



## UvA-DARE (Digital Academic Repository)

### Advances in digital chest radiography: impact on reader performance

De Boo, D.W.

**Publication date**  
2012

[Link to publication](#)

#### **Citation for published version (APA):**

De Boo, D. W. (2012). *Advances in digital chest radiography: impact on reader performance*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Computer-aided Detection of Small Pulmonary Nodules  
in Chest Radiographs: an Observer Study



Diederick W De Boo  
Martin Uffmann  
Michael Weber  
Shandra Bipat  
Eelco F Boorsma  
Maeke J Scheerder  
Nicole J Freling  
Cornelia M Schaefer-Prokop

## Abstract

### Objective

To evaluate the impact of computer-aided detection (CAD, IQQA-Chest; EDDA Technology, Princeton Junction, NJ) used as second reader on the detection of small pulmonary nodules in chest radiography (CXR).

### Material and Methods

A total of 113 patients (mean age 62 years) with CT and CXR within 6 weeks were selected. Fifty-nine patients showed 101 pulmonary nodules (diameter 5-15mm); the remaining 54 patients served as negative controls. Six readers of varying experience individually evaluated the CXR without and with CAD as second reader in two separate reading sessions. The sensitivity per lesion, figure of merit (FOM), and mean false positive per image (mFP) were calculated. Institutional review board approval was waived.

### Results

With CAD, the sensitivity increased for inexperienced readers (39% vs. 45%,  $p < 0.05$ ) and remained unchanged for experienced readers (50% vs. 51%). The mFP nonsignificantly increased for both inexperienced and experienced readers (0.27 vs. 0.34 and 0.16 vs. 0.21). The mean FOM did not significantly differ for readings without and with CAD irrespective of reader experience (0.71 vs. 0.71 and 0.84 vs. 0.87). All readers together dismissed 33% of true positive CAD candidates. False positive candidates by CAD provoked 40% of all false positive marks made by the readers.

### Conclusion

CAD improves the sensitivity of inexperienced readers for the detection of small nodules at the expense of loss of specificity. Overall performance by means of FOM was therefore not affected. To use CAD more beneficial, readers need to improve their ability to differentiate true from false positive CAD candidates.

## Introduction

In clinical practice, small primary lung carcinomas and pulmonary metastases may be missed on two-view chest radiographs (CXR) though they were frequently visible in retrospect. Depending on the study design, miss rates between 20% and 90% have been reported for primary lung carcinomas<sup>(1-7)</sup>. Factors that contribute to detection errors include image quality, size and type of nodules, superposition of anatomical structures, the presence of accompanying abnormalities, and the radiologists' variable experience and perception capacity. The goal of computer-aided detection (CAD) software is to reduce the effects of the latter, namely to lower the number of perception errors. Variable sensitivities of CAD in chest radiography (34%-78%) have been reported<sup>(8-13)</sup>. Several publications only refer to the CAD stand-alone performance<sup>(8,10,11,13)</sup>. In clinical practice, however, CAD is designed to be used as second reader, meaning that the ultimate impact of CAD on the diagnostic outcome will be determined by both, the CAD performance and the readers' diagnostic judgment. In this study we investigated a US Food and Drug Administration (FDA)-approved CAD system. We evaluated how CAD, when used as second reader, affects the detection performance of readers with vastly varying experience. Study group and nodules were selected to challenge perception capabilities: most nodules were of low conspicuity and the majority of the elderly study patients showed increased parenchymal markings on the CXR from smoking history and aging and described as "dirty lung" in the literature<sup>(14)</sup>.

## Material and Methods

### *Study group*

From our institution's data archive, we retrospectively selected 113 patients who had undergone both, a two-view (CXR) and a chest CT within six weeks. All images had been acquired for clinical purposes only. Institutional review board approval was therefore waived by our institution's ethic committee (registration number 08170465). Patients' mean age was 62 years (16-89 years), 65 were male and 48 were female. According to patients' records, 61 patients (54%) had a positive history of smoking, 27 patients (24%) were nonsmokers and for the remaining 25 patients (22%) no data was available. Both smoking and increase of age lead to a variable increase of parenchymal markings on the CXRs also described as "dirty lung" <sup>(14)</sup>. Its presence was subjectively scored on the CXRs in consensus by a board certified chest radiologist (>15 years of experience) and the researcher (third-year resident in radiology) using a score from 0 (none) to 1 (minimal), 2 (moderate) and 3 (severe). In the following, we will refer to this score as "anatomic noise" (AN).

