Clinical Radiology 74 (2019) 67–75

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Radiologist performance in the detection of lung cancer using CT

B. Al Mohammad [a,*], S.L. Hillis [b], W. Reed [a], M. Alakhras [a], P.C. Brennan [a]

[a] *Department of Medical Imaging and Radiation Sciences, Faculty of Health Sciences, The University of Sydney, Cumberland Campus, 75 East St, Lidcombe, NSW, 2141, Australia*

[b] *Departments of Radiology and Biostatistics, The University of Iowa, 3170 ML, 200 Hawkins Drive, Iowa City, IA, 52242-1077, USA*

ARTICLE INFORMATION

Article history:
Received 20 April 2018
Accepted 16 October 2018

AIM: To measure the level of radiologists' performance in lung cancer detection, and to explore radiologists' performance in cancer specialised and non-specialised centres.

MATERIALS AND METHODS: Thirty radiologists read 60 chest computed tomography (CT) examinations. Thirty cases had surgically or biopsy-proven lung cancer and 30 were cancer-free cases. The cancer cases were validated by four expert radiologists who located the malignant lung nodules. Reader performance was evaluated by calculating sensitivity, location sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC). In addition, sensitivity at fixed specificity (0.794) was computed from each reader's estimated ROC curve.

RESULTS: The radiologists had a mean sensitivity of 0.749, sensitivity at fixed specificity of 0.744, location sensitivity of 0.666, specificity of 0.81 and AUC of 0.846. Radiologists in the specialised and non-specialised cancer centres had the following (specialised, non-specialised) pairs of values: sensitivity=(0.80, 0.719); sensitivity for fixed 0.794 specificity=(0.752, 0.740); location sensitivity=(0.712, 0.637); specificity=(0.794, 0.82) and AUC=(0.846, 0.846).

CONCLUSION: The efficacy of radiologists was comparable to other studies. Furthermore, AUC outcomes were similar for specialised and non-specialised cancer centre radiologists, suggesting they have similar discriminatory ability and that the higher sensitivity and lower specificity for specialised-centre radiologists can be attributed to them being less conservative in interpreting case images.

Introduction

Lung cancer is considered the leading cause of cancer deaths worldwide; however, outcomes are significantly better if the cancer is detected at an early stage, with a 10 year survival of stage 1 lung cancer of up to 75%.[1] Lung cancer in Jordan is the second most common cancer among males, accounting for 16.2% of all new cancers; in comparison, lung cancer accounts for 9.6% and 13.6% of all new cancers in Australia and the USA, respectively. Mortality from lung cancer accounts for 21.6% of deaths from cancer in Jordan, compared to 20.2% in Australia and 28.3% in the USA.[2]

In 2011, the largest randomised controlled trial in lung cancer screening conducted in the USA, the National Lung

* Guarantor and correspondent: B. Al Mohammad, Department of Medical Imaging and Radiation Sciences, Faculty of Health Sciences, The University of Sydney, Cumberland Campus, 75 East St, Lidcombe, NSW, 2141, Australia. Tel.: +61 2 8627 1444.

E-mail address: baal8200@uni.sydney.edu.au (B. Al Mohammad).

https://doi.org/10.1016/j.crad.2018.10.008
0009-9260/

Screening Trial NLST,[3] demonstrated a significant 20% reduction in lung cancer deaths. This presented important evidence that annual computed tomography (CT) screening for lung cancer focusing on a high-risk population can prevent a substantial number of lung cancer-related deaths. After the findings of the NLST were published a number of major professional health authorities began recommending annual chest screening with CT for people at high risk of developing lung cancer.[4–6] As a consequence, the USA[7] and China[8] are currently implementing lung cancer screening programmes. Whilst other countries such as Australia and those in Europe have yet to establish a specific screening programme, the introduction of such programmes are expected within the next few years.[9,10] In Jordan, although there are early detection and screening programmes for other cancer types, such as breast cancer, there is currently no lung cancer screening. Furthermore, according to the US preventive services task force recommendations,[11] research needs to be conducted where smoking prevalence and lung cancer incidence is higher: among the socioeconomically disadvantaged populations. Thus if Jordan and other countries that do not have a lung cancer screening programme are to implement one, it is critically important to ensure that radiologist diagnostic performance is of a sufficiently high standard.

Previous literature has shown high variability in sensitivity for radiologists when detecting lung nodules in CT images, ranging from 0.30 to 0.97.[12–16] For example, in an investigation by Rubin *et al.*[14] of the performance of 13 radiologists in the detection of 5 mm nodules embedded randomly in 40 chest CT scans on a 1 mm section thickness reconstruction, sensitivity ranged from 0.30 to 0.73; however, most other studies that have described radiologist performance for detecting lung nodules have used a relatively low number of radiologists (2–5 radiologists), with a maximum of 16 radiologists.[12,14–19] In a study[20] that investigated lung nodule detection, the performance of two experienced radiologists was compared to two inexperienced radiologists in terms of sensitivity. The experienced radiologists with an average sensitivity of 0.54 outperformed the inexperienced radiologists with an average sensitivity of 0.26.

In another study, Roos *et al.*[16] evaluated the performance of three radiologists in lung nodule detection on 20 chest CT examinations, before and after the application of computer-aided detection (CAD) system. The detection sensitivities achieved by the readers without CAD were 0.59, 0.57, and 0.44, compared to sensitivities of 0.67, 0.82, and 0.60 with CAD. Among other limitations, the studies mentioned the need for a larger number of observers and cases. None of these studies include conclusions that generalise to both the reader and case populations, but instead treat either cases or readers as random, but not both.

