Low Dose Computed Tomography of the Chest: Applications and Limitations

Hester Alexandra Gietema

ISBN: 978-90-393-4459-0

Cover photo: Antonio Nunes, African Landscape, Kenia
Lay-out: Laurens van Piggelen
Printed by: Gildeprint drukkerijen BV, Enschede, The Netherlands

Publication of this thesis was financially supported by: Philips Medical Systems, Siemens AG Medical Solutions, Röntgen Stichting Utrecht, Boehringer Ingelheim bv en Pfizer bv

Low Dose Computed Tomography of the Chest: Applications and Limitations

Lage dosis computed tomography van de thorax:
toepassingen en beperkingen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. W.H. Gispen,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op dinsdag 6 maart 2007 des middags te 2.30 uur

door

Hester Alexandra Gietema

geboren op 1 november 1976
te Groningen

Promotoren: Prof. dr. W.M. Prokop
Prof. dr. J-W.J. Lammers

Co-promotor: Dr. P. Zanen

Zolang je durft te dromen, komt vanzelf voor jou dat ene moment…..

Uit de schaduw - IOS

Voor mijn ouders

CONTENTS

PART 3 LOW DOSE CHEST CT IN THE DETECTION OF EMPHYSEMA:

Chapter 1

General introduction

GENERAL INTRODUCTION

Since the introduction of computed axial tomography by Godfrey Newbold Hounsfield in 1971 and the introduction of the whole-body scanner in 1974, computed tomography (CT) has been improved massively. The introductions of the spiral scanner in 1990 and the multidetector-row CT in 1998 have revolutionized CT imaging by making volumetric scanning feasible and limiting the scanning time. These developments have increased the number of applications to for example cardiac CT and CT angiography. As a result of these advances there has been an increase in the number of CT examinations and on average, more and thinner slices are obtained at each examination. In the mean time, the applied radiation dose has increased exponentionally. While CT accounts for about 10% of the radiological procedures, it accounts for about 67% of the medical radiation dose [1].

In contrast to plain X-rays, CT is not hampered by over projection. Although the benefits of more accurate data have been proven, the major disadvantage of CT is the intrinsic risk of radiation dose. With present standard techniques, between 4 and 8 mSv effective dose is delivered (UNSCEAR rapport 2001), which leads to a calculated risk of dying from radiation-induced cancer of 2-4 in 10.000 (30-year old) individuals. The risk, however, is strongly age-dependent and decreases sharply for elderly individuals (up to a factor of 5), while it increases for children (up to a factor of 5).

The exponential increase in applied radiation dose is concerning more and more radiologists [2-6] and has lead to the implementation of the "As Low As Reasonably Achievable" (ALARA) principle [7]. According to this principle, for each patient the applied radiation dose should be kept to the minimum, which is necessary to obtain enough image quality to get the required diagnostic information.

PART 1

In clinical routine, many patients receive a plain chest X-ray for suspected pulmonary diseases. In many cases, this chest X-ray is not conclusive and a chest CT-scan is ordered to get the required diagnostic information. This approach results in more radiation dose applied to the patient. Moreover, many patients with chronic, but benign chest disease receive regular X-rays to follow-up the status of their disease. For these patients a CT-scan could provide more information, but the radiation dose level of CT-scans is a major limiting factor. In chapter 2, we investigate if an ultra-low dose CT-scan can provide the diagnostic information of a standard dose CT-scan at the radiation dose of a plain chest X-ray in two directions. We report the results of a study comparing both plain chest X-ray and ultralow-dose chest CT to the standard dose chest CT performed on the same day, in order to show that ultralow-dose chest CT could be performed in daily practice.

The early detection and the study of the natural progression of emphysema is an other application for low dose CT. Emphysema is anatomically defined as a permanent abnormal enlargement of the alveoli distal to the terminal bronchioles. This anatomical definition makes CT an ideal tool to non-invasively detect the presence and the extent of lung tissue destruction. In 1988, Müller and co-workers described a method to highlight areas within the lungs showing an attenuation below a prefixed threshold [8]. The extent of the highlighted areas can be expressed as percentage of total lung area. They demonstrated that the extent of highlighted low-attenuated areas showed a good correlation with the extent of macroscopic emphysema. The method was originally described and validated for axial images, but Park et al showed a good correlation with three-dimensional data [9]. Since emphysema is a common and debilitating disease, CT could play a role in the early detection and secondary prevention of emphysema. However, also for this purpose the radiation dose should be kept as low as achievable. The detection of emphysema has been shown to be feasible on low dose scans by Shaker et al [10], but the lack of contrast between destructed areas and normal lung parenchyma makes the automated technique sensitive to image noise. In chapter 3, we demonstrate that massive image noise raises emphysema scores, especially for low amounts of lung destruction. For reliable diagnostic results, application of a dedicated noise reduction filter is required before the extent of emphysema can reliably be quantified.

PART 2

For most patients, the benefits of CT will exceed the harms. However, CT is now increasingly being used for screening purposes, being applied to healthy subjects. In this particular setting, the development of low-dose scanning protocols is of importance. With the introduction of the multi-detector row scanners protocols applying low radiation dose have become feasible [2]. The high contrast between air and many pathological structures in the lungs, make the chest an area where substantial dose reduction is achievable. This option is already applied in the lung cancer screening trials performed on several locations around the world [11-13]. In this particular situation the increased noise has been shown to have limited influence on the diagnostic task: the detection and measurement of pulmonary nodules [14].

Part two of this thesis concerns the reproducibility of the semiautomated quantification of the volume of pulmonary nodules, which can show early lung cancer and the management of these nodules. Lung cancer is today the most frequent cause of cancer deaths in the Western world. At time of diagnosis more than two-third of these people have locally advanced or metastatic disease, and their poor prognosis is due to late diagnosis and lack of effective treatment of metastatic disease. Less than 15% of the patients are surviving at least 5 years. However, when lung cancer is diagnosed in early stages and curative treatment is

still possible, 85% of the patients are surviving at least 5 years [15]. The well-described population at-risk, the options to treat lung cancer in early stage and the possibility to detect lung cancer in early stage without doing too much harm, have made screening for lung cancer with low dose CT an attractive way to decrease the lung cancer specific mortality.
The Dutch Belgian Lung Cancer Screening Trial (NELSON) is a population-based randomized multi-center lung cancer screening trial that studies about 16,000 current and former heavy smokers. Selection of participants was performed by sending a questionnaire about smoking history and other health-related issues to citizens between 50 and 75 years of age who lived in the areas around the participating centers. Among the respondents, subjects meeting the inclusion criteria of a minimum of 16 cigarettes/day for 25 years or 11 cigarettes/day for 30 years were asked to participate in the trial. After being informed about, among others, the radiation dose that was exposed to the participants, those who gave written informed consent were equally randomized to either the screening arm or the control arm. Participants with a moderate or poor self-reported health status who were unable to climb two flights of stairs were excluded from participation.
Early lung cancer is detected as small pulmonary nodule. Chapter 4 describes the management of pulmonary nodules detected during the NELSON-trial. Since about 70% of the heavy smoking population show such nodules and only a few of these nodules turn out to be malignant, a simple non-invasive way to distinguish the malignant nodules from the benign ones is mandatory. Therefore, in lung cancer screening trial growth of nodules is used as main feature to detect the nodules with a high suspicion to be malignant. However, measurement errors can interfere with the calculation of growth rates. In chapter 5, we report the influence of the observer on the variability in volume measurements of isolated pulmonary nodules performed with a semiautomated, FDA approved, commercially available and widely used lung nodule volume program. In chapter 6, we quantify the extent of measurement variability of isolated pulmonary nodules caused by variables that are changed when a new CT is performed. This study was performed on patients of the outpatient department of oncology with pulmonary metastases shown on previous imaging.

PART 3

The third part of this thesis concerns the detection of an other major smoking-related disease: chronic obstructive pulmonary disease (COPD). COPD is after cardio-vascular disease the second smoking-induced disease in terms of prevalence of morbidity and mortality [16]. COPD is a heterogeneous disease comprising mucosal thickening of the bronchioles, resulting in airflow limitation and emphysema, anatomically defined as a permanent enlargement of the terminal bronchioles and alveoli [17]. In clinical practice computed tomography (CT) is used to detect emphysema, showing good correlations with histology [18-21].

The population enrolled in lung cancer screening trials is also the population at-risk for developing COPD. Therefore, lung cancer screening trials form a good environment to study the prevalence and the natural course of COPD in an asymptomatic population.
Studies correlating pulmonary pathophysiology examined with function tests and pathology demonstrated on CT, have shown moderate to good results [18;22], but smokers with parenchymal damage demonstrated with computed tomography (CT) often have no reduced airflow function [23;24]. Despite this limitation, the detection of emphysema is still based on the FEV_1 [25].
The method of highlighting voxels with an abnormally low attenuation is hypothesized to be a simple method to monitor lung tissue destruction [5;6], but little is known about the preciseness of automatically obtained results. In chapter 7 the reproducibility of the automated quantification of low-attenuation areas in the heavy current and former smokers participating in the screening arm of the NELSON-trial is reported.
In chapter 8 we describe a study performing pulmonary function tests in a randomly selected sample of the participants of the NELSON-trial who underwent a baseline scan in the University Medical Center in Utrecht. Results were compared to the extent of moderate and severe emphysema detected on CT of these subjects. Diffusion testing is a more dedicated method to detect lung tissue destruction than spirometry, so it was not surprising that D_{CO} turned to be the most sensitive parameter to detect emphysema shown on CT. However, D_{NO} was hypothesized to be even more sensitive, since NO has a stronger affinity for haemoglobin than CO. In chapter 9 we investigate the ability of NO-diffusion to detect emphysema demonstrated on CT and compared the results to capability of D_{CO}/V_A to detect emphysema demonstrated on CT.
Parr et al have shown that the distribution pattern of the lung tissue destruction significantly influences the extent of airflow obstruction and gas exchange impairment in a group of patients with $alpha_1$-antitrypsine deficiency [26]. $Alpha_1$-antitrypsine deficiency is a rare disease, but results in severe emphysema. However, emphysema in these patients results from a different pathophysiologial process than smoking-related emphysema, making the results of this study not simply transferable to the smoking population. We conducted a study investigating the impact of the distribution pattern of smoking-related emphysema on airflow limitation and gas exchange impairment in the population of current and former heavy smokers, participating in the NELSON-project. The results are reported in chapter 10.

REFERENCE LIST

(1) Mettler FA, Jr., Wiest PW, Locken JA et al. CT scanning: patterns of use and dose. J Radiol Prot 2000; 20(4):353-359.
(2) Mayo JR, Hartman TE, Lee KS et al. CT of the chest: minimal tube current required for good image quality with the least radiation dose. AJR Am J Roentgenol 1995; 164 (3):603-607.
(3) Rogers LF. Radiation exposure in CT: why so high? AJR Am J Roentgenol 2001; 177 (2):277.
(4) Haaga JR. Radiation dose management: weighing risk versus benefit. AJR Am J Roentgenol 2001; 177(2):289-291.
(5) Dirksen A, Friis M, Olesen KP et al. Progress of emphysema in severe alpha 1-antitrypsin deficiency as assessed by annual CT. Acta Radiol 1997; 38(5):826-832.
(6) Parr DG, Stoel BC, Stolk J et al. Validation of computed tomographic lung densitometry for monitoring emphysema in {alpha}1-antitrypsin deficiency. Thorax 2006;thx.
(7) Prasad KN, Cole WC, Haase GM. Radiation protection in humans: extending the concept of as low as reasonably achievable (ALARA) from dose to biological damage. Br J Radiol 2004; 77(914):97-99.
(8) Muller NL, Staples CA, Miller RR et al. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988; 94(4):782-787.
(9) Park KJ, Bergin CJ, Clausen JL. Quantitation of emphysema with three-dimensional CT densitometry: comparison with two-dimensional analysis, visual emphysema scores, and pulmonary function test results. Radiology 1999; 211(2):541-547.
(10) Shaker SB, Dirksen A, Laursen LC et al. Short-term reproducibility of computed tomography-based lung density measurements in alpha-1 antitrypsin deficiency and smokers with emphysema. Acta Radiol 2004; 45(4):424-430.
(11) Diederich S, Wormanns D, Heindel W. Lung cancer screening with low-dose CT. Eur J Radiol 2003; 45(1):2-7.
(12) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.
(13) Swensen SJ, Jett JR, Hartman TE et al. Lung cancer screening with CT: Mayo Clinic experience. Radiology 2003; 226(3):756-761.
(14) Karabulut N, Toru M, Gelebek V et al. Comparison of low-dose and standard-dose helical CT in the evaluation of pulmonary nodules. Eur Radiol 2002; 12(11):2764-2769.
(15) The International Early Lung Cancer Action Program. Survival of Patients with Stage I Lung Cancer Detected on CT Screening. N Engl J Med 2006; 355(17):1763-1771.
(16) Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. The Lancet 2003; 362(9387):847-852.
(17) The definition of emphysema. Report of a National Heart, Lung, and Blood Institute, Division of Lung Diseases workshop. Am Rev Respir Dis 1985; 132(1):182-185.
(18) Bankier AA, Madani A, Gevenois PA. CT quantification of pulmonary emphysema: assessment of lung structure and function. Crit Rev Comput Tomogr 2002; 43(6):399-417.
(19) Coxson HO, Rogers RM, Whittall KP et al. A quantification of the lung surface area in emphysema using computed tomography. Am J Respir Crit Care Med 1999; 159(3):851-856.
(20) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.
(21) Uppaluri R, Mitsa T, Sonka M et al. Quantification of pulmonary emphysema from lung computed tomography images. Am J Respir Crit Care Med 1997; 156(1):248-254.
(22) Nakano Yasu, Muro Shig, Sakai Hiro et al. Computed Tomographic Measurements of Airway Dimensions and Emphysema in Smokers . Correlation with Lung Function. Am J Respir Crit Care Med 2000; 162(3):1102-1108.
(23) Clark KD, Wardrobe-Wong N, Elliott JJ et al. Patterns of Lung Disease in a "Normal" Smoking Population : Are Emphysema and Airflow Obstruction Found Together? Chest 2001; 120(3):743-747.

(24) Gurney JW, Jones KK, Robbins RA et al. Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers. Radiology 1992; 183(2):457-463.
(25) www.goldcopd.com 2005.
(26) Parr DG, Stoel BC, Stolk J et al. Pattern of Emphysema Distribution in {alpha}1-antitrypsin Deficiency Influences Lung Function Impairment. Am J Respir Crit Care Med 2004.

Part 1

Low Dose Chest CT: Basic Studies

Chapter 2

Diagnostic information of CT at the radiation dose level of a chest X-ray: A feasibility study

Hester Gietema
Cornelia Schaefer-Prokop
Ieneke Hartmann
Inge van den Berk
Jan-Willem Lammers
Mathias Prokop

Preliminary Report

ABSTRACT

PURPOSE

To investigate whether ultralow-dose computed tomography (LDCT) can provide more reliable information than plain chest X-ray (CXR) in two directions.

MATERIAL AND METHODS

We enrolled 40 consecutive patients from the outpatient department of pulmonology with a wide range of chest abnormalities, referred for chest CT. Four patients had to be excluded from further analysis because of study incompletion. The clinically indicated CT-scans (RCT) were performed on a multirow-detector scanner (16x0.75mm collimation 120kVp;130mAs, 4.6mSv) followed by a low-dose CT-scan (LDCT) performed with identical parameters except for radiation dose (90kVp;20mAs, 0.3mSv). Posterior-anterior and lateral CXR (0.1mSv) were performed on the same day. Chest abnormalities were scored by three observers using a scale from 1 (definitely absent) to 5 (definitely present) for lesions located in the lungs (8 items), mediastinum (3 items) and pleura, chest wall and bones (=3 items). RCT was used as reference standard with only scores of 1 and 5. Both, agreement with RCT and reader confidence were used as surrogate for diagnostic performance of CXR and LDCT. Significance of difference was evaluated by McNemar χ^2-testing.

RESULTS

LDCT and RCT showed a higher agreement than CXR and RCT for pleura, chest wall & bones and mediastinum (all $p<0.001$). This was true for analysis of pooled reader data as well as for the three observers separately. Two of the three readers showed a higher confidence with LDCT (both $p<0.001$), while one observer showed a superior confidence with CXR ($p<0.001$).

CONCLUSION

Using agreement of diagnostic interpretation as surrogate of diagnostic performance we found a superior performance of LDCT as compared to CXR, both obtained at a comparable dose levels of less than 1mSv. Further evaluations are needed to define potential and limitations of LDCT with respect to specific diagnostic indications.

INTRODUCTION

Chest computer tomography (CT) has proven to be superior to plain chest radiographs (CXR) in detecting a broad range of abnormalities [1;2]. In contrast to plain radiographs, CT-scan is not hampered by the superimposition of anatomic structures and has better contrast and resolution. Advantages of CT are earlier detection (e.g., of pulmonary embolism, nodules, cancer screening), correct classification (e.g., of interstitial lung diseases, asbestosis, sarcoidosis), better staging (e.g., of bronchogenic cancer) and a better differential diagnosis (e.g., by high resolution CT) of abnormalities. The introductions of spiral and multirow-detector scanners have further improved the sensitivity of CT. At the moment, a substantial number of patients receive a consecutive CT-scan after CXR, when the chest X-ray has not been decisive for diagnosis.

The major disadvantage of CT is its radiation burden and the associated hypothetical cancer risk to the patient [3;4]. With present standard techniques, between 4 and 8 mSv effective dose is delivered (UNSCEAR rapport 2001), which leads to a calculated risk of dying from radiation-induced cancer of 2-4 in 10.000 (30-year old) individuals. The risk, however, is strongly age-dependent and decreases sharply for elderly individuals (up to a factor of 5), while it increases for children (up to a factor of 5). For this reason, the development of low-dose scanning protocols is of importance, especially as CT is now increasingly being used for screening purposes and imaging of non-malignant disease such as pulmonary embolism [5-8] also in young patients.

New developments in the field of multislice scanning have made it possible to substantially reduce radiation exposure to as little as 0.1-0.2 mSv (ultralow-dose scanning) [9]. This exposure is in the range of a plain posterior-anterior and lateral chest radiographic exam. The associated radiation burden is then in the range of one to two months of natural background radiation. Ultralow-dose CT-scanning can be an option in patients when diagnostic information of a plain chest X-ray is not sufficient, but the additional information potentially does not exceed the risk of increased radiation dose of a standard dose CT-scan.

Aim of this feasibility study was to determine if ultra-low dose CT has the potential to provide more information as a CXR at a radiation dose that remains in the range of a plain posterior-anterior and lateral chest radiographic exam. For that purpose we used the agreement rate with the reference CT, performed with standard radiation dose, as surrogate for diagnostic performance of CXR and LDCT.

MATERIAL AND METHODS

PATIENTS

The study was approved by our institutional review board and written informed consent was obtained from all patients. Between September 2003 and May 2004, we enrolled 40 adult patients (26 men, 14 women, 20-81 years old, mean 57 yrs) from the outpatient department of pulmonology, who were referred for a chest CT because of the following indications: aspergilloma (n=1), malignancy (n=6), COPD/ emphysema (n=5), bronchiectasis (n=4), hemoptysis (n=3), pneumonia (n=2), interstitial disease (n=6), dyspnoea (n=6), coughing (n=1), pneumothorax (n=1), aspiration (n=1), chylothorax (n=1), depression (n=1) and sarcoidosis (n=2).

IMAGE ACQUISITION

First we performed the standard-dose chest CT-scan for clinical purposes immediately followed by an ultra-low dose CT-scan. This way, both scans were performed either with or without contrast-enhancement. No patient received contrast alone for the ultra-low dose CT-scan. All scans were acquired on a 16-slice CT scanner (Mx8000 IDT, Philips Medical Systems, Cleveland, OH) using a spiral mode with 16x0.75mm collimation and 15mm table feed per rotation (pitch = 1.3). The entire chest was scanned in about 10 seconds using a caudo-cranial scan direction. Scans were performed in full inspiration after appropriate instruction of the patients.

Axial images were reconstructed at 1.0mm thickness and 0.7mm increment, using a moderately soft kernel (B, Philips Medical Systems), the smallest field of view (FOV) that included the outer rib margins at the widest dimension of the thorax and a 512x512 matrix. For the additional ultralow-dose examinations (LDCT) we applied the lowest exposure settings possible at our CT scanner: 20mAs and 90kVp (volume CT dose index, CTDIvol = 0.6 mGy) for all patients, independent of patient size. The exposure settings for our clinical chest CT examinations are 150mAs at 120kVp (CTDIvol = 10.5mGy).

All patients underwent a plain chest X-ray on the same day as the CT was performed. The chest X-ray was obtained using state-of-the art digital storage phosphor technology (Optimus, Philips Medical Systems).

IMAGE EVALUATION

All images were blinded before the reading sessions and read without clinical information. The chest X-rays were read on a digital workstation (Philips Medical Systems, Cleveland, USA) connected to the picture archiving computer system (PACS). Observers were asked to read the images as they are used to do in clinical practice, including the use of editing tools available on the workstation. The CTs were transferred to a stand-alone workstation (Philips Medical Systems, Cleveland, USA), which is also used in clinical routine in our hospital along with

the PACS-system. The window width and level and the viewing direction could be manipulated along to the preferences of the reader. The readers were also able to adapt on-line the section thickness and use multiplanar reformats in coronal and sagittal direction if wanted. These tools were equally available for both types of CT examinations.
Three observers with different levels of experience (two board certified radiologists with 20yrs and 3yrs of experience dedicated to chest imaging, respectively and a fourth year resident) evaluated all images for the presence or absence of an abnormality. Chest X-rays were the first examinations to be read, followed by the ultra-low dose CTs in a second session and the reference CTs in a final session. Between the reading sessions there was at least a one-week interval. The order of images varied for the different techniques and the three readers.
According to a predefined questionnaire, the readers assessed the presence or absence of pathology using a five point scale ranging from 1 (definitely absent) and 2 (probably absent) to 3 meaning equivocal and 4 and 5 meaning probably and definitely present, respectively.
Pathologic findings were classified according to their anatomic location as intraparenchymal findings, pleura & chest wall, and mediastinal lymph nodes. Readers were asked to assess the presence or absence of the following lung findings: emphysema, small lung nodules (<1 cm), large lung nodules (1–3 cm), lung masses (>3 cm), non-specific linear or reticular opacities (including plate like atelectasis), consolidation/ atelectasis (not plate like), diffuse interstitial disease and airway disease. For the pleura the readers had to assess the presence of an effusion, thickening, or pneumothorax; in the chest wall, they were asked to document bone lesions. In the mediastinum they scored the presence or absence of hilar lymphadenopathy and mediastinal lymphadenopathy (paratracheal, subcarinal, aortopulmonal). For each of the three anatomic areas (lung, mediastinum and pleura/chest wall) the readers had the option to specifically document findings not fitting the predefined pathology. The observers also had the opportunity to indicate if additional imaging was required.

Data Analysis and Statistical Analysis

All reading data were entered in a spreadsheet program (Excel; Microsoft, Redmond, Wash.). For each observer, we used the data obtained from the standard dose CTs as internal reference standard (RCT). We compared the data from the ultra-low dose scans and from the plain X-rays to that internal standard separately for each observer. The overall objective of the data analysis was to determine the agreement rates between the ultra-low dose CT-scans and the standard dose CT-scans and between the plain chest X-rays and the standard dose CT-scans, respectively.
Secondly, we pooled the data of the three observers for all structures to determine which lesion type or anatomic category was most responsible for adversely affecting the agreement.

Statistical analysis was based on 2x2 tables, the McNemar test for paired data yielding $\chi 2$ statistic with one degree of freedom, was used to assess significance of difference between the off-diagonal elements in the 2x2 tables. Statistical analyses were performed using SPSS version 13.0 (SPSS, Chicago, Ill.). P-values <0.05 were considered to indicate a significant difference.

To produce 2x2 tables reading data from the five-point scale had to be reduced to a dichotomous data set. For that purpose scores 1-3 were grouped as "absent" and scores 4&5 were grouped as "present". Within each table, agreement between the reference standard and the ultralow-dose scans was reflected by the row margins, while the agreement between the reference standard and the chest X-ray was reflected by the column margins. The total number of interpretations was split up in frequency of agreement or disagreement, resulting in four cells. The left upper cell (A) reflected disagreement for both techniques, while the right lower cell (D) represented agreement for both techniques with the reference standard. In both other cells (B and C), one technique agrees with the reference technique, while the other did not. Comparison of the "off-diagonal" cells (A and D) in agreement and disagreement enables assessment whether the agreement of the chest X-ray with the reference technique differs from the agreement of the ultra-low dose CTs with the reference technique.

Finally, we assessed the level of reader confidence per reader and technique. For this purpose, reading data had to be reduced again to a dichotomous system: scores of 2,3 and 4 were replaced by zero and scores 1 and 5 by one. The frequency of zero and one were again placed in a 2x2 table, where disagreement was replaced by "low confidence (=0)" and agreement by "high confidence (=1)".

RESULTS

Thirty-six patients completed all examinations, resulting in 108 examinations to be read. Although the image quality of the LDCT of one very heavy patient was extremely poor, all images were eligible for analysis. All examinations were read by three observers, thus there were 108 (36x3) reports per item. In total 864 (8x36x3) structures were analyzed for the lungs, 324 (3x36x3) structures were analyzed for the chest wall & pleura and 218 (2x36x3) structures were analyzed for the hilar & mediastinal lymph nodes.
Averaged over three observers, for 89% of the reference CTs (97 out of 108 reference CTs) at least one abnormal structure was reported. Table 1 shows the number of structures reported as abnormal on RCT for each observer separately and after pooling the data. The number of reported abnormal structures ranged from 110 for observer 1 to 135 for observer 3. A large mass was the least frequent finding, while non-specific linear or reticular opacities were the most frequent reported finding.

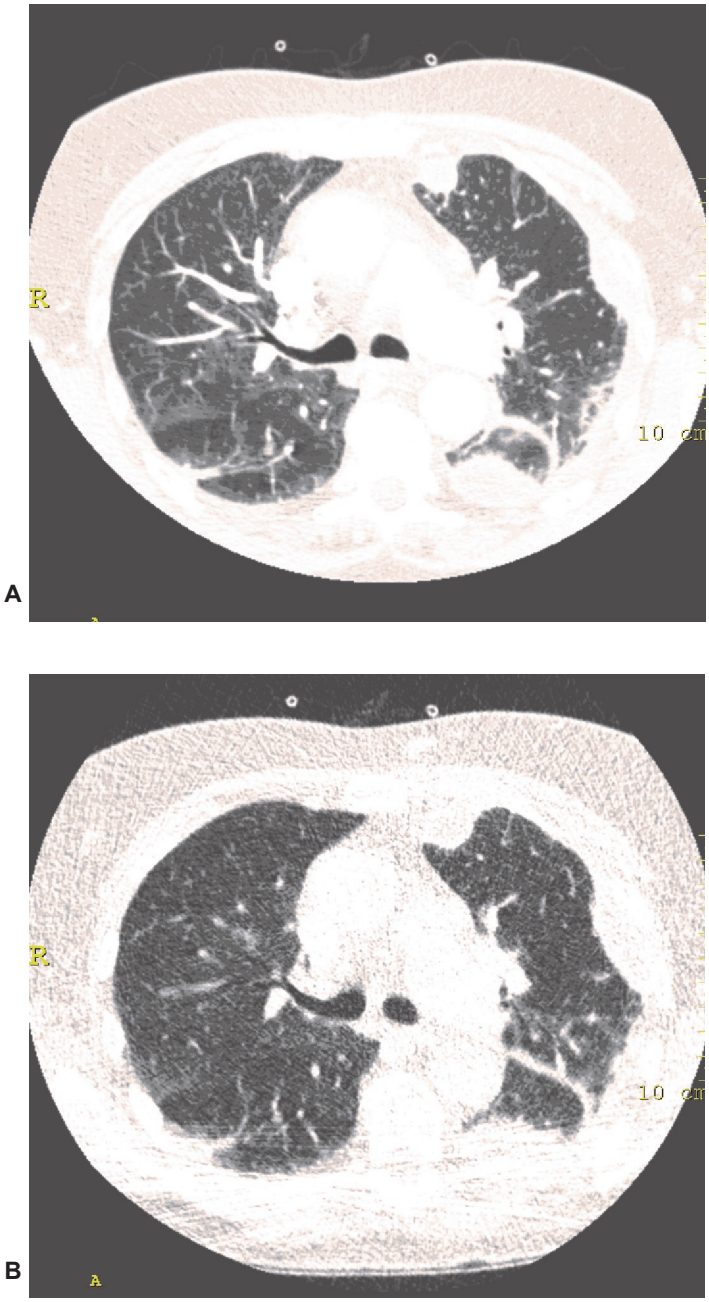


Figure 1

Standard dose (A) and ultra-low dose CT-scan (B) showing a similar image. Note that the mass dorsal in the left lung is obscured by artifacts on the ultra-low dose image, but can still be visualized.

Item	Observer 1	Observer 2	Observer 3	Total (%)
Lung				
Emphysema	7	7	7	21 (19)
Small lung opacities (< 1cm)	13	25	9	47 (44)
Large lung opacities (1-3 cm)	8	7	9	24 (22)
Lung mass (> 3cm)	2	0	1	3 (3)
Non-specific linear or reticular opacities	21	23	31	75 (69)
Consolidation/ Atelectasis (volume loss >1 segment)	6	11	15	32 (30)
Diffuse interstitial disease	5	2	5	12 (11)
Airway disease	22	12	11	45 (42)
Pleura & Chest wall				
Effusion/thickening	10*	19	24	53 (49)
Pneumothorax	2	2	1	5 (5)
Non-degenerative bone lesions	1	9	5	15 (14)
Mediastinum				
Hilar Lymphadenopathy	7	6	3	16 (15)
Mediastinal lymphadenopathy	6	11	13	30 (28)

*only pleural effusion

Table 1

Frequency of positive scores on standard dose CT-scan ("item present") for each observer separately and for all observers together.

For the pooled data of all three observers, we found a significantly higher agreement between LDCT and RCT for intraparenchymal abnormalities, pleural&chest wall abnormalities and the presence/absence of pathologically enlarged mediastinal or hilar lymph nodes (all $p<0.001$).

Analysis of the three observers separately also showed significantly more agreement between LDCT and RCT than between CXR and RCT ($p<0.001$ to $p=0.01$). Further sub analysis for each item revealed that the presence or absence of small pulmonary nodules (<1cm) ($p=0.03$), large pulmonary nodules (1-3 cm) ($p=0.02$) and airway diseases ($p=0.02$), were reported with higher agreement between LDCT and RCT than between CXR and RCT. No significant difference was demonstrated for the interobserver agreement between the CXR and the RCT ($p=0.17$ to $p=0.41$) or between LDCT and RCT ($p=0.09$ to $p=0.97$).

The level of confidence was significantly higher for LDCT compared to chest X-ray for observers 2 and 3 (both $p<0.001$), but significantly lower for observer 1 ($p<0.001$). For observer one, the number of items scored as probably absent (scored as 2) increased from n=45 (9.6%) to n=79 (16.9%), while for observers 2 and 3 the numbers decreased from 97 (20.9%) to 35 (7.5%) and from 103 (22.0%) to 59 (12.7%), respectively. The number of items scored as probably present (scored as 4) decreased for all three observers. For observer 1, the number of items scored as probably present decreased from 22 (4.7%) to 7 (1.5%), for observer 2 the number of items decreased from 23 (4.9%) to 6 (1.3%) and for observer 3 the number of items scored as probably present decreased from 37 (7.9%) to 7 (1.5%).

These results were supported by the number of indicated additional examinations. For the chest X-rays, the three observers reported that in 51/108 cases an additional examination was required, while for LDCTs the number of non-conclusive cases decreased to 35/108 ($p=0.01$).

DISCUSSION

Since CT is not hampered by over projection of several structures, CT can provide more information than plain X-rays, but at the cost of more radiation dose. The diagnostic information of a plain chest X-ray is often sufficient, but in many cases an additional CT is indicated. The main disadvantage of CT is its radiation burden and the associated hypothetical cancer risk to the patient [3], which is concerning more and more radiologists [10;11]. Multiple previous publications showed that in the thorax, where there is a high contrast between the lungs and most abnormalities in the lung such as nodular or reticular patterns, lung masses and air wall thickening, the radiation dose can be substantially reduced, still providing diagnostic information [12]. This technique has been realized in lung cancer screening trials [1;7;13-17], but in daily clinical practice low dose CT-scans are not routinely applied. However, in cases where the diagnostic information of a chest X-ray is likely to be insufficient and radiation dose, however, represents a not negligible risk, such as in young patients or in patients with benign disease or multiple follow ups, ultra-low dose CT may represent a diagnostically more effective alternative. Potential clinical indications include follow-up examinations of sarcoidosis, chronic interstitial lung diseases, bronchiectases and aspergilloma. Since we did not aim to assess the absolute diagnostic accuracy of one or the other technique and there was no absolute standard of truth in our study, we used the agreement with the most superior technique - namely the standard dose CT - as surrogate for diagnostic performance. The results of our study show the clear superior agreement between the LDCT and the RCT as compared to the CXR. This was true not only for lesions located in the lung parenchyma but also for lesions located in chest wall and mediastinum. Accordingly the level of confidence assessed by the relative frequency of definite instead of probable answers was higher with LDCT as with CXR for two of the three readers. One reader showed a higher confidence with CXR although her diagnostic agreement was higher with LDCT and RCT. This might be due to the fact that this reader felt irritated by the relatively high noise level in the LDCT images.

Multiple studies have been published, investigating the minimum dose required to provide enough diagnostic information and studies investigating the effect of dose reduction on image quality. Naidich et al already showed that lung parenchyma could be evaluated on CT-scans performed with 20mAs, without losing diagnostic quality [18], while Zwirewich et al demonstrated no significant difference in lung parenchymal structures between low-dose (40 mAs) and high-dose (400 mAs) thin section CT images, except for the presence of ground glass [12]. Ravenel et al reported that images obtained at 40mAs were all, except for one, deemed “adequate” for diagnostic purposes. While the results of our study are in agreement with these previous results in the respect that diagnostic information might be unchanged also in dose reduced scans, none of these previous studies reduced so drastically as in our study to the dose level of a CXR.

Only one study, published by Lee et al, has compared the performance of chest X-rays to low dose helical CT, delivering a comparable radiation dose [19]. Lee et al used a discontinuous high resolution protocol, performing three image, obtained at the aortic arch, at the tracheal carina, and 1 cm above the right hemidiaphragm. This approach of discontinuous data acquisition is still used today for follow-up of diffuse interstitial lung diseases, but has a limited value for other indications due to the fact that only a fraction of the lung parenchyma is scanned.

The main limitation of our study is the limited number of patients and diseases. The results may be interpreted in a way that the potential of LDCT could be shown, but more evaluation with respect to larger patient groups and specific disease entities is needed to conform our findings and more importantly to show the limitations of the ultra-low dose technique. Only then indications and patient groups can be defined that might take advantage of that technique according to the principle of ALARA.

The order of images was altered for reader and per technique all readers started with CXR, followed by LDCT and RCT at the last. We chose that order to start with the potentially least sensitive method and ending with the most sensitive method to avoid learning effects. Since the patient group, however, was relatively small, learning effects cannot be completely excluded, of so they are, however, unlikely to have affected the basic tendency of our results.

In conclusion, ultra-low dose CT performed at a radiation dose of a chest X-ray in two directions provides more diagnostic information than the plain chest X-ray as indicated by the higher agreement rate between the LDCT and the reference CT. Further analysis is needed to refine indications and patient groups that might take advantage of that technique.

REFERENCE LIST

(1) Kaneko M, Eguchi K, Ohmatsu H et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. Radiology 1996; 201(3):798-802.
(2) Swensen SJ, Aughenbaugh GL, Brown LR. High-resolution computed tomography of the lung. Mayo Clin Proc 1989; 64(10):1284-1294.
(3) de Gonzalez ABDS. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet 2004; 363(9406):345-351.
(4) Brenner DJ, Doll R, Goodhead DT et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc Natl Acad Sci U S A 2003; 100 (24):13761-13766.
(5) Aberle DR, Gamsu G, Henschke CI et al. A consensus statement of the Society of Thoracic Radiology: screening for lung cancer with helical computed tomography. J Thorac Imaging 2001; 16(1):65-68.
(6) Patz EF, Jr., Goodman PC, Bepler G. Screening for lung cancer. N Engl J Med 2000; 343(22):1627-1633.
(7) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.
(8) Tack D, de Maertelaer V, Petit W et al. Multi-Detector Row CT Pulmonary Angiography: Comparison of Standard-Dose and Simulated Low-Dose Techniques. Radiology 2005; 236(1):318-325.
(9) Lucaya J, Piqueras J, Garcia-Pena P et al. Low-dose high-resolution CT of the chest in children and young adults: dose, cooperation, artifact incidence, and image quality. AJR Am J Roentgenol 2000; 175(4):985-992.
(10) Kalra MK, Maher MM, Toth TL et al. Strategies for CT radiation dose optimization. Radiology 2004; 230(3):619-628.
(11) Mayo JR, Aldrich J, Muller NL. Radiation exposure at chest CT: a statement of the Fleischner Society. Radiology 2003; 228(1):15-21.
(12) Zwirewich CV, Mayo JR, Muller NL. Low-dose high-resolution CT of lung parenchyma. Radiology 1991; 180(2):413-417.
(13) Diederich S, Wormanns D, Heindel W. Lung cancer screening with low-dose CT. Eur J Radiol 2003; 45(1):2-7.
(14) Diederich S, Thomas M, Semik M et al. Screening for early lung cancer with low-dose spiral computed tomography: results of annual follow-up examinations in asymptomatic smokers. Eur Radiol 2004; .
(15) Henschke CI, Naidich DP, Yankelevitz DF et al. Early lung cancer action project: initial findings on repeat screenings. Cancer 2001; 92(1):153-159.
(16) Swensen SJ, Jett JR, Hartman TE et al. Lung cancer screening with CT: Mayo Clinic experience. Radiology 2003; 226(3):756-761.
(17) Takashima S, Sone S, Li F et al. Indeterminate solitary pulmonary nodules revealed at population-based CT screening of the lung: using first follow-up diagnostic CT to differentiate benign and malignant lesions. AJR Am J Roentgenol 2003; 180(5):1255-1263.
(18) Naidich DP, Marshall CH, Gribbin C et al. Low-dose CT of the lungs: preliminary observations. Radiology 1990; 175(3):729-731.
(19) Lee KS, Primack SL, Staples CA et al. Chronic infiltrative lung disease: comparison of diagnostic accuracies of radiography and low- and conventional-dose thin-section CT. Radiology 1994; 191(3):669-673.

Chapter 3

Impact of massive image noise on the extent of automated quantified emphysema using multidetector-row computed tomography

Hester Gietema
Arnold Schilham
Bram van Ginneken
Jan-Willem Lammers
Mathias Prokop

Submitted

ABSTRACT

PURPOSE

To demonstrate the impact of massive image noise on the automated quantified extent of low-attenuated areas representing small extents of emphysema using multidetector-row computed tomography, before and after applying a dedicated noise reduction filter.

MATERIAL AND METHODS

Between March 2003 and May 2004, we enrolled 31 patients (16 men, 15 women; mean age 54y, range 19-74y) from the outpatient department of pulmonology referred for a non-contrast-enhanced chest CT. All patients underwent a standard dose CT-scan (SCT) (120kVp, 130mAs, 16 x 0.75mm collimation) followed by a low dose CT (LDCT) (90kVp, 20mAs). Emphysema was quantified for all scans using a fully automated program by calculating emphysema scores (ES) as the extent of low-attenuated areas using three prefixed thresholds: -910HU, -930HU and -950HU, expressed as percentage of total lung volume. Finally, ES for LDCTs was assessed after applying a dedicated noise reduction filter. ES for SCT and LDCT, before and after applying the filter, were compared by paired-samples t-tests.

RESULTS

The extent of emphysema was overestimated for LDCT compared to SCT for all thresholds (all $p<0.05$). Results for both CTs became similar after applying a dedicated noise reduction filter (all $p>0.05$).

CONCLUSION

For low extents of emphysema, massive image noise leads to overestimation of emphysema when automated quantified. Application of a dedicated noise reduction filter can prevent this overestimation.

INTRODUCTION

Smoking-related emphysema is a common disease with high morbidity and mortality [1] and with a well-described population at-risk. Early detection of emphysema may prevent the occurrence of severe airflow obstruction by smoking cessation or medical interventions [2]. Currently, chronic obstructive pulmonary disease (COPD) is staged according to the guidelines provided by the Global initiative on Obstructive Lung diseases (GOLD), which are mainly based on the forced expiratory volume in one second (FEV_1) [3], but emphysema and airflow obstruction have been shown to be only loosely correlated [4]. Moreover, FEV_1 has been shown to be a bad predictor of emphysema mortality [5].

Although the anatomical definition of emphysema as a permanent abnormal enlargement of the airspace distal to the terminal bronchioles without obvious fibrosis [6] actually requires histology for diagnosis, in clinical practice computed tomography (CT) is used to detect emphysema. In 1988, Müller et al have described a technique to quantify emphysema on CT by highlighting pixels with an attenuation below a prefixed threshold [7] and quantifying their area as percentage of the total investigated area in a range from 0% to 100%. The results were shown to have a good correlation with pathology. Müller and co-workers used a single 10mm thick slice, but in 1995 Gevenois validated the technique for high-resolution CTs performing thin slices (1.0 mm) at 10.0 mm intervals [8] and in 2006 together with Madani for multislice scans [9]. Nowadays, the method is automated and provided by many manufacturers, resulting in a technique which is widely available, quick and easy to apply.

Presently, there are several ongoing lung cancer screening trials [10-13]. Since lung cancer and emphysema share smoking as the main risk factor, CTs performed in these trials may provide suitable data for studying the prevalence and natural course of smoking-related emphysema in relatively healthy subjects [14]. Since only a subgroup of heavy smokers develop COPD [15], these data could be used to select groups of smokers in whom more aggressive risk-modifying treatment is necessary to prevent development of severe lung destruction and airflow limitation.

Lung cancer screening trials are being performed using low-dose protocols. Lowering the dose, however, increases image noise. Quantification of the extent of low-attenuated areas in patients suffering from severe emphysema will not be effected by the extent of image noise as shown by Shaker *et al.* [16] except for very low doses. Image noise is mainly limiting diagnostic information when there is a small contrast between the structure of interest and the surrounding tissue as in low extents of emphysema.

The aim of the present study was to demonstrate that for low extents of emphysema, the extent of emphysema is overestimated. Secondly, we aimed to show that this effect can be cancelled out by the application of a dedicated noise reduction filter.

MATERIAL AND METHODS

PATIENTS AND SCANNING PROTOCOL

Between March 2003 and May 2004, we enrolled 31 patients (17 men, 14 women; mean age 54y, range 19-74y) from the outpatient department of pulmonology referred for a non-contrast-enhanced chest CT. Three patients were current smokers, 6 patients were ex-smokers and 22 patients were never smokers. The study was approved by our institutional review board and written informed consent was obtained from all patients.

Indications for referral were sarcoidosis (n=6), interstitial lung diseases (n=8), emphysema (n=6), follow-up of infectious diseases (n=7), pneumothorax (n=2, both turned out to be recovered at the time the CT was performed), chest pain (n=1) and dyspnoea (n=1).

All scans were acquired on a 16-slice CT scanner (Mx8000 IDT, Philips Medical Systems, Cleveland, OH) using a spiral mode with 16x0.75mm collimation and 15mm table feed per rotation (pitch = 1.3). CT scans for clinical purposes were realized in full inspiration using 120kVp, 130mAs ($CTDI_{vol}$ = 8.7mGy), without contrast injection. The standard-dose chest CT (SCT) was followed by a low-dose CT (LDCT) realized with identical parameters except for the radiation dose (90kVp, 20mAs, $CTDI_{vol}$ = 0.6mGy). Axial images were reconstructed at 1.0mm thickness and 0.7mm increment, using a moderately soft kernel (Philips, filter "B") and the smallest field of view (FOV) that included the outer rib margins at the widest dimension of the thorax.

