Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening
Abstract

**Background:** The main challenge in CT screening for lung cancer is the high prevalence of pulmonary nodules and the relatively low incidence of lung cancer. Management protocols use thresholds for nodule size and growth rate to determine which nodules require additional diagnostic procedures, but these should be based on individuals’ probabilities of developing lung cancer. In this retrospective analysis, using data from the NELSON CT screening trial, we aimed to quantify how nodule diameter, volume, and volume-doubling time affect the probability of developing lung cancer within two years of a CT scan, and to propose and evaluate thresholds for management protocols.

**Methods:** Eligible participants in the NELSON trial were those aged 50-75 years, who have smoked 15 cigarettes or more per day for more than 25 years, or 10 cigarettes or more for more than 30 years and were still smoking, or had stopped smoking less than 10 years ago. Participants were randomly assigned to low-dose CT screening at increasing intervals, or no screening. We included all participants assigned to the screening group who had attended at least one round of screening, and obtained data on lung cancer diagnoses from the national cancer registry database. We calculated lung cancer probabilities, stratified by nodule characteristics, by nodule diameter, volume and volume-doubling time and did logistic regression analysis using diameter, volume, volume-doubling time, and multinodularity as potential predictor variables. We assessed management strategies based on nodule threshold characteristics for specificity and sensitivity, and compared them to the American College of Chest Physicians (ACCP) guidelines.

**Findings:** Volume, volume-doubling time and volumetry-based diameter of 9,681 non-calcified nodules detected by CT screening in 7,155 of 7,915 participants in the screening group of NELSON were used to quantify lung cancer probability. Lung cancer probability was low in participants with a nodule volume of 100 mm$^3$ or smaller (0.6% [95% CI 0.4-0.8%]) or maximum transverse diameter smaller than 5 mm (0.4% [0.2-0.7%]), and not significantly different from participants without nodules (0.4% [0.3-0.6], $P=0.17$ and $P=1.00$, respectively. Lung cancer probability was intermediate if nodules had a volume of 100-300 mm$^3$ (2.4% [1.7-3.5%]), or a diameter 5-10 mm (1.3% [1.0-1.8%]). Volume-doubling time further stratified the probabilities: 0.8% (95% CI 0.4-1.7%) for volume-doubling times of 600 days or more, 4.0 (1.8-8.3%) for volume-doubling times 400-600 days, and 9.9% (95% CI 6.9-14.1%) for volume-doubling times of 400 days or fewer. Lung cancer probability was high for participants with nodule volumes 300 mm$^3$ or bigger (16.9% [95% CI 14.1-20.0%]) or diameters 10 mm or bigger (15.2% [12.7-18.1%]), even if these nodules had long volume-doubling times. The simulated ACCP management protocol yielded a sensitivity and specificity of 90.9% (95% CI 89.3-90.7), and 87.2% (86.4-87.9) respectively. A diameter-based protocol with slightly adjusted thresholds (based on lung cancer probability) yielded a higher sensitivity (92.4% [95% CI 83.1-97.1]), and a higher specificity (90.0% [81.2-96.1]). A volume-based protocol (with thresholds based on lung cancer probability) yielded the same sensitivity as the ACCP protocol (90.9% [95% CI 81.2-96.1]), and a very high specificity (94.9% [94.4-95.4]).

**Interpretation:** Small nodules (volume <100 mm$^3$ or diameter <5 mm) are not predictive for lung cancer. Immediate diagnostic evaluation is necessary for large nodules ($\geq$300 mm$^3$ or $\geq$10 mm). Volume-doubling time assessment is advocated only for intermediate-sized nodules (volume of 100-300 mm$^3$ or diameter of 5-10 mm). Nodule management protocols based on these thresholds performed better than the simulated ACCP nodule protocol.
Introduction

Several prominent medical associations have recommended regular low-dose CT screening for asymptomatic smokers and ex-smokers at high risk of developing lung cancer. [1, 2]. The main challenge faced by clinicians doing CT screening for lung cancer is that about half of people screened have one or more pulmonary nodules, but only a small percent of these people have lung cancer [3, 4]. Validated guidelines to determine optimum patient management strategies based on characteristics of detected nodules are urgently needed.

When CT screening for lung cancer first began, the accepted standard of practice was to regard all non-calcified pulmonary nodules as potentially malignant lesions requiring follow-up screening until proven stable for a period of 2 years [5–7]. Later, the Fleischner Society recommended that nodules of 4 mm in diameter or smaller in high-risk individuals (ie, history of smoking or other known risk factors) required no further follow-up if the nodule was unchanged at a 12-month follow-up examination, because the risk of the nodule being malignant was less than 1%. [8]. However, people with nodules 4-8 mm in size were still recommended to undergo two to three follow-up examinations over a period of 2 years. Individuals with nodules larger than 8 mm were recommended to undergo diagnostic work-up, which consisted of more invasive diagnostic procedures [8]. Recently, the results of the Early Lung Cancer Action Project (ELCAP) [9] - which suggested raising of the threshold for initiation of follow-up CT examinations to nodules of 8 mm or larger - were reproduced with data from the National Lung Screening Trial (NLST) [10]. However, the ELCAP analyses were limited to screen detected lung cancers, and only false-positive values and time to diagnosis were taken into account when assessing new thresholds for nodule diameter. Increasing the protocol-screening thresholds for nodule diameter to determine which patients should undergo diagnostic follow-up reduces the potential harms of diagnostic procedures, exposure to ionizing radiation, and costs [11, 12]. However, it might also decrease the sensitivity for cancerous nodules, thus, in turn, increasing lung cancer mortality, and so it is important to balance these potential benefits and harms [4]. Therefore, thresholds for negative, indeterminate, and positive screening results should be based on individual participants’ probability of developing lung cancer, and should be assessed in terms of sensitivity, specificity, number of required CT examinations, and number of required invasive diagnostic procedures.

Recommendations of the latest American College of Chest Physicians (ACCP) guidelines for management of individuals with pulmonary nodules with a diameter of 8 mm or larger were based on the consensus statement of the Fleischner Society [8]. This statement has not been formally validated, and alternative management strategies might yield an improved performance in terms of sensitivity, specificity, and the number of required follow-up scans.

The NELSON trial is a randomized trial to assess whether low-dose CT screening with an increasing length of screening interval (1, 2, and 2.5 years) compared with no screening reduces lung cancer mortality [13]. We used data from NELSON to quantify the probability of developing lung cancer within two years of CT screening, based on measurements of lung nodule diameters, volumes, and volume-doubling times. We used lung cancer probabilities to assess the nodule management protocol recommended by the ACCP, and to propose
improved management protocols [8, 14].

Methods

Study design and participants

Details about the design and conduct of the NELSON trial have been reported previously [13, 15]. Briefly, participants from four centers in the Netherlands and Belgium were enrolled and randomly assigned to receive low-dose CT screening or no screening. Eligible participants were adults aged 50-75 years, who had smoked 15 or more cigarettes per day for more than 25 years or ten or more cigarettes per day for more than 30 years, and were still smoking or had stopped smoking less than 10 years previously. People with self-reported moderate or bad health (with a questionnaire adapted from the SF-36 questionnaire), inability to climb two flights of stairs, bodyweight of 140 kg or more, current or past renal cancer, melanoma, breast cancer, or lung cancer diagnosed less than 5 years ago, or a chest CT examination less than 1 year ago, were excluded.

All participants who were diagnosed with lung cancer were identified from the national cancer registry of the Netherlands. We included all Dutch participants who were randomly assigned to the screening group, who had attended at least one round of screening in the first two screening rounds at 1 and 2 years after baseline screening. We excluded Belgian participants because no data were yet available from the Belgian cancer registry.

The NELSON trial was approved by the Dutch Minister of Health and ethics boards at each participating center. All participants gave written informed consent for participation and evaluation of personal data from hospital charts.

