**Article 25fa pilot End User Agreement**

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. Please note that you are not allowed to share this article on other platforms, but can link to it. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication or parts of it other than authorised under this licence or copyright law is prohibited. Neither Radboud University nor the authors of this publication are liable for any damage resulting from your (re)use of this publication.

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 contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library  
Radboud University  
Copyright Information Point  
PO Box 9100  
6500 HA Nijmegen

You will be contacted as soon as possible.
Computer-aided Detection Improves Detection of Pulmonary Nodules in Chest Radiographs beyond the Support by Bone-suppressed Images

Purpose:
To evaluate the added value of computer-aided detection (CAD) for lung nodules on chest radiographs when radiologists have bone-suppressed images (BSIs) available.

Materials and Methods:
Written informed consent was waived by the institutional review board. Selection of study images and study setup was reviewed and approved by the institutional review boards. Three hundred posteroanterior (PA) and lateral chest radiographs (189 radiographs with negative findings and 111 radiographs with a solitary nodule) in 300 subjects were selected from image archives at four institutions. PA images were processed by using a commercially available CAD, and PA BSIs were generated. Five radiologists and three residents evaluated the radiographs with BSIs available, first, without CAD and, second, after inspection of the CAD marks. Readers marked locations suspicious for a nodule and provided a confidence score for that location to be a nodule. Location-based receiver operating characteristic analysis was performed by using jackknife alternative free-response receiver operating characteristic analysis. Area under the curve (AUC) functioned as figure of merit, and P values were computed with the Dorfman-Berbaum-Metz method.

Results:
Average nodule size was 16.2 mm. Stand-alone CAD reached a sensitivity of 74% at 1.0 false-positive mark per image. Without CAD, average AUC for observers was 0.812. With CAD, performance significantly improved to an AUC of 0.841 (P = .0001). CAD detected 127 of 239 nodules that were missed after evaluation of the radiographs together with BSIs pooled over all observers. Only 57 of these detections were eventually marked by the observers after review of CAD candidates.

Conclusion:
CAD improved radiologists’ performance for the detection of lung nodules on chest radiographs, even when baseline performance was optimized by providing lateral radiographs and BSIs. Still, most of the true-positive CAD candidates are dismissed by observers.
chest radiography still represents the first-line method for diagnosis of lung nodules regardless of its known inferiority to computed tomography (CT). In addition to the fact that a subset of nodules is therefore indeed not visible on a chest radiograph, there is a subset of lesions that are well visible but are overlooked by the observer (1,2). The latter may be due to overprojection by other anatomic structures or may happen due to inattentive blindness by the observer, meaning that lesions are simply overlooked or no appropriate action is triggered once the lesion is noticed. Therefore, the goal of recent image processing techniques is to reduce the risk of overlooking nodular lesions that are well visible in retrospect. According to the literature, the number of lung cancers initially missed on chest radiographs but retrospectively visible amounts to 19%–26% in various study populations (1,2). Overprojection by osseous structures has been reported as one reason leading to reduced detection of lung cancer (3,4). Bone-suppressed images (BSIs) address this issue and have been found to significantly improve observer performance for detection of pulmonary nodules, with an increase of sensitivity ranging from 4% to 17% (5,6).

To further reduce the risk of missing a nodule that is visually discernible on the radiograph, several programs for computer-aided detection (CAD) have been developed. Since the first introduction of CAD, these programs have been continuously improved with respect to sensitivity and specificity. Researchers in observer studies so far have reported variable results. For instance, while investigators in one study (7) demonstrated an improved detection of lung cancers from 68.2% to 76.7%, those in two other studies (8,9) did not find a significant improvement with CAD. As a possible underlying problem, the authors in those studies discussed that observers could not sufficiently discriminate between true-positive (TP) and false-positive (FP) CAD candidate lesions: TP CAD candidate lesions were dismissed, while FP CAD candidate lesions were accepted by the readers (8,9). It is conceivable that combining both efforts, BSIs and CAD, leads to an improved detection performance for small and low-conspicuity nodules. The purpose of our study was to evaluate the added value of CAD for lung nodules in chest radiographs when radiologists have BSIs available.

