Lung nodule assessment in low-dose CT lung cancer screening

Zhao, Yingru

Lung Nodule Assessment in Low-Dose CT Lung Cancer Screening: Validation of Detection and Volumetric Measurement

Yingru Zhao

Lung Nodule Assessment in Low-Dose CT Lung Cancer Screening: Validation of Detection and Volumetric Measurement

Yingru Zhao

Yingru Zhao 赵颖如
Lung nodule assessment in low-dose CT lung cancer Screening:
validation of detection and volumetric measurement
PhD thesis University of Groningen, with a summary in Dutch

Cover design: Yingru Zhao and Xueqian Xie
Layout: Xueqian Xie

The publication of this thesis was financially supported by
University of Groningen, University Medical Center Groningen

Thesis statements

Lung nodule assessment in low-dose CT lung cancer screening: validation of detection and volumetric measurement

1. For indeterminate nodules detected in screening, a short term follow-up after initial CT could exclude a considerable number of benign lesions from further work-up. (This thesis)
2. Using a combination of computed-aided detection and nodule size cut-off in lung cancer screening improves the sensitivity of pulmonary nodule detection, and significantly reduces the false positive rate. (This thesis)
3. The NELSON nodule management regimen has very high negative predictive value for lung cancer in CT lung cancer screening. (This thesis)
4. Using different software packages influences nodule management decisions, especially growth categorization based on consecutive examinations. (This thesis)
5. Further standardization of software for nodule volumetry and volume doubling time assessment is needed to optimize nodule management in lung cancer CT screening. (This thesis)
6. CT features of intermediate-sized nodules cannot sufficiently distinguish between malignant nodules and subsequently resolving nodules. (This thesis)
7. Volumetric three-dimensional measurement is more accurate than two-dimensional evaluation of pulmonary nodules.
8. LungCARE is a very accurate software package for measuring the volume of solid lung nodules.
9. As a fruit needs not only sunshine but cold nights and chilling showers to ripen it, so character needs not only joy but trial and difficulty to mellow it. (H. Black)
10. For a researcher, imagination is more important than knowledge.

Yingru Zhao

August 6, 2013

RIJKSUNIVERSITEIT GRONINGEN

Lung Nodule Assessment in Low-Dose CT Lung Cancer Screening: Validation of Detection and Volumetric Measurement

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
maandag 16 september 2013
om 14:30 uur

door

Yingru Zhao

geboren op 24 januari 1975

te Tianjin, China

Promotor : Prof. dr. M. Oudkerk
Prof. dr. G. H. de Bock

Copromotor : Dr. R. Vliegenthart

Beoordelingscommissie : Prof. dr. J.W.J. Lammers
Prof. dr. W.P.Th.M. Mali
Prof. dr. H.J.M. Groen

ISBN: 978-90-367-6324-0 (book)
978-90-367-6325-7 (e-book)

List of Contents

1

General Introduction

1.1 Lung cancer screening

Lung cancer is a global public health problem of epidemic proportions and remains the leading cause of cancer mortality [1]. Because of the fact that lung cancer is often diagnosed at an advanced stage, in which curation is no longer an option, the long term survival rate of lung cancer is still low. The 5-year survival rate for lung cancer is approximately 16% for all stages combined [2], but increases to 60-75% when patients are diagnosed at an early stage (stage I) [3]. Therefore, a method for early detection in high-risk individuals could potentially reduce mortality from lung cancer. In recent years, low-dose computed tomography (CT) has been proposed as a screening tool. A recent report from the National Lung Screening Trial (NLST) indicates that low-dose CT screening reduces mortality from lung cancer by 20%, compared to chest radiography [4].

The Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON) was launched in 2003. The NELSON trial is a population based multi-centric randomized trial. The purpose is to investigate whether lung cancer screening by low-dose multi-detector CT in high-risk subjects will lead to a decrease of 10-year lung cancer mortality by at least 25% [5]. Included participants are aged 50-75 years and heavy (ex-)smokers. Four rounds of CT screening are performed; at year 1 (baseline round), 2 (the 2nd round), 4 (the 3rd round), and 5.5 (the 4th round). The control arm receives no intervention except for advice on smoking cessation.

1.2 Volume-based evaluation in lung cancer screening

Lung nodules are frequently found in CT screening. Many nodules found in screening have an indeterminate size, meaning that the possibility of being lung cancer cannot be eliminated at first detection. Since these indeterminate lesions are not eligible for immediate histological evaluation, they are often re-examined after an interval. Nodule growth is important in the differentiation between malignancy and benignancy. Although nodule diameter is still the main discriminator in nodule categorization in lung cancer screening trials, the probability that a nodule will demonstrate growth on diameter alone is small [6]. Compared to di-

ameter, change of volume has higher sensitivity and accuracy for the evaluation of nodule growth.

Based on nodule volume, volume doubling time (VDT) was introduced as a crucial diagnostic tool in growth assessment, especially in case of sub-centimeter lesions [7]. The VDTs of most benign pulmonary nodules were found to be more than 450 days, whereas the VDTs of malignant lesions were usually less than 400 days [7]. Of current lung cancer screening trials in Europe and North America, the NELSON study is the first study, in which the probability of malignancy in case of indeterminate-sized nodules is based on volume and VDT. This strategy of screening for lung cancer with the use of volume CT diminished the need for follow-up evaluation in participants with an indeterminate test result. The advantages of volumetric measurements become fully apparent when a volumetric comparison can be made with a previous indeterminate CT scan. It reduced the number of follow-up examinations in participants with an indeterminate test result without reducing the overall sensitivity of the technique. Using volumetric assessment, the NELSON trial generated less false positive cases, compared to the NLST trial based on diameter measurement [8]. That trial recently showed a positive CT in 24.2% of participants, with 96.4% being false positive results. Nowadays, dedicated software is used to assist in evaluation of pulmonary nodules, such as computer-aided detection (CAD) and semi-automated volumetry.

1.3 Computer-aided detection (CAD) system

In lung cancer screening, small pulmonary nodules are extremely common findings, but are uneasy to detect [9-10]. In screening, a radiologist has to read a large number of images in a short time. Fatigue increases the risk of misdiagnosis due to perceptual errors. Meanwhile, although the introduction of low-dose CT in lung cancer screening protocols was found effective in detecting peripheral lung cancers at an early stage [11-13], it may be more difficult for screening radiologists to find lesions in case of increased image noise due to low dose and thin slice thickness. To reduce the number of missed lesions, double reading has been recommended [14]. The utilization of double reading in the setting of lung cancer screening has been debated [15]. Although double reading improves nodule detec-

tion over single reading, double reading is not widely used in clinical routine because of limited human resources and cost-effectiveness [16]. Computer-aided detection (CAD) of pulmonary nodules may help to address that problem by being utilized as an assistant reader [17].

An important application of CAD is to detect lung nodules [17]. For radiologists, various factors affect nodule recognition during screening including reader experience and variability, CT technique and viewing conditions, as well as nodule characteristics [18]. The performance of readers can be influenced by nodule location and its relationship to surrounding anatomical structures [19-20]. For example, in central lung regions, nodules can go undetected because they can be confused with blood vessels imaged in cross sections [21-22]. A lesion not noticed by a reader because of a particular location, may often be detected in retrospective review after being detected on a follow-up scan. In the Mayo Clinic study [23], about one third of nodules were missed at baseline, but were later detected at follow-up screening. Therefore, CAD has been introduced to assist radiologists to improve the efficiency to handle large data sets and the sensitivity of nodule detection.

For more than a decade, considerable effort has been focused on developing automated systems to detect suspicious lesions in thoracic CT images. Numerous studies have demonstrated that the introduction of CAD in radiological practice can significantly improve diagnostic accuracy of pulmonary nodule detection. Using CAD software packages, a wide range in sensitivity, from 54% to 95% has been reported, with 0.55 to 8.3 false positive findings per CT examination [24]. Rubin et al. reported that the mean sensitivity was 50% for an individual reading and 76% after adding CAD [25]. In another study, CAD as second opinion after single reading increased the sensitivity to 79% which proved to be significantly better than double reading [26]. Therefore, despite the high false positive rate, CAD can be valuable to increase sensitivity of nodule detection.

1.4 Semi-automated nodule volumetry

In lung cancer screening, accurate volumetry is essential to assess nodule growth. It is well known that accurate and precise volume measurement depends on a

number of factors, such as CT acquisition protocol, image reconstruction parameters and nodule characteristics [27-29]. In recent years, nodule volume can be semi-automatically measured by using dedicated software. The software algorithm generally relies on complex shape-analysis for nodule segmentation from adjacent structures [30-31]. It has been reported that volumetry software from different vendors had substantial variations in segmentation performance which may cause misdiagnosis [32]. In lung cancer screening, usually multiple centers and CT equipments are involved [6]. Systematic differences in volume measurements between software packages could influence nodule categorization and treatment decisions. Thus, in a setting based on nodule volumetry, the influence of software packages on volume measurement and VDT evaluation needs to be investigated.

1.5 Resolving nodules

With the widespread of thin-slice CT and dedicated software, the number of detected pulmonary nodules has increased enormously, as compared with traditional CT and X-ray radiography. Up to 66% of participants enrolled in screening trials have at least one small-to-medium-sized pulmonary nodule [23]. An important topic in lung cancer screening is the interval of follow-up. The volume doubling time (VDT) of most benign pulmonary nodules were found to be more than 450 days, whereas VDTs of malignant lesions were usually less than 400 days [7]. Currently, in several randomised controlled trails which are underway, the interval of early follow-up imaging is designed as 3-, 6-, or 12-month [33-36].

The majority of indeterminate nodules is not malignant [5], and may represent granulomatous or infectious lesions, or enlarged lymph nodes. These benign nodules are often resolved completely or were reduced in size at short-term follow-up either after therapy with antibiotics or spontaneously [11]. Several studies emphasized the importance of short-term CT follow-up for indeterminate lung nodules. In the study by Felix et al. [37], the mean resolving interval of 32 localized ground glass opacities was 7.6 months. Libby et al. [38] from Early Lung Cancer Action Project (ELCAP) reported that 12% of nodules ≥ 5mm in diameter had complete resolution within 2 months of the initial CT in baseline screening both with and without having received antibiotics. Therefore, for indeterminate nod-

ules detected in screening, a short term follow-up after initial CT could exclude a great amount of benign lesions from malignancy.

1.6 Outline thesis

The issues of detection, volumetric measurement and characterization of pulmonary nodules were investigated in this thesis. In **Chapter 2**, we describe the details regarding participant recruitment, CT acquisition and nodule management protocol in the NELSON trial. In **Chapter 3**, we investigate the results of low-dose CT lung cancer screening from the baseline and second round. Subsequently, in **Chapter 4 and 5**, we investigate the implementation of CAD software, and volumetry variability among different software packages. Finally, in **Chapter 6**, we retrospectively studied the nodule features associated with complete disappearance of solid pulmonary nodules.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. Dec 15 2010;127(12):2893-2917.
2. Society AC. Cancer facts and figures 2007. ACS. 2007.
3. Silvestri GA, Alberg AJ, Ravenel J. The changing epidemiology of lung cancer with a focus on screening. BMJ. 2009;339:b3053.
4. Aberle DR, Adams AM, Berg CD. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395-409.
5. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. N Engl J Med. Dec 3 2009;361(23):2221-2229.
6. Nair A, Hansell DM. European and North American lung cancer screening experience and implications for pulmonary nodule management. Eur Radiol. Dec 2011;21(12):2445-2454.
7. Hasegawa M, Sone S, Takashima S, et al. Growth rate of small lung cancers detected on mass CT screening. Br J Radiol. Dec 2000;73(876):1252-1259.
8. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. Aug 2011;365(5):395-409.
9. Xu DM, van Klaveren RJ, de Bock GH. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. . Eur J Radiol. 2008;68:347-352.
10. Wang Y, de Bock GH, van Klaveren RJ. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. European Radiology. 2010;20:1180-1187.
11. Diederich S, Wormanns D, Semik M, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. Radiology. Mar 2002;222(3):773-781.
12. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet. Jul 10 1999;354(9173):99-105.

13. Nawa T, Nakagawa T, Kusano S, Kawasaki Y, Sugawara Y, Nakata H. Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies. Chest. Jul 2002;122(1):15-20.
14. Wormanns D, Ludwig K, Beyer F, Heindel W, Diederich S. Detection of pulmonary nodules at multirow-detector CT: effectiveness of double reading to improve sensitivity at standard-dose and low-dose chest CT. Eur Radiol. Jan 2005;15(1):14-22.
15. Wang Y, Van Klaveren R, de Bock GH, et al. No benefit for consensus double reading at baseline screening for lung cancer with the use of semiautomated volumetry software Radiology (Epub ahead of print). 2011.
16. Wang Y, van Klaveren RJ, de Bock GH, et al. No benefit for consensus double reading at baseline screening for lung cancer with the use of semiautomated volumetry software. Radiology. Jan 2012;262(1):320-326.
17. Hirose T, Nitta N, Shiraishi J, Nagatani Y, Takahashi M, Murata K. Evaluation of Computer-aided Diagnosis (CAD) Software for the Detection of Lung Nodules on Multidetector Row Computed Tomography (MDCT): JAFROC Study for the Improvement in Radiologists' Diagnostic Accuracy. Academic Radiology. Dec 2008;15(12):1505-1512.
18. Goo JM, Lee JW. Automated lung nodule detection at low-dose CT: preliminary experience. Korean J Radiol. 2003;4:211-216.
19. Naidich DP, Rusinek H, McGuinness G, Leitman B, McCauley DI. Variables affecting pulmonary nodule detection with computed tomography: evaluation with three-dimensional computer simulation. . J Thorac Imaging. 1993;8:291-299.
20. Rusinek H, Naidich DP, McGuinness G, Leitman BS, McCauley DI. Pulmonary nodule detection: low-dose versus conventional CT. Radiology. 1998;209:243-249.
21. Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W. Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. Eur Radiol. May 2002;12(5):1052-1057.

22. Armato SG, 3rd, Li F, Giger ML, MacMahon H, Sone S, Doi K. Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program. Radiology. Dec 2002;225(3):685-692.
23. Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. Am J Respir Crit Care Med. Feb 15 2002;165(4):508-513.
24. Goo JM. A computer-aided diagnosis for evaluating lung nodules on chest CT: the current status and perspective. Korean J Radiol. Mar 2011;12(2):145-155.
25. Rubin GD, Lyo JK, Paik DS, et al. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology. Jan 2005;234(1):274-283.
26. Wormanns D, Beyer F, Diederich S, Ludwig K, Heindel W. Diagnostic performance of a commercially available computer-aided diagnosis system for automatic detection of pulmonary nodules: comparison with single and double reading. Rofo. Jul 2004;176(7):953-958.
27. Goo JM, Kim KG, Gierada DS, Castro M, Bae KT. Volumetric measurements of lung nodules with multi-detector row CT: effect of changes in lung volume. Korean J Radiol. Oct-Dec 2006;7(4):243-248.
28. Larici AR, Storto ML, Torge M, et al. Automated volumetry of pulmonary nodules on multidetector CT: influence of slice thickness, reconstruction algorithm and tube current. Preliminary results. Radiol Med. Feb 2008;113(1):29-42.
29. Gavrielides MA, Kinnard LM, Myers KJ, Petrick N. Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology. Apr 2009;251(1):26-37.
30. Prionas ND, Ray S, Boone JM. Volume assessment accuracy in computed tomography: a phantom study. J. Appl. Clin. Med. Phys. 2010;11(2):168-180.
31. Tao P, Griess F, Lvov Y, et al. Characterization of small nodules by automatic segmentation of X-ray computed tomography images. J Comput Assist Tomogr. May-Jun 2004;28(3):372-377.
32. de Hoop B, Gietema H, van Ginneken B, Zanen P, Groenewegen G, Prokop M. A comparison of six software packages for evaluation of solid lung nodules us-

ing semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. Eur Radiol. Apr 2009;19(4):800-808.

33. Aberle DR, Berg CD, Black WC. The National Lung Screening Trial: overview and study design. Radiology. 2011;258(1):243-253.
34. Baldwin DR, Duffy SW, Wald NJ, Page R, Hansell DM, Field JK. UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. Thorax. 2011;66(4):308-313.
35. Lopes PA, Picozzi G, Mascalchi M. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. Lung Cancer. 2009;64:34-40.
36. Pedersen JH, Ashraf H, Dirksen A. The Danish randomized lung cancer CT screening trial--overall design and results of the prevalence round. J Thorac Oncol. 2009;4(5):608-614.
37. Felix L, Serra-Tosio G, Lantuejoul S. CT characteristics of resolving ground-glass opacities in a lung cancer screening programme. Eur J Radiol. 2011;77(3):410-416.
38. Libby D, Wu N, Lee IJ, et al. CT Screening for Lung Cancer: The Value of Short-term CT Follow-up. Chest. 2006;129:1039-1042.

2

NELSON Lung Cancer Screening Study

Cancer Imaging 2011;11:S79-84

Yingru Zhao[12] / Xueqian Xie[12]
Rozemarijn Vliegenthart[12] / Matthijs Oudkerk[12]

University of Groningen, University Medical Center Groningen,
[1]Department of Radiology, [2]Center for Medical Imaging – North East Netherlands

Abstract

The Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON study) was designed to investigate whether screening for lung cancer by low-dose multi-detector computed tomography (CT) in high risk subjects will lead to a decrease in 10-year lung cancer mortality of at least 25% compared to a control group without screening. Since the start of the NELSON study in 2003, 7557 participants underwent CT screening, with scan rounds in year 1, 2, 4 and 6. In the current review, the design of the NELSON study including participant selection and lung nodule management protocol, as well as results on validation of CT screening and first results on lung cancer screening are described.

Introduction

Lung cancer is the most common cause of cancer-related death in the world [1]. At the time of diagnosis, lung cancer is often already in an advanced stage, with 5-year-survival of only 15% or less [2]. Observational studies in high-risk populations have shown that spiral computed tomography (CT) screening detects more lung cancers than chest X-ray screening [3,4], with 55–85% of CT-detected lung cancers being at a surgically removable stage (stage I). However, observational studies are prone to lead-time, length-time and over-diagnosis bias. Randomized studies are needed to compare disease-specific mortality between a screened and an unscreened population. This was the reason to launch the Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON study) in September 2003. The hypothesis of the NELSON study is that lung cancer screening by low-dose spiral CT will reduce 10-year lung cancer mortality by 25% in high risk (ex-)smokers between 50 and 75 years of age.

NELSON study trial design

Participant selection and recruitment

During the first recruitment phase, men aged 50-75 years from 7 districts in the Netherlands and men and women from 14 municipalities around Leuven in Belgium were sent a questionnaire about health, smoking, cancer history, and other lifestyle and health factors. Based on the smoking history, the estimated lung cancer mortality risk of the respondents was determined. Next, the required sample size including required participation rate was determined. Included were current smokers and former smokers with 10 years or less of cessation, who smoked more than 15 cigarettes daily for over 25 years or more than 10 cigarettes daily for over 30 years. Exclusion criteria were a moderate or bad self-reported health, inability to climb 2 flights of stairs, body weight ≥ 140 kg, lung cancer less than 5 years ago or still under treatment, current or past renal cancer, melanoma or breast cancer, and chest CT less than 1 year. The aim was to include 16,000 participants, half in the screen arm and half in the control arm. The trial was approved by the Dutch

Minister of Health and the ethics board at each participating center. All participants gave written informed consent. For more details on participant selection and recruitment as well as numbers concerning response rates, see van Iersel et al [5].

To conduct this logistically complex multi-centre study, the NELSON management system was developed. This is a web-based interactive database application for data collection and management of all study related processes such as the selection and randomisation of participants, electronic storage of forms, study monitoring, reporting of scan results and scheduling of appointments for follow-up scans.

The remainder of this review concerns the screen arm of the study. Participants randomised to the screen arm were invited to one of the four screening sites (University Medical Center Groningen, University Medical Center Utrecht and Kennemer Gasthuis Haarlem in the Netherlands, and University Hospital Gasthuisberg Leuven in Belgium). Screening rounds took place in year 1, 2, 4, and 6. On one day, participants underwent CT (see below), and, depending on the screening round, blood sampling and pulmonary function testing. Pulmonary function tests included forced expiratory volume in 1 s (FEV_1) and forced vital capacity (FVC) with a pneumotachograph. Participants received a quality of life questionnaire after each visit to the screening site.

CT scan protocol

For chest CT scanning, 16-detector CT scanners (Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA, or Sensation-16, Siemens Medical Solutions, Forchheim, Germany) were used. Scans took about 12 s in spiral mode with 16 × 0.75 mm collimation and 15 mm table feed per rotation (pitch = 1.3). Scans were obtained in a cranio-caudal scan direction, without contrast media, in low-dose setting. Depending on the body weight (<50, 50–80 and >80 kg) the kVp settings were 80–90, 120 and 140 kVp, respectively. To achieve a $CTDI_{vol}$ of 0.8, 1.6 and 3.2 mGy, respectively, the mAs settings were adjusted accordingly dependent on the machine used. To minimise breathing artefacts, scans were performed in inspiration after appropriate instruction of the participants. Data acquisition and

scanning conditions were standard across screening sites and were the same for all rounds of the screening [6]. Data sets were derived from images of the lung with a thickness of 1 mm, reconstructed at overlapping 0.7-mm intervals. Isotropic data sets allowed for volume measurements with good reproducibility, even in case of small lesions [7].

Table 1 Nodule categorization based on size and density (new nodules) and growth rate (existing nodules) in the NELSON trial

Category	Definition		
NODCAT 1	A benign nodule (with fat/benign calcifications) or other benign abnormalities		
NODCAT 2	A nodule, smaller than NODCAT3, not belonging to NODCAT1		
	Solid	Partial solid	Non-solid
NODCAT 3	$50 \leq V \leq 500$ mm^3 Pleural based: $5 \leq d_{min} \leq 10$ mm	Solid component: $50 \leq V \leq 500$ mm^3 Non-solid component: $d_{mean} \geq 8$ mm	$d_{mean} \geq 8$ mm
NODCAT 4	$V > 500$ mm^3 Pleural based: $d_{min} > 10$ mm	Solid component: $V > 500$ mm^3	(non-existing category)
GROWCAT A	VDT > 600 days		
GROWCAT B	400 ≤ VDT ≤ 600 days		
GROWCAT C	VDT < 400 days, or new solid component in non-solid lesion		

V = volume; d_{min} = minimal diameter; d_{mean} = mean diameter; VDT = volume doubling time.

CT reading protocol

Images were read on Siemens Leonardo workstations using the Syngo LungCARE software package (Version Somaris/5 VB 10A-W) for semi-automated volume measurements. Images were interpreted both at lung window and mediastinal settings. The first reading was performed by a reader with experience in reading chest CT scans varying from none to more than 20 years. In case of inappropriate segmentation (i.e., nodules that were attached to a fissure or to a vessel), the reader was allowed to enter manual measurements, which overruled the automatically generated volumes. Baseline and follow-up images were reviewed and displayed simultaneously on one workstation. Data generated by the LungCARE

software were uploaded into the NELSON Management System, which automatically detected whether a nodule was new or had been present previously and which calculated the percentage change in volume and the volume-doubling time in days. Second readings were done by two radiologists with 6 years of experience. The 2nd readers were unaware of the conclusion of the first reader. In case of discrepancy, a third reader made the final decision. More details on the method of evaluation of lung nodules can be found in Gietema et al [7].

Lung nodule definitions and management

A nodule was further evaluated if it did not meet criteria for benign lesions. Nodule volume was obtained semi-automated by LungCARE software; for certain nodules like pleural-based nodules, measurement of diameters from a point perpendicular to the costal pleura was performed manually. In addition to size, nodule characteristics like shape and surface were noted. Growth was defined as change in volume of at least 25% between scans of subsequent scans, based on validation studies with repeated low-dose CT on the same day, in which the measurement error was maximally 25% [7,8]. The volume-doubling time was calculated as described previously [6]. Growing nodules were classified into three growth categories according to their volume-doubling time. The definitions of the different categories of lung nodules are shown in Table 1.

Based on the highest nodule category found, management was determined. Table 2 provides an overview of nodule management for the baseline and incidence scans. NODCAT 3 was defined as an indeterminate test result which required a repeat scan 3 to 4 months later to assess growth. 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. For new nodules, the same classification according to size was made as for the baseline screening round. Follow-up was at shorter interval, however, because at incidence screen new nodules are supposed to have a relatively higher growth rate.

Table 2 NELSON management protocol for non-calcified pulmonary nodules in the different screening rounds

	Year 1	Year 2	Year 4	Year 6
NODCAT 1	Negative test Annual CT	Negative test CT in year 4	Negative test CT in year 6	Negative test End of screening
NODCAT 2	Negative test Annual CT	Indeterminate test CT after 1 year	Indeterminate test CT after 1 year	Indeterminate test End of screening
NODCAT 3	Indeterminate test 3 months follow-up CT	Indeterminate test CT after 6-8 weeks	Indeterminate test CT after 6-8 weeks	Indeterminate test CT after 6-8 weeks
NODCAT 4	Positive test Refer to pulmonologist for work-up and diagnosis	Positive test Refer to pulmonologist for work-up and diagnosis	Positive test Refer to pulmonologist for work-up and diagnosis	Positive test Refer to pulmonologist for work-up and diagnosis
GROWCAT A	Negative test CT in year 2	Negative test CT in year 4	Negative test CT in year 6	Negative test End of screening
GROWCAT B	Negative test CT in year 2	Indeterminate test CT after 1 year	Indeterminate test CT after 1 year	Indeterminate test CT after 1 year
GROWCAT C	Positive test Refer to pulmonologist for work-up and diagnosis	Positive test Refer to pulmonologist for work-up and diagnosis	Positive test Refer to pulmonologist for work-up and diagnosis	Positive test Refer to pulmonologist for work-up and diagnosis

If the highest category was a NODCAT 4 of GROWCAT C, the participant was referred to a chest physician 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). If malignancy was proven, staging was performed, followed by surgical resection of the nodule. The work-up for participants with GROWCAT C was essentially the same as for NODCAT 4, except that for the former nodules a final histological diagnosis had to be obtained either by FNA, video-assisted thoracoscopic surgery, or wedge resection

and examination on frozen section. The workup, staging, and treatment were standard across all screening sites and were performed according to published guidelines.