Procedures

The protocol describing how CT screening was done in the NELSON trial has been previously published [13], and is summarized in the appendix. Briefly, CT screening was done with 16-detector CT scanners in a low-dose setting (effective radiation dose <0.4 mSv, <0.8 mSv and <1.6 mSv, dependent on bodyweight) [13]. Datasets were derived from images of the thorax (slice thickness 1 mm, interval 0.7 mm) and analyzed with software for semi-automated volume measurements (LungCARE, Siemens, version Somaris/5 VB 10A-W).

For any CT screen-detected non-calcified nodules, semi-automatic volumetric software independently then measured volume and maximum transverse diameter. Hence, the diameters used in this study were not measured manually. In cases in which no volume \( V \) could be assessed (eg, in non-solid nodules), volume was estimated with use of a manually measured diameter \( D \), assuming a spherical shape of the nodule according to the formula:

\[
V = \frac{1}{6} \cdot \pi \cdot D^3 \tag{5.1}
\]

When diameter was missing, it was estimated with the inverse of this formula.
We calculated volume-doubling time for the first and second round for all nodules detected on at least two scans. For the assessment of lung cancer probability by volume-doubling time and the volume-based nodule protocol, we used the formula:

$$VDT = \frac{\ln(2)\Delta t}{\ln(V_2/V_1)}$$  \hspace{1cm} (5.2)

in which $\Delta t$ represents time in days between scans. The volume-doubling times of all nodules detected in round one and the newly detected nodules in round two were calculated with the volumes measured on the regular round scan ($V_1$) and the follow-up scan ($V_2$). The volume-doubling times of nodules in round two that had also been detected at baseline were calculated with volumes measured on the baseline scan ($V_1$) and the second round scan ($V_2$). For the evaluation of the diameter-based nodule protocols, the following formula for volume-doubling time was used:

$$VDT = \frac{\ln(2)\Delta t}{3 \cdot \ln(\text{MaxDiamXY}_2/\text{MaxDiamXY}_1)}$$  \hspace{1cm} (5.3)

in which $\Delta t$ represents time in days between scans, and $\text{MaxDiamXY}_1$ and $\text{MaxDiamXY}_2$ are maximum diameters on the X-Y axis at first and second assessment [13]. All analyses were done at the participant level; for participants with more than one nodule, we used the size of the largest nodule and volume-doubling time of the fastest growing nodule (of 50-500 mm$^3$). Using these findings we calculated probabilities of developing lung cancer, stratified by nodule characteristics. Two-year probability was chosen because it is the recommended follow-up time for indeterminate nodules [8, 14]. We predicted lung cancer risk in the 2 years following each screening round using regression analysis stratified with nodule characteristics as potentially predictive variables. Based on these outcomes, we designed nodule management protocols for both nodule volume and diameter. Participants without nodules or with a lung cancer probability not significantly different from those without nodules were classified as negative, and were not recommended to undergo intensified CT surveillance [8] besides screening. Participants with a significantly increased lung cancer probability (but less than about 5%; adopted from ACCP guidelines [14]) were classified as indeterminate, and were recommended to undergo CT surveillance to assess nodule growth; if lung cancer probability based on volume-doubling time was significantly higher than in participants without nodules, the final result was classified as positive, otherwise, it was classified as negative. Participants with a lung cancer probability of more than 5% were directly classified as positive, and recommended to undergo additional diagnostic procedures immediately (adopted from ACCP guideline for nodules with a 5% to 65% risk of malignancy) [14]. Furthermore, the ACCP management protocol [8, 14] (originally designed for manually measured nodules) was simulated as follows: follow-up CT at 12 months for nodules 4 mm or smaller (classified as negative); follow-up CT at 6-12 months and 18-24 months for nodules 4-8 mm in size (classified as indeterminate; final result positive for volume-doubling times $<400$ days, [13] otherwise negative); and additional diagnostic procedures for nodules larger than 8 mm (classified as positive).
Outcomes

The primary endpoint of the NELSON trial is reduction of lung cancer mortality by 25% or more at 10 years after randomisation [15, 16]. The aim of this prespecified analysis was to quantify the probability of developing lung cancer within 2 years after screening, stratified by measured nodule diameters, volumes, and volume-doubling times. The secondary aims were to model lung cancer risk using predictive variables, and to propose and assess thresholds for nodule management protocols.

Statistical analysis

Probabilities of developing lung cancer stratified by different nodule variables were calculated by dividing the number of individuals with cancer by the total number of participants. Differences between lung cancer probabilities were tested using Fisher’s exact test; 95% CIs were calculated using the Agresti-Coull method.

To predict lung cancer risk in the 2 years after each screening round, we did logistic regression analysis using diameter, volume, volume-doubling time, and multinodularity as potential predictor variables. The model only included participants whose largest nodule measured 50-500 mm$^3$ and who had one nodule or more growing in this volume range, because volume-doubling time was available only for this subgroup. We accounted for non-linear effects of the predictor variables using fractional polynomials. For each predictor variable, we included two terms of the form $X^K$, with the value of K chosen from the set (-2, -1, -0.5, 0, 0.5, 1, 2, 3); $X^0$ denoted a logarithmic transformation. The predictor variables in the final model and the non-linear transformations were chosen with backward elimination with a significance level of 5%, on the basis of the multivariable fractional polynomials algorithm [17]. We used a closed-test procedure to control the family-wise type I error rate in a situation with multiple testing [18]. The calibration of the model was assessed with the Hosmer-Lemeshow test.

We estimated test characteristics of all three nodule management protocols using the detection method with a 1-year interval plus all lung cancers detected in the same screening round (appendix). Hence, we estimated sensitivity by dividing the number of true-positive screens by the numbers of true-positive and false-positive screens. We estimated specificity by dividing the number of true-negative screens by the numbers of true-negative and false-negative screens. We estimated positive predictive value by dividing all individuals with a true-positive screening by all individuals with positive screening. We estimated negative predictive value by dividing all participants with a true negative screening by all participants with negative screening (appendix).

All statistical tests were two-sided, used a significance level of 5%, and were done with Stata (version 12), R (version 2.15), and Microsoft Excel (2010). This trial is registered as an International Standard Randomised Controlled Trial at www.trialregister.nl, number ISRCTN63545820.
Results

Participants

A total of 15,822 participants were enrolled in the NELSON trial between Dec 23, 2003, and July 6, 2006. Screening round one was conducted from January, 2004, to December, 2006, and screening round two from January, 2005, to September, 2008. For this study, we excluded 7,907 participants randomly assigned to the no screening group, 477 participants from Belgium (no data were yet available from the Belgian cancer registry), and 283 participants who did not attend their screening examinations (no screening test characteristics could be calculated in the absence of screening). Thus, we included 7,155 participants in our study (7,135 of whom received screening at the first screening round, and 6,889 of whom received screening at the second screening round; Figure 5.1).

Median length of available follow-up of the participants was 8.16 years (IQR 7.56-8.56). Median age was 58 years (IQR 50-66). 1,206 (16%) of 7,438 participants were women, 6,232 (84%) were men, 4,165 (56%) were current smokers, and their median number of pack-years at randomization was 38 (IQR 19-57).

Quantifying lung cancer probability

Two-year lung cancer probability for all included participants was 1.3% (95% CI 1.2-1.5; Table 5.1). Participants without any pulmonary nodule (7,630 [54%] of 14,024 screenings in rounds one and two combined) had a lung cancer probability of 0.4% (0.3-0.6). In all participants with CT-detected nodules, lung cancer probability was 2.5% (2.1-2.9), but individuals’ probabilities depended strongly on nodule volume, diameter and volumedoubling time (Table 5.1).