EMPHYSEMA QUANTIFICATION

Data were transferred to a digital workstation with in-house developed software. Total lung volume was calculated using the following steps. Segmentation of trachea, left and right lung was performed by a fully automated region growing program starting in the trachea, which included all connected areas below -500HU. In a second step, trachea and main bronchi were excluded from the lungs. The algorithm is similar to the one described by Hu [17]. A frequency distribution histogram of voxel attenuation in lung fields was calculated for each CT. Finally, the extent of low-attenuation areas was determined by highlighting voxels with attenuation below a prefixed threshold. Emphysema scores (ES) were calculated as volume of these low-attenuation areas and expressed as percentage of total lung volume in a range from 0% to 100% for three attenuation thresholds often mentioned in literature: -910HU, -930HU and -950HU. The discussion about the most optimal attenuation threshold to apply was beyond the scope of this study.

First of all, we calculated emphysema scores for all scans without reduction of image noise. Secondly, we applied the NOVA denoising filter as post-processing step to the reconstructed 1.0 mm slices of the low-dose scans before the

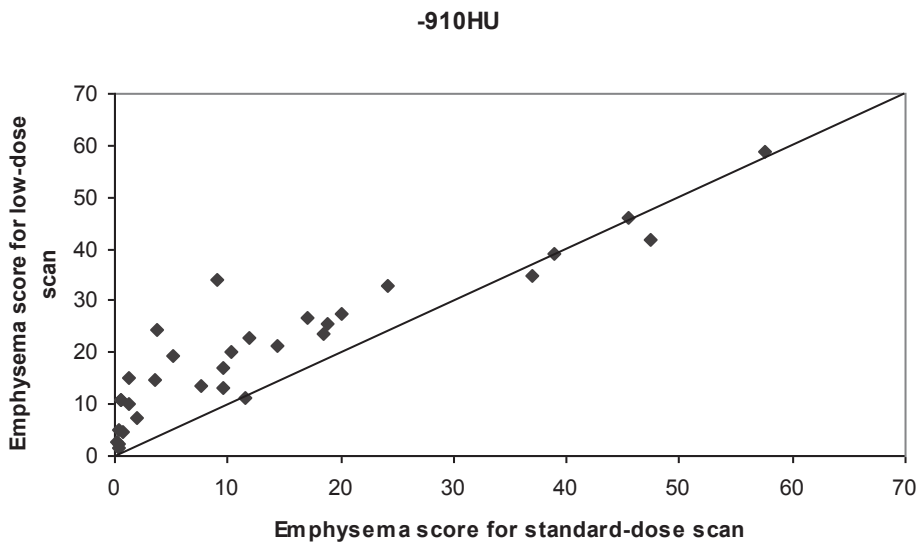
calculation of emphysema scores in order to reduce image noise as described by Schilham et al [18]. This filter was developed specifically for chest CT data, and uses prior knowledge about the noise and tissue distribution to determine at each spatial location the proper amount of local averaging. Emphysema scores of the filtered low-dose scans were compared to results from the unfiltered low-dose scans and the standard-dose scans.

ANALYSIS

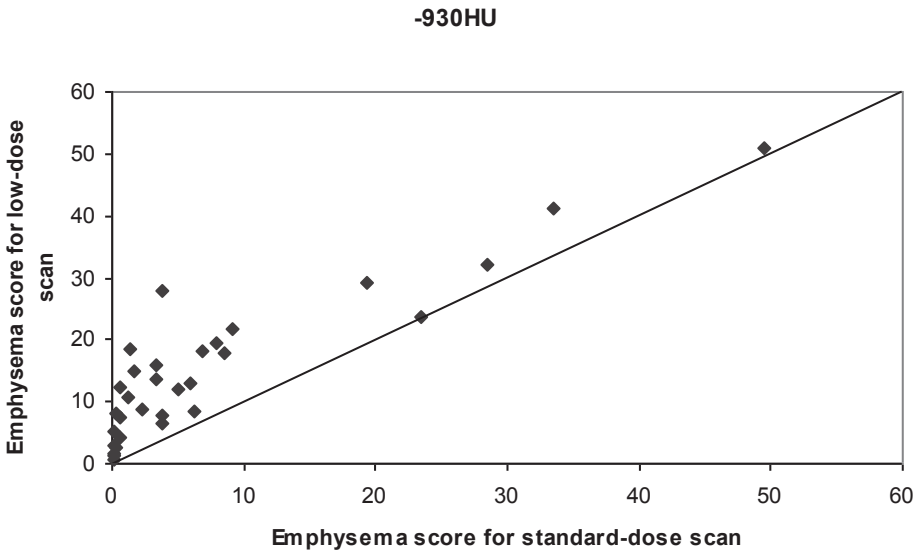
All statistical calculations were performed using SPSS statistical software release 12.0 (SPSS Inc, Chicago, Ill.). Results were shown as median and 25%-75% interquartile ranges for non-normal distributed emphysema scores and lung volumes. Lung volumes of both scans were compared for each patient using paired-samples t-tests in order to make differences in emphysema scores between both techniques caused by variation in inspirational levels less probable. Emphysema scores of the standard-dose scans were compared to emphysema scores of the low-dose scans, before and after the application of the NOVA-filter, with Spearman's correlation coefficients and with paired-samples t-tests. P-values <0.05 were considered significant.

RESULTS

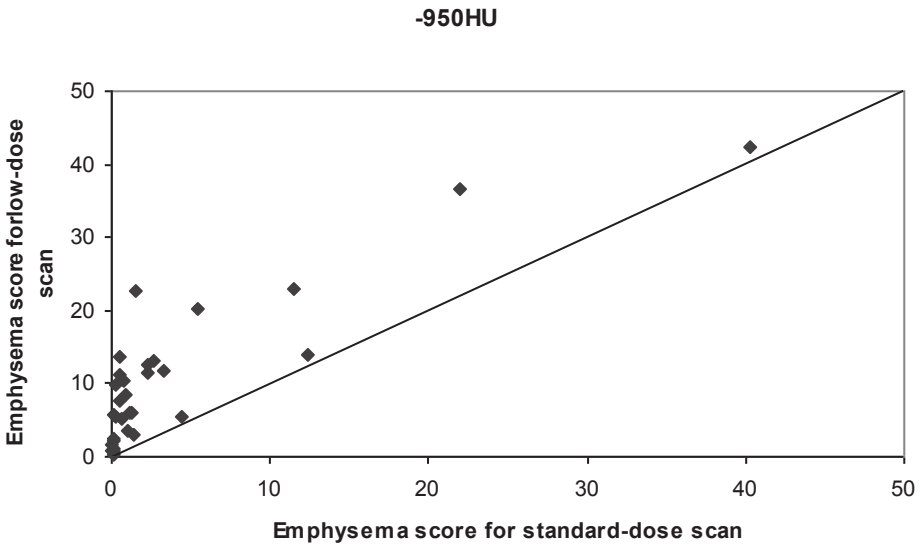
All patients completed both scans and all scans were eligible for analysis. Median lung volume was 6660 ml (interquartile range 5088ml-7822ml) for standard-dose scans (SCT) and 6845 ml (interquartile range 4930ml-8169ml) for low-dose scans (LDCT; p=0.28). Spearman's correlation coefficients for emphysema scores (ES) calculated for LDCTs and SCTs without noise reduction showed good to excellent correlations with coefficients of 0.91, 0.86 and 0.80 for extent of low-attenuation areas below -910HU, -930HU and -950HU, respectively. However, ES for LDCT were significantly higher compared to ES for scans performed at standard radiation dose (p<0.0001 for all thresholds). As shown in Figure 1, emphysema scores raised significantly after dose reduction for low to moderate extents of emphysema, while for severe emphysema (ES>30) scores were similar.



1A



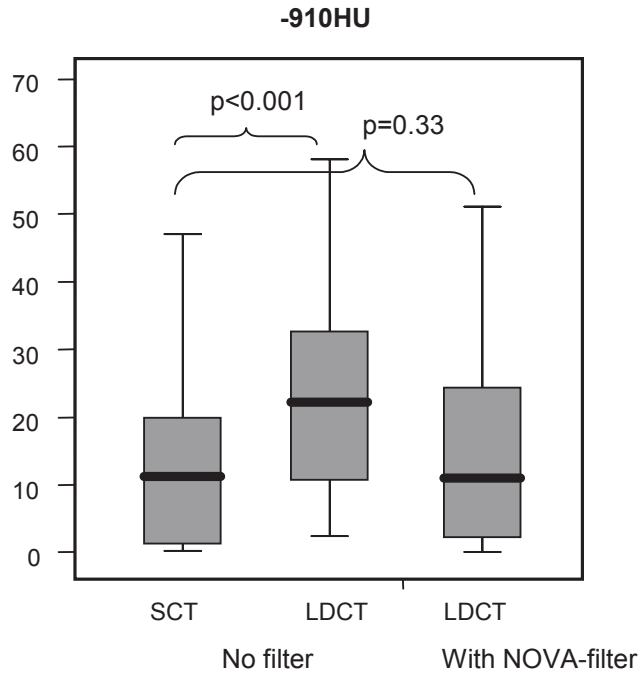
1B



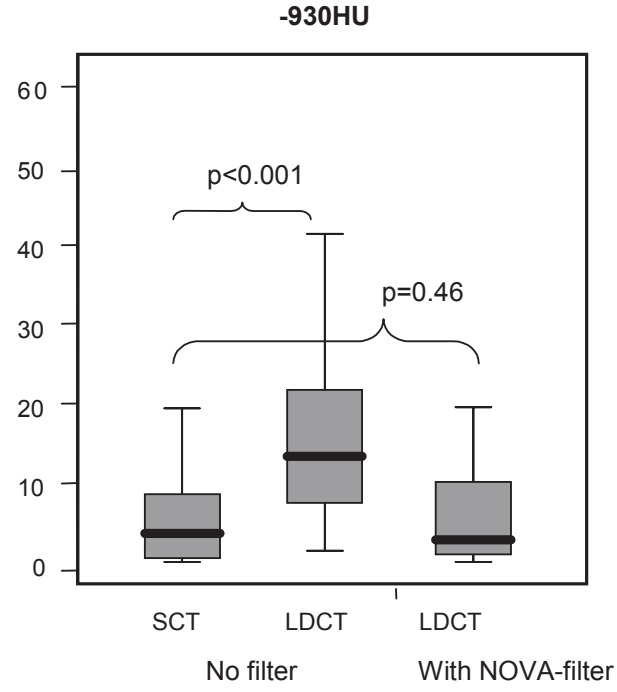
1C

Figure 1

Scatter plots showing emphysema scores performed on standard-dose scans correlated to emphysema scores performed on low-dose scans. The continuous line represents x=y. Note that emphysema scores for low-dose scans are higher than emphysema scores for standard dose scans from the same patients, except for scans with low-attenuation areas below -910HU comprising >30% of total lung volume.


-910HU
70
60
50
40
30
20
10
0
p<0.001
p=0.33
SCT
LDCT
LDCT
No filter
With NOVA-filter

2A


-930HU
60
50
40
30
20
10
0
p<0.001
p=0.46
SCT
LDCT
LDCT
No filter
With NOVA-filter

2B

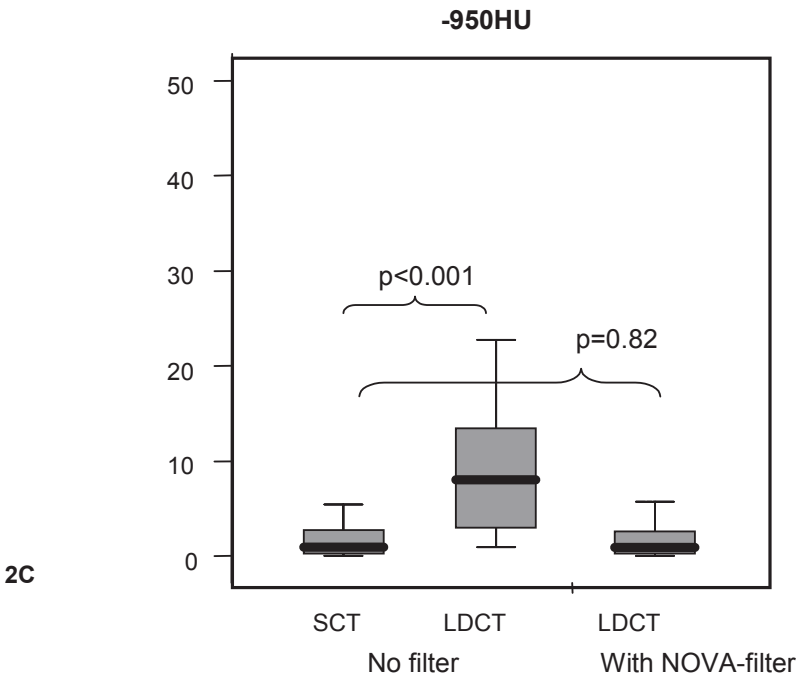


Figure 2

Box plots showing mean emphysema score and standard deviations at three attenuation thresholds for standard-dose and low-dose scans with and without noise reduction (NOVA-filter).

After filtering low-dose scans for image noise, emphysema scores dropped to the range of ES of standard-dose scans ($p>0.05$; Figure 2). This effect becomes more clear when looking at the shape of the histogram showing the number of voxels plotted against CT-values (Figure 3). This figure shows the histograms of a standard-dose CT and a low-dose CT performed for the same patient. The histogram of the low-dose CT is more flattened and shows far more voxels with a CT-number close to the attenuation of air than the standard dose scan, resulting in upraised emphysema scores. Moreover, this figure demonstrates the effect of applying a noise reduction filter, showing that the histogram of the low-dose CT after noise reduction becomes similar to that of the standard-dose CT.

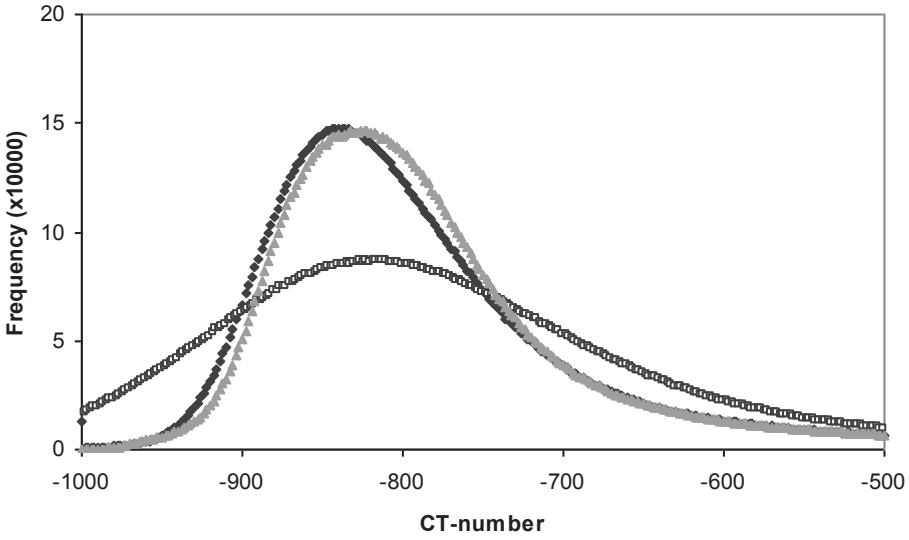


Figure 3

Histograms of the standard-dose (black diamonds) and the low-dose scan (white squares with black line) of one patient. Although the CT-numbers of the peaks of both histograms are similar, the slope of the low-dose scan is much less steep. The area under the curve of the low-dose scan includes obviously more voxels with CT-values close to -1000HU. The grey triangles represent the histogram of the low-dose scan after applying a noise reduction (NOVA) filter. Note that both histograms are almost but not completely similar. Lung volumes were 4782 ml for standard dose scan and 4882 ml for the low-dose scan.

DISCUSSION

Many factors influencing emphysema scores have been described in literature, as inspirational level [19], scanner calibration [20] and CT scanners [21], but to the best of our knowledge the impact of massive image noise on automatically obtained emphysema scores has not been reported yet.

Today, a standard method for emphysema quantification is not yet available [22]. Several studies have been performed to compare extent of emphysema detected at macroscopic and microscopic specimens and these studies recommend different attenuation thresholds as optimal threshold to use for emphysema quantification [7-9], mainly due to differences in applied scan protocols. Although the method of highlighting and quantifying low-attenuated areas has been described and validated against pathology almost twenty years ago [7], advancing technical developments have directed scan protocols applied in clinical practice to thinner collimations and reconstructed slices, and increasing applications for CT. This latter development has also increased the sense of radiation risk and subsequent temptations to decrease radiation dose to the dose "as low as reasonably achievable" (ALARA-principle) without losing diagnostic information. The high contrast in the chest between low-attenuated normal air and high-attenuated abnormal tissue makes large reduction of radiation dose in the chest possible [23;24]. Such low dose scans are already being performed in lung cancer screening studies [11;12;25;26], where the risk of dying from lung cancer highly exceeds the risk of developing cancer from the applied radiation dose [27].

COPD is a common and disabling disease. Nowadays, staging is based on the severity of airflow limitation according to the guidelines provided by the GOLD [3]. However, airflow limitation is a sign of advanced disease [28], while detection of early lung destruction can enable a more aggressive risk-modifying approach in order to slow down the progression to advanced disease or maybe even prevent the stage of airflow obstruction. Since lung cancer and COPD share smoking as the main risk factor and lung cancer screening is performed in healthy smokers, the smoking population participating in lung cancer screening trials is also the population at risk for suffering from early stages of emphysema. Therefore, routine emphysema quantification on CTs performed for lung cancer screening can be an attractive way to detect early emphysema in this high-risk population without applying additional radiation dose. Lung cancer screening trials are performed using low-dose protocols (120 or 140kVp and 20-50mAs), which are sufficient to detect and measure intrapulmonary nodules [10-13]. But for diseased areas with a low contrast to healthy tissue, such as in mild lung destruction, noise hampers the detection and especially the automated quantification of the diseased areas. Our study shows that automated emphysema quantification on low-dose chest CTs is feasible, also for low extents of lung destruction, but only after applying a denoising filter. Most vendors of CT post-processing software provide an automated emphysema quantification program, which is comparable

to the “density mask” method released by GE Healthcare and validated by Müller et al [7]. However, we showed that these programs can not be applied to patients with low extents of lung destruction without an additional filtering step. We used a noise filtering method developed by Schilham et al [18] as post-processing step before calculation of emphysema scores. This NOVA-filter was an in-house developed filter and is not commercially available. But the filter is well-described before [18] and can be reproduced by other manufacturers. Emphysema scores for low-dose scans filtered for image noise with this method were demonstrated to be similar to emphysema scores obtained from standard dose scans showing that, using this NOVA-filter, obtaining reliable emphysema scores from on low-dose scans becomes feasible. The NOVA-filter was developed especially to reduce noise in chest CT-scans performed with low radiation doses and it therefore not surprising that we obtained good results applying this filter on another group of patients, but scanned with the same protocol as the patient group Schilham and co-workers used for testing his NOVA-filter.

The results of our study may also be an explanation for the results recently published by Madani et al, who showed that -960HU and -970HU showed better correlations to macroscopic emphysema for multi-slice scans than -950HU [9], recommended by Gevenois et al for single slice scanners [8]. They applied different radiation doses (140kVp; 80mAs versus 137kVp; 255mAs) which can have resulted in the reported difference in recommended attenuation threshold.

Several other factors than radiation dose influencing emphysema scores have been mentioned in literature [22], but we could exclude crucial factors like changes in inspirational level, sampling bias and scanner calibration, to cause the reported differences in ES between standard-dose and low-dose scans. Although lung volume is shown to influence emphysema scores [19], lung volumes in both scans were similar, which makes changes in lung volume as a explanation of the systematic higher emphysema scores on low-dose CTs less probable. Since we used continuous data sets, sampling error, which turned out the most important error in the study by Stoel et al [29], was also not applicable. Finally, the scans were performed in the same session, excluding a calibration error which can influence the scoring results [20].

Our study also has some limitations. First of all, we demonstrated the effect of massive image noise on emphysema scores, but we did not investigate several radiation doses. Such a study design would provide information about the correlation between radiation dose and emphysema scores. However, performing series of scans in the same patient would lead to massive increases of applied radiation doses for this patient and is therefore unethical. A series of several radiation dose reduction steps could also be performed using dose reduction simulation programs, but it is difficult to reproduce the real clinical situation with simulation programs.

A second limitation of the study was the absence of pathologic specimens that are crucial in the diagnosis of emphysema, by which we did not compare our results

to emphysema quantified at macroscopy. Although we proved that emphysema scores performed after noise reduction produces results that are similar to scores performed with standard-dose scans, no judgment can be done about the accuracy to detect the real extent of emphysema.
A third limitation is that only three patients in this study population were current smokers and 6 patients were former smokers. Since referral for standard dose chest CT was the only inclusion criterion to be included in our study investigating the applications of low-dose scans, the majority of participants turned out to be never smokers. But since the study population showed low extents of lung destruction, like many healthy smokers [14], these patients constituted a appropriate study group to demonstrate the effect of massive image noise in patients with low extents of lung destruction.

CONCLUSION

Radiation dose reduction has a significant effect on the results of emphysema quantification with the technique of highlighting and quantifying low-attenuated areas. Emphysema scoring is feasible for low-dose scans, but only when a dedicated noise reduction filter is applied before quantifying the extent of low-attenuation areas.

REFERENCE LIST

(1) Feenstra TL, van Genugten ML, Hoogenveen RT et al. The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands. Am J Respir Crit Care Med 2001; 164(4):590-596.

(2) Bankier AA, Madani A, Gevenois PA. CT quantification of pulmonary emphysema: assessment of lung structure and function. Crit Rev Comput Tomogr 2002; 43(6):399-417.

(3) www.goldcopd.com 2005.

(4) Clark KD, Wardrobe-Wong N, Elliott JJ et al. Patterns of Lung Disease in a "Normal" Smoking Population : Are Emphysema and Airflow Obstruction Found Together? Chest 2001; 120(3):743-747.

(5) Martinez FJ, Foster G, Curtis JL et al. Predictors of Mortality in Patients with Emphysema and Severe Airflow Obstruction. Am J Respir Crit Care Med 2006;200510-1677OC.

(6) The definition of emphysema. Report of a National Heart, Lung, and Blood Institute, Division of Lung Diseases workshop. Am Rev Respir Dis 1985; 132(1):182-185.

(7) Muller NL, Staples CA, Miller RR et al. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988; 94(4):782-787.

(8) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.

(9) Madani A, Zanen J, de Maertelaer V et al. Pulmonary Emphysema: Objective Quantification at Multi-Detector Row CT--Comparison with Macroscopic and Microscopic Morphometry. Radiology 2006;2382042196.

(10) Church TR. Chest radiography as the comparison for spiral CT in the National Lung Screening Trial. Acad Radiol 2003; 10(6):713-715.

(11) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.

(12) Sone S, Takashima S, Li F et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet 1998; 351(9111):1242-1245.

(13) van Klaveren RJ, Habbema JD, de Koning HJ et al. [Screening for lung cancer in the Netherlands: the role of spiral CT scan]. Ned Tijdschr Geneeskd 2001; 145(11):521-526.

(14) Omori H, Nakashima R, Otsuka N et al. Emphysema detected by lung cancer screening with low-dose spiral CT: prevalence, and correlation with smoking habits and pulmonary function in Japanese male subjects. Respirology 2006; 11(2):205-210.

(15) Viegi G, Scognamiglio A, Baldacci S et al. Epidemiology of chronic obstructive pulmonary disease (COPD). Respiration 2001; 68(1):4-19.

(16) Shaker SB, Dirksen A, Laursen LC et al. Short-term reproducibility of computed tomography-based lung density measurements in alpha-1 antitrypsin deficiency and smokers with emphysema. Acta Radiol 2004; 45(4):424-430.

(17) Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001; 20(6):490-498.

(18) Schilham A, Van Ginneken B, Gietema H et al. Local Noise Weighted Filtering for Emphysema Scoring of Low-dose CT Images. IEEE Transactions on Medical Imaging 2006; 25(4):451-463.

(19) Shaker SB, Dirksen A, Laursen LC et al. Volume adjustment of lung density by computed tomography scans in patients with emphysema. Acta Radiol 2004; 45(4):417-423.

(20) Parr DG, Stoel BC, Stolk J et al. Influence of calibration on densitometric studies of emphysema progression using computed tomography. Am J Respir Crit Care Med 2004; 170(8):883-890.

(21)`Bakker ME, Stolk J, Putter H et al. Variability in Densitometric Assessment of Pulmonary Emphysema With Computed Tomography. Invest Radiol 2005; 40(12):777-783.

(22) Stoel BC, Stolk J. Optimization and standardization of lung densitometry in the assessment of pulmonary emphysema. Invest Radiol 2004; 39(11):681-688.
(23) Mayo JR, Hartman TE, Lee KS et al. CT of the chest: minimal tube current required for good image quality with the least radiation dose. AJR Am J Roentgenol 1995; 164 (3):603-607.
(24) Mayo JR, Kim KI, MacDonald SLS et al. Reduced Radiation Dose Helical Chest CT: Effect on Reader Evaluation of Structures and Lung Findings. Radiology 2004; 232 (3):749-756.
(25) Diederich S, Wormanns D, Heindel W. Lung cancer screening with low-dose CT. Eur J Radiol 2003; 45(1):2-7.
(26) Swensen SJ, Jett JR, Hartman TE et al. Lung cancer screening with CT: Mayo Clinic experience. Radiology 2003; 226(3):756-761.
(27) Brenner DJ. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. Radiology 2004; 231(2):440-445.
(28) Madani A, Keyzer C, Gevenois PA. Quantitative computed tomography assessment of lung structure and function in pulmonary emphysema. Eur Respir J 2001; 18(4):720-730.
(29) Stoel BC, Vrooman HA, Stolk J et al. Sources of error in lung densitometry with CT. Invest Radiol 1999; 34(4):303-309.

Part 2

The Dutch-Belgian Lung Cancer Screening (NELSON)-trial: Management and Reproducibility Studies

Chapter 4

Nodule management protocol of the NELSON randomized lung cancer screening trial

Dong Ming Xu
Hester Gietema
Mathias Prokop
Harry de Koning
René Vernhout
Kristiaan Nackaerts
Carla Weenink
Jan-Willem Lammers
Harry Groen
Matthijs Oudkerk
Rob van Klaveren

Lung Cancer. 2006 Nov;54(2):177-84

ABSTRACT

In December 2003, the Dutch–Belgian NELSON trial, a Dutch acronym for "Nederlands-Leuvens Longkanker Screenings ONderzoek", has been launched. Primary objective of the NELSON trial is to investigate whether screening for lung cancer by 16-detector multi-slice CT with 16 × 0.75 mm collimation and 15mm table feed per rotation (pitch = 1.5) in year 1, 2 and 4 will lead to a decrease in lung cancer mortality in high risk subjects of at least 25% compared to a control group which receives no screening. In this paper, the screening regimen and the classification and management of the screen-detected nodules at baseline and incidence screening is presented. This is the first large lung cancer screening trial in which the nodule management protocol is based on volumetric nodule assessment and the presence or absence of growth. Furthermore, the quality assurance measures and the NELSON management system (NMS) are presented.

INTRODUCTION

Lung cancer is currently a serious public health problem. In Europe alone, an estimated 375,000 people die from lung cancer every year and worldwide 1.4 million per year[1]. At the time of diagnosis, over 75% of persons with lung cancer have loco-regional spread or distant metastases, substantially reducing the chances of survival [2]. Theoretically, primary prevention, quitting smoking or more importantly, measures to reduce starting smoking may totally eliminate the disease, but although several such measures have been successful, the number of lung cancer deaths each year is still unacceptably high. One of the most promising recent preventive measures is early detection using multi-detector low-dose computed tomography (MDCT) screening. Cohort studies have shown that lung cancer can be detected in a much earlier stage, but it is yet unknown if earlier detection will eventually reduce lung cancer mortality [2]. To address this question, in the US the National Lung Screening trial (NLST) has been launched in 2002. It is a very large multi-center trial with 53,476 participants in 46 institutes across the US, comparing CT screening with chest X-ray screening in the control arm [3]. In Europe, the only large randomized trial is the Dutch–Belgian NELSON trial with 15,523 male and to a lesser extent female participants in four institutes, which have been launched in December 2003 [4]. Primary objective of the NELSON trial is to investigate whether screening for lung cancer by MDCTs in year 1, 2 and 4 will lead to a decrease in lung cancer mortality in high risk subjects of at least 25% compared to a control group which receives no screening, and to estimate the cost-effectiveness of this screening program. In collaboration with a single institute in Copenhagen, Denmark, where 4104 participants have been enrolled in a randomized MDCT screening trial with almost the same design as NELSON, the target of 20,000 participants has almost been reached.

Screening is not merely a radiological technique, but instead a complex process of identification and selection of the target population, call and recall of participants, work-up and evaluation of positive screenees and adequate communication of the test results to the participants and all involved health care professionals. The recruitment procedure, the selection criteria and the power calculation used in the NELSON trial have been described elsewhere [4]. This paper deals with the screening algorithm and the classification and management of screen-detected nodules at baseline and incidence screening.

The management of persons with pulmonary nodules detected in a screening context differs markedly from usual clinical practice. Screening deals with asymptomatic 'healthy' individuals, approached by a letter of invitation and health care professionals participating in a screening program carry thus extra responsibility for the information and safety of the individuals included in such a programme. Therefore, in this paper, also attention will be paid to quality assurance aspects and the role of the NELSON management system (NMS) in it. Given the fact that more and more advanced multi-detector CT scanner with

smaller collimations are being used, also outside screening programs, clinicians are more and more faced with the problem of small non-calcified pulmonary nodules. Prevalence rates up to 50% have been reported [5]. New software tools to assess volumes and volume doubling times become rapidly and widely available. Therefore, this management protocol could also be useful for the non-screening setting and provide new tools on how to deal with pulmonary nodules by using volumetric software.

NELSON MANAGEMENT SYSTEM (NMS)

To conduct this logistically complex multi-center study, the NELSON management system (NMS) has been developed. It is a web-based interactive database application used for data collection and management of all study related processes such as the selection and randomization of participants, electronic storage of questionnaires and informed consent forms, completely trackable data collection, study monitoring, reporting of scan results and scheduling of appointments for follow-up scans. Because the system works with action dates, it provides us with a complete overview and control of the planned actions, such as the planning of follow-up scans, sending of invitations to participants, test results and work-up and evaluation of suspicious nodules.

SCREENS

The participants randomized to the screen arm were invited by an invitation letter to one of the four screening sites (University Hospital Groningen, University Hospital Utrecht and Kennemer Gasthuis Haarlem in the Netherlands, and University Hospital Gasthuisberg Leuven in Belgium. The CT scans used were all 16 detector MSCT scanners (M×8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA, or Sensation-16, Siemens Medical Solutions, Forchheim, Germany). All scans were realized in about 12s in spiral mode with 16 × 0.75mm collimation and 15mm table feed per rotation (pitch=1.5), in a cranial–caudal scan direction, without contrast in low-dose setting. Depending on the body weight (<50, 50–80 and >80kg) the kVp settings were 80–90, 120 and 140kVp, respectively and to achieve a CTDIvol of 0.8, 1.6 and 3.2 mGy, respectively, the mAs settings were adjusted, dependent on the machine used. To minimize breathing artifacts, scans were performed in inspiration after appropriate instruction of the participants.

IMAGE READING

Images were read on Siemens workstations using the Syngo Lungcare software package (Version Somaris/5 VB 10A-W) for multi-dimensional image processing and computer viewing. Lung windows were assessed at a width of 1500 and a level of −650 Hounsfield Units. After a first reading, the data were stored locally on the PACS system, and sent overnight via a protected internet connection to

Groningen for second reading and central storage in the radiological database. The second readers were unaware of the conclusion of the first reader and read the images within 3 weeks after the first reading. In case of a discrepancy, a third reader (M.P. and M.O.) made the final decision. One of the second readers was trained for 3 weeks in lung cancer screening at the Department of Radiology Weill Medical College, Cornell University New York (Prof. C. Henschke), the others trained themselves by means of the ELCAP teaching file.

A nodule was defined as a small approximately spherical, non-linear circumscribed focus of abnormal tissue [6]. A nodule was classified as non-calcified (NCN) if it did not show a benign pattern of calcification [6]. For all NCNs found at baseline and annual repeat scan the maximum dimensions in *x*, *y* and *z* direction, minimum, maximum and mean diameter, size, volume, density, location (central versus peripheral, lung segment, slice number) were recorded, as well as the surface characteristics (smooth, spiculated or other).

During CT evaluation, for each evaluable nodule, the surface characteristics, distance to the pleura and the aspect of the nodule (i.e. solid, partial-solid or non-solid) was entered by the radiologist in an electronic data collection form customized for the Lungcare Siemens workstation. Nodules were classified as peripheral if the distance to the thoracic wall was less than one third of the total distance to the lung hilum. Together with the calculated sizes and volumes generated by the Siemens software, these data were automatically uploaded in NMS immediately after completion of the reading for an unlimited number of evaluated nodules per scan. In case of consecutive CT scans, nodules were matched with the same nodules documented on previous scans in order to determine changes in volume and to estimate the volume doubling time (VDT). This could be done either automatically – a matching algorithm in NMS resulted in the most probable match of nodules based on the combination of consistency, size and location – or manually, or both automatically and manually. Based on the matching of nodules, NMS detected whether a nodule was new or already existing, and automatic determination of the nodule category (1–4) and/or growth category (A, B or C) was reported (Table 1 and Table 3). After the second reading of the CT-scan and after reaching consensus about the screen result and the planned actions to be taken, the NMS generated the appropriate standard letter in order to inform both the participant and the general practitioner within 3 weeks after the CT scan. Discrepancies were identified when there was no auto-matching achieved or when the second reader disagreed on nodule number, location or volume.

For solid nodules and for the solid component of partial-solid nodules, volume was calculated by three-dimensional (3D) volumetric computer assessment. In case of inappropriate segmentation, the radiologist was able to enter manual measurements as well, which then overruled the automatically generated volume calculations. For solid pleural-based nodules, the diameter perpendicular to the costal pleura was taken because the volumetric software used was not accurate

enough for pleural-based lesions, due to inappropriate segmentation. Also for non solid lesions, size had to be determined based on two-dimensional (2D) manual measurements, and was defined as the average of length and width (d_{mean}). Length was measured in the *X*/*Y*-axis on a single CT image that showed the maximum length. Width was defined as the longest diameter perpendicular to length on the same CT image. For partial-solid lesions, both the volume (solid part) and d_{mean} (overall size of the nodule) were recorded. Throughout the study, the definition of growth was kept constant, and was defined as a percent volume change (PVC) of 25% or more according to the following formula:

(1) $$PVC(\%) = 100 * \frac{V_2 - V_1}{V_1}$$

Also for NCNs in which only 2D size parameters (d_{min} or d_{mean}) were available, volume and PVC could be estimated based on formula (3) (see below).

BASELINE SCREEN PROTOCOL

NCNs were classified in four nodule categories (NODCAT) based on size, either 3D (solid and partial solid lesions) or 2D (solid pleural lesions and non-solid lesions) or based on growth (GROWCAT) according to formula 1 (Table 2). NODCAT 1 was defined as benign, NODCAT 2 as non-significantly small, NODCAT 3 as indeterminate and NODCAT 4 as potentially malignant. Based on the highest nodule category found, participants with NODCATs 1 and 2 received a negative test result, and were invited for an annual repeat scan (first incidence screen) 1 year later (Table 3) . This design was chosen because the likelihood of malignancy in a NODCAT 2 nodule at baseline is less than 1% [7]. NODCAT 3 was defined as an indeterminate test result which required a repeat scan 3–4 months later to assess growth. If there was no significant growth on the repeat scan, the test result was called negative and participants were scheduled for an annual repeat CT-scan 8–9 months later. If there was significant growth, the test result was positive (GROWCAT C), which means that a histological diagnosis had to be obtained. Also NODCAT 4 was a positive test result which required referral to a pulmonologist for work-up and diagnosis. In case of NODCAT 4 and GROWCAT C, the general practitioner was first of all informed by the radiologist of the screening site by phone about the test results and its consequences, followed by a letter to the participant and the general practitioner.

NODCAT baseline	Definition	
I	Benign nodule (fat / benign calcifications) or other benign characteristics	
II	Any nodule, smaller than NODCAT III and no characteristics of NODCAT I	
III	Solid: Solid, pleural based: Partial solid, any nodule: Partial solid, solid component: Non-solid:	50-500 mm^3 5-10 mm dmin ≥ 8 mm dmean 50- 500 mm^3 ≥8 mm dmean
IV	Solid: Solid, pleural based: Partial solid, solid component:	> 500 mm^3 > 10 mm dmin > 500 mm^3

Table 2

NELSON classification of the different non-calcified nodules according to size at baseline screening

Nodule Type	NODCAT I	NODCAT II	NODCAT III	NODCAT IV
Solid	Negative test Annual CT	Negative Test Annual CT	Indeterminate test 3 month follow-up CT	Positive test Refer to pulmonologist for work-up and diagnosis
Partial solid	Negative test Annual CT	Negative Test Annual CT	Indeterminate test 3 month follow-up CT	Positive test Refer to pulmonologist for work-up and diagnosis
Solid-pleural based	Negative Test Annual CT	Negative Test Annual CT	Indeterminate test 3 month follow-up CT	Positive test Refer to pulmonologist for work-up and diagnosis
Non-solid	Non-existing category	Negative Test Annual CT	Indeterminate test 3 month follow-up CT	Non-existing category

Table 3

NELSON Management protocol for non-calcified nodules at baseline screening

INCIDENCE SCREEN PROTOCOL

At annual repeat screening (incidence screening), there are two possibilities: either an NCN is existing, and comparison with baseline screening is possible, or the NCN is new. For the new nodules, the same classification according to size was made as for the baseline screening round. Follow-up was different, however, because at incidence screen new nodules are supposed to have a relatively higher growth rate (Table 5).

For all existing nodules, except for NODCAT 1, a comparison with the baseline screening round was always made. If in solid nodules or solid components of partial solid nodules the PVC was 25% or more (Table 4), the volume doubling time based on changes in calculated volumes over time (VDT_v) was determined according the formula (2) [8]:

(2) $VDTv\ (days) = [ln\ 2 \times \Delta t]/[ln\ (V_2/V_1)]$

In situations in which a reliable volume estimate could not be made due to software limitations and/or manual measurement was preferred in either one of the two evaluations, changes in volumes based on changes in estimated diameter over time (VDTd) was determined according the formula (3):

(3) $VDTd\ (days) = [ln\ 2 \times \Delta t]/[3ln\ (MaxDiamXY_2\ /\ MaxDiamXY_1]$

where MaxDiamXY = maximum diameter in X/Y-axis.

However, if both scans had to be evaluated by manual measurements, such as for pleural-based solid nodules or non-solid nodules, the following formula for growth determination was applicable:

(4) $VDTd=[ln2xDt]/[ln((MaxDiamXY_2 \times PerpDiamXY_2 \times MaxDiamZ_2)/(MaxDiamXY_1 \times PerpDiamXY_1 \times MaxDiamZ_1)]$

where MaxDiamXY is the maximum diameter in X/Y-axis, PerpdiamXY the maximum diameter perpendicular to MaxDiamXY and MaxDiamZ is the maximum diameter in Z-axis. If MaxDiamZ was missing, then MaxDiamZ equalled 0.7 × |caudal slicenumber − cranial slicenumber|.

According to the VDT, growing NCNs were classified in three growth categories; GROWCAT A with a VDT > 600 days, GROWCAT B with a VDT between 400–600 days and GROWCAT C with a VDT < 400 days. Non-solid nodules in which a new solid component appeared were also classified GROWCAT C (Table 4).

During incidence screening, the test result (negative, indeterminate, positive) was based on the highest GROWCAT or the highest NODCAT in case of a new nodule. Subjects with no growth or GROWCAT A received a negative test result, and they were re-scheduled for a CT scan in year 4. For subjects with GROWCAT B or a new NODCAT 2, the test result was indeterminate and a repeat scan was made 1

year later (year 3) (Table 5). A new NODCAT 3 also resulted in an indeterminate test, which, however, required a repeat scan 6–8 weeks later. Participants with GROWCAT C or a new NODCAT 4 had a positive test result and were referred to a chest physician for work-up and diagnostic assessment.

	Year 1	Year 2	Year 3	Year 4
Volume	V1	V2	V3	V4
Percentage Volume Change: PVC (%) (solid nodules only)		$100 \cdot (V_2\text{-}V_1)/ V_2$	$100 \cdot (V_3\text{-} V_1)/ V_1$	$100 \cdot (V_4\text{-} V_1)/ V_1$
Growth (%)		PVC < 25% : no PVC ≥ 25% : yes	PVC < 25% : no PVC ≥ 25% : yes	PVC < 25% : no PVC ≥ 25% : yes
If growth: Determine Volume Doubling Time (VDT)				
Volume (solid): VDTv (days)		VDTv = $[\ln 2 \times \Delta t] /$ $[\ln (V_2/ V_1)]$	VDTv = $[\ln 2 \times \Delta t] /$ $[\ln (V_3/ V_1)]$	VDTv = $[\ln 2 \times \Delta t] /$ $[\ln (V_4/ V_1)]$
Diameter (part solid, non solid, pleural based, manual measurements) : VDTd (days)		VDTd = $[\ln 2 \times \Delta t] /$ $[3\ln (D_2/D_1)]$	VDTd = $[\ln 2 \times \Delta t] /$ $[3\ln (D_3/ D_1)]$	VDTd = $[\ln 2 \times \Delta t] /$ $[3\ln (D_4/ D1)]$
Select lowest VDT (either VDTv or VDTd)				
VDT > 600 days **GROWCAT A**		Annual CT year 4	Annual CT year 4	Stop
VDT 400-600 days **GROWCAT B**		Annual CT year 3	Annual CT year 4	Stop
VDT < 400 days or new solid component in non-solid lesion **GROWCAT C**		Refer to pulmonologist	Refer to pulmonologist	Refer to pulmonologist

Table 4

NELSON Follow-up protocol for non-calcified nodules at annual repeat screening

Nodule Type	NODCAT I	NODCAT II	NODCAT III	NODCAT IV
Solid	Negative test CT in year 4	Indeterminate test CT in year 3	Indeterminate test CT after 6-8 weeks	Positive test Refer to pulmonologist for work-up and diagnosis
Partial solid	Negative test CT in year 4	Indeterminate Test CT in year 3	Indeterminate test CT after 6-8 weeks	Positive test Refer to pulmonologist for work-up and diagnosis
Solid-pleural based	Negative Test CT in year 4	Indeterminate Test CT in year 3	Indeterminate test CT after 6-8 weeks	Positive test Refer to pulmonologist for work-up and diagnosis
Non-solid	Non-existing category	Indeterminate Test CT in year 3	Indeterminate test CT after 6-8 weeks	Non-existing category

Table 5
NELSON Management protocol for new non-calcified nodules at incidence screening

MANAGEMENT OF NODCAT 4 AND GROWCAT C NODULES

Before describing the work-up and staging procedures for the different nodule categories, it is important to realize that especially in a screening setting unnecessary surgery for benign nodules should be avoided as much as possible. This imposes special problems for the diagnostic strategy. In general, non-invasive diagnostic procedures should be applied before invasive ones if possible, so that the latter can be reserved for lesions with a high probability of malignancy and resources can be used most economically. Another problem is that the national CBO guideline for non-small cell lung cancer only deals with nodules larger than 1 cm, because sub-centimeter lung cancer lesions have been almost non-existing so far. Even though our standard work-up protocol and the national CBO guideline are available and approved by all participating centers [9], all clinical management decisions were taken at an individual level at regular multi-

disciplinary oncology meetings at the four screening sites. In some rare cases, the team decided to deviate from the management protocol described below in particular circumstances, but this was always after consensus of the whole team was obtained.