CT nodule evaluation results in the NELSON study

Since the start of the NELSON study, numerous studies on CT nodule evaluation in the trial have been published. The results of a number of these will be shortly mentioned. The studies can be separated into variability studies and studies on nodule characteristics suggestive of malignancy / benignancy.

An already mentioned study [7] investigated the inter-observer variability of semi-automated volume measurements of small-to-intermediate size lung nodules (NODCAT 2 and 3). Inter-observer correlation was very high (r = 0.99). Nearly 90% of nodules did not show any variation in volume with double reading. In only 3.7% there was a volume difference of > 10%, mostly due to incomplete segmentation due to irregular shape or margins. The variability of volume measurements was, in a further study [9], found to be related to nodule morphology, location and size. Volume disagreement was most likely in case of juxtavascular and irregular nodules. In a third study [10], semi-automated nodule volumes were compared for CT data sets reconstructed with different settings of section thickness and kernel. The repeatability coefficients were found to differ according to setting, depending also on nodule location and morphology. The volume measurement was most repeatable for 1 mm section thickness with soft kernel. In case of serial CT studies, consistent reconstruction parameters were concluded to be essential.

Valuable knowledge about characteristics of lung nodules associated with cancer risk was obtained in the NELSON study [11-13]. Solid nodules at intermediate size (NODCAT 3) were evaluated by CT at 3 months and at 1 year after baseline. Cancers were found to be non-spherical and purely intra-parenchymal, without attachment to vessels, the pleura, or fissures [11]. In non-smooth nodules without attachment, the only predictor of malignancy was size. The results suggest that the risk of malignancy in smooth or attached solid nodules at intermedi-

ate size is extremely low. In a study of the intermediate-to-large size nodules (NODCAT 3 and 4) [12], especially size and to a lesser extent irregular shape and margin were found to increase the likelihood of malignancy. Baseline CT density of the lung nodule was not predictive of malignancy. However, an increase in CT density was suggestive of malignancy [13] in intermediate size nodules (NODCAT 3). Furthermore, the majority of both benign and malignant nodules did not change of characteristics during 1 year of follow-up.

Screening results from the NELSON study

In 2009, the NELSON screening results from the first and second screening round were published in the New England Journal of Medicine [14]. The mean age of the population was 59 years, mean number of pack-years was 42. Of the 7557 participants who underwent CT screening, 1.6% (119) had a positive baseline scan. In addition, of the 19.2% (1451) of participants that had one or more intermediate size nodules (NODCAT 3) and thus an indeterminate test result, 5.3% showed a growing nodule suspected for malignancy (GROWCAT C) on the 3-month follow-up scan. Figure 1 is an example of a GROWCAT C nodule that was proven to be lung cancer. Combined, 2.6% (196) had a positive test result. Seventy were found to have lung cancer, with benign disease or other cancer in 107. Lung cancer detection rate was 0.9%. Sensitivity of the first screening round was 94.6%, negative predictive value 99.7%. There were only 3 interval cancers between the first and second screening round. At the second screening round, 1.8% (128) of participants had a positive result, with 54 turning out to have lung cancer. Sensitivity of the second screening round was 96.4%, with a negative predictive value of 99.9%.

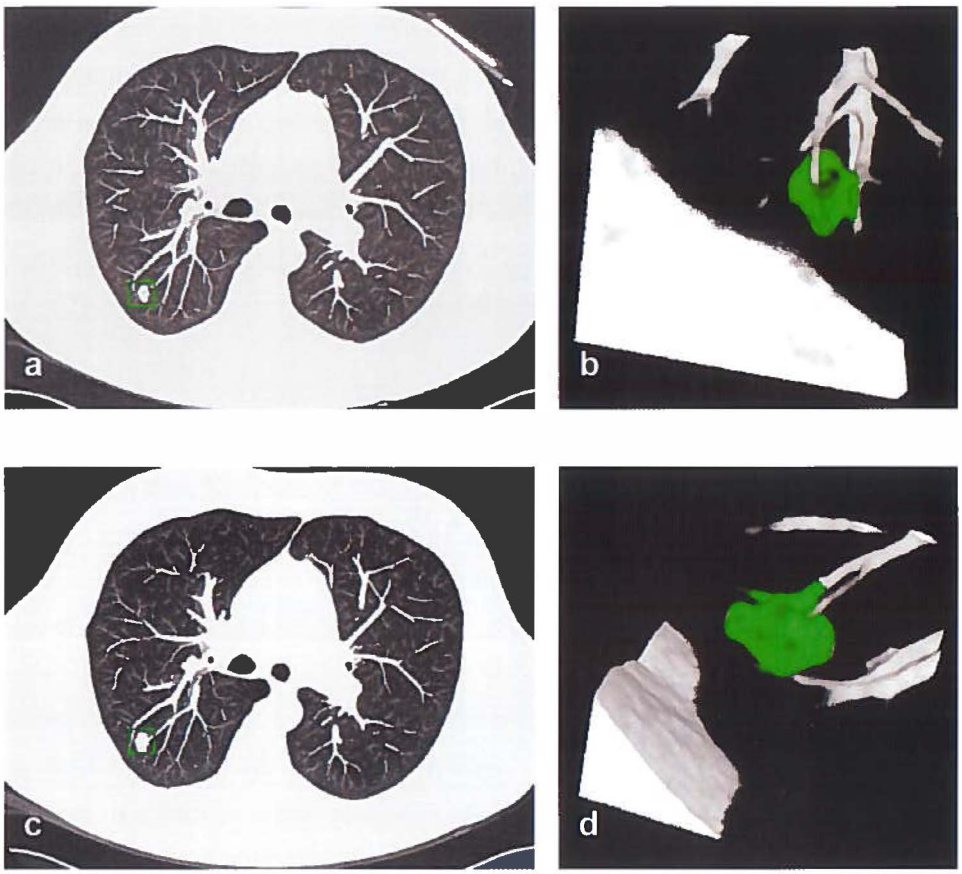


Figure 1 Baseline and 3-month follow-up CT scan images in a 68-year old participant of the NELSON study. Transverse thin-section CT (**a**, **c**) and volume-rendered reconstruction (**b**, **d**) images show a lobulated pulmonary nodule with vessel attachment (boxed on **a**, **c** and green area in **b**, **d**). On the baseline scan (**a**, **b**) the volume was 303 mm^3. On the 3-month follow-up CT (**c**, **d**), the volume was 576 mm^3. This is consistent with a percentage volume growth of 90% and a volume-doubling time of 98 days. Histopathology of the resected nodule: squamous cell carcinoma.

Screening results from the NELSON study

In 2009, the NELSON screening results from the first and second screening round were published in the New England Journal of Medicine [14]. The mean age of the population was 59 years, mean number of pack-years was 42. Of the 7557 partici-

pants who underwent CT screening, 1.6% (119) had a positive baseline scan. In addition, of the 19.2% (1451) of participants that had one or more intermediate size nodules (NODCAT 3) and thus an indeterminate test result, 5.3% showed a growing nodule suspected for malignancy (GROWCAT C) on the 3-month follow-up scan. Figure 1 is an example of a GROWCAT C nodule that was proven to be lung cancer. Combined, 2.6% (196) had a positive test result. Seventy were found to have lung cancer, with benign disease or other cancer in 107. Lung cancer detection rate was 0.9%. Sensitivity of the first screening round was 94.6%, negative predictive value 99.7%. There were only 3 interval cancers between the first and second screening round. At the second screening round, 1.8% (128) of participants had a positive result, with 54 turning out to have lung cancer. Sensitivity of the second screening round was 96.4%, with a negative predictive value of 99.9%.

Conclusion

The first results of the NELSON study show the value of 3D-based lung nodule management for CT lung cancer screening, with an extremely high negative predictive value. The NELSON study has several features that distinguish this trial from f.e. the National Lung Cancer Screening trial [15]. First of all, the nodules detected at baseline and new nodules detected at incidence screening were classified and managed according to volume. At (annual) repeat CT scanning, the first assessment is whether there is growth or not, and if so, a nodule is subsequently classified in one of three growth categories based on VDT. NELSON is the first large lung cancer screening trial in which semi-automated, volumetric nodule assessment is routinely applied and forms an integral part of the nodule management protocol. Volumetric, 3D measurements have been found to be more accurate than 2D evaluation of pulmonary nodules [16,17]. This was confirmed in our study by an extremely low rate of interval cancers.

Another major difference is the differentiated manner with which lung nodules were managed, according to size and density. Although in a screening setting, the sensitivity has to be very high, the specificity has to be high enough to limit the number of false positives. A high false positive rate leads to unnecessary anxi-

ety, costs and morbidity. The NLST screening trial recently showed a positive CT in 24.2%, with 96.4% being false positive results. By adding 3 to 4 month follow-up CT in the NELSON study for nodules of intermediate size, the number of false positive findings could be greatly reduced as many intermediate nodules were found to have resolved or have a non-malignant growth pattern. In the NELSON study, only 2.6% of the participants had a positive baseline screening result, with a false positive rate of 64.3%.

Follow-up of the NELSON study population is ongoing. Within 3 years, the 10-year mortality results are expected, which will provide solid evidence whether lung cancer screening in high risk subjects by low dose CT does decrease lung cancer mortality compared to no screening.

References

1. Shibuya K, Mathers CD, Boschi-Pinto C, Lopez AD, Murray CJL. Global and regional estimates of cancer mortality and incidence by site: II. Results for the global burden of disease 2000. BMC Cancer 2002; 2: 37.
2. Janssen-Heijnen ML, Coebergh JW. Trends in incidence and prognosis of the histological subtypes of lung cancer in North America, Australia, New Zealand and Europe. Lung Cancer 2001; 31: 123-137.
3. Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999; 354: 99-105.
4. Kaneko M, Eguchi K, Ohmatsu H, Kakinuma R, Naruke T, Suemasu K, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. Radiology 1996; 201: 798-802.
5. van Iersel CA, de Koning HJ, Draisma G, Mali WPTM, Scholten ET, Nackaerts K, 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; 120: 868-874.
6. Xu DM, Gietema H, de Koning H, Vernhout R, Nackaerts K, Prokop M, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 2006; 54: 177-184.
7. Gietema HA, Wang Y, Xu DM, van Klaveren RJ, de Koning H, Scholten E, et al. Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. Radiology 2006; 241: 251-257.
8. Wormanns D, Kohl G, Klotz E, Marheine A, Beyer F, Heindel W, et al. Volumetric measurement of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004; 14: 86-92.
9. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, de Bock GH, Gietema HA, Xu DM, et al. Effect of nodule characteristics on variability of semiautomated

volume measurements in pulmonary nodules detected in a lung cancer screening program. Radiology 2008; 248: 625-631.

10. Wang Y, de Bock GH, van Klaveren RJ, van Ooyen P, Tukker W, Zhao Y, et al. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. Eur Radiol 2010; 20: 1180-1187.
11. Xu Dm, van der Zaag-Loonen HJ, Oudkerk M, Wang Y, Vliegenthart R, Scholten ET, et al. Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. Radiology 2009; 250: 264-272.
12. Xu DM, van Klaveren RJ, de Bock GH, Leusveld A, Zhao Y, Wang Y, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. Eur J Radiol 2008; 68: 347-352.
13. Xu DM, van Klaveren RJ, de Bock GH, Leusveld ALM, Dorrius MD, Zhao Y, et al. Role of baseline nodule density and changes in density and nodule features in the discrimination between benign and malignant solid indeterminate pulmonary nodules. Eur J Radiol 2009; 70: 492-498.
14. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. New Engl J Med 2009; 361: 2221-2229.
15. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. New Engl J Med 2011; doi: 10.1056/NEJMoa1102873.
16. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? Radiology 2004; 231:453-458.
17. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 2000; 217:251-256.

3

Management of Lung Nodules Detected by Volume CT Scanning

New England Journal of Medicine 2009; 361:2221-2229

Rob J. van Klaveren[1] / Matthijs Oudkerk[2] / Mathias Prokop[3] / Ernst T. Scholten[4] Kristiaan Nackaerts[5] / Rene Vernhout[1] / Carola A. van Iersel[1,6] Karien A.M. van den Bergh[6] / Susan van 't Westeinde[1] Carlijn van der Aalst[1] Erik Thunnissen[7] / Dongming Xu[2] / Ying Wang[2] / Yingru Zhao[2] / Hester A. Gietema[3] Bart-Jan de Hoop[3] / Harry J.M. Groen[8] Geertruida H. de Bock[9] / Peter van Ooijen[2] Carla Weenink[10] / Johny Verschakelen[11] / Jan-Willem J. Lammers[12] / Wim Timens[13] Dik Willebrand[14] / Aryan Vink[15] / Willem Mali[3] / Harry J. de Koning[6]

[1]Department of Pulmonology, [6]Department of Public Health, Erasmus Medical Center; University of Groningen, University Medical Center Groningen, [2]Department of Radiology, [8]Department of Pulmonology Diseases, [9]Department of Epidemiology, [13]Department of Pathology; [3]Department of Radiology, [12]Department of Pulmonology Diseases, [15]Department of Pathology, University Medical Center Utrecht; [4]Department of Radiology, [10]Department of Pulmonology Diseases, [14]Department of Pathology, Kennemer Gasthuis; [5]Departmens of Pulmonology Diseases, [11]Departmens of Radiology, University Hospital Gasthuisberg Leuven; [7]Department of Pathology, Free University Medical Center Amsterdam

Abstract

Background

The use of multidetector computed tomography (CT) in lung-cancer screening trials involving subjects with an increased risk of lung cancer has highlighted the problem for the clinician of deciding on the best course of action when noncalcified pulmonary nodules are detected by CT.

Methods

A total of 7557 participants underwent CT screening in years 1, 2, and 4 of a randomized trial of lung-cancer screening. We used software to evaluate a noncalcified nodule according to its volume or volume-doubling time. Growth was defined as an increase in volume of at least 25% between two scans. The first-round screening test was considered to be negative if the volume of a nodule was less than 50 mm^3, if it was 50 to 500 mm^3 but had not grown by the time of the 3-month follow-up CT, or if, in the case of those that had grown, the volume-doubling time was 400 days or more.

Results

In the first and second rounds of screening, 2.6% and 1.8% of the participants, respectively, had a positive test result. In round one, the sensitivity of the screen was 94.6% (95% confidence interval [CI], 86.5 to 98.0) and the negative predictive value 99.9% (95% CI, 99.9 to 100.0). In the 7361 subjects with a negative screening result in round one, 20 lung cancers were detected after 2 years of follow-up.

Conclusions

Among subjects at high risk for lung cancer who were screened in three rounds of CT scanning and in whom noncalcified pulmonary nodules were evaluated according to volume and volume-doubling time, the chances of finding lung cancer 1 and 2 years after a negative first-round test were 1 in 1000 and 3 in 1000, respectively.

Introduction

The use of multi-detector computed tomography (CT) has increased the chance of finding non-calcified pulmonary nodules [1,2], and as a result, clinicians often face the problem of deciding on the best course of action with respect to such nodules when they are found in asymptomatic subjects who have an increased risk for lung cancer [3]. This difficulty is especially evident in CT-based screening programs for lung cancer. The current practice is to refer participants in these programs for additional diagnostic evaluation if they have a non-calcified nodule that is larger than 5 mm in diameter [4-9]. In designing the Dutch-Belgian randomized lung cancer screening trial (Nederlands-Leuvens Longkanker Screenings Onderzoek [NELSON]), we adopted a strategy that was meant to provide an inexpensive and simple follow-up process without increasing the false negative rate of the screening test [10]. The strategy entailed the use of the volume and volume-doubling time of a non-calcified nodule as the main criteria for deciding on further action. In this article, we report an evaluation of this strategy, which involved the tracking of individual nodules and the collection of 2-year follow-up data from the screened population of the NELSON trial.

Methods

Participants

We randomly assigned eligible participants in NELSON, who were recruited as described previously [11], to undergo CT screening at baseline (first round), 1 year later (second round), and 3 years later (third round, 2 years after the second round), or no screening. The purpose of the trial is to determine whether at 10 years after randomization CT screening will have reduced mortality from lung cancer by at least 25%. The trial was approved by the Dutch Minister of Health and the ethics board at each participating center. All participants gave written informed consent.

Screening Strategy

A 16-detector CT scanner (Somatom Sensation 16, Siemens Medical Solutions or, at the screening site in Utrecht, Mx8000 IDT or Brilliance-16P, Philips Medical Systems) was used at each of the screening sites. Data sets were derived from images of the lung with a thickness of 1 mm that were reconstructed at overlapping 0.7-mm intervals. Isotropic data sets allowed for volume measurements with good reproducibility, even in the case of small lesions. Data acquisition and scanning conditions were standard across screening sites and were the same for all rounds of screening [10]. At each site, CT data were analyzed on one type of digital workstation (Leonardo, Siemens Medical Solutions) with the use of software for semi-automated volume measurements (LungCare, version Somaris/5 VA70C-W, Siemens Medical Solutions) [13,14]. In the case of inappropriate segmentation (i.e., nodules that were attached to a fissure or to a vessel), the radiologist was allowed to enter manual measurements, which overruled the automatically generated volumes. Data generated by the LungCare software were uploaded into the NELSON Management System, which automatically detected whether a nodule was new or had been present previously and which calculated the percentage change in volume and the volume-doubling time in days (Supplemental Figure A).

A nodule was classified as non-calcified if it did not meet previously specified criteria for a benign lesion [4]. For solid pleural-based and non-solid pulmonary nodules, the diameter was determined manually, and the volume-doubling time was calculated as described previously [10] (Supplemental Figure A). In the case of pleural-based nodules, the diameter was measured at a point perpendicular to the costal pleura. In the case of partially solid lesions, only the volume of the solid region was used. The diameter was defined as the average of the maximum length and width of the nodule. Growth was defined as a change in volume of at least 25% between the first and second scans or between the second and third scans. The 25% threshold was based on three zero-change data sets in which the variation in volume of individual nodules was assessed between two low-dose CT scans. After the first of these scans, the patient returned to the examining table for the second scan to simulate the condition of a repeat examination for the follow-up of a pulmonary nodule. In these studies, the volume measurement error

varied between 20% and 25% [12,14,15] Growing nodules were classified into three growth categories according to their volume-doubling time (<400, 400 to 600, and >600 days).

CT scans were independently read by first and second readers. The experience of the 13 first readers ranged from none to more than 20 years of experience reading thoracic CT scans (median, 6 years); both second readers had 6 years of experience. The second readers matched the nodules they had identified with nodules identified by the first readers according to location and size and compared their results with those of the first readers. If the results were discrepant, the readers re-evaluated the scan to reach a consensus. If no consensus was reached, a third radiologist arbitrated the results.

First-Round (Baseline) Scan

A test was considered to be positive if on the CT scan any non-calcified nodule had a solid component that was more than 500 mm^3 (> 9.8 mm in diameter) and was considered to be indeterminate if the volume of the largest solid nodule or of the solid component of a partially solid nodule was 50 to 500 mm^3 (4.6 to 9.8 mm in diameter) or if the diameter of a nonsolid nodule was greater than 8 mm [10]. In subjects with an indeterminate result, a follow-up scan was obtained 3 months after the baseline scan to assess the growth of the lesion. If at that time the lesion had a volume-doubling time of less than 400 days, the final result was declared to be positive; otherwise, it was considered to be negative. Subjects with positive screening tests were referred to a chest physician for workup and diagnosis. If lung cancer was diagnosed, the participant was treated for the disease and left the screening trial; if no lung cancer was found, the regular second-round CT scan was scheduled for 12 months after the baseline scan.

Second-Round Scan

When one or more new nodules were found on the second-round scan, the interpretation (positive or negative result) was based on the size of the nodule, as it had been in round one; if the result was indeterminate, a follow-up scan was obtained 6 weeks later [10]. In the case of nodules that had been detected previously,

the second-round result was based on the volume-doubling time. If there was no growth, or if the volume-doubling time was more than 600 days, the screen was classified as negative. If the volume-doubling time was less than 400 days, or if a new solid component had emerged in a previously nonsolid nodule, the scan was considered to be positive. When the volume-doubling time was 400 to 600 days, the test result was considered to be indeterminate and a follow-up scan was obtained 1 year after the second-round scan. At that time, if the volume-doubling time was less than 400 days, the final result was considered to be positive; otherwise it was considered to be negative. If both new and existing nodules were present, the nodule with the largest volume or fastest growth determined the result. All participants with a negative second-round test result were invited to undergo the third round of screening 2 years after the second round. A cancer detected on screening was classified as a first-round or second-round cancer if it was diagnosed after a workup during the first year after a positive first-round or second-round screen, respectively. Lung cancers that were detected during the first year after a negative first-round or second-round screening test were classified as interval cancers. They were identified through linkage with the national pathology database, information from participants and general practitioners, and, in the case of round-one interval cancers, linkage with the National Cancer Registry. The workup, staging, and treatment were standard across all screening sites and were performed according to published guidelines [10,16,17].

All the authors contributed to the data collection and the decision to submit the manuscript for publication, and all the authors vouch for the accuracy and completeness of the data.

Statistical Analysis

The diagnostic sensitivity was defined as the ratio between the number of true positive results (participants who were diagnosed with lung cancer during the first year after a positive screening test) and the number of true positive results plus the number of false negative results (interval cancers detected during the same time period). Diagnostic sensitivity, specificity, positive predictive value, and negative predictive value were calculated at the participant level, and 95%

confidence intervals were determined with the use of SPSS software, version 15.0 (SPSS). The standard for a negative baseline or second-round test result was based on the retrospective information that lung cancer was absent 2 years after the first round of screening and 1 year after the second round. Normally distributed data are shown as means ± SD. P values of less than 0.05 were considered to indicate statistical significance.

Results

First round

The mean (±SD) age of the screened participants was 59 ± 6 years, and the mean number of pack-years smoked was 42 ± 19; a total of 16% of the participants were women. The first round of screening was conducted from April 2004 through December 2006 (Supplemental Figure B). Of the 7557 participants, 50.5% had a total of 8623 non-calcified pulmonary nodules, of which 98.0% were solid. Automated volumetric data were manually adjusted in the case of 6.3% of the nodules. The screening results were deter- mined to be negative in 5987 participants (79.2%), indeterminate in 1451 (19.2%), and positive in 119 (1.6%) (Figure 1). A total of 1536 follow-up scans were obtained 100 ± 19 days, on average, after the baseline scan in participants with an indeterminate result. Including the outcome of these follow-up scans, the results from round one of the screening were negative in 7361 participants (97.4%) and positive in 196 (2.6%).

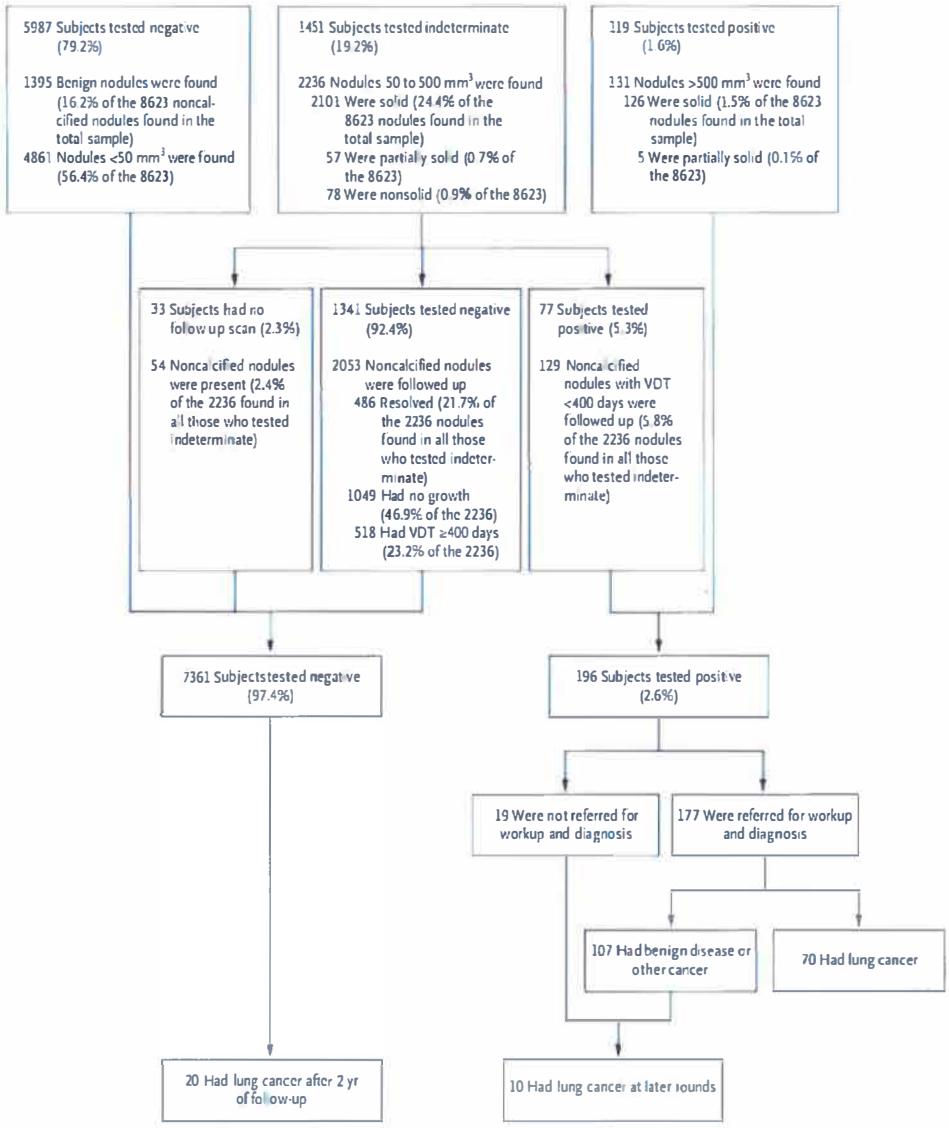


Figure 1 Results of the First Round of Screening

Of the 196 participants with a positive scan, 177 were referred for workup; 19 were not referred (9 because of a decision by the tumor board, 3 because of an administrative error, and 7 because they were already receiving treatment from another specialist). Lung cancer was diagnosed in 70 of the 177 participants who had a positive scan (39.5%); the diagnosis was made mainly by means of an invasive procedure (85.7%). These 70 participants had 72 lung cancers, of which 46 (63.9%) were classified as pathological stage I. In three subjects, no tissue for a histological diagnosis could be obtained. These subjects received high-dose radiotherapy because the lesions were growing and were assessed as positive on a positron-emission tomographic (PET) scan. Of the remaining 107 subjects with a positive scan, 100 had benign disease and 7 had metastases from another cancer. In round one, the proportion of invasive procedures that revealed benign disease was 27.2%.