We used volume, volume-doubling time, and volumetry-based diameter of 9,681 non-calcified nodules detected by CT screening in 7,155 participants in the screening group of NELSON to quantify lung cancer probability (Table 5.1). Lung cancer probability did not significantly differ between participants who had nodules of less than 100 mm$^3$ in volume and participants who had no detected nodules (0.6% [95% CI 0.4-0.8] vs 0.4% [0.3-0.6]; $P = 0.17$). Participants who had nodules between 100-300 mm$^3$ had a significantly greater probability of developing lung cancer compared to participants with no screening-detected nodules (2.4% [95% CI 1.7-3.5]; $P<0.0001$) and so these participants could be regarded as being at intermediate risk for developing lung cancer. Participants who had nodules of 300 mm$^3$ or more also had a significantly greater probability of developing lung cancer compared to participants with no nodules (16.9% [95% CI 14.1-20.0]; $P<0.0001$) and so can be regarded as at a high risk of developing lung cancer.

We noted slightly different thresholds for volumetry-based nodule diameter (Table 5.1). Lung cancer probability was not significantly increased in participants whose nodules measured less than 5 mm compared with those with no nodules (0.4% [95% CI 0.2-0.7]; $P = 1.00$), but was significantly increased for participants whose nodules measured 5-10 mm (1.3% [95% CI 1.0-1.8]; $P<0.0001$), and participants whose nodules measured 10 mm or more (15.2% [95% CI 12.7-18.1]; $P<0.0001$), who could be regarded as being at intermediate and high risk of developing lung cancer, respectively.

The probability of being diagnosed with lung cancer within 2 years after CT scan ac-
<table>
<thead>
<tr>
<th>Volume of largest nodule (mm³)</th>
<th>Round one</th>
<th>Round two</th>
<th>Rounds one and two</th>
<th>Lung cancer probability (95% CI)</th>
<th>P-value†</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥1000</td>
<td>36/137</td>
<td>26/104</td>
<td>62/241</td>
<td>25.7% (20.6-31.6)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>750 to &lt;1000</td>
<td>8/33</td>
<td>4/30</td>
<td>12/63</td>
<td>19.0% (11.1-30.6)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>500 to &lt;750</td>
<td>8/63</td>
<td>4/47</td>
<td>12/110</td>
<td>10.9% (6.2-18.2)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>300 to &lt;500</td>
<td>12/101</td>
<td>6/102</td>
<td>18/203</td>
<td>8.9% (5.6-13.7)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>200 to &lt;300</td>
<td>9/127</td>
<td>5/116</td>
<td>14/243</td>
<td>5.8% (3.4-9.5)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>100 to &lt;200</td>
<td>6/428</td>
<td>7/440</td>
<td>13/868</td>
<td>1.5% (0.9-2.6)</td>
<td>0.0002</td>
</tr>
<tr>
<td>50 to &lt;100</td>
<td>6/800</td>
<td>6/843</td>
<td>12/1643</td>
<td>0.7% (0.4-1.3)</td>
<td>0.07</td>
</tr>
<tr>
<td>25 to &lt;50</td>
<td>6/961</td>
<td>4/1008</td>
<td>10/1969</td>
<td>0.5% (0.3-0.9)</td>
<td>0.44</td>
</tr>
<tr>
<td>&lt;25</td>
<td>3/539</td>
<td>2/515</td>
<td>5/1054</td>
<td>0.5% (0.2-1.1)</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Maximum diameter of largest nodule (mm)§

<table>
<thead>
<tr>
<th>Maximum diameter of largest nodule (mm)</th>
<th>Round one</th>
<th>Round two</th>
<th>Rounds one and two</th>
<th>P-value†</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥30</td>
<td>3/10</td>
<td>3/9</td>
<td>6/19</td>
<td>31.6% (15.2-54.2)</td>
</tr>
<tr>
<td>20 to &lt;30</td>
<td>13/52</td>
<td>9/36</td>
<td>22/88</td>
<td>25.0% (17.1-35.0)</td>
</tr>
<tr>
<td>15 to &lt;20</td>
<td>22/84</td>
<td>7/64</td>
<td>29/148</td>
<td>19.6% (14.0-26.8)</td>
</tr>
<tr>
<td>10 to &lt;15</td>
<td>28/229</td>
<td>21/213</td>
<td>49/442</td>
<td>11.1% (8.5-14.4)</td>
</tr>
<tr>
<td>8 to &lt;10</td>
<td>7/260</td>
<td>9/296</td>
<td>16/556</td>
<td>2.9% (1.7-4.7)</td>
</tr>
<tr>
<td>7 to &lt;8</td>
<td>8/327</td>
<td>4/328</td>
<td>12/655</td>
<td>1.8% (1.0-3.2)</td>
</tr>
<tr>
<td>6 to &lt;7</td>
<td>1/371</td>
<td>2/331</td>
<td>3/702</td>
<td>0.4% (0.1-1.3)</td>
</tr>
<tr>
<td>5 to &lt;6</td>
<td>7/628</td>
<td>5/721</td>
<td>12/1349</td>
<td>0.9% (0.5-1.6)</td>
</tr>
<tr>
<td>4 to &lt;5</td>
<td>3/799</td>
<td>1/776</td>
<td>4/1575</td>
<td>0.3% (0.1-0.7)</td>
</tr>
<tr>
<td>&lt;4</td>
<td>2/429</td>
<td>3/431</td>
<td>5/860</td>
<td>0.6% (0.2-1.4)</td>
</tr>
</tbody>
</table>

Volume doubling time of fastest-growing nodule (days)‡

<table>
<thead>
<tr>
<th>Volume doubling time of fastest-growing nodule (days)</th>
<th>Round one</th>
<th>Round two</th>
<th>Rounds one and two</th>
<th>P-value†</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;100</td>
<td>7/24</td>
<td>2/10</td>
<td>9/34</td>
<td>26.5% (14.4-43.3)</td>
</tr>
<tr>
<td>100 to &lt;200</td>
<td>3/40</td>
<td>3/16</td>
<td>6/56</td>
<td>10.7% (4.7-21.8)</td>
</tr>
<tr>
<td>200 to &lt;400</td>
<td>5/130</td>
<td>7/52</td>
<td>12/182</td>
<td>6.6% (3.7-11.3)</td>
</tr>
<tr>
<td>400 to &lt;600</td>
<td>3/92</td>
<td>4/81</td>
<td>7/173</td>
<td>4.0% (1.8-8.3)</td>
</tr>
<tr>
<td>600 to &lt;800</td>
<td>0/56</td>
<td>0/74</td>
<td>0/130</td>
<td>0.0% (0.0-3.4)</td>
</tr>
<tr>
<td>800 to &lt;1000</td>
<td>0/45</td>
<td>1/63</td>
<td>1/108</td>
<td>0.9% (0.0-5.6)</td>
</tr>
<tr>
<td>≥1000</td>
<td>5/171</td>
<td>2/542</td>
<td>7/713</td>
<td>1.0% (0.4-2.1)</td>
</tr>
</tbody>
</table>

Unless otherwise stated data are n/N, where n refers to participants who were subsequently diagnosed with lung cancer and N refers to number of patients in category. *Probability of malignant disease within 2 years after a CT scan. †P-value refers to lung cancer risk for participants with nodules compared with participants without any nodules. ‡Volume of the largest non-calcified nodule; maximum diameter of the largest nodule in a participant. §Manually measured diameters are less accurate and will overestimate nodule size, which corresponds with lower lung cancer probabilities than presented in this table. ‡Maximum volume doubling time in participants whose largest nodule measured 50-500 mm³.
Figure 5.1: Trial profile.

