General discussion

Partly published as:
Marjolein A Heuvelmans
Matthijs Oudkerk

Management of subsolid pulmonary nodules in CT lung cancer screening

The studies described in this thesis contribute to the optimization of the nodule management protocol in computed tomography (CT) lung cancer screening. The merits and shortcomings of the described studies have been discussed in the previous chapters. This chapter will provide a more general discussion of the main findings, and will consider methodological issues. Furthermore, this chapter will comment on the clinical relevance of the findings, and give suggestions for further research to optimize management of screen-detected lung nodules.

**Background**

In view of the prospective results of the National Lung Screening Trial (NLST), and baseline results of other trials, interest in CT for lung cancer screening in high-risk individuals is increasing. In 2011, the U.S. NLST demonstrated that screening using low-dose CT (LDCT) reduces lung cancer mortality by 20% compared to screening by chest radiography [1]. This result was translated by several U.S. medical associations, including the U.S. Preventive Services Task Force, into a recommendation to screen subjects at high-risk for developing lung cancer by LDCT [2–6]. According to the recommendation of the U.S. Preventive Services Task Force, all individuals between 55 and 80 years old who smoked at least 30 pack-years and quit not longer than 15 years ago are eligible for lung cancer screening. Screening should be discontinued once a person has not smoked for 15 years or develops a health problem that substantially limits life expectancy or the ability or willingness to have curative lung surgery [6].

A drawback of CT screening is the high prevalence of small to intermediate-sized (<500 mm$^3$) lung nodules, most of which are benign. Up to 66% of participants enrolled in CT screening trials has at least one pulmonary nodule [7]. Overview of the differential diagnosis of a pulmonary nodule is given in Table 12.1 [8]. Solid lung nodules are the most common type of nodule found at CT lung cancer screening [9, 10]. Accurate nodule management is required to differentiate between benign and malignant lung nodules. In the Dutch-Belgian lung cancer screening trial (NELSON trial, a Dutch acronym for Nederlands-Leuvens Longkanker Screening Onderzoek), pulmonary nodule management is based on volume of nodules at baseline and newly detected nodules at subsequent rounds, and volumedoubling time (VDT) of pre-existing nodules on repeat scans with a change of volume of 25% between two subsequent CTs (percentage volume change [PVC]) [11]. The NELSON trial is the largest randomized lung cancer screening trial in which lung cancer screening by low-dose chest CT is compared to no screening.

Pulmonary nodules with large volumes or high growth rates are more likely to be malignant. In the NELSON trial, participants with a nodule with volume >500 mm$^3$ or VDT <400 days were referred to a pulmonologist for workup and diagnosis. This management protocol yielded a low rate of positive screening tests compared to other screening trials, while the number of missed cancers was low [9]. Still, the majority of suspicious lung nodules that resulted in referral to a pulmonologist turned out to be benign.

A nodule management strategy should allow sensitive and timely diagnosis of malignant nodules, while for benign nodules, patient anxiety, cost and morbidity associated with unnecessary diagnostic procedures should be minimized. Therefore, the main goal of this
Table 12.1: Differential diagnosis of a lung nodule.* [8]

<table>
<thead>
<tr>
<th>Benign</th>
<th>Malignant (1.1% - 12%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonspecific granuloma (15%-25%)</td>
<td>Adenocarcinoma (47%)</td>
</tr>
<tr>
<td>Hamartoma (15%)</td>
<td>Squamous cell carcinoma (22%)</td>
</tr>
<tr>
<td>Infectious granuloma (15%)</td>
<td>Metastatic disease (8%)</td>
</tr>
<tr>
<td>Aspergillosis</td>
<td>Small cell lung carcinoma (4%)</td>
</tr>
<tr>
<td>Coccidioidomycosis</td>
<td>Others: large cell carcinoma, carcinoid tumors, lymphomas, malignant teratomas</td>
</tr>
<tr>
<td>Cryptococcosis</td>
<td>Others: lung abscess, round pneumonia, bronchogenic cysts, focal hemorrhage, hemangiomas</td>
</tr>
<tr>
<td>Histoplasmosis</td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td></td>
</tr>
</tbody>
</table>

*Incidentally or clinically detected lung nodule, without considering lung cancer pre-test probability.

thesis was to determine the optimal nodule volume and VDT cutoff points for differentiating benign from malignant pulmonary nodules in different rounds of LDCT lung cancer screening, to increase the efficiency of lung cancer detection and to reduce the number of false-positive results. The studies described in this thesis contribute to the optimization of the nodule management protocol in CT lung cancer screening.