The lung-cancer detection rate in round one was 0.9% (70 of 7557 subjects). There were four interval cancers, all of which were stage IV adenocarcinomas; three of these were new non-calcified nodules, and one, which had been seen in the first round, had a volume-doubling time of more than 600 days at the 3-month follow-up. The sensitivity of round-one screening was 94.6% (95% confidence interval [CI], 86.5 to 98.0), the specificity 98.3% (95% CI, 98.0 to 98.6), the positive predictive value 35.7% (95% CI, 29.3 to 42.7), and the negative predictive value 99.9% (95% CI, 99.9 to 100.0). Thus, in a subject with a positive CT screening test, the probability that the lesion would be malignant was 36%; with a negative screening test, the probability that a participant would not have lung cancer was 99.9%.

Among the 7361 negative CT scans in round one, 20 lung cancers were detected during the 2 years of follow-up: 3 were round-one interval cancers, and 17 were detected in the round-two screening. On the basis of this information, the negative predictive value was 99.7% (95% CI, 99.6 to 99.8). All 126 participants with a positive screening result at round one but with a negative workup returned to the screening program. After a mean follow-up of 785±263 days, 10 of these 126 subjects received the diagnosis of pulmonary adenocarcinoma, which appeared to have originated from a suspicious nodule that was detected in round one (Sup-

plemental Table A).

Second Round

In accordance with the trial's protocol, all the participants in the first round of screening, except those in whom lung cancer had been diagnosed, were invited to undergo screening in the second round, 12 which was conducted from April 2005 through April 2008. A total of 7289 participants underwent screening 384 ± 59 days after the round-one screening (Supplemental Figure A). In 1588 (21.8%) of these participants, a total of 2320 new nodules were detected, 29.2% of which had a volume of less than 15 mm^3 or had been missed in round one. Automated volumetric data were manually adjusted in the case of 5.4% of the new nodules and 1.9% of previously existing nodules. The second-round screening result was negative in 6719 participants (92.2%), indeterminate in 480 (6.6%), and positive in 90 (1.2%) (Figure 2). Among participants with an indeterminate result, 276 had a follow-up scan 77 ± 36 days after the second-round screening and 231 had a follow-up scan 364 ± 36 days after the second-round screening. The follow-up scans were positive in 38 subjects, and when the results of these positive follow-up scans were added to the results of the 90 positive screening scans, there were 128 subjects (1.8%) with positive second-round scans. Of these 128 participants, 1 patient died as a result of a metastatic colon carcinoma and 118 were referred for workup; 54 of the 118 who were referred for workup (45.8%) received the diagnosis of lung cancer, mainly after undergoing an invasive procedure (88.9%). The nine participants who were not referred for workup (four because of a decision by the tumor board, four because of an administrative error, and one be-cause the patient was already receiving treatment from another specialist) were invited to participate in the third round of screening 2 years later. In one of these nine, lung cancer was found 23 months after the first detection of the nodule in a nodule that had not been seen previously. Of the remaining 64 subjects with a positive scan, 62 had benign disease and 2 had another cancer (1 a thymoma and 1 lymphoma).

There were two subjects with suspicious lesions from whom no tissue could be obtained for histological diagnosis. These subjects were treated with high-dose radiotherapy because the lesions were new and growing and were positive on a

PET scan. The 54 participants with lung cancer had 57 cancerous nodules, 42 of which (73.7%) were classified as pathological stage I, including 3 that were synchronous double tumors. The lung-cancer detection rate was 0.5% (40 of 7289) during the first year after the second-round screening and 0.8% (57 of 7289) for the entire 2-year period after the second and third rounds of screening. One stage IV small-cell and one stage IV large-cell interval carcinoma, both of which were present in nodules that had been absent at the time of the second-round screening, were diagnosed during the first year after the second-round screening. The sensitivity of the second-round screening was 96.4% (95% CI, 86.8 to 99.1), the specificity was 99.0% (95% CI, 98.7 to 99.2), the positive predictive value was 42.2% (95% CI, 33.9 to 50.9), and the negative predictive value was 99.9% (95% CI, 99.9 to 100.0).

Additional Diagnostic Investigations

The recall rates for CT scans among participants with indeterminate test results were 19.0% and 3.8% in rounds one and two, respectively (Supplemental Table B). No diagnostic PET or PET–CT scanning was performed in participants with positive test results, and fine-needle biopsy procedures were performed in less than 1% of the subjects. The rate of invasive diagnostic procedures was 1.2% in round one and 0.8% in round two.

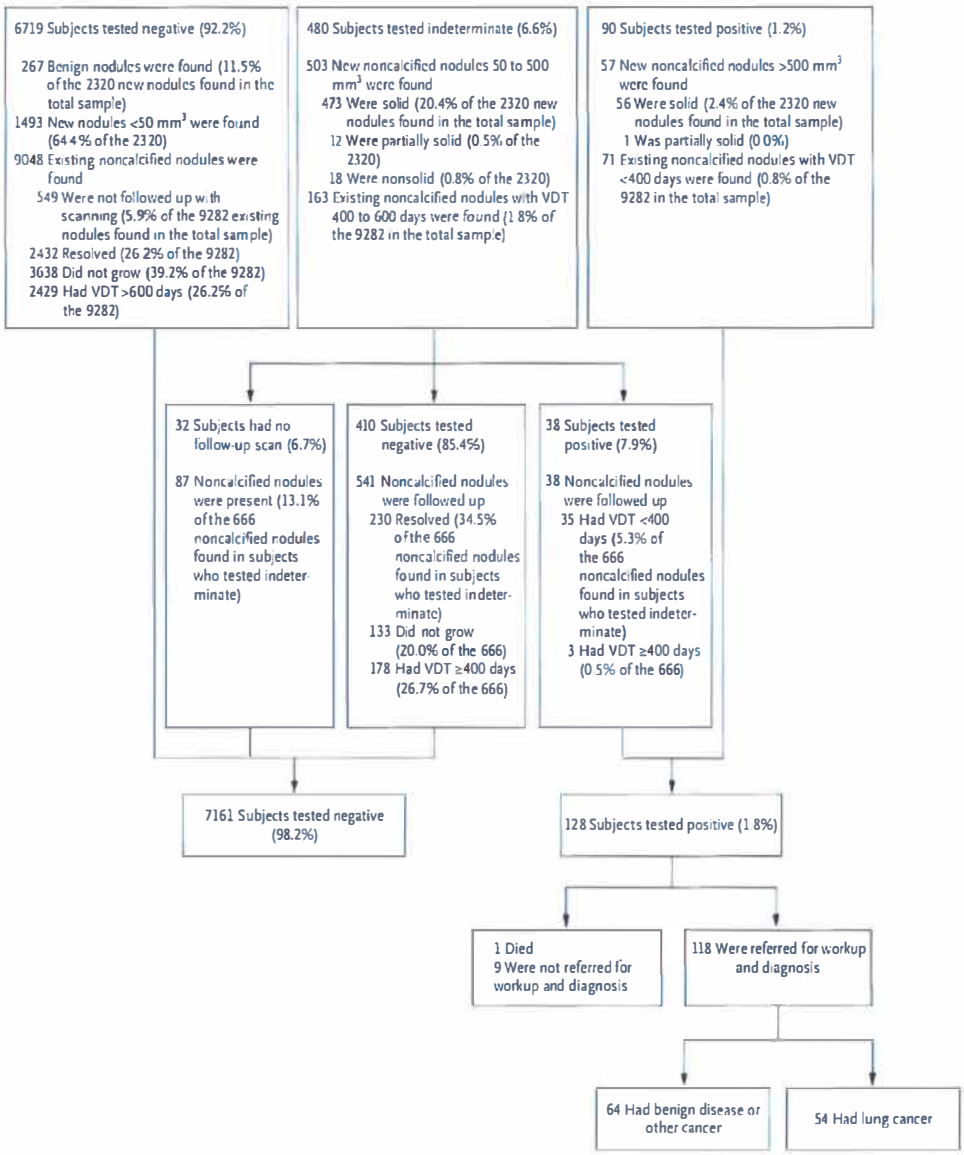


Figure 2 Results of the Second Round of Screening. Some participants had more than one nodule. VDT = volume-doubling time.

Discussion

In a population that was at an increased risk for lung cancer, our strategy of screening for lung cancer with the use of volume CT diminished the need for follow-up evaluation in participants with an indeterminate test result. This strategy was especially useful during the second-round screening. It reduced the number of follow-up examinations in participants with a positive test result without reducing the overall sensitivity of the technique, as compared with that reported in the literature [4-8,18-23]. This report concerns itself only with how to deal with an abnormality that has been detected on a CT scan in this population; it does not address the usefulness of screening for lung cancer with the use of CT scanning.

The rate of interval cancers that were found in participants in our trial was similar to that found in participants in other trials [20]. The proportion of early (stage I) lung cancers detected in round one (63.9%) was similar to that found in other randomized trials [18,19,23], but lower than that found in nonrandomized trials (e.g., the proportion in the International Early Lung Cancer Action Program [I-ELCAP] was 86%, and the proportion in a trial performed at the Mayo Clinic was 75%) [6,7,20]. The lung-cancer detection rate in round one in I-ELCAP was higher than that in NELSON (1.3% vs. 0.9%) [7], despite similar median ages of the participants and a higher number of pack-years smoked by participants in NELSON. The discrepancy was probably due to the fact that the proportion of women, who tend to have slow-growing cancers [24,25], was higher in I-ELCAP than in NELSON. Moreover, in I-ELCAP surgeons removed any nonsolid nodule that was larger than 8 mm, instead of waiting for the nodule to grow before removing it, as was done in NELSON. In our trial of subjects who had an increased risk of lung cancer, we found that the chances of finding lung cancer on a CT scan at 3 months, 1 year, and 2 years after a negative first-round test were 0, 1 in 1000, and 3 in 1000, respectively.

In round one, the proportion of invasive procedures that revealed benign disease was 27.2%, which is similar to that found in other trials [5,6,19,21,22,26-30]. The advantages of volumetric measurements become fully apparent when a volumetric comparison can be made with a previous indeterminate CT scan. Because

there were no comparative CT scans available at round one, the first-round recall rate was almost as high as that in other trials (Supplemental Table B). The LungCare software version that we used is not proprietary and can be used with any CT data set, regardless of the CT system, for evaluation of solid nodules and the solid component of partially solid non-calcified nodules smaller than 500 mm^3. With manual correction, the mean relative deviation from the true lesion volume was only -0.3 ± 6.5% for these types of lesions [13].

As an absolute standard for negative test results, we used the absence of lung cancer after 2 years of follow-up, a period that is considered to be sufficient for concluding that a nodule is benign [2]. The 400-day threshold for volume-doubling time that we used was based on current opinion that lung cancers with a volume-doubling time of 400 days or more are over-diagnosed cases [24,31]. A volume-doubling time of 500 days is regarded as the upper limit for lung cancer, even though some tumors may grow more slowly [32-34]; our upper limit was set at 600 days. If a lower upper limit had been used, the rate of false negatives would have increased, but the rate of false positives would have decreased. Therefore, the ranges for volume-doubling time that we used are not definite and could be improved. Finally, before we can make clinically directive recommendations, our strategy requires validation in an independent study.

References

1. Fischbach F, Knollmann F, Griesshaber V, Freund T, Akkol E, Felix R. Detection of pulmonary nodules by multislice computed tomography: improved detection rate with reduced slice thickness. Eur Radiol 2003;13:2378-83.
2. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology 2005;237:395-400.
3. van Klaveren RJ, de Koning HJ, Mulshine J, Hirsch FR. Lung cancer screening by spiral CT: what is the optimal target population for screening trials? Lung Cancer 2002;38:243-52.
4. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999;354:99-105.
5. 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; 362:593-7.
6. Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: five-year prospective experience. Radiology 2005; 235:259-65.
7. The International Early Lung Cancer Action Program Investigators. Survival of patients with clinical stage I lung cancer detected on CT. N Engl J Med 2006;355:1763-71. [Errata, N Engl J Med 2008;358: 1862, 2008;358:1875, 2008;359:871.]
8. Sone S, Nakayama T, Honda T, et al. Long-term follow-up study of a population-based 1996-1998 mass screening programme for lung cancer using mobile low-dose spiral computed tomography. Lung Cancer 2007;58:329-41.
9. National Lung Screening Trial. Bethesda, MD: National Cancer Institute. (Accessed November 6, 2009, at http://www.nci.nih.gov/NLST.)
10. Xu DM, Gietema HA, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 2006;54:177-84.
11. 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 2007;120:868-74.

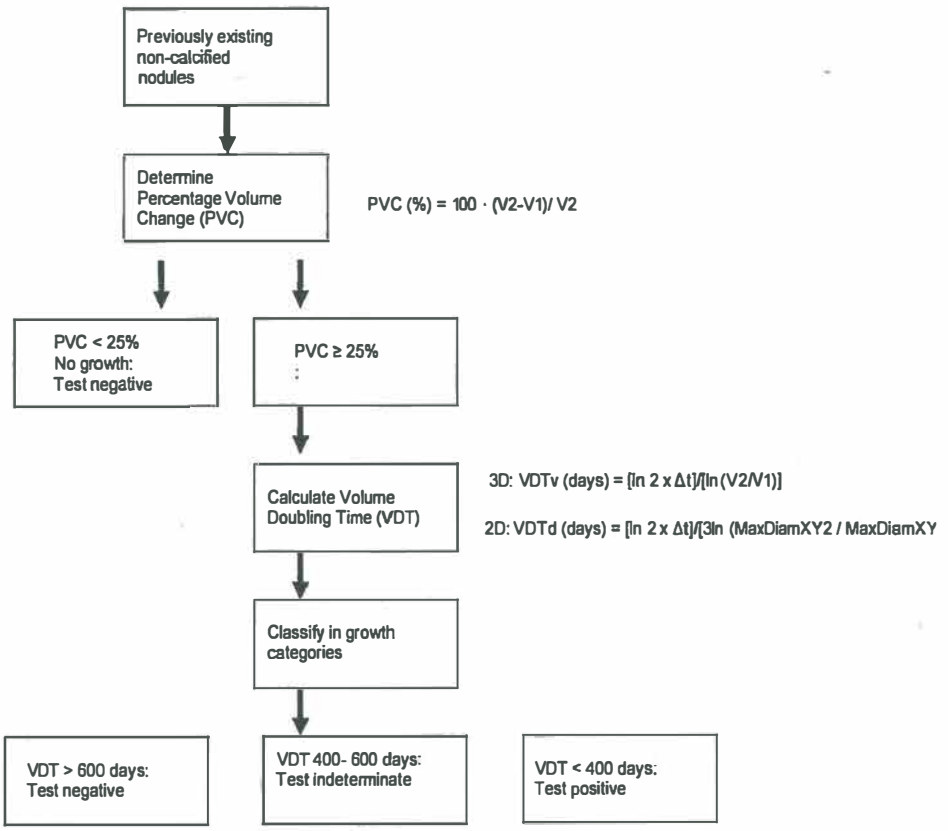
12. Gietema HA, Schaefer-Prokop CM, Mali WP, Groenewegen G, Prokop M. Pulmonary nodules: interscan variability of semi-automated volume measurement with multisection CT — influence of inspiration level, nodule size and segmentation performance. Radiology 2007;245:888-94.
13. Bolte H, Riedel C, Müller-Hülsbeck S, et al. Precision of computer-aided volumetry of artificial small solid pulmonary nodules in ex vivo porcine lungs. Br J Radiol 2007;80:414-21.
14. Wormanns D, Kohl G, Klotz E, et al. Volumetric measurement of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004;14:86-92.
15. Goodman LR, Gulsun M, Washington L, Nagy PG, Piacsek KL. Inherent variability of CT lung nodule measurements in vivo using semi-automated volumetric measurements. AJR Am J Roentgenol 2006;186:989-94.
16. Mountain CF. Revisions in the International System for Staging Lung Cancer. Chest 1997;111:1710-7.
17. CBO guideline — non-small cell lung cancer: staging and treatment. Alphen aan de Rijn, the Netherlands: Van Zuiden Communications BV, 2004.
18. Blanchon T, Bréchot JM, Grenier PA, et al. Baseline results of the Depiscan study: a French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR). Lung Cancer 2007;58:50-8.
19. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. Lung Cancer 2008;59:355-63.
20. Black WC. Computed tomography screening for lung cancer: review of screening principles and update on current status. Cancer 2007;110:2370-84.
21. Wilson DO, Weissfeld JL, Fuhrman CR, et al. The Pittsburgh Lung Cancer Screening Study (PLuSS): outcomes within 3 years of a first CT scan. Am J Respir Crit Care Med 2008;178:956-61.
22. Veronesi G, Bellomi M, Veronesi U, et al. Role of positron emission tomography scanning in the management of lung nodules detected at baseline

computed tomography screening. Ann Thorac Surg 2007;84:959-66.

23. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. Lung Cancer 2009;64:34-40.
24. Lindell RM, Hartman TE, Swensen SJ, et al. Five-year lung cancer screening experience: CT appearance, growth rate, location and histologic features of 61 lung cancers. Radiology 2007;242:555-62.
25. International Early Lung Cancer Action Program investigators. Women's susceptibility to tobacco carcinogens and survival after diagnosis of lung cancer. JAMA 2006;296:180-4. [Erratum, JAMA 2008 299:1775.]
26. 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;126:114-21.
27. Gohagan JK, Marcus PM, Fagerstrom RM, et al. Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. Lung Cancer 2005;47:9-15.
28. Menezes RJ, Roberts HC, Paul NS, et al. Lung cancer screening using low-dose computed tomography in at-risk individuals: the Toronto experience. Lung Cancer 2009 May 6 (Epub ahead of print).
29. Veronesi G, Bellomi M, Mulshine JL, et al. Lung cancer screening with low-dose computed tomography: a non-invasive diagnostic protocol for baseline nodules. Lung Cancer 2008;61:340-9.
30. Veronesi G, Bellomi M, Scanagatta P, et al. Difficulties encountered managing nodules detected during a computed tomography lung cancer screening program. J Thorac Cardiovasc Surg 2008;136:611-7.
31. Yankelevitz DF, Kostis WJ, Henschke CI, et al. Overdiagnosis in chest radiographic screening for lung carcinoma: frequency. Cancer 2003;97:1271-5.
32. Revel MP, Merlin A, Peyrard S, et al. Software volumetric evaluation of doubling times for differentiating benign versus malignant pulmonary nodules. AJR Am J Roentgenol 2006;187:135-42.
33. Winer-Muram HT, Jennings SG, Tarver RD, et al. Volumetric growth rate of

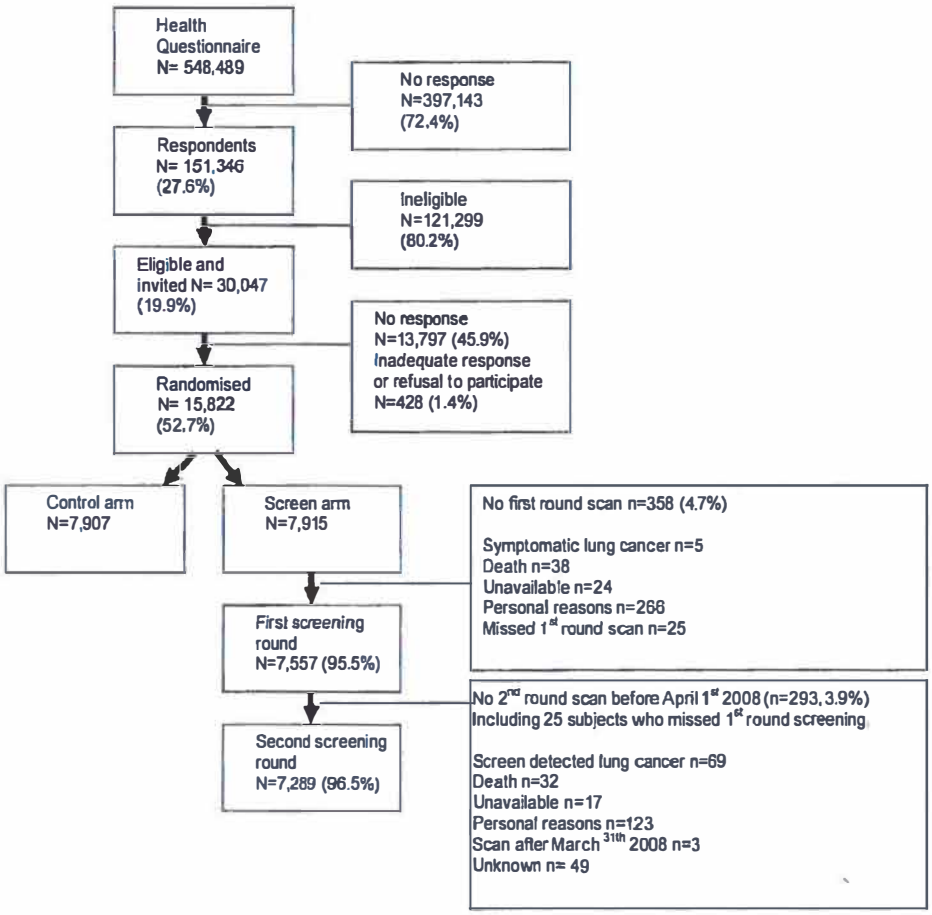
stage I lung cancer prior to treatment: serial CT scanning. Radiology 2002;223:798-805.

34. Lindell RM, Hartman TE, Swensen SJ, Jett JR, Midthun DE, Mandrekar JN. 5-Year lung cancer screening experience: growth curves of 18 lung cancers compared to histological type, CT attenuation, stage, survival and size. Chest 2009 July 6 (Epub ahead of print).



Supplemental Figure A

V1 = volume of the nodule (mm^3) at first detection on CT; V2 = volume of the nodule (mm^3) at subsequent CT evaluation; 3D = volume by three-dimensional volumetry software (VDTv); 2D = volume estimate based on two-dimensional measurements (VDTd); MaxDiamXY1 = maximum diameter in X/Y-axis at first detection on CT; MaxDiamXY2 = maximum diameter in X/Y-axis at subsequent CT evaluation



Supplemental Figure B CONSORT flowchart

Supplemental Table A Subjects with a positive 1st round test result and a false negative work-up in whom lung cancer was diagnosed in the suspicious 1st round nodule at later screening rounds

	1st screening round						Final lung cancer diagnosis			
Case	Moment of first detection	Consistency	Size at first detection	VDT, days	Highest level procedure	Diagnosis	Interval, months	VDT, days	Diagnosis	pTNM
Lung cancer detected during 2nd round										
1	3months FU	Partial-solid	50-500 mm^3	<400	FBR with washing	No malignancy	23	<400	Adeno-cacinoma	T1N0M0
2	3months FU	Non-solid	50-500 mm^3	<400	FBR with washing	Inconclusive	36	<400	Adeno-cacinoma	T1N0M0
3	Baseline	Solid	>500 mm^3	NA	FNA	Fibrosis	27	<400	Adeno-cacinoma	T1N0M0
4	3months FU	Solid	50-500 mm^3	<400	CT	Rest of pneumonia	20	<400	Adeno-cacinoma	T1N0M0
5	3months FU	Solid	50-500 mm^3	400-600	FBR with washing	No malignancy	11	<400	Adeno-cacinoma	T1N1M0
6	Baseline	Solid	>500 mm^3	NA	FBR with washing	No malignancy	26	>600	Adeno-cacinoma	T1N2M0
7	Baseline	Solid	>500 mm^3	NA	No work-up*	NA	14	400-600	Adeno-cacinoma	T1N0M0
8	3months FU	Solid	50-500 mm^3	<400	FBR with washing	No malignancy	24	<400	Adeno-cacinoma	T1N0M0
Lung cancer detected during 3rd round										
9	3months FU	Non-solid	25 mm	NA	FBR with washing	No malignancy	37	NA	Adeno-cacinoma	T2N0M0
10	3months FU	Solid	>500 mm3	NA	FBR with washing	Fibrosis	32	>600	Adeno-cacinoma	T2N0M0

VDT = volume doubling time; FU = follow-up; FBR = flexible bronchoscopy; CT = computed tomography; FNA = fine needle aspirate; NA = not applicable

* Protocol violation.