BASELINE: NODCAT 4

If the highest category was a NODCAT 4, the participant was referred to the chest physician of choice via the general practitioner, usually the chest physician associated with the screening center. Primary objective was to confirm the presence of malignancy by performing routine physical examinations, routine laboratory tests and a bronchoscopy (bronchial washing for cytology and culture, and transbronchial biopsy or brushing on indication). A percutaneous CT-guided fine needle aspirate (FNA) to obtain histology or cytology of the lesion is not a routine procedure in the Netherlands and Belgium, and if the FNA technique is used, it is only for larger peripheral nodules with good access. The FNA result can be malignant, specific benign or non-specific benign. Specific benign diagnosis include tuberculosis, mycoses, nocardia, hamartoma or a benign lymph node. If malignancy was proven, the patient was further staged (see below), followed by surgical resection. A definitively specified benign diagnosis required treatment or just observation. If no diagnosis or a non-specific benign diagnosis was obtained, the follow-up strategy was based on the assessment of nodule growth similar as to NODCAT 3, i.e. a repeat scan after 3–4 months. If at that time there was no growth, the test result was negative and participants were scheduled for an annual repeat CT scan 8–9 months later. If there was growth, the test results was positive (GROWCAT C), which meant that a definitive histological diagnosis had to be obtained. Actually, this work-up was according to our national CBO guidelines for the diagnosis and treatment of non-small cell lung cancer [9], with the exception that a FDG-positron emission tomography (PET) scan was not routinely included in the work-up of a NODCAT 4, primarily because our NELSON trial is a CT screening trial, in which the presence or absence of growth of the nodule is leading, and not the outcome of the PET scan. Furthermore, the pre-test probability of malignancy in this population of current and former smokers is very high, and a substantial proportion of the PET scans is false negative because of bronchiolo-alveolar cell carcinomas (BAC) or adenocarcinomas with BAC features, limiting the diagnostic value of the PET in the context of this CT screening trial [10-12].

BASELINE OR INCIDENCE: GROWCAT C

The work-up for participants with growing lesions (GROWCAT C) was essentially the same as for NODCAT 4, except that for these nodules a final histological diagnosis had to be obtained either by FNA, video-assisted thoracoscopic surgery (VATS), or wedge resection and examination on frozen section, and that further

observation by follow-up CT scans was no longer allowed. If malignant, the nodule had to be surgically removed after appropriate staging. If the outcome of the investigation(s) was that the lesion was benign, the participant was re-scheduled for the next regular annual CT scan.

STAGING

Staging included a standard CT-scan of the chest and upper abdomen including the liver and adrenals with intravenous contrast injection. A bone scintigraphy and MRI brain were only made on clinical indication. If the nodule was a NODCAT 4 or GROWCAT C larger than 500 mm^3, a mediastinoscopy was only performed if the PET scan showed positive mediastinal lymph nodes, if there were enlarged lymph nodes on CT (short axis > 1 cm), and in the presence of a peripheral adenocarcinoma or a centrally located tumor. For nodules between 50–500 mm^3, the role of routine FDG-PET and mediastinoscopy is not yet established and were therefore not routinely recommended.

SURGICAL RESECTION

The treatment of small malignant lesions (T1) found at screening is according to standard practice [8] i.e. if possible at least a lobectomy should be performed due to a high frequency of local recurrence after more limited resections. Only in patients with poor pulmonary function who are judged by the surgeon not to tolerate a lobectomy, a segmentectomy or wedge resection could be performed. This may in some cases be performed as a minimal invasive VATS procedure. Because the small ground glass lesions have turned out to have an excellent prognosis (Noguchi A and B) for these lesions a more limited resection is allowed [11]. During surgery staging of the tumor by systematic lymph node dissection is mandatory. In medically inoperable patients, curative stereotactic 4D radiotherapy is allowed.

QUALITY ASSURANCE

In order to promote the expertise of the investigators and to ensure the lung cancer screening trial's compliance with the quality demands of the National Health Council, several measures were taken. All radiological images are interpreted both locally and centrally in Groningen (double reading), with the intention to promote the quality and to optimise the sensitivity of the screening. To this end, two full-time dedicated radiologists were appointed in Groningen, and a third one joined later. Annual site visits, central quarterly monitoring meetings and an annual investigators' meeting were organised. Taking into account the quality requirements with which the NELSON project must comply, these site visits and monitoring meetings led to specific adjustments in the approach and the formulation of specific areas of attention. Finally, a national panel for pathology review was established, constituted by relevant pathologists at the different screening sites and an international pathology review panel formed by

seven pathologists from the United States and Europe (Dr. Flieder (USA), Prof. Franklin (USA), Prof. Westra (USA), Prof. Brambilla (FR), Dr. Thunnissen (NL), Dr. Kerr and Dr. Guldhammer (DK).

DISCUSSION

With the advent of high resolution CT screening, physicians are faced now with the very early stages of lung cancer among large numbers of insignificant, benign nodules. The optimal management protocol to discriminate between malignant and benign lesions is yet unknown. Several differences exist between the various nodule management protocols used world-wide and also the definitions used vary or are undefined, as for example, what should be regarded as an indeterminate nodule and what the definition of growth is. Both retrospective evaluation and future harmonization of the ongoing nodule management protocols is needed, and will be of great importance for the further evolution and the clinical implementation of MDCT screening for lung cancer.

Our nodule management protocol is primarily based on the Early Lung Cancer Action Project (ELCAP) protocol [13;14], but there are several differences. First of all, the nodules detected at baseline and the new nodules detected at incidence screening are classified on volume and are managed according to size. At (annual) repeat CT scanning, nodules are classified first according to the presence or absence of growth and, when there is growth, they are subsequently classified in 3 growth categories based on VDT. As such, NELSON is the first large lung cancer screening trial in which automated, volumetric nodule assessment is routinely applied and forms part of the nodule management protocol. Hopefully, this will provide an answer to the question what the predictive value of VDT is for the likelihood of malignancy in the pre-operative evaluation of screen-detected lung nodules with the current available software. New software versions for automated nodule detection and improved nodules segmentation and volume assessment will soon be released, so that the volume of non-solid nodules can also be estimated.

In many institutes, a FNA of a small pulmonary nodule is not part of the routine practice and not only requires a skilled interventional radiologist, but, ideally, also a cyto-pathologist on site. Although FNA is an important technique for larger, peripheral nodules, its sensitivity and specificity has yet to be proven for small nodules between 5-10 mm. Reliance on VDT alone might therefore be an attractive option, but although 90% of all solid and part-solid nodules have a VDT of less than 400 days, several open questions remain. Growth may not always be linear, but instead be sigmoid-shaped. Although data are scarce, lung cancer, and especially pulmonary adenocarcinoma precursor lesions, may suddenly change towards a rapid growth phase with invasive characteristics [15]. On the other hand, also benign lesions may demonstrate growth. These factors may potentially limit the value of using VDT in stratifying nodules in potentially benign or malignant.

Another major difference with other ELCAP-based screen protocols is that in the NELSON trial the number of additional radiological investigations in between the planned annual CT scans was limited as much as possible and far less than in the ELCAP protocol, not only to reduce the work load, costs and radiation exposure, but also to enable us to conclude that a reduction in lung cancer

mortality is due to annual CT screening, and not the result of a combined effect of annual screening and numerous repeat scans. The only regular repeat scan allowed according to the protocol is a CT scan after 3-4 months or 6-8 weeks for indeterminate nodules at baseline and at incidence screening, respectively. Whether this is the most appropriate approach not leading to numerous interval cancers or a high rate of additional repeat scans after referral to the pulmonologist, will become evident after evaluation of our baseline results. A 3-4 months interval seems at least long enough for nodules of infectious origin to resolve. Therefore also, we decided not to prescribe broad spectrum antibiotics routinely for indeterminate nodules.

In conclusion, taking into account the ongoing technological evolution, the widespread introduction of multidetector-row CT scanners capable of producing thin slices and the application of volumetric analysis systems, the specific management recommendations for screen-detected lung nodules are likely to change. The NELSON nodule management protocol presented is the first lung cancer screening protocol based on volumetry and designed for a large scale population based screening program without the standard use of FNA.

REFERENCE LIST

(1) Howe HL, Wu X, Ries LA et al. Annual report to the nation on the status of cancer, 1975 -2003, featuring cancer among U.S. Hispanic/Latino populations. Cancer 2006; 107 (8):1711-1742.

(2) Mulshine JL, Sullivan DC. Lung Cancer Screening. N Engl J Med 2005; 352(26):2714-2720.

(3) Gohagan J, Marcus P, Fagerstrom R et al. Baseline Findings of a Randomized Feasibility Trial of Lung Cancer Screening With Spiral CT Scan vs Chest Radiograph: The Lung Screening Study of the National Cancer Institute. Chest 2004; 126(1):114-121.

(4) Van Iersel CA, de Koning HJ, Draisma G et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). Int J Cancer 2006 in press. 2006.

(5) Austin JH, Muller NL, Friedman PJ et al. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. Radiology 1996; 200(2):327-331.

(6) Henschke CI, Yankelevitz DF, Naidich DP et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. Radiology 2004; 231 (1):164-168.

(7) Yankelevitz DF, Reeves AP, Kostis WJ et al. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 2000; 217(1):251-256.

(8) CBO Richtlijn Niet-Kleincellig longcarcinoom: Stadiering en behandeling. Van Zuiden Communications BV, Alphen aan de Rijn, the Netherlands . 2004. Ref Type: Journal (Full)

(9) Lindell RM, Hartman TE, Swensen SJ et al. Lung Cancer Screening Experience: A Retrospective Review of PET in 22 Non-Small Cell Lung Carcinomas Detected on Screening Chest CT in a High-Risk Population. Am J Roentgenol 2005; 185(1):126-131.

(10) Bastarrika G, Garcia-Velloso MJ, Lozano MD et al. Early Lung Cancer Detection Using Spiral Computed Tomography and Positron Emission Tomography. Am J Respir Crit Care Med 2005; 171(12):1378-1383.

(11) Noguchi M, Morikawa A, Kawasaki M et al. Small adenocarcinoma of the lung. Histologic characteristics and prognosis. Cancer 1995; 75(12):2844-2852.

(12) Pastorino U, Bellomi M, Landoni C et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. Lancet 2003; 362 (9384):593-597.

(13) Henschke CI. Computed Tomography Screening for Lung Cancer: Principles and Results. Clin Cancer Res 2005; 11(13):4984s-4987.

(14) Libby DM, Smith JP, Altorki NK et al. Managing the small pulmonary nodule discovered by CT. Chest 2004; 125(4):1522-1529.

(15) Thunnissen FB, Schuurbiers OC, den Bakker MA. A critical appraisal of prognostic and predictive factors for common lung cancers. Histopathology 2006; 48(7):779-786.

Chapter 5

Pulmonary nodules detected at lung cancer screening: Interobserver variability of semiautomated volume measurements

Hester Gietema
Ying Wang
Dongming Xu
Rob van Klaveren
Harry de Koning
Ernst Scholten
Johny Verschakelen
Gerhard Kohl
Matthijs Oudkerk
Mathias Prokop

Radiology 2006;241:251-257

ABSTRACT

PURPOSE

To retrospectively determine interobserver variability of semi automated volume measurements of pulmonary nodules and the potential reasons for variability.

METHODS

The Dutch-Belgian lung cancer screening trial (NELSON) is a lung cancer screening study that includes men between the ages of 50 and 75 years who are current or former heavy smokers. The NELSON project was approved by the Dutch Ministry of Health and the ethics committee of each participating hospital. Informed consent was obtained from all participants. For this study, the authors evaluated 1200 consecutive low-dose computed tomographic (CT) scans of the chest obtained during the NELSON project and identified subjects who had at least one 50–500-mm^3 nodule. One local and one central observer independently evaluated the scans and measured the volume of any detected nodule by using semiautomated software. Non-calcified solid nodules with volumes of 15–500 mm^3 were included in this study if they were fully surrounded by air (intraparenchymal) and were detected by both observers. The mean volume and the difference between both measurements were calculated for all nodules. Intermeasurement agreement was assessed with the Spearman correlation coefficient. Potential reasons for discrepancies were assessed.

RESULTS

There were 232 men (mean age, 60 years; age range, 52–73 years) with 430 eligible nodules (mean volume, 77.8 mm^3; range, 15.3–499.5 mm^3). Interobserver correlation was high (r = 0.99). No difference in volume was seen for 383 nodules (89.1%). Discrepant results were obtained for 47 nodules (10.9%); in 16 cases (3.7%), the discrepancy was larger than 10%. The most frequent cause of variability was incomplete segmentation due to an irregular shape or irregular margins.

CONCLUSION

In a minority (approximately 11%) of small solid intraparenchymal nodules, semiautomated measurements are not completely reproducible and, thus, may cause errors in the assessment of nodule growth. For small or irregularly shaped nodules, an observer should check the segmentation shown by the program.

INTRODUCTION

The introduction of multidetector computed tomography (CT) has led to a substantial increase in the number of incidentally detected small pulmonary nodules. In lung cancer screening trials in particular, only a small proportion of such nodules will be malignant; the vast majority will turn out to be benign [1-3]. The challenge is to correctly identify the few malignant nodules among the large number of benign ones by using noninvasive procedures, thus avoiding unnecessary invasive procedures [4].

Because of the high prevalence and often small size of these nodules, repeat CT scans are often performed to assess growth. The early assessment of growth has been established as the best way to discriminate between malignant and benign lesions [5-9]. Volume measurements obtained with automated segmentation tools have been shown to be more reproducible than diameter measurements [10]. Currently, various software tools are commercially available for that purpose. We have been using one of the commercially available software packages [11-15] for a multicenter lung cancer screening trial [16] and have found that volume measurements obtained with this semi automated nodule segmentation program sometimes differed between the first and second readings of the same CT data set by different observers. The purpose of this study, therefore, was to retrospectively determine interobserver variability of semi automated volume measurements of pulmonary nodules and potential reasons for the variability.

MATERIAL AND METHODS

STUDY GROUP

Data were collected from the NELSON project, a randomized Dutch-Belgian multicenter lung cancer screening trial that includes men aged 50–75 years who are current or former heavy smokers. Subjects were included if they smoked a minimum of 16 cigarettes per day for 25 years or 11 cigarettes per day for 30 years and were able to hold their breath for at least 20 seconds. Subjects who quit more than 10 years ago and/or had a history of cancer were excluded from participation.

The NELSON project is a population study that was approved by the Dutch Ministry of Health and the ethics committee of each participating hospital. Informed consent for participation and use of the collected data was obtained from all participants. Our study was performed by using CT data obtained at one of the participating centers (see below).

Within the NELSON project, all nodules with volumes of more than 15 mm^3 (corresponding mean diameter, 3.1 mm) were prospectively recorded in a database. Noncalcified solid nodules with volumes of more than 500 mm^3 (mean

diameter, larger than 9.8 mm) were considered suspicious for carcinoma, and subjects with such nodules were referred to a pulmonologist for further work-up. Solid nodules with volumes of 15–50 mm^3 (mean diameter, 3.1–4.6 mm) were followed up at standard screening intervals (1 and 3 years). Noncalcified solid nodules with volumes of 50–500 mm^3 were considered indeterminate and followed up with a repeat CT examination after 3–4 months to determine growth. Nonsolid nodules, part-solid nodules, pleural-based nodules, and nodules attached to a vessel were excluded from this interobserver variability study because the software we used for volume measurements is not yet designed for calculating reliable volume estimates for these kinds of nodules.

For this study, we included the first 1200 consecutive participants scanned between April and December 2004 at one of the participating centers (University Medical Center, Utrecht, the Netherlands). We identified all individuals with at least one solid nodule with a volume of 50–500 mm^3. All patients were referred to short-term follow-up according to the trial set-up. We included all patients with nonc-alcified solid nodules that were not attached to vessels or pleura with volumes of 15–500 mm^3. This volume range included the smaller nodules of 15–50 mm^3 found in this patient group because we wanted to acquire data for a wider range of nodule sizes.

CT Scanning

CT was performed with a 16–detector row scanner (Mx8000 IDT; Philips Medical Systems, Best, the Netherlands). All scans were performed in a spiral mode with 16 x 0.75-mm collimation and 15-mm table feed per rotation (pitch, 1.3). We used a caudocranial scan direction and scanned the entire chest in approximately 10 seconds. Contrast material was not injected. Exposure settings were 30 mAs at 120 kVp for patients weighing 80 kg or less and 30 mAs at 140 kVp for those weighing more than 80 kg. This corresponds to CT dose index values of 2.2 and 3.5 mGy, respectively. We reconstructed 1.0-mm-thick transverse images at 0.7-mm increments by using the smallest field of view to include the outer rib margins at the widest dimension of the thorax. A soft kernel was used for reconstruction (B filter; Philips Medical Systems).

Nodule Detection and Volume Measurements

CT scans were evaluated exclusively by using digital workstations (Leonardo workstation; Siemens Medical Solutions, Erlangen, Germany) with U.S. Food and Drug Administration–approved, commercially available software for semi automated volume measurements (LungCare; Siemens Medical Solutions). Potential nodules were identified by a human observer (H.A.G., with 1 year of experience in radiology) on transverse thin-slab maximum intensity projections (slab thickness, 5 mm), which were reviewed with standardized window level and width settings of 1500 and –500 HU, respectively.

After the observer marks a candidate nodule with a mouse click, the program automatically defines a volume of interest around the candidate nodule, which can be further analyzed by using volume-rendered displays or multiplanar reformations. The candidate nodule can then be either approved or discarded.
The evaluation of a nodule with a second mouse click initiates an automated volume measurement program, which includes the following steps that are performed in the background [4]. First, a fixed-attenuation threshold of –400 HU is applied, and a three-dimensional connected "structure of interest" is extracted. This structure of interest consists of the nodule and, if present, connected structures such as vessels or parts of the chest wall. Subsequently, a small spherical three-dimensional template that originates from the click point is gradually expanded; its cross-correlation with the segmented nodule is computed for each step. The peak value of the cross-correlation curve is determined, and an empirical cutoff value close to the peak value is used to separate the nodule from potential adjacent structures.
In this manner, an optimum three-dimensional template is generated that represents the nodule in the most optimal way. This template includes as much of the nodule as possible without the inclusion of surrounding structures. Finally, the nodule is segmented by fusing the optimum three-dimensional template and the structure of interest; this is followed by spatial reasoning to remove adjacent structures [4]. The segmented nodule is then shown in yellow on the volume-rendered display of the volume of interest. If the proposed overlay does not match the nodule, the observer can interactively change which point on the cross-correlation curve is taken as the best proposal for a three-dimensional template (by using the "modify" option). In this way, the user can shrink or grow the overlaid volume to a user-defined extent. In addition, three-dimensional cutoff planes can be set, defining the space that must be excluded from the overlay volume.
The program calculates the volume; the diameters along the x-, y-, and z-axes; and the minimum and maximum diameters of the segmented nodule to two decimal places.

NODULE EVALUATION

After the first reading by one local observer (H.A.G.), scans were transferred for central reading to University Hospital Groningen, where another observer (Y.W. or D.X., both with 6 years of experience in radiology) performed the second reading. The same software algorithm was used at both centers. All nodule measurements were performed once locally and once at the central reading area without modification of the output of the program. Finally, the segmented volumes were stored in a central database (the NELSON Management System).
Nodules for which volume measurements yielded different results underwent further work-up. Two observers (H.A.G. and M.P.)—both from the local institution—classified the morphology of these nodules with regard to their outer

contours (smooth, irregular, polylobulated, or spiculated) and forms (round, oval, or elliptic). This classification was performed in consensus and on the basis of visual assessment. For each of these nodules, the same two observers visually compared the two segmented volumes (marked in yellow) and the appearance of the nodule on the volume-rendered images to qualitatively assess the discrepancies in the segmented volumes. This routine procedure was performed for all nodules as part of the NELSON project.

DATA ANALYSIS

Differences in volume were calculated by subtracting the volume measured by the local observer from that measured by the central observer. Differences in volume measurements of one nodule were normalized with respect to the underlying nodule size. This was done separately for each nodule by calculating the ratio between the difference in volume measurements (numerator) and the mean volume (denominator). Agreement in measured volume was shown visually by using Bland-Altman plots with 95% confidence intervals (CIs) (limits of agreement) [17]. Because nodule size showed a non-normal distribution, inter-measurement agreement was determined by calculating the Spearman correlation coefficient.

We determined the number of nodules for which there were discrepant volume measurements. The degree of variation was calculated as the difference between the first and second measurements relative to the mean volume measured by both observers. We did this for the entire group of nodules and for the following three nodule-volume subgroups: 15–50 mm^3, greater than 50 to 200 mm^3, and greater than 200 to 500 mm^3. Differences between subgroups were calculated by using one-way analysis of variance.

In addition, we calculated descriptive statistical parameters for the relative differences between the two measurements for (a) all nodules and (b) only those nodules in which discrepancies in volume were found. The difference in the frequency of discrepancies in nodule size between the three nodule size categories was determined with the Spearman correlation coefficient for non–normally distributed populations. A P value of less than .05 was considered to indicate a statistically significant difference.

To calculate the degree of variation in those nodules for which discrepant volume measurements were obtained, we used the positive value of the relative difference between both measurements. All statistical analyses were performed with software (Excel 2000, Microsoft, Redmond, Wash; and SPSS, version 12, SPSS, Chicago, Ill.).

RESULTS

NODULES INCLUDED

From among patients who underwent the first 1200 baseline scans performed in our hospital as part of the NELSON project, we identified 232 men who had at least one nodule with a volume of 50–500 mm^3. In these 232 men (mean age, 60 years; age range, 52–73 years), we identified 450 nodules with volumes of 15–500 mm^3 that had been detected and measured by both observers. Of these 450 nodules, 430 fulfilled the inclusion criteria of being solid and fully surrounded by lung tissue. Of the 20 nodules that were excluded, one was excluded because it was not solid, six were excluded because they were attached to a vessel, and 13 were excluded because they were attached to the costal pleura. The mean volume (± standard deviation) of these 430 nodules was 77.8 mm3 ± 71.7 (range, 15.3–499.5 mm^3). One hundred eighty-eight nodules (43.7%) had a volume of 15–50 mm3, 213 (49.5%) had a volume of 50–200 mm^3, and 29 (6.7%) had a volume of 200–500 mm^3.

MEASUREMENTS

Spearman correlation coefficient analysis revealed that inter-measurement (interobserver) agreement was excellent (r = 0.99) (Figure 1). Identical volumes were measured by both observers in 383 of the 430 nodules (89.1%; Figure 2). Six nodules showed a relative difference of less than –10%, nine showed a relative difference of –0.1% to –10%, 22 showed a relative difference of 0.1%–10%, and 10 showed a relative difference of more than 10%. This means that the positive difference in volume measurements was less than 10% in 31 nodules (7.2%) and more than 10% in 16 (3.7%). In five nodules (1.2%), the positive difference was more than 25%. The mean positive difference between the measurements for all 430 nodules was 1.2% ± 4.3. If we consider only those 47 nodules for which we found interobserver variability, these values increase to 10.2% ± 8.4 (Figure 3). Results of analysis according to nodule size demonstrated that the percentage of nodules for which there were discrepant volume measurements increased with larger nodule size: Although only 7.4% (14 of 188 nodules) with volumes of 15–50 mm^3 were affected, the percentage increased to 13.1% (28 of 213 nodules) for nodules with volumes of 50–200 mm^3 and to 17.2% (five of 29 nodules) for nodules with volumes of 200–500 mm^3 ($P < 0.001$). In the latter group of five nodules, the mean positive difference in measured volume was 16.5% ± 9.9.

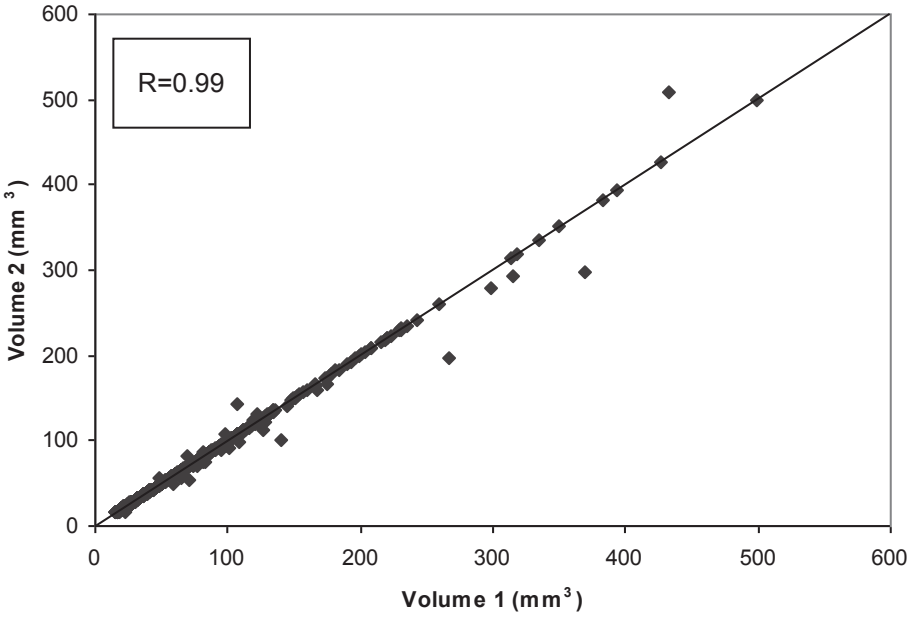
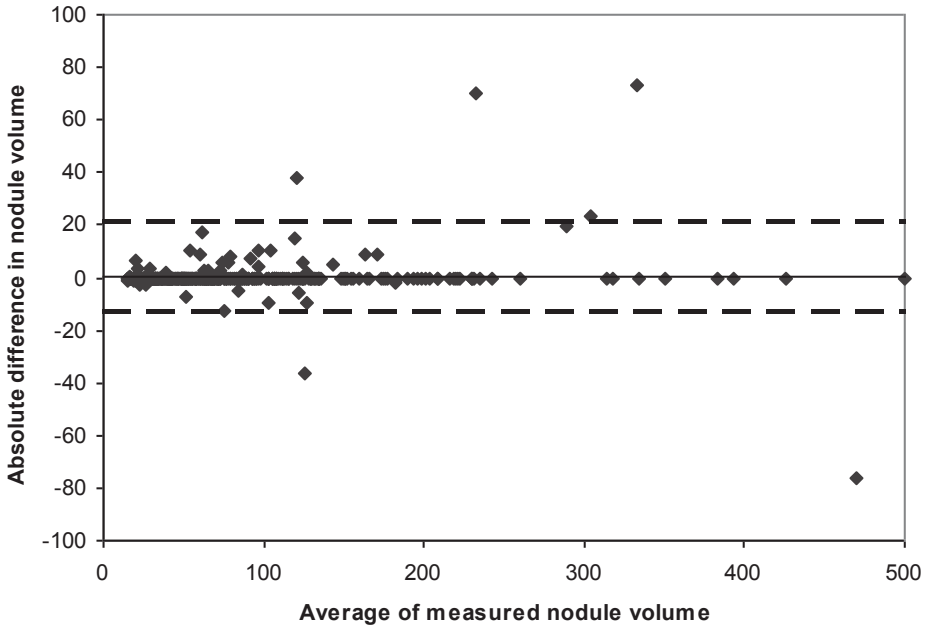
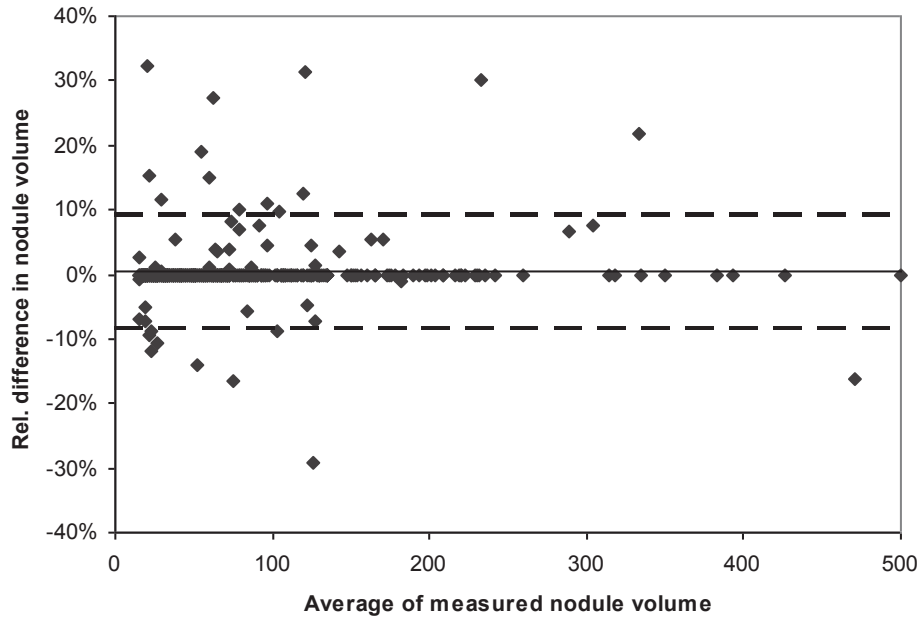


Figure 1

Volumetric results for each nodule. Volume 1 was measured by the local observer while volume 2 was measured by the central observer. The line indicates identical volumes measured by both readers.


Absolute difference in nodule volume
100
80
60
40
20
0
-20
-40
-60
-80
-100
0
100
200
300
400
500
Average of measured nodule volume

2A



2B

Figure 2

Bland and Altman plot for all 430 nodules included in the study. Absolute (A) and relative (B) interobserver variability is shown as a function of the mean nodule volume. The 95% confidence interval is indicated by a dashed lines; the bias (average difference) is stippled line. Note that although the mean difference is close to zero, a substantial interobserver variability can be seen as shown by the 95% confidence interval .

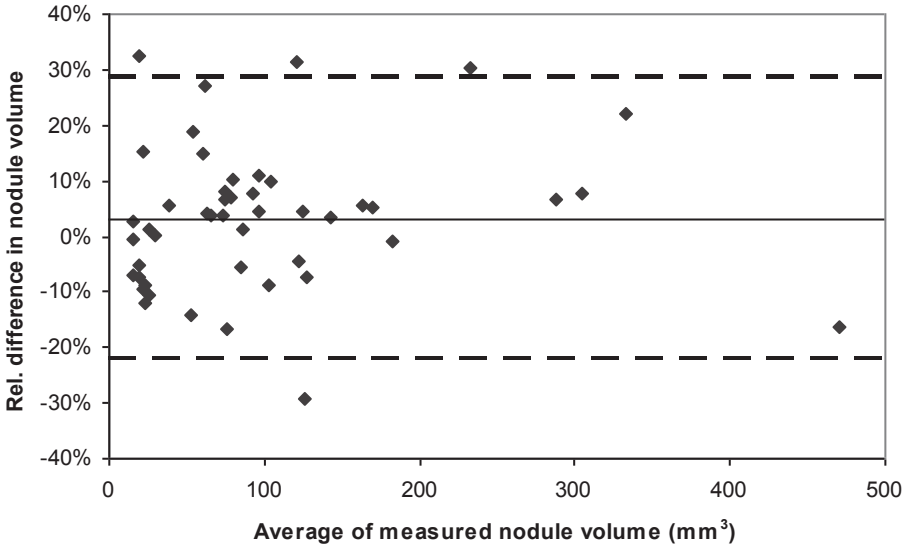


Figure 3

Bland and Altman plot for the 47 nodules in which there was a difference in volume measurements between central and local reading. Interobserver variability is shown as a function of the mean nodule volume. The 95% confidence interval is indicated by a dashed lines; the bias (average difference) is continuous line.

REASONS FOR DISCREPANCIES

We identified the following potential reasons for discrepancies in volume measurements between observers. In very small nodules with a volume of less than 30 mm^3, the entire nodule that is visible on the volume-rendered image is not actually included in the threshold used for analysis. Instead, the outer voxels, which contain part nodule and part surrounding lung, are excluded. This was the cause of discrepant measurements in 12 nodules. The user, however, can interactively modify the segmented volume to improve segmentation.

In larger nodules, an irregular shape may cause problems because the separation between a nodule and its surroundings might be dependent on how well the template matches the actual nodule. As a result, parts of the nodule may not be included in the initial segmentation. Among the remaining 35 nodules for which measurements varied, we found an irregular margin in 27 (Figure 4), a lobulated shape in three, spiculated margins in three, and an elongated shape in two (Figure 5).

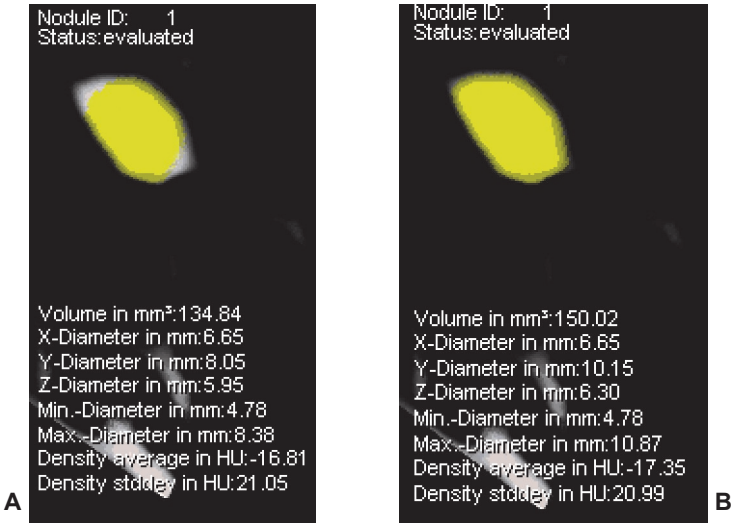


Figure 4

Nodule with a polygonal shape that was incompletely segmented during one of the two measurements. Nodule volume varied from 134.8 mm^3 (A) to 150.0 mm^3 (B).

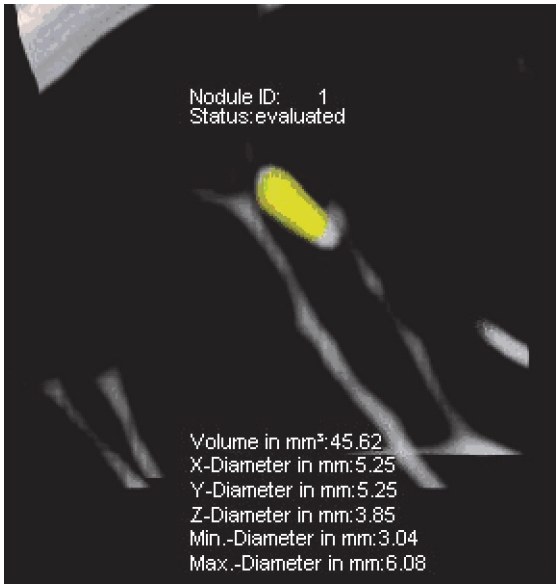


Figure 5

Elongated nodule in which there was incomplete segmentation (indicated in yellow), which can lead to interobserver variability.

DISCUSSION

Although semi automated measurements were highly reproducible for the vast majority of nodules, our results showed that there is a subset of nodules (approximately 11%) in which measured volume varies. The use of growth rates to detect malignant nodules, as is done in most lung cancer screening studies [1-3], necessitates reproducible volumetric measurements. It is known that interobserver variability is one of the factors that can influence the reproducibility of volume measurements [4;11;18].

The observation that there is a subset of nodules for which measurement varies between observers basically indicates that a very small measured difference in volume should not automatically imply real growth as the only source. When a repeat scan is performed after 6 weeks, the volume of a malignant nodule (with a volume doubling time of 300 days) that shows slow but substantial growth will increase about 10%. Referral of all subjects in whom nodules have shown a 10% increase in volume to a pulmonologist would imply that about 4% of the stable nodules will turn out to be false-positive findings. In lung cancer screening trials, in which thousands of nodules are being detected, this will concern a substantial number of nodules. A 3-month interval was recently recommended by the Fleischner Society for the follow-up of nodules with diameters of more than 6 mm (corresponding volume, 216 mm^3) [19]. After 3 months, a malignant nodule with a volume doubling time of 300 days will have increased 23% in volume. With use of a 25% increase in volume as the threshold for referral to a pulmonologist, the number of false-positive findings will decrease to 1.2%. When a repeat scan is performed after 6 months, no stable nodules will be depicted as showing substantial growth.

Currently, the combination of fully automated nodule detection and consecutive volume measurements is not yet commercially available. However, semiautomated software in which a nodule is digitally marked by an observer to initiate an automated volume measurement should already yield highly reproducible estimates because the only nonautomated part of the procedure is a manually indicated starting point for nodule segmentation. Results of in vitro validation studies of this software with artificial (rounded) nodules [14] demonstrated that the segmentation approach used with the software tools is very precise for rounded nodules, showing that even a change of 100 μm can be detected on CT images. Because in vitro reproducibility was 100%, there seemed to be little use in meticulously checking the segmentation of simple intraparenchymal nodules.

Semiautomated volume estimates of rounded nodules with a smooth margin appear to be extremely reproducible in vivo. In fact, we found no such nodule among those with discrepant measurements. This was expected because the software uses a spherical template to separate the nodule from surrounding structures. In our lung cancer screening trial, however, a substantial amount of nodules had a more irregular shape or margins, and some nodules had

spiculation. These problems were seen most often in the group of intermediate-sized nodules with volumes of 50–500 mm^3. The segmentation of such nodules often turns out to be incomplete and variable, due in part to the partial volume effect on the margins of these nodules, which leads to a lowered attenuation in part of the nodule and exclusion of this part of the nodule from the segmentation. The irregular shape can lead to problems with the spherical template: Depending on the location of the seed point, there may be different peak values for matching the spherical template with the real nodule, and, thus, the measured volume may vary. This is important because this group of irregular nodules can be expected to include the most malignant ones. With the current semiautomated software, however, it is possible to modify the segmented volumes and, consequently, adapt the segmented part to the shape of the nodule. In this way, the observer can match the segmentation shown as a yellow overlay with the visual assessment of the nodule. In cases of ill-defined margins, it can be difficult to identify the correct border lines that separate the entire nodule from its surroundings.

Although we focused on software from one particular vendor for this study, most other volume measurement software for lung nodules also use seed points to indicate a nodule and require some type of paradigm to separate a nodule from neighboring structures. This makes it probable that other programs may also be affected by the factors described earlier.

A potential limitation of this study was the fact that the on-site reviewer had only 1 year of radiology experience. This reader, however, was well trained for the specific task of identifying and selecting nodules. Because this was not a nodule detection study but a study in which semiautomatic software was used to calculate the volume of nodules, the influence of user experience should be minimal. Another potential limitation of this study was the fact that it was based on one specific software program; thus, our results are strictly applicable to this specific software only. Any other software with which any manual user input is required will also be subject to potential interobserver variations. The amount of these variations, however, will be dependent on the software.

In conclusion, the results of our study demonstrate that although modern software used to measure lung nodule volume yields very high interobserver correlation ($r = 0.99$), volume measurements may vary in approximately 11% of cases. Because the detection of minor differences in nodule volume is important for determining whether a nodule is growing, it is essential to improve reproducibility even further. These tools, however, cannot be used in a fully automated approach. In fact, it is important to check the segmentation for completeness, particularly when a nodule has an irregular shape or irregular margins. Early published in vitro data [14] gave hope that checking the segmentation does not appear necessary; our results, however, suggest that a meticulous check of the segmentation should be performed in a clinical (in vivo) setting: In case of an incomplete segmentation, one should try to reposition the initial seed point to ensure visually complete segmentation and, when necessary,

modify the segmented volume to match segmentation to the visual assessment of the nodule of interest.

REFERENCE LIST

(1) Diederich S, Wormanns D, Heindel W. Lung cancer screening with low-dose CT. Eur J Radiol 2003; 45(1):2-7.
(2) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.
(3) Swensen SJ, Jett JR, Hartman TE et al. Lung cancer screening with CT: Mayo Clinic experience. Radiology 2003; 226(3):756-761.
(4) Wormanns D, Kohl G, Klotz E et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004; 14(1):86-92.
(5) Diederich S, Thomas M, Semik M et al. Screening for early lung cancer with low-dose spiral computed tomography: results of annual follow-up examinations in asymptomatic smokers. Eur Radiol 2004; .
(6) Henschke CI, Naidich DP, Yankelevitz DF et al. Early lung cancer action project: initial findings on repeat screenings. Cancer 2001; 92(1):153-159.
(7) Takashima S, Sone S, Li F et al. Indeterminate solitary pulmonary nodules revealed at population-based CT screening of the lung: using first follow-up diagnostic CT to differentiate benign and malignant lesions. AJR Am J Roentgenol 2003; 180(5):1255-1263.
(8) Usuda K, Saito Y, Sagawa M et al. Tumor doubling time and prognostic assessment of patients with primary lung cancer. Cancer 1994; 74(8):2239-2244.
(9) Wormanns D, Diederich S. Characterization of small pulmonary nodules by CT. Eur Radiol 2004; 14(8):1380-1391.
(10) Revel MP, Lefort C, Bissery A et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. Radiology 2004; 231(2):459-466.
(11) Goo JM, Lee JW, Lee HJ et al. Automated lung nodule detection at low-dose CT: preliminary experience. Korean J Radiol 2003; 4(4):211-216.
(12) Hasegawa M, Sone S, Takashima S et al. Growth rate of small lung cancers detected on mass CT screening. Br J Radiol 2000; 73(876):1252-1259.
(13) Reeves AP, Kostis WJ. Computer-aided diagnosis for lung cancer. Radiol Clin North Am 2000; 38(3):497-509.
(14) Wormanns D, Fiebich M, Saidi M et al. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. Eur Radiol 2002; 12(5):1052-1057.
(15) Zhao B, Gamsu G, Ginsberg MS et al. Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. J Appl Clin Med Phys 2003; 4(3):248-260.
(16) van Klaveren RJ, Habbema JD, de Koning HJ et al. [Screening for lung cancer in the Netherlands: the role of spiral CT scan]. Ned Tijdschr Geneeskd 2001; 145(11):521-526.
(17) Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307-310.
(18) Kostis WJ, Yankelevitz DF, Reeves AP et al. Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT. Radiology 2004; 231(2):446-452.
(19) MacMahon H, Austin JHM, Gamsu G et al. Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society. Radiology 2005; 237(2):395-400.

Chapter 6

Interscan variability of semiautomated volume measurements in intraparenchymal pulmonary nodules using multidetector-row computed tomography: Influence of inspirational level, nodule size and segmentation performance

Hester Gietema
Cornelia Schaefer-Prokop
Willem Mali
Gerard Groenewegen
Mathias Prokop

Accepted for publication in Radiology

ABSTRACT

PURPOSE

To prospectively assess the reproducibility of semiautomated volume measurements of pulmonary nodules in low-dose multidetector-row computed tomography (CT) and to investigate the influence of nodule size, segmentation algorithm and inspirational level.

MATERIAL AND METHODS

The study was approved by our institutional review board and written informed consent was obtained from all patients. Between June 2004 and March 2005, 20 patients (15 men, 5 women, aged 40-84yrs, mean 57yrs) referred for chest CT for known lung metastases, received two additional low-dose non-contrast-enhanced chest CTs (16 x 0.75mm collimation). Between these scans, patients got off and on the table to simulate a follow-up scan. Non-calcified solid pulmonary nodules between 15 and 500 mm^3, not abutting with vessel or pleura, were measured on both scans using widely applied commercial semiautomated software. Inter-scan variability was established by the Bland-and-Altman approach. We assessed the impact of nodule shape (spherical or non-spherical) to measurement variability using one-way ANOVA, while the contributions of mean nodule volume and change in lung volume were investigated with univariate linear regression for completely (group A) and incompletely segmented (group B) nodules.

RESULTS

We evaluated 218 eligible nodules (range 16.4 to 472.7 mm^3, 106 spherical, 112 non-spherical shape). The 95% confidence interval for difference in measured volumes was -21.2% to 23.8% (mean difference 1.3%). The preciseness of nodule segmentation was highly dependent on nodule shape ($p<0.001$) and weakly related to inspirational level for completely segmented nodules ($r=-0.20$; $p<0.047$), while mean nodule volume did not show any impact ($p=0.15$ and $p=0.81$ for group A&B respectively).

CONCLUSION

Variation of semiautomated volume measurements of pulmonary nodules can be substantial. Segmentation represents the most important factor contributing to measurement variability, while change in inspirational level has only a weak impact for completely segmented nodules.