Main findings

Part I: Current status of CT lung cancer screening in Europe

In 2013, an overview of all European randomized CT lung cancer screening (EUCT) trials was published [12]. EUCT includes the NELSON trial, the United Kingdom Lung Cancer Screening trial (UKLS), the Danish Lung Cancer Screening Trial (DLCST), the German Lung Cancer Screening Intervention Study (LUSI), the Multi-centre Italian Lung Detection Trial (MILD), the Italian Lung cancer Computed Tomography screening trial (ITALUNG), and the Detection and screening of early lung cancer by Novel imaging Technology and molecular assays (DANTE). By pooling of the data, EUCT hopes to be able to provide additional information for the discussion of some important sub questions regarding the implementation of lung cancer screening by LDCT, for instance for the determination of the optimal screen population, for the comparison between a volume- and diameter-based nodule management protocol, and for the determination of optimal screen intervals [12]. Pooling for the primary endpoint of the studies - decrease in lung cancer probability - is not recommended due to the limited comparability of the different trials. So far, none of the finished European lung cancer screening trials could reproduce a decrease in lung cancer mortality comparable to the NLST. Actually, in these three European trials, no significant decrease in lung cancer mortality was found, although the individual trials may have been underpowered [13–15]. The mortality results of the NELSON study, the largest European randomized lung cancer screening trial, are expected for 2016 / 2017.
Recently, a collaborative paper of the European Respiratory Society and the European Society of Radiology summarized key elements necessary for a comprehensive lung cancer screening program in Europe including minimum requirements and recommended refinements [16, 17]. These societies recommend a longitudinal comprehensive screening program at accredited medical centers for high-risk individuals (aged 55-80 years, smoked at least 30 pack-years, quit smoking maximum 15 years ago) without comorbidities precluding curative therapy. The following prerequisites were proposed:

1. Standardized procedures for image acquisition, nodule evaluation, positive results, their management and follow-up;
2. Computer-assisted nodule evaluation and documentation;
3. Identical measurement software for follow-up examinations;
4. Volumetric measurements preferred over diameter;
5. Use of at least 16-row multi-detector CT; isotropic high spatial resolution; slice thickness of 1 mm, and effective dose of 1-3 mSv;
6. Collection and submission of screening data to a lung screening registry.

These recommendations contrast with the published results of European trials so far, that all showed no advantage for lung cancer screening [13–15]. Until this consensus document, there were no positive recommendations regarding implementation of lung cancer screening in routine clinical practice or reimbursed screening programs in Europe. Although the prerequisites largely reflect CT screening procedures in the European trials, it seems these recommendations are premature, as results of the larger European trials should be awaited.

Part II: Quantitative evaluation of CT lung cancer screening

Considering the millions of individuals that are potentially eligible for lung cancer screening [18, 19], efforts should be made to design a screening program with minimal harms and costs with preserved efficiency [20]. For this, the experience in the NLST and in the other lung cancer screening trials should be used. Recently, the question was raised what target population should be invited for LDCT screening [21, 22]. However, before that question can be answered, an optimal nodule management protocol needs to be established, in terms of management of different types of nodules (solid, subsolid), measurement technique for quantification of nodule size and growth, cutoff values for nodule size and growth rate, and in terms of management for nodules, newly detected after baseline.

Management of solid and subsolid nodules in CT lung cancer screening

Early stage lung cancer most often presents as a small solid lung nodule. In about 50% of screenees, at least one pulmonary nodule is found on the screening examination. The large majority of these nodules is benign. A nodule management protocol should be sensitive for lung cancer detection, with a false-positive rate as low as reasonably achievable. In order to obviate the problem of high false-positive rates, in the NELSON study, a third
screening classification of indeterminate was defined, in addition to negative and positive screen results [11]. The introduction of a volume-based nodule management protocol and a short-term follow-up result for indeterminate nodules, instead of direct referral to a pulmonologist, led to a tenfold decrease of false-positive rate in the NELSON trial (at baseline 1.7% in NELSON versus 26.6% in NLST) [1, 9]. A number of other, European, lung cancer screening trials followed this principle [23–26]. An overview of different management protocols used in lung cancer screening trials for screen-detected solid lung nodules is given in Table 2.1.

Different U.S. associations have recently developed guidelines with positive recommendations for implementation of lung cancer screening in high-risk individuals [27–29]. In these guidelines, a distinction is made between solid and subsolid pulmonary nodules. Solid nodules are by far the most common type of screen-detected lung nodules. An overview of management guidelines for solid lung nodules at initial detection, detected incidentally or as part of screening, is given in Table 12.2. From this table can be concluded that some disagreement exist between the different guidelines on the management of solid lung nodules. In case lung cancer screening will become more widespread available, it will be important to deploy more uniform nodule management guidelines, updated based on the latest knowledge on nodule management and CT parameters.

### Table 12.2: Management of incidental and screen-detected solid pulmonary nodules in high-risk individuals, at initial detection [30].