Materials and Methods
Riverain Technologies (Miamisburg, Ohio) paid for this study, and one author (C.M.S.) served on an advisory board for Riverain Technologies. Study design, data acquisition and analysis, as well as manuscript writing, were completely controlled by the authors without any influence by Riverain Technologies.

Data
Selection of study images and study setup was reviewed and approved by the institutional review boards. Written informed consent was waived. We retrospectively selected 300 digital posteroanterior (PA) and lateral chest radiographs by reviewing the image archives of four institutions (three academic and one nonacademic hospital), as follows: Radboud University Medical Center (Nijmegen, the Netherlands), University Medical Center (Utrecht, the Netherlands), Academic Medical Center (Amsterdam, the Netherlands), and Meander Medical Center (Amersfoort, the Netherlands). All images were derived from clinically indicated examinations.

Inclusion criteria were the presence of a solid solitary nodule (<30 mm in diameter) and the availability of a PA and lateral chest radiograph and a chest CT scan obtained within 3 months. Radiographs showing signs of other disease, except for chronic obstructive pulmonary disease, were not included. All subjects had to be older than 40 years. For control patients, absence of disease was ascertained by a chest radiograph with negative findings and a CT scan with negative findings within 6 months of the chest radiograph.

To ensure a wide range of lesion conspicuousities with a sufficient number of low-conspicuity lesions, visibility of the nodules was assessed by an expert radiologist and a clinical researcher (C.M.S. and S.S.) in consensus. Nodules had to be visible on the PA radiograph and, on

**Advances in Knowledge**
- Computer-aided detection (CAD) improves observer performance for the detection of lung nodules on chest radiographs, beyond the application of bone suppression alone; CAD detected 53% (127 of 239) of the nodules that initially were missed by the readers; readers accepted 43% (57 of 127) of these true-positive CAD candidates.
- The beneficial effect of CAD is limited by the insufficient ability of the observers to differentiate true-positive from false-positive CAD candidates.

**Implication for Patient Care**
- Combination of CAD and bone suppression in chest radiography improves detection of potentially early lung cancers.

**Published online before print**
10.1148/radiol.14131315 Content code: CH

**Radiology 2014; 272:252–261**

**Abbreviations:**
AFROC = alternative free-response receiver operating characteristic  
AUC = area under the curve  
BSI = bone-suppressed image  
CAD = computer-aided detection  
FP = false-positive  
PA = posteroanterior  
TP = true-positive

**Author contributions:**
Guarantors of integrity of entire study, S.S., C.M.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, S.S., B.v.G., N.K., C.M.S.; clinical studies, S.S., B.v.G., E.K., A.M.T., R.W., N.K.; statistical analysis, S.S., A.M.T., N.K., C.M.S.; and manuscript editing, S.S., B.v.G., M.M.S., R.W., N.K., C.M.S.

Conflicts of interest are listed at the end of this article.
the basis of visual inspection, were classified into four categories, ranging from well visible (category 1) to moderately subtle (category 2), subtle (category 3), and very subtle (category 4). Lesions were annotated on the PA chest radiograph, after a coronal projection image of the CT scan was reviewed.

Nodule volume (in cubic millimeters) was calculated from annotations made on CT scans (10). Subsequently, diameter of the nodule was calculated from the nodule volume, assuming each nodule to be a sphere. If known, pathologic findings and follow-up data were derived from clinical records.

Image Acquisition

All chest radiographs were obtained with a digital technique by using storage phosphor plates (CR; Agfa Healthcare, Mortsel, Belgium), a selenium drum (Thoravision; Philips Healthcare, Hamburg, Germany), and flat-panel-detector digital radiography systems (Siemens, Erlangen, Germany). Image postprocessing was applied as recommended by the manufacturer and used in clinical routine.

Image Processing Software

Computer-detection output was generated by a commercially available CAD system (ClearRead +Detect 5.2; Riverain Technologies). The CAD system is optimized for detection of nodules between 9 and 30 mm in diameter, although larger and smaller nodules are also marked if they are localized with the software. Candidate lesions identified with the software were marked by circles. CAD marks could be displayed on both the original radiograph and the PA BSIs. Only PA radiographs are processed with the system.

BSIs were computed by using software (ClearRead Bone Suppression 2.4; Riverain Technologies). The software generates PA radiographs identical to the original image with respect to size and gradation characteristics, with the difference that overprojections of ribs and clavicle were digitally removed. Both software products are commercially available and have U.S. Food and Drug Administration approval.