### *Intrapulmonary nodules*

We followed the glossary of terms for thoracic imaging by the Fleischner society that defines a nodule as a well or poorly defined nodular opacity with a diameter between 3 and 30 mm. Based on this definition, 113 nodules were present in 59 patients. The tested CAD software is optimized for detecting nodules between 5 and 15 mm, therefore the 12 nodules exceeding 15 mm in diameter were excluded from further data analysis. Thus the study group comprised 101 nodules present in 59 patients with a mean and median diameter of 10 mm and a range of 5-15 mm. The 54 patients that had no nodules, as proven by CT, served as normal controls. Seventy-four of the 101 nodules had smooth margins on CT and well-defined contours on the CXR. Twenty-seven nodules had spiculated margins on CT and poorly defined contours on the CXR. None of these nodules were calcified. Histological proof was available in 24% (18/74) of the well-defined nodules (M) and referred to primary non-small-cell lung cancer (8), metastases of renal cell carcinoma (2) or osteosarcoma (1); seven biopsies revealed no malignancy. The remaining 56 well-defined nodules were interpreted as metastases because of a known history of extrathoracic malignancy and the fact that follow-up studies had revealed nodule growth. Histological proof was available in 78% (21/27) of the poorly defined nodules (T) and referred to primary non-small-cell lung cancer (11),

bronchoalveolar cell carcinoma (1), neuro-endocrine tumors (2) and a nonmalignant histology (7). The remaining six poorly defined nodular opacities were considered nonmalignant because they decreased in size over time under antibiotic and / or anti-inflammatory treatment (Table 1).

### *Nodule conspicuity*

Nodule conspicuity on the CXR was subjectively graded by a board-certified chest radiologist (>15 years of experience) and the researcher (third-year resident in radiology) after the reading had been completed and ranged from 1) high to 2) moderate, 3) low and 4) very low.

### *Chest radiography*

CXRs were obtained in digital technique using a dedicated chest stand (Thoravision Philips Medical Systems, Hamburg, Germany). Images were processed using non-linear multifrequency processing (Unique, Philips Medical Systems). Processing parameters were chosen according to the recommendations of the manufacturer and represented the standard processing used in clinical routine at our institution.

### *CAD*

The CAD software (IQQA-Chest; EDDA Technology, Princeton Junction, NJ) is designed to detect nodules in a size range from 5 to 15 mm in diameter. Only the posteroanterior (PA) radiographs are analyzed. The software is running in the background and only on demand between zero and five candidates are marked by semitransparent circles that are centered on a focal density. By visual side-by-side comparison between the marked and unmarked radiograph or by toggling the semitransparent overlay the reader can decide to accept or dismiss the CAD candidate.

### *Image evaluation*

Three radiology residents (first-, second- and third-year training) and three board certified radiologists (two general radiologist, one chest radiologist, all with more than 10 years of experience) independently interpreted all 113 CXR using high resolution LCD monitors (Barco, MDCG 2121-CB, 1.2K x 1.9K matrix) that are subject to regular quality control. PA and lateral radiographs of the 113 patients were evaluated twice in two separate reading sessions, once without CAD and once with the availability of CAD. Half of the cases were seen first without CAD; the other half was first interpreted with the use of CAD. When using CAD, readers were specifically instructed to first analyze the images unassisted before taking the result of the CAD analysis into account and having the candidate circles available.