INTRODUCTION

Precise assessment of change in size of pulmonary nodules on follow-up scans is pivotal for evaluation of nodules in lung cancer screening trials [1-6] but also in clinical practice where change in size is used to evaluate response to therapy [7]. Today, in an oncological setting two-dimensional measurements are being performed using electronic calipers, in order to determine the longest diameter of the target lesion, as recommended by the RECIST-group [7]. However, three-dimensional measurements have been shown to be more accurate [4;5] and are therefore often applied in lung cancer screening trials [1;3]. Moreover, RECIST-criteria apply to nodules over 10mm, while in a lung cancer screening trials, only nodules less than 10mm are being followed-up.

Volume doubling times of malignant pulmonary nodules may vary between 30 and 400 days [6]. For small or moderately growing nodules, the increase in size over a follow-up period of up to one year will be only in the range of a few voxels [2;6]. Therefore, it is crucial to distinguish between real but slow change in size and other factors influencing volume measurements. Patient position, heart pulsation and inspirational levels have been hypothesized to influence the assessment of nodule size [3]. Inter-scan variability has already been shown to be an important factor by Wormanns et al, but solely patient positioning was considered in that study [8]. Kostis et al evaluated the reproducibility of repeat volume measurements of stable small pulmonary nodules that showed no increase or decrease in the size of the nodule for more than 2 years. As a consequence of the inclusion criteria, the vast majority of nodules was smaller than 5mm in diameter and no distinction could be made between truly identical volume over the follow-up period or minor changes in size during this period [3].

Objective of our study was to evaluate the preciseness of a commercially available and widely used semiautomated volume measurement algorithm (LungCare©, Siemens Medical Solutions, Erlangen, Germany) for intrapulmonary nodules with a diameter of 3 mm to 10 mm (corresponding to a volume range of 15 mm^3 to 500 mm^3). For this purpose, we quantified variation of semiautomated volume measurements of intrapulmonary nodules in repeat scans. Moreover, we investigated the contributions of nodule size, the performance of the segmentation algorithm and of variation in inspirational level to the extent of interscan variability.

MATERIALS AND METHODS

PATIENTS AND NODULE SELECTION

The study was approved by our institutional review board and written informed consent was obtained from all patients after explanation about the risks of the additional radiation dose. Between June 2004 and March 2005, we enrolled 20 consecutive adult patients (15 men, 5 women, 40-84 years old, mean 57 yrs) in this study who had known pulmonary metastases. All patients visited the oncology outpatient department and were referred for a chest CT for clinical indications. The presence of lung metastases had been previously shown on chest CT or chest radiography. The majority of patients (n =13) were referred for chest CT to monitor the effect of anticancer therapy. The other patients (n=7) were referred for a baseline chest CT before the start of anticancer therapy. These patients had pulmonary metastases shown on chest X-ray. Underlying primary tumors were melanoma (n=3), renal cell carcinoma (n=6), colorectal cancer (n=5), breast carcinoma (n=2), prostate cancer (n=1), seminoma (n=1), medullar thyroid cancer (n=1) and esophageal adenocarcinoma (n=1).

Even though the patients also had larger lesions we only included nodules smaller than 500 mm^3, a size that corresponds to a mean diameter of approximately 10 mm (exact number: 9.85 mm), because the commercially available algorithm we applied (LungCare©, Siemens Medical Solutions, Erlangen, Germany) has been released for nodules less than 10mm in diameter. The minimum nodule volume we included in this evaluation was 15mm^3 (corresponding to a diameter of about 3 mm). We included not only nodules suspected for being metastases but also nodules that could potentially have a benign histology. Completely calcified nodules, however, were excluded. We included only solid nodules in our evaluation since the current software is not released for volume measurements of non-solid and part-solid nodules.

IMAGE ACQUISITION

We performed two low-dose non-contrast-enhanced chest CTs, followed by a contrast-enhanced standard-dose chest CT for clinical purposes. Between the two low-dose scans patients were asked to get off and on the table to simulate the conditions of a repeat scan for follow-up of a pulmonary nodule. Using that set up, growth or decrease in size of the lung lesions could reliably be excluded.

All scans were acquired on a 16-slice CT scanner (Mx8000 IDT, Philips Medical Systems, Cleveland, OH) using a spiral mode with 16 x 0.75mm collimation. The entire chest was scanned in about 10 seconds using a caudo-cranial scan direction. Scans were performed in full inspiration after appropriate instruction of the patients. In order to reproduce the standard situation in a screening setting, we used no spirometric control or respiratory belt. Exposure settings for the additional low-dose examinations were 30mAs and 120kVp (volume CT dose

index, $CTDI_{vol}$=2.2mGy) for patients weighing ≤80kg, and 30mAs at 140kVp for those weighing over 80kg ($CTDI_{vol}$=3.5mGy). Axial images were reconstructed at 1.0 mm thickness and 0.7mm increment, using a moderately soft reconstruction kernel (kernel "B"), the smallest field of view (FOV) that included the outer rib margins at the widest dimension of the thorax and a 512x512 matrix.

SEMIAUTOMATED VOLUME MEASUREMENTS OF PULMONARY NODULES

Data were transferred from the CT scanner to a digital workstation (Leonardo Workstation, Siemens Medical Solutions, Erlangen, Germany) with commercially available software for semiautomated volume measurements (LungCare©, Siemens Medical Solutions Erlangen, Germany). Nodules were identified by a single observer (one year of experience in radiology and trained for this specific task) using axial thin slab maximum intensity projections (slab thickness 10mm) displayed with window/center settings of 1500/-500HU. After manually marking a candidate nodule with a mouse click in the center of the nodule, the program automatically defines a volume of interest (VOI) around the nodule, which can be analyzed using volume rendered displays. Quantitative evaluation of a nodule is initiated with a second mouse click starting an automated volume measurement program described previously [8]. In this step the nodule of interest is segmented and the volume of the segmented area is calculated. This segmented area is shown by a yellow overlay on the nodule. No manual interaction was carried out to correct mismatches.

To minimize the influence of the separation process that distinguishes between the nodule itself and adjacent structures we only included lesions that had no direct contact with the pleura or vessels. That way it was assured that we solely focused on precisenessthe volume of an isolated solid lesion. For each patient, all nodules with a volume of 15-500 mm^3 meeting the inclusion criteria were included for evaluation.

Nodules on the second scan were identified with knowledge of the first scan matching the nodules by the combination of slice number, lung segment and distance to the pleura.

EVALUATION OF THE EFFECT OF SEGMENTATION PERFORMANCE

We visually determined the preciseness of the measurement software by assessing if the nodule was completely segmented. Nodules were categorized in two groups (A and B) based on whether the yellow overlay completely matched the nodule (A) or whether visual assessments determined a mismatch (B). Mismatch was defined as a visually obvious exclusion of a part of the investigated nodule from the segmented area.

Evaluation of Nodule Shape

In a separate reading session, an experienced observer (MP, 20 years of experience in radiology) visually assessed nodule shape and nodules were categorized into three subgroups. A nodule was defined as spherical when it showed a constant radius and as lobular when it showed a variable radius but with smooth outer margins. It was defined as irregular when the outer margins of the nodule were not smooth.

To investigate the effect of nodule shape on the performance of the segmentation algorithm, we determined how many nodules were spherical, non-spherical or irregular for completely (group A) or incompletely segmented nodules (group B).

Evaluation of Inspirational Level and Lung Volume

To compare the inspirational levels of the two scans, we calculated the lung volume for every scan using completely automated in-house developed software. The algorithm is similar to the one described previously [9]. The lungs are segmented from adjacent soft tissue structures (e.g., the mediastinum, vascular structures and the chest wall) by region growing from an automated starting point in the trachea, including all connected areas below -500HU. In a second step, trachea and main bronchi are excluded from the lungs. The number of voxels within the segmented lungs is multiplied by the size of a voxel to calculate total lung volume.

Natural variation in inspirational level between the two low dose scans was established as ratio between the lung volume on the second scan and the lung volume on the first scan.

Statistical Evaluation

Nodule volumes are given as mean ± standard deviation. Reproducibility of volume measurements was assessed by correlating nodule volumes in both scans using the Spearman's correlation coefficient for non-normally distributed populations.

Differences in volume (ΔV) were calculated by subtracting the volume measured on the

$$\Delta V = \frac{V_2 - V_1}{(V_2 + V_1)/2}$$

scan first (V_1) from the volume measured on the second scan (V_2). These differences were plotted against the mean nodule volume, using the approach described by Bland and Altman [10]:

The differences in volume measurements were normalized with respect to mean nodule volume to assess relative differences:

$$\Delta V_{rel} = 100\% * \frac{V_2 - V_1}{(V_1 + V_2)/2}$$

Limits of agreement were given as 95% confidence intervals (CI). Inter-scan variability was defined as the 95% CI of the relative differences. An increase in nodule volume above these upper limits of agreement can, with 95% confidence, be attributed to real growth.

Since the shape of a nodule was a binary variable, an ANOVA-test, comparing the variances of both groups, was performed to assess the impact of nodule shape on the relative difference in measured volumes. To assess the effects of inspirational level, we performed univariate regression for both groups with the relative difference as dependent variable and the ratio of lung volumes as independent variable. To assess the impact of nodule volume on measurement variation for both groups, we performed univariate linear regression with the relative SD (SD divided by mean volume) of both measurements as dependent variable and the mean nodule volume as independent variable. All statistics were calculated with SPSS statistical software package version 12.0 (SPSS, Chicago, USA). P-values <0.05 were considered significant.

RESULTS

NODULE CHARACTERISTICS

A total of 218 non-calcified, solid intraparenchymal nodules with a volume between 15 mm^3 and 500 mm^3 were eligible for analysis. Twelve calcified nodules and twelve pleura-based nodules were excluded from analysis. No vessel-attached nodules were detected.

Mean volume of all nodules was 123.0 mm^3 (±101.9 mm^3; range 16.4 mm^3 to 473.0 mm^3, median 82.7 mm^3). The number of eligible nodules per patient ranged from 0 (no metastases visible after therapy) in four patients to 62 nodules in one patient. The median number of nodules per patient was 6. None of the patients had only nodules that were larger than 500 mm^3.

Hundred six nodules were spherical, 30 nodules showed a non-spherical shape and 82 nodules showed irregular margins.

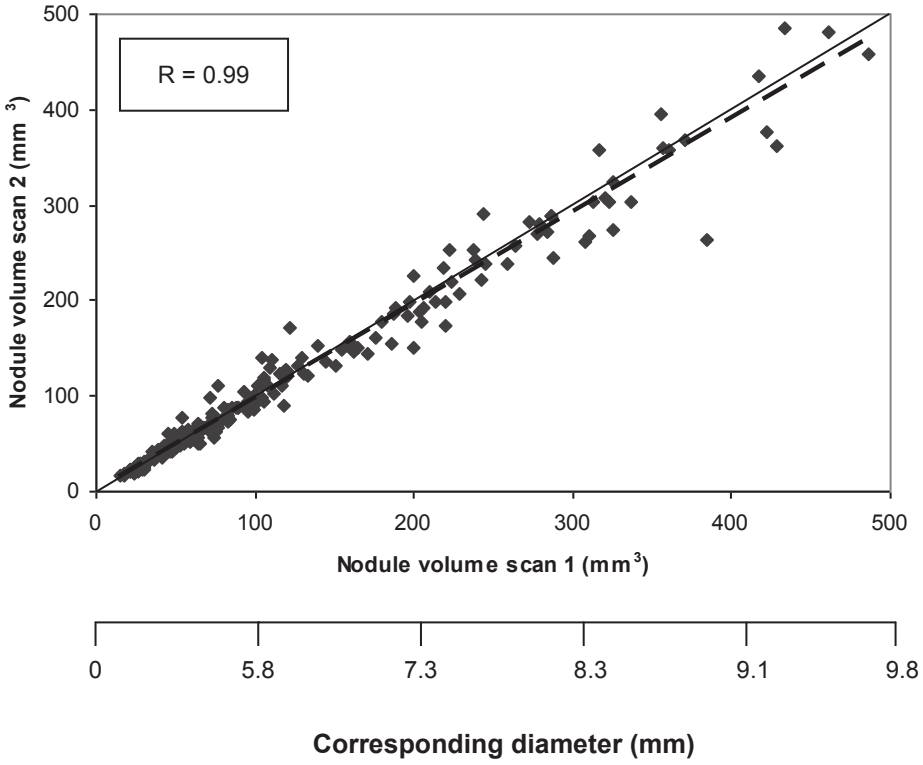


Figure 1

Volumetric results of each nodule. On the x-axis the volume measured on the first scan, on the y-axis the volume measured on the second scan. The dashed line shows the correlation between both measurements of the same nodule. Note that this line is almost equal to the (continuous) x=y line. Spearman's correlation coefficient = 0.99.

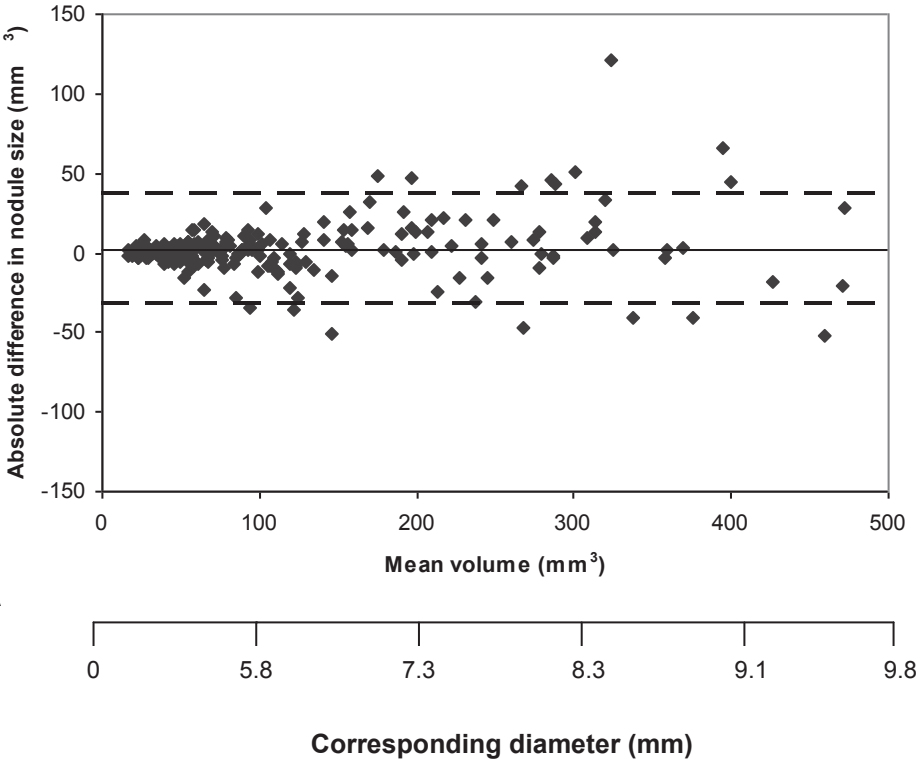
REPRODUCIBILITY OF VOLUME MEASUREMENTS

The reproducibility of volume measurements for the whole group was excellent with a Spearman's correlation coefficient of 0.99 (Figure 1). For the total group of nodules the mean difference in volume measurements amounted to 2.4mm^3, ranging from -53.0mm^3 to 120.8mm^3 (95% CI -32.0mm^3 to 36.7mm^3, Figure 2A). The mean relative difference amounted to 1.3% with a 95% CI of -21.2% to 23.8% and a range of -37.1% to 37.3% (Figure 2B).

Group	Number of nodules	Size (mm3)	Change in volume (mm3)	Mean relative difference (±SD)	Mean positive difference (range)
Completely segmented	106	16.4 – 369.2	0.04 – 15.35	0.28% (±6.2%)	4.2% 0.0%-34%
Incompletely segmented	112	21.4 - 472.7	2.5 – 120.79	1.61% (±14.5%)	12.4% 1.0%-37%

Table 1

Influence of algorithm performance on absolute and relative difference in volume


Absolute difference in nodule size (mm 3)
150
100
50
0
-50
-100
-150
0
100
200
300
400
500
Mean volume (mm3)
0
5.8
7.3
8.3
9.1
9.8
Corresponding diameter (mm)

2A

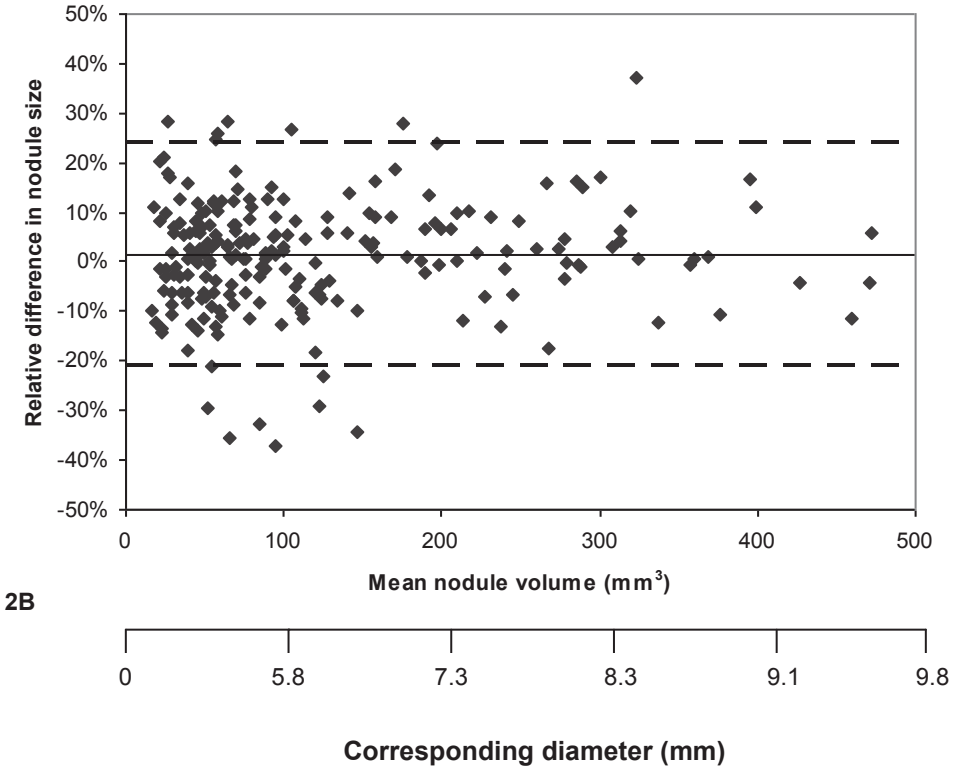


Figure 2

Inter-scan agreement of volume measurements of all nodules. The absolute (A) and relative (B) differences between both measurements are plotted against mean nodule volume. Mean difference is shown by the continuous line; upper and lower limits of agreement are shown by the dashed lines.

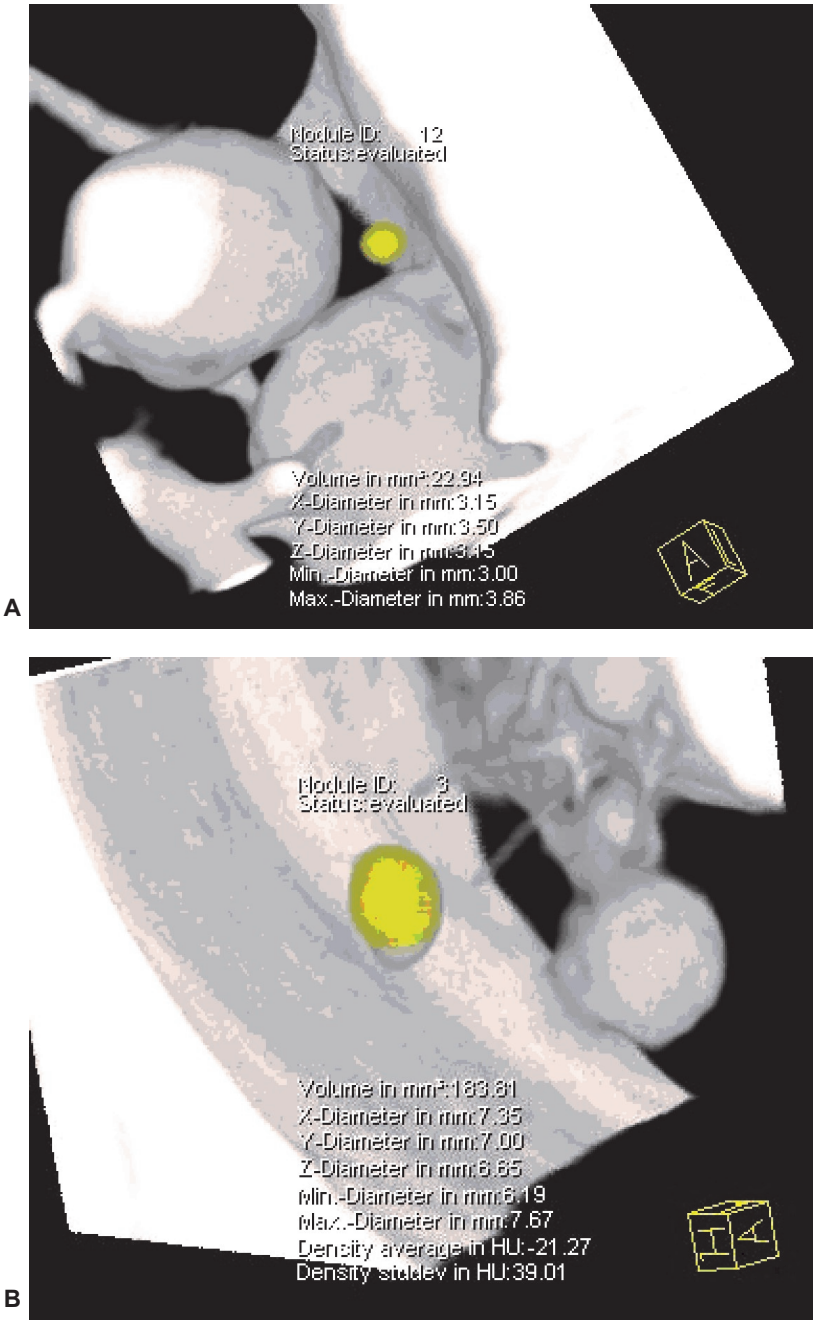


Figure 3

A complete (A) and an incomplete segmented nodule (B). Note that the yellow overlay does not cover the whole nodule in case of incomplete segmentation.

Influence of Nodule Size, Segmentation Performance and Inspirational Level

Analysis of the segmentation performance revealed that 106 of the 218 nodules (48.6%) had been completely segmented (group A, nodule size, 16.4mm^3 to 369mm^3) while 112 nodules (51.4%) were incompletely segmented (group B, nodule size, 21.4mm^3 – 473mm^3 (Table 1). All nodules were categorized to the same subgroup on both scans.

The group of completely segmented nodules (group A) contained only spherical lesions. All nodules with irregular margins (n=82) or a non-spherical shape (n=30) were incompletely segmented (group B). Group B did not contain any spherical nodules.

Inter-scan variability was -11.9% to 12.4% for completely segmented nodules. For incompletely segmented nodules this interval was more than twice as large: -26.8% to 30.0%. The maximum absolute difference between measured volumes on consecutive scans was 15.4mm^3 for completely segmented nodules and 121mm^3 for incompletely segmented ones. The maximum relative difference between volumes was 15% for completely segmented and 37% for incompletely segmented nodules. The measurement variability as given by the width of the 95% confidence interval standard deviations was 2.3 times smaller for completely segmented nodules than for incompletely segmented ones. This difference was statistically significant (F-test, $p< 0.0015$).

Ratio in lung volume between the second and the first scan ranged from 88% to 116% (Figure 4). Lung volume calculation failed in five patients due to the high number of nodules (n=2) or fibrosis (n=3), while lung volumes of the four patients without nodules could not be used for analysis.

No significant impact of mean nodule volume on the relative standard deviation was demonstrated for complete segmented nodules ($p=0.15$) or incomplete segmented nodules ($p=0.81$).

For complete segmented nodules, the relative difference decreased with increasing ratio of inspirational level ($r=-0.20;p=0.047$), while for incomplete segmented nodules we found no correlation ($p=0.67$).

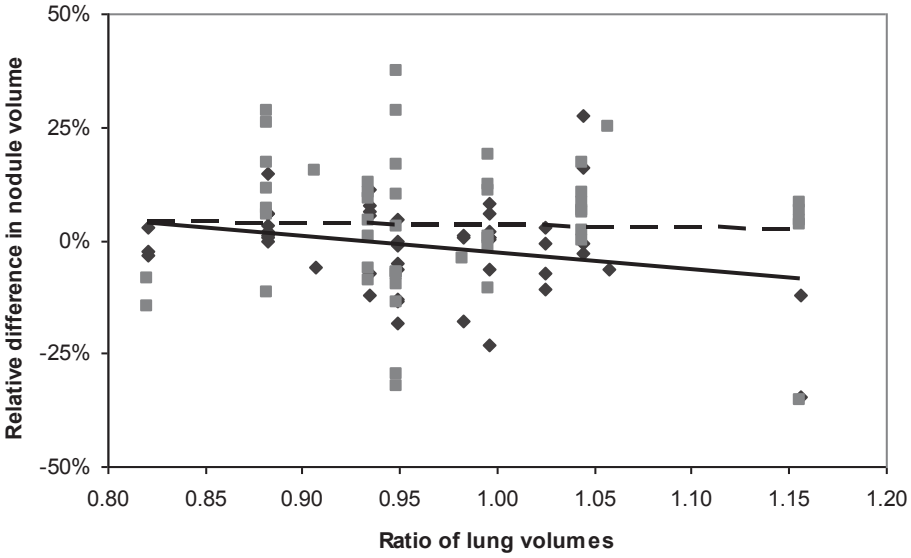


Figure 4

Correlation between ratios of lung volumes (lung volume on 2nd scan divided by lung volume on the 1st scan) and differences in nodule size. Black diamonds show completely segmented nodules, grey squares show incompletely segmented nodules. For incompletely segmented nodules, no correlation could be demonstrated (dashed line), while for completely segmented nodules, nodule volume decreased moderately, but significantly with increase in lung volume ($r=0.20$, $p<0.001$) as shown by the continuous line.

DISCUSSION

In this study we tested the preciseness of a commercially available and widely used algorithm for the assessment of nodule volumetry. To minimize effects of segmentation from adjacent structures, only nodules completely surrounded by lung tissue were included in the study group. Previous studies have shown that the applied software was very accurate for small spherical nodules [11], but our results suggest that the preciseness may substantially vary with nodule morphology. While the preciseness was extremely high for spherical solid nodules, the preciseness went down for non-spherical nodules and nodules with irregular shapes. This is especially noteworthy considering the fact that many nodules detected in a lung cancer screening setting do not show perfectly spherical shape with smooth margins. Thus their volume assessment using the semiautomated software we tested might be prone to considerable variations.

Wormanns et al applied a study design comparable to ours, using the same type of semiautomated software [8]. Although we used thinner collimations and included a higher number of nodules, we found comparable limits of agreement: -21.2% to 23.8% in our study versus -20.6% to 21.9% in the study by Wormanns. Wormanns and colleagues discussed that both, ill-defined shape and attachment to pleura or vessels of some of the included nodules could explain why reported variability was higher "in vivo" than seen in phantom studies. However, they did not further analyze the contribution of these factors separately. We specifically excluded pleural-based and vessel-attached nodules to be able to assess these factors separately. The reason why we still found similar limits of agreement despite these exclusion criteria is likely to be due to the higher number of irregularly shaped nodules in our study as compared to the group of nodules analyzed by Wormanns et al.

While Lungcare© uses a global thresholding method for the segmentation of a candidate nodule, the algorithm reported by Kostis et al uses a more sophisticated segmentation approach, which is developed to deal with irregular shapes [12]. With their algorithm, Kostis et al also found an effect of the initial nodule size on the extent of variability, while Wormanns had reported similar limits of agreement for different size ranges of nodules. We could not show a significant effect of lesion size on the variability of measurements for nodules that had been perfectly segmented.

One could argue that the preciseness of the segmentation algorithm could have been improved by manual modification. Lungcare© offers the possibility to adapt the segmented area by modifying the cut-off value of the crossection curves. The means of that adaptation process, however, are limited with respect to the fact that modification of the cut-off value is valid for the complete circumference of the nodule. As a consequence, moving the cut-off value of the crossection curve to the right (resulting in increased overlaid volume) helps to include parts of the nodules originally not included, but at the expense of including surrounding parts, leading

to an overestimation of the volume. Other important aspects are that the use of any manual adaptation or modification of the volumetric measurement would not only increase the radiologist's reading time but would also introduce another variable.

Changes of the inspiratory level turned out to be another source of variability. We found that higher inspirational level led to a decrease in measured volume. This finding is most likely due to the fact that the attenuation of the surrounding lung parenchyma changes with inspiratory level leading to an alteration of the cut-off value and thus the volume assessment. Although the contribution of inspiratory level was statistically significant, the quantities of volume changes introduced by it were only small.

Other potential sources of variability are interobserver and intra-observer variability. Since in our study the same single observer measured all nodules, interobserver variability did not contribute to the variability of our measurements. We did not specifically evaluate intra-observer variability but Wormanns and colleagues had already shown that the extent of intraobserver variability was negligible compared to the effect of interscan variability [8].

The main limitation of our study is that all results reported are valid only for that particular software release. Although this program is widely used, our results are not transferable to other algorithms. The limitation we found is related to the fact that a global thresholding method is used for the segmentation of a candidate nodule. Algorithms that use a more sophisticated segmentation algorithm are likely to achieve a higher preciseness also for irregularly shaped nodules.

CONCLUSION

The preciseness of the commercially available software tool tested is dependent on nodule morphology and was found to be less reproducible for non-spherical than for spherical solid nodules. The extent of variability decreases with increasing nodule size and higher inspirational level.

Taking the reported variation into account, the threshold to call an increased measured volume a real volume increase with a 95% confidence lies at 30% increase for an irregular shaped lesion. For spherical nodules this threshold can be lowered to 15%.

REFERENCE LIST

(1) Gietema H, Wang Y, Xu D et al. Pulmonary Nodules detected by Lung Cancer Screening: Inter-Observer Variability of Semiautomated Volume Measurements. 2006.

(2) Hasegawa M, Sone S, Takashima S et al. Growth rate of small lung cancers detected on mass CT screening. Br J Radiol 2000; 73(876):1252-1259.

(3) Kostis WJ, Yankelevitz DF, Reeves AP et al. Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT. Radiology 2004; 231(2):446-452.

(4) Revel MP, Lefort C, Bissery A et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. Radiology 2004; 231(2):459-466.

(5) Revel MP, Bissery A, Bienvenu M et al. Are Two-dimensional CT Measurements of Small Noncalcified Pulmonary Nodules Reliable? Radiology 2004; 231(2):453-458.

(6) Usuda K, Saito Y, Sagawa M et al. Tumor doubling time and prognostic assessment of patients with primary lung cancer. Cancer 1994; 74(8):2239-2244.

(7) Therasse P, Arbuck SG, Eisenhauer EA et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. J Natl Cancer Inst 2000; 92(3):205-216.

(8) Wormanns D, Kohl G, Klotz E et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004; 14(1):86-92.

(9) Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001; 20(6):490-498.

(10) Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307-310.

(11) Wormanns D, Fiebich M, Saidi M et al. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. Eur Radiol 2002; 12(5):1052-1057.

(12) Kostis WJ, Reeves AP, Yankelevitz DF et al. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Trans Med Imaging 2003; 22(10):1259-1274.

Part 3

Low Dose Chest CT in the Detection of Emphysema

Chapter 7

Monitoring of smoking-induced emphysema with multidetector-row computed tomography in a lung cancer screening setting: What is the minimum increase in emphysema scores required to distinguish real increase in extent of emphysema from interscan variability?

Hester Gietema
Arnold Schilham
Bram van Ginneken
Rob van Klaveren
Jan-Willem Lammers
Mathias Prokop

Accepted for publication in Radiology

ABSTRACT

PURPOSE

To retrospectively establish the minimum increase in emphysema scores (ES) required for detection of real increase in extent of emphysema with 95% confidence, using multidetector-row computed tomography (MDCT) in a lung cancer screening setting.

MATERIAL AND METHODS

The trial was approved by the Dutch ministry of health and by the ethics committee of each participating hospital. For our part of the study a waiver was received. Of the 1684 men screened at our hospital with low-dose MDCT (30mAs; 16 x 0.75mm collimation) between April 2004 and March 2005, we included only participants, who underwent repeat MDCT after three months on the same CT-scanner because of an indeterminate pulmonary nodule. Extent of emphysema was considered to remain stable in this short period. Extent of low-attenuated areas representing emphysema was computed for repeat and baseline scans as percentage of lung volume below three attenuation thresholds (-910HU, -930HU and -950HU). Limits of agreement were determined with the Bland-and-Altman approach and upper limits were used to deduce the minimum increase in ES required for detecting increase in extent of emphysema with 95% probability. Factors influencing the limits of agreement were determined.

RESULTS

In total 157 men (52-77y, mean 60y) were included in our study. The limits of agreement for differences between repeat and baseline scans were -13.4% to +12.6% at -910HU, -4.7% to +4.2% at -930HU and -1.3% to +1.1% at -950HU (percentages of total lung volume). Differences in ES showed weak to moderate correlation with variation in level of inspiration ($r=0.20$ to $r=0.49$, $p<0.05$). Scanner calibration could be excluded as factor contributing to variation in ES.

CONCLUSION

Increase in ES required to detect increase in extent of smoking-related emphysema with 95% probability varies between 1.1% at -950HU and 12.6% of total lung volume at -910HU for low-dose MDCT.

INTRODUCTION

To our knowledge, the most frequently used computer-based method to detect emphysema on CT is highlighting and quantifying low-attenuated areas [1-3], first described by Müller et al [4]. Disappearance of lung tissue produces a relative increase of air within a voxel, which results in a lowered attenuation within the voxel. The percentage volume of the highlighted voxels can be calculated relative to total lung volume, resulting in a voxel index or emphysema score (ES) in the range from 0% to 100%.

Presently there are several ongoing lung cancer screening trials [5-8]. Since lung cancer and emphysema share smoking as the main risk factor, CTs performed in these trials may provide suitable data for studying the prevalence and natural course of smoking-related emphysema in relatively healthy participants [9]. These data could be used to select groups of smokers in whom more aggressive risk-modifying treatment is necessary to prevent development of severe lung destruction and airflow limitation. Before an automated method can be used for screening and monitoring purposes, more information about an issue as interscan variation and the effect of factors that have been shown to influence emphysema scores, such as level of inspiration [10] and scanner calibration [11], is required. Data about the interscan variation are useful to distinguish real progression of the extent of emphysema from measurement variation.

Thus, the aim of our study was to retrospectively establish the minimum increase in emphysema scores required for detection of real increase in extent of emphysema with 95% confidence using multidetector-row computed tomography (MDCT) in a lung cancer screening setting.

MATERIALS AND METHODS

PARTICIPANTS

The NELSON-project is a population-based randomized Dutch-Belgian multi-center lung cancer screening trial that studies 16,000 current and former heavy smokers. The trial was approved by the Dutch ministry of health and by the ethics committee of each participating hospital. For our part of the study a waiver was received. Selection of participants was performed by sending a questionnaire about smoking history and other health-related issues to citizens between 50 and 75 years of age who lived in the areas around the participating centers. Among the respondents, subjects meeting the inclusion criteria of a minimum of 16 cigarettes/day for 25 years or 11 cigarettes/day for 30 years were asked to participate in the trial. Since men had a better chance to meet this inclusion criterion having smoked this minimum number of cigarettes during their life, we started to recruit men. After being informed about, among others, the radiation dose that was exposed to the participants, those who gave written informed consent were equally randomized to either the screening arm or the control arm. Participants in the screening arm underwent baseline CT to assess the prevalence of lung cancer in this population and will undergo two further CT scans in years two and four to establish the one year and three year incidence of lung cancer in this population. Participants with a moderate or poor self-reported health status who were unable to climb two flights of stairs were excluded from participation.

Between April 2004 and March 2005, 1684 male participants received baseline screening in our center. From these participants, we included all participants who, according to the trial protocol, received a short-term repeat chest CT after three months because of an indeterminate nodule (50-500mm^3) found on the baseline scan. Extent of emphysematous lung destruction was considered to remain stable in this short period. To test this assumption, results of pulmonary function tests, performed in subgroup of the investigated population on the same days as the baseline and repeat CTs were performed, were compared. No medical intervention was applied. Since CT scans were performed on various 16-detector-row scanners, we selected only participants scanned twice on the same scanner.

CT SCANNING

Scanning was performed using 16-detector-row CT (either one of two Mx8000 IDT scanners or one Brilliance 16P scanner, Philips Medical Systems, Cleveland, OH). All scans were realized in about 12 seconds in helical mode with 16 x 0.75mm collimation and 15mm table feed per rotation (pitch = 1.3), in a caudo-cranial scan direction to minimize breathing artifacts. Participants were asked to take a deep breath and to hold their breath. No spirometric gating was applied because spirometric gating is not standard in a lung cancer screening setting and spirometric gating would therefore make the results less applicable in a standard lung cancer screening setting. No intravenous contrast was injected. Exposure settings were 30mAs at 120kVp for patients weighing ≤80kg, and 30mAs at 140kVp for those weighing more than 80kg without dose modulation. All participants received the same radiation dose during both scans. Axial images of 1.0mm thickness were reconstructed at 0.7mm increment, using the smallest field of view (FOV) that included the outer rib margins at the widest dimension of the thorax. All scans were reconstructed with a 512x512 matrix and a moderately soft "B" kernel.

EMPHYSEMA QUANTIFICATION

Data were transferred from the CT scanner to a digital workstation. The extent of low-attenuated areas was fully automated for quantification with in-house developed software. Total lung volume was calculated using the following steps. Firstly, segmentation of trachea, left and right lung was performed by a region growing program starting in the trachea, which included all connected areas below -500HU. In a second step, trachea and main bronchi were separated from the lungs. The algorithm is similar to the one described by Hu et al [12]. After segmenting the lung, the data were subjected to a median noise-reducing filter [13].
The extent of low-attenuation areas was determined by computing lung volume with CT attenuation below a certain attenuation threshold as percentage of total lung volume. We studied the three attenuation thresholds often mentioned in literature: -910HU, -930HU and -950HU [4;14-16].

Influence of Level of Inspiration

CT numbers of voxels representing the lungs are lowered when a participant reaches a higher level of inspiration due to increasing relative amount of air per voxel as described by Kalender et al. [17]. So, the extent of low-attenuated areas increases not only when the extent of emphysema increases, but also when a participant reaches a deeper level of inspiration during a repeat scanning than in baseline scanning. For this reason, the correlation between natural variation in level of inspiration of baseline and repeat scan of each participant and changes in emphysema scores was evaluated. Total lung volume as calculated from each scan was used as surrogate for level of inspiration.

Quality Control

Since the method of highlighting and quantifying low-attenuated areas starts from a fixed threshold, the method is sensitive to CT number shifts due to, for example, X-ray tube ageing. We performed weekly air calibrations and completed the screening scans within 24 hours after calibration. In addition, we performed scans of a quality control phantom before and after each data acquisition session. This phantom consisted in a foam body (mimicking emphysematous areas) of 320mm in diameter including two cylinders, each 80mm in diameter (Figure 1). One cylinder contained air; the other cylinder was filled with plastic. The phantom was scanned at 120kVp with otherwise identical parameters as applied to the participants.

Average CT numbers for each structure were measured in a circular region of interest (ROI; 100mm^2) manually drawn by one observer (HG, three years of experience in radiology) in the center of both cylinders and in the periphery of the foam body, 2 cm from the outside border. We performed five scans of the phantom in one session at the start of the study and after three months in order to assess the variation in CT numbers within one session. Changes in average CT numbers within the ROI during the period of data collection were determined.

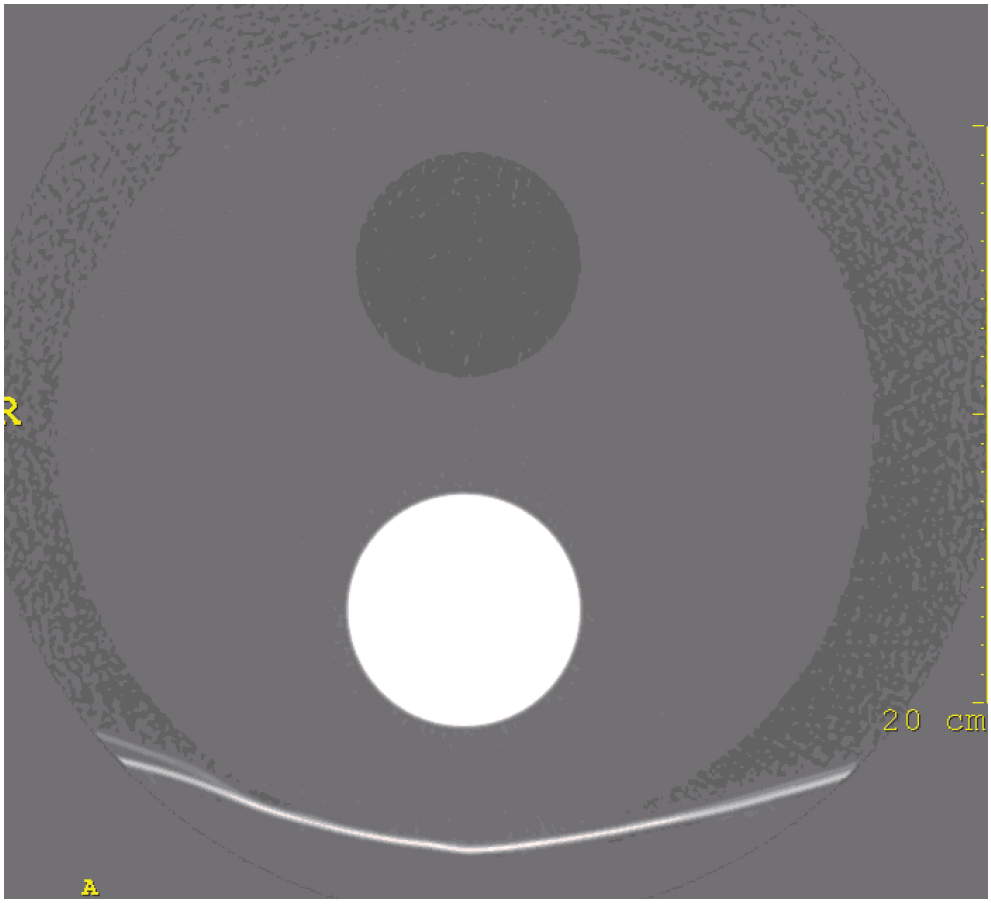


Figure 1

Phantom used to monitor CT-values during the study. The phantom consists in a foam body, 320 mm in diameter, and two free spaces of 80mm in diameter. One of the free spaces is filled with plastic, while the other contains only air.

STATISTICAL ANALYSIS

All statistical calculations were performed using SPSS statistical software release 13.0 (SPSS Inc, Chicago, Ill, USA). We calculated means, standard deviations (SD) and 95% confidence intervals (CI) for normal distributed differences in ES and medians and interquartile ranges for non-normal distributed emphysema scores. Changes in emphysema scores were given as percentages of total lung volume.
Forced expiratory volumes in one second (FEV_1) performed on the day of baseline scanning (FEV_{1A}) and on the day of repeat scanning (FEV_{1B}) were compared after logistic transformation:

$$e^{(\ln FEV1B - \ln FEV1A)} = (FEV_{1B}) / (FEV_{1A})$$

The results of ln (FEV_{1A}) and ln (FEV_{1B}) were compared using paired samples t-testing.
Differences in emphysema scores (ΔES) were calculated by subtracting the ES measured on the baseline scan (ES_1) from the ES measured on the repeat scan (ES_2). These differences were plotted against the mean of both ES, using the approach described by Bland and Altman [18]:

$$\Delta ES = \frac{ES_2 - ES_1}{(ES_2 + ES_1)/2}$$

Limits of agreement were given as 95% confidence intervals (CI). For monitoring purposes, an increase in emphysema score above these upper limits of agreement can, with 95% confidence, be attributed to real increase in extent of emphysema.
To asses the repeatability of the quantification of the extent of low-attenuated areas, we calculated coefficients of variation as ratio of the within subject SD to the mean of both measurements. In order to determine if these coefficients of variation were related to the extent of emphysema, represented by the mean of the two measurements, we calculated Spearman's correlation coefficient for each attenuation threshold.