<table>
<thead>
<tr>
<th>Referral for workup*</th>
<th>3-month FU</th>
<th>6-month FU</th>
<th>1-year FU</th>
<th>No FU required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleischner [31]</td>
<td>&gt;8mm</td>
<td>&gt;6-8mm</td>
<td>&gt;4-6mm</td>
<td>≤4mm</td>
</tr>
<tr>
<td>NCCN [27]</td>
<td>&gt;8mm</td>
<td>&gt;6-8mm</td>
<td>&gt;4-6mm</td>
<td>≤4mm</td>
</tr>
<tr>
<td>ACCP [28]</td>
<td>&gt;8mm</td>
<td>&gt;6-8mm</td>
<td>&gt;4-6mm</td>
<td>≤4mm</td>
</tr>
<tr>
<td>Lung-RADS [29]</td>
<td>≥15mm new</td>
<td>≥8-&lt;15mm new</td>
<td>≥6-&lt;8mm new</td>
<td>&lt;6mm baseline</td>
</tr>
<tr>
<td></td>
<td>≥8mm new</td>
<td>≥6-&lt;8mm new</td>
<td>≥4-&lt;6mm new</td>
<td>&lt;4mm</td>
</tr>
</tbody>
</table>

ACCP = American College of Chest Physicians. FU = follow-up. NCCN = National Comprehensive Cancer Network.  
*Chest CT with or without contrast, PET/CT and/or tissue sampling depending on the probability of malignancy and comorbidities.

Subsolid pulmonary nodules (SSNs) are a specific subtype of lung nodule found in a small number of screening participants. In the NELSON trial, nodule management of SSNs was diameter based, because of the inaccuracy of software to semi-automatically determine nodule volume. SSNs should be dealt with differently compared to solid nodules, because of their non-aggressive behavior in contrast to solid nodules. Despite the relatively high risk of malignancy in these nodules, especially in part-solid nodules, progression to cancer stage beyond stage I is very rare. The major challenge in the management of SSNs in LDCT lung cancer screening is to avoid overdiagnosis and overtreatment of non-aggressive, indolent lung cancers, but to also timely identify and avoid potential increase in cancer stage. The distinct appearance and behavior of SSNs have resulted in separate recommendations for the management of solitary SSNs at initial detection, both for incidentally detected [32] as well as for screen-detected nodules [27–29], as shown in Table 12.3. In the
different guidelines regarding screen-detected SSNs, nodule management is based on the
diameter of the largest nodule, with recommendation to repeat imaging within one year
for small-to-intermediate size SSNs [27–29]. The guidelines have been based primarily on
consensus/expert opinion, and there is little consensus between the different guidelines
(Table 12.3).

Table 12.3: Management of incidental and screen-detected SSNs in high-risk individuals,
at initial detection [30].

<table>
<thead>
<tr>
<th>Referral for workup*</th>
<th>3-month FU</th>
<th>6-month FU</th>
<th>1-year FU</th>
<th>No FU required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleischner [32]</td>
<td>Part-solid</td>
<td>GGN &gt;10mm</td>
<td>GGN &gt;5mm</td>
<td>GGN ≤5mm</td>
</tr>
<tr>
<td></td>
<td>Part-solid</td>
<td>GGN &gt;10mm</td>
<td>Part-solid</td>
<td>Part-solid</td>
</tr>
<tr>
<td></td>
<td>Part-solid</td>
<td>GGN &lt;10mm</td>
<td>GGN ≤5mm</td>
<td>GGN ≤5mm</td>
</tr>
<tr>
<td>NCCN [27]</td>
<td>Part-solid</td>
<td>GGN &gt;8mm</td>
<td>GGN &gt;10mm</td>
<td>GGN &lt;5mm</td>
</tr>
<tr>
<td></td>
<td>Part-solid</td>
<td>GGN &gt;8mm</td>
<td>GGN &lt;10mm</td>
<td>GGN &lt;4mm</td>
</tr>
<tr>
<td></td>
<td>Part-solid</td>
<td>GGN &lt;8mm</td>
<td>GGN &gt;5mm</td>
<td>GGN ≤5mm</td>
</tr>
<tr>
<td>ACCP [28]</td>
<td>Part-solid</td>
<td>GGN ≥15mm</td>
<td>Part-solid</td>
<td>Part-solid</td>
</tr>
<tr>
<td></td>
<td>Part-solid</td>
<td>GGN ≥15mm</td>
<td>Part-solid</td>
<td>Part-solid</td>
</tr>
<tr>
<td></td>
<td>Part-solid</td>
<td>GGN ≥15mm</td>
<td>GGN &gt;5mm</td>
<td>GGN ≤5mm</td>
</tr>
<tr>
<td>Lung-RADS [29]</td>
<td>Part-solid</td>
<td>GGN ≥6mm; SC ≥6mm; GGN ≥6mm; SC &lt;6mm; GGN &gt;20mm</td>
<td>Part-solid</td>
<td>GGN &lt;20mm</td>
</tr>
</tbody>
</table>


*Chest CT with or without contrast, PET/CT and/or tissue sampling depending on the probability of malignancy and comorbidities.