Readers were asked to document localization of nodules and diagnostic confidence after consideration of CAD. For both conditions, readers were allowed to use processing tools such as windowing or magnification according to their preferences. There was an at least 6 weeks time interval between the two reading sessions and images were evaluated in different random orders. The presence or absence of a nodule was scored using a five point scale of confidence ranging from 5 = definite pulmonary nodule, 4 = probable pulmonary nodule, 3 = equivocal, 2 probably no nodule, and 1 = definitely no nodule. For the confidence ratings 3, 4 and 5, readers were asked to indicate the anatomic location of the suspected nodule on a separate data sheet. This was done separately for each patient; per patient more than one nodule could be marked. The readers were informed that images could contain none, a solitary or multiple nodules but did not know the percentage of each subgroup. They were instructed to ignore calcified nodules and nodules smaller than 5 mm. All readers evaluated 10 training cases with CAD before conducting the study; none of the readers had experience with CAD in daily routine.

### *Statistical analysis*

The patients with and patients without intrapulmonary nodules were compared with regard to gender and smoking history by chi-square test, with regard to age by student t-tests as the data were normally distributed and with regard to AN by chi-square test for trend. For smoking history, the missing data (22%) were excluded. All reader markings, without and with CAD, were determined to be true or false positive by comparing the markings on the separate data sheets with the original CXR and the corresponding CT. This was done in consensus by the researcher and an experienced chest radiologist after all readings had been completed. The data were analyzed using the jackknife free-response receiver operating characteristic (JAFROC) method especially developed to analyze observer free-response tasks<sup>(15-17)</sup>. JAFROC software 2.3a was used to calculate a figure of merit (FOM). The FOM is defined as the probability on a scale from 0 to 1.0 that true positive marks for nodules are rated with higher confidence than false positive (non-lesions) marks on control CXRs<sup>(16)</sup>. A FOM of 0.5 means that the confidence of marks for nodules on abnormal CXR are equal to marks for nonlesions (false positive marks) on negative control CXR. A FOM of 1.0 describes the ideal situation where all nodules are correctly marked with no false positives in the control images. Sensitivity was calculated on a per-nodule basis. For these calculations, reader ratings 4 and 5 were considered true positive if made for correctly

localized nodules. CAD candidates were considered true positive if the candidate was centralized above a CT proven nodule. Specificity was described as mean number of false positive marks per image (mFP). For these calculations, reader ratings 4 and 5 were considered false positive, if made in locations, where CT did not show a nodule. CAD candidates only partially covering a nodule or indicating a nodule not proven by CT were considered false positive. Sensitivities and mFPs were calculated for CAD as stand-alone and for each reader separately without and with use of CAD. Significance of differences was tested using McNemar’s test. We also assessed the number of rating differences made in the two reading sessions without and with the availability of CAD. A rating difference was considered beneficial if a nodule was missed during reading without CAD but was correctly detected during the reading with CAD. The scenario, that the reader made a false positive mark without CAD but made no false positive mark with the availability of CAD, was also considered a beneficial rating difference. It was considered a detrimental rating difference if the reader detected the nodule during the reading without CAD but missed it in the reading session with the availability of CAD or if the reader made a new false positive mark during the reading with CAD. Agreement between the readers and the stand-alone performance of the CAD-system was calculated pairwise using non-weighted Kappa statistics. A Fleiss Kappa statistics was used to evaluate the interobserver agreement during readings without and readings with CAD as second reader. All analyses were performed in SPSS 15.0.1. Significance was assumed at  $p < 0.05$ .

**Table 1**  
Characteristics of the 101 nodules.

Characteristics		
<b>Shape</b>	M	27
	T	74
<b>Conspicuity</b>	high (1)	21
	moderate (2)	39
	low (3)	20
	very low (4)	21
<b>Size</b>	<10mm	42
	≥10mm	59

M: round nodules, T: nodules with irregular margins

## Results

### *Study groups and nodule characteristics*

An increased AN was scored present in 69% of all patients and varied from mild in 42% to moderate in 18% and severe in 9% of patients. Of the 59 patients with nodules 38 had a solitary nodule, 10 had two nodules and 11 had more than two nodules. The nodules were located in the upper lobes in 58%, in the middle lobe in 10% and in the lower lobes in 32%. Nodule conspicuity was rated very low in 21%, low in 39%, moderate in 20% and high in 21%. Nodule conspicuity was negatively correlated with AN ( $p = 0.04$ ) and nodule size ( $p = 0.024$ ). There were no differences between patients with and without nodules with regard to age, gender, AN or history of smoking ( $p = 0.85, 0.52, 0.49, 0.20$ , respectively).