Lung cancer detection is considered one of the most difficult radiological tasks, with approximately 300 axial images contained within a single CT examination of very complex lung parenchyma surrounding the nodules. With such previously reported variability in radiologist performance, it is essential that the diagnostic ability of radiologists in Jordan is established before the commencement of any screening programme, so that any necessary level and type of education and training can be considered in advance. The primary purpose of this study, therefore, is to measure the level of radiologists' performance in Jordan; a secondary purpose is to estimate the difference in radiologists' diagnostic efficacy between one specialised and several non-specialised cancer centres.

Materials and methods

This study investigated radiologist performance in lung cancer detection. Ethics approval from the University of Sydney was obtained. Informed consent from each participating radiologist was obtained before conducting this observer performance study. Permission from the hospital where radiologists performed the readings was provided.

Participants

The study included 30 radiologists with varying experience levels (median=7 years, minimum=2 years and maximum=30 years). There were 14 female and 15 male readers, ranging in age between 27 and 59 with a median value of 35 years. Demographics of the participating readers are presented in Table 1. Readers were recruited from different hospitals in Jordan to provide the range of expertise found in general radiology. Invitational emails were sent to the Radiology Department Chairs in the hospitals about the study, after which interested radiologists were contacted. A brief presentation was given to the interested radiologists summarising the aims, methods, and significance of the study. Information regarding each participant's

Table 1
Demographics of the participating radiologists.

Variable	All participants	Specialised cancer centre	Non-specialised cancer centre[a]
Age	35 (27, 56)	35 (35, 56)	35 (27, 56)
Gender	F=14 M=16	F=6 M=5	F=8 M=11
Years of experience	7 (2, 30)	12 (7, 30)	3 (2, 30)
Years reading chest CT scans	7 (2, 30)	7 (7, 30)	3 (2, 30)
Hours spent reading chest CT/week	15 (3, 28)	28 (3, 28)	11 (3, 28)
No. of scans read annually	1,500 (500, 6000)	1,500 (500, 4,500)	1,500 (500, 6000)

All variables except for gender are summarised using the format median (minimum, maximum).
F, female; M: male; CT, computed tomography.
[a] Non-specialised cancer centre readers are from the following hospitals: public (*n*=3), private (*n*=3), military (*n*=7), and university (*n*=6).

age, gender, and experience was collected using a questionnaire.

The participants in this study were divided into two subgroups: the first group consisted of 11 radiologists from a specialised cancer centre, and the second group included 19 radiologists from non-specialised cancer centres (general hospitals) which included six from a university, three from a private, seven from a military, and three from a public hospital. The specialised cancer centre[21] is an independent non-governmental comprehensive cancer care centre for cancer care, training, education, and research, and is fully accredited as an oncology centre equipped to diagnose and treat all types of cancer.

Image test set

All readers independently read the same test set of 60 chest CT cases (30 cancer cases and 30 cancer-free cases). The mean number of images per case was 230. All cases were randomised and de-identified using DicomSort software. The cancer cases were obtained from the LIDC-IDRI (Lung Image Database Consortium) image collection, which is publicly available from the Cancer Imaging Archive (TCIA).[22,23] The 30 cancer cases in the present study were chosen from 1,018 cases, which had been acquired using different scanner types, a tube voltage ranging from 120–140 kV, tube current range 40–627 mA, and reconstruction interval mean of 1.74 mm.[23] For the purpose of this study, only biopsy- or surgically-proven primary lung cancer cases, with four or fewer nodules per case were included. Locating the nodules was initially performed by four experienced thoracic radiologists in a two-phase image annotation process. The first phase was the blinded read phase where each radiologist independently reviewed the scans and located the nodules. In the following unblinded phase the radiologists independently reviewed their nodule markings in addition to the anonymised marks of the other three radiologists to make a decision, without forced consensus.[23] Cancer cases had varying numbers of nodules with one, two, three, and four nodules being present in 13, 11, three, and three cases respectively. The diameter of the nodules ranged from 3 to 30 mm, with 33 nodules between 3 and 10 mm, and 23 nodules >10 mm. Seven nodules that had a diameter <5 mm were all in scans that contained multiple nodules, in which there was at least one nodule >5 mm. This was in agreement with the British Thoracic Society (BTS) guidelines in which risk assessment of scans with multiple nodules <5 mm depend on that of the largest nodule.[24] There was also variation in the location of the nodules: 18 in the right upper lobe, three in the middle, 11 in the right lower lobe, 10 in the left upper lobe, and 14 in the left lower lobe of the lungs. Details are presented in Table 2.

Cancer-free cases were chosen and independently confirmed by two expert cardiothoracic radiologists with an average of 16 years thoracic CT experience. A number of the cancer-free cases (like the cancer cases) had benign pulmonary conditions such as pneumonia, pleural effusion, or benign nodules. The following criteria were used to exclude the cases in the non-cancer data set: large nodule size, nodules with suspicious morphology, such as lobulation or spiculation, irregular walls, multiplicity, inhomogeneous density, and high-risk patients. Nodules that were considered benign were solitary nodules, with smooth walls, rounded or oval in shape, with no lobulation or spiculation, with benign features, such as containing fat or calcification and were <6 mm in diameter. These benign nodules were only recorded with low-risk patients.[25]

Table 2
Number of malignant nodules according to size, location, and lung cancer type.