Reading Methods

Five radiologists, one of whom was an author (M.M.S., with 5 years of experience) and four of whom were not authors (with 13, 3, 17, and 17 years of experience), and three residents (E.K., a 4th-year resident; R.W., a 2nd-year resident; and A.M.T., a 4th-year resident) evaluated the radiographs in different randomized orders. None of the readers had any previous experience with BSIs or CAD.

Readers reviewed the cases first without and subsequently with the use of CAD marks within one reading session. BSIs were available at all times. To familiarize the readers with the software used in the study, a training session of 40 cases, with instant feedback from the researcher, was provided in advance. The training set consisted of 22 patients with a nodule and 18 control patients without a nodule, which were not used in the study. Training images contained representative lesions and other abnormalities (eg, old rib fractures) to familiarize the radiologists with the BSIs and the CAD output. The setup of the training was similar to the review of study images, including the two-stage scoring, without and with CAD.

In the observer study, readers were able to mark and score suspicious regions in the PA chest radiograph. A ruler on the screen with a continuous scale between 0 and 100 was used to document the reader's degree of suspiciousness (confidence) that a nodule was present (0, not suspicious; 100, definitely suspicious). Observers were allowed to score multiple regions that were suspicious for a nodule per image. All readings with respect to localization and confidence scale were digitally documented. Readers did not have the ability to change previous annotations.

Readings took place at a 30-inch Digital Imaging and Communications in Medicine–calibrated liquid crystal display monitor (FlexscanSX3031 W; Eizo, Ishikawa, Japan), with a native screen resolution of $2560 \times 1600$ in a darkened room, mimicking clinical reading conditions. Processing tools, such as adjustment of window and level, zoom in and zoom out, and gray-scale inversion, were available. Observers could review both PA and lateral images side by side. BSIs could be visualized with a key on the keyboard and appeared at the exact same location as the PA radiograph. In this way, the reader could toggle between the original chest radiograph and the BSI to easily review corresponding areas in the radiograph. After the first scoring without CAD but with BSIs, CAD marks were automatically displayed and could also be toggled on and off, if desired, by the readers. Subsequently, the readers were asked to score location and degree of suspicion for the presence of a nodule for the second time: The readers were allowed to score new lesions, remove old lesions, or modulate previously determined scores after they had seen the CAD marks.

Radiologists had been given the information that a maximum of one nodule was present in each examination. They knew the study set consisted of more control patients than patients with a nodule, but they did not know the exact numbers. Radiologists were assigned to read the chest radiographs in different random order, encouraging them to review chest radiographs with a reading speed they would have in clinical conditions.

Statistics

For statistical analysis, multireader multiple-case jackknife alternative free-response receiver operating characteristic (AFROC) analysis was performed (11,12). A finding by an observer was considered a TP finding when the marking was within 1 cm of the center of the ground-truth annotation. As input for jackknife AFROC analysis, only one reader score per image is used for analysis. For cases with negative (normal) findings, this is the FP finding with the highest score. For cases with positive (abnormal) findings, markings of nonlesion locations are ignored, and only TP markings (if present) are used. In that way, we ensured that readers could not be rewarded for marking nonlesion locations in cases with positive findings. Area under the curve (AUC), which represents the probability that a lesion is rated higher than nonlesion...
locations in images with negative findings, was calculated by using the trapezoidal integration method, also known as the Wilcoxon rank-sum test. AUCs without and with help of CAD were compared with the Dorfman-Berbaum-Metz method (DBMMRMC package, version 2.33, Medical Image Perception Laboratory, Department of Radiology, Carver College of Medicine, University of Iowa, Iowa City, Iowa; http://perception.radiology.uiowa.edu), which includes reader, case, and treatment variance. Also partial AUC, corresponding to a specificity range of 80%–100%, was calculated from the AFROC curves.

Further, sensitivity was calculated by dividing the number of correctly localized lesions by the total number of lesions. Specificity was calculated by dividing the number of nonmarked cases by the total number of cases with negative findings. These calculations were performed, taking all scores of suspiciousness into account, and were performed for all lesions, as well as for the subcategories of various conspicuities. Sensitivities and specificities were also calculated for different thresholds of suspiciousness scores given by the observers to investigate the shift in sensitivity and specificity with use of CAD. Differences in patient characteristics were compared by using a χ² test with respect to sex distribution and an unpaired t test with respect to age distribution. A significant difference was defined as P < .05.