### *Stand-alone performance of CAD*

The CAD system demonstrated a stand-alone sensitivity of 47%. The sensitivity was 70% for nodules with moderate and high conspicuity and 30% for nodules with low and very low conspicuity. The CAD stand-alone sensitivity was not associated with nodule shape, nodule size, anatomic location or AN. A total of 194 false positive candidates were produced with a mFP of 1.7. There was no association between the number of produced false positives and the AN.

**Table 2**

Individual reader outcome without and with use of CAD.

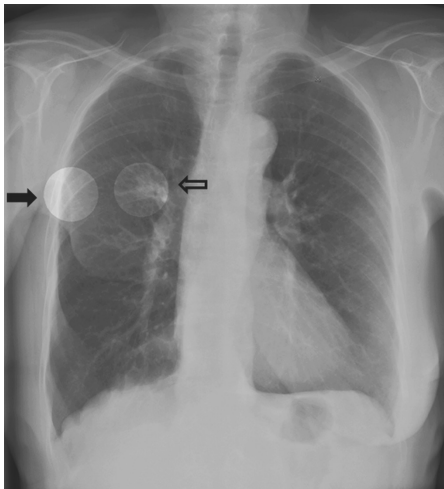
	FOM		Sensitivity (in%)		mFP	
	Baseline	CAD	Baseline	CAD	Baseline	CAD
<b>Inexperienced readers (n = 3)</b>	<b>0.71</b>	<b>0.71</b>	<b>39</b>	<b>45</b>	<b>0.27</b>	<b>0.34</b>
1	0.52	0.53	39	46	0.54	0.67
2	0.77	0.79	35	39	0.08	0.13
3	0.84	0.80	42	50	0.18	0.21
<b>Experienced readers (n = 3)</b>	<b>0.84</b>	<b>0.87</b>	<b>50</b>	<b>51</b>	<b>0.16</b>	<b>0.21</b>
4	0.85	0.91	45	44	0.15	0.18
5	0.80	0.83	53	54	0.20	0.24
6	0.87	0.88	53	57	0.12	0.20

Baseline: reading without CAD, CAD: reading with computer-aided detection as second reader, FOM: figure of merit, mFP: mean false positives per image

### Reader performance without CAD

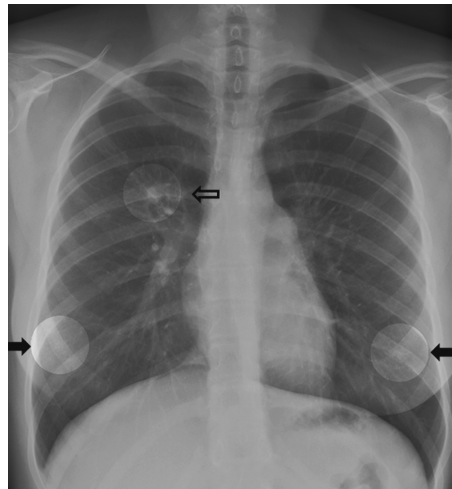
The mean sensitivity for the inexperienced readers was 39% with a mFP of 0.27. The mean sensitivity for the experienced readers was 50% with a mFP of 0.16. The average FOM for the inexperienced readers was 0.71 compared with 0.84 for the experienced readers (Table 2). For both reader groups, there was no significant association between the number of false positives and AN. Looking at subgroups of nodules as function of conspicuity, size and anatomic background noise, the readers' sensitivity was generally lower for nodules of low conspicuity (22% vs. 76%) and smaller size (25% vs. 58%) and for nodules located in lungs with increased anatomic noise (31% vs. 52%; Table 3).

**Figure 1**



The three inexperienced readers missed the right perihilar nodule without computer-aided detection (CAD) but accepted the true positive CAD candidate (open arrow). All readers dismissed the false positive CAD candidates located more laterally (black arrow).

**Figure 2**



Four readers dismissed the true positive computer-aided detection (CAD) candidate for the right perihilar nodule (open arrow). The CAD candidates in the right and left lower zone represented false positives.