Nodule characteristics	No. of nodules
Nodule size	
3–10 mm	33
11–20 mm	18
21–30 mm	5
Nodule location	
Right upper lobe	18
Right middle lobe	3
Right lower lobe	11
Left upper lobe	10
Left lower lobe	14
Nodule (cancer) type	
Non-small cell lung cancer	33
Small cell lung cancer	23
Total number of nodules	56

Readings

Readings were performed on HP Z440 workstations (CPU Intel Processor, Memory Seagate RAM 8 GB, Storage Space 500 GB, Network Interface Broadcom) and the images were displayed on two calibrated, 19.5 inch, 3 (MP) megapixel, greyscale, display monitors available with each workstation. The scans were reviewed in a reading/reporting room 5×5.5 m in size, painted with a matt light blue colour. All reading conditions were kept constant for all the readings.

Readers were asked to identify and locate all perceived malignant nodules, and provide a score of confidence for each perceived nodule. Readers were instructed to give cases that appeared to be normal, i.e., did not have any nodules, a score of 1. A perceived benign nodule was to be given a score of 2, and a perceived malignant nodule was to be given a score ranging from 3 to 5, with a higher score indicating higher confidence of malignancy. The readers were requested to ignore abnormalities other than lung nodules or any other findings suspicious for lung cancer. It was not possible for radiologists to mark their findings on the displayed image because of the large number of images present in each case (an average of 230 images/case); therefore, the radiologists recorded their decisions using a paper-based method. Each radiologist was given full instructions before the readings commenced. The readers were asked to place the screen curser over each perceived nodule to identify its unique three-dimensional location, followed by filling in a table with the nodule location information and confidence score.

Fig 1 shows an example of the viewing software that the readers used to locate the nodules. The reader listed a

Figure 1 The information needed to locate lung nodules on the axial chest CT section. Arrow A is pointed towards a nodule, arrow B shows the section number and arrows C show the x and y coordinates of the centre of the nodule after the curser has been placed in the middle of the nodule.

section number where a malignancy was noted, as well as the x and y coordinates for that nodule and the confidence score for the nodule. By placing the curser in the centre of the nodule, the reader obtained the x and y coordinates. For example, for the nodule shown in Fig 1, the reader filled in the table section number 85 (as the nodule is visible in slices from 78 to 89, any section number within this range is considered correct). The same applies for the x and y coordinates; by placing the curser in the centre of the nodule, the reader obtained the x and y coordinates for that nodule, which were x=370 and y=368 for the example. Furthermore, as the nodule has visible edges in the x direction from 350 to 380 and visible edges in the y direction from 358 to 387, any x and y values within this range was considered correct. An additional 10 pixels margin of error was added in either direction for each nodule.

Radiologists were blinded to the scanning protocol and the patient clinical history. No information on cancer cases prevalence was given and radiologists were able to use all the post-processing tools typically available in a clinical setting. No time restrictions were imposed and readers could interpret images over several sessions if required.

Data analysis

Each radiologist's performance was evaluated using sensitivity, specificity, sensitivity at fixed specificity, location sensitivity, and area under the receiver operating characteristic (ROC) curve (AUC). These measures were defined as follows: (1) sensitivity: the proportion of cancer cases correctly identified, i.e., given a confidence score of 3 or above for any perceived nodule, regardless of whether the nodule was correctly located; (2) specificity: the proportion of cancer-free cases that were correctly identified as either normal (confidence score=1), or if any perceived nodules were identified as possibly benign (confidence score=2); (3) location sensitivity: the proportion of actually malignant nodules that were correctly identified (confidence score $\geq$3) and located by the radiologist; (4) AUC computed using the PROPROC[26–28] procedure. The ROC curve is a plot of sensitivity versus 1 − specificity. For a randomly selected pair of cases, one having cancer and one being cancer free, the AUC estimates the probability that the reader will correctly discriminate between the two cases.[28] The PROPROC procedure produces smooth semiparametric ROC curves that are proper, and thus have no undesirable "hooks", which can occur when ROC curves are estimated assuming a latent binormal model. The PROPROC method is preferred to the non-parametric (or empirical) method because the latter tends to underestimate the AUC for data such as the present, where the readers are forced to use only a few ordinal ratings.[29] The reader AUCs, their standard deviation (SD), and the 95% confidence interval (CI) for the difference of the population AUCs between the centres were calculated using the software package *OR-DBM MRMC 2.50*[30] which is freely available from http://perception.radiology.uiowa.edu/. Sensitivity at fixed specificity, with fixed specificity equal to 0.794; 0.794 was chosen because it is the mean specificity for readers from the specialised cancer centre. This measure is determined from each reader's estimated ROC curve. It allows us to compare the estimated performance of the readers in specialised and non-specialised cancer centres when they are operating at the same specificity.

The data were tested for normality using the Shapiro–Wilk test, and 95% CIs based on Student's t-test distribution were used to describe overall reader performance and to compare the specialised and non-specialised cancer centre readers. The two groups were considered to be significantly different if the 95% CI for the difference in values between the two groups did not include zero, which corresponds to a p-value <0.05 using the two-sample t test.