*283 Dutch participants were randomly assigned but did not respond to the invitation for the baseline CT.
† 27 Dutch participants missed the second round CT, but were screened in the third round due to: participant declined (n=3), participant unattainable or repeated no show (n=16), still in diagnostic work-up round one (n=3), administrative error (n=5). The remaining 239 Dutch participants underwent no screening in the second round due to lung cancer (n=61), death (n=25), participant declined (n=110), participant unattainable or repeated no show (n=42), still in diagnostic work-up round one (n=1).
‡ 20 participants missed the baseline CT due to late return of their informed consent.
§ One patient did not have initial scan in round 2.

According to nodule volume-doubling time for the participants whose largest nodule measured 50-500 mm³ is presented in Table 5.1. Participants with slowly-growing (volume-doubling time ≥600 days), stable, shrunken, or resolved (ie, disappeared by follow-up CT) nodules had a low probability of lung cancer (0.0% to 1.0%). Lung cancer probability was not significantly increased for participants with nodule volume-doubling times of 600 days or more (0.8% (appendix)). However, in the total study population, the proportion of lung cancers varied as the amount of nodules per participant increased [95% CI 0.4-1.7]; P = 0.06). Lung cancer probability was significantly increased for participants with nod-
ule volume-doubling times of 400-600 days (4.0% [1.8-8.3]; \( P < 0.0001 \)), who could be regarded at low risk of developing lung cancer, and for participants with a nodule volume-doubling time of 400 days or fewer (9.9% [6.9-14.1]; \( P < 0.0001 \)), who could be regarded at high risk of developing lung cancer. Probabilities of developing lung cancer according to other categories of nodule volume and volume-doubling time (eg, volume-doubling time \( \leq 0 \), nodule resolved at follow-up, or no follow-up CT done and participants not referred for diagnostic workup) were done, but did not significantly differ from the probability for individuals without any pulmonary nodules (Table 5.1).

**Predicting lung cancer probability**

We did logistic regression analyses to predict lung cancer probability; nodule diameter, volume, volume-doubling time, and multinodularity were used as potential predictors. All four candidate predictors were significant univariate predictors (data not shown). Nodule volume, nodule volume-doubling time, and multinodularity were also significant multivariate predictors (appendix). However, the relationship between multinodularity and lung cancer risk was ambiguous: for those participants whose nodules were growing and measured 50-500 mm\(^3\), the relative proportion of participants with lung cancer decreased as the numbers of nodules per participant increased (appendix Figures 5.3 and 5.4). Therefore, we thought it appropriate to do further studies to unravel the association between multinodularity and lung cancer risk before inclusion of multinodularity in the prediction model and nodule management protocols, and so did not analyse multinodularity further in this study. Figure 5.2 shows the combined effect of nodule volume and volume-doubling time (with the final prediction model) on lung cancer probability; the interaction between volume and volume-doubling time was not statistically significant (\( P = 0.95 \)). In participants with nodules of 300 mm\(^3\) in size or larger, the lung cancer probability was substantial (from 5.9% to >50%), even in case of slow nodule growth. In participants with nodules sized 100-300 mm\(^3\), lung cancer probability ranged from less than 3% to 20%, dependent on the volume-doubling time.

Nodule management protocols, designed with either nodule volume or diameter thresholds based on lung cancer probability, or using the simulated ACCP management protocol, are presented in Table 5.2. After the first screening round (for a 1-year interval), the protocol that used nodule volume had a sensitivity of 90.9% (95% CI 81.2-96.1), and a specificity of 94.9% (95% CI 94.4-95.4). Due to its high specificity, relatively few patients would have had follow-up CT examinations (555 [8%] of 7,135) and additional diagnostic procedures (418 [6%]) compared to the other protocols. The protocol that used (volumetry-based) nodule diameter had a lower specificity than the volume protocol (90.0% [95% CI 89.3-90.7]), which would have led to more follow-up examinations (1586 [22%]), and additional diagnostic procedures (769 [11%]), but had a slightly higher sensitivity for lung cancer (92.4% [95% CI 83.1-97.1]). The simulated ACCP protocol had a sensitivity of 90.9% (95% CI 81.2-96.1), and the lowest specificity of the three evaluated protocols (87.2% [95% CI 86.4-87.9], and would have led to the most follow-up CT examinations (2,125 [30%]) and additional diagnostic procedures (968 [14%]). Performance of the lung cancer probability-based volume and diameter protocols in the second screening round with the same thresholds is provided in the appendix.
### Table 5.2: Performance assessment of simulated nodule management protocols for CT-detected nodules at the first screening round.

<table>
<thead>
<tr>
<th>Screening result†</th>
<th>Management protocol based on volumetry</th>
<th>Management protocol based on diameter*</th>
<th>Simulated management protocol of the ACCP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>volume $\geq$ 300 mm$^3$</td>
<td>diameter $\geq$ 10 mm</td>
<td>diameter $\geq$ 8 mm</td>
</tr>
<tr>
<td></td>
<td>volume $\geq$ 100 to $\leq$ 300 mm$^3$</td>
<td>diameter $\geq$ 5 to $&lt;10$ mm</td>
<td>diameter $\geq$ 4 to $&lt;8$ mm†</td>
</tr>
<tr>
<td></td>
<td>volume $&lt;$ 100 mm$^3$</td>
<td>diameter $&lt;$ 5 mm</td>
<td>diameter $\leq$ 4 mm</td>
</tr>
<tr>
<td>Indeterminate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Screen test results**

<table>
<thead>
<tr>
<th></th>
<th>Management protocol based on volumetry</th>
<th>Management protocol based on diameter*</th>
<th>Simulated management protocol of the ACCP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct referral due to positive result (n=7,135)</td>
<td>334 (5%)</td>
<td>375 (5%)</td>
<td>635 (9%)</td>
</tr>
<tr>
<td>Follow-up examination due to indeterminate result (n=7,135)</td>
<td>555 (8%)</td>
<td>1586 (22%)</td>
<td>2125 (30%)</td>
</tr>
<tr>
<td>- positive result after follow-up examination</td>
<td>84 (1%)</td>
<td>394 (6%)</td>
<td>333 (5%)</td>
</tr>
<tr>
<td>- negative result after follow-up examination</td>
<td>471 (7%)</td>
<td>1192 (17%)</td>
<td>1792 (25%)</td>
</tr>
<tr>
<td>Detected lung cancers (n=7,135)</td>
<td>60 (91%)</td>
<td>61 (92%)</td>
<td>60 (91%)</td>
</tr>
</tbody>
</table>

**Screen test parameters**

<table>
<thead>
<tr>
<th></th>
<th>Management protocol based on volumetry</th>
<th>Management protocol based on diameter*</th>
<th>Simulated management protocol of the ACCP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>90.9 (81.2 - 96.1)</td>
<td>92.4 (83.1 - 97.1)</td>
<td>90.9 (81.2 - 96.1)</td>
</tr>
<tr>
<td>Specificity</td>
<td>94.9 (94.4 - 95.4)</td>
<td>90.0 (89.3 - 90.7)</td>
<td>87.2 (86.4 - 87.9)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>14.4 (11.3 - 18.1)</td>
<td>7.9 (6.2 - 10.1)</td>
<td>6.2 (4.8 - 7.9)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>99.9 (99.8 - 100.0)</td>
<td>99.9 (99.8 - 100.0)</td>
<td>99.9 (99.8 - 99.9)</td>
</tr>
</tbody>
</table>