Recently, two studies were published regarding SSNs detected in LDCT lung cancer screening trials, including management advice [33, 34]. It was concluded that immediate re-
section of SSNs is often not desirable; close follow-up of SSNs by annual LDCT usually
is sufficient because of non-aggressiveness of subsolid lung cancers. Implementation of
the results of Yankelevitz et al. [34] and Scholten et al. [33] contributes to the optimization
of management of screen-detected SSNs, in terms of reduction of overdiagnosis and overtreatment. Future research should show if biannual instead of annual follow-up of
screen-detected SSNs does not increase overall mortality and morbidity rates.

Accuracy of diameter and volume measurements

In the NLST, radiologists manually measured the diameter of all non-calcified nodules [35].
This method is also applied in current U.S. guidelines for incidentally and screen-detected
lung nodules (Tables 12.2 and 12.3). The term diameter assumes that a nodule can be
fairly represented by a sphere (or perhaps ellipsoid). Since Nature usually does not cre-
ate perfect geometrically shaped pulmonary nodules, errors in the estimation of nodule
size may result. The alternative method is to measure nodule volume using software for
semi-automated measurements. This software enables an accurate estimation of nodule size via contour finding of the lesions, after three-dimensional reconstruction. The latter was applied in most European lung cancer screening trials [11, 23–25]. They have chosen to adopt the, at that time novel, volumetric measurement technique for its higher precision and smaller measurement bias [36–39], and higher reliability (better agreement among tests) [36, 37, 40] compared to manual diameter nodule measurements. This is especially relevant for nodule growth assessment, for example at the short-term repeat CT for indeterminate-sized nodules, part of the volume-based nodule management algorithms [11, 23–25]. As recently elucidated [41], diameter assessment is already on theoretical grounds less sensitive to detect growth of lung nodules, compared to volumetry. Add to this the reduced reproducibility in size assessment based on diameters (manual versus semi-automated measurement, and variability in diameter assessment as translated in volume), and it is nearly self-evident that volume measurements are to be preferred for lung nodule evaluation.

Most diameter-based nodule management protocols are based on mean nodule diameter, defined as the mean of maximum axial diameter and perpendicular diameter in the XY-plane [31, 32]. In practice, radiologist’s adherence to guidelines on lung nodule management is moderate [42, 43], and it is not clear if radiologists always measure nodule diameters according to the protocol in case of mean diameter-based nodule management. It is likely that in a number of cases, management is based only on the maximum transverse diameter, instead of mean diameter, since maximum transverse cross-sectional measurements are the most commonly used nodule measurements at most clinics, and reflects practice in treatment response evaluation [44].

Previous research showed that two-dimensional CT measurements of small-to-intermediate sized nodules (<20 mm) are unreliable [39]. In the referenced study, three independent readers performed three serial measurements of maximum transverse diameter of 54 lung nodules. Both intra- and inter-reader agreement were found to be poor. Even the most consistent of the three readers had a minimum measurement error of 1.32 mm. This minimum measurement error increased to 1.73 mm in case repeated measurements of a single nodule were performed by two different readers. In the Lung-RADS guidelines for the management of screen-detected nodules [29], lung nodule growth is defined as an increase in nodule diameter of 1.5 mm between two subsequent screening examinations, which lies around the minimum measurement error for a single observer. In case two different readers would perform subsequent measurements, the likelihood that the measured difference in nodule diameter corresponds to an apparent change in nodule size is smaller than 5%, based on the previous study [39].

When directly comparing the approach of diameter-based versus volume-based measurement and management criteria in CT lung cancer screening, we found a range in mean nodule diameter of 3.0 mm to 15.4 mm for indeterminate nodules with volume between 50 and 500 mm$^3$ (Chapter 6). Diameter-based volume overestimated semi-automated nodule volume by 47.2% (95%-CI: 44.7–49.7%), based on mean diameter, and by 85.1% (mean, 95%-CI: 81.2–89.0%), based on maximum transverse diameter. Since nodules have an infinitive number of diameters, but only one volume, it is inevitable that future management of incidentally detected and screen-detected pulmonary nodules will be based on (semi-)automated volume measurements.
Measurement of nodule growth

A wait-and-see principle is not commonly used when lung cancer is suspected, because of the usual aggressiveness of the disease. In-vivo information on growth patterns of solid lung cancers, from small nodules barely detectable by imaging techniques to histologically proven lung cancers, is therefore scarce. Nodule growth is often defined in terms of VDT, under the assumption of exponential growth. However, this pattern had never been quantified in actual patient data, and a qualitative study on lung cancer growth patterns questioned the exponential growth pattern [45]. In that study, in which growth patterns in 18 lung cancers were visually evaluated, the conclusion was based on diameter-based volumes and the majority of lung cancers was subsolid, causing inaccurate estimation of nodule size. In a smaller study (13 lung cancers), growth rates appeared approximately constant on the log scale, consistent with exponential growth [46]. However, analysis of lung cancer growth curves was limited, and restricted to a plot of the growth curves on a log scale and visual analyses.