Correlation coefficients between difference in total lung volumes and difference in emphysema scores for each pair of scans were best after logarithmic transformation of lung volumes as described by Shaker et al. [10]. We determined the corresponding Pearson's correlation coefficients for each attenuation threshold in order to assess if a correction factor for the level of inspiration could be calculated.
P-values <0.05 were considered significant.

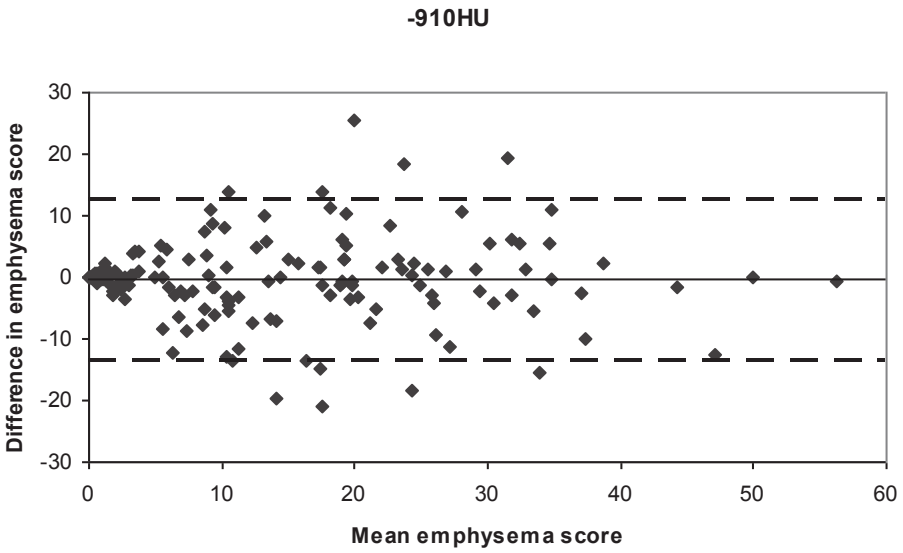
RESULTS

EMPHYSEMA SCORES

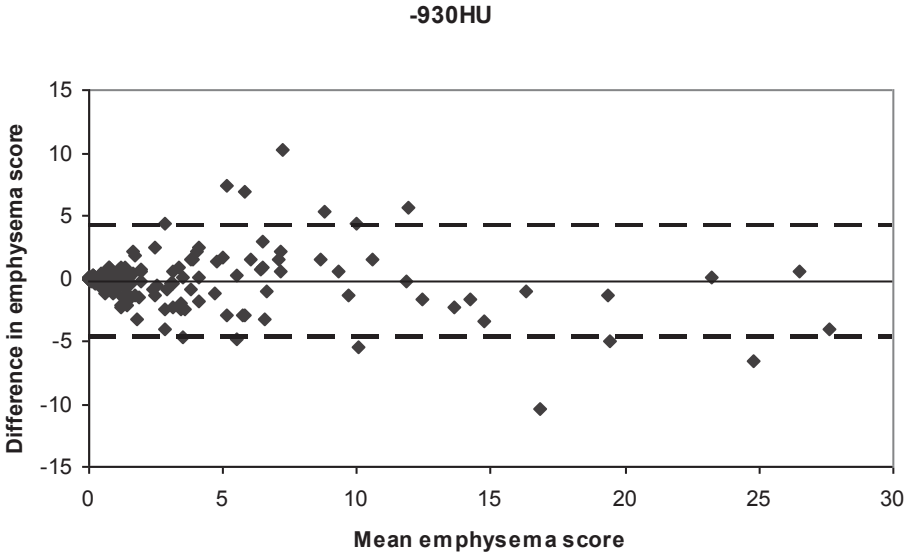
Between April 2004 and March 2005, 249 consecutive male participants underwent a baseline and a 3-month follow-up scan. The group of participants that received both scans on the same scanner included 157 participants (52-77 yrs, mean 60 yrs). These participants were further analyzed (Figure 2).

Sixty subjects underwent pulmonary function testing on the day of baseline scanning and again on the day of repeat scanning. The FEV_1 did not change significantly in the three months interval (p=0.311). The 95% confidence interval of the ratio of the FEV_1 during both tests ranged from 0.99 to 1.03, showing that the variation in FEV_1 was only 4% in this three months interval.

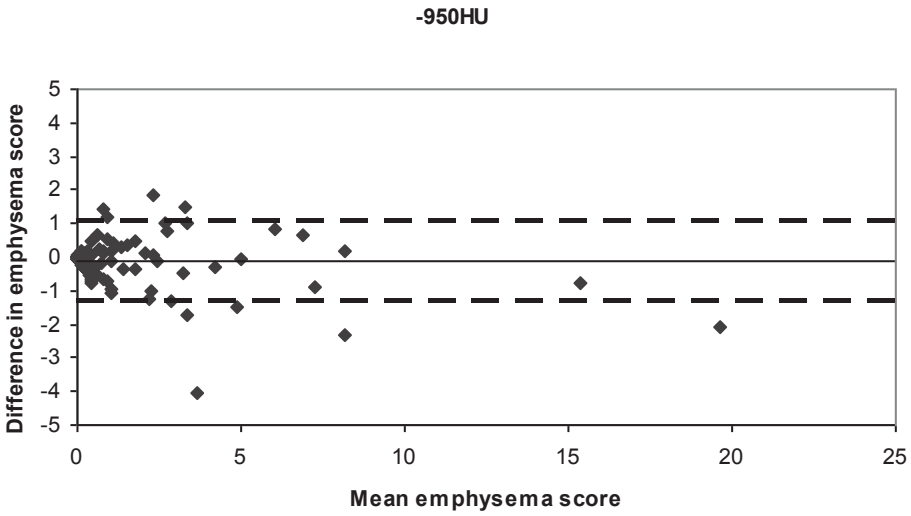
Emphysema scores ranged from 0.0% for volume with an attenuation below -950HU to 56.5% for volume with an attenuation below -910HU for baseline scans (Table 1). Median emphysema scores ranged from 0.08% for volume with an attenuation below -950HU for repeat scans to 11.8% for volume with an attenuation below -910HU for baseline scans (Table 1) . Coefficients of variation ranged from 0.0% to 141% and decreased with increasing extent of emphysema (Table 1). Mean difference in emphysema scores ranged from -0.1% for volume with an attenuation below -950HU to -0.41% for volume with an attenuation below -910HU for baseline scans (Figure 2).


-910HU
Difference in emphysema score
30
20
10
0
-10
-20
-30
0
10
20
30
40
50
60
Mean emphysema score

2A


-930HU
Difference in emphysema score
15
10
5
0
-5
-10
-15
0
5
10
15
20
25
30
Mean emphysema score

2B



2C

Figure 2

Bland-Altman plots for emphysema scores at -910HU (A), -930HU (B) and -950HU (C) The x-axes show the means of emphysema scores on the baseline scan and the repeat scan, the y-axes show emphysema scores on the repeat scan minus emphysema scores on the baseline scan expressed as percentage of total lung volume. The mean differences are shown by the continuous lines, the limits of agreement are shown by the dashed lines. An increase in emphysema score more than the upper limit of agreement or a decrease below the lower limit of agreement has an 95% likelihood to be a real progression or regression of emphysema.

	Threshold		
	-910HU	**-930HU**	**-950HU**
Baseline scan			
Median	11.8%	1.5%	0.17%
Interquartile range	2.7%-21.9%	1.3%-4.9%	0.05%-0.79%
Range	0.0%-56.5%	0.0%-29.7%	0.0%-20.7%
Repeat scan			
Median	8.9%	1.2%	0.08%
Interquartile range	2.3%-22.0%	0.17%-4.8%	0.02%-0.81%
Range	0.0%-55.8%	0.0%-26.8%	0.0%-18.6%
Difference			
Mean (95% confidence interval)	-0.41% (-13.4% to 12.6%)	-0.23% (-4.7% to 4.2%)	-0.1% (-1.3% to 1.1%)
Coefficient of variation			
Median	34%	23%	58%
Interquartile range	8%-55%	8%-35%	19%-98%
Range	0.0%-139%	0.0%-69%	0.0%-141%
Spearman's correlation coefficient	-0.42	-0.42	-0.34

Table 1

Emphysema scores for the study population. The emphysema scores describe the percentage of lung tissue below the designated threshold. Differences represent percentages of total lung. volume. Coefficients of variation represent percentage of the mean score for an individual at a designated threshold. Note that the correlation between coefficient of variation decreases with increasing extent of emphysema as shown by Spearman's correlation coefficients ($p<0.0001$ for all thresholds).

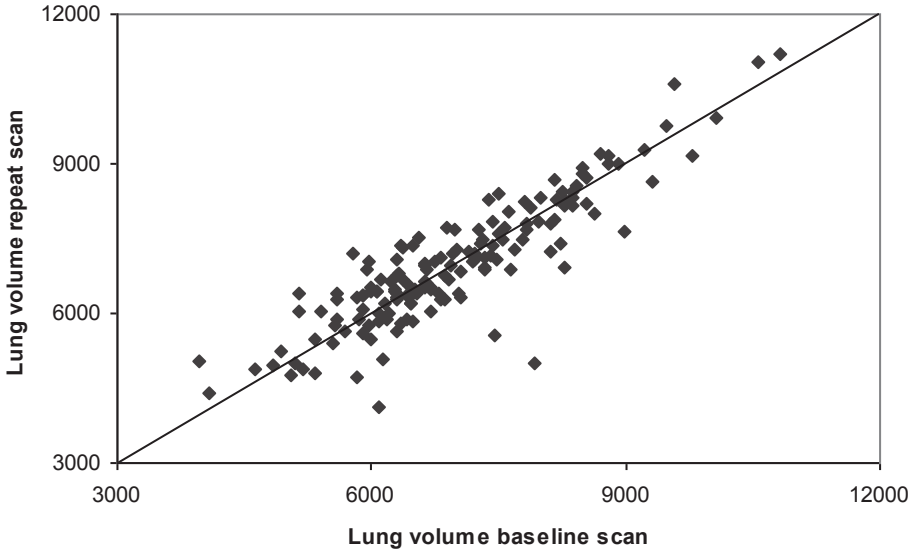
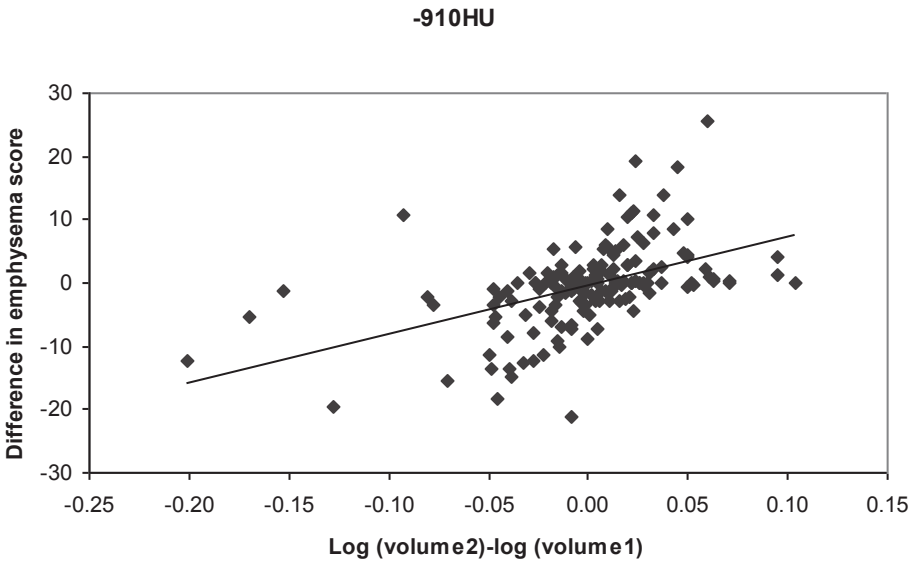


Figure 3

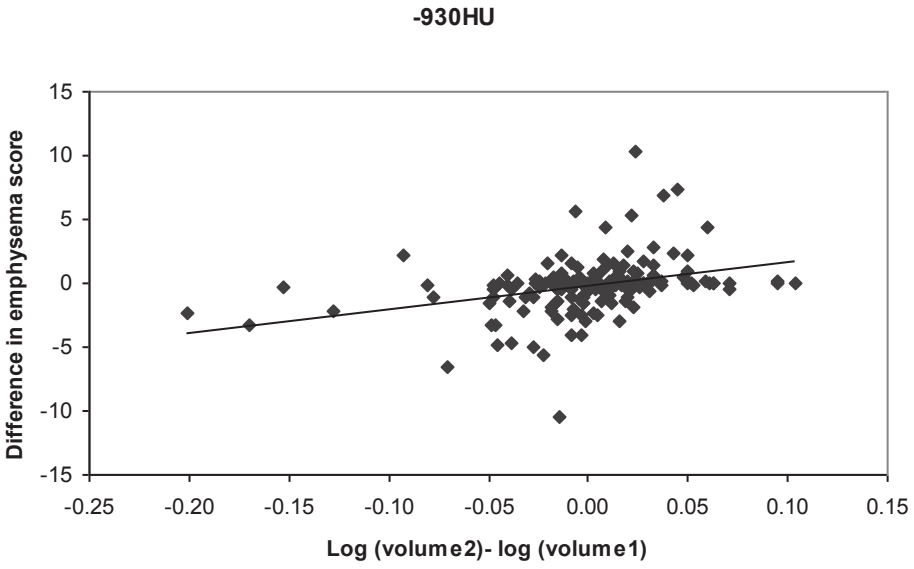
Correlation between lung volumes on the baseline scans and lung volumes on repeat scans. Identical volumes are demonstrated by the continuous line. No systematic difference between lung volume on baseline scan and repeat scan could be demonstrated.

LEVEL OF INSPIRATION

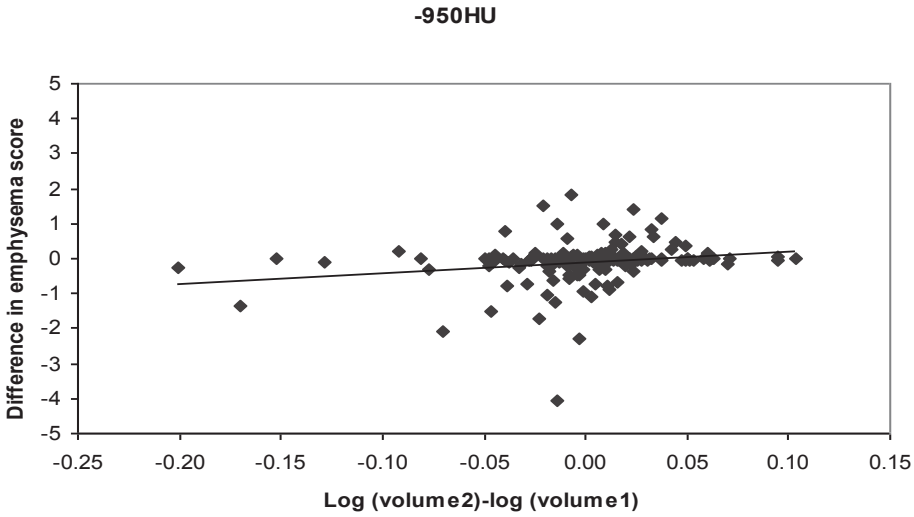
Mean total lung volume was 6935ml (± 1267ml) for the baseline scans and 6945ml (± 1322ml) for the repeat scans. Although many participants were not able to repeat the inspirational volume of the baseline scan during repeat scanning, the level of inspiration was not statistically different for both scans on cohort level (p=0.8, Figure 3). After logarithmic transformation we could demonstrate a significant ($p<0.001$ for -910HU and $p<0.001$ for -930HU, $p<0.01$ for -950HU) but weak to moderate correlation between changes in level of inspiration (lung volume) and emphysema scores for all thresholds ($r = 0.49$ for -910HU, $r = 0.33$ for -930HU and $r = 0.20$ for -950HU; Figure 4).


-910HU
Difference in emphysema score
30
20
10
0
-10
-20
-30
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
Log (volume2)-log (volume1)

4A


-930HU
Difference in emphysema score
15
10
5
0
-5
-10
-15
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
Log (volume2)- log (volume1)

4B



4C

Figure 4

Correlation between difference between inspiration level on repeat scans (volume 2) after logarithmic transformation and inspiration level on baseline scans (volume 1) after logarithmic transformation and difference in emphysema scores between repeat scans (ES2) and baseline scans (ES1). Significant, but low to moderate correlations could be demonstrated.

QUALITY CONTROL

The average CT-value for foam was -967.9HU (±2.0HU) for the 5 scans performed in succession at the beginning of our study and -969.2 HU (±2.2HU) for the 5 scans performed in succession after 3 months, while the average CT-value from April 2004 to March 2005 was 968HU (±2.7HU). Variation in measured CT numbers was independent of the time of the day. The SD of 2.7HU is within the range of tolerance reported by the vendor (0-4 HU).

DISCUSSION

Our results provide information of interscan variation of the quantification of low-attenuated areas, representing emphysema, in a cohort of current and former heavy smokers participating in a lung cancer screening trial. This screening trial aims to detect lung cancer in a curable stage, so participants have to be able to undergo surgery. For this reason, participants with severe airflow limitation were excluded from participation, resulting in a population under investigation with a relatively low extent of emphysema. While in the early nineties the quantification of low-attenuated areas had to be performed slice-by-slice and took hours per scan [19], nowadays, it takes less than 5 minutes for a complete CT and can be applied in large groups of patients. Shaker et al and Gierada et al have demonstrated that determining the extent of emphysema is highly repeatable on a cohort level in patients with large areas of destructed tissue [19;20], but they did not report the limits of agreement, while knowledge about the interscan variability is mandatory to distinguish between real increase in extent of emphysema and measurement variability in a monitor setting. With data reported in our study an increase in emphysema score of more than the corresponding upper limit of agreement can, with 95% likelihood, be subjected to real increase in extent of emphysema.

Since Müller et al introduced the quantification of the extent of emphysema highlighting low-attenuated areas on CT images [4], this method has been used for several scanning techniques. Müller et al. validated the technique to macroscopic histology for a single contrast-enhanced 10mm slice and found that -910HU was the best threshold to detect the extent of macroscopic emphysema. Gevenois et al determined the optimum attenuation threshold for high-resolution CT and recommended -950HU for both microscopicly and macroscopicly detected emphysema [15;21]. The difference in optimum attenuation threshold was subjected to variation in slice thickness and this effect was also investigated by Kemerink et al. [22]. Park et al. reported a high correlation between emphysema quantification on 2D an 3D datasets, making the technique also applicable to volumetric data [23]. Recently Madani et al. compared the extent of both microscopic and macroscopic detected emphysema to the quantification of the extent of low-attenuated areas with multidetector-row CT and reported -960HU to -970HU as optimum attenuation threshold for MDCT [24]. However, they applied less radiation dose than Gevenois et al (140kVp; 80mAs versus 137kVp; 255mAs) and Mishima et al. already showed the impact of applied radiation dose on the extent of low-attenuated areas [25]. Finally Parr et al. investigated several attenuation thresholds for monitoring purposes and recommended -930HU as optimum attenuation threshold to monitor the progression or regression in the extent of emphysema [16]. To our knowledge, there is no consensus about the optimum scanning technique for the quantification of the extent of emphysema by calculating the extent of low-attenuated areas and no consensus on the optimum attenuation threshold [3].

Therefore, we investigated the limits of agreement of the extent of low-attenuated areas on repeated scans for three attenuation thresholds often used in literature so far. Our results do not provide any information about the accuracy of emphysema scores for detecting lung destruction since we studied healthy participants and did not have any histological specimens available.

The effect of level of inspiration on lung attenuation is well described [17;26;27]. Shaker et al reported a large variability of correlation coefficients between the emphysema score and the total lung volume for the lower range of emphysema scores, while for more severe emphysema a more stable correlation could be reported [10]. We demonstrated a low to moderate but significant correlation between natural variation in level of inspiration and changes in emphysema scores, but also a large variation in this effect. Spirometric control could have narrowed the limits of agreement, but would also have limited our results to spirometric controlled CTs. Since spirometric controlled scanning is not available in many hospitals, we have performed the CTs in end-inspiratory volume as usual in clinical routine.

In our study a low-dose protocol was applied, since radiation dose has an intrinsic risk of inducing neoplasm. For a structure of interest with high contrast to its surroundings such as a pulmonary nodule in lung parenchyma, the detection and segmentation of this lesion are not affected by accompanying increase in image noise [28;29]. But for emphysema, especially for low extents of lung destruction, there is a low contrast between the destructed areas and the normal lung parenchyma. In that situation the increased image noise raises emphysema scores, which can be reduced, but not excluded, by the application of a noise reduction filter [30].

Quality control showed that regular scanner calibration for air resulted in stable CT numbers. The small variations in CT numbers were within the range of tolerance of our scanner, but still may contribute to variations in emphysema scores.

Although our results can provide useful information for monitoring high-risk participants in a screening setting, the study has also an important limitation. Our study has been performed in a lung cancer screening setting and the results are therefore useful in a low-dose setting but not necessarily applicable to a clinical setting with standard radiation dose. However, to our knowledge, emphysema quantification in large cohorts is mainly performed for study purposes with low-dose scans [9;20;31].

CONCLUSION

Although emphysema scores in a lung cancer screening setting are highly reproducible on cohort level, individual variation can be substantial. An increase in emphysema score of at least 1.1% for -950HU to 12.6% for -910HU is required for detection of increase of extent of emphysema with 95% confidence when monitoring smoking-induced emphysema with low-dose CT in a lung cancer screening setting.

REFERENCE LIST

(1) Arakawa A, Yamashita Y, Nakayama Y et al. Assessment of lung volumes in pulmonary emphysema using multidetector helical CT: comparison with pulmonary function tests. Comput Med Imaging Graph 2001; 25(5):399-404.

(2) Kinsella M, Muller NL, Abboud RT et al. Quantitation of emphysema by computed tomography using a "density mask" program and correlation with pulmonary function tests. Chest 1990; 97(2):315-321.

(3) Madani A, Keyzer C, Gevenois PA. Quantitative computed tomography assessment of lung structure and function in pulmonary emphysema. Eur Respir J 2001; 18(4):720-730.

(4) Muller NL, Staples CA, Miller RR et al. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988; 94(4):782-787.

(5) Church TR. Chest radiography as the comparison for spiral CT in the National Lung Screening Trial. Acad Radiol 2003; 10(6):713-715.

(6) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.

(7) Sone S, Takashima S, Li F et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet 1998; 351(9111):1242-1245.

(8) van Klaveren RJ, Habbema JD, de Koning HJ et al. [Screening for lung cancer in the Netherlands: the role of spiral CT scan]. Ned Tijdschr Geneeskd 2001; 145(11):521-526.

(9) Omori H, Nakashima R, Otsuka N et al. Emphysema detected by lung cancer screening with low-dose spiral CT: prevalence, and correlation with smoking habits and pulmonary function in Japanese male subjects. Respirology 2006; 11(2):205-210.

(10) Shaker SB, Dirksen A, Laursen LC et al. Volume adjustment of lung density by computed tomography scans in patients with emphysema. Acta Radiol 2004; 45(4):417-423.

(11) Stoel BC, Stolk J. Optimization and standardization of lung densitometry in the assessment of pulmonary emphysema. Invest Radiol 2004; 39(11):681-688.

(12) Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001; 20(6):490-498.

(13) Hilts M, Duzenli C. Image filtering for improved dose resolution in CT polymer gel dosimetry. Med Phys 2004; 31(1):39-49.

(14) Dirksen A, Friis M, Olesen KP et al. Progress of emphysema in severe alpha 1-antitrypsin deficiency as assessed by annual CT. Acta Radiol 1997; 38(5):826-832.

(15) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.

(16) Parr DG, Stoel BC, Stolk J et al. Influence of calibration on densitometric studies of emphysema progression using computed tomography. Am J Respir Crit Care Med 2004; 170(8):883-890.

(17) Kalender WA, Rienmuller R, Seissler W et al. Measurement of pulmonary parenchymal attenuation: use of spirometric gating with quantitative CT. Radiology 1990; 175(1):265-268.

(18) Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307-310.

(19) Gierada DS, Yusen RD, Pilgram TK et al. Repeatability of Quantitative CT Indexes of Emphysema in Patients Evaluated for Lung Volume Reduction Surgery. Radiology 2001; 220(2):448-454.

(20) Shaker SB, Dirksen A, Laursen LC et al. Short-term reproducibility of computed tomography-based lung density measurements in alpha-1 antitrypsin deficiency and smokers with emphysema. Acta Radiol 2004; 45(4):424-430.

(21) Gevenois PA, De Vuyst P, De M, V et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1996; 154(1):187-19

(22) Kemerink GJ, Kruize HH, Lamers RJ et al. CT lung densitometry: dependence of CT number histograms on sample volume and consequences for scan protocol comparability. J Comput Assist Tomogr 1997; 21(6):948-954.
(23) Park KJ, Bergin CJ, Clausen JL. Quantitation of emphysema with three-dimensional CT densitometry: comparison with two-dimensional analysis, visual emphysema scores, and pulmonary function test results. Radiology 1999; 211(2):541-547.
(24) Madani A, Zanen J, de Maertelaer V et al. Pulmonary Emphysema: Objective Quantification at Multi-Detector Row CT--Comparison with Macroscopic and Microscopic Morphometry. Radiology 2006;2382042196.
(25) Mishima M, Itoh H, Sakai H et al. Optimized scanning conditions of high resolution CT in the follow-up of pulmonary emphysema. J Comput Assist Tomogr 1999; 23(3):380-384.
(26) Gotway MB, Lee ES, Reddy GP et al. Low-dose, dynamic, expiratory thin-section CT of the lungs using a spiral CT scanner. J Thorac Imaging 2000; 15(3):168-172.
(27) Lamers RJ, Thelissen GR, Kessels AG et al. Chronic obstructive pulmonary disease: evaluation with spirometrically controlled CT lung densitometry. Radiology 1994; 193 (1):109-113.
(28) Itoh S, Ikeda M, Arahata S et al. Lung cancer screening: minimum tube current required for helical CT. Radiology 2000; 215(1):175-183.
(29) Revel MP, Lefort C, Bissery A et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. Radiology 2004; 231(2):459-466.
(30) Schilham A, Van Ginneken B, Gietema H et al. Local Noise Weighted Filtering for Emphysema Scoring of Low-dose CT Images. IEEE Transactions on Medical Imaging 2006; 25(4):451-463.
(31) Parr DG, Stoel BC, Stolk J et al. Validation of computed tomographic lung densitometry for monitoring emphysema in {alpha}1-antitrypsin deficiency. Thorax 2006;thx.

Chapter 8

Early detection of emphysema: Computed tomography versus pulmonary function testing

Hester Gietema
Ivo van der Lee
Pieter Zanen
Bram van Ginneken
Arnold Schilham
Rob van Klaveren
Carola van Iersel
Jules van den Bosch
Mathias Prokop
Jan-Willem Lammers

Submitted

ABSTRACT

OBJECTIVE

To establish the extent of moderate and severe emphysema detected on high-resolution computed tomography (HRCT) in a lung cancer screening setting for participants with and without pulmonary function impairment or gas exchange impairment.

METHODS

Between April 2004 and March 2005, we included 545 male current and former heavy smokers (51-74y, mean 62y) participating in a lung cancer screening trial (NELSON) with baseline low-dose HRCT (16x0.75mm slice collimation) in who also flow-volume curves and diffusion capacity testing were assessed. Moderate emphysema was determined as areas with an attenuation between -910 Hounsfield units (HU) and -950HU, where as areas with an attenuation below -950HU represented severe emphysema. Both were expressed as emphysema score (ES), representing percentage of total lung volume. The extent of moderate and severe emphysema was assessed for participants with and without pulmonary function impairment or gas exchange impairment.

RESULTS

Twelve percent lung volume with moderate emphysema appears to be the optimal cut-off to result in a lowered FEV_1/VC, while 9% of total lung volume with moderate emphysema appears to be the optimal cut-off to result in a lowered $Tlco/V_A$ ratio. The optimal cut-off for severe emphysema is 0.15% for both a lowered FEV_1/VC ratio and a lowered $Tlco/V_A$ ratio.

CONCLUSION

The probability of moderate emphysema to result in a pulmonary function impairment or gas exchange impairment is low, while small amounts of severe emphysema already resulted in pulmonary function impairment and gas exchange impairment.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is the most frequent chronic disease in developed countries and is predicted to be the third cause of death in 2020 [1]. COPD is functionally defined on the extent of airflow obstruction, which can be detected by pulmonary function testing (PFT): impairment of the forced expiratory volume in 1 second (FEV_1) is fundamental for the diagnosis according to the guidelines of the Global initiative on Obstructive Lung Diseases (GOLD) [2]. Emphysema is anatomically defined as an abnormal permanent enlargement of the airspace distal to the terminal bronchioles without fibrosis [3]. Several investigators correlated the extent of emphysema determined via CT with pulmonary function parameters and reported that the FEV_1/VC and $Tlco/V_A$ ratios were the best correlating parameters [4;5], but relations were not very strong. However, the detection of early changes can enable more aggressive risk-modifying interventions in this group of patients [6]. Moreover, emphysema can cause airflow obstruction, but emphysema can also exist without impairment of the FEV_1 [7].
Because of the anatomical definition, histology is required for the diagnosis of emphysema, but computed tomography (CT) can non-invasive provide anatomical information and the extent of emphysema detected with CT has been shown to correlate well with histology [8-11]. Therefore, CT can be an attractive alternative to detect emphysema before it reaches the symptomatic stage causing airflow obstruction. To our knowledge, the most frequently used technique is the one firstly described by Müller et al [12] highlighting low-attenuated areas, representing emphysema. The extent of emphysema is expressed as percentage of total lung volume in a range from 0% to 100%. This method has been validated for high-resolution CT against pathology for both microscopic and macroscopic techniques by Gevenois and co-workers [10;13].

Since pulmonary function testing (PFT) is more easily performed and to lower costs in a large population at-risk than CT-scanning, the aim of our study was to assess the extent of moderate (loss of lung tissue) and severe emphysema (complete destruction of lung tissue) in participants of a lung cancer screening trial that elicits pulmonary function impairment or gas exchange impairment.

MATERIAL AND METHODS

PARTICIPANTS

The NELSON-project is the Dutch-Belgian multi-center lung cancer screening trial, studying current and former heavy smokers 14. The trial was approved by the Dutch ministry of health and by the ethics committee of each participating hospital. Selection of participants was performed by sending a questionnaire about smoking history and other health related questions to people between 50 and 75 years of age, living in the areas around the participating centers. Current and former male smokers meeting the inclusion criteria of having smoked a minimum of 16 cigarettes/day for 25 years or 11 cigarettes/day for 30 years, who gave written informed consent, were equally randomized to either the screening or the control arm. Persons with a self-reported moderate or bad health status, who were unable to climb two flights of stairs were excluded. Persons with current or past renal cancer, melanoma, breast cancer or with lung cancer diagnosed less than 5 years before recruitment were also excluded. From the participants who underwent baseline screening in our hospital, randomly one out of three participants was selected for pulmonary function testing on the same day. For the present study, we included participants who underwent baseline screening between April 2004 and February 2005.

CT SCANNING AND CALCULATION OF EMPHYSEMA SCORES

CT scanning was performed by a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH) with 16 x 0.75mm collimation. A caudo-cranial scan direction was applied and the entire chest was scanned in approximately 10 seconds. No intravenous contrast injection was used. Exposure settings were 30mAs at 120kVp for patients weighing £80 kg and 30mAs at 140kVp for those weighing >80 kg. We reconstructed axial images of 1.0 mm thickness at 0.7 mm increment, using the smallest field of view (FOV) to include the outer rib margins at the widest dimension of the thorax. All scans were reconstructed with a soft kernel (Philips “B”) at 512x512 matrix.

EMPHYSEMA QUANTIFICATION

Extent of low-attenuation areas was determined, using in-house developed software (imageXplorer (iX), Image Sciences Institute, Utrecht, The Netherlands). Segmentation of trachea, left and right lung was performed by a fully automated region growing program starting in the trachea, which included all connected areas below -500HU. In a second step, trachea and main bronchi were separated from the lungs. The algorithm is similar to the one described by Hu and co-workers 15. The number of voxels within the segmented area was multiplied by the size of a voxel to calculate total lung volume. Finally, segmented lungs were subjected to a noise reduction filter [16].

Emphysema scores (ES) were calculated as volume with an attenuation below a fixed attenuation threshold as percentage of total lung volume. For the definitions of moderate and severe emphysema, we used the criteria described by the National Emphysema Treatment Trial (NETT) 17: areas with an attenuation below -950HU represented severe emphysema, areas with an attenuation below -910HU represented moderate emphysema.

Pulmonary Function Tests

Pulmonary function tests (PFT) included forced expiratory volume in one second (FEV1) and vital capacity (VC) with a pneumotachograph followed by assessment of diffusion capacity (Tlco), according to ERS guidelines [18]. Upon arrival, participants rested for 15 minutes after which non-forced spirometry was performed, immediately followed by recording the FEV1 by a flow-volume curve. The best of three temptations was selected for analysis. No reversibility testing was applied.

Diffusing capacity measurements were performed after spirometry. The inhalation mixture contained 0.3% CO and 10% He with balanced air. A breath-holding period of 10 seconds was used. Participants were asked to refrain from smoking, but the Tlco was not corrected for Hb, because in a normal population such correction is not useful [19]. For analysis Tlco was corrected for alveolar volume (Tlco/VA). Abnormal pulmonary function parameters were defined as values below the lower limit of normality (LLN), i.e. ≤-1.64 standard deviations below reference values [18]. Participants were staged according to updated GOLD guidelines [20].

Statistics

We calculated means, standard deviations and 95% confidence intervals (CI) for normal distributed parameters and medians and 25%/75% quartiles for non-normal distributed ones. Spearman's correlation coefficients were used to assess a relationship between lung function parameters and ES for all participants and for a subgroup of participants fulfilling the criteria of GOLD stage II and more. Kruskal-Wallis tests were performed to detect differences between GOLD-stages in both pack years smoked (one pack of cigarettes a day during one year) and ES.

Using the presence or absence of a lowered PFT as outcome variable, the area under the receiver-operator characteristic (ROC) curve of moderate and severe ES was estimated: this area denotes the probability to correctly diagnose the presence or absence of a lowered PFT. From that ROC analysis an optimal cut-off value for moderate and severe ES can be derived, which is that value showing the combined highest sensitivity and specificity. We also used logistic regression with the presence or absence of a lowered PFT as independent and moderate and severe ES as continuous dependent variable to further chart the relation between ES and the probability of a lowered PFT. This analysis was performed for both

lowered FEV1 and lowered Tlco/VA.
All statistics were calculated with SPSS statistical software package version 13 (SPSS, Chicago, Ill.). P-values <0.05 were considered significant.

RESULTS

Five hundred forty-five participants (50-74y, mean 62y), 185 smokers and 360 former smokers, underwent CT scanning and pulmonary function testing on the same day. Characteristics of the study participants are shown in Table X. None of the participants fulfilled the criteria for GOLD stage IV.

GOLD stage	age (years) (SD)	VC (SD)	FEV1 (SD)	FEV1/ VC (SD)	Tlco (SD)	Tlco/ VA (SD)	Pack years (25th-75th percentile)
All	59.8 (5.5)	105.1 (13.6)	97.3 (17.9)	90.6 (12.1)	83.7 (17.8)	90.0 (24.5)	37.8 (27.3 - 48.3)
0 (At risk) (n=339)	59.3 (5.4)	106.2 (13.1)	104.9 (13.9)	97.8 (6.0)	87.1 (14.6)	94.2 (15.7)	37.8 (27.3 – 48.3)
I (Mild) (n=135)	60.5 (5.4)	112.4 (11.3)	94.3 (9.1)	83.3 (5.9)	80.4 (17.3)	81.3 (16.1)	42.6 (33.3 – 48.3)
II (Moderate) (n=62)	61.2 (6.4)	94.7 (11.5)	69.5 (7.2)	73.1 (8.3)	75.0 (18.6)	83.3 (19.7)	44.8 (33.3 – 58.8)
III (Severe) (n=9)	60.7 (2.4)	90.5 (10.8)	44.1 (2.2)	49.0 (6.8)	47.4 (14.3)	50.8 (13.8)	37.8 (37.8 – 57.0)

Table 1
Descriptive statistics shown as mean values (±SD) according to GOLD stage. All pulmonary function parameters are expressed as percentage of the predicted value.

Correlation between emphysema and PFT parameters were low but significant, as shown in Table 1. The coefficients calculated in subsample of participant fulfilling the criteria for GOLD stage II and III were considerably higher for FEV_1 and Tlco, while for FEV_1/VC ratio and Tlco/V_A ratio only moderate changes were found (Table 2). This indicates that selection bias can and will influence the correlation coefficients: in more severe disease stronger relations will be found. The FEV_1/VC and Tlco/V_A ratio correlated best with the ES scores and so these two parameters were selected for further analysis. Hundred forty-three participants (26.2%) showed a FEV_1/VC ratio below the LLN, 210 participants (38.5%) showed a lowered TL_{co}/V_A.

GOLD stage	VC	FEV_1	FEV_1/VC	TI_{co}	TI_{co}/V_A
All	0.23*	-0.16*	-0.48*	-0.28*	-0.47*
II & III	0.08	-0.35*	-0.44*	-0.48*	-0.49*

Table 2

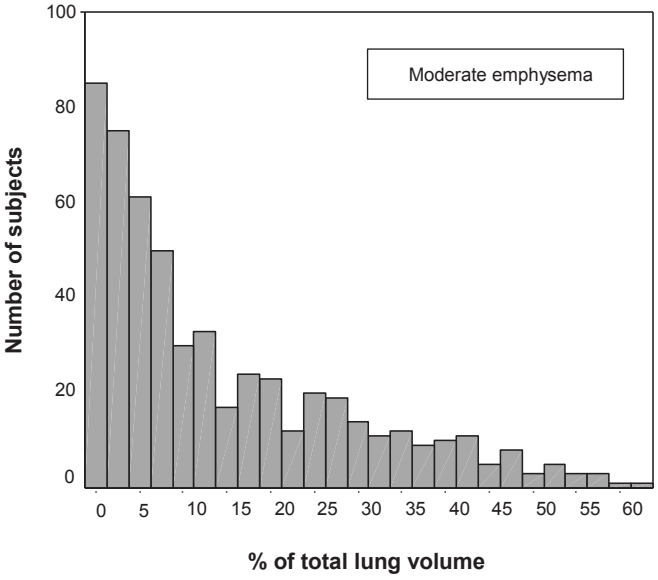
Non-parametric correlation coefficients between emphysema scores and pulmonary function parameters, expressed as percentage of predicted results, calculated for the total study sample and for a subgroup of subjects with GOLD II and III.

	Median ES (25^{th}-75^{th} percentile)	
GOLD stage	**-910HU to -950HU (moderate emphysema)**	**-950HU (severe emphysema)**
All	8.6% (2.8-21.7%)	0.1% (0.04-0.45%)
0 (At risk)	5.8% (1.6-13.5%)	0.08% (0.04-0.19%)
I (Mild)	17.8% (7.3-30.1%)	0.3% (0.11-0.96%)
II (Moderate)	15.0% (4.6-26.4%)	0.4% (0.17-1.5%)
III (Severe)	18.9% (15.5-29.3%)	4.2% (0.5-16.7%)

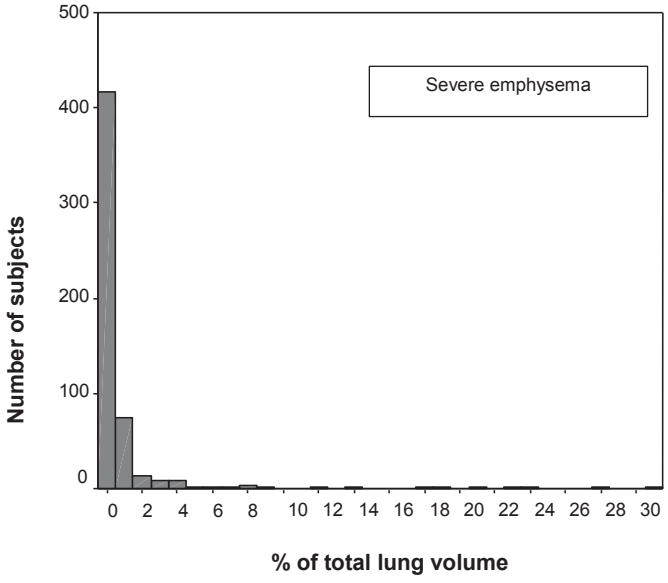
Table 3

Median tissue destruction (percentage of total lung volume) and 25^{th}-75^{th} percentile for mild and severe emphysema, according to GOLD stage.

In Figure 1 the extent of emphysema is illustrated in a frequency plot. The median ES for severe emphysema was 2.7% (inter-quartile range: 1.4% to 6.9%), the median ES for moderate emphysema was 22.9% (inter-quartile range: 15.2% to 31.1%). Median emphysema scores according to GOLD stage are shown in Table 3. Figure 2 shows the scatterplots of the FEV_1/VC and $Tlco/V_A$ values (as percentage of the predicted value) versus the extent of moderate and severe ES.