Supplemental Table B Additional diagnostic evaluations in participants with an indeterminate or positive test result following first and second round screening of the NELSON trial in comparison with the literature

	NELSON [10]			PluSS [21]	Cosmos [22,29, 30]	Toronto [28]	LSS† [26,27]		
Variable	All no. (%)	No lung cancer no. (%)	Lung cancer no. (%)	All no. (%)	All no. (%)	All no. (%)	All no. %)	No lung cancer no. (%)	Lung cancer no. (%)
Round one screening	7557 (100)	7487 (100)	70 (100)	3642 (100)	5203 (100)	3352 (100)	1586 (100)	1556 (100)	30 (100)
Clinical evaluation	181 (2)	111 (2)	70 (100)	1477 (41)	NA	NA	244 (15)	217 (14)	27 (90)
Recall chest CT scan	1438 (19)	1419 (19)	19 (27)	821 (23)[§]	482 (9)	628 (19)	325 (21)	305 (20)	20 (67)
Recall chest CT scans/subject	1.1	1.1	1.2	1.4	NA	NA	1.0	1.0	1.0
Chest X-ray	55 (1)	27 (0)	28 (40)	NA	NA	NA	92 (6)	80 (5)	12 (40)
PET or PET/CT	0 (0)	0 (0)	0 (0)	NA	160 (3)	NA	NA	NA	NA
MRI	5 (0)	2 (0)	3 (0)	NA	NA	NA	NA	NA	NA
Lung function test	147 (2)	78 (1)	69 (99)	NA	NA	NA	73 (5)	55 (4)	18 (60)
Bronchoscopy	149 (2)	84 (1)	65 (93)	NA	NA	NA	29 (2)	16 (1)	13 (43)
FNA	13 (0)	5 (0)	8 (11)	NA	4 (0)	57 (2)	46 (3)	18 (1)	28 (93)
Invasive procedure*	92 (1)	32 (0)	60 (86)	90 (3)	106 (2)	48 (1)	53 (3)	23 (2)	30 (100)
Round two screening	7289 (100)	7235 (100)	54 (100)	3423 (100)	4867 (100)	2686 (100)	1398 (100)	1390 (100)	8 (100)
Clinical evaluation	125 (2)	71 (1)	54 (100)	1450 (42)	NA	NA	NA	NA	NA
Recall chest CT scan	275 (4)	267 (4)	8 (15)	1386 (41)[§]	142 (3)	NA	NA	NA	NA
Recall chest CT scans/subject	1.1	1.1	1.4	1.1	NA	NA	NA	NA	NA
Chest X-ray	35 (0)	17 (0)	18 (33)	NA	NA	NA	64 (4)	NA	NA
PET or PET/CT	0 (0)	0 (0)	0 (0)	NA	66 (1)	NA	NA	NA	NA
MRI	0 (0)	0 (0)	0 (0)	NA	NA	NA	NA	NA	NA
Lung function test	103 (1)	55 (1)	48 (89)	NA	NA	NA	70 (4)	NA	NA
Bronchoscopy	98 (1)	46 (1)	46 (85)	NA	NA	NA	14 (1)	NA	NA
FNA	3 (0)	3 (0)	0 (0)	NA	NA	16 (1)	18 (1)	NA	NA
Invasive procedure*	61 (1)	13 (0)	48 (89)	NA	NA	NA	NA	NA	NA

◄ FNA = fine needle aspirate; PET = positron emission tomography; CT = computed tomography; MRI = magnetic resonance imaging; NA = not available

† Diagnostic follow-up information available for 316/325 and 351/360 participants with a positive test result at 1st and 2nd round screening, respectively.

* Includes: lung biopsy or wedge resection, video-assisted thoracotomy, thoracotomy, mediastinoscopy and mediastinotomy.

$ Includes: PET and PET-CT.

4

Performance of Computer-Aided Detection of Pulmonary Nodules in Low-Dose CT: Comparison to Double Reading by Nodule Volume

European Radiology 2012; 22:2076-2084

Yingru Zhao[12] / Geertruida H. de Bock[3] / Rozemarijn Vliegenthart[12]
Rob J. van Klaveren[4] / Ying Wang[12] / Luca Bogoni[5] / Pim A. de Jong[6]
Willem P. Mali[6] / Peter M.A. van Ooijen[12] / Matthijs Oudkerk[2]

University of Groningen, University Medical Center Groningen,
[1]Department of Radiology, [2]Center for Medical Imaging – North East Netherlands,
[3]Department of Epidemiology;
[4]Department of Pulmonology, Lievensberg Hospital;
[5]CAD Group, Siemens Medical Solutions USA;
[6]Department of Radiology, University Medical Center Utrecht

Abstract

Objective

To evaluate performance of computer-aided detection (CAD) beyond double reading for pulmonary nodules on low-dose computed tomography (CT) by nodule volume.

Methods

A total of 400 low-dose chest CT examinations were randomly selected from the NELSON lung cancer screening trial. CTs were evaluated by two independent readers and processed by CAD. A total of 1,667 findings marked by readers and/or CAD were evaluated by a consensus panel of expert chest radiologists. Performance was evaluated by calculating sensitivity of pulmonary nodule detection and number of false positives, by nodule characteristics and volume.

Results

According to the screening protocol, 90.9% of the findings could be excluded from further evaluation, 49.2% being small nodules (less than 50 mm^3). Excluding small nodules reduced false-positive detections by CAD from 3.7 to 1.9 per examination. Of 151 findings that needed further evaluation, 33 (21.9%) were detected by CAD only, one of them being diagnosed as lung cancer the following year. The sensitivity of nodule detection was 78.1% for double reading and 96.7% for CAD. A total of 69.7% of nodules undetected by readers were attached nodules of which 78.3% were vessel-attached.

Conclusions

CAD is valuable in lung cancer screening to improve sensitivity of pulmonary nodule detection beyond double reading, at a low false-positive rate when excluding small nodules.

Introduction

The rapid development of multi-detector CT (MDCT) has increased the amount of data for radiologists to analyze. Reviewer's fatigue increases the risk of false-negative diagnosis due to perceptual error. In addition, although the introduction of low-dose CT in lung cancer screening protocols was found effective in detecting peripheral lung cancers at an early stage [1-4], it may be more difficult for screening radiologists to find lesions in case of increased image noise due to low dose and thin slice thickness. To reduce the number of missed lesions, double reading has been recommended [5]. Previous studies have found that more pulmonary nodules are detected by double reading than by single reading [5, 6]. However, double reading is not widely used in clinical routine because of limited human resources and cost-effectiveness [7, 8].

Computer-aided detection (CAD) of pulmonary nodules may help address this problem by being utilized as an assistant reader [9-13]. A significant improvement in sensitivity was shown in pulmonary nodule detection, albeit at the disadvantage of a large increase in false-positive (FP) findings. Previous studies have found that CAD increases the sensitivity of pulmonary nodule detection compared to that of single human reading [14, 15]. In one rather small study, true-positive (TP) findings identified with the aid of CAD complemented radiologists' TP findings to a greater extent than those contributed by second readers [16].

In lung cancer screening, small pulmonary nodules are extremely common findings. Previous studies using low-dose CT for early detection of asymptomatic lung cancer in populations at risk reported that more than 95% of nodules 10 mm or smaller were benign [1, 2, 17]. Avail-able data indicate that less than 1% of very small (less than 5 mm, corresponding to 65.4 mm^3) nodules were malignant [2, 18, 19]. Therefore, a size cut-off in CAD could be more efficient in helping radiologists make diagnoses. In recent years, volume instead of diameter has become an important factor to evaluate nodule size and growth. As this measure is more accurate for evaluating growth [20-22], a volume cut-off is likely more precise in distinguishing probably malignant and probably benign nodules.

The purpose of our study was to assess the performance of CAD for detection of pulmonary nodules as a complementary tool in a large-scale, low-dose CT lung cancer screening study compared to double reading, with stratification according to nodule volume. Double reading is the original design of nodule evaluation in our lung cancer screening trial. The hypothesis of the current study was that CAD increases sensitivity of lung nodule detection beyond double reading.

Materials and methods

Study population

The subjects in this study were participants of the four screening sites of the Dutch–Belgian randomized trial for lung cancer screening (NELSON) by low-dose MDCT. The protocol required participants to be current or former smokers with a smoking history of more than 15 cigarettes / day for longer than 25 years or more than 10 cigarettes / day for longer than 30 years. The NELSON study was approved by the medical ethics committees of all institutions and all participants provided written informed consent [23] that also covered the current analysis. As a side-study of the NELSON project, we randomly selected 400 out of 4,280 base-line CTs from 2005 using a statistical program (SPSS 16.0 for Windows, SPSS, Inc., Chicago, IL, USA).

CT protocol

At all screening sites 16-detector CT was used (3 Sensation-16, Siemens Medical Solutions, Forchheim, Germany and 1 Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA). CT of the entire chest was per-formed, in caudo-cranial direction. CT data were acquired with 16 × 0.75 mm collimation and pitch of 1.3. No intravenous contrast medium was used. Low-dose settings were applied depending on body weight (less than 50 kg, 50–80 kg and greater than 80 kg), with corresponding kVp settings of 80–90, 120 and 140 kVp, to achieve a CT dose index volume of approximately 0.8, 1.6 and 3.2 mGy, respectively. The mAs settings were adjusted accordingly depending on the machine used. To minimize breathing artifacts, CT data acquisition was performed at suspended maximal in-

spiration after appropriate instructions were given to the subjects. Data were reconstructed at 1.0-mm slice thickness, with 0.7-mm reconstruction increments and soft kernel (Siemens B30 filter, Siemens Medical Solutions, Forchheim, Germany). The Siemens B30 kernel is the standard soft tissue reconstruction kernel. Transversal, 6-mm-thick maximum intensity projections (MIP) reconstructions were used to identify pulmonary nodules.

Evaluation of CT examinations by double reading

At the time of acquisition (2005), all CT images of the lungs from each examination were independently read by first and second readers (double reading) as part of the NELSON protocol [6, 23]. The first reading was performed by 13 readers (experience in reading thoracic CT varying from 0 to 20 years); the second reading was performed by two readers, each with 6 years of experience. Upon identifying a finding as a pulmonary nodule, volume measurement was performed by the individual readers as part of the double reading. The LungCARE© software package (Leonardo© workstation, Somaris/5 VB10A, Siemens Medical Solutions, Erlangen, Germany) designed to aid readers in measuring and characterizing pulmonary nodules was used in addition to visual readings by all readers. Nodule diameter and volume were automatically calculated using this three-dimensional (3D) volumetric assessment tool. In case of inappropriate segmentation, the radiologists could perform manual two-dimensional (2D) measurements using a calliper.

Lung CAD algorithm

The lung CAD algorithm evaluated in this study was a commercial software version available since 2006 (LungCAD VB10A, Siemens AG Healthcare) [24]. This is an extensively validated CAD software, designed as a multi-step approach aiming to detect parenchymal lesions at high sensitivity and specificity, focusing on solid lesions larger than 3 mm. All CT images were processed by this LungCAD software package to mark potential lesions. The findings were reviewed both as 2D-axial images and 3D rendered views obtained with LungCARE. The MIP reconstruction settings used in LungCAD were equal to those in LungCARE.

Evaluation of findings by consensus panel

Retrospectively, a consensus panel of two expert radiologists with at least 8 years of experience in reading thoracic CT reviewed the CAD-marked images and the results obtained from double reading were entered into the NELSON management system, and compared the findings in LungCARE [23]. The consensus panel did not search for potential additional nodules. This reference standard was similar to previously reported practices [14, 25]. The consensus panel labeled the findings as "nodule" according to the definitions in the NELSON protocol [23]. Upon identifying a finding as a pulmonary lesion, volume measurement was performed by the consensus panel. Conforming with the image reading protocol used by the readers in the double reading, nodules smaller than 15 mm^3 were not assessed whereas larger non-calcified solid nodules were classified into three categories based on size (negative nodule, smaller than 50 mm^3; indeterminate nodule, 50–500 mm^3; positive nodule, larger than 500 mm^3) [26]. A cut-off of 50 mm^3 (4.6 mm diameter) was chosen as previous studies have shown that the possibility of malignancy in these small nodules is negligible [18, 19]. Because consistent volume measurement was not possible in non-/part-solid nodules, the calliper was used to measure the largest axial diameter of these lesions. Non-solid and part-solid nodules with non-solid component at least 8 mm as well as part-solid nodules with solid part larger than 50 mm^3 were considered indeterminate nodules.

Findings were divided into two groups based on NELSON's nodule categories: findings that could be excluded from further evaluation and those that needed further evaluation. Findings excluded from further evaluation were subdivided into three sub-groups: negative nodule (smaller than 50 mm^3), benign lesion or non-lesion. Calcifications and abnormal findings not presenting as nodule shapes, e.g. pleural plaque, fissure thickening or fibrosis, were recorded as benign lesions. A finding was assigned as "non-lesion" if the finding was due to normal anatomy or artifact. Findings needing further evaluation consisted of indeterminate and positive nodules. These findings were subsequently characterized by the consensus panel in terms of location (peripheral or non-peripheral), consistency (solid or non-/part-solid), attachment (intraparenchymal, fissure-attached, vessel-attached or pleural-based), shape (spherical or non-spherical) and edge (smooth or non-

smooth) [23, 27]. Nodules were classified as peripheral if the distance to the thoracic wall was less than one third of the total distance from the thoracic wall to the lung hilum, and as non-peripheral otherwise.

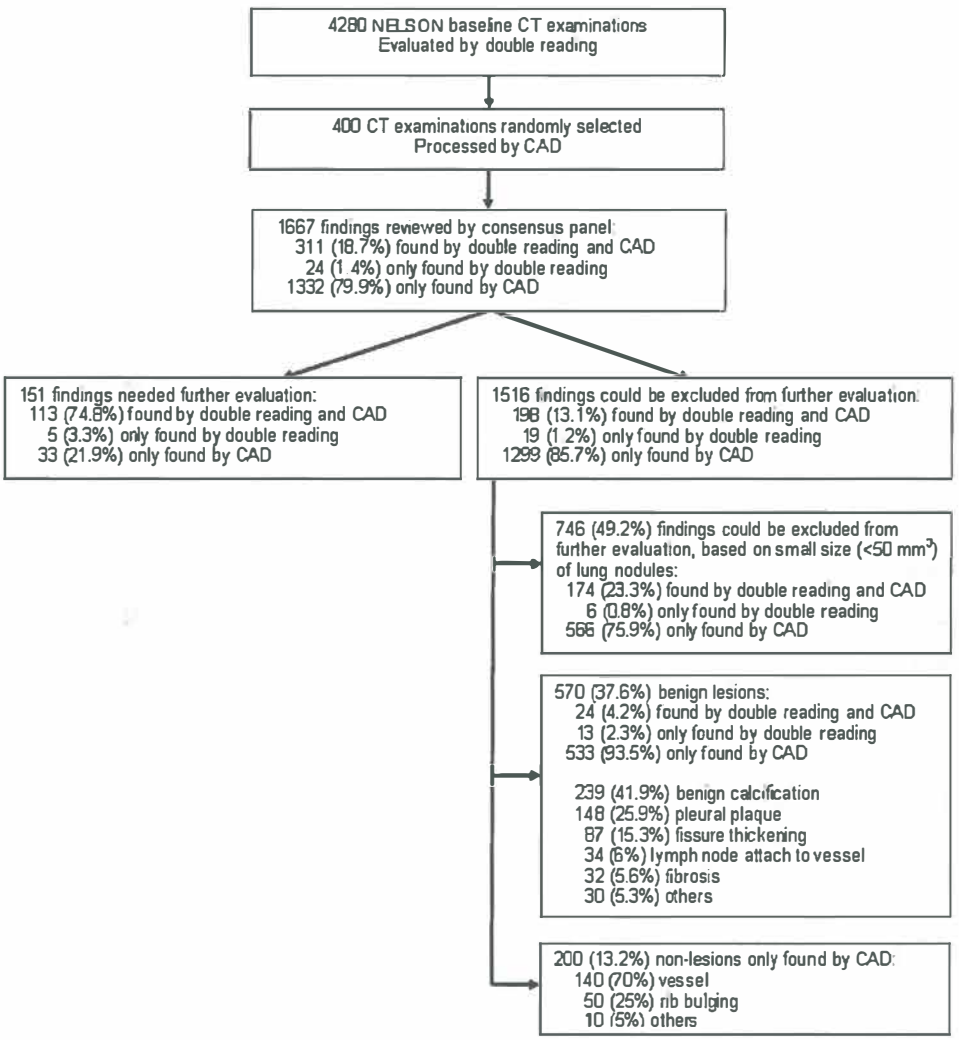


Figure 1 Flow chart of nodule detection and evaluation

Data analysis

Findings from double reading and from CAD were labelled either as TP, if they were determined by the consensus panel as findings needing further evaluation, or otherwise as FP. Sensitivity for pulmonary nodules from double reading and of CAD findings was calculated using the consensus panel as the reference standard. The FP rate presented with respect to nodule volume (less than or at least 50 mm^3) was computed as the number of FP detections per CT. Additionally, the positive predictive value of findings detected by CAD was calculated. The probability of detecting pulmonary nodules according to nodule characteristics was tested between CAD and double reading by the McNemar test. All statistical analyses were performed using SPSS 16.0.

Results

The mean age of the 400 participants was 59 ± 6 years (range 51 to 76 years). On 332 of the 400 baseline CT examinations at least one finding was reported (Figure 1). A total of 1,667 findings were detected by the readers and CAD system. A total of 90.9 % (n = 1,516) of the identified findings could be excluded from further evaluation (Figure 1). In this study, these findings were considered as FP findings. The FP rate was 3.7 per CT for CAD and 0.5 per CT for readers (Table 1). Excluding small nodules (less than 50 mm^3) and regarding benign lesions and non-lesions as FP findings only, the FP rate for CAD decreased to 1.9. By using 50 mm^3 as the cut-off below which pulmonary nodules were disregarded, the positive predictive value of CAD increased from 8.9 to 16.2%. The positive predictive value of double reading was 35.2% and 76.1% for all nodules and nodules larger than 50 mm^3, respectively.

According to the consensus panel, 151 (9.1 %) of 1,667 findings needed further evaluation. Of these 151 nodules, 113 were found both by readers and CAD, and 33 and 5 were found only by CAD or readers, respectively (Figure 1). The overall sensitivity for potentially significant pulmonary nodules (indeterminate and positive pulmonary nodules) was 78.1% for readers and 96.7% for CAD (Table 1).

Table 1 Sensitivity, false positive (FP) rate and positive predictive value (PPV) of nodule detection for double reading and CAD in all nodules and nodules larger than 50 mm^3

	All nodules		Nodules > 50 mm^3	
	Double reading	CAD	Double reading	CAD
Sensitivity, %	78.1	96.7	78.1	96.7
FP/examination, n	0.5	3.7	0.1	1.9
PPV, %	35.2	8.9	76.1	16.2

CAD = computer-aided detection; FP = false positive; PPV = positive predictive value

Table 2 presents an overview of the nodules found by readers or by CAD. Among the 151 indeterminate and positive pulmonary nodules, 76.6% were located peripherally, 96.7% were solid nodules, 49.7% were intraparenchymal and 76.2% were spherical and 87.4% were smooth. The median volume of 146 solid nodules was 85.4 mm^3 (range 50.0 to 1,672.4 mm^3). Consistent volume measurement was not possible in the 5 non-/part-solid nodules. CAD was better in detecting most types of nodules, namely peripheral and non-peripheral nodules, solid nodules, intraparenchymal nodules, and spherical and non-spherical nodules. Some differences could not be tested for significance as some cells were empty (for non-/part-solid, vessel-attached, non-smooth and positive nodules).

Only 37.9% (11/29) of vessel-attached nodules were detected by readers. A total of 69.7% of 33 nodules not detected by readers were attached nodules, and 78.3% of these were vessel-attached (Figure 2). Of the non-peripheral, vessel-attached nodules, 7 out of 11 were missed by readers but all were detected by CAD. Of 33 nodules missed by readers at baseline, 24 were detected at 3-month or 1-year follow-up CT examinations. Lung cancer was diagnosed in one solid intraparenchymal nodule, found to have grown at the second-year screening CT. The baseline volume of this missed nodule was 160.7 mm^3.

One fissure-attached and two pleural-based nodules were missed by CAD. Two of five nodules missed by CAD were non-/part-solid. A solid pleural-based nodule with volume 217.8 mm^3 missed by CAD was diagnosed as lung cancer after

it was found to be growing on the 3-month follow-up CT examination, with volume doubling time less than 400 days (Figure 3).

Table 2 Characteristics of 151 pulmonary nodules needing further evaluation, found by CAD and/or double reading

Variable	*n*	Nodules found by		*P* value
		CAD (%)	Double reading (%)	
Location				
Peripheral	116	112 (96.6)	94 (81.0)	<0.01
Non-peripheral	35	34 (97.1)	24 (68.6)	<0.01
Consistency				
Solid	146	143 (97.9)	113 (77.4)	<0.001
Non-/part-solid	5	3 (60.0)	5 (100)	NA
Attachment				
Intraparenchymal	75	73 (97.3)	65 (86.7)	<0.05
Fissure-attached	18	17 (94.4)	17 (94.4)	NS
Vessel-attached	29	29 (100)	11 (37.9)	NA
Pleural-based	29	27 (93.1)	25 (86.2)	NS
Shape				
Spherical	115	111 (96.5)	92 (80.0)	<0.001
Non-spherical	36	35 (93.1)	26 (72.2)	<0.05
Edge				
Smooth	132	130 (98.5)	99 (75.0)	<0.001
Non-smooth	19	16 (84.2)	19 (100)	NA
Volume*				
50-500 mm^3	141	138 (97.9)	108 (76.6)	<0.001
>500 mm^3	5	5 (100)	5 (100)	NA

* For 5 non-/part-solid nodules, volume was not available.

NA = not applicable. NS = not significantly different.

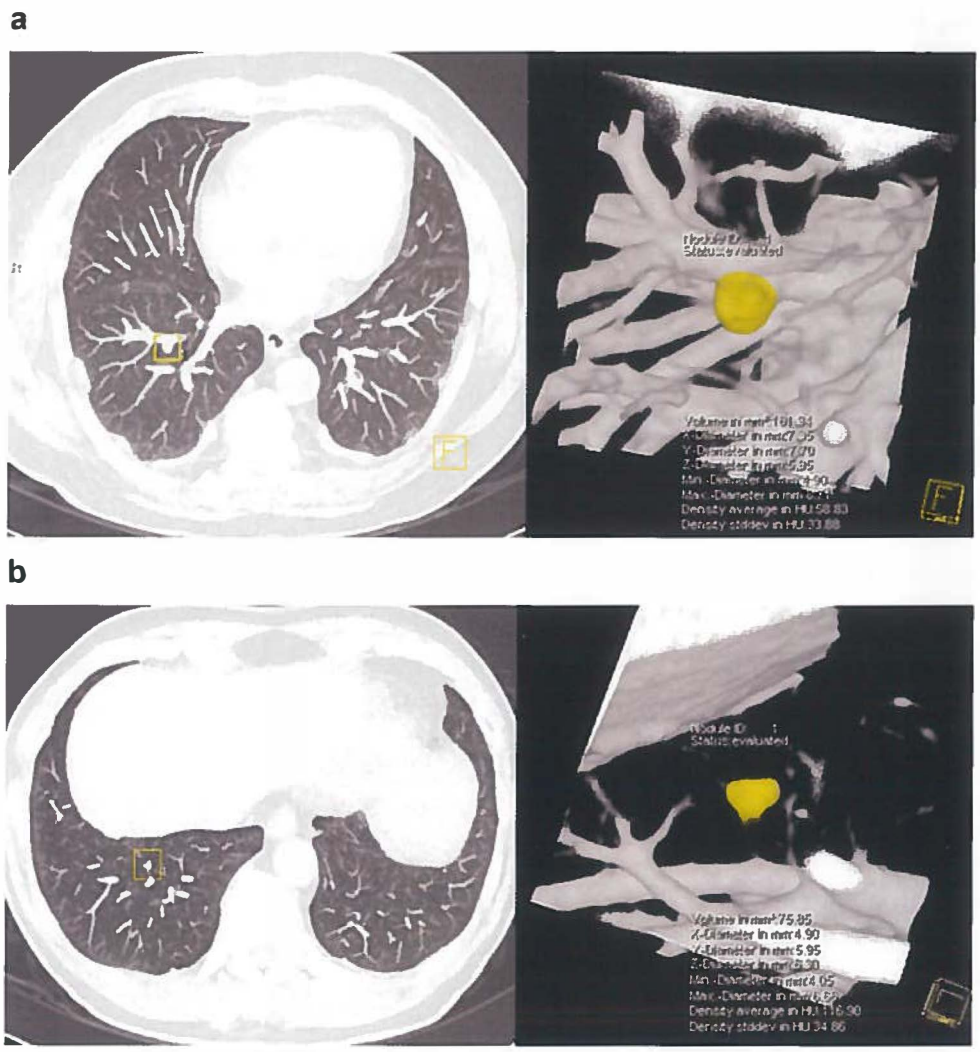


Figure 2 Examples of pulmonary nodules needing further evaluation that were missed by double reading. Vessel-attached nodule with baseline volume of 161.9 mm^3 (**a**), intraparenchymal nodule with volume of 75.9 mm^3 (**b**).

Through the fourth screening round (the 7^{th} year), in total 7 lung cancers (all adenocarcinomas) have been diagnosed. None of the lung cancers originated from FP results from CAD. Three of the cancers were proven by biopsy during the baseline round, 3 during the second screening round, and 1 at the fourth round screen-

ing. None of the benign-appearing pulmonary nodules presented with malignant behavior during subsequent screening rounds.