We found that solid lung nodules, found at at least three subsequent screening CT examinations before lung cancer diagnosis was made, usually evolve at an exponential pattern with excellent fit (Chapter 8). Therefore, VDT can be used to describe the growth of lung nodules detected at LDCT lung cancer screening. This finding also endorses the inaccuracy of a fixed increase in diameter as growth criterion, as used in a.o. Lung-RADS [29]. Due to limited measurement points, and therefore scarce information among growth pattern, only solid nodules that had been found at three or more LDCT screening examinations before lung cancer diagnosis was made could be excluded. Thus, we could not gather information on growth patterns of-usually faster-growing- nodules found at only one or two LDCTs before cancer diagnosis. However, the results as found in the included participants strongly suggest an exponential growth pattern in solid screen-detected lung cancer.

Optimization of cutoff values in the current nodule management protocol

After determination of most accurate size and growth measurement, it is important to determine the most optimal cutoff values to differentiate between benign and malignant lung nodules, to thereby decrease the number of false-positive screen results. Considering all benign and malignant lung nodules in the first two screening rounds of the NELSON study, it was found that by altering the volume cutoffs for an indeterminate screen result from 50 - 500 mm$^3$ to 100 - 300 mm$^3$, test characteristics can be improved [47].

Optimization of the VDT cutoff to differentiate between benign and malignant lung nodules can decrease the false-positive rate as well (Chapter 4). The three-month interval for follow-up of indeterminate nodules after baseline was based on the 25% increase in volume criterion and VDT of 400 days, which was initially chosen as the cutoff value to identify fast-growing lung nodules [11]. We found that lowering the VDT cutoff to 232 days for fast growing lung nodules (VDT < 400 days) at three-month follow-up CT after baseline could reduce false-positives by 33% [48]. VDT cutoff of fast-growing nodules referred after the regular second-round CT could not be lowered from 400 days. This might be explained by the fact that the median VDT for these cancers was higher (98 days three months after baseline versus 205 days after the regular second-round CT), and 14/20 cancers already had a three-month follow-up CT after baseline with VDT > 400.
Hypothetically, two types of lung cancer growth rate can be distinguished. First there are very fast growing cancers: Cancer in nodules detected newly after baseline (defined as incident nodules), cancers diagnosed at time of first follow-up CT after baseline (after three months or one year, depending of baseline size), and interval cancers developed in case of longer screen interval between two CT examinations. Secondly there are slower growing cancers: Cancers diagnosed at baseline, or after being followed with multiple subsequent CT examinations before cancer diagnosis was made. The second group of cancers seems to be more often adenocarcinoma, and detected more often at early stage. For this group of cancers, a more conservative protocol by follow-up CTs - until the nodule has reached a certain size - might be considered.

We also evaluated VDT cutoff optimization, considering not only fast-growing cancers but all cancers detected in the first two rounds of the NELSON study [47]. From that study (Chapter 5), it was concluded that most optimal management protocol characteristics in the first two screening rounds, in terms of sensitivity, specificity, negative predictive value and positive predictive value could be obtained with VDT cutoff of 600 days. However, the 25% increase in nodule volume, to assure real increase in nodule volume [49] between two subsequent CT examinations, implies a maximum VDT of 280 days in case of a three-month time interval between two screening examinations. Thus, considering the minimal volume increase needed to exclude measurement variation, a VDT cutoff of maximum 280 days might be regarded for nodules detected after a three-month time interval, which supports the optimal VDT cutoff for fast-growing lung cancers (232 days). In case of longer or shorter time intervals between subsequent CT examinations, or if volume measurements can be done even more precise, maximum possible VDT cutoffs change accordingly. After 365 days, with the requirement percentage volume change >25%, maximum VDT cutoff is 1134 days. The proposed VDT cutoff of 600 days for the regular second round therefore can truly improve nodule management test characteristics.