Data are n (%) or %, 95% CI (n/N). ACCP=American College of Chest Physicians. The test characteristics were estimated using the detection method; using a one-year interval plus all lung cancers detected in the same screening round (details are provided in the appendix). † In case of multiple nodules, the size of the largest nodule determines the screening result. * Estimates based on diameters assessed using semi-automated volumetry. Manually measured diameters are less accurate and will overestimate nodule size. As a result, the performance of the presented nodule algorithm based on diameter will be worse when manually measured diameters are used to calculate nodule size and nodule VDT. ‡ Subjects with an indeterminate screening result should undergo a follow-up scan after three months to assess the VDT; a VDT < 400 days is a positive screening result and leads to referral for diagnostic work-up. Although the lung cancer probability of nodules with VDTs of 400-600 days is intermediate (4.1% in two years), it is not possible for this analysis to classify this subgroup as indeterminate because every participant must have a definite screening test results (positive or negative) to be able to determine whether lung cancer was detected by screening or not and to calculate the test characteristics of the screening algorithm. Semi-automatically assessed nodule diameters were used for calculation of the VDT. The use of manually measured nodule diameters for the calculation of the VDT is less accurate and will affect the sensitivity and specificity of the algorithm.
Discussion

In this analysis, we used NELSON trial data to calculate the probability developing lung cancer within two years after a low-dose CT scan, and stratified this risk by nodule volume, diameter, and volume-doubling time. We used lung cancer probability to design and assess nodule management protocols. Our findings show that screened participants with nodules with volumes of 100 mm$^3$ or smaller, or diameters of 5 mm or smaller, have a lung cancer risk that is not significantly different from that in participants without nodules and should not undergo additional CT examinations. Individuals with nodules of 100-300 mm$^3$ in volume or 5-10 mm in diameter represent an indeterminate subgroup for whom assessment of volume-doubling time is appropriate (<600 days warrants follow-up evaluation). Those participants whose largest nodules’ volume measured 300 mm$^3$ or more, or had a diameter of 10 mm or more, should have immediate diagnostic evaluation.

Figure 5.2: Contour plot of the effect of the combined effect of nodule volume and VDT on the two-year lung cancer probability. The risk isolines represent the percentage of NELSON participants that will be diagnosed with lung cancer within 2 years according to the volume of their largest nodule and volume doubling time of the fastest growing nodule in the 50-500 mm$^3$ range.

In more than half of the included participants, no pulmonary nodules were detected. Their 2-year probability developing lung cancer was 0.4%, which suggests that a screening interval of at least 2 years might be safe to apply in these individuals.

Our findings support previous evidence that the probability of small nodules (volume <50 mm$^3$ or diameter <4 mm) being, or developing into, lung cancer is low: 0.6% or lower, similar to the previously reported values of less than 1% [7, 14, 19–22]. Moreover, the two-year probability of developing lung cancer in participants whose nodules measured 50-100 mm$^3$ or 4-5 mm was also low, and did not significantly differ from that in participants without nodules. At present, guidelines recommend two to four follow-up scans for such
nODULES [8, 14, 23]. Omission of these CT surveillance schedules for this patient population should be considered, because the risk of malignancy does not justify harms of ionising radiation (effective dose estimated at 10 mSv per full-dose CT) [11], psychological distress (clinically relevant increase in lung-cancer-specific distress as shown by van den Bergh and colleagues [24], and confusion, distress and frustration as reported by Wiener and colleagues [25]), and associated pressure on financial resources [26, 27].

Participants whose nodules measured 100-300 mm$^3$ (or 5-10 mm in diameter) had a significantly higher two-year lung cancer risk than did participants without nodules, which, according to current guidelines [8, 14, 23], justifies additional CT examinations. Because lung cancer risk of participants with nodules between 5 mm and 8 mm is similar (0.9% to 1.8%) [21] a uniform CT surveillance schedule could be applied, with volume-doubling time assessed at CT surveillance used to reassess lung cancer probability. Participants with slowly growing (volume-doubling time of $\geq 600$ days), stable, shrunken or resolved nodules were at low risk of developing lung cancer, and could withdraw from intensified CT surveillance [8] and return to regular screening [1, 2]. By contrast, participants whose nodules had a volume-doubling time of less than 600 days had a significantly increased risk of lung cancer which justifies intensified CT surveillance [8] and additional diagnostic procedures [1]. Participants whose nodules had a volume-doubling time of 400-600 days could be regarded as at intermediate risk, because their lung cancer probability was 4.0% (95% CI 1.8-8.3) over two years. Hence, a follow-up CT scan at short notice to reassess nodule size and growth might be a better initial option instead of more invasive diagnostic procedures.

These findings lend support to the notion that people with large nodules have a high probability of developing lung cancer, reported to be more than 10% in previous studies [8, 19, 22, 28], and 8.9% (95% CI 5.6-13.7) or higher for volumes 300 mm$^3$ or greater, or to 11.1% (8.5-14.4) diameters 10 mm or greater in this study. Risk for these large nodules remained high even when they grew slowly (Figure 5.2). However, risk of developing lung cancer for participants with large nodules that had shrunken or resolved within 2 years was very low. Although classification of large slow-growing nodules as possibly malignant might add to overdiagnosis, the risk of large nodules (defined as those measuring $\geq 300$ mm$^3$ or $\geq 10$ mm) being or developing into lung cancer is thought to be too high to delay diagnosis. Therefore, follow-up CT examinations to assess growth for large nodules provide little additional information, but may delay lung cancer diagnosis. Hence, immediate diagnostic work-up is suggested instead.

We did logistic regression analyses to predict lung cancer probability, and found that nodule diameter, volume, volume-doubling time, and multinodularity were significant univariate predictors. Nodule volume, nodule volume-doubling time, and multinodularity were also significant multivariate predictors. The interaction between nodule volume and volume-doubling time was not statistically significant; these two variables were included in the final lung cancer prediction model. The relationship between multinodularity and lung cancer risk was ambiguous; lung cancer probability varied as the number of nodules per subject increased. These findings contradict those of McWilliams and colleagues [29], who demonstrated an increased lung cancer risk for one, two, and three nodules per participant, and a decreased risk for more than four nodules per participant. Therefore, we thought it appropriate to do further studies to unravel the association between multinodularity and lung cancer risk before inclusion of multinodularity in the prediction model and nodule
LUNG CANCER PROBABILITY IN PATIENTS WITH LUNG NODULES

Based on these findings, we proposed and evaluated nodule management protocols, based on a two-step management approach as described above. Participants without nodules, or nodules smaller than the lower thresholds were to be classified as negative, and receive no additional diagnostic procedures. Participants whose nodules measured between the lower and upper thresholds were to be classified as indeterminate. Participants whose nodules are larger than the upper size threshold were to be classified as positive, and were directly referred for diagnostic work-up to diagnose or rule lung cancer. Participants who were classified as indeterminate were to undergo another low-dose CT examination to determine their final screening test result based on nodule growth using a single volumedoubling time threshold. The advantage of nodule management protocols using a two-step approach compared to protocols that use just one nodule evaluation (e.g., as used in the ELCAP [7] and the NLST [4] trials) is a single low-dose CT examination is given at short notice (for example after three months) for indeterminate nodules, instead of 2-3 CT scans in two years [8]. Further, this approach allows for a better risk stratification by nodule volume-doubling time, which is a strong lung cancer predictor [3, 5, 14].

The protocol that used lung cancer probability-based diameter thresholds was more sensitive than the simulated ACCP protocol, and would have led to fewer CT examinations and additional diagnostic procedures. Nonetheless, these results imply that the simulated ACCP nodule management protocol performs well, but improvements are possible.

The protocol that used lung cancer probability-based thresholds for nodule volume had high specificity, and would have led to substantially fewer follow-up CT examinations and additional diagnostic procedures than would the simulated ACCP protocol. Moreover, this protocol was as sensitive as the simulated ACCP protocol. However, if manual diameter measurements had been used instead of volumetry-based measurements, as recommended in the ACCP protocol, it is unlikely that such high sensitivity values would have been reached due to the intrinsic unreliability of manual measurements [30]. We believe that the advantages of an increase in specificity of the volume protocol indicate that lung cancer screening should be performed using volumetric software, despite the fact that volumetry demands more advanced CT equipment and takes more time than manual nodule measurements. Moreover, the use of volumetry enables reliable nodule growth assessment at short notice, which is not possible when manual nodule measurements are used, due to the lower sensitivity for actual nodule growth as a result of measurement error.