The two groups were also compared with respect to their summary ROC curves, where the summary ROC curve is defined as a plot of average sensitivity, computed across readers, for each possible fixed value of 1 − specificity. To illustrate the computation of the summary curve, consider 1 − specificity=0.10. For each reader, the corresponding sensitivity was computed from the estimated ROC curve and the average computed. The AUC for each summary ROC curve is the same as the mean AUC (See Chen and Samuelson[31] for a description of this method of computing a summary ROC curve).

Results

Individual reader results

The individual reader estimates for each of the outcomes are shown in Fig 2, grouped according to centre (specialised compared to non-specialised). Fig 2 reveals that one reader was an extreme outlier with respect to specificity, having a value of 0.23, whereas all other readers had values exceeding 0.67. Thus this reader was dropped from all subsequent analyses.

Overall performance

The mean and corresponding 95% CIs, SD, and minimum and maximum values for the sensitivity, sensitivity (with

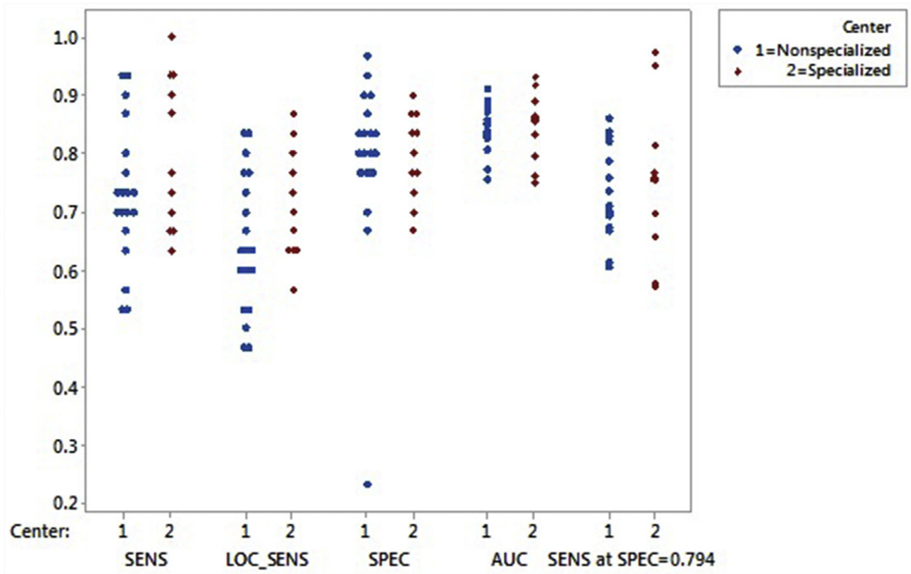

Figure 2 Individual reader performance outcomes by centre (specialised compared to non-specialised). The reader with specificity equal to 0.23 was dropped from all subsequent analyses. Notes: SENS=sensitivity; LOC_SENS=location sensitivity; SPEC=specificity; AUC=area under the ROC curve; SENS at SPEC=.0794 is the estimated sensitivity for specificity=0.794, computed from the estimated ROC curve. For the last two outcomes, the ROC curve was estimated using the PROPROC method.

fixed specificity=0.794), location sensitivity, specificity and AUC, calculated across readers, are presented in Table 3. The means are as follows: sensitivity=0.749, sensitivity (with fixed specificity=0.794)=0.744, location sensitivity=0.666, specificity=0.81 and AUC=0.846.

Table 4 is similar to Table 3, except that results are given for each of the two subgroups of readers and for their difference. The means for the specialised cancer-centre readers are as follows: sensitivity=0.80, sensitivity (with fixed specificity=0.794)=0.752, location sensitivity=0.712, specificity=0.794 and AUC=0.846. The means for the non-specialised cancer centre readers are as follows: sensitivity=0.719, sensitivity (with fixed specificity=0.794)=0.74, location sensitivity=0.637, specificity=0.82 and AUC=0.846.

Although there was no statistically significant difference between the two groups of readers in terms of sensitivity (p=0.0826), location sensitivity (p=0.0755), sensitivity at fixed specificity (p=0.7496), specificity (p=0.3698) and AUC (p=1), the 95% CI of some of the performance parameters, such as sensitivity (-0.012, 0.175), location sensitivity (-0.009, 0.159), and specificity (-0.086, 0.033) were too wide to come to the conclusion that the groups had similar performance.

The individual reader ROC curves are displayed in Fig 3 and the summary ROC curves for the two groups, as well as the observed (1 − specificity, sensitivity) values, are displayed in Fig 4. As previously noted, each summary ROC curve is a plot of average sensitivity, computed across readers, for each possible fixed value of 1 − specificity. Fig 4 shows that the summary ROC curves are quite similar and that both of the observed (1 − specificity, sensitivity) values are very close to their corresponding summary curves.

Table 3
Overall reader-performance summary statistics.

	Mean (95% CI)	SD	Minimum–maximum
Sensitivity	0.749 (0.702, 0.796)	0.124	0.533–1.0
Sensitivity at Specificity=0.794	0.744 (0.707, 0.781)	0.097	0.572–0.973
Location Sensitivity	0.666 (0.623, 0.708)	0.111	0.467–0.867
Specificity	0.81 (0.782, 0.839)	0.075	0.667–0.967
AUC	0.846 (0.828, 0.863)	0.046	0.749–0.931

The numbers in parentheses are the 95% confidence-interval (CI) lower and upper bounds. Sensitivity at specificity=0.794 was computed from each reader's estimated ROC curve, with 0.794 chosen because it is the mean specificity for readers from the specialised cancer centre.
SD, standard deviation; AUC, area under the curve.