**Results**

**Patient Characteristics**

Three hundred subjects were selected for the study, and images from these subjects included 189 radiographs with negative findings and 111 radiographs with a solitary pulmonary nodule. Average age was 65 years (range, 44–88 years) for patients with a nodule and 64 years (range, 41–88 years) for control subjects. Differences in average age were not significant (P = .22). One hundred seventy-seven subjects were male (average age, 64 years; range, 44–87 years), and 123 were female (average age, 64 years; range, 41–88 years). Difference in age between male and female subjects was not significant (P = .76). There was an even age distribution over the groups (node group: 66 male subjects [age, 66 years; range, 44–83 years] and 43 female subjects [age, 63 years; range, 66–88 years] vs control group: 111 male subjects [age, 63 years; range, 44–87 years] and 78 female subjects [age, 64 years; range, 41–88 years]). Average nodule diameter as determined on the CT scan was 16.2 mm (median, 15.1 mm; range, 7.8–35 mm). When lesions were measured on the CT scans, five lesions exceeded 30 mm in diameter, herewith representing a mass and not a nodule according to the Fleischner glossary (13). The conspicuity of these five lesions was categorized as being obvious (n = 1), moderately subtle (n = 3), and subtle (n = 1). Seventy-eight lesions were malignant, and malignancy was histologically proved in 67 cases and was based on clinical history in 11 cases. Twenty nodules were benign, and the pathologic findings in 13 nodules were unknown.

**Stand-alone CAD**

CAD reached a stand-alone sensitivity of 74% (82 of 111) at 1.0 FP mark per image (range, 0–5 marks per image). Sensitivity of CAD was 91% (29 of 32) for well-visible nodules and 88% (28 of 32) for moderately subtle nodules. Sensitivities for subtle and very subtle nodules were 62% (18 of 29) and 39% (seven of 18), respectively. Ninety-one percent (21 of 23) of the nodules larger than 20 mm in diameter were detected by CAD. For smaller nodules, detection rates were lower, with a detection rate of 62% (20 of 32) for nodules between 15 and 20 mm in diameter and a detection rate of 77% (36 of 47) for nodules.
between 10 and 15 mm in diameter. The CAD system localized 56% (five of nine) of the nodules smaller than 10 mm in diameter. On this data set, CAD reached an area under the AFROC curve of 0.656. Pooled over all readers, the readers missed 239 nodules without CAD. CAD detected 53% (127 of 239) of these nodules (Table 1).

In total, CAD generated 196 FP detections in cases with negative finding (n = 189). Most FP findings of the CAD system were provoked by the anterior contour of the first rib (n = 29) or by hilar vascular shadows (n = 69) (Fig 1). Seventy-one (38%) of the cases without a nodule (normal cases) and 10 (9%) of the cases with a nodule did not contain any CAD mark.

**Observer Performance**

Total area under the AFROC curve for the observers was 0.812 without CAD versus 0.841 with CAD (P < .0001) (Fig 2). Observers improved their performance at a specificity that ranged between 80% and 100%, shown by an increase of the partial AUC from 0.125 to 0.133. All eight observers improved their performance with use of CAD (Figs 3, 4); individually, the differences reached significance for five observers (Table 2).

Comparing sensitivity and specificity without and with CAD at different thresholds of suspicion, the largest increase of sensitivity was seen at high confidence scores above 90 (Table 3). This was associated with the smallest loss in specificity. Going to lower thresholds, increase of sensitivity became smaller and loss of specificity became larger.

CAD detected 53% (127 of 239) of the nodules that were missed by the readers. Readers dismissed 55% (70 of 127) of these TP CAD candidates. If CAD led to modification of reader scores, it was beneficial in most cases (eg, leading in cases with a nodule to placement of new labels in 57 occasions and an increase of suspiciousness score in 220 occasions). These positive
Mean scores of suspicion for the presence of a lesion were 80, 66, 46, and 12 (on a scale from 0 to 100) for the four subsets of nodules with decreasing conspicuity. This factor indicates that effects counteracted the negative effects caused by placing new labels \( (n = 92) \) or in an increase of suspiciousness \( (n = 66) \) in cases without a nodule (normal cases) (Table 4).