### Reader performance with CAD as second reader

For the group of the inexperienced readers, the mean sensitivity increased from 39% to 45% ( $p = 0.008$ ; example in Figure 1) with a nonsignificant increase of the mFP from 0.27 to 0.34. For the experienced readers the mean sensitivity remained unchanged (50% vs. 51%) with a mFP that also nonsignificantly increased from 0.16 to 0.21. When separately analyzed, sensitivity increased for five of six readers but differences were not significant. The mean FOM was higher for the experienced readers but for both reader groups, the mean FOM did not significantly change with the availability of CAD (0.71 vs. 0.71 and 0.84 vs. 0.87, respectively).

### *Reader-CAD interaction*

Altogether, 33% (94/282) of the true positive CAD candidates were dismissed by the readers, meaning readers did not realize that CAD correctly indicated existing nodules which had been primarily missed by the readers. The readers subsequently disregarded those CAD candidates though they were true positive (Figure 2). Sixty-eight percent of these dismissed true positive CAD candidates were for nodules of low to very low conspicuity. Inexperienced readers declined 51, whereas experienced readers declined 43 true positive CAD candidates (Table 4). When comparing the readings without and with CAD we found a total of 286 rating differences that affected the correct diagnosis. 54% (154/286) of them were detrimental and 46% were beneficial. 40% (61/154) of the detrimental rating differences represented false positive ratings likely to have been provoked by a false positive CAD candidate in identical location (Figure 3). Maximum benefit from CAD would be established if the readers accepted all true positive CAD candidates and declined only false positive CAD candidates. In this theoretical scenario all readers would show a substantial increase in sensitivity to a mean overall sensitivity of 70% with a slight inferiority of the inexperienced readers (66%) compared to the experienced readers (74%).

### *Interobserver agreement and agreement with CAD*

The agreement between the stand-alone performance of CAD and the six readers was poor to fair with Kappa values ranging from 0.15 to 0.32. The interobserver agreement for the readers during reading without CAD was 0.36 (fair) and increased slightly to 0.39 (fair) when CAD was used as second reader.

**Table 3**  
Sensitivity (in %) of CAD stand-alone and of readers without and with CAD for subgroups of nodules, dependant on shape, conspicuity, size and presence of anatomic noise.

	Shape		Conspicuity			Size			AN	
	T	M	1+2	3+4	<10mm	≥10mm	0+1	2+3		
<b>CAD stand-alone</b>	48	46	70	30	40	51	51	39		
<b>All readers (n=6)</b>										
baseline	38	46	76	22	25	58	52	31		
CAD	42	50	84	23	29	61	56	31		
<b>Inexperienced readers (n=3)</b>										
baseline	30	41	69	17	19	52	45	27		
CAD	37	47	82	19	25	58	52	30		
<b>Experienced readers (n=3)</b>										
baseline	46	51	83	27	31	63	58	34		
CAD	47	52	85	28	33	64	61	33		

AN: anatomic noise, Baseline: reading without CAD, CAD: reading with computer-aided detection as second reader,

M: round nodules, T: nodules with irregular margins

**Table 4**  
Dismissed true positive CAD candidates (in absolute numbers and in percentage of all CAD true positives between brackets) pooled over all readers and over subgroups of experienced and inexperienced readers.

	Shape		Conspicuity			Size			AN	
	T	M	1+2	3+4	<10mm	≥10mm	0+1	1+2		
<b>All readers (n=6)</b>	34 (44%)	60 (29%)	21 (12%)	73 (68%)	52 (51%)	42 (23%)	50 (25%)	44 (52%)		
<b>Inexperienced readers (n=3)</b>	19 (49%)	32 (31%)	11 (13%)	40 (74%)	28 (55%)	23 (26%)	29 (29%)	22 (52%)		
<b>Experienced readers (n=3)</b>	15 (38%)	28 (27%)	10 (11%)	33 (61%)	24 (47%)	19 (21%)	21 (21%)	22 (52%)		

AN: anatomic noise, M: round nodules, T: nodules with irregular margins

**Figure 3**

*Four false positives were generated by the computer-aided detection (CAD) software. Three readers accepted the left lower false positive (arrow). The other false positives were dismissed by all readers.*

