Lung cancer screening: evaluation of radiologists and different computer assisted detection software (CAD) as first and second readers for lung nodule detection at different CT dose levels.

Poster No.: C-2614  
Congress: ECR 2013  
Type: Scientific Exhibit  
Authors: A. Christe\(^1\), L. Leidolt\(^2\), A. Huber\(^1\), P. Steiger\(^1\), Z. Szucs-Farkas\(^3\), J. ROOS\(^4\), J. Heverhagen\(^1\), L. Ebner\(^1\); \(^1\)Bern/CH, \(^2\)Bern, De/CH, \(^3\)Biel/ Bienne/CH, \(^4\)Winterthur/CH  
Keywords: Radioprotection / Radiation dose, Thorax, Respiratory system, CAD, CT, Computer Applications-Detection, diagnosis, Dosimetry, Cancer, Multidisciplinary cancer care  
DOI: 10.1594/ecr2013/C-2614

Any information contained in this pdf file is automatically generated from digital material submitted to EPOS by third parties in the form of scientific presentations. References to any names, marks, products, or services of third parties or hypertext links to third-party sites or information are provided solely as a convenience to you and do not in any way constitute or imply ECR’s endorsement, sponsorship or recommendation of the third party, information, product or service. ECR is not responsible for the content of these pages and does not make any representations regarding the content or accuracy of material in this file.

As per copyright regulations, any unauthorised use of the material or parts thereof as well as commercial reproduction or multiple distribution by any traditional or electronically based reproduction/publication method is strictly prohibited.

You agree to defend, indemnify, and hold ECR harmless from and against any and all claims, damages, costs, and expenses, including attorneys’ fees, arising from or related to your use of these pages.
Please note: Links to movies, ppt slideshows and any other multimedia files are not available in the pdf version of presentations.
www.myESR.org
Purpose

The purpose of this study was the evaluation of two radiologists and three different CAD-systems as first and second readers for lung cancer screening at various CT dose levels including low kilo-voltage settings, to find the best reader pairing at the lowest acceptable radiation exposure level. Sensitivity of CT for lung nodules depends on many things like nodule size, nodule density, nodule location, image quality, reader and post-processing tools [1-7]. Most of these given variables may not be changed, except for the last ones. Therefore, it makes sense to improve these variables in order to increase detectability by a second reader, either with a second radiologist or computer assisted detection system. Wormanns et al. reported a 11 to 16% higher sensitivity, when a second reading was performed by a different radiologist. Previous investigators stated the usefulness of a computer assisted detection software (CAD) [6-10]. An increase of sensitivity between 2 and 22% was reported, when an additional CAD was used [6, 9, 11, 12]. However, Lee et al. [13] showed that the sensitivity of a CAD system as standalone tool (81%) did not significantly differ from that of radiologists (85%). Radiologists were more sensitive at detecting nodules attached to other structures, whereas CAD was better at detecting isolated nodules and those that were # 5 mm in diameter [13]. But until now there is little results available on how accurate a CAD is working at low dose levels in combination with radiologists or with a second CAD system: Hein et al. described a feasibility of low dose CAD at 5mAs/120kVp and Das at 10 mAs/120kVp.

Furthermore, previous studies have shown, that diagnostic image quality does not suffer at lower dose levels [1,3,14]. To maintain diagnostic quality in the detection of lung nodules tube currents can be reduced to well below 100 mAs at a constant voltage of 120 kVp [2,3,14,15,16]. Some authors even proposed threshold tube currents of 20 mAs (at 120 kVp) to detect ground glass nodules, alveolar consolidation and lung nodules [3,16]. To decrease tube voltage is another strategy to reduce radiation dose in CT. There are still high voltages (100-140 kVp) widely applied in radiology, although results from vascular studies suggest that a reduction to 80 kVp is possible [17-20]. There is very little data on the combination of low kVp and mAs for lung nodule detection. In screening CT the prevalence of part solid or ground glass nodules is reported in the literature between 13% to 40% [21-23]. Furthermore, peripheral pulmonary adenocarcinoma represents up to 35 % of all lung cancer and usually demonstrates as GGN in the beginning [24, 25]. Therefore it is important to intersperse ground glass nodules (GGN) to simulate a screening setting.