Discussion

Of the 1,667 findings on lung cancer screening CT by readers and CAD, presented to the consensus panel, 90.9% could be excluded from further evaluation with small size of pulmonary nodules being the main reason (49.2%). The false-positive findings by CAD decreased from 3.7 to 1.9 per CT by using a nodule volume cut-off (larger than 50 mm^3). Given the 151 (9.1%) findings that needed further evaluation, 33 nodules (21.9%) would have been missed if CAD was not applied. The sensitivity of nodule detection by readers could have increased by 18.6% (from 78.1% to 96.7%) if CAD had also been used. However, only one lung cancer missed by readers was detected by CAD.

Numerous studies have demonstrated that the introduction of CAD in radiological practice can significantly improve the diagnostic accuracy of pulmonary nodule detection. The reported sensitivity of CAD ranged from 38 to 100% [28-35]. Our study indicates a high sensitivity of greater than 95% when LungCAD software is used. By using CAD, an extra 18.6% of nodules were detected. A general comparison between our study and previous studies is, however, not possible due to the differences in methods, e.g. regarding the CT technique and the threshold of nodule size for CAD. These types of differences may explain the wide range in sensitivity reported.

High FP rate is still a considerable drawback of CAD. In this study, the FP rate of CAD was low compared to that of other studies (range 1.3 to 13.4/case, average 4.7/case) [28-35]. However, it is still higher than the FP rate for double reading (0.5/case). Over 80% of the FPs in this study was reported only by CAD, of which nearly half could be excluded from further evaluation if nodule size was considered. Using a nodule volume cut-off of greater than 50 mm^3, the FP rate decreased to 1.9 per CT. The use of 50 mm^3 (equal to 4.6 mm diameter for a spherical nodule) as cut-off volume for pulmonary nodules is supported by the NELSON results [14]. As published in the New England Journal of Medicine [14],

the NELSON nodule evaluation protocol with a negative screening result in case of nodules smaller than 50 mm^3 had a sensitivity for lung cancer of 94.6%, whereas none of the scarce interval cancers between the first and second year screening were due to malignancy in pulmonary nodules smaller than 50 mm^3. The chance of finding lung cancer in a participant on a second-round screening CT after a negative baseline test was only 1 in 1,000, confirming the safety of the current approach and the negligible 1-year risk of lung cancer in very small pulmonary nodules (smaller than 50 mm^3). Use of CAD led to one additional lung cancer being detected, whereas one malignant pulmonary nodule was missed by CAD. Both nodules had a volume greater than 50 mm^3.

Among all FP findings identified by CAD, nearly 40% were considered benign lesions by the consensus panel, e.g. fissure thickening and pleural plaque. In a previous study by Wormanns et al. [35] concerning nodules adjacent to the pleural surface, none of the 21 pleural-based findings detected by CAD were regarded as true pulmonary nodules. Given the high rate of this type of CAD finding in our study, one may conclude that CAD has difficulty in distinguishing pleura-based nodules and pleural plaques. This may be caused by the image segmentation component of the algorithm which may regard a part of the chest wall as a nodule and include it in further image processing. On the other hand, the one lung cancer missed by CAD was a pleural-based nodule. A considerable number of FP findings for CAD concerned vessels and rib bulging which were frequently misinterpreted owing to their nodule-like appearance in cross sections. Another principal problem was the difficulty in establishing a density value as threshold for lesion detection as a result of partial volume effects and motion artifacts. All non-lesions were easily distinguished from nodules by the readers, particularly when 3D visualization was used in the pulmonary nodule evaluation platform.

Various factors affect nodule recognition during screening including reader experience and variability, CT technique and viewing conditions, as well as nodule characteristics [36]. The performance of readers can be influenced by nodule location and its relationship to surrounding anatomical structures [37, 38]. The radiologist has little difficulty in finding peripheral and subpleural nodules even if they are small because there are no vessels of similar size near the pleural surface

[39]. In central lung regions, however, nodules can go undetected because they can be confused with blood vessels imaged in axial cross sections [35, 40]. A lesion not noticed by a reader because of a particular location, may often be detected in retrospective review after being detected on a subsequent CT. We found that vessel-attached nodules in particular can be missed by human readers. Although Marten et al. [41] reported that readers recognized more of the nodules with vascular attachment, Naidich et al. [37] showed that nodules either over-lapping or superimposing blood vessels were harder for radiologists to identify (sensitivity 32.5%). In our study, 30.3% of the attached nodules were not detected by human readers, and 78.3% of these missed nodules were vessel-attached. Furthermore, the mean size of vessel-attached nodules missed by readers was larger than that of other subtypes (data not shown). This indicates that contact with vessels increases the difficulty of detection by radiologists. The study by Naidich [37] demonstrated a significant relationship between nodule location and detectability by human reader sensitivity: peripheral 73.9%, central 48.6%, perihilar 36.7%). Of the non-peripheral nodules, two-thirds were found by double reading, considerably higher than in the aforementioned article. However, of the non-peripheral, vessel-attached nodules, 7 out of 11 were missed by readers. CAD was significantly more sensitive for these types of nodules detecting all 11 of them. In our study, the percentage of sub-solid nodules was low (5 nodules, 3.3%), similar to the relatively low prevalence in our entire lung cancer screening study (2%) [26]. Two of the five non-/part-solid nodules were missed by CAD but none were missed by readers. Most CAD systems so far are designed and optimized for solid nodules. The obstacle of adequate detection of sub-solid nodules is primarily caused by the setting of an attenuation range. The selection of texture features will affect the diagnostic performance of the final CAD scheme [42]. In a small study by Armato [40], four of six lung cancers not detected by automated detection were non-solid and two were part-solid. A computerized scheme based on the application of artificial neural networks to selected texture features and Gaussian curve fitting features may hold promise for facilitating detection of localized sub-solid nodules in CT [42]. The CAD used in our study does not support the detection of nodules with non-solid components. Furthermore, the number of sub-solid nodules was small.

Evaluation of the benefit of new CAD systems with improved sensitivity for subsolid nodules should be conducted in future studies with larger numbers of subsolid nodules.

A limitation of our study is that nodule diagnosis was in most cases (intermediate-sized nodules) not directly proven by biopsy but by evaluation of nodule growth on a short-term follow-up CT examination. However, the aim of the current study was to assess the performance of CAD versus double reading by human readers for detecting potentially relevant pulmonary nodules, which is the first step on the road to diagnosing early stages of lung cancer. The reference standard for defining the presence of a pulmonary nodule was an experienced consensus panel. Reader experience and variability could have affected the results. However, by using the interpretation of the sum of all findings by a consensus panel as the reference standard the effect of reader variability was reduced if not minimized. Also, the consensus panel did not perform a free search for potential additional findings. It is theoretically possible that the consensus panel could have found one or more additional pulmonary nodules. However, in view of the extremely high sensitivity of CAD for pulmonary nodules [32, 34], this was considered unlikely. As a result of the small numbers of certain nodule types, logistic regression could not be reliably performed for all nodule characteristics. Although we have demonstrated the benefits of CAD complementary to double reading compared to double reading alone, timing of the two modes still can be investigated; this actually depends on the efficiency of the workflow for CAD mark review. Lastly, the current analysis was based on a certain type of CAD software and a specific protocol for double reading and nodule evaluation. Whether the results can be generalized to other types of CAD software was not deter-mined; however, the results are in line with previous reports on the improved sensitivity of pulmonary nodule detection by CAD compared to that of human readers [15, 32].

In conclusion, using a combination of CAD and nodule size cut-off in lung cancer screening improves the sensitivity of pulmonary nodule detection compared to that of double reading, without missing lung cancers. Adding a nodule volume cut-off of 50 mm^3 to CAD leads to nearly half the FP rate (1.9 versus 3.7 FP/CT) with an increase in positive predictive value.

References

1. Diederich S, Wormanns D, Semik M et al (2002) Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. Radiology 222:773-781
2. Henschke CI, McCauley DI, Yankelevitz DF et al (1999) Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 354:99-105
3. Nawa T, Nakagawa T, Kusano S, Kawasaki Y, Sugawara Y, Nakata H (2002) Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies. Chest 122:15-20
4. The National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD et al (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365:395-409
5. Wormanns D, Ludwig K, Beyer F, Heindel W, Diederich S (2005) Detection of pulmonary nodules at multirow-detector CT: effectiveness of double reading to improve sensitivity at standard-dose and low-dose chest CT. Eur Radiol 15:14-22
6. Wang Y, Van Klaveren R, de Bock GH et al (2012) No benefit for consensus double reading at baseline screening for lung cancer with the use of semiautomated volumetry software. Radiology 262:320-326
7. Brown J, Bryan S, Warren R (1996) Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms. BMJ 312:809-812
8. Leivo T, Salminen T, Sintonen H et al (1999) Incremental cost-effectiveness of double-reading mammograms. Breast Cancer Res Treat 54:261-267
9. Lee IJ, Gamsu G, Czum J, Wu N, Johnson R, Chakrapani S (2005) Lung nodule detection on chest CT: evaluation of a computer-aided detection (CAD) system. Korean J Radiol 6:89-93
10. Peldschus K, Herzog P, Wood SA, Cheema JI, Costello P, Schoepf UJ (2005) Computer-aided diagnosis as a second reader: spectrum of findings in CT studies of the chest interpreted as normal. Chest 128:1517-1523

11. Wormanns D, Beyer F, Diederich S, Ludwig K, Heindel W (2004) Diagnostic performance of a commercially available computer-aided diagnosis system for automatic detection of pulmonary nodules: comparison with single and double reading. Rofo 176:953-958
12. Teague SD, Trilikis G, Dharaiya E (2010) Lung nodule computer-aided detection as a second reader: influence on radiology residents. J Comput Assist Tomogr 34:35-39
13. White CS, Pugatch R, Koonce T, Rust SW, Dharaiya E (2008) Lung nodule CAD software as a second reader: a multicenter study. Acad Radiol 15:326-333
14. Beyer F, Zierott L, Fallenberg EM et al (2007) Comparison of sensitivity and reading time for the use of computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. Eur Radiol 17:2941-294
15. Roos JE, Paik D, Olsen D et al (2010) Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance. Eur Radiol 20:549-557
16. Rubin GD, Lyo JK, Paik DS et al (2005) Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology 234:274-283
17. Swensen SJ, Jett JR, Sloan JA et al (2002) Screening for lung cancer with low-dose spiral computed tomography. Am J Respir Crit Care Med 165:508-513
18. Henschke CI, Yankelevitz DF, Naidich DP et al (2004) CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. Radiology 231:164-168
19. Swensen SJ, Jett JR, Hartman TE et al (2003) Lung cancer screening with CT: Mayo Clinic experience. Radiology 226:756-761
20. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI (2000) Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 217:251-256
21. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C (2006) Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. Eur Radiol 16:781-790

22. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G (2004) Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? radiology 231:453-458
23. Xu DM, Gietema H, de Koning H et al (2006) Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 54:177-184
24. Bogoni L, Bi J, Florin C et al (2010) Lung nodule detection. In: Croft WB, Müller H, Clough P, Deselaers T, Caputo B (eds) ImageCLEF. Information retrieval, vol 32. Springer, Berlin, pp 415-434
25. Jankowski A, Martinelli T, Timsit JF et al (2010) Pulmonary nodule detection on MDCT images: evaluation of diagnostic performance using thin axial images, maximum intensity projections, and computer-assisted detection. Eur Radiol 17:3148-3156
26. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K et al (2009) Management of lung nodules detected by volume CT scanning. N Engl J Med 361:2221-2229
27. Xu DM, van der Zaag-Loonen HJ, OudkerkM et al (2009) Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. Radiology 250:264-272
28. Beigelman A (2007) Computed-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. AJR Am J Roentgenol 189:948-955
29. Bellotti R, De Carlo F, Gargano G et al (2007) A CAD system for nodule detection in low-dose lung CTs based on region growing and a new active contour model. Med Phys 34:4901-4910
30. Brown MS, Goldin JG, Rogers S et al (2005) Computer-aided lung nodule detection in CT: results of large-scale observer test. Acad Radiol 12:681-686
31. Das M, Muhlenbruch G, Heinen S et al (2008) Performance evaluation of a computer-aided detection algorithm for solid pulmonary nodules in low-dose and standard-dose MDCT chest examinations and its influence on radiologists. Br J Radiol 81:841-847

32. Fraioli F, Bertoletti L, Napoli A et al (2007) Computer-aided detection (CAD) in lung cancer screening at chest MDCT: ROC analysis of CAD versus radiologist performance. J Thorac Imaging 22:241-246
33. Giger ML, Bae KT, MacMahon H (1994) Computerized detection of pulmonary nodules in computed tomography images. Invest Radiol 29:459-465
34. Ozekes S, Osman O, Ucan ON (2008) Nodule detection in a lung region that's segmented with using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding. Korean J Radiol 9:1-9
35. Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W (2002) Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. Eur Radiol 12:1052-1057
36. Goo JM, Lee JW (2003) Automated lung nodule detection at low-dose CT: preliminary experience. Korean J Radiol 4:211-216
37. Naidich DP, Rusinek H, McGuinness G, Leitman B, McCauley DI, Henschke CI (1993) Variables affecting pulmonary nodule detection with computed tomography: evaluation with three-dimensional computer simulation. J Thorac Imaging 8:291-299
38. Rusinek H, Naidich DP, McGuinness G et al (1998) Pulmonary nodule detection: low-dose versus conventional CT. Radiology 209:243-249
39. Yuan R, Vos PM, Cooperberg PL (2006) Computer-aided detection in screening CT for pulmonary nodules. AJR Am J Roentgenol 186:1280-1287
40. Armato SG 3rd, Li F, Giger ML, MacMahon H, Sone S, Doi K (2002) Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program. Radiology 225:685-692
41. Marten K, Engelke C, Seyfarth T, Grillhosl A, Obenauer S, Rummeny EJ (2005) Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. Clin Radiol 60:196-206
42. Kim KG, Goo JM, Kim JH et al (2005) Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: initial experience. Radiology 237:657-661

5

Comparison of Three Software Systems for Semi-Automatic Volumetry of Pulmonary Nodules on Baseline and Follow-Up CT Examinations

Accepted by Acta Radiologica

Yingru Zhao[1,2] / Peter M.A. van Ooijen[1,2] / Monique D. Dorrius[1,2]
Marjolein Heuvelmans[1] / Geertruida H. de Bock[3]
Rozemarijn Vliegenthart[1,2] / Matthijs Oudkerk[1]

University of Groningen, University Medical Center Groningen,
[1]Center for Medical Imaging – North East Netherlands,
[2]Department of Radiology, [3]Department of Epidemiology

Abstract

Purpose

To compare volumetric measurements of solid pulmonary nodules on baseline and follow-up CT scans as well as the volume doubling time (VDT) for three software packages.

Materials and Methods

From a Lung Cancer Screening study (NELSON), 50 participants were randomly selected from the baseline round. The study population comprised participants with at least one pulmonary nodule at the baseline and consecutive CT examination. The volume of each nodule was determined for both time points using three semi-automated software packages (P1, P2 and P3). Manual modification was performed when automated assessment was visually inaccurate. VDT was calculated to evaluate nodule growth. Volume, VDT and nodule management were compared for the three software packages, using P1 as the reference standard.

Results

In 25 participants, 147 nodules were present on both examinations (volume: 12.0 to 436.6 mm^3). Initial segmentation at baseline was evaluated to be satisfactory in 93.9% of nodules for P1, 84.4% for P2, and 88.4% for P3. Significant difference was found in measured volume between P1 and the other two packages ($p < 0.001$). P2 overestimated the volume by 38 ± 24%, and P3 by 50 ± 22%. At baseline, there was consensus on nodule size categorization in 80% for P1&P2 and 74% for P1&P3. At follow-up, consensus on VDT categorization was present in 47% for P1&P2 and 44% for P1&P3.

Conclusion

Software packages for lung nodule evaluation yield significant differences in volumetric measurements and VDT. This variation affects the classification of lung nodules, especially in follow-up examinations.

Introduction

Early diagnosis of lung cancer in a treatable stage is the main purpose of lung cancer screening by computed tomography (CT). Besides morphological characteristics of a lung nodule, nodule size is an important factor for predicting the risk of malignancy. In ongoing low-dose CT lung cancer screening studies in high-risk populations, the prevalence of cancer varied between 0.1-1% for nodules less than 5 mm, 1-30% for nodules measuring 5-10 mm, and 30-80% for nodules over 10 mm [1]. Many indeterminate lesions (5-10 mm diameter, or 50-500 mm^3) detected in low dose CT are not suitable for immediate evaluation by positron emission tomography, contrast-enhanced CT, or percutaneous needle biopsy. Thus, they are often re-scanned after an interval to assess growth.

Based on nodule volume, volume doubling time (VDT) was introduced as a crucial diagnostic tool to differentiate between malignant and benign nodules, especially in case of subcentimeter lesions [2]. The VDT of most benign pulmonary nodules was found to be more than 450 days, whereas VDT of malignant lesions was usually less than 400 days [3]. Furthermore, the prognosis of lung cancer correlates well with the tumor growth rate [3,4]. Therefore, in lung cancer screening trials, nodule volumetry is increasingly used for follow-up of indeterminate nodules in order to detect growth and thus, identify lesions with an increased risk of malignancy [5].

Accurate three-dimensional (3D) size and growth measurements are essential in the assessment of those indeterminate nodules. Variability in volumetric results may cause false-positive or false-negative diagnosis. Nowadays, nodule volumes can be calculated by using semi-automated volumetric software. First, automated techniques are used to make two-dimensional (2D) measurements on a single CT section. This method is followed by the 3D estimation of nodule volume and growth. However, substantial variations in segmentation performance are reported between current lung nodule software packages [6]. Systematic differences in volume measurements between packages could influence nodule categorization and treatment decisions. Therefore, high accuracy of segmentation is a requisite in software volumetric evaluation.

The purpose of this study is to assess the influence of lung nodule volumetric software packages on volume measurement at baseline and on change in volume over time. Additionally the impact of the software results is investigated on subsequent decisions concerning nodule management.

Materials and Methods

Study population

The subjects of this study were participants of the Dutch-Belgian randomised trial for lung cancer screening (NELSON) who underwent baseline screening for lung cancer by low-dose multi-detector CT [7]. The mean (±SD) age of the screened participants was 59 ± 6 years, and the mean number of pack-years smoked was 42 ± 19 [5]. The NELSON study was approved by the Medical Ethical Committees of all participating institutions and all subjects provided their written informed consent [7] that also covered the current analysis.

In this study, 50 participants with at least one consecutive scan were randomly selected from the baseline round of the NELSON study.

CT scanning protocol

The four participating screening sites all used 16-detector CT scanners (3 Sensation-16, Siemens Medical Solutions, Forchheim, Germany and 1 Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA). Scanning of the entire chest was performed in caudal-cranial direction. Scan data were obtained in spiral mode, with 16 × 0.75 mm collimation and pitch of 1.3. No contrast medium was administered. Low-dose settings were applied depending on body weight (< 50 kg, 50-80 kg and > 80 kg), with kVp settings of 80-90, 120 and 140 kVp, respectively, to achieve a Computed Tomography Dose Index Volume of approximately 0.8 mGy, 1.6 mGy and 3.2 mGy. The mAs settings were adjusted accordingly, depending on the acquisition system used. To minimize breathing artefacts, scans were performed at suspended maximal inspiration after appropriate instruction of the subjects. Data were reconstructed at 1.0 mm slice thickness, with 0.7 mm reconstruction increment and soft kernel (Siemens B30 filter). The Siemens B30

kernel is the standard soft tissue reconstruction kernel. Repeat scans were performed with the same technical parameters as used for the baseline scans in low-dose setting.

Nodule measurement

Due to the fact that current software packages do not support volume measurement of nodules with non-solid components, only solid nodules were included in this study. All nodule measurements were performed once by a radiologist with experience in reading thoracic CT scans for over 10 years. In a previous investigation we have shown that volume disagreement in case of repeated measurements, using semi-automated software, is negligible for smooth and spherical nodules (which constituted 95% of the detected nodules) [8]. Thus, we did not repeat measurement of the nodule volumes.

Nodule volumes were measured by using three different semi-automated software packages: Syngo LungCARE (Somaris / 5VB 10A, Siemens, Forchheim, Germany), OncoTREAT (v1.6, MEVIS, Bremen, Germany) and Vitrea (v2.1, Vital images, Minneapolis, MN, USA). P1 was assigned to the LungCARE software package, which is used for nodule evaluation in the NELSON study. Recent results from a phantom study indicate that the differences between the measured results obtained by LungCARE software and the actual volumes are very small (on average -4.9%) (9). We therefore chose to use P1 as reference for comparison. This choice is made for easy comparison of the other software packages and does not imply that one of the software packages performs better than the other since the actual volume of the nodules in screened subjects is unknown. The results from the other two packages were compared to P1. For the purpose of anonymization, the characters P2 and P3 were randomly assigned to the other two packages.

In all algorithms, initial segmentation was started by clicking in the center of a nodule. After a fully automated evaluation, the software produced a visual 3D presentation of the detected nodule highlighting the voxels of the nodule for which the initial volume was calculated. The segmented region was displayed by a colored overlay on the nodule or a colored line indicating the edge of the nodule.

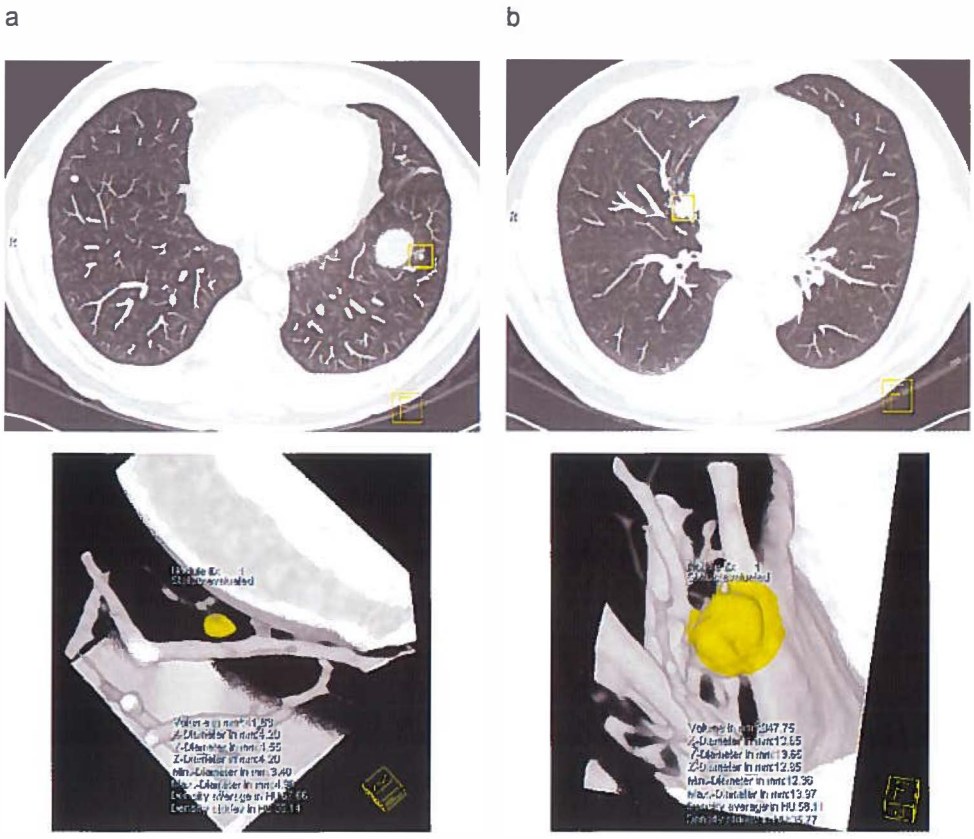


Figure 1 Example of initial segmentation of a lung nodule in 2 screening participants. Low-dose, non-contrast medium enhanced CT of the thorax. Display as maximum-intensity projection (MIP) transverse section (upper images), and as volume rendered image (lower images), both within software package P_1. Satisfactory initial segmentation (**a**). Unsatisfactory initial segmentation (**b**).

The segmentation was judged visually by the radiologist for accuracy and qualitatively categorized as satisfactory or unsatisfactory based on the perceived reliability of volume measurement [5]. Segmentation was assigned satisfactory if the segmented region matched the nodule well and included no surrounding structures such as vessels and pleura, or if the mismatched region between the segmented region and nodule was visually evaluated not to exceed 20% in volume; otherwise, it was considered unsatisfactory (Figure 1).

For unsatisfactory segmentations, all packages offered the possibility to modify the initial segmentations manually. Modified volume instead of initial volume was regarded as final measuring volume when manual modification was performed, assuming that the modified segmentation was more correct than the initial segmentation.

All nodules were measured on baseline and follow-up scans. A comparison of nodule volume was made between two time points. VDT based on change in calculated volumes over time was determined according the formula 1:

$$VDT(days) = \frac{[\ln 2 \times \Delta t]}{[\ln(v_1 \div v_0)]} \quad (1)$$

(Δt: the days of interval between baseline and follow up scans)

In our study, the VDTs were calculated based on nodule final measuring volumes.

Nodule categorization and characteristics

According to the NELSON protocol, solid nodules were classified into 3 categories based on their volumes at baseline (negative: < 50 mm^3, indeterminate: 50-500 mm^3, or positive: > 500 mm^3) and 3 categories based on their VDT at follow-up (GROWCAT A: > 600 days, GROWCAT B: 400-600 days, or GROWCAT C: < 400 days) [7].

Additionally, nodules were classified into four subgroups based on their location (peripheral or non-peripheral), attachment (intraparenchymal, fissure-attached, vessel-attached or pleural-based), shape (spherical or non-spherical), and edge (smooth or non-smooth) [7,10]. Nodules were classified as peripheral when the distance to the thoracic wall was less than one third of the total distance to the lung hilum, and non-peripheral otherwise. A nodule was defined as intraparenchymal when a nodule was surrounded by lung parenchyma. A nodule was considered attached when the length of the contact surface with other pulmonary structures (fissure, vessel or pleura) was more than 50% of the diameter of the nodule at volume-rendered reconstruction or on transverse images. A nodule was regarded spherical when its maximum diameter was smaller than twice its minimum diameter; otherwise, it was considered non-spherical. A nodule was regarded non-smooth when its margin was lobulated, irregular or spiculated.