## Discussion

Various CAD systems with and without FDA approval have been developed for the detection of nodules in CXRs. Reported stand-alone sensitivities of these systems vary from 34% to 78% dependant on lesion selection and study group<sup>(8-13)</sup>. In our study, the stand-alone sensitivity of the tested CAD software was 47% and this was slightly better than the mean sensitivity of 44% achieved by the readers. Thus both readers and CAD showed a relatively low baseline performance, which is likely to be the result of the low to very low conspicuity of the majority of the studied nodules. For a successful application of CAD, it seems warranted that CAD detects nodules, which the radiologist tends to miss. As indicated by the Kappa analysis, we found a relatively lower agreement between CAD and the readers as between the readers, supporting the fact that CAD has the potential to find nodules, the readers have not seen. Also Bley et al., who evaluated the same CAD software, reported an only moderate agreement between CAD and the readers underlining the capacity of this CAD software to detect different nodules as the readers<sup>(10)</sup>. Application of CAD ideally leads to an increase of sensitivity without loss of specificity. We found a significant increase in sensitivity for the inexperienced readers with the use of CAD. Experienced readers, however, did not take any advantage of CAD. Yet, the increase of sensitivity of the inexperienced readers went along with a loss of specificity that, though not statistically significant, impeded

an increase of the FOM. With regard to the localization of false positive marks by the readers, an interaction with CAD candidates took place: 40% of the detrimental rating differences referred to false positive marks that corresponded to false positive CAD candidates in these locations and thus were likely to have been provoked by CAD. On the other hand our study could not prove a significant increase of the mFP. These findings suggest that false positives seen by the readers without CAD did not attract the same grade of suspicion when CAD indicated different candidate areas. In fact, in a substantial amount of readings (61/154) CAD candidates entrapped the readers to call an area suspicious. The number of false positive marks possibly provoked by CAD were equally distributed between experienced and inexperienced readers and showed no significant association with the amount of anatomic noise. Though, our results could not prove that CAD improves detection performance for small intrapulmonary nodules, there are some findings indicating that CAD may have potential to increase detection rates if applied differently. In that context, it is noteworthy that all six readers together declined 33% of the true positive CAD candidates. The majority of the dismissed true positive candidates referred to the low conspicuous nodules and to nodules smaller than 10 mm in diameter. These nodules represent indeed the type of pathology that are most difficult to detect by the readers and might take the highest advantage of a CAD system. If the readers would have accepted all true positive CAD candidates and declined only false positive CAD candidates, all readers would have shown a substantial increase in performance from a baseline sensitivity of 47% to a mean overall sensitivity of 70%, with a slight inferiority of the inexperienced readers (66%) compared to the experienced readers (74%). These findings indicate the potential of CAD if used optimally but also indicate the readers' severe difficulty to distinguish true positive from false positive CAD candidates. For further improvement of reader performance with CAD, two aspects seem important. Firstly, the number of false positive CAD candidates should be decreased. Recent publications described a significantly reduced mFP when CAD was applied on energy subtracted soft-tissue images<sup>(18,19)</sup>. Whether this indeed will increase reader performance yet remains to be proven by an observer study. Second, additional tools appear to be needed to help the observer to accept true positive candidates even when they indicate low conspicuous nodules. Whether this can be achieved by means of additional displays such as grey-scale reversal or rib subtraction or whether it requires additional software tools, such as indication of the grade of suspicion based on the CAD analysis, is subject of ongoing research. The latter was found quite promising for reading mammography: reader detection performance could be further