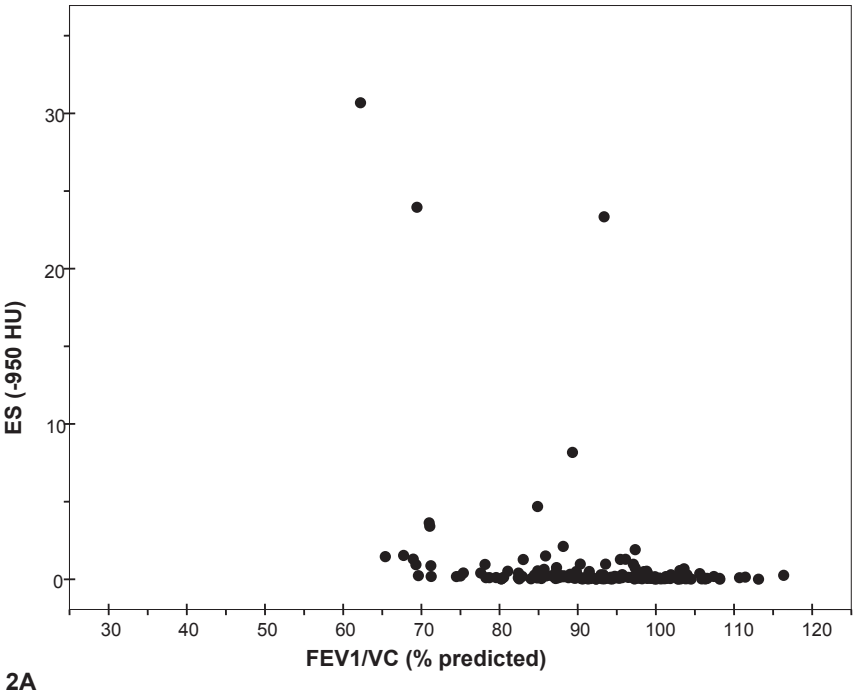
1A



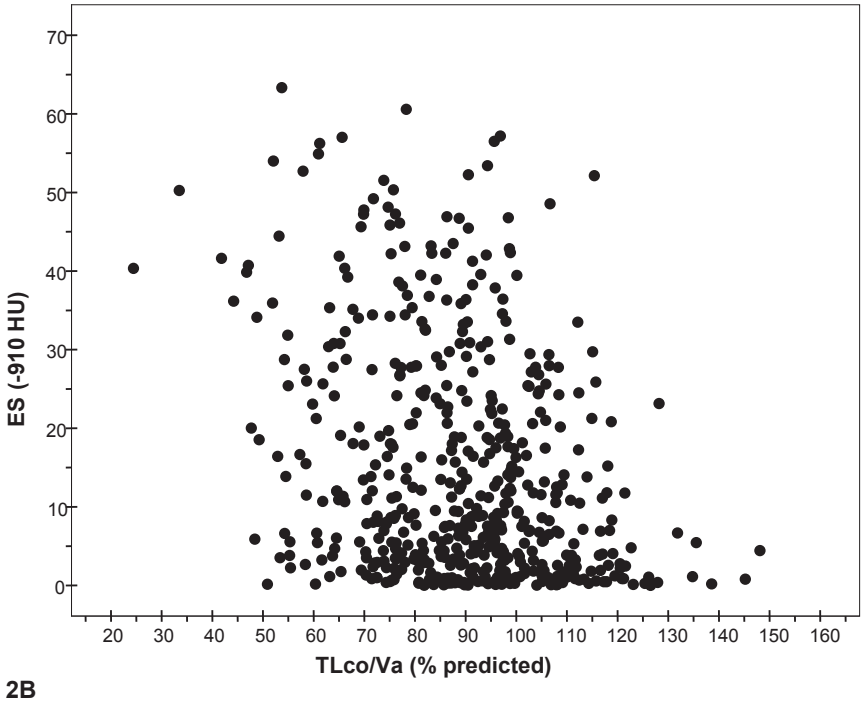
1B

Figure 1

Extent of moderate (A) and severe (B) emphysema plotted against the number of participants


ES (-950 HU)
30
20
10
0
30
40
50
60
70
80
90
100
110
120
FEV1/VC (% predicted)

2A


ES (-910 HU)
70
60
50
40
30
20
10
0
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
TLco/Va (% predicted)

2B

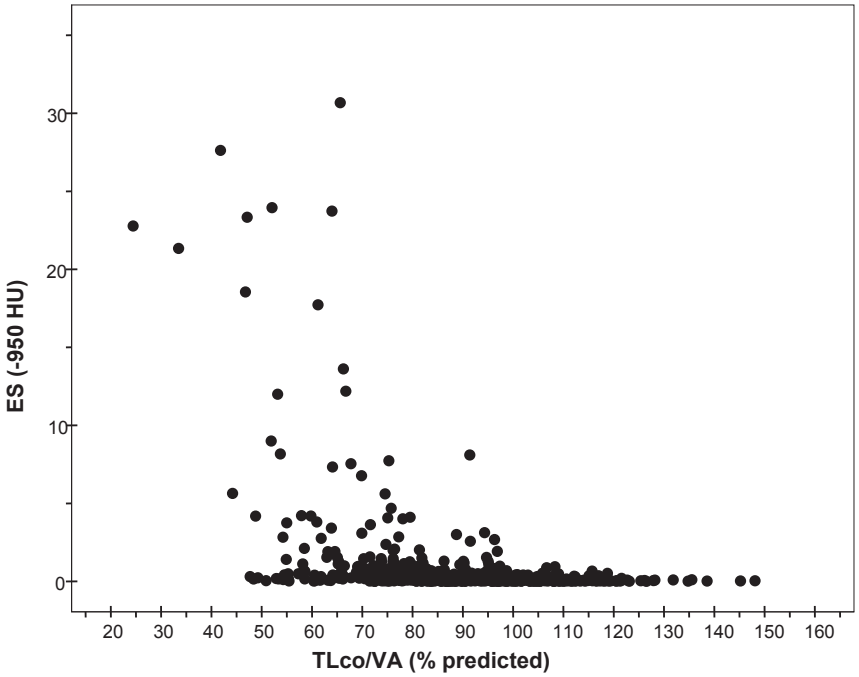
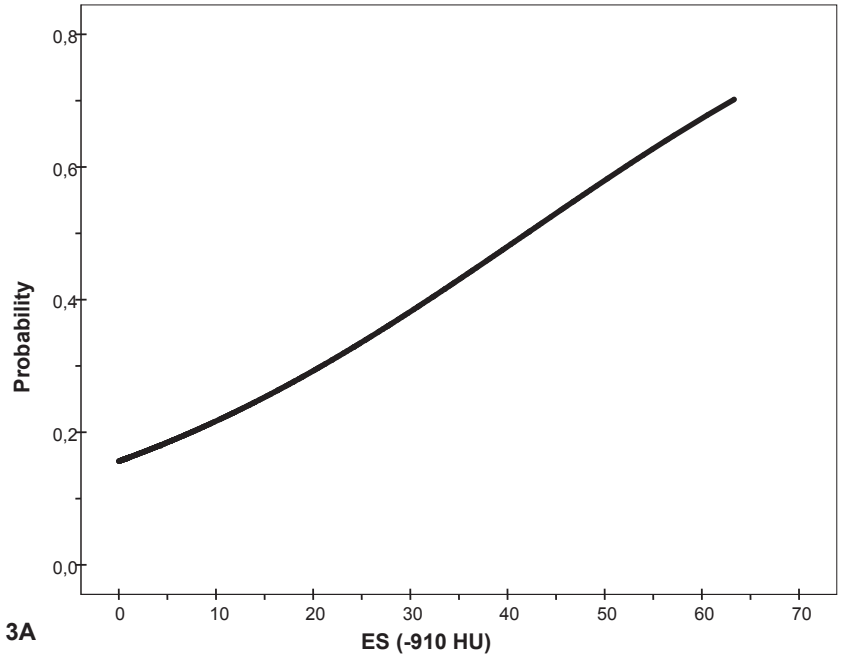


Figure 2

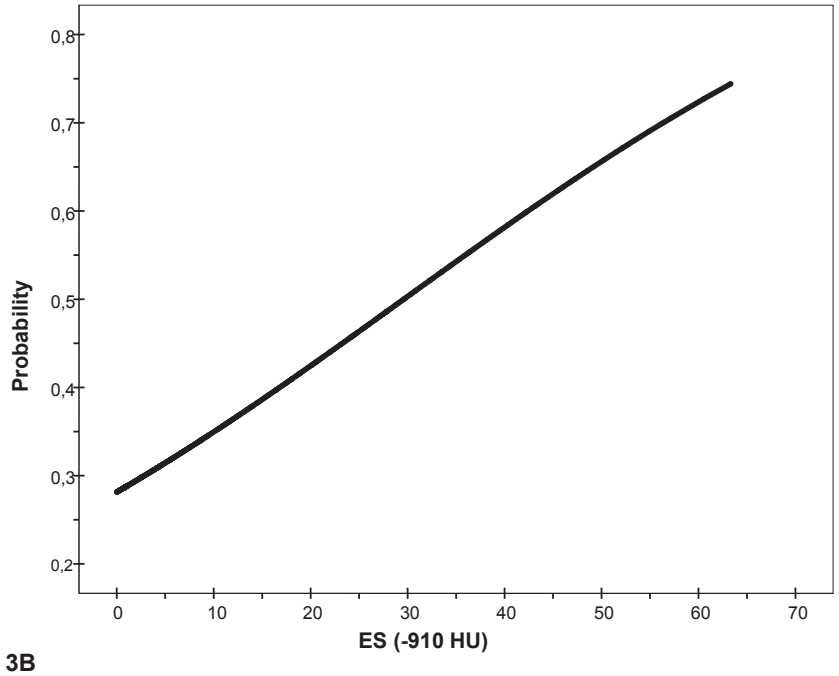
Scatterplots depicting the relation between moderate or severe emphysema scores and FEV_1/VC or TL_{co}/V_A (as percent of the predicted value)

The area under ROC curve for moderate emphysema, predicting the presence or absence of a lowered FEV_1/VC or $Tlco/V_A$ ratio is 0.698 (95% CI 0.650 -0.749) and 0.623 (95% CI 0.575 – 0.672) respectively. Twelve percent of total lung volume with moderate emphysema appears to be the optimal cut-off to result in a lowered FEV_1/VC, while 9% of total lung volume with moderate emphysema appears to be the optimal cut-off to result in a lowered $Tlco/V_A$ ratio. For severe emphysema, the area ROC curve, predicting the presence or absence of a lowered FEV_1/VC and $Tlco/V_A$ ratio is 0.723 (95% CI 0.673 – 0.773) and 0.742 (95% CI 0.698 – 0.786) respectively. For severe emphysema, 0.15% of the total lung volume appears to be the optimal cut-off to result in FEV_1/VC both a lowered FEV_1/VC ratio and a lowered $Tlco/V_A$ ratio.
When we use the GOLD cut-off value of a FEV_1/FVC ratio <0.7, the area under the ROC curve for moderate emphysema, predicting the presence or absence of a FEV_1/FVC <0.70 is 0.732 (95% CI 0.689 -0.775). Moderate emphysema covering ≥11% of the total lung volume appears to be the optimal cut-off in case of both FEV_1/FVC. The area under the ROC curve for severe emphysema, predicting the presence or absence of a FEV_1/FVC <0.70 is 0.765 (95% CI 0.724 - 0.806). Severe emphysema covering 0.15% of total lung volume appears to be the optimal cut-off in case of both FEV_1/FVC.

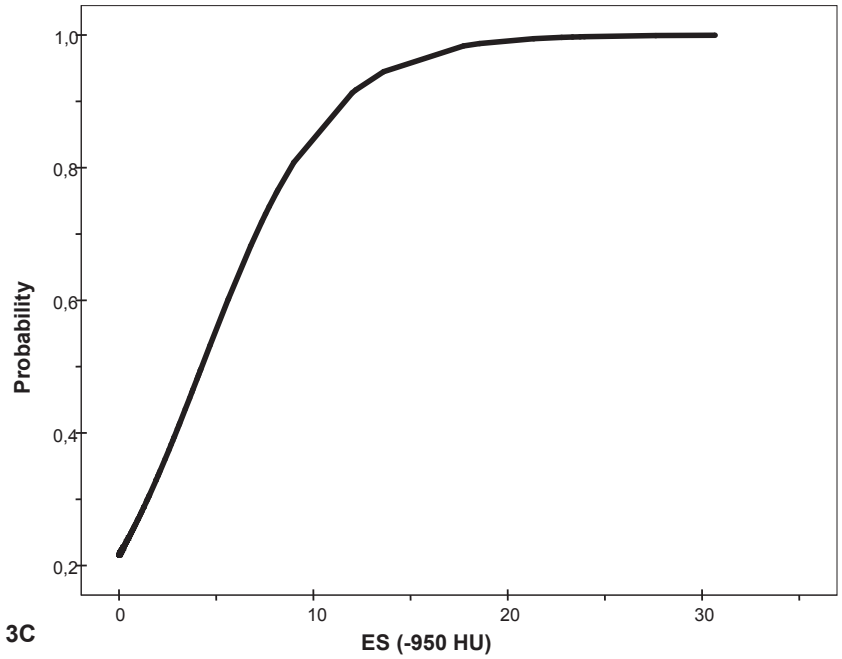
In Figure 3 we demonstrate the relationship between moderate and severe emphysema and the probability of a lowered FEV_1/VC and $Tlco/V_A$ ratio below the lower limit as established by GOLD.


0,8
0,6
0,4
0,2
0,0
Probability
0
10
20
30
40
50
60
70
ES (-910 HU)

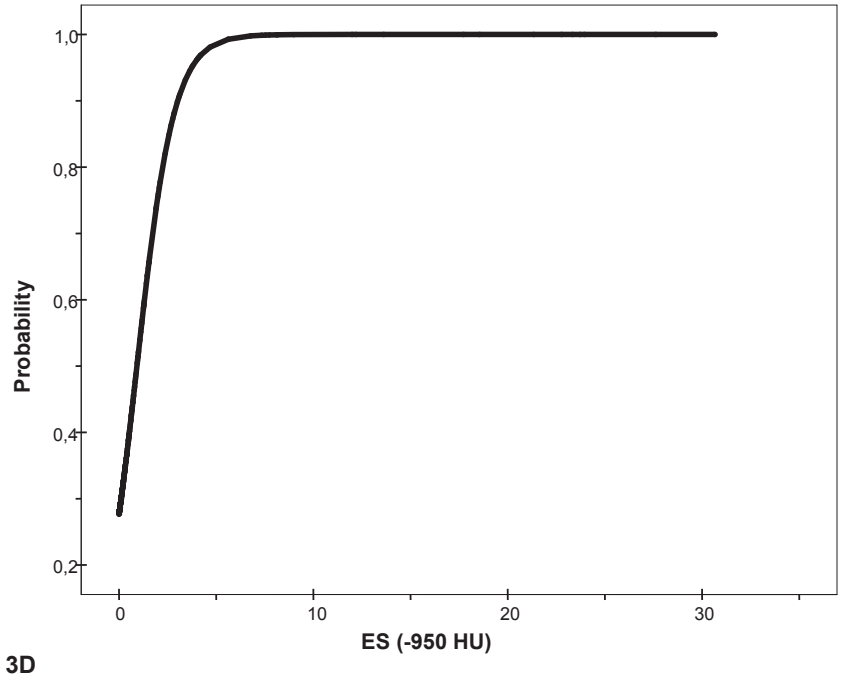
3A


0,8
0,7
0,6
0,5
0,4
0,3
0,2
Probability
0
10
20
30
40
50
60
70
ES (-910 HU)

3B



3C



3D

Figure 3

Graphs relating the probability to detect a FEV_1/VC ratio <LLN to the degree of moderate/severe emphysema (upper left and right) or a TL_{co}/V_A ratio to the same (lower left and right).

DISCUSSION

We demonstrated that the presence of significant amounts of moderate emphysema detected on HRCT does not necessarily result in pulmonary function impairment, while small amounts of severe emphysema already resulted in pulmonary function impairment and gas exchange impairment. Especially for moderate emphysema large areas of destructed lung tissue are required to elicit a high probability of a pulmonary function impairment or gas exchange impairment.

The fact that PFT can still be normal even when CT shows moderate tissue destruction over large areas of the lungs, indicates that pulmonary obstruction is prevented via the non-emphysematous lung parts, which apparently retained their functional characteristics till massive lung destruction is present. In any case, these participants form a substantial subgroup in the spectrum of smoking related lung diseases, which we like to define as 'emphysema with normal pulmonary function' or 'emphysema without obstruction'. Longitudinal studies are required to answer the question whether these participants with smoking related emphysema will also develop obstructive disease or not. If so, screening high-risk participants with low-dose HRCT can be very useful to detect lung destruction before it progresses to a stage with pulmonary function impairment.

The finding that pulmonary function tests are frequently abnormal when tissue destruction is absent or minimal, points at a significant role of apparently ill-functioning, but 'non-destroyed' lung parenchyma, which can not be detected by highlighting low attenuated areas. In line with above, we can define these participants as 'abnormal pulmonary function without emphysema' or as 'obstruction without emphysema'. The Tlco/VA can for example be jeopardized through pathology present at the level of the pulmonary vascular bed [21]. When this damage now precedes gross alveolar destruction, a dissociation between CT and pulmonary function findings is to be expected: small airway disease and respiratory bronchiolitis can lower Tlco/VA ratios in a smoking population [22]. These abnormalities have been demonstrated also to be present in asymptomatic smokers [23;24]. FEV1 and (F)VC are sensitive to mucosal thickening or loss of elasticity resulting from airway inflammation and remodeling. These phenomena can elicit airway obstruction without lung destruction, while not resulting in low-attenuated areas and therefore not being detected by that technique. However, techniques measuring wall thickness of bronchi and bronchioli could provide more insight in this mechanism. Orlandi et al showed significant correlation between both air wall thickness with FEV_1/VC ratio and with D_{CO} in patients with a previous diagnosis of COPD [25] with and without chronic bronchitis (CB). Since patients with COPD and without CB showed a significant higher extent of emphysema, the mechanisms resulting in COPD were supposed to be different. In subjects with CB, the airflow limitation was due to intrinsic bronchial changes resulting in thicker bronchial walls, while in patients without CB the extrinsic

changes such as loss of elastic recoil due to lung tissue destruction result in airflow limitation.
The GOLD-criteria to diagnose COPD are based on an absolute value, not corrected for age, height and sex (FEV_1/FVC <0.7) while others start their diagnostic scheme from the notion that a FEV_1/VC should be lower than the lower than the lower limit of 90% confidence interval for normal results. The data from this study show that both strategies are similar in their relation to moderate or severe emphysema in smokers. More important than choosing between the two is the fact that both approaches are not well suited to detect all aspects of COPD.
This problem might be the result of for example not optimally defined FEV_1/FVC cut off values. Lowering the ratio to values further below 0.70 is rather contra productive because more participants will be depicted as 'healthy', which is not realistic. Moreover we already showed that the correlations between pulmonary function and emphysema will improve with increasing severity of disease: the outcome in terms of ROC areas is a predictable increase. So, for the discussion, we investigated the effects of increasing the FEV_1/FVC cut off values to either 0.75 or 0.80, defining less severe obstruction as already diseased. The area under the ROC curve for moderate and severe emphysema using <0.7 as cut-off value was respectively was 0.732 and 0.765 and increasing the cut off to either 0.75 made the areas decrease by respectively 0.021 and 0.018. Using a FEV-1/FVC cut off values of 0.8 has moderate effects: an increase of moderate emphysema ROC area of 0.002 and a decrease by 0.008 for severe emphysema . These changes are not warranting change of cut off values.
For the diagnosis of emphysema, actually histology is required [3]. No histology was available in the present study, but CT has shown to be able to detect lung tissue destruction on two dimensional images, based on a good correlation with histology, rendering CT-scanning a reliable surrogate marker for pathology [9;10;26-29]. The density mask technique has been reported to be also reliable on three dimensional CTs [4;5]. The main disadvantages of CT are the costs and radiation burden [30;31], but introduction of low-dose protocols as used in our study has reduced the radiation risk substantially [31;32]. The increase of image noise on low-dose scans can influence results of the density mask as shown by Schilham et a [33], but they also showed that emphysema scores performed on low-dose scans filtered with a noise reduction filter revealed results that were similar to ES performed on standard-dose scans realized in the same session. Therefore, our scans were subjected to a noise reduction filter before emphysema scores were calculated.

We here examined the relationship between emphysema and lowered pulmonary function, which is possible because there is consensus on the definition of a impaired pulmonary function. The presence or absence of a lowered function hence easily can act as gold standard. It might be interesting to examine the capability of pulmonary function testing to detect emphysema by reversing the

gold standard, i.e. use the presence of absence of emphysema, too. That definition is pivotal, but unfortunately no consensus exist of the threshold for pathological amounts of emphysema. Ageing in non-smokers could elicit already small amounts of emphysema-like alterations and these have to be separated from smoking induced emphysema. Up to now only small samples of non-smokers were scanned and from those studies no clear cut threshold values could be defined. For the detection of severe emphysema, Kinsella et al proposed a threshold of >1% [34]. The obvious approach is to scan the lungs from a large healthy population, but the question arises whether the radiation risk of CT makes it ethical to obtain these data.

In conclusion, we demonstrated the probability of especially moderate emphysema to result in a pulmonary function impairment or gas exchange impairment is low, while small amounts of severe emphysema already resulted in pulmonary function impairment and gas exchange impairment. Moreover, many current and former smokers suffer from moderate emphysema without clinical diagnosis of COPD.

REFERENCE LIST

(1) Feenstra TL, van Genugten ML, Hoogenveen RT et al. The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands. Am J Respir Crit Care Med 2001; 164(4):590-596.

(2) Pauwels RA, Buist AS, Calverley PMA et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease . NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. Am J Respir Crit Care Med 2001; 163(5):1256-1276.

(3) The definition of emphysema. Report of a National Heart, Lung, and Blood Institute, Division of Lung Diseases workshop. Am Rev Respir Dis 1985; 132(1):182-185.

(4) Arakawa A, Yamashita Y, Nakayama Y et al. Assessment of lung volumes in pulmonary emphysema using multidetector helical CT: comparison with pulmonary function tests. Comput Med Imaging Graph 2001; 25(5):399-404.

(5) Park KJ, Bergin CJ, Clausen JL. Quantitation of emphysema with three-dimensional CT densitometry: comparison with two-dimensional analysis, visual emphysema scores, and pulmonary function test results. Radiology 1999; 211(2):541-547.

(6) Calverley PMA. COPD : Early Detection and Intervention. Chest 2000; 117(90052):365S-371.

(7) Clark KD, Wardrobe-Wong N, Elliott JJ et al. Patterns of Lung Disease in a "Normal" Smoking Population : Are Emphysema and Airflow Obstruction Found Together? Chest 2001; 120(3):743-747.

(8) Bankier AA, Madani A, Gevenois PA. CT quantification of pulmonary emphysema: assessment of lung structure and function. Crit Rev Comput Tomogr 2002; 43(6):399-417.

(9) Coxson HO, Rogers RM, Whittall KP et al. A quantification of the lung surface area in emphysema using computed tomography. Am J Respir Crit Care Med 1999; 159(3):851-856.

(10) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.

(11) Uppaluri R, Mitsa T, Sonka M et al. Quantification of pulmonary emphysema from lung computed tomography images. Am J Respir Crit Care Med 1997; 156(1):248-254.

(12) Muller NL, Staples CA, Miller RR et al. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988; 94(4):782-787.

(13) Gevenois PA, De Vuyst P, De M, V et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1996; 154(1):187-192.

(14) Xu DM, Gietema H, de Koning H et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 2006; 54(2):177-184.

(15) Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001; 20(6):490-498.

(16) Hilts M, Duzenli C. Image filtering for improved dose resolution in CT polymer gel dosimetry. Med Phys 2004; 31(1):39-49.

(17) Naunheim KS, Wood DE, Krasna MJ et al. Predictors of operative mortality and cardiopulmonary morbidity in the National Emphysema Treatment Trial. J Thorac Cardiovasc Surg 2006; 131(1):43-53.

(18) Quanjer PH, Tammeling GJ, Cotes JE et al. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J Suppl 1993; 16:5-40.

(19) Stam H, Hrachovina V, Stijnen T et al. Diffusing capacity dependent on lung volume and age in normal subjects. J Appl Physiol 1994; 76(6):2356-2363.

(20) www.goldcopd.com 2005.

(21) Voelkel NF, Cool CD. Pulmonary vascular involvement in chronic obstructive pulmonary disease. Eur Respir J 2003; 22(46_suppl):28S-32.

(22) Heyneman LE, Ward S, Lynch DA et al. Respiratory bronchiolitis, respiratory bronchiolitis-associated interstitial lung disease, and desquamative interstitial pneumonia: different entities or part of the spectrum of the same disease process? AJR Am J Roentgenol 1999; 173(6):1617-1622.
(23) Remy-Jardin M, Remy J, Boulenguez C et al. Morphologic effects of cigarette smoking on airways and pulmonary parenchyma in healthy adult volunteers: CT evaluation and correlation with pulmonary function tests. Radiology 1993; 186(1):107-115.
(24) Remy-Jardin M, Edme JL, Boulenguez C et al. Longitudinal Follow-up Study of Smoker's Lung with Thin-Section CT in Correlation with Pulmonary Function Tests. Radiology 2002; 222(1):261-270.
(25) Orlandi I, Moroni C, Camiciottoli G et al. Chronic Obstructive Pulmonary Disease: Thin-Section CT Measurement of Airway Wall Thickness and Lung Attenuation. Radiology 2005; 234(2):604-610.
(26) Madani A, Keyzer C, Gevenois PA. Quantitative computed tomography assessment of lung structure and function in pulmonary emphysema. Eur Respir J 2001; 18(4):720-730.
(27) Bergin C, Muller N, Nichols DM et al. The diagnosis of emphysema. A computed tomographic-pathologic correlation. Am Rev Respir Dis 1986; 133(4):541-546.
(28) Newell JD, Jr., Hogg JC, Snider GL. Report of a workshop: quantitative computed tomography scanning in longitudinal studies of emphysema. Eur Respir J 2004; 23 (5):769-775.
(29) Coddington R, Mera SL, Goddard PR et al. Pathological evaluation of computed tomography images of lungs. J Clin Pathol 1982; 35(5):536-540.
(30) Haaga JR. Radiation dose management: weighing risk versus benefit. AJR Am J Roentgenol 2001; 177(2):289-291.
(31) Brenner DJ, Doll R, Goodhead DT et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc Natl Acad Sci U S A 2003; 100 (24):13761-13766.
(32) Brenner DJ. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. Radiology 2004; 231(2):440-445.
(33) Schilham A, Van Ginneken B, Gietema H et al. Local Noise Weighted Filtering for Emphysema Scoring of Low-dose CT Images. IEEE Transactions on Medical Imaging 2006; 25(4):451-463.
(34) Kinsella M, Muller NL, Abboud RT et al. Quantitation of emphysema by computed tomography using a "density mask" program and correlation with pulmonary function tests. Chest 1990; 97(2):315-321.

Chapter 9

The nitric oxide transfer factor as a tool for the early diagnosis of emphysema

Ivo van der Lee
Hester Gietema
Pieter Zanen
Mathias Prokop
Jan-Willem Lammers
Jules van den Bosch

Submitted

ABSTRACT

The transfer factor for nitric oxide (Tlno) is independent of pulmonary capillary blood volume due to the very high affinity of nitric oxide (NO) to hemoglobin, in contrast to the transfer factor for carbon monoxide (Tlco). Therefore the sensitivity for the detection of alveolar destruction is supposed to be higher.

We measured flow volumes curves, Tlno, Tlco, the transfer coefficients Kno and Kco and performed high-resolution computed tomography (HRCT) in 263 randomly selected heavy smokers. Participants with areas ≥1% of the total lung volume showing an attenuation <-950 Hounsfield Units were considered to have emphysema. In 36 participants emphysema was diagnosed with HRCT, an abnormal Kno was present in 94 participants, and in 95 participants a FEV_1/FVC ratio <70% was seen. The area under the ROC curve for the Kno was 0.894 and for the Kco 0.822. The Kno therefore showed a slightly higher sensitivity to detect emphysema, compared to the Kco. The positive predictive value of K_{NO} however was low (34.7%), while the negative predictive value of Kno was very high (98.2%), indicating an exclusion test. The Tlno/Tlco ratio is significantly higher compared to normal participants, indicating a strong influence of a decreased microvascularity.

INTRODUCTION

The prevalence of chronic obstructive pulmonary disease (COPD) is expected to increase in the next decades, leading to a decreased quality of life in older participants, as well as an increased financial burden to society [1]. Estimations of the percentage of smokers developing COPD vary from 0.3 to 8.5%, depending on diagnostic criteria [2-4].

Emphysema is a component of COPD and is defined as an abnormal enlargement of the terminal bronchioles and alveoli [5]. Gevenois et al [6] demonstrated that the amount of abnormally low attenuation areas on high resolution computed tomography (HRCT) scans correlated well with the amount of emphysema present in lung specimens. Several investigators showed that the transfer factor (Tlco) and transfer coefficient (Kco) for carbon monoxide (CO) showed a stronger correlation with HRCT emphysema indices than the FEV_1 and FEV_1/FVC ratio [7-9]. HRCT as an instrument for mass screening for COPD has the disadvantage of high costs. The Tlco and Kco are simple and cheap parameters, available in any pulmonary function laboratory, and may be used as a tool for early detection of emphysema.

The Tlco is dependent on the thickness and surface area of the alveolocapillary membrane, its solubility in water and binding to hemoglobin. This has been formulated by Roughton and Forster [10] in 1959: $1/Tlco=1/Dmco + 1/\theta co*Vcap$, where Dmco is the membrane diffusing capacity for CO, θco the CO uptake by erythrocytes and Vcap the pulmonary capillary blood volume. Since 15 years studies have been performed with nitric oxide (NO) instead of CO [11;12]. NO has a much stronger affinity for hemoglobin, so θ_{NO} is very high, leading to a negligible value for $1/\theta_{NO}*Vcap$. Therefore, the transfer factor for NO (Tlno) is supposed to represent the true membrane diffusing capacity [13]. A predominantly vascular disease will lower the Tlco, but not the Tlno, as it is not influenced by erythrocyte NO uptake (=decreased Vcap). The Tlno/Tlco will tend to increase when predominantly vascular disease is present [14]. On the other hand [15], a predominantly membranous disturbance will affect both Tlco and Tlno, and the alleged high sensitivity of the Tlno for membranous disturbances will tend to sharply decrease Tlno. The Tlco, being partly dependent on membranous damage will not change that sharply: the ratio will tend to decrease.

The aim of this study was to determine whether the Tlno is a better screening tool for the detection of emphysema than the Tlco, and if the Tlno/Tlco ratio differs from healthy participants.

METHODS

PARTICIPANTS

All participants of the current study were participating in the NELSON-project, a Dutch-Belgian multi-center lung cancer screening trial. The participants were all male, 50 to 75 years of age with a smoking history of at least 16 cigarettes/day for 25 years or 11 cigarettes/day for 30 years. Exclusion criteria were current or past melanoma, renal breast or lung cancer diagnosed <5 years before recruitment, a chest CT scan <1 year before recruitment, a body weight ≥140 kilogram and quitting smoking >10 years before start of the trial.
This study was performed in the University Medical Center Utrecht, one of the participating hospitals in the NELSON study. The NELSON-project was approved by the Dutch ministry of health and by the ethics committee of the participating hospital; informed consent was obtained from all participants. From three thousand participants who underwent the screening in our hospital, randomly one out of three screened participants was selected for pulmonary function testing on the same day.

PULMONARY FUNCTION TESTING

Spirometry and flow-volume curves measurements were obtained via pneumotachography, according to ERS guidelines [16]. No reversibility testing was done. In line with the GOLD classification [17], a FEV1/FVC ratio <70% was labeled as abnormal. A simultaneous single breath Tlno and Tlco test was performed directly after spirometry, as described earlier [18]. At least two measurements per participant were obtained and a difference <10% in Tlco was acceptable. All pulmonary function values are given as a mean with standard deviation (SD), and as a percentage of predicted values [16;19]. For Tlno we used references equations from an earlier study. Values <-1.64 times the standard deviation were considered abnormal. The Tlno/Tlco and FEV1/FVC ratio's are expressed as absolute values.

CT SCANNING AND EMPHYSEMA QUANTIFICATION

CTs were performed on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH), scan time was 12 seconds, in spiral mode with 16 x 0.75mm collimation, 1.0 mm reconstruction thickness, without contrast-injection. After automatic lung segmentation connecting all areas below -500 Hounsfield units (HU) starting in the trachea and excluding the main bronchi, quantification of emphysema was done highlighting low attenuated areas [20] using in-house developed software (ImageXplorer, Image Sciences Institute, Utrecht, The Netherlands), with a threshold of -950 HU. Emphysema scores (ES) were calculated as the volume with an attenuation <-950 HU indexed to total lung volume. Participants with an ES≥1% were considered to suffer from emphysema [7;21;22].

STATISTICS

Parameters are depicted as means with standard deviations (SD). Spearman correlation coefficients were used to assess significant relationships; the area under the curve (AUC) of the receiver operator curves (ROC) were used to assess the capability of the pulmonary function parameters to signal the absence or the presence of emphysema. One-way ANOVA was used to compare differences between groups. All statistics were calculated with SPSS for Windows release 13 (SPSS Inc, Chicago, Ill.).

RESULTS

Between October 2004 and April 2005 263 male participants were included in the study. Several characteristics of the study population are given in Table 1. Thirty-six (13.6%) participants were shown to suffer from emphysema (ES≥1).

Significant negative correlations were observed between all diffusion capacity parameters, the FEV_1/FVC ratio and the ES (Table 2). The mean Tlno/Tlco ratio was 4.9, which is higher than the 4.3 value reported by Borland et al [11] in healthy participants and also significantly higher ($p<0.001$) than the 4.6 value in 65 males or the 4.4 value in men and women pooled, derived from the previous data.

The AUC of the ROC curves showed the highest values for Kno (as percentage of predicted) to detect emphysema, with Kco (also as percentage predicted) and FEV_1/FVC in second respectively in third highest position (Table 3). The differences between the AUC ROC for the Kno and Kco were small and clinically irrelevant.

The sensitivity and specificity of all parameters to detect an ES≥1% are given in Table 4, as well as the positive (PPV) and negative (NPV) predictive values. The NPV is especially high indicating that a normal Kco, Kno or FEV1/FVC virtually excludes an ES≥1%. The low PPV values indicate that an abnormal Kco, Kno or FEV1/FVC only points at an ES≥1% in a minority of the cases.

Participants with low Kno values, participants with emphysema on HRCT and participants with a FEV1/FVC ratio lower than 70% form partially overlapping groups, which can be illustrated with a Venn diagram (Figure 1).

	Mean	Range	SD
Age, years	60.3	52.3-76.9	5.4
Height, m	1.78	1.61-2.00	0.07
VC, %pred	105.5	61.1-147.5	13.2
FEV_1/FVC ratio, %pred	93.6	42.3-113.9	11.7
FEV_1, %pred	97.7	43.0-140.8	16.8
FEV_1/FVC ratio	71.5	32.4-86.9	9.0
Tlco, %pred	87.4	49.8-140.4	16.1
K_{CO}, %pred	84.4	46.6-140.2	15.9
Tlno, %pred	87.5	45.3-121.3	13.5
K_{NO}, %pred	90.4	53.7-121.6	12.4
Tlno/Tlco ratio	4.9	3.8-6.4	0.4
emphysema score	0.6	0.0-14.7	1.5

Table 1

Characteristics of the study population, n = 263

	Emphysema score
FEV_1/FVC ratio	-0.43*
Tlco, %predicted	-0.26*
K_{CO}, %predicted	-0.38*
Tlno, %predicted	-0.29*
K_{NO}, %predicted	-0.50*

* = p< 0.01

Table 2

Correlation coefficients of the emphysema score versus pulmonary function data

	AUC	p=	95%CI
FEV_1, %pred	0.656	<0.003	0.551-0.761
FEV_1/FVC, %pred	0.795	<0.001	0.710-0.880
Tlco, %pred	0.727	<0.001	0.622-0.833
K_{CO}, %pred	0.822	<0.001	0.757-0.887
Tlno, %pred	0.711	<0.001	0.608-0.815
K_{NO}, %pred	0.894	<0.001	0.850-0.938

Table 3

Area under the curve (AUC) with 95% confidence intervals (95% CI) of the receiver operator curves (ROC).

	Sensitivity	Specificity	PPV	NPV
TIco	58.3%	81.5%	33.3%	92.5%
TIno	50.0%	81.9%	30.5%	91.2%
Kco	88.9%	57.3%	24.8%	97.0%
Kno	91.7%	72.7%	34.7%	98.2%
FEV_1/FVC	77.8%	70.1%	29.5%	95.2%

Table 4

Sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV) of the measured parameters to detect emphysema.

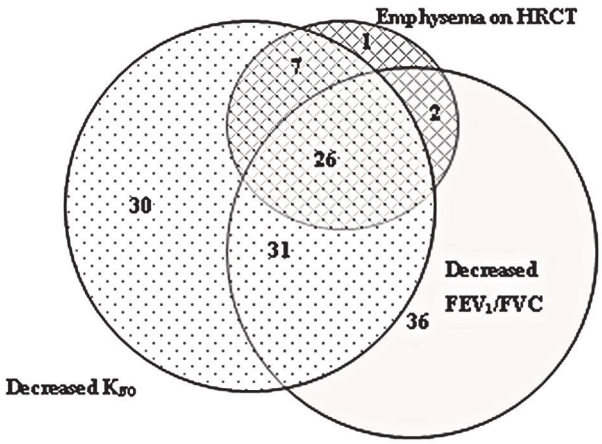


Figure 1

Venn diagram: the presence of emphysema on HRCT (n=36), a decreased Kno (n=94) and a decreased FEV_1/FVC ratio (n=95) are partially overlapping entities.

DISCUSSION

We measured the single breath transfer factor and transfer coefficient for CO and NO in a large sample of current and former heavy smokers. The hypothesis was that the Tlno would better detect emphysema on HRCT than the Tlco, based on its alleged higher sensitivity for alveolar membrane destruction. However, this proved not to be the case.

The Tlno/Tlco ratio in these smokers is significantly higher than in healthy participants, and reaches values seen in participants with pulmonary arterial hypertension [14;15], suggesting that in this group of smokers vascular damage is an important phenomenon and strongly influences the diffusion capacity. In the data from Borland et al a Tlno/Tlco ratio of 5.0 in participants with PAH versus 4.5 in healthy participants was reported [23].

The study outcome indicates that lung function is not very sensitive to detect emphysema, even if the parameter used is maximally influenced by a loss of alveolar membranes, like the Tlno and the Kno. Apparently the early presence or a low degree of emphysema does not influence the lung function significantly and other processes, not directly linked to emphysema, are more important for the changes in lung function. We conclude that, at least in the beginning, the status of the non-emphysematous parts of the lung influences lung function parameters more than the emphysematous parts. One must acknowledge that the lung function is influenced by the status of the entire lungs and that emphysematous parts are only subdivisions of that total, so small amounts of emphysema will easy go undetected. The CT-scan can however detect such small amounts of emphysema more easily.

The reason why the Tlno and Kno are only marginally better in detecting emphysema compared to the Tlco and KCO is probably due to the fact that in our study population the majority of the tested participants showed either no or only small amounts of emphysema. These small amounts of loss of alveolar tissue and the higher sensitivity of the Tlno and Kno is apparently not sufficient to overcome the problem of the low emphysema expression in lung function. When the non-emphysematous parts of the lungs dictate the decrease in function strongly the Tlno and Kno will lose their theoretical advantage.

As a result the correlation between CT and lung function apparently can not be strong. In older reports discrepancies between the pulmonary function testing and the presence of CT-based emphysema have been reported in participants with COPD [20] and in candidates for lung resection or transplantation [7]. This confirms our conclusion.

In this population based study in heavy (former) smokers the correlation between pulmonary function tests and CT based assessment of emphysema is even weaker than reported earlier [20]. This is not unexpected, because our participants were not selected on the presence of COPD via lung function testing. In studies in which such work-up bias was present, one might expect stronger correlations, due the

presence of more severe disease. As a consequence, the correlation between for example the FEV_1/FVC ratio and the emphysema score is low: 8 out of 36 participants with emphysema on HRCT had a FEV_1/fvc ratio above 70%.
Similar arguments can be used to explain why the spirometric parameters (FEV1 and FEV_1/FVC) also show rather low AUC's of the ROC curve values. As with the diffusion parameters: they are less useful to diagnose emphysema. Again we must conclude that the presence or the degree of emphysema does not influence spirometry significantly.
As mentioned before in the introduction, a high Tlno/Tlco ratio points at a significant influence of diffusion parameters on vascular damage. The Tlno is not influenced to the same extent by vascular damage as the Tlco: the ratio therefore will increase if such damage is significant. Now we calculated this ratio for the entire sample, and most participants had no emphysema detectable on CT. The reduction of the diffusion capacity parameters therefore seems to be caused to a large extent by damage of the alveolar vascular compartment by smoking. This functional damage may be renamed as vascular malfunctioning. The fact that the high Tlno/Tlco ratio also characterizes the non-emphysematous participants points at the fact that functional alveolar vascular damage precedes the loss of alveolar tissue measurable via CT-scanning. This is not an illogical approach: overt alveolar tissue loss will be small in early disease. It is conceivable that alveolar vessel function already is impaired in such an early stage of the disease, followed by the more overt loss of tissue later. The loss of alveolar vascular function will of course coincide with a loss of recoil because it is hard to conceive that only the vessels in the alveolar membrane will suffer from smoking.
Spirometry, gas transfer and CT-scans are complementary in the detection of emphysema and partially overlap (Figure 1). In clinical practice this can be used in excluding emphysema: only when combining all measures a complete description of the pathophysiology of the lung will be obtained.
The commonly used GOLD criteria [17] are mainly based on the presence of airflow limitation. The FEV_1 has been chosen as the major determinant because abundant data are available to correlate the FEV_1 with symptoms, prognosis and mortality. Such data are not available for HRCT-detected emphysema, and are scarce for diffusion parameters. Emphysema, however, is not always accompanied by significant airflow limitation. Therefore, one must realize that the COPD is more than just airflow limitation. In the GOLD criteria stage 0 is defined as "at risk": this compasses a FEV_1/FVC ratio >70% combined with complaints of coughing and increased phlegm production. By strictly using the FEV_1/FVC cut off point of 70% and ignoring the transfer factor (or tissue destruction on HRCT), selection bias is introduced.

LIMITATIONS

Although it was recommended that the participants refrained from smoking 24 hours prior to testing, the question is always open whether this advice was

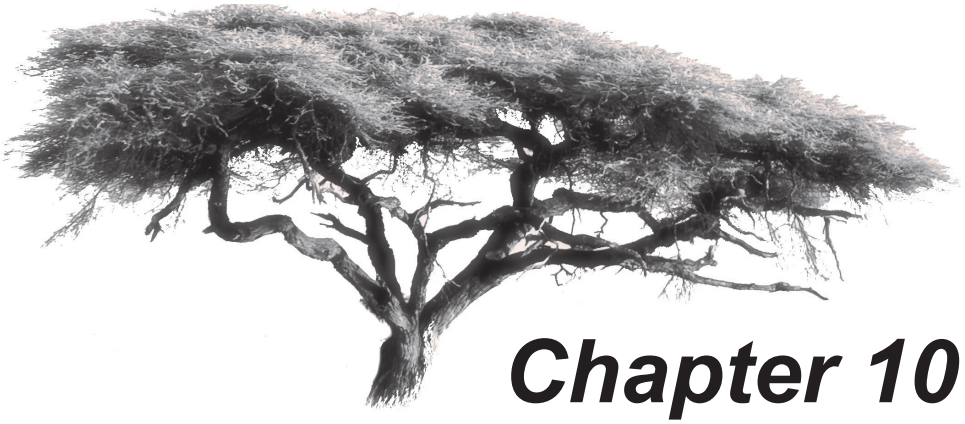
followed. This could have led to a small decrease in the Tlco based on higher CO-backpressure. In our believe this could not alter the main results, because the smoking effect is very small: the Tlco decreases by 1% for each percent COHb present. So when the presence of COHb is 7%, the Tlco decrease by approximately 7% [25]. Such high levels of COHb are seldom.
The Tlno references equations are based on a previous study in 124 participants, of which were 65 males with a mean age of 40.1 years. The mean age in this study was higher. However, the Tlco values in that previous study match exactly with the ECCS vales 19, so gross deviations in estimating age, and height effects are unlikely. The extrapolation to older participants is therefore possible.
The cut off value for the amount of emphysema on HRCT scan suffers from a lack of consensus: we defined a cut-off point of 1% at -950 HU based 1 mm slice techniques, based on the study of Kinsella et al [22].

Spirometry is very important for the diagnosis and classification of emphysema, but only measures airway obstruction. The Kno is a sensitive measure for the detection of emphysema on HRCT. Pathophysiological changes in emphysema also compass microvascular changes and parenchymal loss. It is not unlikely that microvascular disease precedes extensive parenchymal loss eventually leading to airway obstruction. Longitudinal studies are needed to further explore this concept.

REFERENCE LIST

(1) Chapman KR, Mannino DM, Soriano JB et al. Epidemiology and costs of chronic obstructive pulmonary disease. Eur Respir J 2006; 27(1):188-207.
(2) Fletcher C, Peto R. The natural history of chronic airflow obstruction. Br Med J 1977; 1 (6077):1645-1648.
(3) Lundback B, Lindberg A, Lindstrom M et al. Not 15 but 50% of smokers develop COPD? --Report from the Obstructive Lung Disease in Northern Sweden Studies. Respir Med 2003; 97(2):115-122.
(4) Zielinski J, Bednarek M. Early detection of COPD in a high-risk population using spirometric screening. Chest 2001; 119(3):731-736.
(5) Mitchell RS, Stanford RE, Johnson JM et al. The morphologic features of the bronchi, bronchioles, and alveoli in chronic airway obstruction: a clinicopathologic study. Am Rev Respir Dis 1976; 114(1):137-145.
(6) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.
(7) Gevenois PA, De Vuyst P, De M, V et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1996; 154(1):187-192.
(8) Klein JS, Gamsu G, Webb WR et al. High-resolution CT diagnosis of emphysema in symptomatic patients with normal chest radiographs and isolated low diffusing capacity. Radiology 1992; 182(3):817-821.
(9) Tylen U, Boijsen M, Ekberg-Jansson A et al. Emphysematous lesions and lung function in healthy smokers 60 years of age. Respir Med 2000; 94(1):38-43.
(10) Roughton FJ, Forster RE. Relative importance of diffusion and chemical reaction rates in determining rate of exchange of gases in the human lung, with special reference to true diffusing capacity of pulmonary membrane and volume of blood in the lung capillaries. J Appl Physiol 1957; 11(2):290-302.
(11) Borland CD, Higenbottam TW. A simultaneous single breath measurement of pulmonary diffusing capacity with nitric oxide and carbon monoxide. Eur Respir J 1989; 2(1):56-63.
(12) Guenard H, Varene N, Vaida P. Determination of lung capillary blood volume and membrane diffusing capacity in man by the measurements of NO and CO transfer. Respir Physiol 1987; 70(1):113-120.
(13) Tamhane RM, Johnson RL, Jr., Hsia CC. Pulmonary membrane diffusing capacity and capillary blood volume measured during exercise from nitric oxide uptake. Chest 2001; 120(6):1850-1856.
(14) Harris RS, Hadian M, Hess DR et al. Pulmonary artery occlusion increases the ratio of diffusing capacity for nitric oxide to carbon monoxide in prone sheep. Chest 2004; 126 (2):559-565.
(15) van der L,ee I, Zanen P, Grutters JC et al. Diffusing capacity for nitric oxide and carbon monoxide in patients with diffuse parenchymal lung disease and pulmonary arterial hypertension. Chest 2006; 129(2):378-383.
(16) Quanjer PH, Tammeling GJ, Cotes JE et al. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J Suppl 1993; 16:5-40.
(17) Pauwels RA, Buist AS, Calverley PMA et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease . NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. Am J Respir Crit Care Med 2001; 163(5):1256-1276.
(18) van der Lee, I, Zanen P, Biesma DH et al. The effect of red cell transfusion on nitric oxide diffusing capacity. Respiration 2005; 72(5):512-516.

