

NELSON lung cancer screening study

Ying Ru Zhao^a, Xueqian Xie^a, Harry J de Koning^b, Willem P Mali^c, Rozemarijn Vliegenthart^a and Matthijs Oudkerk^a

^aDepartment of Radiology, University Medical Center Groningen, University of Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands; ^bDepartment of Public Health, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands; ^cDepartment of Radiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Corresponding address: Matthijs Oudkerk, Department of Radiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, PO Box 30.001, 9700 RB Groningen, The Netherlands.
Email: m.oudkerk@umcg.nl

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 multidetector computed tomography (CT) in high-risk subjects will lead to a decrease in 10-year lung cancer mortality of at least 25% compared with a control group without screening. Since the start of the NELSON study in 2003, 7557 participants underwent CT screening, with scan rounds in years 1, 2, 4 and 6. In the current review, the design of the NELSON study including participant selection and the lung nodule management protocol, as well as results on validation of CT screening and first results on lung cancer screening are described.

Keywords: Lung cancer; screening; multidetector computed tomography; population; pulmonary nodules; volume measurement.

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 radiography 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 overdiagnosis bias. Randomized studies are needed to compare disease-specific mortality between a screened and an unscreened population. This was the reason for launching 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 seven 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 two 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 centre. 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 multicenter 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 randomization of participants, electronic storage of forms, study monitoring, reporting of scan results and scheduling of appointments for follow-up scans.

This review concerns the screen arm of the study. Participants randomized 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 years 1, 2, 4, and 6. On 1 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 (FEV₁) 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, four 16-detector CT scanners (three Sensation-16, Siemens Medical Solutions, Forchheim, Germany M × 8000 IDT and one Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA) 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, in a 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 depending on the machine used. To minimize 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].

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 second 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 such as pleural-based nodules, measurement of diameters from a point perpendicular to the costal pleura was performed manually. In addition to size, nodule characteristics such as shape and surface were noted. Growth was defined as change in volume of at least 25% between 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.

Management was determined based on the highest nodule category found. 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 3–4 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 a shorter interval, however, because at incidence screen new nodules are supposed to have a relatively higher growth rate.

If the highest category was a NODCAT 4 or GROWCAT C, the participant was referred to a chest physician via the general practitioner, usually the chest physician associated with the screening centre. The primary objective was to confirm the presence of

Table 1 Nodule categorization based on size and characteristics (new nodules) and growth rate (existing nodules) in NELSON study

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		
NODCAT 3	Solid	Partial solid	Non-solid
	$50 \leq V \leq 500 \text{ mm}^3$	Solid component: $50 \leq V \leq 500 \text{ mm}^3$	$d_{\text{mean}} \geq 8 \text{ mm}$
NODCAT 4	Pleural based: $5 \leq d_{\text{min}} \leq 10 \text{ mm}$	Non-solid component: $d_{\text{mean}} \geq 8 \text{ mm}$	
	$V > 500 \text{ mm}^3$	Solid component: $V > 500 \text{ mm}^3$	Non-existent category
GROWCAT A	Pleural based: $d_{\text{min}} > 10 \text{ mm}$		
GROWCAT B	VDT > 600 days		
GROWCAT C	$400 \leq \text{VDT} \leq 600 \text{ days}$		
	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.

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

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 fine-needle aspiration, video-assisted thoracoscopic surgery, or wedge resection and examination on frozen section. The work-up, 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 mentioned shortly. The studies can be separated into variability studies and studies on nodule characteristics suggestive of malignancy/benignancy.

Gietema et al.^[7] investigated the interobserver variability of semi-automated volume measurements of small-to-intermediate size lung nodules (NODCAT 2 and 3). Interobserver correlation was very high ($r = 0.99$). Nearly 90% of nodules did not show any variation in volume with double reading. There was a volume difference of >10% in only 3.7%, mostly due to incomplete segmentation due to irregular shape or margins. In a further study^[9], the variability of volume measurements was found to be related to nodule morphology, location and size. Volume disagreement was most likely in the 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