increased when the CAD software indicated a likelihood of suspicion for an area of interest selected by the observer instead of showing the reader a certain amount of equally weighted ROIs<sup>(20)</sup>. Previous literature reports controversial results about the effect of CAD on the detection of nodules. While Song et al. only state the impact on sensitivity without considering the specificity<sup>(21)</sup>, other studies more precisely report changes of Az that way statistically taking into consideration both sensitivity and specificity. Most interestingly, all studies that report a significantly increased performance (increase of Az) also stated a higher CAD stand-alone performance with sensitivities between 52% and 100% as compared to the stand-alone sensitivity of 47% found in our study<sup>(12,22,23,24,25)</sup>. The majority of the CAD systems evaluated were not FDA approved at time of study conductance. There are four studies with a stand-alone sensitivity below 50%<sup>(8,10,22,26)</sup>: two of them assessed reader performance and found no increase of reader performance measured as Az or FOM, conform our results. We conclude from this that in addition to reader experience the conspicuity of nodules represents a determining factor with respect to the impact of CAD on reader performance. As seen in our results (Table 3), more conspicuous nodules (here classified as 1 and 2), if missed by the reader because of perception error, were more easily accepted by the observer when indicated by CAD. Low conspicuous nodules (classified as 3 and 4), however, that require high perception and interpretation skills, were less easily accepted by the observer even when correctly indicated by CAD. Our study suffers from a number of limitations:

- a) We evaluated a selected group of patients with a higher prevalence of nodules than normally seen in clinical routine. Readers were specifically asked to look for small nodular densities which probably lowered their overall specificity.
- b) Images were evaluated without and with CAD as second reader in two reading sessions introducing the additional factor of intrareader variability. We chose to do so to keep reading conditions possibly realistic and not to interrupt the readers' visual analysis by requiring nodule documentation separately without and with CAD. However, we cannot exclude that intrareader variability may have partially ameliorated the effect of CAD on reader performance.
- c) Though the readers were appropriately introduced to the use of CAD by a number of training cases, which were not included in the study, none of them had experience with CAD for a longer period of time or within routine application. For mammography a substantial positive impact of learning effects on the performance with CAD is assumed to be two years<sup>(27)</sup>; whether this also holds for chest radiography remains to be evaluated.

d) We evaluated a specific type of CAD algorithm. Our results only pertain to that particular CAD application and to that particular study group and are not directly transferrable to another CAD software and to other study groups of lesions.

In summary we conclude that the sensitivity of the inexperienced readers significantly increased with the use of CAD as second reader. Overall performance (FOM) for both inexperienced and experienced readers, however, was not affected, meaning that the increase in sensitivity came with a decrease in specificity. The impact of CAD on the detection of small and low conspicuous nodules seems to be largely impeded by the readers' inability to differentiate true from false positive CAD candidates. There is potential for further improvement of reader performance with CAD, given the high rate of dismissed true positive CAD candidates and the number of accepted CAD false positives. Our findings underline the importance to further decrease the number of false positive CAD candidates in the future and the need for further research to find out whether a longer learning process or additional visual or analytic tools that come along with the CAD output can further improve the readers' ability to distinguish true from false positive CAD candidates.