Methods and Materials

Image acquisition:
All CT scans were performed on a 64-row multi-detector CT scanner Somatom Sensation 64 (24x1.2mm, pitch 0.8, slice thickness 1.5mm by Siemens, Forchheim, Germany). Automatic current modulation (CareDose4D) was used for raw data acquisition and filtered back projection was used for image reconstruction. Three tube current-time products of 25, 50 and 100 mAs (Miliampère-second) were combined with 80, 100, 120 kVp (peak Kilovoltage), resulting in 9 different pairings: 8 low dose levels were compared against the standard dose CT of 100mAs/120kVp.

Phantom and nodules

An anthropomorphic lung phantom (Chest Phantom N1, Kyoto Kagaku Co., Ltd, Kyoto, Japan) was used based on an average Japanese man with a body weight of 70 Kg. It is an accurate life-size anatomical model of a male human torso with synthetic pulmonary vessels, which are spatially traceable (right and left), heart, trachea and abdomen (diaphragm) block. Phantom size is 43 cm x 40 cm x 48cm (height). The soft tissue material (polyurethane, gravity 1.06) and synthetic bones (epoxy resin) have x-ray absorption rates close to those of human tissues. Arms-abducted position of the torso suits CT scanning.

Four artificial solid nodules with a density of +100 Hounsfield units [HU] (5, 8, 10, 12 mm) and 4 artificial ground glass nodules (GGN) with a density of -630 HU (5, 8, 10, 12 mm) by Kyoto Kagaku Co. (Ltd, Kyoto, Japan) were used. For each dose level 100 nodules were randomly distributed into 40 phantoms. The lung-segment, -side and the size were also randomly assigned to each nodule. Therefore, the average nodule size reached the mean of the four sizes (8.75 mm). The 100 nodules consisted of 75 solid and 25 ground glass nodules, simulating the screening prevalence. To prevent recognition bias nodules in the phantom were rearranged for each exposure level, equaling a total of 900 nodules (675 solid and 225 GGN) for the 9 examined protocols. Each nodule was registered on an excel-sheet, indicating exposure level, exact location by slice position of nodule center, side and density of nodule (answer key).

Image analysis

Images were sent to a PACS-workstation (Picture Archiving and Communication System R11.4.1, 2009; Philips, Best, Netherlands; Sectra, Linkoping, Sweden) and 3 CAD-workstation: CAD1 was syngoCT-CAD (Siemens, Erlangen, Germany), CAD2 was LMS-Lung/Track (Lesion-measurement-solution Version 6, Median-Technologies, Valbonne, France) and CAD3 was Lung Nodule CAD (Prototype, Philips, Best, Netherlands). Two blinded radiologist with 4 and 2 years experience in chest CT imaging read the scans on the PACS-workstation. Readers had to register the lung nodules by indicating the location (side of the lung, exact slice position of the nodule center). A third radiologist with 2 years of chest CT experience run the CAD analysis. He fed the CAD with the 1.5 mm slices and documented the true positive, false negative and the false positive nodules that were found by each CAD by comparing each CAD-positivity with the answer key.
Statistic analysis.

The 2 radiologists and the 3 CADs were analyzed individually or as standalone tool as first reader at each dose level. Sensitivities of these 5 first readers were compared using McNemar test [26]. The sensitivity of each low dose level was then tested against the standard dose level using the Chi-square test (Z-test of proportions) [27].

Each of the 5 readers was paired with the other 4 readers to test the double-reading outcome at each dose level: the number of nodules detected by at least one reader of the pairing defined the combined sensitivity. The combined sensitivity of the ten pairings was analyzed with McNemar test for each dose level. Each low dose level was tested against the standard dose with the chi-square test (Z-test of proportions) [27] to find the lowest acceptable dose level without loss of sensitivity compared to the standard CT.

