final examination. The third limitation is the lack of validation of the hotspot tool. However, this study was designed to be more of a case study setup to understand the implications of interventions, in terms of group or individualized lectures in the curriculum.

The significance of the utility of our results must be explored in more detail in future studies. The utility of chest radiographs in mass screening of at-risk populations needs to be further examined with larger samples across multiple sites. Additional information on the effectiveness of group and individualized lectures will help educators to design the most effective curriculum for chest radiograph interpretation. In addition, the hotspot technology must be compared to conventional education methods to determine its efficacy. In addition, examining the numerous applications of hotspot technology in other specialties and settings will help to uncover its full potential in medical education.

Conclusion

In summary, we were able to measure the detection rate of lung cancer from chest radiographs using a new tool comprising hotspot detection. Our data indicate optimal methods for educating residents to detect lung cancer from chest radiographs. Further studies are required to evaluate the reasons behind the differences between group and individual lessons, as well as to identify the most appropriate education programs for each department.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval After obtaining approval from the institutional review board, 52 resident physicians from Uijeongbu St. Mary's Hospital, Catholic University of Korea were recruited for the study. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. NHS England Analytical Services. Diagnostic imaging dataset annual statistical release 2013/14. https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2014/06/Annual-Statistical-Release-2013-14-DID-pdf-1118KB.pdf. Published November 06, 2014. Accessed May 14, 2018
2. Van Ginneken B, ter Haar Romeny BM, Viergever MA (2001) Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imaging 20:1228–1241
3. Shiraishi J, Abe H, Engelmann R, Aoyama M, MacMahon H, Doi K (2003) Computer-aided diagnosis to distinguish benign from malignant solitary pulmonary nodules on radiographs: ROC analysis of radiologists' performance—initial experience. Radiology 227:469–474
4. Chen JJ, White CS (2008) Use of CAD to evaluate lung cancer on chest radiography. J Thorac Imaging 23:93–96
5. Shiraishi J, Li Q, Appelbaum D, Doi K (2011) Computer-aided diagnosis and artificial intelligence in clinical imaging. Semin Nucl Med 41:449–462
6. Mazzone PJ, Obuchowski N, Phillips M, Risius B, Bazerbashi B, Meziane M (2013) Lung cancer screening with computer aided detection chest radiography: design and results of a randomized, controlled trial. PLoS One 8:e59650
7. National Comprehensive Cancer Network, Inc. *National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology.* Lung cancer screening. Version 2. 2016
8. Fontana R, Sanderson DR, Woolner LB et al (1991) Screening for lung cancer: a critique of the Mayo Lung Project. Cancer 67:1155–1164
9. Kubik A, Polak J (1986) Lung cancer detection: results of a randomized prospective study in Czechoslovakia. Cancer 57:2428–2437
10. Strauss GM, Gleason RE, Sugarbaker DJ (1997) Screening for lung cancer. Another look; a different view. Chest 111:754–768
11. Brett GZ (1969) Earlier diagnosis and survival in lung cancer. Br Med J 4:260–262

12. Monnier-Cholley L, Carrat F, Cholley BP, Tubiana JM, Arrivé L (2004) Detection of lung cancer on radiographs; receiver operating characteristic analyses of radiologists', pulmonologists', and anesthesiologists' performance. Radiology 233:799–805
13. Kuritzky L, Hardy RI, Curry RW (1987) Interpretation of chest roentgenograms by primary care physicians. South Med J 80: 1347–1351
14. Kaufman B, Dhar P, O'Neill D et al (2001) Chest radiograph interpretation skills of anesthesiologists. J Cardiothorac Vasc Anesth 15: 680–683
15. Grosvenor LJ, Verma R, O'Brien R et al (2003) Does reporting of plain chest radiographs affect the immediate management of patients admitted to a medical assessment unit? Clin Radiol 58:719–722
16. IBM Corp (2016) IBM SPSS statistics for windows, version 24.0. IBM Corp, Armonk, NY
17. Keijzers G, Sithirasenan V (2011) The effect of a chest imaging lecture on emergency department doctors' ability to interpret chest CT images: a randomized study. Eur J Emerg Med 19:40–45
18. Torre DM, Simpson D, Sebastian JL, Elnicki DM (2005) Learning/feedback activities and high-quality teaching: perceptions of third-year medical students during an inpatient rotation. Acad Med 80: 950–954
19. van Geel K, Kik EM, Dijkstra J, Robben SG, van Merrienboer JJ (2017) Teaching systematic viewing to final-year medical students improves systematicity but not coverage or detection or radiologic abnormalities. J Am Coll Radiol 14:235–241
20. Chen Y, James JJ, Dong L, Gale AG (2017) Measuring performance in the interpretation of chest radiographs: a pilot study. Clin Radiol 72:230–235
21. Rozenshtein A, Pearson GD, Yan SX, Liu AZ, Toy D (2016) Effect of massed versus interleaved teaching method on performance of students in radiology. J Am Coll Radiol 13:979–984