Data analysis

The rate of satisfactory segmentation was compared between P1 and the other two software packages (P2 and P3) by using the Fisher exact test. The Kolmogorov-Smirnov test was used to determine the normality of the distribution of nodule volumes. Nodule volume and VDT were compared between P1&P2 and between P1&P3 by using the Wilcoxon signed ranks test. The percent difference in volume between P1 and the other two packages was defined as a percentage of the difference between two measuring volumes divided by the mean of the two values. The kappa test was used to determine the agreement for nodule size categories and growth categories between P1 and the other two software packages (P2 and P3). All statistical analyses were performed using SPSS software (version 18.0; SPSS, Chicago, IL, USA).

Table 1 Overview of satisfactory initial segmentation of baseline lung nodules by software package P_1, P_2 and P_3, according to nodule characteristics

	n	Satisfactory initial segmentation			p-value	
		P_1	P_2	P_3	P_1 & P_2	P_1 & P_3
Location						
Peripheral	125	120 (96.0)	106 (84.8)	110 (88.0)	n.s.	<0.05
Non-peripheral	22	18 (81.8)	18 (81.8)	20 (90.9)	<0.05	n.s.
Attachment						
Intraparenchymal	78	76 (97.4)	73 (93.6)	78 (100.0)	n.s.	n.a.
Fissure-attached	11	9 (81.8)	11 (100.0)	11 (100.0)	n.a.	n.a.
Vessel-attached	33	28 (84.8)	18 (54.5)	17 (51.5)	n.s.	n.s.
Pleural-based	25	25 (100.0)	22 (88.0)	24 (96.0)	n.a.	n.a.
Shape						
Spherical	141	133 (94.3)	120 (85.1)	125 (88.7)	<0.05	<0.01
Non-spherical	6	5 (83.3)	4 (66.7)	5 (83.3)	n.s.	n.s.
Edge						
Smooth	141	134 (95.0)	121 (85.8)	126 (89.4)	<0.05	<0.01
Non-smooth	6	4 (66.7)	3 (50.0)	4 (66.7)	n.s.	n.s.
Total	147	138 (93.9)	124 (84.4)	130 (88.4)	<0.01	<0.05

Unless otherwise indicated, data are numbers of nodules, with percentages in parentheses. P-values for agreement based on Fisher exact test, comparing P_1&P_2, and P_1&P_3. n.s. = not significant. n.a. = not available (due to cell with 100% value).

Results

In the 50 participants randomly selected from the baseline round of the NELSON study, 44 subjects (38 male / 6 female) had at least one lung nodule on the baseline CT examination. 147 nodules in 25 participants were found on both baseline and follow-up CT examinations. The mean (± SD) age of 25 participants was 62 ± 7 years. The number of nodules per subject ranged from 1 to 19. Thirteen participants had more than 5 nodules. Of the 147 nodules, 125 were located in the periphery, 141 were spherical and 141 had a smooth edge (Table 1). Besides 78 intraparenchymal nodules, 11, 33 and 25 nodules were fissure-attached, vessel-attached and pleural-based, respectively. Initial segmentation at baseline was evaluated to be satisfactory in 93.9% of 147 nodules using software P_1, 84.4 % using P_2, and 88.4% using P_3. Significant difference was found in percentage of satisfactory segmentation between the three packages ($p < 0.01$), based on evaluation of agreement in the nodules that were satisfactorily segmented by the different software packages. Only half of vessel-attached nodules were satisfactorily segmented by software P_2 (54.5%) and P_3 (51.5%) versus 84.8% for P_1. Especially in case of smooth and spherical nodules, the reference software yielded a better initial segmentation rate than P_2 and P_3 (Table 1).

Table 2 Comparison of nodule volume at baseline and volume doubling time (VDT) at follow-up between three software packages

Software package	Median Initial volume*, mm^3	Median final volume*, mm^3	Median VDT, days
P_1	37.7 (21.4-62.8)	37.3 (21.4-64.7)	297 (-590.3-777.5)
P_2	56.0 (38.0-89.0)	52.0 (37.0-87.0)	356 (-599.3-914.2)
P_3	64.4 (42.7-108.4)	62.2 (427.-93.8)	265 (-670.8-918.8)

Unless otherwise indicated, data are volumes of nodules with 25th and 75th percentile in parentheses.

* Wilcoxon signed ranks test; comparison between P_1&P_2, and between P_1&P_3; for both comparisons, p value less than 0.001

The distribution of the nodule volumes was skewed for all three software packages (Kolmogorov-Smirnov test: p<0.001). The median of the initial and final baseline volume as well as the VDT was determined for all three software packages (Table 2). The mean percent difference between initial and final volume was 0.5% for software P1, 4.9% for P2 and 7.2% for P3. Significant difference was found in the volume between P1 and the other two packages ($p < 0.001$). Compared to the reference software P1, P2 overestimated the volume by 38 ± 24%, and P3 by 50 ± 22%. No significant difference was found in VDT between P1&P2 and P1&P3; however the interquartile range in VDT was wide.

The median of the final baseline volume for the three software packages by nodule characteristics is shown in Table 3. For nearly all nodule characteristics, the volume derived from P2 and P3 was significantly larger than that obtained for the reference software. Only for non-smooth nodules (in total 6 in the current selection), no significant volume difference was found.

Table 3 The median final volume on baseline scans for the software packages by nodule characteristics.

Nodule characteristic	*n*	Median final volume, mm^3			P-value	
		P_1	P_2	P_3	P_1&P_2	P_1&P_3
Location						
Peripheral	125	46.1 (21.3-61.1)	64.3 (37.8-86.0)	73.4 (42.5-93.8)	<0.001	<0.001
Non-peripheral	22	79.4 (29.3-101.3)	89.9 (36.8-111.5)	98.8 (41.9-132.4)	<0.01	<0.01
Attachment						
Intraparenchymal	78	45.7 (20.3-52.1)	63.3 (36.8-78.3)	70.0 (41.5-85.6)	<0.001	<0.001
Fissure-attached	11	58.7 (16.7-87.6)	87.0 (35.0-124.0)	90.4 (37.6-138.6)	<0.01	<0.01
Vessel-attached	33	60.9 (31.5-76.5)	71.0 (42.0-92.5)	92.1 (56.3-115.3)	<0.01	<0.001
Pleural-based	25	51.6 (23.0-63.3)	71.1 (35.5-90.5)	73.9 (39.6-93.8)	<0.001	<0.001
Shape						
Spherical	141	50.6 (21.7-63.7)	67.2 (37.0-87.0)	76.3 (42.7-93.8)	<0.001	<0.001
Non-spherical	6	62.8 (18.8-111.2)	88.8 (40.8-160.8)	97.4 (36.8-201.8)	<0.05	<0.05
Edge *						
Smooth	141	46.4 (21.1-60.1)	64.0 (37.0-83.5)	72.8 (42.0-91.7)	<0.001	<0.001
Non-smooth	6	161.8(53.1-247.2)	165.3 (78.0-237.0)	180.9 (70.4-254.9)	n.s.	n.s.

Data are median volumes of nodules with 25th and 75th percentile in parentheses.

Wilcoxon Signed ranks test was used to compare the median final volume between P_1&P_2 and P_1&P_3 within every individual nodule characteristic.

Table 4 shows the categorization of pulmonary nodules according to size and VDT, and agreement between software packages. At baseline, there was consensus on nodule size categorization for 117 nodules (79.6%) when comparing software P1& P2 and 109 (74.1%) when comparing P1& P3 (Table 4). At follow-up, P1& P2 showed consensus on nodule VDT categorization in 40 (47.1%) nodules, and P1& P3 in 38 (44.2%) nodules. There was moderate agreement for nodule size categorization at baseline and growth determination at follow-up between P1 and the other two software packages (P2 and P3). The categorization according to VDT at follow-up showed fair agreement between P1&P2 and P1&P3.

Table 4 Distribution of nodule size categories and growth categories according to three software packages

	P_1	P_2	P_3	Consensus* between P_1&P_2	Consensus* between P_1&P_3	Kappa-value		p-value	
						P_1&P_2	P_1&P_3	P_1&P_2	P_1&P_3
Size category at baseline						0.598	0.514	<0.001	<0.001
Negative (< 50 mm^3)	96	70	58	68	58				
Indeterminate (50-500 mm^3)	51	77	89	49	51				
Positive (> 500 mm^3)	0	0	0	0	0				
Growth in follow-up						0.539	0.427	<0.001	<0.001
No growth	62	61	61	45	41				
Growth	85	86	86	69	65				
VDT in follow-up						0.388	0.330	<0.001	<0.001
GROWCAT A (>600 days)	43	52	51	23	23				
GROWCAT B (400-600days)	23	13	13	3	1				
GROWCAT C (<400 days)	19	21	22	14	14				

Unless otherwise indicated, data are numbers of nodules.

* Consensus in categorization between software packages

Discussion

This study shows that satisfactory nodule segmentation was automatically obtained in the majority of lung nodules for all three software packages. We recently found only very small differences between nodule volumes for software package P1 and actual volumes in a phantom study [9]. Thus, P1 was used as reference for

comparison. The measured nodule volumes were significantly different for the other two software packages with overestimation of nodule volumes of 38% to 50% compared to the reference software package. Despite volume differences, agreement in baseline categorization of nodules according to volume was still obtained in 74-80% of cases. However, the consensus on the VDT categorization at follow-up was less than 50 % when P2 and P3 were compared to the chosen reference, P1.

Because of recent reports on the value of VDT for the determination of malignancy [4, 11-13], there is now widespread interest in the use of volumetric software. Compared to nodule volumetry, the repeatability of diameter measurements has been found to be suboptimal [14,15]. Likely, this is related to the fact that two-dimensional measurements are relatively insensitive to size change. For example, 1 mm increase in the cross-sectional diameter of a 10-mm spherical nodule on two consecutive scans corresponds to a 10% increase in diameter but a 33% increase in volume. If the two examinations were performed 3 months apart, the volume increase indicates a VDT compatible with malignancy (VDT <400 days). However, the diameter change would still be considered to fall in the range of measurement variation. In addition, volume changes estimated from two-dimensional measurements may miss asymmetric growth [16]. The most common approach for measuring the volume of lung nodules is based on a grey-level threshold, that allows a user to segment lung nodules from the background voxels [17]. Since attenuation information is not sufficient to distinguish nodule boundaries from attached vessels, most semi-automated methods incorporate morphologic operators [18].

The only article so far that has assessed different software packages for volumetric evaluation of lung nodules, is by de Hoop et al. [6]. Both in our study and in the study by de Hoop et al., the software types were anonymized. Thus, we cannot directly compare our results for specific software packages to those of de Hoop. In this study, six software packages were compared, in which the accuracy of segmentation varied from 71% to 86% prior to manual correction. In our study, initial segmentation by the three software packages was visually judged to be satisfactory for most of the nodules (range, 84.4%-93.9%). The segmentation rates

for software packages in the study by de Hoop were somewhat lower than in our study. This is likely due to a higher percentage of smooth, round, intraparenchymal nodules in our study. Although the visual appearance of segmentation was correct in the majority of cases for all software packages in our study, P_2 and P_3 resulted in larger nodule volumes when compared to P_1. The most likely reason for the difference in volume lies in differences in segmentation of border tissue where a 1 pixel increase in segmented surface can already have considerable implications for the measured volume.

It is obvious that segmentation that includes surrounding structures or does not include part of a nodule may lead to inaccurate measurements and wrong management decisions [6]. A considerable part of segmentation errors in completely automatic measurement is due to adjacent structures. For example, juxtapleural and juxtavascular nodules have been shown to exhibit higher volume measurement variability than well-circumscribed intraparenchymal nodules [8,15]. In this study, for two of the three software packages, the segmentation of nearly half of the vessel-attached nodules was initially unsatisfactory. Compared to purely-intraparenchymal nodules, attached nodules have been found to have a lower risk of malignancy. In a NELSON sub-study by Xu et al. [10] reported that no malignancies were found at 1-year follow-up in smooth or attached solid indeterminate non-calcified nodules. As cancer risk may be negligible the importance of measurement variability in attached nodules is likely less in the clinical practice. In some studies, irregular or spiculated nodules showed increased volume measurement variability compared to smooth nodules [19,20]. We found that the percentage of satisfactory segmentations of non-smooth nodules was low for all three packages. However, only 6 non-smooth nodules were included in the current study. Nevertheless, our further results indicate that nodule morphology influences volume measurement variability,

In the study by de Hoop et al. [6], significant differences in nodule volume were found among software packages, although the actual differences in volumetry between software packages were not reported in detail. Similarly, our results indicate that the volumes obtained from nodule evaluation software packages were significantly different. Recent results from a phantom study indicate

only a small underestimation of nodule volume (overall, -4.9%) by LungCARE software (P1) compared to actual volumes [9]. These differences between LungCARE results and actual volumes were an order of magnitude smaller than the differences found between LungCARE and the other two software packages in the current study. This may suggest that software P1 yields more accurate volumetry results than P2 and P3. However, the findings from the phantom study cannot be extrapolated with certainty to clinical practice, as the actual size of the detected lung nodules in subjects is unknown in contrast to a phantom situation. Thus, we cannot determine a reference standard based on our results. Our findings do point to a significant influence of software packages on nodule categorization and follow-up management. Using P2 and P3, more nodules ended up in a higher category of nodule management, potentially leading to unnecessary diagnostic procedures, patient anxiety, morbidity and costs. To reduce influence in categorization based on changes in software package, consecutive screening evaluations should be performed with the same semi-automated software package.

There is limited public knowledge of the algorithms used to perform nodule segmentation. As described in the Material and Methods, all software packages used a similar one-click approach for the segmentation. Based on published details about the different algorithms they all three rely on a quite similar and comparable multi-step approach [21-23]. The first step uses intensity/density based algorithm using the tissue densities in Hounsfield unit values combined with a region growing or connected components algorithm starting from the click point. The second step consists of constrained morphological operations that limit the actual volume and separate the nodule from the surrounding structures based on properties such as roundness. Based on the segmentation result, the volume is determined in a third and final step. In this final step the description of the OncoTreat software [22] deviates from the other two in that it specifically mentions the use of a variation of the partial volume method for volume calculation which involves a three region approach to avoid partial volume effects on the volume. Other differences in the actual measurement of the volume are most probably due to subtle differences in the thresholds and specific criteria used in the different steps of which some are also user definable and correctable when a non-optimal

segmentation occurs. A definite conclusion about the influence of the different steps of the algorithms used is however not feasible since the exact algorithms used to determine the segmented area are not disclosed by the respective companies.

On serial examinations, lung nodule measurements derived from CT are used to evaluate size change to predict the likelihood of malignancy and to monitor the interval of follow-up [3]. A change in size is considered highly important for suggesting a diagnosis or management for follow-up since malignant nodules generally grow fast. Although no significant difference in VDT was found between the software packages, the agreement in VDT categorization was only fair. Therefore, using different software packages could greatly influence nodule management decisions especially for intermediate-sized nodules on consecutive examinations. We consider validation and calibration of software for lung nodule volumetry of utmost importance to limit variability and inaccuracy in nodule management. The validity and reliability of software packages could be improved by calibration against a publicly available database of CT datasets with nodules of known size.

In this study all nodule measurements were performed by one radiologist with more than 10 years experience, because no significant benefit has been found for consensus double reading of lung nodule volume measurements [24] and the inter-observer variability is negligible by using semi-automated volumetry [25]. Furthermore, the intra- and inter-observer reproducibility of visual assessment of segmentation accuracy is high [6].

The current study concerns lung nodules found as part of a CT lung cancer screening program. Our results do not apply to a non-screening situation. We have previously shown that compared to size, morphological characteristics of screen-detected nodules are of minor importance to distinguish between benign and malignant nodules [10, 26]. However, in clinical decision making, other factors can play an important role. For example, a recent clinical study questioned the use of CT-derived growth rates of as the only determinant to decide whether or not to perform biopsy [27]. Further investigations are needed to evaluate the role of volumetry and VDT of pulmonary nodules in clinical settings.

The main limitation of the current study is the relatively small sample of individuals. However, the number of nodules was still considerable, and there was a fair distribution in the types of nodules commonly encountered in lung cancer CT screening. We suggest larger studies should confirm our findings. A second limitation is that we only included solid lung nodules. These are the most prevalent type of nodules in lung cancer screening. Management of solid nodules is based on size and VDT. Features that need to be assessed for determination of management of sub-solid nodules are still under investigation. The vast majority of the nodules in our study were smooth and round. Lung cancer cases more often have an irregular contour and lobulated or spiculated shape [26]. However, this reflects findings in lung cancer screening, as this was also the most prevalent nodule type in a larger sample of 658 screened individuals with 891 indeterminate lung nodules [10]. The fact that the percentage of smooth, round nodules was somewhat higher in the current study is considered to be due to random sampling. Also, as the pathology of the nodules was unknown, we cannot draw conclusions on the accuracy of VDT assessment for distinguishing between benign and malignant nodules. A study on optimization of VDT cutoff criteria was recently published [28]. Lastly, we only evaluated three software packages. Whether the results found for these software packages are a good representation of the results of other available lung nodule volumetry software, is unknown.

In conclusion, software packages for lung nodule evaluation yield significant differences in volumetric measurements and VDT. This variation affects the classification of lung nodules in baseline and, especially, follow-up examinations. Overestimation of volumetry may result in false-positive conclusions with potential serious consequences for the patient. Further standardization of software used for nodule volumetry and VDT assessment is needed to optimize nodule management in lung cancer CT screening.

References

1. Beigelman-Aubry C, Hill C, Grenier PA. Management of an incidentally discovered pulmonary nodule. Eur Radiol 2007;17:449-466.
2. Arai T, Kuroishi T, Saito Y, et al. Tumor doubling time and prognosis in lung cancer patients: evaluation from chest films and clinical follow-up study. Japanese Lung Cancer Screening Research Group. Jpn J Clin Oncol 1994;24:199-204.
3. Hasegawa M, Sone S, Takashima S, et al. Growth rate of small lung cancers detected on mass CT screening. Br J Radiol 2000;73:1252-1259.
4. Usuda K, Saito Y, Sagawa M, et al. Tumor doubling time and prognostic assessment of patients with primary lung cancer. Cancer 1994;74:2239-2244.
5. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. N Engl J Med 2009;361:2221-2229.
6. de Hoop B, Gietema H, van Ginneken B, et al.. A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. Eur Radiol 2009;19:800-808.
7. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 2006;54:177-184.
8. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. Radiology 2008;248:625-631.
9. Xie X, Zhao Y, Snijder RA, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. Eur Radiol 2013;23:139-147
10. Xu DM, van der Zaag-Loonen HJ, Oudkerk M, et al. Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. Radiology 2009;250:264-272.

11. Lillington GA. Management of solitary pulmonary nodules. Dis Mon. 1991;37:271-318.
12. Yankelevitz DF, Henschke CI. Does 2-year stability imply that pulmonary nodules are benign? AJR Am J Roentgenol. 1997;168:325-328.
13. Awai K, Fujikawa K, S. N. Serial changes in CT findings of small peripheral pulmonary adenocarcinomas followed up for more than one year. Jpn J Lung Cancer 1998;38:10.
14. Revel MP, Bissery A, Bienvenu M, et al. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? Radiology 2004;231:453-458.
15. Revel MP, Lefort C, Bissery A, et al. Pulmonary nodules: preliminary experience with three-dimensional evaluation. Radiology 2004;231:459-466.
16. Yankelevitz DF, Reeves AP, Kostis WJ, et al. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 2000;217:251-256.
17. Iwano S, Okada T, Koike W, et al. Semi-automatic volumetric measurement of lung cancer using multi-detector CT effects of nodule characteristics. Acad Radiol 2009;16:1179-1186.
18. Gavrielides MA, Kinnard LM, Myers KJ, et al. Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology 2009;251:26-37.
19. Petrou M, Quint LE, Nan B, Baker LH. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. AJR Am J Roentgenol 2007;188:306-312.
20. Winer-Muram HT, Jennings SG, Meyer CA, et al. Effect of varying CT section width on volumetric measurement of lung tumors and application of compensatory equations. Radiology 2003;229:184-194.
21. Wormans D, Kohl G, Klotz E, et al. Volumetric Measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 2004;14:86-92.
22. Bornemann L, Kuhnigk J-M, Dicken V, et al. New Tools for Computer Assistance in Thoracic CT Part 2. Therapy Monitoring of Pulmonary Metastases. Radiographics 2005;25:841-848.

23. Wilson DO, Ryan A, Fuhrman C, et al. Doubling Times and CT Screen-Detected Lung Cancers in the Pittsburgh Lung Screening Study. Am J Respir Crit Care Med 2012;185:85-89.
24. Wang Y, van Klaveren RJ, de Bock GH, et al. No benefit for consensus double reading at baseline screening for lung cancer with the use of semiautomated volumetry software. Radiology 2012;262:320-326.
25. Bolte H, Jahnke T, Schafer FKW, et al. Interobserver-variabilty of lung nodule volumetry considering different segmentation algorithms and observer training levels. Eur J Radiol 2007;64:285-295.
26. Xu DM, van Klaveren RJ, de Bock GH, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. Eur J Radiol 2008;68:347-352.
27. Korst RJ, Lee BE, Krinsky GA, Rutledge JR. The utility of automated volumetric growth analysis in a dedicated pulmonary nodule clinic. J Thorac Cardiovasc Surg 2011;142:372-377.
28. Heuvelmans MA, Oudkerk M, de Bock GH, et al. Optimisation of volume-doubling time cutoff for fast-growing lung nodules in CT lung cancer screening reduces false-positive referrals. Eur Radiol 2013;23:1836-1845.

6

Features of Resolving and Non-Resolving Indeterminate Pulmonary Nodules on Follow-Up CT: the NELSON Study

Accepted by Radiology

Yingru Zhao[12] / Marjolein A. Heuvelmans[12] / Monique D. Dorrius[12]
Peter M. A. van Ooijen[12] / Ying Wang[12] / Geertruida H. de Bock[3]
Matthijs Oudkerk[2] / Rozemarijn Vliegenthart[12]

University of Groningen, University Medical Center Groningen,
[1]Department of Radiology, [2]Center for Medical Imaging–North East Netherlands, [3]Department of Epidemiology,

Abstract

Purpose

To retrospectively identify features that predict disappearance of solid indeterminate (size 50-500 mm^3), intraparenchymal nodules detected at baseline in a lung cancer computed tomography (CT) screening study among individuals at high risk for lung cancer

Materials and Methods

The study was institutional review board approved. Participants gave informed consent. Participants with at least one non-calcified solid, indeterminate, intraparenchymal nodule (size 50-500 mm^3) at baseline were included (n=964 nodules in 750 participants). According to protocol, indeterminate nodules were re-examined by 3-month follow-up CT. Regular repeat screening rounds were at year 2 and 4. A nodule was defined as resolving if it had disappeared on a subsequent CT. Nodule resolution was regarded as spontaneous, not the effect of treatment. CT features of resolving nodules and non-resolving (stable and malignant) nodules were compared by generalized estimating equations analysis.

Results

During subsequent screening rounds, 10.1% (97/964) of the nodules disappeared, 77.3% (n = 75) of these before the 3-month CT and 94.8% (n = 92) before the second-round screening. Non-peripheral nodules were more likely to resolve than peripheral nodules (odd Ratio [OR]: 3.16; 95% CI: 1.76-5.70) Compared to smooth nodules, nodules with spiculated margin showed the highest probability of disappearance (OR: 4.36; 95% CI: 2.24-8.49).

Conclusions

About 10% of solid intraparenchymal pulmonary nodules of intermediate size found at baseline lung cancer screening resolves during follow-up. Three quarters of resolving nodules have already disappeared at the 3-month follow-up CT which is performed for intermediate size lung nodules. Resolving pulmonary nodules share CT features with malignant nodules.

Introduction

With the widespread use of multi-detector computed tomography (CT) in daily clinical practice and its use in lung cancer screening, the number of detected pulmonary nodules has increased as compared to standard chest X-ray examinations. Up to 66% [1] of participants enrolled in CT screening trials has at least one small-to-intermediate-size pulmonary nodule. Solid lung nodules are the most common type of nodules found in lung cancer screening [2, 3]. Most indeterminate nodules are benign [2], and may represent granulomatous or infectious lesions, or enlarged lymph nodes. The question arises whether it is possible to identify specific features of nodules that will subsequently resolve in order to avoid unnecessary repeat CT scans and work-up as well as public health costs and anxiety.

Four studies studied resolving nodules [4-7]. Only one focused on solid lung nodules [4]. A substantial part of the follow-up examinations in this study were performed with thicker slices, which may miss small rest lesions. Also, there was no comparison to non-resolving nodules.

The Dutch-Belgian randomised lung cancer screening trial (Dutch acronym NELSON) is the first in which nodule management is based on CT-derived volume and volume-doubling time (VDT) assessment [8]. Volumetric measurement is more accurate than diameter measurements [9, 10]. At baseline screening, a nodule with volume > 500 mm^3 led to referral for workup. For intermediate sized solid nodules (50–500 mm^3, corresponding to 4.6-9.8 mm in diameter), so-called indeterminate nodules, repeat CT was performed after 3 months. Indeterminate nodules were then divided into those that showed no or less than 25% growth, leading to a regular next-round screening, and those that showed greater than 25% growth, leading to referral for workup and diagnosis [8, 11]. Next-round CT results (year 2, 4 and 6) were based on volume measurements for newly detected nodules and growth evaluation of previously detected nodules.

The purpose of our study was to retrospectively identify features that predict complete resolution of non-calcified solid indeterminate, intraparenchymal nod-

ules detected at baseline in a lung cancer CT screening study among individuals at high risk for lung cancer.