Inter-observer agreement was calculated for the detection of the true positive nodules and the classification into ground glass and solid nodules were taken into account for each dose level separately. Radiologists were compared among each other and agreement between CADs was calculated separately. In addition, agreement of each radiologist with each CAD was calculated. Mean agreements for the radiologists, for the CADs and the radiologists against CAD were determined. Inter-observer comparison was performed calculating agreements levels using Fleiss' Kappa statistics [28, 29]; Kappa strength of agreement: < 0.20 poor, 0.21 - 0.40 fair, 0.41 - 0.60 moderate, 0.61 - 0.80 good, 0.81 - 1.00 very good agreement [30]. Inter-observer agreement between 'radiologist-radiologist' and 'radiologist-CAD' was tested using the comparison of correlation coefficient [27].

Results

Individual nodule detection (first reader):

Individual standard CT sensitivities for lung nodules were 92.2%, 90.3%, 82.0%, 83.3% and 67.6% for reader 1, 2, CAD1, CAD2, CAD3, respectively (Table 1). Only the sensitivity of CAD3 is significantly lower than radiologists' sensitivity (p<0.001). Sensitivity at lowest dose level decreases to 87.8%, 82.4%, 76.6%, 64.1%, and 67.8% for reader 1, 2, CAD1, CAD2, CAD3, respectively (Fig. 1). Only CAD2 showed a significant loss of sensitivity at lowest dose level (64.1%, p-value=0.028).

Combined nodule detection (first and second reader)

Combining reader 1 and 2 sensitivity increased to 96.8%, which was a significant rise for reader 2 (p=0.016) but not for reader 1 (p=0.063). This combined sensitivity dropped insignificantly to 94.0% at the lowest dose level. Highest sensitivities at standard dose level - between 96.6% and 99.0% - were achieved by combining any radiologist
with any CAD with a significant increased detection rate for both sides (p<0.05). This combination worked at all dose levels with increased sensitivities for both sides (Table 1). No significant loss of detectability with this combination was found at lower dose levels compared to standard CT. Combining any two CADs, sensitivities were significantly lower (between 83.3% and 89.4% for standard CT and between 75.6% and 86.4% for 25 ref mAs/80 kVp; p<0.001). Sensitivity for solid nodules was around 90% for both radiologists and CADs, but CADs detected significantly less GGN compared to radiologists: At standard CT sensitivity for GGN was 96.1%, 88.3%, 46.2%, 59.1% and 8.0% for reader 1, 2, CAD1, CAD2, CAD3, respectively.

Inter-observer agreement

Inter-observer agreement of the two radiologists was excellent with a Kappa level of 0.91±0.05 (standard error) at standard CT (Table 2). Kappa strength of agreement did not drop significantly at lower dose levels (Kappa: 0.8 to 0.93). Inter-observer agreement was lower between radiologists and CADs. The mean Kappa strength of agreement between radiologists and CADs was 0.58 ±0.11 (p<0.0001) at standard CT. This agreement did not change significantly for the other dose levels. For all dose levels combined CAD3 scored the lowest agreements with the radiologists (0.45±0.1); CAD1 and CAD2 demonstrated an agreement of 0.67±0.09 and 0.59±0.11, respectively.

Conclusion

Our results suggest that lung cancer screening should be performed by a radiologist using any CAD. Combined CT-sensitivity does not drop significantly at lower dose levels, including 25mAs/80kVp with an average dose length product of 22.3 mGycm, equaling an equivalent dose of 0.3 Millisievert. In a previous study (under review), we stated that a minimum dose level of 25 ref mAs/100kVp is needed to detect both solid and ground glass nodules. This threshold level applies to GGNs when analyzed separately, but in a lung cancer screening setting where the prevalence of GGNs is set to 25% [21-23] the voltage could be reduced to 80 kVp.