Discussion

The present study investigated the lung cancer detection performance of radiologists examining CT chest examinations in Jordan and involved a higher number of participating radiologists compared to similar previous studies[13,20,32–35] performed in other countries. Moreover, radiologist' performance in the present study is evaluated and described in a detailed and comprehensive manner, reporting sensitivity, location sensitivity, specificity, sensitivity at a fixed specificity, and AUC. Sensitivity in the present study (mean of 0.749) was higher than in most other studies,[13,14,16,19,20,32,33,36] and comparable to others.[12,37] The AUC for readers in the present study (mean of 0.846) was similar to other studies.[13,20,32–34]

The performance of radiologists in specialised cancer centres and non-specialised cancer centres was also investigated. Although the preliminary analysis did not show significant differences between the two groups, 95% CIs

Table 4
Reader-performance summary statistics by centre.

Parameters	Specialised centre Mean (95% CI)	SD	Non-specialised centre Mean (95% CI)	SD	95% CI of the difference
Sensitivity	0.80 (0.713, 0.887)	0.13	0.719 (0.663, 0.774)	0.112	(-0.012, 0.175)
Sensitivity (at 0.794 specificity)	0.752 (0.665, 0.839)	0.129	0.740 (0.702, 0.777)	0.076	(-0.065, 0.09)
Location sensitivity	0.712 (0.648, 0.776)	0.096	0.637 (0.581, 0.693)	0.113	(-0.009, 0.159)
Specificity	0.794 (0.744, 0.844)	0.074	0.82 (0.783, 0.858)	0.076	(-0.086, 0.033)
AUC	0.846 (0.807, 0.885)	0.058	0.846 (0.826, 0.865)	0.039	(-0.036, 0.037)

The numbers in parentheses are the 95% CI lower and upper bounds.
Sensitivity at specificity=0.794 was computed from each reader's estimated ROC curve, with 0.794 chosen because it is the mean specificity for readers from the specialised cancer centre.
CI, confidence interval; SD, standard deviation; AUC, area under the curve.

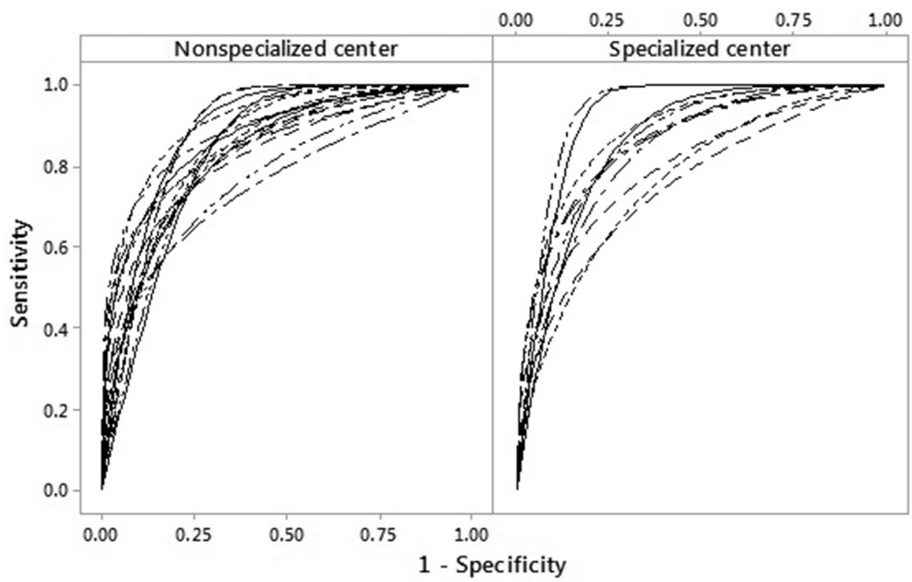

Figure 3 Individual reader ROC curves by centre.

were too wide to conclude that the groups were clinically similar with respect to sensitivity, location sensitivity, and specificity. Mean sensitivity was 8 percentage points higher in the specialised centre (0.80 compared to 0.719), with most of the confidence interval (-0.012, 0.175) for the difference (specialised minus non-specialised) being to the right of zero; these results suggest that a significant difference might be found with more precise estimates from a larger study. In contrast, mean specificity was 2.6 percentage points higher for the non-specialised centre (0.794 compared to 0.82), with more of the confidence interval (-0.086, 0.033) being to the left of zero.

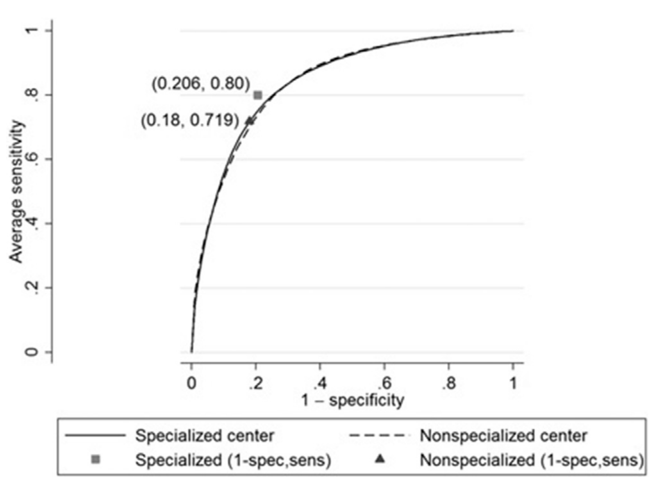

Figure 4 Summary ROC curves and observed (1−specificity, sensitivity) points by centre.