Analyses in this study were done at the participant level by using the largest and fastest growing nodule in participants with multiple nodules. This approach is recommended by the ACCP [14], and accounts for the fact that some interval cancers could not be matched to a nodule previously detected by screening. Lung cancer probability of the largest or fastest growing nodule in a participant could be a slight overestimate, as lung cancer was not always diagnosed in this nodule. Also, the presented lung cancer probabilities may be slightly overestimated due to advancing lung cancer diagnoses by screening in the 2-year follow-up. However, the probabilities may also be slightly underestimated because some lung cancers diagnosed as the two-year follow-up period may not have been present at the time of screening.

A limitation of this study is the inability of the LungCARE software to calculate volume of
sub-solid nodules, and so we had to estimate some volumes based on manually measured diameters, which may have introduced some inaccuracies. Another limitation may be the length of follow-up, which was limited to two years. As a result, we cannot provide results to aid decision making for nodule management for periods longer than 2 years. Moreover, presented lung cancer probabilities may only be extrapolated to populations with a comparable prevalence of lung nodules (about 50%) [3], and a comparable lung cancer risk (about 1.3% in 2 years) [15].

Lastly, presented lung cancer probabilities, volume-doubling times, and nodule protocols were all estimated and evaluated using a data set of nodule measurements that were mainly assessed using volumetry. Evaluation of two nodule management protocols using diameter was done under the assumption that nodule diameters measured using semi-automatic volumetry software were comparable to manually measured nodule diameters. However, measurement error of manual measurement of nodule diameter is larger than measurement error of the volumetry-based diameters we used in this study [30–34]. Further, calculations of volume-doubling time based on manually measured nodule diameters are less accurate than calculations of volume-doubling time based on semi-automated volumetry. As a result, the relationship between nodule diameter and lung cancer probability may be weaker for manually measured nodule diameters. In addition, when results of this study are applied to manually measured diameters, presented sensitivities and specificities of protocols using diameter are likely to be too high, and the false-positive rate, number of follow-up CTs and diagnostic work-ups are likely to be too low. These discrepancies could be reduced by using the mean transverse nodule diameter instead of maximal nodule diameter. Nonetheless, the aforementioned theoretical discrepancies in lung cancer probability and performance characteristics are probably limited in practice, as our estimates of lung cancer probability are comparable to the probabilities published by the ELCAP, NLST, and the Pan-Canadian Early Detection of Lung Cancer Study, which used manual measurements of nodule diameters for analyses [9, 10, 22, 29]. Since our conclusions are restricted to volumetry-based diameter analysis, it remains unclear whether the protocol using manually-measured diameters with the thresholds of 5 mm and 10 mm, can be applied to situations in which it is not possible to use semiautomatic volumetric software.

In the current study, nodule size and volume-doubling time were used to determine an individual’s lung cancer probability. Other nodule characteristics, such as nodule attenuation and multiplicity, and background characteristics, such as age and smoking history, may also affect lung cancer probability [29]. Future studies need to determine whether we could include such characteristics in our prediction model to estimate an individual’s lung cancer probability more accurately. Further, validation of presented lung cancer probabilities on a large, reliable data set would be valuable.

Conclusion

We designed improved management protocols for CT detected nodules, using thresholds for nodule size and VDT that are based on lung cancer probability. Subjects with nodules \( \leq 100 \text{ mm}^3 \) or \( \leq 5 \text{ mm} \) have a lung cancer risk that is not significantly different from that in subjects without nodules and should not undergo additional CT examinations. Individuals with nodules 100–300 mm\(^3\) or 5-10 mm represent an indeterminate subgroup for whom assessment of VDT is appropriate (<600 days warrants follow-up evaluation). Lung cancer risk of subjects whose nodules measure >300 mm\(^3\) or >10 mm demands
immediate diagnostic evaluation.

**Research in context: Systematic Review**

A systematic review was done as part of planning for this trial. To identify all relevant articles on management of solitary pulmonary nodules, we searched PubMed with the terms “lung neoplasms” [MeSH] AND “solitary pulmonary nodule” [MeSH] AND “tomography, x-ray computed” [MeSH] AND “probability” [MeSH]; limits: humans, adults; published in the past 10 years, in English, in core clinical journals, or MEDLINE. To identify all articles of lung cancer CT screening trials that described pulmonary nodules, we used the terms “lung neoplasms” [MeSH] AND “early detection of cancer” [MeSH] AND “tomography, x-ray computed” [MeSH] AND “epidemiologic study characteristics as topic” [MeSH]. The search was limited to studies done in adults, and published from Jan 1, 2000, in English. Titles and abstracts of articles that were identified with these search terms were scanned to select articles relevant for this study. Reference lists of relevant articles were checked to identify more relevant articles. Current clinical practice guidelines on management of pulmonary nodules use thresholds for nodule diameter to determine appropriate follow-up strategy. In addition, use of prediction models to assess individual lung cancer risk is recommended by some guidelines. Data used to design current clinical practice guidelines is mainly obtained from published results of lung cancer screening cohort studies conducted in the 1990s.

**Interpretation**

Published probabilities of lung cancer stratified by nodule size were comparable to the probabilities estimated in our study. However, none of the published studies provided estimates for such small ranges of diameters, as in our study. Moreover, no estimates of lung cancer probability were published for nodule volume and nodule VDT. This retrospective analysis showed that the simulated ACCP guidelines performed well when volumetry-based diameter measurements were used, but also that improvements were possible. By small adjustments of thresholds for nodule size and growth rate, which were determined based on the associated lung cancer probability, sensitivity and specificity of the simulated ACCP protocol may be increased. Further, this study evaluated a nodule management protocol with lung cancer probability-based thresholds for nodule volume and volume-doubling time, which yielded the same sensitivity as the simulated ACCP guideline and a substantially higher specificity. These results imply that use of lung cancer probability-based thresholds for nodule size and growth and volumetry in nodule management protocols can improve lung cancer detection, and reduce unnecessary follow-up CTs, invasive diagnostic procedures and costs.

**Acknowledgements**

We thank M. Quak, R. Faber, F. Santegoets, L. van Dongen, A. de Bruijn (Erasmus University Medical Center Rotterdam, the Netherlands), H. Ziengs (University Medical Center Groningen, the Netherlands), A. Hamersma, S. van Amelsvoort-van de Vorst (University Medical Center Utrecht, the Netherlands), L. Peeters (University Hospital Gasthuisberg
Leuven, Belgium). Finally, we would like to thank the Dutch Cancer Registry for the data linkages that identified the interval lung cancers. The NELSON trial is supported by Zorg Onderzoek Nederland-Medische Wetenschappen (ZonMw) and KWF Kankerbestrijding. Roche diagnostics provided a grant for the performance of proteomics-research. Siemens Germany provided four digital workstations and LungCARE for the performance of 3D-measurements.