(19) Cotes JE, Chinn DJ, Quanjer PH et al. [Standardization of the measurement of transfer factor (diffusing capacity). Work Group on Standardization of Respiratory Function Tests. European Community for Coal and Steel. Official position of the European Respiratory Society]. Rev Mal Respir 1994; 11 Suppl 3:41-52.
(20) Muller NL, Staples CA, Miller RR et al. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988; 94(4):782-787.
(21) Gevenois PA, Scillia P, De M, V et al. The effects of age, sex, lung size, and hyperinflation on CT lung densitometry. AJR Am J Roentgenol 1996; 167(5):1169-1173.
(22) Kinsella M, Muller NL, Abboud RT et al. Quantitation of emphysema by computed tomography using a "density mask" program and correlation with pulmonary function tests. Chest 1990; 97(2):315-321.
(23) Borland C, Cox Y, Higenbottam T. Reduction of pulmonary capillary blood volume in patients with severe unexplained pulmonary hypertension. Thorax 1996; 51(8):855-856.
(24) Park KJ, Bergin CJ, Clausen JL. Quantitation of emphysema with three-dimensional CT densitometry: comparison with two-dimensional analysis, visual emphysema scores, and pulmonary function test results. Radiology 1999; 211(2):541-547.
(25) Sansores RH, Pare PD, Abboud RT. Acute effect of cigarette smoking on the carbon monoxide diffusing capacity of the lung. Am Rev Respir Dis 1992; 146(4):951-958.



Chapter 10

Distribution of moderate and severe emphysema in heavy smokers joining a population-based lung cancer screening trial: Impact on pulmonary function

Hester Gietema
Pieter Zanen
Arnold Schilham
Bram van Ginneken
Rob van Klaveren
Mathias Prokop
Jan-Willem Lammers

Submitted

ABSTRACT

PURPOSE

To investigate the impact of the distribution pattern of both moderate and severe emphysema, detected on volumetric high-resolution CT, on the severity of airflow limitation and gas exchange impairment in a large population of current and former heavy smokers participating in a lung cancer screening trial.

MATERIALS AND METHODS

Between April 2004 and February 2005, 1386 male current and former heavy smokers underwent low-dose CT (16 x 0.75mm slice collimation) in our center as part of a population-based lung cancer screening trial. In 545 subjects pulmonary function testing was performed on the same day. Severe emphysema was defined as lung volume with an attenuation <-950 Hounsfield units (HU), quantified objectively, while areas with an attenuation between -910HU and -950HU represented moderate emphysema. Impact of distribution on the severity of pulmonary function impairment was investigated.

RESULTS

Corrected for extent of emphysema, for both moderate and severe emphysema an apical distribution was associated with more airflow obstruction and gas exchange impairment than a basal distribution.

CONCLUSION

In an asymptomatic heavy smoking population, an apical distribution is associated with more severe pulmonary function impairment than a basal distribution.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of morbidity and mortality in the developed countries [1;2], but little is known about its early stages. Since COPD is functionally defined, diagnosis and staging is based on the guidelines provided by the Global initiative on Obstructive Lung Diseases (GOLD) guidelines [3]. These guidelines are based on results of spirometric testing, using forced expiratory volume in one second (FEV_1) and FEV_1/forced vital capacity (FVC) as primary parameters. But COPD is a heterogeneous disease comprising not only mucosal thickening of the bronchioles, resulting in airflow limitation but also emphysema, anatomically defined as a permanent enlargement of the terminal bronchioles and alveoli. Although emphysema can cause airflow obstruction and correlation between pulmonary pathophysiology examined with pulmonary function tests and pathology found with CT has shown moderate to good results [4;5], demonstrating more pulmonary function impairment with increasing amounts of emphysema, emphysema can also exist without impairment of the FEV_1 [6]. Because of the anatomical definition, histology is required for the diagnosis of emphysema, but computed tomography (CT) can non -invasive provide anatomical information and the extent of emphysema detected with CT has been shown to correlate well with histology [4;7-9]. Therefore, CT can be an attractive alternative to investigate the natural course of emphysema before it reaches the symptomatic stage causing airflow obstruction.

Presently there are several ongoing lung cancer screening trials [10-13]. Since lung cancer and emphysema share smoking as the main risk factor, CT-scans performed in these trials may provide suitable data for studying the natural course of smoking-related emphysema in relatively healthy subjects [14]. These data could be used to select groups of smokers in whom more aggressive risk-modifying treatment is necessary to prevent development of severe lung destruction and airflow limitation.

Previous studies have suggested that subjects with similar amounts of parenchymal destruction can show different degrees of severity of airflow limitation and gas exchange impairment according to the distribution of the damage in either the apical or the basal parts of the lungs [15-17].

The impact of these pattern of emphysema distribution has been investigated in patients with α_1-antitrypsine deficiency and severe emphysema[16] and these data can not be extrapolated easily to smoking induced COPD. In a study with unselected smokers [15], only the extent of subjectively quantified lung destruction showed an association between distribution pattern (i.e. apical or basal predominance of emphysema) and pulmonary function, while the extent of objectively quantified emphysema did not. However, subjective scoring has been demonstrated to result in an overestimation of the extent of lung destruction [18]. Moreover, only two slices, one above and one below the carina, were used for analysis. Therefore, the results of this study could have suffered from selection

bias.
The aim the current study was to investigate the impact of the distribution pattern of both moderate and severe emphysema, detected on volumetric high-resolution CT, on the severity of airflow limitation and gas exchange impairment in a large population of current and former heavy smokers participating in a lung cancer screening trial.

MATERIAL AND METHODS

SUBJECTS

The NELSON-project is the population based randomized Dutch-Belgian multi-center lung cancer screening trial, studying current and former heavy smokers. The trial was approved by the Dutch ministry of health and by the ethics committee of each participating hospital. Selection of participants for the trial was performed by sending a questionnaire about smoking history and other health-related issues to people between 50 and 75 years of age, living in the areas around the participating centers. Subjects meeting the inclusion criteria of a minimum of 16 cigarettes/day for 25 years or 11 cigarettes/day for 30 years, who gave written informed consent were equally randomized to either the screening arm or the control arm. Before inviting eligible subjects, persons with a moderate or poor self-reported health status who were unable to climb two flights of stairs were excluded from participation.
From three thousand participants who underwent baseline screening in our hospital, randomly one in three screened subjects was selected for pulmonary function testing on the same day. For the current study, we included participants who underwent both baseline screening and pulmonary function testing on the same day between April 2004 and February 2005.

CT SCANNING AND CALCULATION EMPHYSEMA SCORES

CT scanning was performed on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA). All scans were realized within 12 seconds, in spiral mode with 16 x 0.75mm collimation, 1.0mm reconstruction thickness, without contrast-injection. Exposure settings were 30mAs at 120kVp for subjects weighing ≤80kg or 30mAs at 140kVp for those weighing over 80kg. Scans were performed in end-inspiration after appropriate instruction of the subjects. They were asked to take a deep breath and to hold their breath.
Scans were transferred to a digital workstation with in-house developed software (ImageXplorer (iX) Image Sciences Institute, Utrecht, The Netherlands). Lung volume was calculated after segmenting both lungs using a fully automated region-growing program starting from the trachea and connecting all areas with a

attenuation below -500HU. In a next step, central airways were separated. The algorithm is similar to the one described by Hu et al [19]. Finally, segmented lungs were subjected to a noise reduction filter [20]. Emphysema score (ES) was calculated for volume with an attenuation below two attenuation thresholds (-950HU and -910HU) as percentage of total lung volume. Lung areas with an attenuation between -910 Hounsfield units (HU) and -950HU represented moderate emphysema, while lung volume with an attenuation below -950HU represented severe emphysema as described by the National Emphysema Treatment Trial (NETT) [21]. Emphysema scores were given as percentage of total lung volume in a range from 0% to 100%.

Pulmonary Function Tests

Pulmonary function tests (PFT) included forced expiratory volume in one second (FEV_1) and vital capacity (VC) with a pneumotachograph and assessment of diffusion capacity (Tlco), according to ERS guidelines [22]. Upon arrival, subjects rested for 15 minutes after which non-forced spirometry was performed, immediately followed by recording the FEV_1 by a flow-volume curve. The best of three temptations was selected for analysis. No reversibility testing was performed. Diffusing capacity measurements were performed after spirometry. The inhalation mixture contained 0.3% CO and 10% He with balance air.
Abnormal pulmonary function parameters were defined as values ≤-1.64 standard deviations below reference values [22] and classified according to the updated GOLD guidelines [3]. Results were expressed as percentages of predicted values.

Distribution of Emphysema and Statistics

The lungs were divided in three parts with equal volumes (top, middle and lower part) and the ES for each part was calculated. Since no consensus how to distinguish apical, basal and homogeneous distributions patterns has been established yet, the effects of apical and basal predominance of emphysema on lung function were evaluated via multiple linear regression. However, the extent of emphysema in these three parts of the lungs was highly correlated and therefore multiple linear regression became less reliable. We solved this by using principal component analysis to obtain perfectly uncorrelated predictors based on the principal components scores. These new predictors were incorporated into the multiple regression procedures and the height of the resulting regression coefficients reflect the impact of that predictor for pulmonary function impairment.
In a second step, we added a model-dependent approach forming two groups based on the area with the highest amount of emphysema: a subgroup with mainly apical emphysema and a subgroup with a basal predominance. While mainly subjects with an apical predominance benefit from lung volume reduction surgery, the subjects with the highest emphysema score in the upper third of the

lungs were formed into the apical group, while all other subjects were grouped as subjects with a basal predominance. Analysis of variance (ANOVA) was used to further examine the effects of distribution patterns on lung function parameters.
We calculated means, standard deviations (SD) and 95% confidence intervals (CI) for normal distributed differences in ES and medians and interquartile ranges for non-normal distributed emphysema scores. Changes in emphysema scores were given as percentages of total lung volume. All statistics were calculated with SPSS statistical software package version 13 (SPSS, Chicago, Ill.). P-values <0.05 were considered significant.

RESULTS

SUBJECTS

Between April 2004 and February 2005, 1386 subjects underwent baseline low-dose chest CT and 545 of them (all male) underwent both baseline CT and pulmonary function testing. Descriptive data are shown in Table 1.

The FEV_1 was abnormally low in 60 (11%) subjects, the FEV_1/VC ratio in 143 (26%) subjects, the Tlco in 167 subjects (31%) and the Tlco/$V_{A\ (alveolar\ volume)}$ in 210 (38%) subjects. Since FEV_1/VC and Tlco/V_A ratio detected more subjects with pulmonary function impairments, we used these parameters for further analysis.

MODERATE EMPHYSEMA

In subjects with moderate emphysema, the principal component analysis delivered the following two variables: the total extent of emphysema and the difference between top and basal ES. Multiple regression analysis incorporating these new variables showed that both components significantly reduced the FEV_1/VC ratio. The regression coefficients were respectively -4.88, and -1.65 (both $p<0.001$), showing a strong effect of the total amount of emphysema and a weaker impact of the difference between the ES in top and basal part of the lungs. For the Tlco/V_A ratio a similar effect was detected: the regression coefficients were -5.33 and -4.50 respectively (both $p<0.001$). The model-independent approach so reported a major effect of distribution patterns on pulmonary function and the impact of the patterns on pulmonary function was further investigated, comparing the apical and basal emphysema subgroups.
An apical predominance of moderate emphysema was associated with a lower FEV_1/VC ratio compared to basal distribution ($p<0.001$). Subjects with an apical predominance showed a FEV_1/VC ratio that was 5.84% (95% CI 3.29% - 8.40%) lower than the FEV_1/VC ratio in subjects with a basal predominance. An apical predominance was also associated with a lower Tlco/V_A, compared to a basal predominance ($p<0.001$). Subjects with an apical predominance showed a Tlco/V_A ratio that was 13.04% (95% CI 9.36% - 16.72%) lower than the Tlco/V_A ratio in subjects with a basal predominance.

SEVERE EMPHYSEMA

In subjects with severe emphysema, the principal component analysis also delivered the total extent of emphysema and the difference between top and basal ES as uncorrelated variables influencing the severity of pulmonary function impairment. The resulting regression coefficients for the two components FEV_1/VC ratio were -4.86 and -1.33 ($p<0.001$ and $p=0.005$) for the total extent of emphysema and the difference between top and basal ES respectively, while for the Tlco/V_A ratio the regression coefficients were -7.33 and -2.23 ($p<0.001$ and

p=0.001).

Apical emphysema was associated with a lower FEV_1/VC and $Tlco/V_A$ ratio ($p<0.001$). Subjects with an apical predominance showed a FEV_1/VC ratio, which was 5.85% (95% CI 2.90% - 8.79%) lower than the ratio in subjects with a basal distribution pattern. For the $Tlco/V_A$ ratio the difference was 8.94% (95% CI 4.34%- 13.55%).

	Apical predominance	Basal predominance
Moderate emphysema		
Number of subjects	167	378
Emphysema score Median Interquartile range	 13.5% 3.76%-32.6%	 7.6% 2.65%-20.2%
FEV1 Median Interquartile range	 97.6% 84.3%-110.6%	 98.9% 87.6%-117.7%
Vital capacity Median Interquartile range	 107.3% 99.0%-116.5%	 103.3% 94.7%-112.5%
FEV1/FVC Median Interquartile range	 70.9% 64.4%-76.2%	 73.9% 67.4%-77.5%
Tlco/VA Median Interquartile range	 81.2% 68.9-96.9%	 92.5% 81.7%-102.5%
Severe emphysema		
Number of subjects	209	336
Emphysema score Median Interquartile range	 0.15% 0.037%-0.94%	 0.13% 0.056%-0.31%
FEV1 Median Interquartile range	 97.8% 86.2%-108.3%	 99.0% 86.6%-110.9%
Vital capacity Median Interquartile range	 103.4% 95.8%-112.9%	 105.0% 95.7%-114.4%
FEV1/FVC Median Interquartile range	 96.0% 86.6%-101.4%	 94.8% 87.8%-100.6%
Tlco/VA Median Interquartile range	 86.9% 72.5%-97.6%	 92.2% 80.4%-100.6%

Table 1

Descriptive statistics shown as median values and 25%-75% interquartile ranges. Emphysema scores represent percentages of total lung volume. All lung function parameters are expressed as percentage of the predicted value. Moderate emphysema is detected as lung volume with an attenuation between -950HU and -910HU; severe emphysema is detected as lung volume with an attenuation below -950HU.

DISCUSSION

In this study investigating current and former heavy smokers mainly with a low extent of lung destruction, we demonstrated that for objectively quantified emphysema an apical predominance of lung destruction is accompanied with more severe airflow limitation and gas exchange impairment than a basal predominance. When using the GOLD-guidelines, subjects with an apical distribution have therefore a higher probability to be diagnosed as suffering from functional COPD than subjects with a similar extent of emphysema, but with a basal predominance. A similar argument holds when the diffusion capacity is used as diagnostic parameter.

The literature on the relation between PFT and emphysema distribution shows varying outcomes. Gurney reported a study investigating 59 heavy smokers that the Tlco and TLC showed stronger correlations with basal emphysema than with apical emphysema, but the FEV_1 and FEF_{25-75} showed the highest correlations with apical emphysema [15]. However, these differences were only statistical significant for subjectively quantified emphysema not for objectively quantified emphysema. In a subgroup of 15 patients with both a lowered Tlco and FEV_1, they showed that the Tlco, TLC and FVC correlated significantly better with lower lung zone emphysema, while Saitoh reported in a study in 62 subjects with a prior diagnosis of emphysema, that the FEV_1/VC showed the strongest correlations for with lower lung emphysema and the $Tlco/V_A$ with upper lung emphysema [23]. Haraguchi included 25 subjects with proven emphysema and reported that the Tlco correlated slightly better with middle and basal emphysema, while the FEV_1 showed the best correlation (r=0.64) with basal emphysema only [24]. Nakano concluded in their study of 73 male patients with a prior diagnosis of COPD that the $Tlco/V_A$ was stronger influenced by upper-inner and middle-inner emphysema and the FEV_1/VC by lower-inner and lower-outer emphysema [25]. The major drawback of this study was the use of stepwise backward multiple regression in which multi-collinearity could strongly have influenced the results. Aziz could not demonstrate any significant correlations in a retrospective study in 101 subjects with evidence of emphysema [26]. Finally, the study by Parr in patients with α_1-antitrypsine deficiency showed that the $Tlco/V_A$ ratio was stronger influenced by upper zone emphysema and the FEV_1 with lower zone emphysema [16]. However, these results could not be repeated in a later study [17].

Several reasons for these variations in results can be mentioned. The quantification method of emphysema was not standardized: some authors quantified the extent of emphysema subjectively, while other groups applied objective methods. The inclusion criteria were variable: some investigators included patients with a prior diagnosis of COPD or emphysema on HRCT, while others included subjects at risk to develop COPD. Selection bias can have influenced the outcomes seriously. In studies with a small sample size, outliers

and power can be a problem, while other studies suffered from statistical problems as multi-collinearity and some studies even excluded subjects from analysis.

In the present study we investigated two stages of emphysema (moderate and severe) and we showed similar effects: the stage of emphysema did not influence the impact of the distribution pattern on the extent of pulmonary function impairment. We avoided power problems by investigating a large group of subjects and included all subjects in the analysis. We dealt with multi-collinearity problems via principal component analysis model-independent analysis. Finally, we included smokers without a prior diagnosis of COPD.

The diffusing capacity is influenced by several parameters. In COPD not only the loss of alveolar membrane surface plays a role, but also the change in alveolar capillaries and the magnitude of the ventilation/perfusion (V/Q) mismatch have an impact on the gas exchange capacity. We reported that loss of alveolar tissue was not always accompanied by reduced diffusing capacity. The destruction of lung tissue in the lower lung zones apparently can be more extensive before pulmonary function impairment develops than the amount of apical emphysema. The sensitivity of the diffusing capacity to detect basal lung tissue destruction is lower than the sensitivity to the detect apical emphysema. Even in emphysematous lungs, the lower parts of the lung receive more blood than the upper parts do due to the effects of gravity. An increased capillary blood volume in the lower lung zones reduces the red blood cell resistance in the Roughton-Forster equation. This effect will increase the diffusing capacity and more severe pathologic changes in the alveolar capillary membrane are required to lower the diffusion capacity below normal values. Therefore we hypothesize that the remaining non-emphysematous lower lung tissue can more longer compensate changes in gas exchange due to lung destruction. This compensation mechanism does not occur in the upper lung zones to a similar extent, since the blood supply in this area is lower and lung destruction will result earlier in gas exchange impairment [27].

For the spirometry, the role of small airway disease in limiting airflow is important. Kim et al showed that two types of lung destruction can develop within smokers and that these types behave mechanically differently [28]. They reported that lung function in subjects with centrilobular emphysema (CLE) was more reduced compared to those with panlobular emphysema (PLE) and they reported that this was caused by the degree of small airway disease (SAD). In CLE the FEV_1/FVC ratio was negatively correlated to the severity of SAD, while in PLE SAD did not contribute to airflow reduction. Based on the data reported by Cosio [29], Saetta concluded that “there is also a definite tendency for small airways in the upper lung to be more diseased than those in the lower lung, a topographical distribution that corresponds to centrilobular emphysema”[30]. These observations agree with the results we reported in this study, that a location of lung destruction has a major impact on the extent airflow and gas exchange limitation,

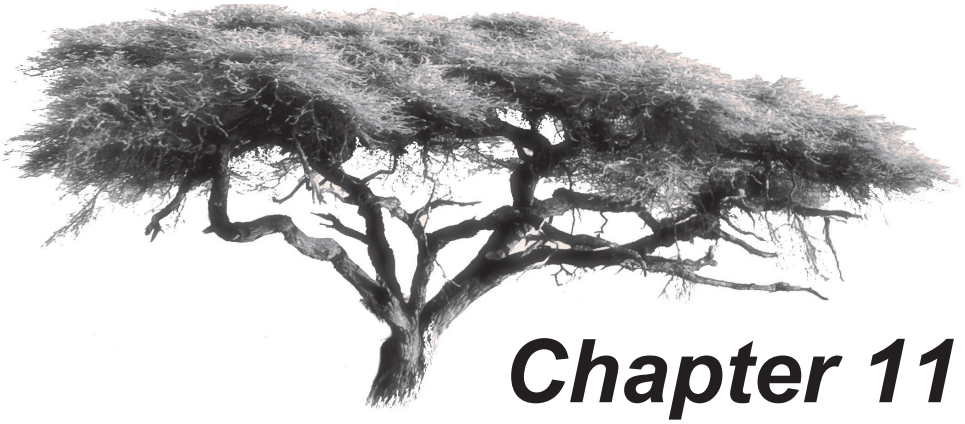
next to the total amount of emphysema. Since small airways disease can lead to airtrapping, this can be an explanation for the reduction of gas exchange in patients with SAD.

In conclusion, subjects with smoking-induced emphysema mainly located in the apical lung zones have a higher risk for developing pulmonary function impairment than those with more basal located emphysema. For this reason, smokers with apical lung destruction should be subjected to a more aggressive risk-modifying approach to prevent them from severe COPD.

REFERENCE LIST

(1) Feenstra TL, van Genugten ML, Hoogenveen RT et al. The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands. Am J Respir Crit Care Med 2001; 164(4):590-596.
(2) Viegi G, Scognamiglio A, Baldacci S et al. Epidemiology of chronic obstructive pulmonary disease (COPD). Respiration 2001; 68(1):4-19.
(3) www.goldcopd.com 2005.
(4) Bankier AA, Madani A, Gevenois PA. CT quantification of pulmonary emphysema: assessment of lung structure and function. Crit Rev Comput Tomogr 2002; 43(6):399-417.
(5) Nakano Yasu, Muro Shig, Sakai Hiro et al. Computed Tomographic Measurements of Airway Dimensions and Emphysema in Smokers. Correlation with Lung Function. Am J Respir Crit Care Med 2000; 162(3):1102-1108.
(6) Clark KD, Wardrobe-Wong N, Elliott JJ et al. Patterns of Lung Disease in a "Normal" Smoking Population : Are Emphysema and Airflow Obstruction Found Together? Chest 2001; 120(3):743-747.
(7) Coxson HO, Rogers RM, Whittall KP et al. A quantification of the lung surface area in emphysema using computed tomography. Am J Respir Crit Care Med 1999; 159(3):851-856.
(8) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.
(9) Uppaluri R, Mitsa T, Sonka M et al. Quantification of pulmonary emphysema from lung computed tomography images. Am J Respir Crit Care Med 1997; 156(1):248-254.
(10) Church TR. Chest radiography as the comparison for spiral CT in the National Lung Screening Trial. Acad Radiol 2003; 10(6):713-715.
(11) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.
(12) Sone S, Takashima S, Li F et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet 1998; 351(9111):1242-1245.
(13) van Klaveren RJ, Habbema JD, de Koning HJ et al. [Screening for lung cancer in the Netherlands: the role of spiral CT scan]. Ned Tijdschr Geneeskd 2001; 145(11):521-526.
(14) Omori H, Nakashima R, Otsuka N et al. Emphysema detected by lung cancer screening with low-dose spiral CT: prevalence, and correlation with smoking habits and pulmonary function in Japanese male subjects. Respirology 2006; 11(2):205-210.
(15) Gurney JW, Jones KK, Robbins RA et al. Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers. Radiology 1992; 183(2):457-463.
(16) Parr DG, Stoel BC, Stolk J et al. Pattern of Emphysema Distribution in {alpha}1-antitrypsin Deficiency Influences Lung Function Impairment. Am J Respir Crit Care Med 2004.
(17) Parr DG, Stoel BC, Stolk J et al. Validation of computed tomographic lung densitometry for monitoring emphysema in {alpha}1-antitrypsin deficiency. Thorax 2006;thx.
(18) Bankier AA, de Maertelaer V, Keyzer C et al. Pulmonary Emphysema: Subjective Visual Grading versus Objective Quantification with Macroscopic Morphometry and Thin-Section CT Densitometry. Radiology 1999; 211(3):851-858.
(19) Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001; 20(6):490-498.
(20) Hilts M, Duzenli C. Image filtering for improved dose resolution in CT polymer gel dosimetry. Med Phys 2004; 31(1):39-49.
(21) Martinez FJ, Foster G, Curtis JL et al. Predictors of Mortality in Patients with Emphysema and Severe Airflow Obstruction. Am J Respir Crit Care Med 2006;200510-1677OC.

(22) Quanjer PH, Tammeling GJ, Cotes JE et al. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J Suppl 1993; 16:5-40.
(23) Saitoh T, Koba H, Shijubo N et al. Lobar distribution of emphysema in computed tomographic densitometric analysis. Invest Radiol 2000; 35(4):235-243.
(24) Haraguchi M, Shimura S, Hida W et al. Pulmonary function and regional distribution of emphysema as determined by high-resolution computed tomography. Respiration 1998; 65(2):125-129.
(25) Nakano Y, Sakai H, Muro S et al. Comparison of low attenuation areas on computed tomographic scans between inner and outer segments of the lung in patients with chronic obstructive pulmonary disease: incidence and contribution to lung function. Thorax 1999; 54(5):384-389.
(26) Aziz ZA, Wells AU, Desai SR et al. Functional Impairment in Emphysema: Contribution of Airway Abnormalities and Distribution of Parenchymal Disease. Am J Roentgenol 2005; 185(6):1509-1515.
(27) DOLLERY CT, DYSON NA, SINCLAIR JD. Regional variations in uptake of radioactive CO in the normal lung. J Appl Physiol 1960; 15:411-417.
(28) Kim DS, Kim YS, Jung KS et al. Prevalence of Chronic Obstructive Pulmonary Disease in Korea: A Population-based Spirometry Survey. Am J Respir Crit Care Med 2005; 172 (7):842-847.
(29) Cosio MG, Hale KA, Niewoehner DE. Morphologic and morphometric effects of prolonged cigarette smoking on the small airways. Am Rev Respir Dis 1980; 122(2):265-21.
(30) Saetta M, Finkelstein R, Cosio MG. Morphological and cellular basis for airflow limitation in smokers. Eur Respir J 1994; 7(8):1505-1515.



Chapter 11

Summary and general discussion

INTRODUCTION

The introduction of spiral, multidetector-row computed tomography (CT) has expanded the applications of CT by increasing the amount of obtained data within one scanning session and limiting the scanning time for a specific area of interest. For example, scanning of the entire chest within one breath hold of about 10 s can be combined with the reconstruction of thin 1.0 mm slices. However, this development was accompanied by an exponential increase in applied medical radiation dose and thereby possibly a larger hypothetical risk for the patient of developing radiation-induced cancer [1;2]. With present standard techniques, an effective dose between 4 and 8 mSv is delivered during one scanning session (UNSCEAR rapport 2001), which leads to a calculated risk of dying from radiation -induced cancer of 2-4 in 10.000 (30-year old) individuals.
Although the correlation between radiation exposure and radiation-induced cancer is still unclear for relatively low amounts of radiation, one should aim to keep the radiation dose applied during a medical examination to a minimum. For this reason the International Commission on Radiological Protection has introduced the As Low As Reasonably Achievable (ALARA) principle [3]. In the past decade radiologists started to pay attention to the possibilities of dose reduction and individually adjusted protocols have become available. Reduction of radiation dose has been shown to be feasible while maintaining enough image quality to gain the required diagnostic information.

In this thesis, we showed that low dose CT is feasible for several indications, in which the increased image noise has more or less impact on the diagnostic quality and the validity of automated quantifications of the size of abnormal areas, such as pulmonary nodules and emphysema. For quantification of structures which show high contrast with the surrounding tissue, such as a lung nodule within normal lung tissue, the diagnostic information is not limited by the amount of image noise. However, in case of a low intrinsic contrast between the area of interest and the surrounding tissue, for example in case of low extents or diffuse located emphysema, noise has a massive impact on the diagnostic process.

PART 1

In areas with a high intrinsic contrast such as the chest, radiation dose can further be reduced for specific indications. In clinical practice, many benign diseases are being followed-up by imaging. This is mainly performed with plain X-rays. Although CT scans provide more information than plain radiography, plain radiography is often the first choice due to the high radiation dose and the high costs of CT scans. However, in about 10% of the patients, who require a diagnosis for new symptoms, a CT scan is indicated after plain chest X-ray performed, because CXR did not provide the required diagnostic information. This approach

leads to more radiation dose, more costs and a larger burden for the patient. We showed in a small sample of patients from the outpatient department of pulmonology that ultralow-dose CT is feasible and can provide more information than CXR in two directions, while radiation dose is similar. More patients have to be studied to support these results and to detect particular subgroups of patients for which the combination CXR and standard chest CT could be replaced by an ultralow-dose chest CT scan.
The main disadvantage of radiation dose reduction is the accompanying increase in image noise. The detection of abnormalities with a high contrast to the surrounding normal tissue is not limited by the amount of image noise on the CT scans performed with the current minimum radiation dose [4]. However, when the structure of interest shows a low contrast to the surrounding normal tissue, image noise can hamper the detection and especially the automated size measurement of the abnormality. We demonstrated the effect of massive dose reduction on the results of the automated quantification of volume with an attenuation below a fixed threshold. This effect is mainly seen for low extents of lung destruction with a low number of voxels below the chosen attenuation threshold and a steep histogram at the level of the chosen threshold. We showed that an increase in image noise results in overestimation of the emphysema score compared to standard dose CT, but also that the application of a dedicated noise reduction filter to the reconstructed data before the automated quantification of emphysema can prevent this overestimations. We conclude that automated emphysema quantification on low dose CTs is feasible even for early stage emphysema, but only when a dedicated noise reduction filter is applied. This chapter forms the reasoning for the standard application of a noise reduction filter in chapters 7-10. Our standard dose chest CTs also show a certain amount of image noise. Since we did not have a gold standard available like histopathological specimens or emphysema scores at maximal dose, we were not able to judge the accuracy of emphysema scoring on our standard dose scans.

PART 2

The development of low dose protocols has made screening for cancer with chest CT feasible. In a screened population of current and former heavy smokers, the risk of the applied radiation dose is largely exceeded by the risk of dying from lung cancer. Detection of lung cancer in an early, curable stage could reduce lung cancer-related mortality in this population, but the benefit of screening for lung cancer has not yet been proven. The Dutch-Belgian Lung Cancer Screening Trial (NELSON) will be the first randomized controlled trial to show the benefits and limitations of screening for lung cancer in a high-risk population of current and former heavy smokers, using a highly sensitive imaging modality, i.e. low-dose , and no intervention in the control arm.
Lung cancer is today the most frequent cause of cancer death in the Western

world. On a global basis it is estimated that 1.2 million people are diagnosed with this disease every year (12.3% of the total number of cancer diagnosed), and about 1.1 million people are dying of this disease yearly (17.8% of the total cancer death) [5]. More than two-third of these people are diagnosed with locally advanced or metastatic disease, and their poor prognosis is due to a late diagnosis and lack of effective treatment of metastatic disease. Less than 15% of the patients are surviving after 5 years, and in several European countries the 5-year survival is far less.

Since the first publications of the use of the spiral CT in the early diagnosis of lung cancer, and especially the report from the Early Lung Cancer Action Program (ELCAP) in 1999, the interest in lung cancer screening by low-dose spiral CT has increased tremendously[6]. Henschke *et al* recently showed that annual screening with low dose CT can detect lung cancer in a curable stage [7].

Low dose CT has been shown to be extremely sensitive to detect early lung cancer[8;9], but the technique is not very specific[6]. About 70% of the screened subjects showed at least one non-calcified nodule at baseline-screening, which has the potential to be malignant. The vast majority of these nodules will turn out to be benign as proven by other trials [6;10;11]. We reported the way these nodules are managed within the NELSON-trial.

Growth is the most used feature to distinguish the few malignant nodules from the benign ones, but to detect growth and to calculate reliable volume doubling times, precise size measurements are of utmost importance. Revel et al showed that volumetric measurements are more accurate than two-dimensional size measurements [12], but in agreement with Wormanns et al [13], we showed that for Lungcare©, a commercially available and widely used semi-automated volume measurement program, the interobserver variability was low, but the interscan variability could be substantial. The reported limits of agreement have to be taken into account when interpreting an increase in measured nodule size on a repeat scan. We showed that the performance of the algorithm to segment the nodule completely was the most important factor contributing to the variability, using an algorithm that started from a spherical shape of a nodule. Since many nodules detected in a lung cancer screening setting are not spherical, the segmentation step often failed to include the whole nodule. We reported that an increase in volume on a repeat scan of more than 25% for non-spherical nodules and 15% for spherical nodules can with 95% likelihood be applied to a real increase in volume.

More sophisticated nodule segmentation algorithms have already been developed and they are now being released. Our study should be repeated using these new algorithms to set a new reference for measurement error. Although we did not compare the volume measurements on low dose CTs to the results on standard dose CTs, Karabulut et al already showed similar nodule size for both protocols [14].

PART 3

Since chronic obstructive pulmonary disease (COPD) and lung cancer share smoking as main risk factor, lung cancer screening trials provide a good opportunity to study the early stages and natural progression of COPD. Ezatti *et al* reported in 2003 that in the Western countries the number of people that each year dies from COPD is larger than the number of people that dies from lung cancer [15]. COPD is predicted to become the third cause of death from in developed countries in 2020 [16]. COPD is defined as an airflow limitation that is not fully reversible. However, airflow limitation is a relative late symptom of a cascade of smoking related changes caused by the exposure to noxious agents. This cascade comprises lung inflammation, lung tissue destruction, impairment of defense mechanisms that serve to limit the destruction and disruption of the repair mechanisms that may be able to restore the tissue structure in the face of some injuries (www.goldcopd.com). The results of tissue destruction are mucus hypersecretion, airway narrowing and fibrosis, destruction of the lung parenchyma (emphysema) and vascular changes.

Emphysema can easily be detected and automated be quantified on CT by highlighting voxels with an abnormally low X-ray attenuation. Several groups have shown that the results of this method demonstrate a good correlation with the gold standard, the extent of lung destruction in histopathological samples [17-20]. The method is based on the relative increase of air within a voxel due to tissue destruction. Since air has a lower attenuation than tissue, the destruction of lung tissue results in a lowered attenuation within the voxels. The number of voxels with such a lowered X-ray attenuation can be quantified and expressed as percentage of the total number of voxels within the lungs (0-100%). We have called the resulting parameter the emphysema score, but this parameter was first described for 2D datasets and known as the pixel index. Nowadays, many groups use 3D datasets and present the results as the voxel index. However, this method also suffers from some variables influencing the results other than the extent of lung destruction. The impact of slice thickness has been described by Kemerink *et al* [21]. The smaller the size of a voxel, the more different structure will end up in different voxels, resulting in more voxels at the end of the spectrum. Thinner slices result in smaller voxels and in more voxels at the lower end of the frequency histogram of CT-numbers, resulting in higher emphysema scores at a fixed attenuation threshold compared to the conventional CT reconstructed at 10mm slices.

Another variable influencing the emphysema score is the level of inspiration. Lamers *et al* [22] and Kalender *et al* [23] demonstrated the relation between the level of inspiration and the resulting emphysema score. The amount of air within the lung will increase at a higher level of inspiration and this effect results in a lower mean CT-number. We investigated the reproducibility of the quantification of emphysema scores and showed that the level of inspiration was one of the factors

influencing the reproducibility. However, the correlation between the difference in emphysema score and the difference in inspiration level was weak.
To assess the progression of the extent of moderate and severe emphysema, again knowledge of the extent of measurement error between scans is required to distinguish between progression of emphysema and measurement error. We provided the limits of agreement for three common attenuation thresholds for emphysema scores obtained in a lung cancer screening setting. The reported coefficients of variations were high, especially when compared to the coefficients of variation reported by Revel *et al* [24] for volume measurements of non-calcified nodules detected on low dose chest CT. We hypothesize that the main explanation for this difference in coefficient in variation is the difference in contrast of the abnormality of interest to the surrounding normal tissue as mentioned above. Although we used a dedicated noise reduction filter, the automated quantification technique remains sensitive to noise or small changes in CT-numbers. We scanned a phantom before and after a day of scanning NELSON participants to be able to detect and correct for changes in scanner calibration. We detected small variations in the CT-numbers in a homogeneous region of interest measured on the two scans performed on one day, but also during the NELSON-trial and when performing 5 consecutive scans. Since these variations were within the limits provided by the vendor (Philips) and independent of the time of scanning, we were not able to correct the scans performed as part of the NELSON-trial for these small fluctuations. However, these fluctuations can contribute to the coefficients of variations and the broad limits of agreement. Sophisticated methods to detect small shifts in the frequency histogram of CT-numbers may be able to correct for the level of inspiration, but also for small fluctuations in CT-numbers due to variation caused by the CT scanner.
Finally, the anatomical information about the extent and location of emphysema has been compared to the functional information provided by pulmonary function tests. Currently, pulmonologists treat only patients with pulmonary function impairment, as recommended by the Global initiative for chronic Obstructive Lung Disease (GOLD). Staging of COPD is mainly based on the percentage of predicted forced expiratory volume in one second (FEV_1), a marker of airflow limitation and to a lesser extent to the forced vital capacity (FVC). The GOLD-guidelines recommend management to:

- prevent disease progression
- relieve symptoms
- improve exercise tolerance
- improve health status
- prevent and treat complications
- prevent and treat exacerbations
- reduce mortality

The staging according to the GOLD-guidelines also describes a population at risk to develop COPD, this is the population suffering from chronic cough and sputum production, but with a normal lung function. However, not all subjects at risk will develop COPD. In this stage of disease lung function is normal, but other features of smoking related disease may already be present. One of these features is lung destruction. Detection of early lung destruction before COPD becomes symptomatic may prevent the patients from progression to the symptomatic stage. The population enrolled in the NELSON-trial is the population at risk to develop COPD, but without severe COPD, since they have to be fit enough to undergo surgery.

Although COPD is predicted to be the third cause of morbidity and mortality in Western countries in 2020 [16], little is known about the early stages of disease. A lot of research is being performed on development of airway inflammation, but little is known about the progression in disease from "at risk" to airflow limitation. Moreover, little is known about the subjects who will develop COPD and the subjects who will not. Emphysema scoring for each performed chest CT could detect emphysema in an early stage. We showed that many subjects have destructed parenchyma, but a preserved lung function. However, the gas exchange is often already impaired in these subjects, but gas exchange is not one of the hallmarks of COPD (yet). The diffusion of D_{NO} is shown to be a marker of vascular changes, which can precede macroscopic lung tissue destruction and can be impaired before CT can detect emphysema. We showed that D_{NO} is a good sensitive marker to detect emphysema.

The main limitation of studying early emphysema on low dose scans, is the lack of information about the accuracy of the emphysema scores to represent real lung destruction. In a screening study, scanning healthy participants, typically no histological tissue is available. For this reason, we could not judge the results of our emphysema scoring program. Müller *et al* [19] and Gevenois *et al* [17;18] have shown that the extent of emphysema quantified on CT shows a good correlation with the amount of lung destruction detected by macroscopic and microscopic techniques, but we were not able to support these data for low dose scans. A study comparing the emphysema scores obtained from low dose scans performed in lung cancer screening trials to histopathological samples, for example from patients undergoing lobectomy for a malignant nodule, can provide information about the accuracy of emphysema scores to represent real lung destruction.

In the final chapter, we showed that the distribution pattern of emphysema has an impact on the extent of airflow impairment. Participants of the NELSON-project with an apical predominance of emphysema showed more severe airflow limitation than subjects with an equal emphysema score, but a basal predominance of lung destruction. Longitudinal studies investigating the progression of emphysema according to the distribution of emphysema can answer the question if subjects with an apical distribution show more progression

of disease or if subjects showing progression of emphysema will demonstrate a shift to an apical distribution. When subjects with an apical distribution show more progression of disease, these subjects can comprise a subgroup for which a more aggressive risk-modifying approach is required. Only with this information one can predict the clinical relevance and course of low amounts of emphysema in asymptomatic subjects.

Overall Conclusion and Future Perspectives

In this thesis, we have shown that low dose chest CT is feasible and cannot only be applied for lung cancer screening, but also in daily routine and for early detection of lung destruction. We demonstrated that the increased image noise is a limiting factor for the detection of early emphysema, showing low contrast to the surrounding normal tissue, but also that this limitation can be overcome by the use of a dedicated noise reduction filter.

The detection of growth is the main feature to distinguish malignant non-calcified nodules from potentially benign ones, but inter-scan and to a lesser extent inter-observer variability can limit the detection of growth. However, when a nodule shows more than 25% increase in volume, a criterion which will be fulfilled by most malignant nodules on a short-term follow-up CT after three months, it has a likelihood of 95% to be really grown. New segmentation algorithms considering potentially malignant pulmonary nodules as non-spherical may reduce the inter-scan variability, decreasing the upper threshold of agreement. With these algorithms the growth rate can more precisely be calculated and real growth can more easily be distinguished from measurement variation.

This thesis will be the start of further research dealing with the detection and quantification of low-attenuation areas on low dose chest CTs performed in asymptomatic, but high-risk subjects within the scope of the NELSON-trial. We described the baseline emphysema scores in a screening population, but the results of the follow-up study will be more interesting. This data will provide insight in the natural course of emphysema and the impact of the extent of mild and severe emphysema at baseline, the distribution of emphysema at baseline and the impact of quitting smoking before the start of the trial, quitting smoking during the trial and continuous smoking on the progression of emphysema.