The mean AUC estimates were the same (0.846) for both subgroups with the confidence interval (-0.036 to 0.037) for the specialised minus non-specialised centre difference being narrower than for the other outcomes. Thus there is 95% confidence that the population mean AUC for specialised-centre radiologists is between 0.036 less and 0.037 more than for non-specialised-centre radiologists. Hence the present results indicate that the absolute difference in AUC for the two populations is not >0.037. These results demonstrate that the discriminatory ability of radiologists in the two groups is similar. One way to describe the similarity is that, on average, radiologists from the two groups differ by <0.037 probability of being able to distinguish between a randomly selected non-diseased case and a diseased case.

Mean sensitivity at fixed specificity (0.794) was slightly higher for the specialised centre: 0.752 versus 0.740, with the specialised minus non-specialised CI (-0.065, 0.09) for the centre difference being wider than that for the AUC. It can be concluded with 95% confidence that the population mean sensitivity for fixed specificity (fixed at 0.794) for specialised-centre radiologists is between 0.065 less and 0.09 more than for non-specialised-centre radiologists. Although there is not a statistically significant difference between the two outcomes (because the confidence interval contains zero) the width of the CI prevents the conclusion that the outcomes are relatively similar for the two groups.

Taken together, the present results based on the ROC curves – specifically, the closeness of the AUC, and sensitivity for fixed specificity (0.794) estimates, the similarity of the summary ROC curves, and the relatively small width for the CI for the AUC difference–suggest that that the two groups have similar discriminatory ability and that the observed higher sensitivity and lower specificity for specialised-centre radiologists can be predominantly attributed to specialised-centre radiologists being less conservative in interpreting case images. The advantage of comparisons based on the ROC curve is that ROC analysis disentangles the inherent discriminatory capacity of the reader from the particular level of conservatism at which the reader is operating.[38]

It is interesting to note that whilst hours of reading per week was much higher in the specialised centre, the number of cases per year were similar. It is difficult to understand fully why this discrepancy exists at this time, but possible explanations are that in the specialised centre either the reading time per case is higher or that the radiologists in the non-specialised centre do not read lung cancer cases every week. This will require further investigation. Readers' characteristics and practices, such as years of experience, the number of chest CT examinations read, and the time spent reading the scans, in addition to nodule features, such as size and location, may affect the detectability of the nodules.[20,39–41] This will be further investigated in a future study as these features are critically important when trying to optimise nodule detection.

Differences in methodologies used and metrics measured make comparisons between studies difficult. In previous literature, radiologist performance has been described in a number of ways: some only reported sensitivity,[14–16,37] while others focused on AUC[13,20,32,33] and specificity.[15] Furthermore, the reported performances varied according to the task asked of the readers: in a study by Kang *et al.*[15] radiologists' overall sensitivity and specificity was 0.75 and 0.695, respectively, when searching for metastatic lung nodules, but a 0.34 sensitivity and 0.93 specificity for the lung metastatic osteosarcoma patients compared with a 0.97 sensitivity and 0.33 specificity for the lung metastatic non-osteosarcoma group.

Detection sensitivities depend on the nodule size threshold used for detection: increasing the nodule size threshold results in improved detection performance. In the present study, a 3 mm nodule diameter was used as the minimum size, whereas elsewhere[37] a higher nodule size threshold of ≥5 and ≥7 mm in diameter led to higher sensitivities of 0.78 and 0.85, respectively.

The CT protocol, acquisition, and reconstruction affect radiologist's performance. Using a thinner section thickness, preferably ≤2.5 mm, with reconstruction intervals less than or equal to the section thickness improves lung nodule detection.[42,43] One study[44] assessed the influence of section thickness on the ability of six radiologists to detect lung nodules on low-dose CT, in which more nodules were detected using the 2 mm section thickness than the 10 mm section thickness chest CT. An earlier study[45] demonstrated an improved detection rate when using thin section by analysing the interpretations and diagnostic confidence of two radiologists while reviewing 100 chest CT examinations.

One of the limitations of the majority of the previous papers is that they have depended on expert radiologists to define the pathological truth without necessarily having either surgical or histological evidence. The use of only the assessment of one or more expert readers to establish the truth may be a questionable design, depending on the experts. In the present study, to overcome this limitation, the reference standard for truth was histopathology results indicating if a nodule was malignant.

Although a CAD system was not used in the present study, according to a recent review of CAD and lung cancer detection,[46] the authors reported that CAD can potentially assist radiologists by decreasing the rate of missed lung nodules, reducing the time required to read CT examinations, and measuring nodule growth thus facilitating assessment of nodule progression. They also mentioned that a major concern of using a CAD system is the high false-positive rate, resulting from the CAD falsely identifying either benign nodules or non-nodule structures as potential malignancies; however, the performance of a CAD system depends largely on the interaction between the radiologist and the system[47,48] and the radiologist's baseline performance,[49] the latter being the focus of the present study.

If an annual low-dose CT lung cancer screening programme is to be implemented, careful planning is required. A successful screening programme depends on having certified, multidisciplinary medical centres, guidelines for the selection of population at risk, having a standard protocol for imaging high-risk patients, and nodule evaluation and management, in addition to making sure that highly performing readers are available. The present study creates the first steps towards establishing a screening programme in Jordan by establishing that radiologists in Jordan have a lung detection performance comparable to that of experienced radiologists in other studies.