Besides volume and VDT cutoff optimization, nodule management can be optimized in terms of more stringent management for incident nodules. In Lung-RADS, a stricter follow-up of nodules newly detected after baseline, although based on nodule diameter, has already been recommended [29]. The NELSON trial reported an overall participant cancer rate of approximately 2.6% in the incident rounds (200/7,582, including 50 new nodule cancer [50]. In participants with incident nodules, a considerably higher cancer rate of 6.2% was found in a new solid nodule (49/787 of participants with new solid nodules; 50 cancers, (Chapter 7)). Other lung cancer screening trials that studied lung cancer rate in incident nodules published similar results, although these rates were based on different methodology used in the various screening trials, which complicates comparison [20]. In I-ELCAP lung cancer probability for individuals with an incident nodule was 5.3% (1,460/27,456), in ELCAP it was 3.4% (40/1,184), and in the PluSS trial it was 7.5% [51–53]. The Mayo trial reported a higher rate of 13.1% (191/1,464) [7].

From these results, we can conclude that new solid nodules should be followed more aggressively than nodules detected at baseline screening, for example by using a lower volume cut-off (30 mm³) or shorter follow-up interval. The latter is not possible when manual diameter measurements are used, since diameter-based detection of little growth in a very short time period is impossible. In case a nodule’s diameter increases by 25%, the nodule’s volume almost doubles [41]. Although the current management protocol does
not provide separate guidelines for incident nodules, we found that meticulous screening does enable detection of a new malignant nodule often at an early, still curable stage. More research concerning incident nodules is necessary to determine how to optimize the management of these nodules in CT lung cancer screening, aiming to diagnose cancers in these nodules even earlier.

Part III: Qualitative analysis of CT imaging in lung cancer screening

Sensitive nodule detection

In lung cancer screening, failures in lung cancer diagnosis are most often due to errors of nodule detection rather than interpretation. To optimize screening efficiency, sensitive lung nodule detection and accurate nodule classification are two important issues [54–56]. The NELSON study is, up to now, the only lung cancer screening study that published in-depth results on interval and post-screen cancers. In the NELSON study, 35 lung cancers (20 interval cancers and fifteen post-screening cancers) were not diagnosed through screening [57, 58]. Retrospective radiological evaluation of these cancers showed various reasons for the failure of lung cancer diagnosis, the largest being detection error. In twenty cases, nodules suspicious of lung cancer were not found on the LDCT examination by the radiologist due to detection error. In two cases, nodules suspicious of lung cancer were found, but no further follow-up for these nodules was performed, due to interpretation errors by the radiologist. In thirteen cases, nodules suspicious of lung cancer were found, but not diagnosed as lung cancer due to failure of the protocol (two cases), non-compliance with the protocol by the participant (eight cases), or non-compliance with the protocol by the radiologist since previous diagnostic work-up of the particular nodules revealed no evidence for malignancy (three cases).

One way to improve sensitivity of nodule detection might be by using radiographers as part of the CT reading process, comparable to the role of radiographers in breast cancer screening [59–61]. This could also solve the problem of the numerous thoracic radiologists necessary to read all screening CTs, in case lung cancer screening will be implemented in routine clinical care. Recently, it was demonstrated that radiographers’ performance in lung nodule detection is comparable to radiologists as published in the literature, but lower than that of radiologists reading the same scans, and with higher false-positive rates [62]. It was concluded that radiographers are not sensitive enough to be used as first readers in CT lung screening. However, during the experiment, radiographers showed improvement in their performance both in more sensitive detection and decrease in the number of false-positive results. It was suggested that radiographers could play an additional role in nodule detection, comparable to the role of software for computer-aided detection of nodules, with the advantage of the learning effect of radiographers.

Little is known about the impact of readers on screening efficiency, especially in terms of reduction of false-positive results. In the NELSON study, we found that in baseline lung cancer screening, 5.9% of the baseline CT screen results based on the NELSON nodule management protocol were adjusted by the radiologist (Chapter 9). About 95% of screen results were adjusted downwards. In total, this led to 5.6% less short-term follow-up CT procedures or referrals after the baseline CT, while none of the nodules turned out to be cancer during two years after baseline. Therefore, it was concluded
that radiologists’ expertise can improve nodule classification in addition to a volume-based nodule management protocol. Further research on the performance of (trained) radiographers and radiologists in the reading of screening CTs is recommended.

Overdiagnosis in lung cancer screening - resolving nodules

One of the biases in lung cancer screening is overdiagnosis, defined as cancers identified by screening that, if not resected, would never be lethal because of slow growth rate or competing age-related risk for death. Another example of overdiagnosis in lung cancer screening is redundant work-up for nodules that will disappear naturally, or nodules with an obvious benign nature like completely calcified nodules and perifissural nodules [63, 64]. We have studied the prevalence and CT characteristics of resolving nodules. We found that about 10% of solid intraparenchymal nodules of intermediate size found at baseline lung cancer screening disappears during follow-up (Chapter 10). Since resolving nodules share CT features with nonresolving nodules, nodule characteristics cannot sufficiently distinguish intermediate-sized nodules that subsequently disappear. Future research should evaluate if there are features that predict nodule resolution, in order to avoid (invasive) workup for these specific nodules. Nevertheless, participants with a resolving nodule or a nodule with obvious benign nature should continue to be screened according to the protocol, since new - possible malignant - nodules may arise.