Materials and Methods

Study population

This study was performed in the context of the NELSON trial, (trial registration number: ISRCTN63545820), which was approved by the Dutch Healthcare Committee and the ethics board at each participating center. All participants gave written informed consent at study entry. The current retrospective evaluation fell under the terms of the informed consent. Participants were between 50 and 75 years of age and were recruited via population registries through mail. Only current or former smokers with a smoking history of > 15 cigarettes/day for > 25 years or > 10 cigarettes/day for > 30 years were included. People with a history of pneumonectomy, breast cancer, melanoma or hypernephroma were not included. People with a history of other types of cancer were only eligible if they were curatively treated at least 5 years ago without signs of recurrence at the time of inclusion [12]. Participants underwent low-dose multidetector computed tomography (MDCT) screening at baseline (first round), 1 year later (second round) and 3 years later (third round), and received extra low-dose follow-up MDCT in case of an indeterminate lung nodule. Previously, the NELSON screening protocol was published in detail [8].

Our study was based on all baseline examinations of NELSON project. In total, 7,557 participants underwent baseline screening between April 2004 and December 2006 [2]. According to NELSON protocol, non-calcified solid nodules were classified into categories based on size [8]. A previous study showed that the rate of malignancy in attached indeterminate lung nodules was negligible [13]. Therefore, in the current study, only solid intraparenchymal (i.e. surrounded by lung parenchyma) nodules with volume between 50 – 500 mm^3 (i.e. intermediate size) at the baseline screening were included. Larger nodules were referred to the pulmonologist and smaller nodules did not receive extra follow-up. An indeterminate result led to an extra follow-up CT 3 months after baseline. If no or slow

growth of the nodule was found, subjects subsequently underwent the standard repeat screening examination. Indeterminate nodules without significant growth at least two years after baseline or with benign result on histological analysis were regarded as benign. Subjects with a fast growing nodule (volume doubling time [VDT] <400 days) were referred to pulmonologists for further diagnosis.

Nodules with less than 2 years follow-up after baseline, were excluded. Also participants with malignancies other than primary lung cancer were excluded.

CT scanning protocol

At all four screening sites 16-MDCT scanners were used (Sensation-16, Siemens Medical Solutions, Forchheim, Germany, or Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA). Scanning of the entire chest was performed in caudo-cranial direction. Scan data were obtained in spiral mode, with 16 × 0.75 mm collimation and 1.5 pitch. No contrast media was used. Low-dose settings were applied depending on body weight (<50 kg, 50-80 kg and >80 kg), with kVp settings of 80-90, 120 and 140 kVp, respectively, to achieve a volume CT dose index ($CTDI_{vol}$) of approximately 0.8 mGy, 1.6 mGy and 3.2 mGy, respectively. The mAs settings were adjusted accordingly, depending on the system used. To minimise breathing artefacts, CT-scans were performed at suspended maximal inspiration after appropriate instruction of the subjects. Data were reconstructed at 1.0-mm slice thickness, with 0.7-mm reconstruction increment. Repeat examinations were performed with the same technical parameters in low-dose setting as used at baseline.

Image Reading

All CT images were read twice independently [2, 8]. First readings were done by one of thirteen radiologists with experience in thoracic CT varying from 1 year to more than 20 years. Second readings were done by one of two radiologists (Y.W., Y.R.Z.) with at least 6 years experience. In case of discrepancy between first and second reader, a third radiologist (M.O.) with more than 15 years experience in thoracic CT arbitered. Discrepancy in nodule categorization between first and second reading was found for 43 lesions in 37 individuals.

The Syngo Lungcare© (Leonardo© workstation, Somaris/5 VB 10A Siemens Medical Solutions, Erlangen, Germany) software package designed to aid radiologists in diagnosing pulmonary nodules was used in addition to visual evaluation. Baseline and follow-up images were reviewed and displayed simultaneously on one workstation. Lung windows were assessed at a width of 1600 and a level of -700 Hounsfield Units. All images were interpreted both in lung window and mediastinal settings. First, the reader had to detect and then mark the pulmonary nodule by a mouse click. Subsequently, the program automatically defined the volume of interest of the nodule. A three-dimensional template was generated, optimally representing the nodule. If needed, manual modification of the segmentation was performed. A second mouse click initiated the automated volume measurement program. Semi-automated measurements are highly reproducible for the vast majority of nodules [14]. In 86% of >4000 screen-detected solid nodules, double reading yielded the same volume. Volume differences > 15% were found in only 4% of nodules [14]. If measured volume differed between first and second reader, results from the second reader were used for further analyses.

Nodule characteristics

Nodules were classified as benign or malignant based on histology or benign based on stable volume for > 2 years after baseline. In addition, they were classified based on distance to costal pleura (peripheral or non-peripheral), shape (spherical or non-spherical), and margin (smooth, lobulated, spiculated or irregular) [8, 13, 15].

The distance to costal pleural was < 1/3 from total distance hilum-costal pleura for peripheral nodules, > 1/3 for non-peripheral nodules. A nodule was regarded non-smooth when its margin was lobulated, irregular or spiculated, and smooth otherwise [15, 16]. A nodule was regarded spherical when its maximum diameter was smaller than twice its minimum diameter; otherwise, it was considered non-spherical.

Nodule resolution

At follow-up examinations, images were compared with the previous screening round. A nodule was defined as completely resolving if it had disappeared on a follow-up examination, otherwise, it was considered non-resolving. In the NELSON study, 25% change in nodule volume is used to differentiate real change from measurement variation [11]. Thus, decrease in volume ≥ 25% was regarded as actual decrease in size. Nodules that decreased in size, but did not disappear, were regarded as non-resolving as nodules that decrease at some point, can still eventually turn out to be malignant [17]. Nodule resolution was regarded as spontaneous, not the effect of any treatment, since antibiotic therapy was not part of the nodule management protocol.

Statistical Analysis

Generalized Estimating Equation analyses with a logit link function and a binomial distribution were performed to assess whether nodule characteristics were related to disappearance at 3-month follow-up CT, and at the regular screening CT rounds at year 2 and 4. For distance to costal pleura, shape and margin the odds ratios (ORs) and 95% confidence intervals (95%-CI) were estimated in univariate analyses. Then, multivariate analysis of the combined factors was performed, adjusting for the potential confounding effects of age and gender. Chi-square testing was used to compare the rate and timing of disappearance for nodules with maximal transverse diameter < 8 mm and ≥ 8 mm [18], based on semi-automated volumetry. A $p \leq 0.05$ was considered to indicate a statistically significant difference. All statistical analyses were performed using SPSS 20.0.

Results

At the baseline examination, 1059 solid intraparenchymal nodules with volume between 50 - 500 mm^3 were found in 805 participants. 95 nodules in 55 participants were excluded. A total of 964 nodules in 750 (648 men, 102 women) participants could be included (Figure 1). The mean (± standard deviation) age of these participants was 60 ± 6 years. Ninety-seven (10.1%) nodules in 75 participants disappeared during follow-up. In 61 of 75 participants, one resolving nodule was identified. The other subjects had 2-5 resolving nodules. Of the nodules that resolved, 75/97 (77.3%) had disappeared at 3 months, another 17/97 (17.5%) at year 2 screening, and another 5/97 (5.2%) at year 4 screening.

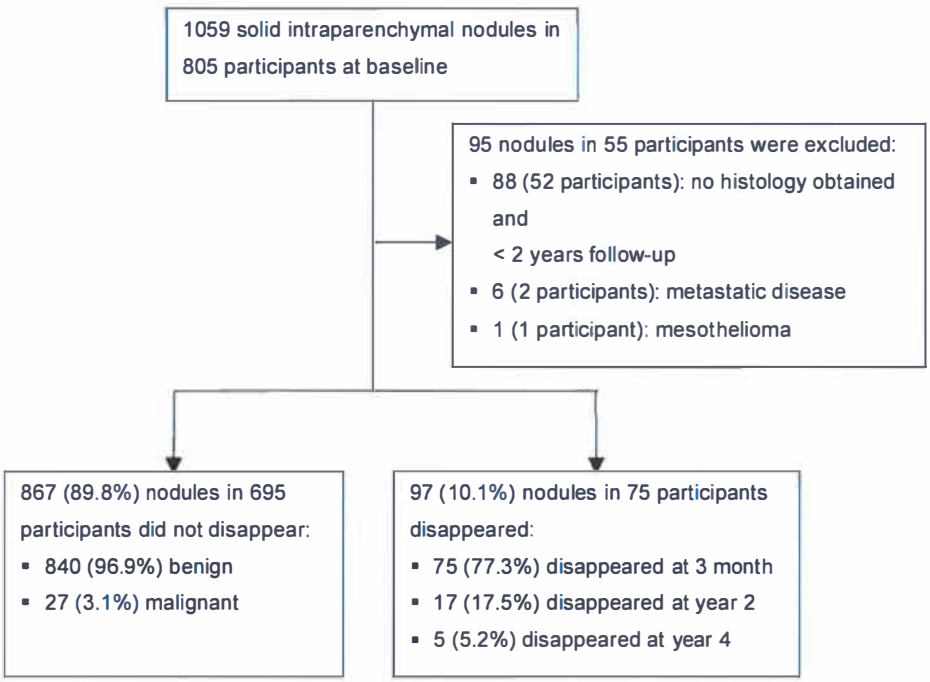


Figure 1 Overview of nodule selection from the NELSON study

Of 867 non-resolving indeterminate nodules at baseline, 105 (12.1%) showed >25% volume decrease during follow-up. Forty-seven (5.4%) nodules had additional workup due to short VDT in follow-up rounds. In 27 nodules (3.1%) lung cancer was diagnosed; 9 at the examination three months after baseline, 9 in the regular second-round examination, 2 one year after the regular second-round screen, 6 in the regular third-round examination and 1 one year after the regular third-round screen. The remaining 20 nodules comprised false-positive results; no malignancy was confirmed and the regular next-round CT scan was scheduled (Figure 1).

Table 1 Characteristics of resolving and non-resolving nodules (*n* in total=964)

	Total	Resolving	Non-resolving	Non-resolving benign	Non-resolving malignant
Total	964	97 (10.1)	867 (89.9)	840 (96.9)	27 (3.1)
Distance to costal pleural					
Peripheral	828	68 (8.2)	760 (91.8)	735 (96.7)	25 (3.3)
Non-peripheral	136	29 (21.3)	107 (78.7)	105 (98.1)	2 (1.9)
Shape					
Spherical	816	88 (10.8)	728 (89.2)	703 (96.6)	25 (3.4)
Non-spherical	148	9 (6.1)	139 (93.9)	137 (98.6)	2 (1.4)
Margin					
Smooth	680	53 (7.8)	627 (92.2)	618 (98.6)	9 (1.4)
Lobulated	195	22 (11.3)	173 (88.7)	165 (95.4)	8 (4.6)
Spiculated	63	17 (27.0)	46 (73.0)	36 (78.3)	10 (21.7)
Irregular	26	5 (19.2)	21 (80.8)	21 (100)	0 (0)

Unless otherwise indicated, data are numbers of nodules, with percentages in parentheses.

The characteristics of resolving and non-resolving nodules are shown in Table 1. Most lung nodules, whether resolving or not, were peripheral, smooth and spherical. Resolving nodules were more frequently non-peripheral than non-resolving nodules. Relatively less resolving nodules were smooth and more had a

spiculated margin. Non-resolving malignant nodules, however, also tended to be less often smooth, and more frequently spiculated (for examples of resolving nodules see Figure 2).

Table 2 Odds ratios showing the association between baseline characteristics and nodule disappearance of solid intraparenchymal nodules detected at baseline (*n* = 964)

Nodule characteristic at baseline	Odds Ratio*	
	Univariate	Multivariate
Distance to costal pleural	p < 0.001	p < 0.001
Peripheral	1	1
Non-peripheral	3.03 (1.73-5.29)	3.16 (1.76-5.70)
Shape	P = 0.12	P = 0.13
Spherical	1	1
Non-spherical	0.54 (0.24-1.19)	0.53 (0.23-1.21)
Margin	†	‡
Smooth	1	1
Lobulated	1.50 (0.83-2.73)	1.59 (0.89-2.86)
Spiculated	4.37 (2.25-8.48)	4.36 (2.24-8.49)
Irregular	2.82 (0.86-9.22)	3.12 (0.75-12.99)

* Data in parentheses are 95% Confidence Intervals. In the Multivariate Analyses, the nodule characteristics were included, as well as participant age and gender as potential confounders.

† Pairwise comparison within the univariate analysis showed a statistically significant difference between spiculated and lobulated nodules (P<0.01 (P=0.005)) and between spiculated and smooth nodules (P<0.001) ((Smooth lobulated P=0.18, smooth irregular P=0.09, lobulated irregular 0.32, speculated irregular 0.49)).

‡ Pairwise comparison within the multivariate generalized mixed model showed a statistically significant difference between spiculated and lobulated nodules (P<0.01 (P=0.008)) and between spiculated and smooth nodules (P<0.001) No significant differences were found between smooth and lobulated (P=0.12), smooth and irregular (P=0.12), lobulated and irregular (P=0.37), and speculated and irregular (P=0.66) nodules.

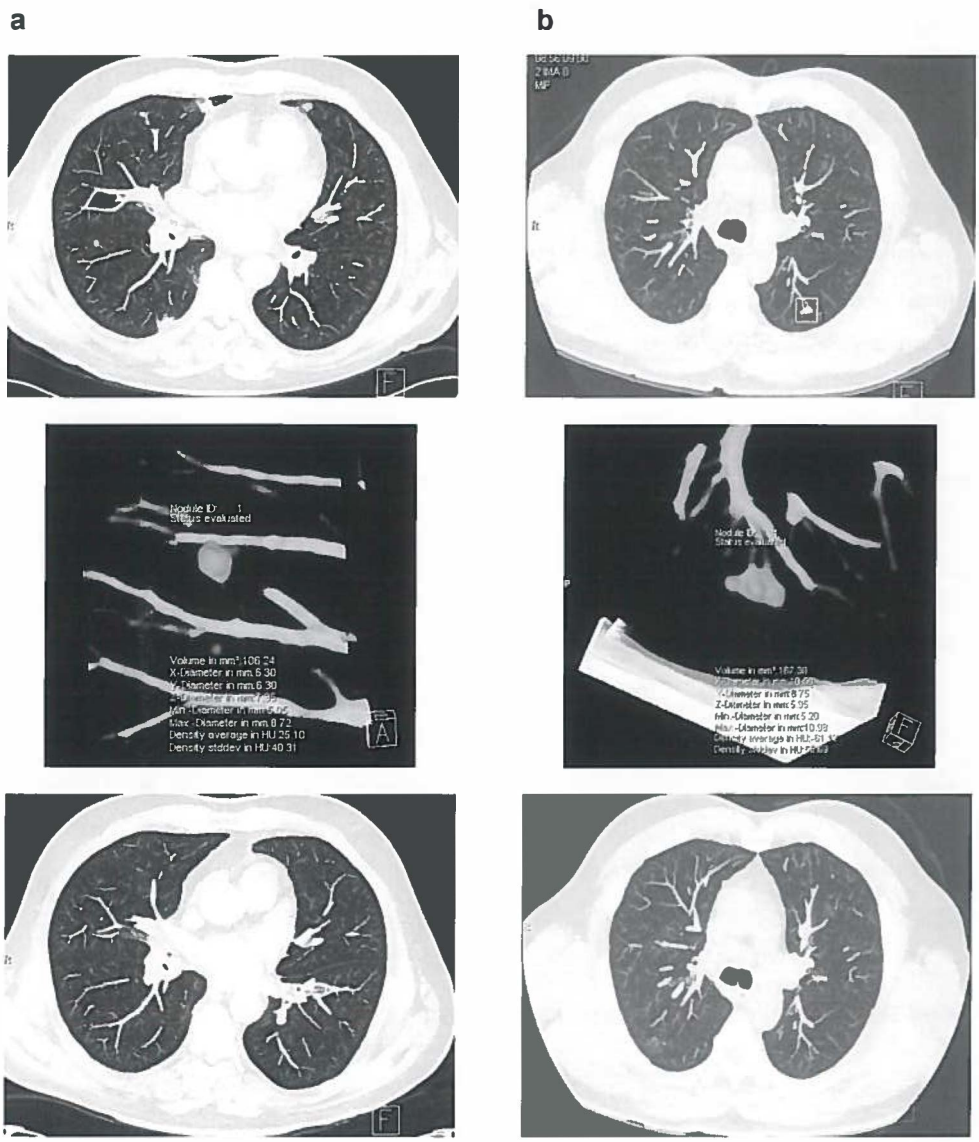


Figure 2 Examples of resolving nodules (**a**) a smooth and round nodule with baseline volume 106.2 mm^3 disappeared at 3-month follow up, (**b**) a lobulated nodule with baseline volume 167.3 mm^3 disappeared at 3-month follow up.

Odds ratios for factors related to nodule disappearance are shown in Table 2. Significant differences were found in distance to costal pleura and margin. Non-peripheral nodules had a 3.16 times higher chance to disappear at 3 months than

peripheral nodules (95%CI: 1.76-5.70). Pairwise comparison showed significant difference in probability of resolution between spiculated and lobulated nodules ($p < 0.01$) and between spiculated and smooth nodules ($p < 0.001$). Although not statistically significant, nodules with non-spherical shape tended to have a lower chance of resolution than spherical nodules (OR: 0.53; 95%CI: 0.23-1.21).

In analysis by maximum diameter (< 8 mm versus ≥ 8 mm), the rate of disappearance was lower in nodules < 8 mm than in nodules ≥ 8 mm (Table 3A). Larger nodules tended to disappear already in higher percentage before the short-term follow-up CT compared to the smaller nodules (Table 3B).

Table 3A Nodule resolution according to nodule size

	Total	Resolving	Non-resolving	Non-resolving malignant
Total	964	97 (10.1)	867 (89.9)	27 (3.1)
Maximal transverse diameter				
< 8 mm	751	58 (7.7)	693 (92.3)*	14 (2.0) †
≥ 8 mm	213	39 (18.3)	174 (81.7)	13 (7.4)

* Significant difference between resolving and non-resolving nodules according to nodule size ($P < 0.001$)

† Significant difference between non-resolving and non-resolving malignant nodules according to nodule size ($P<0.001$)

Table 3B Moment of resolution according to nodule size

	Total resolving	3 month	Year 2	Year 4
Total	97 (10.1)	75 (75.8)	17 (17.2)	5 (5.0)
Maximal transverse diameter				
< 8 mm	58 (7.7)	40 (69.0)	14 (24.1)	4 (6.9)
≥ 8 mm	39 (18.3)	35 (89.7)	3 (7.7)	1 (2.6)

Discussion

From 2004 to 2006, 805 of 7557 participants (10.7%) in the NELSON study had at least one solid intraparenchymal nodules with volume between 50 – 500 mm^3. Of the 964 nodules that were included, 97 (10.1%) had disappeared in the follow-up examinations. While the majority of solid indeterminate, intraparenchymal pulmonary nodules found at baseline lung cancer screening does not resolve, three quarters of the nodules that do resolve can be identified by short-term repeat CT. Non-peripheral nodules were three times more likely to resolve than peripheral nodules. Spiculated nodules had a four times higher chance to disappear than smooth nodules.

Only few studies in the field of lung cancer screening have focused on disappearing nodules. Diederich et al.[4] studied 107 resolving nodules in 56 individuals, Lee et al.[5] studied 126 resolving part-solid nodules in 93 subjects, Felix et al.[6] evaluated 32 resolving ground-glass opacities in 18 subjects, and Mario et al.[7] assessed 18 out of 76 ground-glass opacities resolving. Non- and part-solid nodules have other characteristics than solid nodules.

In the study by Diederich et al.[4], the number of resolving nodules per individual was 2.38 for participants who had at least one resolving nodule. In that study, the maximum diameter of completely resolving nodules was ≤ 5 mm in 56 of the 107 (52%) nodules. In case of nodules < 5 mm, even those that persist have negligible risk of malignancy [19]. In our study, including nodules with volume > 50 mm^3 (corresponding to 4.6 mm diameter), only 2/964 (0.002%) nodules were 5 mm or less. Diederich et al. [4] found the majority of completely resolving nodules in participants of young age (< 50 years). The risk of lung cancer development increases with aging [20]. In this study, the main age of subjects was 60 ± 6 years, with an age distribution comparable with the overall age distribution of the NELSON study [2]. Therefore, our results mainly concern the behaviour of lung nodules at intermediate-risk (based on size and age) of being malignant.

In the study by Diederich et al. more than one resolving nodule was found in 34% individuals, and in one subject, > 13 resolving nodules were identified. In our study, due to the exclusion of small and large nodules, the number of resolving

nodules per individual was lower (1.32 per individual). Twenty percent of the subjects had more than one nodule that disappeared (range 2-5) [4]. Besides the explanations of nodule size and age, another explanation could be that these nodules are the end stage from benign diseases as multiple nodules are often seen in emphysema or inflammation.

Even in patients at high-risk to develop lung cancer the vast majority of incidentally detected nodules are benign [21]. These benign nodules are probably caused by focal infection or inflammation and often resolve completely or decrease in size at short-term follow-up either after therapy with antibiotics or spontaneously [22]. Libby [23] from the Early Lung Cancer Action Project (ELCAP) reported that 12% of nodules ≥ 5 mm in diameter in participants who had received antibiotics had completely resolved within two months after the initial CT in baseline screening. Antibiotic therapy was not part of the NELSON protocol. Libby et al.[23] reported direct referral for nodules >15mm, so the size of their group of nodules (5 – 15 mm), and thereby the risk of infection or inflammation, did not differ much from our study (5 – 10 mm). Since the nodule disappearance in our study was not the effect of any treatment, our results show a lower percentage (8%) of resolving nodules at short term follow-up (3 months).

It has been demonstrated that a solid, peripheral, subpleural nodule is a specific benign lesion, and mostly may represent intrapulmonary lymph nodes [16]. The main proportion of resolving nodules in this study were peripheral. Both resolving and non-resolving nodules were mainly spherical. Nodules with smooth margin were numerous in both the resolving and the non-resolving group. However, the results showed significant differences between resolving and non-resolving nodules in nodule characteristics. Distance to costal pleural and margin were correlated with nodule disappearance.

The characteristics of nodule edge are one of the important factors in determining whether a lesion is benign or malignant. However, nodules in benign conditions, such as lipoid pneumonia, tuberculoma, and progressive massive fibrosis, may have spiculated margins as malignant nodules [24]. Moreover, a lobulated outline is often associated with malignancy, but may be seen in up to 25% of benign nodules [25]. Furuya et al.[26] analysed margin characteristics of pulmonary

nodules at thin-section CT and found that 80% of the polygonal nodules were the result of inflammatory change and 20% represented primary lung cancer. In the study of Takashima et al.[27], concave margin and polygonal shape were both specific to benign lesions. Our results also show that nodules with non-smooth edges disappeared more frequently than smooth nodules. However, non-smooth edges are also more frequently found in malignant nodules [28]. So, based on the characteristics of nodule edge, no differentiation can be made between resolving and non-resolving malignant nodules.

The rate and speed of disappearance was higher in baseline-detected nodules with a larger diameter (≥ 8 mm vs < 8 mm). However, the rate of malignancy was also higher. Some benign conditions, like inflammations, more commonly have a larger diameter. This may be an explanation for the increased probability of disappearance in nodules with larger maximal transverse diameter. Further stratification of indeterminate nodules dependent on diameter did not help in differentiating between resolving and malignant nodules, and, based on our results, cannot substitute the NELSON volume-based protocol.

An important topic in lung cancer screening is the interval of follow-up. The VDTs of most benign pulmonary nodules are more than 450 days, whereas VDTs of malignant lesions are usually less than 400 days [29]. In several randomised controlled trails which are underway, the interval of early follow-up imaging is 3-, 6-, or 12-months [30-33]. According to our screening protocol, indeterminate nodules (50 – 500 mm^3) had an extra repeat CT three months after the baseline examination to detect growth. Our results show that more than 75% of the resolving nodules disappeared at 3-month follow-up. Therefore, for indeterminate nodules detected in screening, a short term follow-up after initial CT could exclude a considerable number of benign lesions from further work-up.

A limitation of the current study was that the precise time point of nodule resolving could not be ascertained, but only the period between the first CT at which the nodule was detected (in this study the baseline examination) and the first CT after the nodule was completely disappeared. Meanwhile, histological evidence could not be obtained for those resolving nodules. Whether our results can be generalized to incidentally found nodules in a non-screening setting, still

needs to be proven. Further investigations should evaluate the applicability of the nodule management protocol as used in the NELSON study in clinical settings.

In conclusion, about 10% of solid intraparenchymal nodules of intermediate size (volume, 50 – 500 mm^3) found at baseline lung cancer screening disappears during follow-up. Our findings provide further support for a 3-month follow-up CT for these indeterminate lung nodules. Short-term follow-up CT is not only valuable to detect fast growth as determined by VDT, but also, as this study shows, to identify three quarters of resolving nodules. Unfortunately, resolving pulmonary nodules share CT features with malignant nodules. Thus, nodule characteristics cannot sufficiently distinguish intermediate-sized nodules that subsequently disappear.