The present study has some limitations. First, readers were not given any information on patient medical history and clinical history, which may have had an impact on the eventual diagnosis. Second, although conclusions from the present analyses generalise to the respective populations of readers, they do not generalise to the respective population of cases. That is, the conclusions generalise to the reader populations, provided that they are reading the specific cases used in this study. Although it is generally preferable to have conclusions generalise to both reader and case populations, doing such an analysis with these data would have resulted in even wider CIs, resulting in more ambiguous conclusions: there would be no significant results, and CIs would be too wide to make clinically meaningful conclusions. More importantly, the primary interest in the present study was in the performance of the radiologists when presented with a representative case sample. Thus this is not considered to be a major limitation, and none of the other studies cited in this paper generalise to both the reader and case populations. Third, although reading time was not measured to be correlated with readers' performances, it is acknowledged that it would be useful information when planning to conduct a screening programme in terms of resource allocation. In the current work, however, gathering timing data was not possible as reading took place in an actual clinical setting, with constant interruptions from medical consultations. Finally, the present study provides the data needed for sizing a larger future study that would generalise to both the reader and case populations.

In conclusion, the results of the present study show that radiologists in Jordan have a similar lung cancer detection performance as demonstrated by the AUC and a higher sensitivity to that demonstrated in other studies. Furthermore, the initial findings suggest that it may be feasible to perform lung cancer screening in multiple non-specialised locations.

Conflict of interest

The authors would like to declare no conflict of interests.

References

1. Henschke CI, Yankelevitz DF, Libby DM, *et al*. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;**355**(17):1763–71.
2. World Health Organization. GLOBOCAN 2012: estimated cancer incidence, mortality and prevelence worldwide in 2012. Available at: http://globocan.iarc.fr/Pages/fact_sheets_population.aspx. Accessed Jan 2018.
3. Team NLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;**365**(5):395.
4. American College of Radiology. *ACR statement on USPSTF draft recommendation for CT lung cancer screening*. 2014, https://www.usprev entiveservicestaskforce.org/Page/Document/RecommendationStatemen tFinal/lung-cancer-screening. Accessed Jan 2018.
5. Wender R, Fontham ET, Barrera E, *et al*. American Cancer Society lung cancer screening guidelines. *CA Cancer J Clin* 2013;**63**(2):106–17.
6. Detterbeck FC, Mazzone PJ, Naidich DP, *et al*. Screening for lung cancer: diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest J* 2013;**143**(5_suppl):e78S–92S.
7. Association AL. Providing guidance on lung cancer screening to patients and physicians. Available at: http://www.lung.org/assets/documents/lung-cancer/lung-cancer-screening-report.pdf. Accessed Feb 2018.
8. Zhou Qh, Fan Yg, Bu H, *et al*. China national lung cancer screening guideline with low-dose computed tomography (2015 version). *Thorac Cancer* 2015;**6**(6):812–8.
9. Pedersen JH, Ashraf H. Implementation and organization of lung cancer screening. *Ann Transl Med* 2016;**4**(8).
10. Standing Committee on Screening. Position statement: lung cancer screening using low-dose computed tomography. Available at: http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/lung-cancer-screening. Accessed Jan 2018.
11. Moyer VA. Screening for lung cancer: US preventive services task force recommendation statement. *Ann Intern Med* 2014;**160**(5):330–8.
12. Nair A, Screaton NJ, Holemans JA, *et al*. The impact of trained radiographers as concurrent readers on performance and reading time of experienced radiologists in the UK Lung Cancer Screening (UKLS) trial. *Eur Radiol* 2017:1–9.
13. Rubin GD, Lyo JK, Paik DS, *et al*. Pulmonary nodules on multidetector row CT scans: performance comparison of radiologists and computer-aided detection. *Radiology* 2005;**234**(1):274–83.
14. Rubin GD, Roos JE, Tall M, *et al*. Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: elucidation with eye tracking. *Radiology* 2014;**274**(1):276–86.
15. Kang MC, Kang CH, Lee HJ, *et al*. Accuracy of 16-channel multi-detector row chest computed tomography with thin sections in the detection of metastatic pulmonary nodules. *Eur J Cardiothorac Surg* 2008;**33**(3):473–9.
16. Roos JE, Paik D, Olsen D, *et al*. Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance. *Eur Radiol* 2010;**20**(3):549–57.
17. Gierada DS, Pilgram TK, Ford M, *et al*. Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening. *Radiology* 2008;**246**(1):265–72.
18. Marten K, Engelke C, Seyfarth T, *et al*. Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. *Clin Radiol* 2005;**60**(2):196–206.
19. Wormanns D, Ludwig K, Beyer F, *et al*. Detection of pulmonary nodules at multirow-detector CT: effectiveness of double reading to improve sensitivity at standard-dose and low-dose chest CT. *Eur Radiol* 2005;**15**(1):14–22.
20. Marten K, Grillhösl A, Seyfarth T, *et al*. Computer-assisted detection of pulmonary nodules: evaluation of diagnostic performance using an expert knowledge-based detection system with variable reconstruction slice thickness settings. *Eur Radiol* 2005;**15**(2):203–12.
21. Abdel-Razeq H, Attiga F, Mansour A. Cancer care in Jordan. *Send Hematol Oncol Stem Cell Ther* 2015;**8**(2):64–70.
22. Clark K, Vendt B, Smith K, *et al*. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013;**26**(6):1045–57.
23. Armato III SG, McLennan G, Bidaut L, *et al*. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys* 2011;**38**(2):915–31.
24. Callister M, Baldwin D, Akram A, *et al*. British Thoracic Society guidelines for the investigation and management of pulmonary nodules: accredited by NICE. *Thorax* 2015;**70**(Suppl. 2):ii1–54.
25. MacMahon H, Naidich DP, Goo JM, *et al*. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology* 2017:161659.
26. Pan X, Metz CE. The "proper" binormal model: parametric receiver operating characteristic curve estimation with degenerate data. *Acad Radiol* 1997;**4**(5):380–9.
27. Metz CE, Pan X. "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *J Math Psychol* 1999;**43**(1):1–33.
28. Hillis SL. Equivalence of binormal likelihood-ratio and bi-chi-squared ROC curve models. *Stat Med* 2016;**35**(12):2031–57.
29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**(1):29–36.
30. Schartz KM, Hillis SL, Pesce LL,, *et al*. *OR-DBM MRMC*. Version 2.50. [Computer software]. Available for download from, http://perception.radiology.uiowa.edu. [Accessed 22 July 2017].
31. Chen W, Samuelson FW. The average receiver operating characteristic curve in multireader multicase imaging studies. *Br J Radiol* 2014;**87**(1040):20140016.
32. Fraioli F, Bertoletti L, Napoli A, *et al*. Computer-aided detection (CAD) in lung cancer screening at chest MDCT: ROC analysis of CAD versus radiologist performance. *J Thorac Imaging* 2007;**22**(3):241–6.
33. Nagatani Y, Takahashi M, Murata K, *et al*. Lung nodule detection performance in five observers on computed tomography (CT) with adaptive