**Lesion Conspicuity**

Considering the markings by all eight observers, they located 649 of 888 nodules (eight readers \( \times \) 111 nodules) without CAD and 704 nodules with CAD.

**Figure 3**: A 12-mm adenocarcinoma in the left upper lobe in a 62-year-old man. The lesion was categorized as subtle (category 3). Without CAD, only one observer marked the tumors. With CAD, the tumor was correctly identified, and this action triggered marking of the cancer by four other observers, as well. Right hilar vascular shadow caused an FP marking of the CAD system. (a) Original radiograph. (b) BSI. (c) Radiograph with CAD marks (circles). (d) Cross section of coronal CT scan.
subjective categorization of conspicuity well correlated with the degree of observers’ suspicion. For the four subcategories of conspicuities, readers reached a mean sensitivity of 96% (247 of 256), 86% (220 of 256), 65% (150 of 232), and 22% (32 of 144) for subtlety categories 1–4, respectively. With CAD, sensitivities increased

Figure 4: A 23-mm non–small-cell lung cancer in the apex of the right lung in a 65-year-old woman. The lesion was categorized as moderately subtle (category 2). None of the observers marked the lesion on the basis of the standard chest radiograph and the BSI. After CAD review, four observers marked the tumor. (a) Original radiograph. (b) BSI. (c) Radiograph with CAD marks (circle). (d) Cross section of coronal CT scan.
to 98% (252 of 256), 92% (235 of 256), 75% (175 of 232), and 29% (42 of 144), respectively. AFROC analysis showed that improvement for moderately subtle nodules (categories 2 and 3, respectively) reached significance, with \( P = .01 \), respectively. AFROC analysis showed that improvement for moderately subtle nodules (categories 2 and 3, respectively) reached significance, with \( P = .01 \), respectively.

The results of this study show that CAD provides additional value beyond previously documented beneficial effect of BSIs alone. Several previous studies had homogeneously documented a significant improvement of the detection of lung nodules by using BSIs (5,6,14). The results of this study also demonstrate that this CAD system provides further support for the detection of lung nodules when the initial image analysis was performed with BSIs. This point is noteworthy because previous studies in which researchers evaluated CAD alone had found variable results. Some found an increase in accuracy for the detection of lung nodules (7,15,16). Others found an increase in sensitivity but also a decrease in specificity, resulting in a negligible effect of CAD (8,9,17,18).

Investigators in three of the previously published CAD observer studies have evaluated a previous version of the same software we used in this study; in one study, they found a positive effect of CAD (16), and in the other two, they could not prove significant improvement by using CAD (9,18). In this study, we used an upgraded version of the CAD algorithm with a lower FP rate. The significant improvement with CAD that we found now, therefore, may be attributable to both a further improvement of the CAD algorithm and the availability of BSIs. On the other hand, it is remarkable that CAD was able to help further increase performance from an already high baseline performance with use of BSIs. The fact that, with CAD, we found nodules that were missed by all readers (three of seven lesions) or one study, they found a positive effect of CAD (16), and in the other two, they could not prove significant improvement by using CAD (9,18). In this study, we used an upgraded version of the CAD algorithm with a lower FP rate. The significant improvement with CAD that we found now, therefore, may be attributable to both a further improvement of the CAD algorithm and the availability of BSIs. On the other hand, it is remarkable that CAD was able to help further increase performance from an already high baseline performance with use of BSIs. The fact that, with CAD, we found nodules that were missed by all readers (three of seven lesions) or by most of the readers (12 of 25 lesions), even with study conditions in which readers were especially vigilant to detect nodules, underscores the potential of the CAD algorithm.