References


Appendix

NELSON nodule management protocol

Below is a more detailed description of the NELSON nodule management protocol. At the first detection of a pulmonary nodule, it is classified according to its size:

NODCAT 1:
- Only nodules with benign features (e.g. benign calcification patterns, fat component)

NODCAT 2:
- Solid nodules <50 mm$^3$
- Solid pleural based nodules <5 mm in minimal diameter
- Part-solid nodules, non-solid component <8 mm in mean diameter
- Part-solid nodules, solid component <50 mm$^3$
- Non-solid nodules <8 mm in mean diameter

NODCAT 3:
- Solid nodules 50-500 mm$^3$
- Solid pleural based nodules 5-10 mm in minimal diameter
- Part-solid nodules, non-solid component ≥8 mm in mean diameter
- Part-solid nodules, solid component 50-500 mm$^3$
- Non-solid nodules ≥8 mm in mean diameter

NODCAT 4:
- Solid nodules >500 mm$^3$
- Solid, pleural based nodules >10 mm in minimal diameter
- Part-solid nodules, solid component >500 mm$^3$

If a nodule is detected at the second and later screenings, it is classified according to its growth rate. First the percentage volume change is calculated. If this percentage volume change is >25%, VDT is calculated, which categorizes the nodules as follows:

GROWCAT A
- VDT >600 days

GROWCAT B
- VDT 400-600 days

GROWCAT C
Referral algorithm of the first screening round:
Negative:
- NODCAT 1
- NODCAT 2
- NODCAT 3 with GROWCATs A or B at follow-up examination
Positive:
- NODCAT 3 with GROWCAT C at follow-up examination
- NODCAT 4

Referral algorithm of the second screening round:
Negative:
- NODCAT 1
- NODCAT 2 with GROWCATs A or B at follow-up examination
- NODCAT 3 with GROWCATs A or B at follow-up examination
Positive:
- NODCAT 2 with GROWCAT C
- NODCAT 3 with GROWCAT C
- NODCAT 4

The screening result could be negative (invitation for the next screen round), indeterminate (invitation for a repeat scan to determine the VDT), or positive (referral for diagnostic work-up). Nodule volume determined the screen result for newly detected nodules: <50 mm$^3$ was negative, 50-500 mm$^3$ was indeterminate, and >500 mm$^3$ was positive. For previously detected nodules, VDT was calculated and determined the screening result: >600 days was negative, 400-600 days was indeterminate and <400 days was positive. The protocol allowed radiologists to adjust the screening result in case of inaccurate measurements by LungCARE, high suspicion of malignancy (e.g. new solid component in non-solid nodule), or high suspicion of benignity (e.g. benign calcification pattern).

Framework for evaluating alternative nodule management protocols

The referral decisions made in the NELSON trial were based on the aforementioned formal NELSON protocol. Using the results of the NELSON trial, we can also assess how alternative nodule management protocols would have performed, if they had they been implemented in the NELSON trial. A complication in the analysis is that if an alternative protocol advised follow-up scanning to assess VDT, this VDT could only be calculated for subjects who received a follow-up scan in the NELSON trial. Below we describe the
framework we used to estimate the lung cancer probabilities and the test characteristics of the evaluated nodule management protocols.

The evaluated protocols differ in several important ways from the original NELSON protocol. First, a single set of nodule size thresholds based on volume or diameter was used for all nodule types. Also, for nodules for which the volume could not be calculated using the volumetric software, the volume \( V \) was imputed using the maximal diameter \( D \) (formula: \( V = \frac{4}{3} \pi r^3 \)). For part-solid nodules, only the solid component was used to determine the nodule size category. Finally, the criterion that the percentage volume change should be \( >25\% \) before calculating the VDT was ignored.

Each evaluated protocol uses a nodule size threshold for a negative screening and a nodule size threshold for a positive screening. These two thresholds are based on either the volume or the diameter of a nodule. In each protocol, each detected nodule was classified as negative, indeterminate, or positive according to the following rules.

Negative: Nodules with benign features (e.g. benign calcification patterns, fat component; NODCAT 1 in the NELSON protocol) and nodules with volume/diameter below the nodule size threshold for a negative screening. The VDT is not relevant for these nodules since the participant is not referred even when the nodule is growing fast. Hence, when VDT was missing, it was not imputed.

Indeterminate: Nodules with volume/diameter above the threshold for a negative screening and below the threshold for a positive screening. For the participants with at least one indeterminate nodule and no positive nodules, the VDT determines whether the participant should be referred. For newly detected nodules, the VDT was calculated using a comparison of the volume on the initial scan and the first available follow-up scan in the same round; if no follow-up scan was available or if no growth was observed, the VDT could not be calculated. For nodules observed on the second round scan that had previously been seen on the baseline scan, we calculated the VDT by comparing the volumes on the baseline scan and the second round scan.

Positive: Nodules with volume/diameter above the threshold for a positive screening. The VDT is not relevant for these nodules since the participant should be referred, even in case of slow nodule growth. Hence, when VDT was missing, it was not imputed.

Participants with at least one positive nodule should be referred and participants with no nodules or only negative nodules should not be referred. For the remaining participants (i.e. participants with at least one indeterminate nodule and no positive nodules), the referral decision was based on the following rules:

1. For the evaluation of the simulated ACCP algorithm: participants with at least one indeterminate nodule with a VDT \( \leq 400 \) days are referred; participants in whom all indeterminate nodules have VDT \( >400 \) days are not referred. For the evaluation of the two new algorithms: participants with at least one indeterminate nodule with a VDT \( \leq 600 \) days are referred; participants in whom all indeterminate nodules have VDT\( >600 \) days are not referred.

2. If the VDT of a nodule could not be calculated because the nodule had not grown or was not visible on the follow-up scan, this did not lead to a decision to refer the participant. If the VDT could not be calculated because no follow-up scan had been made in the NELSON trial, the decision to refer the patient was imputed using
the referral decision made by the radiologists in the NELSON trial. This approach was necessary in approximately 15% of the subjects with the largest nodule in the 50-500 mm$^3$ range, e.g. due to manual adjustments of the screening result by the radiologists.

**Methods for estimating screening test characteristics**

The nodule management algorithms that were evaluated in this study classified each scan result as positive, indeterminate, or negative. In all evaluated algorithms, subjects with an indeterminate screening result receive a second CT examination and the result of this scan was either positive ($VDT < 400$ days) or negative ($VDT \geq 400$ days). Summarizing, all scans have a ‘final’ screening result that was either positive or negative.

Next, whether a lung cancer was present at the time of the CT examination was determined as follows. A screening was classified as being done in the presence of lung cancer if:

- The diagnostic work-up, which was initiated for a positive ‘final’ screening result, led to the diagnosis of lung cancer (true positive screening results).
- A lung cancer diagnosis was made during the period from the first CT examination of the screening round to either the next screening round or one year later, whichever came first (false negative screening results).

Via linkages with the national cancer registry, which has complete national coverage, all lung cancer diagnoses made outside the screening trial were obtained. If the screening was not classified as being done in the presence of lung cancer, it was defined as being done in the absence of lung cancer.

Finally, definitions of the screening test parameters were defined as follows:

- Sensitivity was estimated by dividing the number of true positive screens by the numbers of true positive and false positive screens (positive screens in the absence of lung cancer).
- Specificity was estimated by dividing the number of true negative screens (negative screens in the absence of lung cancer) by the numbers of true negative and false negative screens.
- The positive predictive value was estimated by dividing all subjects with a true positive screening by all subjects with positive screening.
- The negative predictive value was estimated by dividing all subjects with a true negative screening by all subjects with negative screening.

All screening test parameters were presented with 95% binomial confidence intervals (95%CI), which were calculated using the Agresti-Coull method.
Lung cancer diagnoses not confirmed by histological specimens

Lung cancer diagnoses in the first three rounds of the NELSON trial were based on histology or cytology in 174 out of 187 cases (93.0%). The basis for the diagnosis in the 13 participants without histology or cytology was:

1. Tumor in the right upper lobe, volume 1,502 mm$^3$, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to cardiac impairment.

2. Tumor in left lower lobe, volume 2,687 mm$^3$, PET positive, cT1aN0M0, patient did not undergo thoracic surgery due to COPD stage IV.

3. Tumor in left lower lobe, volume 2,792 mm$^3$, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to COPD and renal failure.

4. Tumor in right upper lobe, volume 580 mm$^3$, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to metastasized prostate carcinoma.

5. Tumor in right lower lobe, volume 2,793 mm$^3$, PET-positive, cT1bN0M0, patient did not undergo thoracic surgery due to poor pulmonary function.