Methodological considerations

Study participants

In the NELSON study, in contrary to for example the NLST, no information was gathered on a participant’s ethnicity. It is known that lung cancer incidence is higher among certain ethnicities. In the United States, in 2012 highest lung cancer rates among men were described for black men, followed by white, Asian/Pacific Islander, American Indian/Alaska Native, and Hispanic men. Among women, white women had the highest rate of getting lung cancer, followed by black, American Indian/Alaska Native, Asian/Pacific Islander, and Hispanic women [65]. Different distribution of ethnicities among different lung cancer screening trials might therefore limit comparability of the outcomes. Furthermore, different lung cancer trials used different age and pack-years criteria for eligibility of lung cancer screening. This leads to various pre-screen lung cancer probabilities, which again limits comparability of the outcomes of different screening trials. The final mortality results of the NELSON trial should show whether a lung cancer screening program in the Netherlands would be effective and desirable.

Study protocol

One of the major strengths of the NELSON trial is that the same nodule management protocol was used during all screening rounds, as described by Xu et al [11]. In the first and second screening round, 16-row multi-detector CT scanners in low-dose setting were used. In the last two screening rounds, more modern 64-row multi-detector CT scanners, again in low-dose setting, had replaced these scanners. The influence of the use of different
scanners in the NELSON trial has not been studied yet, but it can be expected that by the improvement in CT scanners, and expected further improvement in the future, even smaller pulmonary nodules can be detected in lung cancer screening. This indicates the need of excellent nodule management protocols, optimized based on current knowledge obtained from screening trials, in case lung cancer screening will become available in daily practice.

**Nodule measurement and segmentation**

This thesis aimed to optimize management of screen-detected pulmonary nodules. Accurate three-dimensional size and growth measurements are essential in the assessment of these nodules. Variability in volume measurements may cause false-positive or false-negative screen results. Substantial variations in segmentation performance between current software packages for semi-automated measurements of lung nodule volume have been described [66]. In the referenced study, in which semi-automated volume measurements of six different software packages were compared, a variability range in measured volume of 18.5% - 25.6% was found for nodules with diameter <8 mm [67]. Within the same software package, variability in measured volume was described as well [49], therefore in the NELSON study for actual growth a PVC of at least 25% was required. This finding was confirmed in a phantom study [40].

Segmentation that includes surrounding structures, or does only include a nodule partly, may lead to inaccurate measurements and wrong management decisions, logically [66]. Segmentation errors in semi-automated volume measurements are often due to adjacent structures; a higher nodule volume variability was found in juxtapleural and juxtavascular nodules compared to well-circumscribed intraparenchymal nodules [68, 69]. Systematic differences in volume measurements between packages could influence nodule categorization and treatment decisions. High accuracy of segmentation is a requisite in software volumetric evaluation.

In the NELSON study, the same version of the same software package was used during all screening rounds, in all screening centers. Therefore, and because of the PVC condition, the discussed issues on software packages could not have influenced results of the NELSON study. In the near future, little as possible measured volume variability between different software packages for segmentation of nodule volume should be sought. For now, it is recommended to use the same software package for follow-up of pulmonary nodules in a single person, preferably the one with highest accuracy [67].

**Generalizability of results and clinical implications**

The NELSON study was performed in a population of older (50 - 75 years), current or former (quit maximum 10 years ago) heavy smokers (smoked > 15 cigarettes a day for >25 years or >10 cigarettes a day for > 30 years), without co-morbidity precluding curative therapy [70]. A lung cancer mortality rate of 3.4 per 1000 pack-years was estimated for this specific population. About 15-25% of the general (Dutch) population, aged 50 - 75 years, would be candidate for lung cancer screening, based on these inclusion criteria [70]. In the Netherlands, this currently corresponds to 783,538 to 1,305,897 persons [71]. In 2030, it is expected that 845,153 to 1,408,589 Dutch persons will meet the eligibility criteria.
(4.8% - 7.9% of total estimated population (17.7 million)) [72]. So, results of this thesis are generalizable to a sufficient part of the Dutch population.

Currently, management of incidentally detected lung nodules, outside the scope of a screening program, is based on mean diameter measurements [31, 32]. Based on studies from this thesis, it is likely that management of incidentally detected nodules can be improved by using a protocol based on semi-automated nodule volume measurements to estimate nodule size, and VDT for nodule growth determination. A prerequisite for volume-based nodule management is the availability of software for semi-automated volume measurements. These software packages are increasingly available in routine clinical practice.