## References

- 1 Muhm JR, Miller WE, Fontana RS, Sanderson DR, Uhlenhoop MA. Lung cancer detected during a screening program using 4-month chest radiographs. *Radiology* 1983; 148:609-615
- 2 Heelan RT, Flehinger BJ, Melamed MR, et al. Non-small-cell lung cancer: results of the New York screening program. *Radiology* 1984; 151:289-293
- 3 Austin JH, Romney BM, Goldsmith LS. Missed bronchogenic carcinoma: radiographic findings in 27 patients with potentially resectable lesion evident in retrospect. *Radiology* 1992; 182:115-122
- 4 Quekel LG, Kessels AG, Goei R, Engelshoven JM. Miss rate of lung cancer on the chest radiograph in clinical practice. *Chest* 1999; 115:720-724
- 5 Monnier-Cholley L, Arrivé L, Porcel A, et al. Characteristics of missed lung cancer on chest radiographs: a French experience. *Eur Radiol* 2001; 11:597-605
- 6 Shah PK, Austin JH, White CS, et al. Missed non-small cell lung cancers: Radiographic findings of potentially resectable lesions evident only in retrospect. *Radiology* 2003; 226:235-241
- 7 Wu HM, Gotway MB, Lee TJ, et al. Features of non-small cell lung carcinomas overlooked at digital chest radiography. *Clinical Radiology* 2008; 63:518-528
- 8 Li F, Engelmann R, Metz CE, Doi K, MacMahon H. Lung cancers missed on chest radiographs: results obtained with a commercial computer-aided detection program. *Radiology* 2008; 246:273-280
- 9 van Beek EJ, Mullan B, Thompson B. Evaluation of a real-time interactive pulmonary nodule analysis system on chest digital radiographic images: a prospective study. *Acad Radiol* 2008; 15:571-575
- 10 Bley TA, Baumann T, Saueressig U, et al. Comparison of radiologist and CAD performance in the detection of CT-confirmed subtle pulmonary nodules on digital chest radiographs. *Invest Radiol* 2008; 43:343-348
- 11 White CS, Flukinger T, Jeudy J, Chen JJ. Use of a computer-aided detection system to detect missed lung cancer at chest radiography. *Radiology* 2009; 252:273-281
- 12 Sakai S, Soeda H, Takahashi N, et al. Computer-aided detection on digital chest radiography: validation test on consecutive T1 cases of resectable lung cancer. *J Digit Imaging* 2006; 19:376-382
- 13 Hardie RC, Rogers SK, Wilson T, Rogers A. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. *Med Image Anal* 2008; 12:240-258
- 14 Gückel C, Hansell DM. Imaging the 'dirty lung' – has high resolution computed tomography cleared the smoke? *Clinical Radiology* 1998; 53:717-722
- 15 Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. *Med Phys* 2004; 31:2313-2330
- 16 Chakraborty DP. Analysis of location specific observer performance data: validated extensions of the jackknife free-response (JAFROC) method. *Acad Radiol* 2006; 13:1187-1193
- 17 Vikgren J, Zachrisson S, Svalkvist A, et al. Comparison of chest thomosynthesis and chest radiography for the detection of pulmonary nodules: human observer study of clinical cases. *Radiology* 2008; 249:1034-1041
- 18 Szucs-Farkas Z, Patak MA, Yuksel-Hatz S, Ruder T, Vock P. Improved detection of

- pulmonary nodules on energy-subtracted chest radiographs with a commercial computer-aided diagnosis software: comparison with human observers. *Eur Radiol* 2010; 20:1289-96
- 19 Balkman JD, Mehandru S, DuPont E, Novak RD, Gilkeson RC. Dual energy subtraction digital radiography improves performance of a next generation computer-aided detection program. *J Thorac Imaging* 2010; 25:41-47
  - 20 Samulski M, Hupse R, Boetes C, Mus RD, den Heeten GJ, Karssemeijer N. Using computer-aided detection in mammography as a decision support. *Eur Radiol* 2010; 20:2323-2330
  - 21 Song W, Fan L, Xie Y, Qian JZ, Jin Z. A study of inter-observer variations of pulmonary nodule marking and characterizing on DR images. *Proc SPIE* 2005; 5749:272-280
  - 22 Kasai S, Li F, Shiraishi J, Doi K. Usefulness of computer-aided diagnosis schemes for vertebral fractures and lung nodules on chest radiographs. *AJR Am J Roentgenol* 2008; 191:260-265
  - 23 MacMahon H, Engelmann R, Behlen FM, et al. Computer-aided diagnosis of pulmonary nodules: results of a large scale observer test. *Radiology* 1999; 213:723-726
  - 24 Freedman MT, Lo SCB, Osicka T, et al. Computer-aided detection of lung cancer on chest radiographs: effect of machine CAD false-positive locations on radiologists' behavior. *Proc SPIE* 2002; 4684:1311-1319
  - 25 Shiraishi J, Abe H, Li F, Engelmann R, MacMahon H, Doi K. Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs: ROC analysis of radiologists' performance. *Acad Radiol* 2006; 13:995-1003
  - 26 de Hoop B, De Boer DW, Gietema HA, et al. Computer-aided detection of lung cancer on chest radiographs: effect on observer performance. *Radiology* 2010; 257:532-540
  - 27 Nishikawa R. Increased CAD use prompts look at advantages, drawbacks. In *RSNA news* February 2007. [http://www.rsna.org/Publications/rsnanews/feb07/upload/RSNANews\\_Feb07\\_CAD\\_Usage.pdf](http://www.rsna.org/Publications/rsnanews/feb07/upload/RSNANews_Feb07_CAD_Usage.pdf)