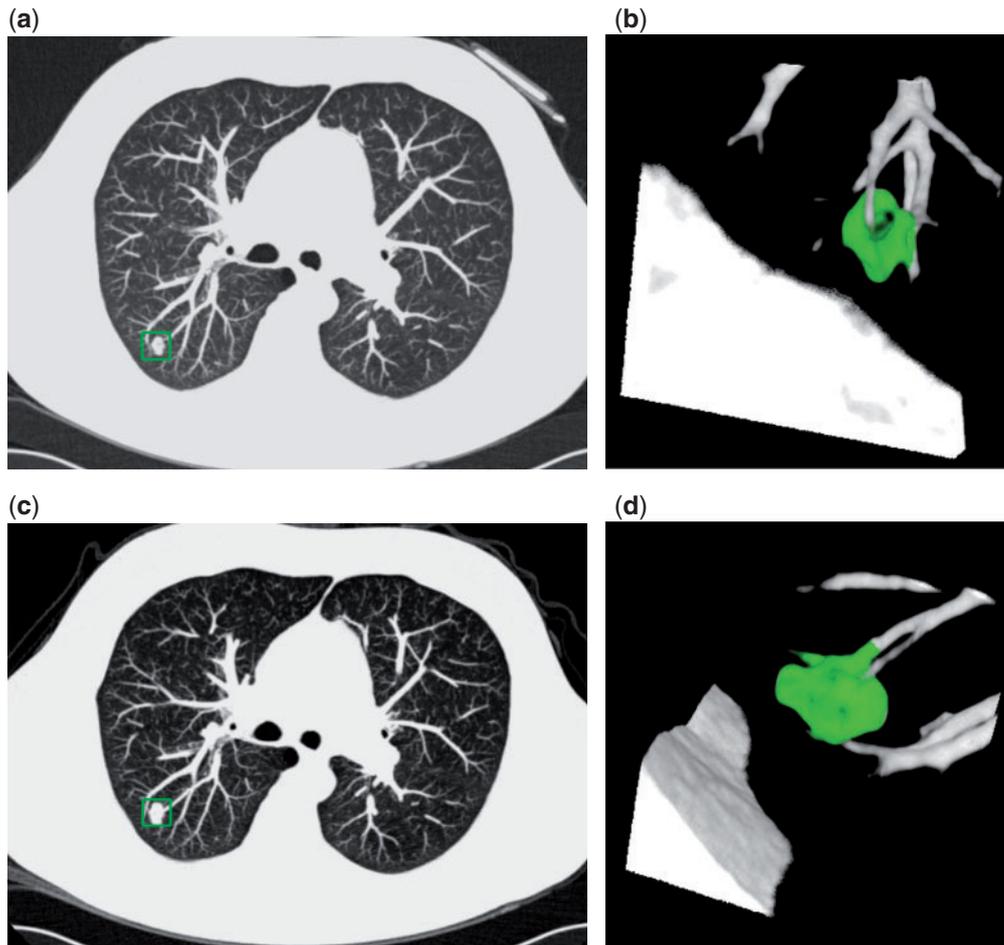


Figure 1 Baseline and 3-month follow-up CT images in a 68-year-old participant of the NELSON study. Transverse thin-section CT (a,c) and volume-rendered reconstruction (b,d) images show a lobulated pulmonary nodule with vessel attachment (boxed on a,c and green area in b,d). On the baseline scan (a,b) 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.

nodule location and morphology. The volume measurement was most repeatable for 1 mm section thickness with soft kernel. In the case of serial CT studies, it was concluded that consistent reconstruction parameters essential.

Valuable knowledge about the 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 1 year after baseline. Cancers were found to be non-spherical and purely intraparenchymal, 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 intermediate 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 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 and the 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 who had one or more intermediate size nodules (NODCAT 3) and thus an indeterminate test result, 5.3% showed a growing nodule suspicious for malignancy (GROWCAT C) on the 3-month follow-up scan. Fig. 1 is an example of a

GROWCAT C nodule that was proved 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. The lung cancer detection rate was 0.9%. The sensitivity of the first screening round was 94.6% and the negative predictive value was 99.7%. There were only three 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 found to have lung cancer. The 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 a very high negative predictive value. The NELSON study has several features that distinguish this trial from the National Lung Cancer Screening trial^[15]. First, 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 volume-doubling time. 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 anxiety, costs and morbidity. The National Lung Cancer Screening trial recently showed a positive CT in 24.2%, with 96.4% being false-positive results. By adding a 3–4 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 (1) whether lung cancer screening in high-risk subjects by low-dose CT does decrease lung cancer mortality compared with no screening and (2) if a CT protocol based on volumetry and volume-doubling time only is more effective in terms

of detection rate, morbidity, mortality, recall rate, and cost-effectiveness, than other approaches.

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. doi:10.1186/1471-2407-2-37.
- [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: 123–37. doi:10.1016/S0169-5002(00)00197-5.
- [3] 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. doi:10.1016/S0140-6736(99)06093-6.
- [4] Kaneko M, Eguchi K, Ohmatsu H, 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, 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: 868–74. doi:10.1002/ijc.22134.
- [6] Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54: 177–84. doi:10.1016/j.lungcan.2006.08.006.
- [7] Gietema HA, Wang Y, Xu DM, et al. Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. *Radiology* 2006; 241: 251–7. doi:10.1148/radiol.2411050860.
- [8] 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. doi:10.1007/s00330-003-2132-0.
- [9] 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: 625–31. doi:10.1148/radiol.2482070957.
- [10] Wang Y, de Bock GH, van Klaveren RJ, et al. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. *Eur Radiol* 2010; 20: 1180–7. doi:10.1007/s00330-009-1634-9.
- [11] 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: 264–72. doi:10.1148/radiol.2493070847.
- [12] 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: 347–52. doi:10.1016/j.ejrad.2007.08.027.
- [13] Xu DM, van Klaveren RJ, de Bock GH, 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: 492–8. doi:10.1016/j.ejrad.2008.02.022.
- [14] van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361: 2221–9. doi:10.1056/NEJMoa0906085.

- [15] The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365: 395–409.
- [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: 453–8. doi:10.1148/radiol.2312030167.
- [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: 251–6.