References

1. Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. Am J Respir Crit Care Med 2002;165(4):508-513.
2. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. N Engl J Med 2009;361(23):2221-2229.
3. Veronesi G, Bellomi M, Mulshine JL, et al. Lung cancer screening with low-dose computed tomography: a non-invasive diagnostic protocol for baseline lung nodules. Lung Cancer 2008;61(3):340-349.
4. Diederich S, Hansen J, Wormanns D. Resolving small pulmonary nodules: CT features. Eur Radiol 2005;15(10):2064-2069.
5. Lee SM, Park CM, Goo JM, et al. Transient part-solid nodules detected at screening thin-section CT for lung cancer: comparison with persistent part-solid nodules. Radiology 2010;255(1):242-251.
6. Felix L, Serra-Tosio G, Lantuejoul S, et al. CT characteristics of resolving ground-glass opacities in a lung cancer screening programme. Eur J Radiol 2011;77(3):410-416.
7. Mario S, Nicola S, Carmelinda M, et al. Long-Term Surveillance of Ground-Glass Nodules: Evidence from the MILD Trial. J Thorac Oncol 2012;7(10):1541-1546.
8. 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.
9. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? Radiology 2004;231(2):453-458.
10. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 2000;217(1):251-256.
11. Gietema HA, Schaefer-Prokop CM, Mali WP, Groenewegen G, Prokop M. Pulmonary nodules: Interscan variability of semiautomated volume

measurements with multisection CT-- influence of inspiration level, nodule size, and segmentation performance. Radiology 2007;245(3):888-894.

12. 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 2007;120(4):868-874.
13. Xu DM, van der Zaag-Loonen HJ, Oudkerk M, et al. Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. Radiology 2009;250(1):264-272.
14. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. Radiology 2008;248(2):625-631.
15. Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. Radiology 1993;186(2):415-422.
16. Takashima S, Sone S, Li F, et al. Small solitary pulmonary nodules (< or =1 cm) detected at population-based CT screening for lung cancer: Reliable high-resolution CT features of benign lesions. AJR Am J Roentgenol 2003;180(4):955-964.
17. Lindell RM, Hartman TE, Swensen SJ, Jett JR, Midthun DE, Mandrekar JN. 5-year lung cancer screening experience: growth curves of 18 lung cancers compared to histologic type, CT attenuation, stage, survival, and size. Chest 2009;136(6):1586-1595.
18. Henschke CI, Yip R, Yankelevitz DF, Smith JP, International Early Lung Cancer Action Program Investigators*. Definition of a positive test result in computed tomography screening for lung cancer: a cohort study. Ann Intern Med 2013;158(4):246-252.
19. 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.

20. Gadgeel SM, Ramalingam S, Cummings G, et al. Lung cancer in patients < 50 years of age: the experience of an academic multidisciplinary program. Chest 1999;115(5):1232-1236.
21. Wormanns D, Diederich S. Characterization of small pulmonary nodules by CT. Eur Radiol 2004;14(8):1380-1391.
22. Diederich S, Wormanns D, Semik M, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. Radiology 2002;222(3):773-781.
23. Libby DM, Wu N, Lee IJ, et al. CT screening for lung cancer: the value of short-term CT follow-up. Chest 2006;129(4):1039-1042.
24. Soubani AO. The evaluation and management of the solitary pulmonary nodule. Postgrad Med J 2008;84(995):459-466.
25. Lee HJ, Goo JM, Lee CH, Yoo CG, Kim YT, Im JG. Nodular ground-glass opacities on thin-section CT: size change during follow-up and pathological results. Korean J Radiol 2007;8(1):22-31.
26. Furuya K, Murayama S, Soeda H, et al. New classification of small pulmonary nodules by margin characteristics on high-resolution CT. Acta Radiol 1999;40(5):496-504.
27. Takashima S, Sone S, Li F, Maruyama Y, Hasegawa M, Kadoya M. 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.
28. Xu DM, van Klaveren RJ, de Bock GH, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. Eur J Radiol 2008;68(2):347-352.
29. 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.
30. National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. Radiology 2011;258(1):243-253.

31. Baldwin DR, Duffy SW, Wald NJ, Page R, Hansell DM, Field JK. UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. Thorax 2011;66(4):308-313.
32. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. Lung Cancer 2009;64(1):34-40.
33. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish randomized lung cancer CT screening trial--overall design and results of the prevalence round. J Thorac Oncol 2009;4(5):608-614.

Summary

Among cancers, lung cancer is the leading cause of death. Lung cancer continues to have a high mortality rate, despite advances in treatment. Early detection by imaging tests could improve the survival in lung cancer. Lung nodules are the predominant radiological finding of lung cancers. Computed tomography (CT) has become the main imaging modality for the detection, characterization and follow-up of lung nodules. Quantification of pulmonary nodules by volume and/or diameter assessment is by now standard procedure.

The Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON) was launched in 2003. The NELSON trial focuses on nodule volume and volume-doubling time (VDT), in contrast to other trials based on nodule diameter assessment. In **Chapter 2**, we described the details on participant recruitment, CT acquisition and nodule management protocol. Valuable knowledge about the presence and characteristics of lung nodules on CT, and associated cancer risk has been obtained in the NELSON study. Small nodules (< 50 mm^3) were shown to have a negligible cancer risk. The results suggest that the risk of malignancy in smooth or attached solid nodules with intermediate size (50 – 500 mm^3) is also very low. In non-smooth nodules without attachment, the only predictor of malignancy was size. Baseline CT density of the lung nodules was found not predictive of malignancy. However, an increase in CT density was suggestive of malignancy in intermediate sized nodules. Among nodule features, size of solid nodules can be considered the main factor related to cancer risk.

A major difference between the NELSON study and other trials is the differentiated manner which lung nodules were managed, according to volume, VDT and density (solid, part-solid, non-solid). The screening result of the solid nodule, the most common type of nodule, was determined by nodule volume at first detection, and VDT on follow-up examinations. Management was determined based on the highest nodule category found. A negative screen result meant that the participant was invited for the regular next screening round. In case of an indeterminate result, a short-term repeat CT was performed to assess nodule growth.

A positive screen result led to referral to a pulmonologist for further work-up and treatment.

In **Chapter 3**, we investigated the NELSON nodule management strategy and the results of the baseline and second round screening. A total of 7,557 participants at high risk of lung cancer based on age and smoking behavior underwent CT screening. The majority of screened individuals had one or more lung nodules. In the baseline round, 2.6% of the participants had a positive test result. The positive and negative predictive value was 36% and 99.9%, respectively. The probability of finding lung cancer one and two years after a negative baseline round test was very low. Nodules with a volume of 50 - 500 mm^3 led to an indeterminate test result which required a repeat scan 3 months later to assess growth. In the baseline round, 92.4% of these indeterminate nodules yielded a negative result in the short-term follow-up CT scan. In a population at increased risk of lung cancer, this strategy of CT screening for lung cancer with management based on volume and VDT diminished the need for further work-up in participants with an indeterminate test result. This strategy was especially useful during the second-round screening. It reduced the number of follow-up examinations needed in participants with a positive test result without reducing the overall sensitivity of the technique.

In the NELSON study, CT scans were independently read by first and second readers. However, double reading is not widely used in clinical routine. Thus, we investigated the utilization of Computer-aided detection (CAD). In **Chapter 4**, we compared the performance of CAD versus double reading in randomly selected CT examinations from the NELSON trial. Based on a consensus panel as reference, the sensitivity of nodule detection was about 20% higher for CAD than for double reading. CAD detected nearly all lung nodules. One lung cancer was missed by readers, but was detected by CAD. Various factors affected nodule recognition during screening including reader experience and variability, CT technique and viewing conditions, as well as nodule characteristics. Contact with vessels increased the difficulty of detection by radiologists. CAD was significantly more sensitive for this type of nodules. On the other hand, high false positive (FP) rate is a considerable drawback of CAD. Over 80% of the FPs in this study was re-

ported only by CAD. A considerable part of FP findings for CAD concerned vessels, pleural plaques and rib bulging which were frequently misinterpreted due to their nodule-like appearance in cross-sectional images. To reduce the FP rate of CAD, we studied the use of a volume cut-off of 50 mm^3. The use of 50 mm^3 as cut-off volume for pulmonary nodules is supported by the NELSON results, as none of the interval cancers between the first and second year screening were due to malignancy in lung nodules < 50 mm^3. Nearly half of the FP cases presented by CAD could be excluded from further evaluation if nodule size was considered, without missing the malignant nodules. Thus, using a combination of CAD and nodule size cut-off in lung cancer screening improves the sensitivity of pulmonary nodule detection compared to that of double reading, and significantly reduces the false positive rate.

NELSON is the first large lung cancer screening trial in which semi-automated, volumetric nodule assessment is routinely applied and forms an integral part of the nodule management protocol. Volumetric three-dimensional measurements have been found to be more accurate than two-dimensional evaluation of pulmonary nodules. In **Chapter 5**, we randomly selected screening CT examinations from the NELSON trial, to investigate different software tools that can be used to assess pulmonary nodules . Specifically, the software packages LungCARE, OncoTREAT and Vitrea were studied. The software packages yielded significantly different values for nodule volumes. Nodule morphology and adjacent structures influenced volume measurement variability. The percentage of satisfactory segmentations of non-smooth nodules was low for all three packages. Nodules adjacent to pleura and vessels showed higher volume measurement variability than well-circumscribed intraparenchymal nodules. Although no significant difference in VDT was found between the software packages, the agreement in VDT categorization was only fair in this study. Thus, variations between software results may lead to false-positive or false-negative screening conclusion. Using different software packages can influence nodule management decisions especially for intermediate-size nodules on consecutive examinations. Further standardization of software used for nodule volumetry and VDT assessment is needed

to optimize nodule management in lung cancer CT screening. At least different software tools should not be used in a single screening trial.

In lung cancer screening, most of the intermediate-sized nodules (size 50-500 mm^3) are not malignant, and may represent granulomatous or infectious lesions, or enlarged lymph nodes. These benign nodules can resolve completely or decrease in size without intervention. The question arose whether it would be possible to identify specific features of nodules that will subsequently resolve, in order to avoid unnecessary follow-up CT. In **Chapter 6**, we retrospectively investigated the nodule features that predict complete disappearance of solid intraparenchymal nodules detected at baseline in the NELSON trial. During subsequent screening rounds, 10% of indeterminate nodules disappeared. Non-peripheral and spiculated nodules showed a higher probability of disappearance compared to peripheral and smooth nodules. Thus, resolving pulmonary nodules share CT features with malignant nodules.

An important topic in lung cancer screening is the interval of follow-up for indeterminate nodules. According to our screening protocol, indeterminate nodules had an extra repeat CT at three months after the baseline examination to evaluate growth. Our results show that by adding a 3 month follow-up CT for nodules of intermediate size, the number of false-positive findings could be greatly reduced as many intermediate nodules were found to have resolved or have a non-malignant growth pattern. Therefore, for indeterminate nodules detected in screening, a short term follow-up after initial CT could exclude a considerable number of benign lesions from further work-up.

In conclusion, this thesis is based on data from the NELSON trial, the first lung cancer screening trial managing nodules using nodule volume and VDT. Our study confirmed that a volume-based 3D measurement is more accurate than diameter-based 2D evaluation by showing an extremely low rate of interval cancers. Using a combination of CAD and nodule size cut-off improves the sensitivity of pulmonary nodule detection compared to double reading. CAD can act as a second reader to assist radiologists in screening work and further in daily clinical work. Different software tools may result in deviant nodule volumetry and VDT assessment and thus, nodule categorization. Thus, further standardization of

nodule evaluation software is needed to optimize nodule management in lung cancer screening. For now, use of a single software package in a lung cancer screening study seems prudent. For indeterminate solid nodules, a short term follow-up after initial CT excludes a considerable number of lesions from further work-up.

Samenvatting

Longkanker is de voornaamste kanker-gerelateerde doodsoorzaak. Ondanks verbeteringen in therapie, is longkanker nog steeds geassocieerd met hoge mortaliteit. Vroege detectie van longkanker middels beeldvorming kan de overleving mogelijk doen toenemen. Longnodulen zijn de belangrijkste radiologische bevinding wijzend op longkanker. Computertomografie (CT) heeft zich ontwikkeld tot de beeldvormende modaliteit bij uitstek voor de detectie, karakterisering en follow-up van longnodulen. Kwantificatie van longnodulen door beoordeling van volume en/of diameter is tegenwoordig standaard.

Het Nederlands-Leuvens Longkanker Screenings Onderzoek (acronym: NELSON) werd gestart in 2003. De NELSON studie richt zich op evaluatie van nodule volume en volume verdubbelingstijd (VDT), in tegenstelling tot andere trials, gebaseerd op nodule diameter meting. In **Hoofdstuk 2** beschrijven we de details met betrekking tot de gerecruteerde deelnemers, het CT scanprotocol en het nodule management regime. De NELSON studie heeft al waardevolle kennis over het voorkomen en de kenmerken van longnodulen op CT, en geassocieerd kankerrisico opgeleverd. Kleine nodulen (< 50 mm^3) bleken een verwaarloosbaar klein kankerrisico te hebben. Resultaten wezen uit dat het risico op maligniteit in gladde nodulen en nodulen vastzittend aan b.v. pleura en vaten, met intermediaire grootte (50 – 500 mm^3) ook zeer laag is. In niet-gladde nodulen in het longparenchym was de grootte de enige voorspeller van kanker. De densiteit van de longnodule op de eerste screenings CT (baseline) bleek niet voorspellend voor het risico op maligniteit. Aan de andere kant, toename in CT densiteit suggereerde wel de aanwezigheid van kanker in nodulen met intermediair volume. Van de nodule kenmerken kan grootte van de solide nodule worden beschouwd als de belangrijkste voorspellende factor voor maligniteit.

Een groot verschil tussen de NELSON studie en andere trials op het gebied van CT longkanker screening, is het management protocol voor de gedetecteerde longnodulen. Bij de NELSON studie vond dit plaats aan de hand van volume, VDT en densiteit (solide, deels solide, niet solide). Het screenings resultaat van de solide nodule, het meest voorkomende type nodule, hing af van het nodule volume

bij de eerste detectie, en van de VDT bij vervolg screenings onderzoeken. Management regime per deelnemer werd bepaald aan de hand van de hoogste nodule categorie op de CT scan. Een negatieve uitslag betekende dat de deelnemer werd uitgenodigd voor de gebruikelijke screening in de volgende ronde. In geval van een indeterminate resultaat werd op korte termijn een herhaal CT verricht om de groei van de longnodule te beoordelen. Een positieve uitslag leidde tot doorverwijzing naar de longarts voor nadere work-up en behandeling.

In **Hoofdstuk 3** werd de nodule management strategie in de NELSON studie onderzocht voor de baseline en tweede screeningsronde. Meer dan 7500 deelnemers met hoog risico op longkanker aan de hand van leeftijd en rookgedrag ondergingen baseline CT screening. De meerderheid van de gescreende deelnemers had één of meer longnodulen. In de baseline ronde had 2.6% van de deelnemers een positieve screeningsuitslag. De positief en negatief voorspellende waarde van de screenings test was 36% en 99.9%, respectievelijk. De kans op het optreden van longkanker één of twee jaar na een negatief baseline screeningsresultaat was zeer laag. Nodulen met een volume van 50 - 500 mm^3 resulteerden in een indeterminate testresultaat en een herhaal CT screening na 3 maanden om groei te evalueren. In de eerste screenings ronde leverde deze korte termijn herhaal screening in meer dan 90% van deelnemers met een indeterminate nodule een negatief resultaat op. In een populatie met verhoogd risico op longkanker verminderde de NELSON strategie de noodzaak voor nadere diagnostiek in deelnemers met een indeterminate screenings resultaat. Deze strategie was in het bijzonder effectief tijdens de tweede screenings ronde. De nodule management strategie op basis van volume en VDT reduceerde het aantal benodigde vervolg onderzoeken in deelnemers met een positief testresultaat, zonder de gevoeligheid van de techniek voor detectie van longkanker omlaag te brengen.

In de NELSON studie werden CT scans onafhankelijk door twee beoordelaars bekeken. Echter in de klinische praktijk komt deze dubbele evaluatie (double reading) niet vaak voor. Daarom werd de inzet van computer-ondersteunde detectie ('Computer-Aided Detection', CAD) onderzocht. In **Hoofdstuk 4** vergeleken we de accuratesse van CAD ten opzichte van dubbele evaluatie, gebruikmakend van een steekproef uit de NELSON studie. De gevoeligheid voor nodule de-

tectie was ongeveer 20% hoger voor CAD dan voor dubbele evaluatie, met het oordeel van een consensus panel als referentie. CAD vond bijna alle longnodulen. Eén geval van longkanker werd door menselijke beoordelaars gemist, maar wel gevonden met behulp van CAD. Factoren die het herkennen van longnodulen beïnvloedden waren onder andere ervaring van de beoordelaars, CT techniek en omgeving waarin beoordeeld werd, alsmede nodule kenmerken. Radiologen hadden meer moeite om nodulen te detecteren als deze vast zaten aan vaten. CAD was gevoeliger voor detectie van dit soort nodulen. Aan de andere kant, gebruik van CAD leverde veel meer fout positieve bevindingen op, structuren die geen longnodulen bleken te betreffen. Meer dan 80% van de fout positieve bevindingen werden alleen door CAD gedetecteerd. Een belangrijk deel van de fout positieve bevindingen betroffen vaten, pleurale verdikkingen en rib uitsteeksels, die door CAD verkeerd waren geïnterpreteerd als longnodulen. Om het percentage fout positieve bevindingen voor CAD te verminderen, werd een afkapwaarde voor het volume van te detecteren nodule onderzocht. Hiervoor werd een volume van tenminste 50 mm^3 gebruikt. Deze volume afkapwaarde wordt ondersteund door resultaten in de NELSON studie, aangezien longnodulen < 50 mm^3 niet leidden tot intervalkankers tussen de eerste en tweede screeningsronde. Bijna de helft van alle fout positieve bevindingen voor CAD konden worden geëxcludeerd als de nodule grootte werd betrokken in de beoordeling, zonder dat maligne nodulen gemist werden. Het gebruik van een combinatie van CAD en nodule grootte afkapwaarde in longkanker screening verbetert de gevoeligheid van longnodule detectie vergeleken met dubbele evaluatie, en vermindert het aantal fout positieve bevindingen.

NELSON is de eerste longkanker screening trial waarin semi-automatische berekening van nodule volume routinematig wordt toegepast, en integraal deel uitmaakt van het nodule management protocol. Drie-dimensionale, volume metingen zijn accurater gebleken dan twee-dimensionale, diameter evaluatie van longnodulen. In **Hoofdstuk 5** werden verschillende software pakketten voor meting van longnodule grootte gebruikt om een steekproef van screenings CT onderzoeken te beoordelen. Het ging specifiek om LungCARE, OncoTREAT and Vitrea. De waarden voor nodule volume die voor de verschillende software pakketten

werden gevonden, weken van elkaar af. Morfologie van nodulen en aangrenzende structuren beïnvloedden de variabiliteit in volume meting. Het percentage adequate segmentatie van niet-gladde nodule was laag voor alledrie de software pakketten. De variabiliteit in volumetrie was hoger voor nodulen aangrenzend aan pleura en vaten, dan voor scherp afgrensbare nodulen in het longparenchym. Alhoewel er geen significant verschil in VDT gevonden werd tussen de software pakketten, was de overeenkomst in categorisatie van nodulen op basis van VDT slechts matig. Variatie in software metingen kunnen leiden tot fout positieve of fout negatieve screenings resultaten. Het gebruik van verschillende software pakketten kan beslissingen over de management van longnodulen verschillend doen uitvallen, vooral in geval van vervolg CT onderzoeken voor nodulen met intermediair volume. Verdere standardisatie van software voor nodule volume en VDT bepaling is nodig om de nodule management strategie in longkanker CT screening te optimaliseren. Ten minste zou in een longkanker screening studie gebruik gemaakt moeten worden van één en hetzelfde software pakket.

De meeste longnodulen met intermediaire grootte (50 - 500 mm^3) die worden gevonden op CT screenings onderzoeken, zijn geen kanker. Deze nodulen kunnen b.v. granulomateuze of infectieuze lesies betreffen, of vergrote lymfklieren. Deze goedaardige nodulen kunnen verdwijnen of kleiner worden zonder enige behandeling. De vraag kwam op of bepaalde nodule kenmerken erop wijzen dat de nodule op vervolg onderzoeken verdwenen is, met als doel om onnodige herhaal CT scans te voorkomen. In **Hoofdstuk 6** werd retrospectief onderzocht welke nodule kenmerken het verdwijnen van een solide, in het longparenchym gelegen nodule voorspelt. Hiervoor werden nodulen geëvalueerd die waren ontdekt bij de baseline screenings ronde. Bij volgende screenings ronden bleek 10% van de indeterminate nodulen verdwenen. Niet-perifere nodulen en gespiculeerde nodulen verdwenen vaker dan perifere en gladde nodulen. Derhalve overlappen de kenmerken van nodulen die verdwijnen, met de kenmerken van maligne nodulen.

Een belangrijk onderwerp in longkanker screening is het interval voor opvolgen van indeterminate nodulen. Volgens het NELSON protocol voor de baseline ronde werd dan een extra herhaal CT verricht na 3 maanden, om nodule groei te

beoordelen. Onze resultaten tonen aan dat door deze korte termijn herhaal CT scan voor indeterminate nodule, het aantal fout positieve bevindingen in longkanker screening naar beneden gebracht kan worden. Op het herhaal CT onderzoek bleken veel nodulen met initieel intermediaire grootte namelijk verdwenen of een niet-maligne groeipatroon te hebben. Daarom kan een korte termijn herhaal CT onderzoek nadere diagnostiek voor veel deelnemers met indeterminate screenings resultaat voorkomen.

Concluderend beschrijft dit proefschrift resultaten van de NELSON studie, het eerste longkanker screenings onderzoek waarin longnodule management gebaseerd is op nodule volume en VDT. Onze studie bevestigt dat nodule volume bepaling accurater is dan diameter meting, gezien het lage percentage intervalkankers tussen de screenings ronden. Een combinatie van CAD en volume afkapwaarde verbetert de gevoeligheid voor nodule detectie vergeleken met dubbele evaluatie. CAD kan als een tweede beoordelaar fungeren, om radiologen te assisteren in het screeningswerk en in de klinische praktijk. Verschillende software pakketten kunnen leiden tot verschillen in nodule volumetrie en VDT bepaling, en dus, tot verschillen in nodule categorizatie en management. Verdere standardisatie van nodule evaluatie software is nodig om het nodule management regime in longkanker screening te optimaliseren. Het is aan te bevelen in een longkanker screening studie gebruik te maken van één software pakket. Een korte termijn herhaal CT onderzoek na de eerste screening kan een aanzienlijk deel van indeterminate, solide longnodulen identificeren die geen nadere diagnostiek behoeven.

Acknowledgements

I would like to express my gratitude to all those who helped me during the writing of this thesis.

My deepest gratitude goes first and foremost to Professor Matthijs Oudkerk, my promoter, for his constant encouragement and patient guidance. He has walked me through all the stages of the writing of this thesis. His standards of academic excellence have made the present form of this thesis.

Second, I would like to express my heartfelt thanks to my co-promoter, Dr. Rozemarijn Vliegenthart, who has taken her precious time off from her tight schedule, reading through each draft carefully and offered me precious criticism. Thank you for the enormous email contact and valuable suggestions.

I would also like to acknowledge my indebtedness to Professor Truuske de Bock whose valuable instruction on statistics has benefited me a great deal.

Dear Peter van Ooijen and Wim Tukker, thank you for the technical support and valuable comments on my thesis.

Special thanks should go to dear Stella. It is really my luck to have you around in the past years. Thank you for giving me generous support and helpful advice on work and on life. Without you pushing me ahead, the completion of this thesis would be impossible.

I also wish to sincerely thank my colleagues and friends.

Particularly, I feel grateful to dear Anne, my first Dutch friend and office mate. At the beginning of those days I was in Holland, your kindness and hospitality made me less homesick. I miss the days we shared one office.

Dear great colleagues, Thank you for contributing your time, thoughts, skills and encouragement to this thesis. Ying, Dongming, Xueqian, Gonda, Hildebrand, Daniël, Monique, Wisnu, Marjolein, Astri, Volkan, Kadek, etc, the time we spent together will be one of my best memories in my life.

Dear wonderful friends, Hao and Pieter, Yongqing, Annie, Gaifen, Wenli, Danna, Cheng, Qu Ning, Hongwei, etc, our friendship made my days in Groningen happy and colorful. Although some of you are far away from me currently, I wish our friendship may last forever.

I would also like to thank the members of the reading committee, Professor J.W.J. Lammers, Professor W.P.Th.M. Mali, Professor H.J.M. Groen, for your critical reading and precious time.

Dear Professor Runxian Bao, Professor Ying Wang, Professor Zhaoxiang Ye and Professor Peifang Liu, thank you for supporting me in my study and work in all those years. Your attitude in the scientific area will keep inspiring me in the future.

Finally I wish to devote this thesis to my beloved Mum and Dad, who have given me life and love. Without your loving encouragement and continuous support, I can not stay alone for many years in a country so far away from home. After all these years, the happiest thing for me is to be back home and be close to you again.

Curriculum Vitae

Yingru Zhao was born on January 24, 1975 in Tianjin, China. After graduating from Tianjin Shiyan high school, she started Medical School in 1993 at Tianjin Medical University. She finished her internship in Tianjin First Center Hospital and Tianjin General Hospital, and gained a Bachelor Degree in Radiology in 1998. Hereafter, she started working as a resident in the radiology department of Tianjin Medical University Cancer Institute & Hospital. In 2000, she got her certification of medical doctor. In 2004, she completed her residency and became a radiologist. From 2004 to 2006, she followed a Master course in Tianjin Medical University. She was supervised by Professor Peifang Liu on the research project "Relationship between mammography findings and histological grades in non-special type invasive breast carcinoma". In 2006, she moved to the Netherlands to start her PhD project on the NELSON-study (Dutch Belgium Lung Cancer Screening Study) at the University Medical Center Groningen. This work resulted in the thesis entitled "Lung nodule assessment in low-dose CT lung cancer screening: validation of detection and volumetric measurement".

Since March 2012, she is working again as radiologist in the Tianjin Medical University Cancer Institute & Hospital.