6. Tumor in right upper lobe, volume 891 mm$^3$, PET indeterminate, cT1aN0M0, patient died due to bowel ischemia just before intended thoracic surgery.

7. Tumor in right lower lobe, volume 731 mm$^3$, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to poor pulmonary function.

8. Tumor in left lower lobe, volume 108 mm$^3$, VDT 125 days, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery because he also participated in another study and was randomized to the radiotherapy treatment arm.

9. Tumor in right upper lobe, volume 383 mm$^3$, VDT 289 days, PET indeterminate, cT1aN0M0, patient did not undergo thoracic surgery because he refused, he was treated with stereotactic radiotherapy instead.

10. Tumor in left lower lobe, diameter 10 mm, PET positive, cT1aN1M0, patient did not undergo thoracic surgery due to poor pulmonary function.

11. Tumor in right upper lobe, diameter 13.2x11.6 mm, PET positive, cT1aN0M0, patient did not undergo thoracic surgery due to poor pulmonary function and general condition.

12. Tumor in right upper lobe, diameter 19.2x12.7 mm, PET positive, cT1bN0M0, patient did not undergo thoracic surgery due to poor general condition.
Table 5.3: Multivariable logistic regression model for the probability to be diagnosed with lung cancer.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodule volume</td>
<td>2.12 (1.64-2.75)*</td>
</tr>
<tr>
<td>Nodule VDT</td>
<td>0.45 (0.35-0.60)*</td>
</tr>
<tr>
<td>Constant</td>
<td>1.35 (0.24-7.79)</td>
</tr>
</tbody>
</table>

In this model, only the participants in whom the largest detected nodule had a volume of $\geq 50 \text{ mm}^3$ and $< 500 \text{ mm}^3$ and who had at least two screenings were included. The dependent variable indicates whether a diagnosis of lung cancer has occurred during the follow-up period; the independent variables are volume, VDT, and a constant term. Hosmer-Lemeshow goodness-of-fit test: $P=0.66$.

Abbreviations: VDT = volume-doubling time, 95% CI = 95% confidence interval using the Agresti-Coull method.

* Linear effect: nodule volume was defined as the volume in $\text{mm}^3$ divided by 100.

+ Logarithmic effect: nodule VDT was defined as the natural logarithm of VDT in days.

* $P$-value $< 0.001$.

Table 5.4: Two-year lung cancer probability by nodule volume and volume doubling-time.

<table>
<thead>
<tr>
<th>Nodule volume doubling time</th>
<th>Nodule volume &lt; 600 days</th>
<th>$\geq 600$ days</th>
<th>shrunken or resolved</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt; 100 \text{ mm}^3$</td>
<td>3.5%</td>
<td>0.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$\geq 100$ to $&lt; 300 \text{ mm}^3$</td>
<td>9.2%</td>
<td>0.8%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$\geq 300 \text{ mm}^3$</td>
<td>21.3%</td>
<td>5.9%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

*Subjects with an indeterminate screening result should undergo a follow-up scan after three months to assess the VDT; a VDT $< 600$ days is a positive screening result and leads to referral for diagnostic work-up. Although the lung cancer probability of nodules with VDTs of 400-600 days is intermediate (4.1% in two years), it is not possible for this analysis to classify this subgroup as indeterminate because every participant must have a definite screening test results (positive or negative) to be able to determine whether lung cancer was detected by screening or not and to calculate the test characteristics of the screening protocol. Semi-automatically assessed nodule diameters were used for calculation of the VDT. The use of manually measured nodule diameters for the calculation of the VDT is less accurate and will affect the sensitivity and specificity of the protocol.
Figure 5.3: Relationship multi-nodularity and lung cancer probability in all subjects with nodules.

Figure 5.4: Relationship multi-nodularity and lung cancer probability in subjects whose largest nodule measure 50-500 mm$^3$ and have a VDT >0 days.
Table 5.5: Performance assessment of simulated nodule management protocols for CT-detected nodules at the second screening round.

<table>
<thead>
<tr>
<th>Screening result†</th>
<th>Management protocol based on volumetry</th>
<th>Management protocol based on diameter*</th>
<th>Simulated management protocol of the ACCP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>volume ≥300mm³</td>
<td>diameter ≥10 mm</td>
<td>diameter ≥8mm</td>
</tr>
<tr>
<td></td>
<td>volume ≥100 to ≤300mm³†</td>
<td>diameter ≥5 to &lt;10mm†</td>
<td>diameter &gt;4 to &lt;8mm‡</td>
</tr>
<tr>
<td></td>
<td>volume &lt;100mm³</td>
<td>diameter &lt;5mm</td>
<td>diameter ≤4mm</td>
</tr>
<tr>
<td>Indeterminate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Screen test results**

<table>
<thead>
<tr>
<th></th>
<th>Percentage (95%CI)</th>
<th>Percentage (95%CI)</th>
<th>Percentage (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct referral due to positive result (of total)</td>
<td>4.1</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>Follow-up examination due to indeterminate result (of total)</td>
<td>8.1</td>
<td>24.3</td>
<td>24.3</td>
</tr>
<tr>
<td>- positive result after follow-up examination</td>
<td>1.1</td>
<td>5.4</td>
<td>5.4</td>
</tr>
<tr>
<td>- negative result after follow-up examination</td>
<td>7.0</td>
<td>19.0</td>
<td>18.9</td>
</tr>
<tr>
<td>Detected lung cancers (of total)</td>
<td>83.1 (71.3 - 90.7)</td>
<td>86.4 (75.2 - 93.2)</td>
<td>88.1 (77.2 - 94.4)</td>
</tr>
</tbody>
</table>

**Screen test parameters**

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>83.1 (71.3 - 90.7)</td>
<td>95.5 (95.0 - 96.0)</td>
<td>13.8 (10.6 - 17.8)</td>
<td>99.9 (99.7 - 99.9)</td>
</tr>
<tr>
<td></td>
<td>86.4 (75.2 - 93.2)</td>
<td>90.6 (89.9 - 91.3)</td>
<td>7.4 (5.6 - 9.6)</td>
<td>99.9 (99.7 - 99.9)</td>
</tr>
<tr>
<td></td>
<td>88.1 (77.2 - 94.4)</td>
<td>88.6 (87.8 - 89.3)</td>
<td>6.3 (4.8 - 8.1)</td>
<td>99.9 (99.8 - 99.9)</td>
</tr>
</tbody>
</table>

Data are n (%) or %, 95% CI (n/N). ACCP=American College of Chest Physicians. The test characteristics were estimated using the detection method; using a one-year interval plus all lung cancers detected in the same screening round. † In case of multiple nodules, the size of the largest nodule determines the screening result. * Estimates based on diameters assessed using semi-automated volumetry. Manually measured diameters are less accurate and will overestimate nodule size. As a result, the performance of the presented nodule algorithm based on diameter will be worse when manually measured diameters are used to calculate nodule size and nodule VDT. ‡Subjects with an indeterminate screening result should undergo a follow-up scan after three months to assess the VDT; a VDT<600 days is a positive screening result and leads to referral for diagnostic work-up. Although the lung cancer probability of nodules with VDTs of 400-600 days is intermediate (4.1% in two years), it is not possible for this analysis to classify this subgroup as indeterminate because every participant must have a definite screening test results (positive or negative) to be able to determine whether lung cancer was detected by screening or not and to calculate the test characteristics of the screening algorithm. Semi-automatically assessed nodule diameters were used for calculation of the VDT. The use of manually measured nodule diameters for the calculation of the VDT is less accurate and will affect the sensitivity and specificity of the algorithm. ‡ Subjects with an indeterminate screening result should undergo a follow-up scan after three months to assess the VDT; a VDT<400 days is a positive screening result and leads to referral for diagnostic work-up, according to the ACCP guideline [14].