### Table 2

<table>
<thead>
<tr>
<th>Observer*</th>
<th>Without CAD</th>
<th>With CAD</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>Partial AUC</td>
<td>AUC</td>
</tr>
<tr>
<td>Radiologists</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer A, 13 years</td>
<td>0.812</td>
<td>0.125</td>
<td>0.840</td>
</tr>
<tr>
<td>Observer B, 3 years</td>
<td>0.820</td>
<td>0.126</td>
<td>0.853</td>
</tr>
<tr>
<td>Observer C, 5 years</td>
<td>0.857</td>
<td>0.143</td>
<td>0.879</td>
</tr>
<tr>
<td>Observer D, 17 years</td>
<td>0.748</td>
<td>0.110</td>
<td>0.792</td>
</tr>
<tr>
<td>Observer F, 17 years</td>
<td>0.830</td>
<td>0.117</td>
<td>0.848</td>
</tr>
<tr>
<td>Average</td>
<td>0.814</td>
<td>0.124</td>
<td>0.842</td>
</tr>
<tr>
<td>Residents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer E, 4 years</td>
<td>0.780</td>
<td>0.120</td>
<td>0.817</td>
</tr>
<tr>
<td>Observer G, 2 years</td>
<td>0.792</td>
<td>0.119</td>
<td>0.831</td>
</tr>
<tr>
<td>Observer H, 4 years</td>
<td>0.856</td>
<td>0.140</td>
<td>0.868</td>
</tr>
<tr>
<td>Average</td>
<td>0.809</td>
<td>0.126</td>
<td>0.839</td>
</tr>
<tr>
<td>All</td>
<td>0.812</td>
<td>0.125</td>
<td>0.841</td>
</tr>
</tbody>
</table>

Note.—The AUC and partial AUC were at a high specificity range between 80% and 100% without and with CAD. \( P \) values of the differences without and with CAD were computed with the Dorfman-Berbaum-Metz method for the whole AUC.

* Number of years is the number of years of experience or the years of residency.

### Table 3

<table>
<thead>
<tr>
<th>Decision Threshold*</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without CAD</td>
<td>With CAD</td>
<td>Difference</td>
<td>Without CAD</td>
</tr>
<tr>
<td>All markings</td>
<td>73.1</td>
<td>79.3</td>
<td>6.2</td>
<td>77.1</td>
</tr>
<tr>
<td>Observer markings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;30</td>
<td>68.4</td>
<td>73.6</td>
<td>5.2</td>
<td>86.2</td>
</tr>
<tr>
<td>&gt;50</td>
<td>61.5</td>
<td>67.2</td>
<td>5.7</td>
<td>92.0</td>
</tr>
<tr>
<td>&gt;70</td>
<td>53.7</td>
<td>59.6</td>
<td>5.9</td>
<td>96.3</td>
</tr>
<tr>
<td>&gt;80</td>
<td>41.8</td>
<td>48.4</td>
<td>6.6</td>
<td>97.6</td>
</tr>
<tr>
<td>&gt;90</td>
<td>23.6</td>
<td>32.4</td>
<td>8.8</td>
<td>99.0</td>
</tr>
</tbody>
</table>

* The decision threshold is for confidence scores of observer markings. For instance, a confidence level of greater than 30 for nodoles that were missed by observers and CAD were in the low-

### Discussion

The results of this study show that CAD provides additional value beyond previously documented beneficial effect of BSIs alone. Several previous studies had homogeneously documented a significant improvement of the detection of lung nodules by using BSIs (5,6,14). The results of this study also demonstrate that this CAD system provides further support for the detection of lung nodules when the initial image analysis was performed with BSIs. This point is noteworthy because previous studies in which researchers evaluated CAD alone found variable results. Some found an increase in accuracy for the detection of lung nodules (7,15,16). Others found an increase in sensitivity but also a decrease in specificity, resulting in a negligible effect of CAD (8,9,17,18).

Investigators in three of the previously published CAD observer studies have evaluated a previous version of the same software we used in this study; in one study, they found a positive effect of CAD (16), and in the other two, they could not prove significant improvement by using CAD (9,18). In this study, we used an upgraded version of the CAD algorithm with a lower FP rate. The significant improvement with CAD that we found now, therefore, may be attributable to both a further improvement of the CAD algorithm and the availability of BSIs. On the other hand, it is remarkable that CAD was able to help further increase performance from an already high baseline performance with use of BSIs. The fact that, with CAD, we found nodules that were missed by all readers (three of seven lesions) or by most of the readers (12 of 25 lesions), even with study conditions in which readers were especially vigilant to detect nodules, underscores the potential of the CAD algorithm.
Our results exceed the findings recently reported in a study in which the effect of BSIs and that of dual energy, both secondarily aided by CAD, were compared (19). In that study, the authors found a further increase of sensitivity when they added CAD as an additional tool; however, no significant increase of figure of merit was seen, indicating that the increased sensitivity was nullified by a loss of specificity. Other methodological differences between that study and our study refer to reading methods, statistical evaluation, and version of CAD software.