iterative dose reduction using three-dimensional processing (AIDR 3D) in a Japanese multicentre study: comparison between ultra-low-dose CT and low-dose CT by receiver-operating characteristic analysis. *Eur J Radiol* 2015;**84**(7):1401–12.

34. Fujita M, Higaki T, Awaya Y, *et al*. Lung cancer screening with ultra-low dose CT using full iterative reconstruction. *Jpn J Radiol* 2017;**35**(4):179–89.
35. Gierada DS, Pinsky P, Nath H, *et al*. Projected outcomes using different nodule sizes to define a positive CT lung cancer screening examination. *J Natl Cancer Inst* 2014;**106**(11):dju284.
36. Beigelman-Aubry C, Raffy P, Yang W, *et al*. Computer-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. *AJR Am J Roentgenol* 2007;**189**(4):948–55.
37. Kakinuma R, Ashizawa K, Kobayashi T, *et al*. Comparison of sensitivity of lung nodule detection between radiologists and technologists on low-dose CT lung cancer screening images. *Br J Radiol* 2012;**85**(1017):e603–8.
38. Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press; 2003. p. 86.
39. Awai K, Murao K, Ozawa A, *et al*. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology* 2004;**230**(2):347–52.
40. Scholten ET, Horeweg N, de Koning HJ, *et al*. Computed tomographic characteristics of interval and post screen carcinomas in lung cancer screening. *Eur Radiol* 2015;**25**(1):81–8.
41. Gierada DS, Pinsky PF, Duan F, *et al*. Interval lung cancer after a negative CT screening examination: CT findings and outcomes in National Lung Screening Trial participants. *Eur Radiol* 2017;**27**(8):3249–56.
42. Kazerooni EA, Austin JH, Black WC, *et al*. ACR–STR practice parameter for the performance and reporting of lung cancer screening thoracic computed tomography (CT): 2014 (Resolution 4). *J Thorac Imaging* 2014;**29**(5):310–6.
43. Rubin GD. Lung nodule and cancer detection in CT screening. *J Thorac Imaging* 2015;**30**(2):130.
44. Sinsuat M, Saita S, Kawata Y, *et al*. Influence of slice thickness on diagnoses of pulmonary nodules using low-dose CT: potential dependence of detection and diagnostic agreement on features and location of nodule. *Acad Radiol* 2011;**18**(5):594–604.
45. Fischbach F, Knollmann F, Griesshaber V, *et al*. Detection of pulmonary nodules by multislice computed tomography: improved detection rate with reduced slice thickness. *Eur Radiol* 2003;**13**(10):2378–83.
46. Al Mohammad B, Brennan P, Mello-Thoms C. A review of lung cancer screening and the role of computer-aided detection. *Clin Radiol* 2017;**72**(6):433–42.
47. White CS, Pugatch R, Koonce T, *et al*. Lung nodule CAD software as a second reader: a multicentre study. *Acad Radiol* 2008;**15**(3):326–33.
48. Nietert PJ, Ravenel JG, Taylor KK, *et al*. Influence of nodule detection software on radiologists' confidence in identifying pulmonary nodules with computed tomography. *J Thorac Imaging* 2011;**26**(1):48.
49. Bogoni L, Ko JP, Alpert J, *et al*. Impact of a computer-aided detection (CAD) system integrated into a picture archiving and communication system (PACS) on reader sensitivity and efficiency for the detection of lung nodules in thoracic CT exams. *J Digit Imaging* 2012;**25**(6):771–81.