REFERENCE LIST

(1) de Gonzalez ABDS. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet 2004; 363(9406):345-351.
(2) Brenner DJ, Doll R, Goodhead DT et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc Natl Acad Sci U S A 2003; 100 (24):13761-13766.
(3) Prasad KN, Cole WC, Haase GM. Radiation protection in humans: extending the concept of as low as reasonably achievable (ALARA) from dose to biological damage. Br J Radiol 2004; 77(914):97-99.
(4) Itoh S, Ikeda M, Arahata S et al. Lung cancer screening: minimum tube current required for helical CT. Radiology 2000; 215(1):175-183.
(5) Landis SH, Murray T, Bolden S et al. Cancer statistics, 1999. CA Cancer J Clin 1999; 49(1):8-31.
(6) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.
(7) The International Early Lung Cancer Action Program. Survival of Patients with Stage I Lung Cancer Detected on CT Screening. N Engl J Med 2006; 355(17):1763-1771.
(8) Kaneko M, Eguchi K, Ohmatsu H et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. Radiology 1996; 201(3):798-802.
(9) Henschke CI, Naidich DP, Yankelevitz DF et al. Early lung cancer action project: initial findings on repeat screenings. Cancer 2001; 92(1):153-159.
(10) Sone S, Takashima S, Li F et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet 1998; 351(9111):1242-1245.
(11) Swensen SJ, Jett JR, Hartman TE et al. Lung cancer screening with CT: Mayo Clinic experience. Radiology 2003; 226(3):756-761.
(12) Revel MP, Bissery A, Bienvenu M et al. Are Two-dimensional CT Measurements of Small Noncalcified Pulmonary Nodules Reliable? Radiology 2004; 231(2):453-458.
(13) Wormanns D, Kohl G, Klotz E et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004; 14(1):86-92.
(14) Karabulut N, Toru M, Gelebek V et al. Comparison of low-dose and standard-dose helical CT in the evaluation of pulmonary nodules. Eur Radiol 2002; 12(11):2764-2769.
(15) Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. The Lancet 2003; 362(9387):847-852.
(16) Feenstra TL, van Genugten ML, Hoogenveen RT et al. The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands. Am J Respir Crit Care Med 2001; 164(4):590-596.
(17) Gevenois PA, De Vuyst P, De M, V et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1996; 154(1):187-192.
(18) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.
(19) Muller NL, Staples CA, Miller RR et al. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988; 94(4):782-787.
(20) Madani A, Zanen J, de Maertelaer V et al. Pulmonary Emphysema: Objective Quantification at Multi-Detector Row CT--Comparison with Macroscopic and Microscopic Morphometry. Radiology 2006;2382042196.
(21) Kemerink GJ, Kruize HH, Lamers RJ et al. CT lung densitometry: dependence of CT number histograms on sample volume and consequences for scan protocol comparability. J Comput Assist Tomogr 1997; 21(6):948-954.
(22) Lamers RJ, Thelissen GR, Kessels AG et al. Chronic obstructive pulmonary disease: evaluation with spirometrically controlled CT lung densitometry. Radiology 1994; 193 (1):109-113.
(23) Kalender WA, Rienmuller R, Seissler W et al. Measurement of pulmonary parenchymal attenuation: use of spirometric gating with quantitative CT. Radiology 1990; 175(1):265-268.

(24) Revel MP, Lefort C, Bissery A et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. Radiology 2004; 231(2):459-466.

NEDERLANDSE SAMENVATTING

De introductie van de multislice spiraal CT-scanner heeft de mogelijkheid geschapen voor een toenemend aantal toepassingen door de hoeveelheid data in een scansessie te vergroten en de scantijd voor een specifiek deel van het lichaam te verminderen. Het is bijvoorbeeld mogelijk om aansluitende 1.0mm dunne plakken te reconstrueren voor de borstkas, terwijl de patiënt slechts 15 seconden of minder zijn adem hoeft in te houden. Deze ontwikkelingen gaan echter gepaard met een exponentiële toename van de hoeveelheid röntgenstraling waaraan de patiënten blootgesteld worden en het daarmee samengaande hypothetische risico op het ontwikkelen van kanker [1;2]. Met de huidige standaard technieken wordt een patiënt gedurende één scan sessie blootgesteld aan 4 tot 8 mSv effectieve dosis (UNSCEAR rapport 2001), wat leidt tot een berekend risico op overlijden van 2-4 op 10.000 personen ten gevolge van stralingsgeïnduceerde maligniteiten.
Hoewel het verband tussen de dosis röntgenstraling en het ontstaan van kanker vooralsnog onduidelijk is voor lage doses straling, dient men voor medische doeleinden te streven naar een minimale dosis toegediende röntgenstraling. Daarom heeft de Internationale Commissie voor Radiologische Bescherming het zogenaamde ALARA ("As Low as Reasonaly Achievable") principe geïntroduceerd [3], waarbij gestreefd wordt de dosis zo laag als mogelijk te houden. In het afgelopen decennium zijn steeds meer radiologen aandacht gaan besteden aan de mogelijkheden van dosisreductie en zijn protocollen beschikbaar gekomen, waarbij rekening gehouden wordt met de individuele patiënt. De dosis röntgenstraling blijkt te kunnen worden gereduceerd, terwijl de kwaliteit van de afbeeldingen voldoende is om de vereiste diagnostische informatie te verkrijgen.

In dit proefschrift hebben we aangetoond dat het voor verschillende indicaties mogelijk is CT-scans te maken met lage dosis röntgenstraling, waarbij de toename van beeldruis in wisselende mate invloed heeft op de diagnostische kwaliteit en de validiteit van de automatische kwantificatie van afwijkende gebieden, zoals longnodules of emfyseem. Voor kwantificatie doeleinden waarbij een groot contrast is tussen de afwijkende structuur en het omgevende normale weefsel, zoals in het geval van een longnodule, wordt de diagnostische informatie niet beperkt door de hoeveelheid beeld ruis. Voor doeleinden waarbij er laag intrinsiek contrast is tussen het abnormale gebied en het omgevende weefsel, bijvoorbeeld in geval van kleine gebieden van emfyseem, heeft ruis echter een groot effect op het diagnostische proces.

DEEL 1

In gebieden met een hoog intrinsiek contrast zoals de longen, kan de stralingsdosis sterk worden gereduceerd voor specifieke indicaties. In de dagelijkse praktijk worden veel goedaardige ziekten vervolgd door middel van beeldvorming. Dit gebeurt meestal met conventionele röntgenfoto's. Hoewel CT-

scans meer informatie opleveren, is de longfoto door de hoge stralingsdosis en de kosten van CT-scans toch eerste keus. Bij patiënten met nieuwe symptomen wordt in ongeveer 10% van de gevallen vervolgens toch een CT-scan vervaardigd, omdat de conventionele foto geen uitsluitsel biedt. Deze aanpak leidt tot meer röntgenstraling, meer kosten en meer belasting voor de patiënt. Wij hebben in een kleine groep patiënten van de polikliniek voor longziekten aangetoond, dat het mogelijk is om met een ultra-lage stralingsdosis te scannen en toch meer informatie te verkrijgen dan met behulp van longfoto's. Om dit resultaat verder te ondersteunen zullen echter grotere groepen patiënten onderzocht moeten worden, zodat specifieke subgroepen aangewezen kunnen worden, voor de combinatie conventionele longfoto & standaard CT-scan vervangen kan worden door een ultra -lage dosis CT-scan.

Het grootste nadeel van de vermindering van röntgenstraling is dat dit gepaard gaat met een toename van beeldruis [4]. De detectie van afwijkingen met een groot contrast met het omgevende normale weefsel wordt niet beperkt door de hoeveelheid beeld ruis op de CT-scans gemaakt met de huidige minimale hoeveelheid straling. Wanneer een afwijkende structuur weinig contrast met het omliggende weefsel vertoont, kan ruis de detectie en met name de automatische metingen verstoren. Wij hebben laten zien, wat het effect van veel ruis is op de resultaten van automatische kwantificatie van volumes met een abnormaal lage densiteit. Dit effect is met name aan de orde voor kleine hoeveelheden gedestrueerd longweefsel met een laag aantal voxels onder de gekozen drempelwaarde en een steile helling van de curve van het frequentie histogram ter plaatse van de gekozen drempelwaarde. We hebben aangetoond dat een toename van de ruis leidt tot een overschatting van de hoeveelheid emfyseem in vergelijking met de CT-scan gemaakt met een normale dosis röntgenstraling, maar ook dat de toepassing van een speciaal ontworpen ruisfilter op de gereconstrueerde data voordat het emfyseem automatisch gekwantificeerd wordt, overschatting kan voorkomen. Uit deze resultaten hebben wij geconcludeerd dat automatische kwantificatie van emfyseem in een vroeg stadium ook mogelijk is in het geval van een CT-scan gemaakt met een lage dosis röntgenstraling, maar alleen wanneer een speciaal ontworpen ruisfilter wordt toegepast om de hoeveelheid ruis terug te dringen. Deze resultaten zijn de onderbouwing van het toepassen van een ruisfilter in de studies beschreven in de hoofdstukken 7-10. Onze CT-scans gemaakt met een dosis röntgenstraling zoals gebruikt in onze dagelijkse klinische praktijk is niet vrij van ruis. Aangezien wij niet beschikken over weefsel van de patiënten of een emfyseemscore verkregen van een CT-scan waarbij de maximale röntgendosis is toegepast, waren we niet in staat om een uitspraak te doen over de betrouwbaarheid van de emfyseemscores van de CT-scans gemaakt volgens het standaard protocol.

DEEL 2

De ontwikkeling van scan protocollen met een lage dosis röntgenstraling heeft er toe geleid dat screening voor longkanker met behulp van CT-scan van de thorax mogelijk is. In de gescreende populatie van voormalige en actieve zware rokers wordt het risico van overlijden aan stralingsgeïnduceerde tumoren ruimschoots overschreden door het risico op overlijden aan longkanker. Opsporen van longkanker in een vroeg stadium, wanneer het nog behandelbaar is, zou de sterfte aan longkanker in deze populatie sterk terug kunnen dringen, maar het voordeel van screenen op longkanker is vooralsnog niet bewezen. Het Nederlands-Leuvens Longscreenings ONderzoek (NELSON) zal het eerste gerandomiseerde onderzoek zijn zonder interventie in de controlearm zijn, dat de voordelen en beperkingen van screening op longkanker in een hoogrisico populatie middels lage dosis CT-scan aan kan tonen.

Vandaag de dag is longkanker de meest voorkomende oorzaak van dood door kanker in de Westerse wereld. Wereldwijd is geschat dat circa 1,2 miljoen mensen (12,3% van het totaal aantal kankers dat gediagnosticeerd wordt) elk jaar de diagnose longkanker krijgen en dat circa 1,1 miljoen mensen (17,8% van het totaal aantal doden door kanker) elk jaar aan deze ziekte overlijden [5]. Meer dan twee derde van deze mensen hebben lokaal of op afstand gemetastaseerde ziekte ten tijde van de diagnose en hun slechte prognose wordt met name bepaald door de late diagnose en het gebrek aan effectieve behandeling voor metastatische ziekte. Minder dan 15% van de patiënten overleeft minimaal 5 jaar en in een groot aantal Europese landen is de 5-jaars overleving nog veel slechter.

Sinds de eerste publicaties over het gebruik van de spiraal CT-scan voor de vroege opsporing van longkanker en met name de rapportage van de “Early Lung Cancer Action Project” (ELCAP) in 1999, is de interesse in longkanker screening met behulp van lage dosis CT-scans enorm toegenomen [6]. Henschke en anderen hebben recent aangetoond dat jaarlijkse screening met lage dosis CT-scan longkanker in een vroeg stadium op kan sporen [7].

Er is aangetoond dat lage dosis CT een extreem sensitieve methode is om longkanker in een vroeg stadium op te sporen [8;9], maar de techniek is niet erg specifiek [6]. Ongeveer 70% van de gescreende individuen had minimaal één niet-verkalkte nodule tijdens de uitgangsscan screening, welke in theorie maligne zou kunnen zijn. Het overgrote deel van deze nodules zal benigne blijken te zijn, zoals aangetoond in andere onderzoeken [6;10;11]. Wij beschreven hoe binnen NELSON met de nodules wordt omgegaan.

Groei is het meest gebruikte kenmerk om de paar maligne nodules te onderscheiden van de grote groep benigne nodules, maar om groei te kunnen detecteren en om betrouwbare volume verdubbelingstijden te kunnen berekenen, zijn precieze volumemetingen van groot belang. Revel e.a. hebben laten zien dat volumemetingen nauwkeuriger zijn dan twee dimensionale metingen [12], maar in navolging van Wormanns e.a [13] hebben wij aangetoond voor Lungcare©, een

commercieel verkrijgbaar en veel gebruikt semi-automatisch programma voor het meten van volumes van nodules, dat de interobserver variatie laag is, maar dat de interscan variatie substantieel kan zijn. De beschreven grenzen van het betrouwbaarheidsinterval moeten in acht genomen worden wanneer een toename in volume van een nodule op een herhaalscan wordt geïnterpreteerd. Wij hebben aangetoond dat de in hoeverre het algoritme de gehele nodule heeft gesegmenteerd de meest bepalende factor is voor de variatie, waarbij we gebruik gemaakt hebben van een algoritme dat uit gaat van een sferisch gevormde nodule. Aangezien in een long screenings setting niet mooi rond zijn, wordt in de segmentatie stap vaak niet de hele nodule geïncludeerd. Wij hebben beschreven dat een toename in volume op een herhaalscan van meer dan 25% voor een niet-sferische nodule en meer dan 15% met 95% betrouwbaarheid kan worden toegeschreven aan een echte toename in volume.
Ondertussen zijn er meer geavanceerdere algoritmen ontwikkeld, welke momenteel op de markt gebracht worden. Onze studie zou moeten worden herhaald met de algoritmen om een nieuwe referentie te bepalen voor de meetfout van deze algoritmen. Hoewel we de volumemetingen van de nodules gescand met lage dosis niet hebben vergeleken met resultaten van CT-scans gemaakt met standaard dosis, hebben Karabulut e.a. al aangetoond dat de metingen vergelijkbaar waren voor deze protocollen [14].

DEEL 3

Aangezien roken de belangrijkste risicofactor is voor het krijgen van zowel chronisch obstructieve long ziekte (COPD), waar emfyseem deel van uitmaakt, als longkanker, vormen vormt screening op longkanker met een goede gelegenheid om de vroege stadia van emfyseem en het natuurlijke verloop van de ziekte te bestuderen. Ezzati e.a. rapporteerden in 2003 dat het aantal mensen dat jaarlijks in de Westerse wereld overlijdt aan COPD groter is dan het aantal mensen dat overlijdt aan longkanker [15]. Er is voorspeld dat COPD de derde doodsoorzaak in de eerste wereld wordt in 2020 [16]. COPD is gedefinieerd als luchtwegobstructie dat niet geheel reversibel is. Luchtwegobstructie is echter een laat symptoom in een cascade van veranderingen die geïnduceerd worden door de blootstelling aan schadelijke stoffen. Deze cascade bestaat uit longontsteking, weefselschade, verslechtering van verdedigingsmechanismen die de schade zouden moeten beperken en de verstoring van reparatiemechanismen, welke de weefselstructuur zouden kunnen herstellen bij sommige beschadigingen (www.goldcopd.com). De resultaten van weefseldestructie zijn mucus hypersecretie, luchtwegvernauwing en fibrose, destructie van longparenchym (emfyseem) en vasculaire veranderingen.
Emfyseem kan vrij simpel worden opgespoord en automatisch worden gekwantificeerd op CT-scans door middel van het markeren van voxels met een abnormaal lage densiteit. Verschillende onderzoeksgroepen hebben aangetoond

dat deze methode een goede correlatie vertoont met de gouden standaard, de hoeveelheid weefselschade in histologische samples [17-20]. De methode is gebaseerd op het relatieve toename van lucht in een voxels als gevolg van weefsel destructie. Aangezien lucht een lagere densiteit heeft dan weefsel, de destructie van longweefsel leidt tot een lagere densiteit in de voxels. Het aantal voxels met zo'n verlaagde densiteit kan worden gekwantificeerd en uitgedrukt als percentage van het totaal aantal voxels in de longen (0-100%). De methode heeft echter last van een aantal andere variabelen dan de hoeveelheid emfyseem die resultaten beïnvloeden. De invloed van plakdikte is beschreven door Kemerink e.a. [21]. CT beelden zijn niet oneindig precies, maar bestaan uit voxels. De grootte van een voxel wordt bepaald door de plakdikte, de grootte van het beeld en de reconstructie matrix. Dunnere plakken leiden tot kleinere voxels en meer voxels met een densiteit in de buurt van lucht, waardoor er meer voxels een densiteit zullen hebben onder een bepaalde drempelwaarde in vergelijking met CT-scans gereconstrueerd tot 10.0mm dikke plakken.
Een andere variabele die de hoogte van emfyseemscores beïnvloedt is het inspiratieniveau. Lamers e.a. [22] en Kalender e.a. [23] hebben de relatie tussen het inspiratieniveau en de hoogte van de emfyseemscores aangetoond. De hoeveelheid lucht in de long neemt toe bij een hoger inspiratieniveau resulterend in gemiddeld lager CT-nummer. We hebben onderzocht hoe reproduceerbaar de emfyseemscores zijn en hebben aangetoond dat de diepte van inspiratie een van de factoren was die de reproduceerbaarheid beïnvloeden. De relatie tussen beide factoren was echter zwak.
Om progressie van matig en ernstig emfyseem te bepalen is het weer belangrijk om te weten hoe groot de meetvariatie is om progressie te kunnen onderscheiden van meetvariatie. Wij hebben in ons longkanker screeningsonderzoek de grenzen van het betrouwbaarheidsinterval bepaald voor drie veel gebruikte drempelwaarden om emfyseem te scoren. De gerapporteerde variatie coëfficiënten waren erg hoog, met name in vergelijking met de variatie coëfficiënten die Revel e.a. hebben gerapporteerd voor volume metingen van niet-verkalkte nodules [24] gedetecteerd op lage dosis CT-scans. Wij vermoeden dat de belangrijkste verklaring voor dit verschil het verschil in contrast van de afwijking tot het omliggende gezonde weefsel is, zoals hierboven beschreven. Hoewel we een speciaal ontworpen ruisfilter hebben, blijft de automatische emfyseem kwantificatie gevoelig voor ruis. Wij hebben een fantoom gescand voor en na een dag scannen in het kader van het NELSON-project om veranderingen in scanner calibratie op te sporen en om in staat te zijn hiervoor te corrigeren. We ontdekten kleine variaties in de CT-waarden in een homogeen gebied gemeten op twee CT-scans gemaakt op dezelfde dag, maar ook gedurende de NELSON-project en wanneer we 5 scans achtereen maakten. Aangezien deze variatie binnen de grenzen valt, die opgegeven zijn door de fabrikant (Philips) en onafhankelijk van het moment van scannen zijn, waren we niet in staat om de CT-scans gemaakt voor het NELSON-project te corrigeren voor deze kleine fluctuaties. Deze variaties

kunnen echter wel bijdragen aan de grote meetvariaties. Speciale methoden die kleine verschuivingen in het frequentie histogram kunnen ontdekken kunnen misschien corrigeren voor het inspiratieniveau, maar ook voor kleine variaties veroorzaakt door de CT-scanner.

Tenslotte hebben we de anatomische informatie over de uitgebreidheid en de locatie van emfyseem vergeleken met functionele informatie van longfunctietesten. Op dit moment behandelen de longartsen alleen patiënten met een verminderde longfunctie zoals aangeraden door de "Global initiative for chronic Obstructive Lung Disease (GOLD). Stadiëring is voornamelijk gebaseerd op het percentage van voorspeld voor leeftijd en geslacht, van het geforceerde expiratoire volume in 1 seconde (FEV1), een marker voor luchtwegobstructie en minder op de geforceerde vitale volume (FVC). De GOLD-richtlijnen raden beleid aan ten aanzien van:

- het voorkomen van progressie van ziekte
- het verminderen van symptomen
- het verbeteren van de inspanningstolerantie
- het verbeteren van de gezondheidsstatus
- het voorkomen en behandelen van complicaties
- het voorkomen en behandelen van exacerbaties
- het reduceren van mortaliteit

De stadiëring volgens de GOLD-criteria beschrijft ook de populatie die risico loopt om COPD te ontwikkelen. Dit is de populatie die last heeft van chronisch hoesten en sputum productie, maar met een normale longfunctie. Niet iedereen in deze groep zal echter COPD ontwikkelen. In dit beginstadium van de ziekte is de longfunctie nog normaal, terwijl er al wel andere tekenen van roken gerelateerde veranderingen aanwezig kunnen zijn. Een van deze tekenen is longdestructie. Het opsporen van vroege longdestructie kan patiënten beschermen tegen progressie naar een symptomatisch stadium. De populatie geïncludeerd in de NELSON studie is de populatie die risico loopt om COPD te ontwikkelen, maar zonder ernstig COPD, aangezien zij fit genoeg moeten zijn om chirurgie te ondergaan.
Hoewel voorspeld is dat COPD de derde oorzaak zal zijn van morbiditeit en mortaliteit in de Westerse wereld in 2020 [16], is er weinig bekend over de vroege stadia van de ziekte. Er wordt veel onderzoek gedaan naar de manier waarop luchtweg ontsteking ontstaat, maar er is weinig bekend over de progressie van "at risk" naar luchtweg obstructie. Daarnaast is er weinig bekend over welke mensen COPD zullen ontwikkelen en welke niet. Door voor elke CT-scan van de thorax de hoeveelheid longweefselschade te bepalen, zou emfyseem in een vroeg stadium opgespoord kunnen worden. Wij hebben aangetoond dat veel personen al beschadigd longweefsel hebben, maar een normale longfunctie. De gas uitwisseling kan echter al gestoord zijn bij deze mensen. Gas uitwisseling is

echter (nog) geen speerpunt van COPD volgens de GOLD-richtlijnen.
Er is aangetoond dat de diffusie van stikstofoxide een marker is van vasculaire veranderingen, elke vooraf kunnen gaan aan macroscopische schade aan longweefsel en welke gestoord kunnen zijn voordat emfyseem ontdekt kan worden op een CT-scan.
De belangrijkste beperking van het bestuderen van emfyseem in een vroeg stadium met behulp van lage dosis CT-scans, is het gebrek aan informatie over de betrouwbaarheid van emfyseem scores om daadwerkelijk emfyseem weer te geven. In een screeningsonderzoek waarbij gezonde vrijwilligers worden gescand, is er meestal geen weefsel beschikbaar. Daarom konden wij de resultaten van ons programma niet beoordelen op betrouwbaarheid. Müller e.a. [20] en Gevenois e.a. [17;18] hebben aangetoond dat de hoeveelheid emfyseem gekwantificeerd middels CT -scans een goede correlatie vertoont met de hoeveelheid longschade opgespoord door middel van macroscopische en microscopische technieken, maar wij zijn niet in staat geweest om deze data te onderbouwen voor lage dosis CT-scans. Een studie die emfyseem scores verkregen van lage dosis CT-scans gemaakt in het kader van longkanker screeningsonderzoeken met histologische samples, bijvoorbeeld bij patiënten die een lobectomie ondergaan in verband met een maligne nodule, kan informatie verschaffen in hoeverre emfyseem scores werkelijke longschade weergeven.
In het laatste hoofdstuk, hebben we aangetoond dat de verdeling van emfyseem over de longen van invloed is op de ernst van de luchtwegobstructie. Deelnemers van het NELSON-project met emfyseem dat voornamelijk in de bovenvelden gelokaliseerd is bleken ernstigere luchtwegobstructie te hebben dan deelnemers met eenzelfde hoeveelheid emfyseem, dat voornamelijk in de ondervelden gelokaliseerd was. Longitudinale studies kunnen aantonen in hoeverre de snelheid van progressie van emfyseem afhankelijk is van de verdeling van het emfyseem over de longen. Wanneer personen met een emfyseem dat voornamelijk in de bovenvelden gelokaliseerd is een snellere progressie vertonen, zouden deze mensen een subgroep kunnen vormen, voor wie een meer agressieve risico reducerende aanpak benodigd is. Alleen deze informatie kan voorspellen wat de klinische relevantie is van kleine hoeveelheden emfyseem in asymptomatische personen.

CONCLUSIE EN TOEKOMSTPERSPECTIEF

In dit proefschrift hebben aangetoond dat lage dosis CT-scans mogelijk is en kan worden toegepast voor longkanker screening, maar ook in de dagelijkse praktijk en voor het opsporen van longweefsel schade. We hebben aangetoond dat een toename in beeldruis een beperkende factor is in de opsporing van vroeg emfyseem, dat een laag contrast vertoont met het omliggende weefsel, maar ook dat men met deze beperking kan omgaan door het gebruik van een speciaal ruisreductie filter.

De detectie van groei is het belangrijkste kenmerk om benigne, niet-verkalkte longnodules te onderscheiden van mogelijk maligne nodules, maar de interscan en in mindere mate de interobserver variatie kunnen de detectie van groei belemmeren. Wanneer een nodule echter meer dan 25% toename in volume vertoont, een criterium waaraan de meeste maligne nodules zullen voldoen ten tijde van een herhaalscan na drie maanden, is het met 95% zeker dat de nodule ook echt gegroeid is. Nieuwe segmentatie algoritmen die er vanuit gaan dat een maligne nodule niet sferisch is, kan de interscan variabiliteit terug brengen en daarmee de bovenste limiet van de meetvariatie verlagen. Met deze algoritmen kan de verdubbelingstijden preciezer worden berekend en kan werkelijke groei gemakkelijker worden onderscheiden van meetvariatie.
Dit proefschrift is het begin van een onderzoekslijn naar de opsporing en kwantificatie van gebieden met een abnormaal lage densiteit met behulp van lage dosis CT-scans in asymptomatische personen met een hoog risico in het kader van het NELSON-project. Wij hebben de uitgangsscores beschreven van onze onderzoekspopulatie, maar de resultaten van de vervolgstudie zullen interessanter zijn. Deze dat zullen inzicht geven in het natuurlijk verloop van emfyseem en de invloed van matig en ernstig emfyseem tijdens de uitgangssituatie, de verdeling tijdens de uitgangssituatie en de invloed van de stoppen met roken voor de start van het onderzoek, tijdens het onderzoek en het doorgaan met roken op de progressie van emfyseem.

REFERENCE LIST

(1) Brenner DJ, Doll R, Goodhead DT et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc Natl Acad Sci U S A 2003; 100 (24):13761-13766.
(2) de Gonzalez ABDS. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet 2004; 363(9406):345-351.
(3) Prasad KN, Cole WC, Haase GM. Radiation protection in humans: extending the concept of as low as reasonably achievable (ALARA) from dose to biological damage. Br J Radiol 2004; 77(914):97-99.
(4) Itoh S, Ikeda M, Arahata S et al. Lung cancer screening: minimum tube current required for helical CT. Radiology 2000; 215(1):175-183.
(5) Landis SH, Murray T, Bolden S et al. Cancer statistics, 1999. CA Cancer J Clin 1999; 49(1):8-31.
(6) Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354(9173):99-105.
(7) The International Early Lung Cancer Action Program. Survival of Patients with Stage I Lung Cancer Detected on CT Screening. N Engl J Med 2006; 355(17):1763-1771.
(8) Henschke CI, Naidich DP, Yankelevitz DF et al. Early lung cancer action project: initial findings on repeat screenings. Cancer 2001; 92(1):153-159.
(9) Kaneko M, Eguchi K, Ohmatsu H et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. Radiology 1996; 201(3):798-802.
(10) Sone S, Takashima S, Li F et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet 1998; 351(9111):1242-1245.
(11) Swensen SJ, Jett JR, Hartman TE et al. Lung cancer screening with CT: Mayo Clinic experience. Radiology 2003; 226(3):756-761.
(12) Revel MP, Bissery A, Bienvenu M et al. Are Two-dimensional CT Measurements of Small Noncalcified Pulmonary Nodules Reliable? Radiology 2004; 231(2):453-458.
(13) Wormanns D, Kohl G, Klotz E et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004; 14(1):86-92.
(14) Karabulut N, Toru M, Gelebek V et al. Comparison of low-dose and standard-dose helical CT in the evaluation of pulmonary nodules. Eur Radiol 2002; 12(11):2764-2769.
(15) Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. The Lancet 2003; 362(9387):847-852.
(16) Feenstra TL, van Genugten ML, Hoogenveen RT et al. The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands. Am J Respir Crit Care Med 2001; 164(4):590-596.
(17) Gevenois PA, De M, V, De VP et al. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1995; 152(2):653-657.
(18) Gevenois PA, De Vuyst P, De M, V et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 1996; 154(1):187-192.
(19) Madani A, Zanen J, de Maertelaer V et al. Pulmonary Emphysema: Objective Quantification at Multi-Detector Row CT--Comparison with Macroscopic and Microscopic Morphometry. Radiology 2006;2382042196.
(20) Muller NL, Staples CA, Miller RR et al. "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 1988; 94(4):782-787.
(21) Kemerink GJ, Kruize HH, Lamers RJ et al. CT lung densitometry: dependence of CT number histograms on sample volume and consequences for scan protocol comparability. J Comput Assist Tomogr 1997; 21(6):948-954.
(22) Lamers RJ, Kemerink GJ, Drent M et al. Reproducibility of spirometrically controlled CT lung densitometry in a clinical setting. Eur Respir J 1998; 11(4):942-945.
(23) Kalender WA, Rienmuller R, Seissler W et al. Measurement of pulmonary parenchymal attenuation: use of spirometric gating with quantitative CT. Radiology 1990; 175(1):265-268.

(24) Revel MP, Lefort C, Bissery A et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. Radiology 2004; 231(2):459-466.

DANKWOORD

Hoewel onderzoek vaak met voorspoed en veel tegenslagen gaat, blijft het geweldig om te doen. Dit proefschrift zou nooit tot stand gekomen zijn, als ik niet zoveel hulp van vele lieve collega's en vrienden gehad zou hebben. Daarvoor wil ik iedereen erg bedanken. Uiteraard wil ik een aantal mensen graag met name noemen.

Geachte professor Prokop, beste Mathias. Er zijn weinig hoogleraren zo laagdrempelig en zo goed gehumeurd als jij. Als een dag 48 uur zou duren, zou je 'm nog vol krijgen. Bedankt voor de vrijheid die je me gaf om mijn eigen ideeën uit te voeren.

Geachte professor Lammers, beste Jan-Willem. Altijd goed voor de klinische kijk op de zaken. Met name als het NELSON-project te veel vanuit de wetenschappelijke cijfers in plaats van het menselijke aspect werd bekeken, was u er altijd om daarin te bemiddelen.

Geachte dr. Zanen, beste Pieter. Toen ik begon met onderzoek had ik nog het romantische beeld van filosoferen over hypotheses en data in mijn hoofd. Bij jou kon ik altijd mijn ei kwijt en samen konden we uren nadenken over de artikelen en interpretatie van de data.

Dr.R.J. van Klaveren. Beste Rob, zonder jou had het NELSON-project niet bestaan en dit boekje er niet gelegen. Ik heb bewondering voor je enorme inzet om eindelijk een antwoord te kunnen geven op de vraag of we door middel van het screenen van zware rokers de mortaliteit van longkanker kunnen reduceren. Ook bedankt voor het plaats nemen in de beoordelingscommissie.

Prof. Dr. W.P.Th.M. Mali. Beste Willem, u was niet direct betrokken bij mijn promotie, maar voor het NELSON-studie altijd een steun in het oplossen van alle problemen waar we binnen het project tegen aan liepen.

Lieve Cornelia, thanks for reading the ultra-low dose scans and for all your contributions on my manuscripts. I hope that we can continue to work together in the future.

Beste Ieneke, altijd een luisterend oor bij promotieperikelen! Bedankt voor het scoren van mijn scans. Het was altijd erg gezellig als we weer twee van de weinige Nederlanders waren tijdens congressen over thorax-radiologie. Beste Inge, ook jij bedankt voor het scoren!

Prof. Dr. Matthijs Oudkerk, bedankt voor het mogelijk maken van het NELSON-project en het beoordelen van het manuscript. Prof. J.J. Lagendijk, prof. Dr E.E. Voest en prof. Dr K.G. Moons, bedankt voor het beoordelen van het manuscript.

Anneke Hamersma, je bent geweldig!! Weinig mensen zijn zo efficiënt als jij. Zonder jou hulp had het NELSON-project in Utrecht niet zo goed gelopen. Cees Haring, je was altijd bereid om de zoveelste uitdraai te maken met gemaakte CTs of verslagen.

Saskia, als jij niet een groot deel van de logistiek van me had overgenomen, had ik dit boekje nooit kunnen schrijven.

Een grote multi-center studie betekent veel kleine dingetjes waar je tegen aan loopt. Rozemarijn, ik was heel blij dat ik jou altijd kon bellen en mailen om te overleggen hoe jullie in Groningen met dit soort dingen om gingen. Succes met je opleiding.

Lieve Carola, hoewel ik veel kluswerk heb gedaan, heb jij echt heel veel voor de studie gedaan! Ook jou kon ik altijd bellen voor vragen. Heel veel succes met het afronden van je onderzoek. Ook Karien heel veel succes met je onderzoek. Ik ben benieuwd naar de resultaten. Dear Ying and Dongming. You did a tremendous job reading thousands of CTs for nodules! I wish you good luck while finishing your theses and a good time back in China.
Ernst Scholten, het was altijd fijn om te horen hoe jullie in Haarlem met de radiologische dilemma's omgingen.

Bram en Arnold. Zonder jullie had dit onderzoek nooit uitgevoerd kunnen worden. Bedankt voor het schrijven van de benodigde software. Hopelijk gaat het lukken om voor alle verstorende factoren te corrigeren. Michiel, bedankt voor het opstarten van het schrijven van de software. Jij, Eva, Yulia en Ivana, succes verder met jullie onderzoek.

Ivo, bedankt voor de samenwerking. Jij hebt dit allemaal al doorgemaakt. Succes met de laatste publicaties.

Vanessa en Desiree, bedankt voor alle hulp en het maken en vooral verzetten van vele afspraken. Gabrielle, Linda en Meike, ook jullie bedankt. Karel en Sven, altijd hulp in geval van computernood! Jan, Roy, Eugene en Karin, bedankt voor jullie hulp bij het bewerken van plaatjes en posters.

Angelique, Karin, Michel, Yannick, Bob en alle andere CT laboranten, bedankt voor het scannen van zowel duizenden NELSON-deelnemers als een flink aantal patiënten. Een extra scan kon er bijna altijd wel bij en de patiënten vonden bijna alles goed.
Paul Munnik en alle longfunctielaboranten, jullie hebben geweldig werk gedaan door vele zaterdagen longfunctie af te nemen bij de NELSON-deelnemers. Daniela, Dieuwertje, Karen, Martijn, Menno, Narges, Niek, Sanna, Stephanie, Tanca, bedankt voor het prikken, uitzoeken van stickers en het converteren van CT-scans. Niek, succes met je onderzoek. Waar roeiers niet goed voor zijn…..

Gerard Groenewegen en George Nossent, bedankt voor het aanleveren van de patiënten voor de verschillende studies.

Collega's Annemarieke, Annet, Irene, Everine, Lisa en Nicky, het valt niet te ontkennen, radiologie wordt een vrouwenvak. Annemarieke, ik heb bewondering voor de rust die jij kunt bewaren als ik weer eens gefrustreerd werd van de promotieperikelen. Nicky, heel veel succes met de laatste loodjes. Lisa, ook al zo lekker relaxed. Annet, nog steeds een geweldige collega, nu als assistenten in opleiding. Kamergenoten Audrey en Peter-Jan, het was gezellig. Auke, Stephanie, Sander en Matthijs en alle epirad dames en heren, succes met het afronden van jullie onderzoeken.

Lieve Jeroen, ik heb bewondering voor je manier van onderzoek doen. Ook al was ik het niet altijd met je eens, je maakt me wel aan het lachen.

Bart-Jan, aan jou de nobele taak om de automatische metingen van zowel het emfyseem als de longnodules te verbeteren. Je bent een geweldige opvolger en ik heb er veel vertrouwen in dat we over een paar jaar de betrouwbaarheidsintervallen sterk versmald hebben. Succes!

Schoolvriendinnen Anne, Hester, Maaike, Maartje, Maria en Marieke. Een deel van jullie heeft deze promotiestress al doorgemaakt. Heel bijzonder dat we elkaar nog steeds zien. Ik heb nu al zin in ons volgende weekendje samen. Anne, jij succes met je onderzoek!

Roeivriend(inn)en Helen, Froukje, Barbara, Heleen & Bert, Noor, Margreet, Danielle, Danielle en de dames van Canottaggio. Roeien blijft een geweldige uitlaatklep.

Paranimfen Margreet en Anna, bedankt voor de steun. En Afrika blijft een prachtig continent en leert je in tijden van tegenslag relativeren. Hoezo stress? Pole pole…..

Lieve Vera & Ubbo, Renee & Sander, Anne & Jelmer, Hans & Madeleine en Onno. Ik kan me geen betere schoonfamilie wensen. Het is altijd gezellig en iedereen is altijd welkom.

Lieve papa en mama., jullie hebben me altijd gesteund. Papa, de discussies zijn altijd leuk. En mama, ook jij kunt goed 30 uur per etmaal vullen. Knap hoe jij werk, studie en een gezin kon combineren. Arnoud en Miranda, het is altijd gezellig! Gaan we nog eens samen fietsen?

Lieve Laurens, bedankt voor de lay-out! Super relaxed, jij bent gewoon de liefste…..

CURRICULUM VITAE

Hester Alexandra Gietema was born on November 1st, 1976 in Groningen, The Netherlands. She graduated from secondary school at the Stedelijk Gymnasium Nijmegen in 1994. From 1994 to 1995 she studied medicine in at the University of Leuven. In 1995 she continued her study medicine at the University of Utrecht. In 2002 she graduated for her medical training and became a PhD student at the department of surgery of the VU Medical Center in Amsterdam under supervision of prof. Dr. P.A.M. van Leeuwen. In 2003 she quitted to start as a PhD student at the department of Radiology at the University Medical Center Utrecht under supervision of prof. Dr. W.M. Prokop, prof. Dr. J-W.J. Lammers and Dr. P. Zanen. In June 2006, she started as a resident in Radiology at the Meander Medical Center Amersfoort under supervision of Dr. H.J. Baarslag. In 2009, she will return to the University Medical Center Utrecht to finish her residency.

LIST OF PUBLICATIONS

- Clinical outcome of stage I/II melanoma patients after selective sentinel lymph node dissection: Long-term follow-up results - RJCLM Vuylsteke, PAM van Leeuwen, MG Statius Muller, *HA Gietema,* DR Kragt and S Meijer; *J Clin Oncology 2002*

- Predicting nonsentinel lymph node involvement in stage I/II melanoma, letter to the editor - *HA Gietema*, RJCLM Vuylsteke, PJ van Diest, S Meijer, PAM van Leeuwen; *Annals of surgical oncology 2003*

- Sentinel lymph node investigation in melanoma: detailed analysis of the yield from step section and immunohistochemistry - *HA Gietema*, RJCLM Vuylsteke, IA de Jonge, PAM van Leeuwen, BG Molenkamp, JRM van der Sijp, S Meijer, PJ van Diest; *J Clin Pathology 2004*

- Local administration of Granulocyte/Macrophage Colony-stimulating Factor increases the number and activation state of dendritic cells in the sentinel lymph node of early-stage melanoma - RJCLM Vuylsteke, BG Molenkamp, *HA Gietema*, PAM van Leeuwen, PGJTB Wijnands, W Vos, PJ van Diest, RJ Scheper, S Meijer and TJ de Gruijl; *Cancer research 2004*

- Metastatic tumor load in the sentinel lymph node in stage I/II melanoma: A strong predictor of additional lymph node involvement and survival - RJCLM Vuylsteke, PAM van Leeuwen, PJ Borgstein, *HA Gietema*, MG Statius Muller, PJ van Diest, JRM van der Sijp, S Meijer; *Annals of surgical oncology 2005*

- Local noise weighted filtering for emphysema scoring of low-dose CT images - AM Schilham, B van Ginneken, *HA Gietema*, M Prokop, *IEEE Trans Med Imaging. 2006*

- Pulmonary nodules detected at lung cancer screening: Interobserver variability of semiautomated volume measurements - *HA Gietema*, Y Wang, DM Xu, R van Klaveren, H de Koning, E Scholten, J Verschakelen, G Kohl, M Oudkerk, M Prokop, *Radiology 2006;241:251-257*

- Nodule management protocol of the NELSON randomised lung cancer screening trial - DM Xu, *HA Gietema*, Mathias Prokop, Harry de Koning, René Vernhout, Kristiaan Nackaerts, Carla Weenink, Jan-Willem Lammers, Harry Groen, Matthijs Oudkerk, Rob van Klaveren, *Lung Cancer. 2006 Nov;54(2):177-84*

- Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners - M Das, J Ley-Zaporozhan, *HA Gietema*, A Czech, G Muhlenbruch, AH Mahnken, M Katoh, A Bakai, M Slganicoff, S Diederich, M Prokop, HU Kauczor, RW Gunter, JE Wildberger *Eur Radiol. 2007 Jan 6; (Epub ahead of print)*

- Monitoring of smoking-induced emphysema with multidetector-row computed tomography in a lung cancer screening setting: What is the minimum increase in emphysema scores required to distinguish real increase in extent of emphysema from interscan variability? - *HA Gietema*, AM Schilham, B van Ginneken, R van Klaveren, J-WJ Lammers, M Prokop, *accepted for publication in Radiology*

- Interscan variability of semiautomated volume measurements in pulmonary nodules using multislice computed tomography: Influence of inspirational level, nodule size and segmentation performance - *HA Gietema*, CM Schaefer-Prokop, WPThM Mali, G Groenewegen, M Prokop, *accepted for publication in Radiology*

- Early diagnosis of emphysema: computed tomography versus pulmonary function testing - *HA Gietema*, I van der Lee, P Zanen, B van Ginneken, AM Schilham, RJ van Klaveren,C van Iersel, JMM van den Bosch, M Prokop, J-WJ Lammers, *submitted*

- Impact of massive image noise on the extent of automated quantified emphysema using multidetector-row computed tomography - *HA Gietema*, AM Schilham, B van Ginneken, J-WJ Lammers, M Prokop, *submitted*

- The nitric oxide transfer factor as a tool for the early diagnosis of emphysema - I van der Lee, *HA Gietema*, P Zanen, M Prokop, J-WJ Lammers, JJ van den Bosch, *submitted*

- Distribution of moderate and severe emphysema in heavy smokers joining a population-based lung cancer screening trial: impact on pulmonary function – *HA Gietema*, P Zanen, AM Schilham, B van Ginneken, R van Klaveren, M Prokop, J-WJ Lammers, *submitted*

- Diagnostic information of CT at the radiation dose level of a chest X-ray: A feasibility study - *HA Gietema*, C Schaefer-Prokop, IJC Hartmann, I van den Berk, J-WJ Lammers, M Prokop *In preparation*

ONDERZOEKERS NELSON-PROJECT

Erasmus Medisch Centrum, Rotterdam
Dr. R.J. van Klaveren, longarts
Dr. H.J. de Koning, epidemioloog
Mw. drs. C.A. van Iersel, onderzoeker
Mw. drs. K. van den Bergh, onderzoeker
Drs. R. Vernhout
T. de Jongh

Universitair Medisch Centrum Groningen
Prof. dr. M. Oudkerk, radioloog
Prof. dr. H. Groen, longarts
Prof. dr. W. Timens, patholoog
Mw. drs. D.M. Xu, Radioloog
Drs. Y. Wang, radioloog
Mw. drs. Y.R. Zhao, radioloog
Mw. dr. H.J. van der Zaag-Loonen, epidemioloog
Mw. drs. A.L.M. Leusveld, arts-onderzoeker
Mw. drs. J.C.M. van de Wiel, arts-onderzoeker
Dr. ir. P.M.A. van Ooijen
Dr. P.P. Taverne, radioloog/studie-coördinator

Universitair Ziekenhuis Gasthuisberg, Leuven
Prof. dr. J. Verschakelen, radioloog
Prof. dr. G. Marchal, radioloog
Dr. W de Wever, radioloog
Drs. F. Cheng, radioloog
Prof. dr. K. Nackaerts, longarts
Prof. dr. P. de Leyn, thorax-chirurg
Prof. dr. E. van Limbergen, Leuvens universitair centrum voor kankerpreventie

Kennemer Gasthuis, Haarlem
Drs. E. Th. Scholten, radioloog
Dr. R. Kruyt, radioloog
Drs. L.P. Driessen, radioloog
Dr. W. Monye, radioloog
Mw. dr. C. Weenink, longarts
Drs. H. Willebrandt, patholoog
H. Pruiksma, studiecoördinator

Universitair Medisch Centrum, Utrecht
Prof. dr. W.P.Th.M. Mali, radioloog
Prof. dr. M. Prokop, radioloog
Mw. drs. H.A. Gietema, arts-onderzoeker
Drs. B.J. de Hoop, arts-onderzoeker
Prof. dr. J-W. J. Lammers, longarts
Dr. P. Zanen, epidemioloog/long-fysioloog
Dr. M. van Oosterhout, patholoog
Mw. dr. P. Bakker, thorax-chirurg
Mw. A. Hamersma
Mw. drs. S. van de Vorst

Canisius Wilhelmina Ziekenhuis, Nijmegen
Dr. E. Thunissen, patholoog