The improvement of reader performance with CAD that we found was significant for the full range of specificities (0%–100%) but also yielded uniform improvement while preserving a high specificity, as indicated by the pAUC at specificities between 80% and 100%. It has to be noted that it is ruled out by our statistical analysis that readers could improve their performance by being able to mark several locations, although they knew that the radiographs contained only one nodule: A high number of cases with normal findings with FP marks inevitably would have led to a performance decrease by using AFROC analysis. Also, AFROC analysis correlated lesion location with locations of reader marks. We offered the option to mark several locations to encourage the readers to scrutinize the image for potential lesions completely and minimize the effects of “satisfaction of search.”

With respect to lesion conspicuity, we found the largest effect of CAD for moderately subtle and subtle lesions. Well-visible and very subtle nodules were relatively less affected by the availability of CAD. These findings can be explained by the fact that the sensitivity of the CAD system for the moderately subtle and the subtle nodules was quite high (ie, 88% and 62%, respectively), providing sufficient chances for further reader improvement. For well-visible nodules, reader performance without CAD and CAD stand-alone performance were both high. For the category of very subtle lesions, the sensitivity of CAD was much better than that of the readers (39% vs 22%). Therefore, readers could have taken advantage of the availability of CAD similarly, as they did for the moderately subtle and subtle lesions. The fact that readers were not able to use CAD more beneficially appears to confirm the previously discussed issue that readers find it difficult to differentiate TP from FP lesions, especially for very subtle lesions.

In total, readers dismissed a substantial number of TP CAD candidates (n = 70) (Table 1), accounting for 55% of the TP CAD candidates (n = 127) made on lesions missed at baseline (n = 239). This percentage is similar to previously published figures ranging from 55% to 80% (9,16,20). These high percentages of dismissed TP CAD candidates show potential of further improvement with use of CAD. Although the majority of FP CAD marks may be easily dismissed, discrimination between a TP CAD mark for a low- conspicuity lesion and an FP CAD mark is apparently very difficult for human observers. Availability of BSIs might have been helpful for differentiating TP from FP candidates, but it did not help to minimize or even eliminate the fact that TP candidates were not accepted by the readers. We, therefore, plan to explore other means to use CAD (eg, by adding weighting factors or likelihood factors to strengthen the influence of CAD on the reader decision).

Our study had some limitations. The study group had a much higher prevalence of subjects with disease (subjects with a solitary nodule) than that encountered in clinical practice. Furthermore, study conditions in which readers are focused on a specific detection task do not reflect the clinical situation.

The effect of BSIs on evaluating images with other than nodular disease is yet unknown. It could be stated that oversight represents a bigger issue in clinical practice than do study conditions, where observers are explicitly focusing on the search for nodules. With respect to failure of detection due to oversight, it is therefore likely that beneficial effects of BSIs and CAD become more prominent in clinical practice. Finally, none of the readers had experience with CAD in chest radiography. Although we provided a training set of 40 cases with instant feedback, this might have been insufficient to help the readers become familiar with BSIs and CAD.

We demonstrated that CAD has an additional beneficial effect on the detection of pulmonary nodules beyond the effect of BSIs alone. Even though baseline performance was optimized by the availability of BSIs, radiologists were able to uniformly improve their detection performance.

Acknowledgments: The authors acknowledge L. Meiss, MD (Meander Medical Center, Amersfoort, the Netherlands), M. indoïf-Tas, MD, and L. Peters-Bax, MD, PhD (both from Radboud University Medical Center, Nijmegen, the Netherlands), and L. G. B. A. Quekel, MD, and L. Peters-Bax, MD, PhD (both from Meander Medical Center, Amersfoort, the Netherlands), for their participation in the observer study.

Disclosures of Conflicts of Interest: S.S. Financial activities related to the present article: participated in a research project funded with a research grant from Riverain Technologies. Financial activities not related to the present arti-
Computer-aided Detection Improves Pulmonary Nodule Detection

THORACIC IMAGING: Computer-aided Detection Improves Pulmonary Nodule Detection Schalekamp et al


References