The individual performance of both radiologists and CADs did not drop significantly at lowest dose level, except for CAD3. The interspersed GGNs were significantly less detected by CAD, especially by CAD3, which probably strongly influenced the significance level. Otherwise, our results are consistent with previous studies stating that low dose CT is not significantly affecting the performance of CAD [5, 15, 31]. Sensitivity of radiologists is higher due to the better detectability of GGNs compared to CAD. Best double-reading pair was a radiologist with a CAD; sensitivity was always over 96% at standard CT independent of the radiologist and the CAD, when at least one of them had to detect the nodule. This pairing was better than any combination of CADs or the pairing of two radiologists. The radiologists - although scoring high individual sensitivities - detected the same nodules (excellent inter-observer agreement), therefore the addition of a
second radiologist was lower. The CADs had lower sensitivities; therefore a combination of two CADs did not reach the combined sensitivity of a radiologist and a CAD.

Inter-observer agreement between radiologists was significantly higher than the agreement between radiologists and CADs, suggesting that more additional information (sensitivity) would come from a CAD as a second reader than from a second radiologist, as long as the sensitivity of the CAD is acceptable. Although radiologists had better sensitivities for lung nodules, the combined sensitivity rose only for one reader significantly when a second radiologist performed the double reading. The benefit of double reading by radiologists (+4.5% and 6.5%) was not as high as described by Wormanns et al (11-15%) probably because of their lower baseline sensitivity of 63% [5], which leaves more room for additional sensitivity. When the CAD was the second reader, the combined sensitivity was significantly higher than the individual sensitivity for both radiologists almost for all dose levels. The low agreement between CAD and radiologists overcomes the slightly lower sensitivity of CAD to detect lung nodules. The prototype CAD of Philips scored the lowest sensitivities and the lowest agreement with the radiologists and reached the same combined sensitivities as the other CADs. We therefore confirm the additional value of a CAD stated by several authors [6, 9, 11, 12]. In our study the mean radiologists' sensitivity gained using a CAD was 6.9% and 10.3% for standard and lowest dose examined. This additional sensitivity is not as much as the 14% stated by Li et al [6], but their baseline sensitivity was 52%, compared to our baseline sensitivity of 82-88% at lowest dose and 90-92% at standard dose. Furthermore, the added GGNs with their lower detectability may have influenced the benefit for the worse.

Low inter-observer agreement and high sensitivities on both observer sides would be the optimal reader pairing and the ratio of sensitivity divided by the agreement (mean sensitivity of both sides/Kappa) could be an important indicator for additional nodule detectability of a CAD-system to radiologists.

Limitations:

We did not analyze the additional value of maximum intensity projection as second reader. This is subject of ongoing studies at our institution.

At the time of our data acquisition there was only filtered back projection available for image reconstruction at our institution. Meanwhile the iterative image reconstruction reduces radiation exposures in our daily routine: preliminary results indicate a dose reduction of 50% at the lower dose levels with the same image quality and noise; meaning a potential low dose CT feasibility below 10 ref mAs/80kVp.

Acknowledgement:

The Bernese Cancer League, the Jubilee Foundation of SwissLife and the Swiss foundation "Fight against Cancer" financially supported this study.
References


17 Szucs-Farkas Z, Schibler F, Cullmann J et al. Diagnostic accuracy of pulmonary CT angiography at low tube voltage: intraindividual comparison of a normal-dose protocol at 120 kVp and a low-dose protocol at 80 kVp using reduced amount of contrast medium in a simulation study. Am J Roentgenol 2011;197(5)

18 Holmquist F, Nyman U. Eightypeak kilovoltage 16-channel multidetector computed tomography and reduced contrast-medium doses tailored to body weight to diagnose pulmonary embolism in azotaemic patients. Eur Radiol 2006; 16:1165-1176


23 Li F, Sone S, Abe H, MacMahon H, Doi K. Malignant versus Benign Nodules at CT Screening for Lung Cancer: Comparison of Thin-Section CT Findings. Radiology 2004;23:3 793-798


30 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174


**Personal Information**

Christe A\(^1\), Leidolt L\(^1\), Huber A\(^1\), Steiger P\(^1\), Szucs-Farkas Z\(^2\), Roos JE\(^3\), Heverhagen J\(^1\), Ebner L\(^1\)

\(^1\)Department of Radiology, University Hospital, Insel, Bern, Switzerland

\(^2\)Radiology, Hospital Center of Biel, Switzerland

\(^3\)Department of Radiology, Duke University, Durham, North Carolina, USA