**Clinical implications**

Since a general population differs from a high-risk screening population in terms of age and number of smoked pack years, it is likely that optimal volume cutoff values for incidentally detected nodules will differ from cutoffs used in a screening setting. The pre-test probability that an incidentally detected lung nodule is benign is, in a population with low-to-intermediate risk for lung cancer, higher than in persons who meet inclusion criteria for lung cancer screening. Although nodule management based on semi-automatically derived nodule volume and VDT is very promising, the value of volume and VDT for incidentally detected lung nodules as biomarkers for lung cancer detection in routine clinical care should be evaluated in future research.

In case lung cancer screening will become available in daily practice in the Netherlands, lung cancer stage at time of detection will shift from mostly lethal stage (stage III or IV) to most often early stage (stage I or II). This will result in a shift in treatment options for curative therapy, as smaller screen-detected lesions can be treated even less invasive than larger clinically detected early stage cancers. It is expected that more lung cancer patients, for instance those at high risk for surgical complications with stage I lung cancer, could be cured via stereotactic radiotherapy [73]. In fit patients with peripheral stage I lung cancer, surgery remains the preferred treatment. Minimal invasive thoracic surgery, like sublobar resection, may become a more important surgical option in lung cancer therapy, with less complications and lower post-operative mortality [74]. Further research on the optimal therapy of (very) early stage lung cancers should be performed, in order to allow patients as much as possible to benefit from the early lung cancer detection.
Conclusion and future perspectives

This thesis focused on the optimization of nodule management in CT lung cancer screening. It was found that nodule management could be improved in terms of:

- the use of semi-automated volume measurements instead of manually-measured diameter measurements to determine nodule size;
- the use of VDT to determine nodule growth;
- optimization of current volume and VDT cutoffs to distinguish between negative, indeterminate, and positive screen results;
- establishing separate, more stringent guidelines for incidence nodules;
- reading of screening examinations by experienced readers, and allow them to adjust the screen result in certain cases.

Lung cancer screening is a promising method to increase lung cancer survival, although the best measure for the prevention of future lung cancer cases obviously is smoking prevention. The publication of the positive results of the largest lung cancer screening study worldwide, the U.S. NLST, confirms the auspiciousness of lung cancer screening [1]. In the USA, lung cancer screening is already becoming more widespread available. However, the health-care system in the USA is very different to that in Europe, and up to today most European countries decided to wait for the final results of the NELSON study before implementing lung cancer screening by low-dose chest CT in their country. Several uncertainties remain within Europe to preclude formulation of a universal policy for screening, a.o. about 1) the optimal screen population, 2) the optimal nodule management protocol, 3) workup of suspected screen-detected nodules, 4) treatment of patients with screen-detected lung cancer, 5) the optimal screen interval and 6) cost-effectiveness of lung cancer screening. In the upcoming years, the future of lung cancer screening in clinical care in Europe will take shape.

In health care, development of models for determination of eligible patients for a certain treatment option or screening program, or for determination of the most suitable treatment option for a specific patient, becomes increasingly important. In a recent paper on radiotherapeutic oncology, it was demonstrated that medical doctors are unable to choose the most suitable treatment option for their patients if they base the treatment on their professional expertise. This form of decision making is regular medical practice, and not based on more sophisticated clinical decision algorithms [75]. Over the last decades, multiple studies in the scope of lung cancer risk prediction models have been performed. These studies indicate that use of high-quality risk prediction models to identify individuals eligible for screening will improve efficacy of lung cancer screening in terms of lower number of false-positives, saving additional lives, and increasing cost-effectiveness [76–79]. The UKLS was the first screening trial actually using a model to select participants eligible for lung cancer screening [25]. However, a great deal of further research on lung cancer risk prediction modeling remains to be done to optimize model use and take advantage of their potential [80]. Enhancement of these risk prediction models could be achieved by using improved modeling methods, and by incorporating additional lung cancer predictors such as molecular biomarkers [81] into the model.
Besides models to elucidate the optimal screen population, other models for the malignancy risk stratification of a nodule, detected in lung cancer screening have been established [74, 82, 83]. These models aim to accurately identify malignant lesions, and thereby to reduce false-positive screen results. Online applications to calculate malignancy risk in solitary pulmonary nodules are increasingly available. In the calculator published by McWilliams et al. [82], for example, the probability that a lung nodule in a specific patient will be diagnosed as cancer within a two to four year follow-up period is estimated, implementing both radiological and patient characteristics. The enormous amount of data gathered by the different lung cancer screening trials will help to further optimize and validate such models in big data source analysis. Such models combined with a robust radiological protocol, highly preferably based on semi-automated volume measurements, will further reduce the number of false positives and help with management decisions about indeterminate nodules [41].

In the NELSON study, low-dose CT scans have been used for lung cancer screening. In the past decade, multiple improvements in the field of CT scanners were established. Ultra low-dose CT scanners have been developed, and an increasing number of papers regarding the applicability of these ultra low-dose scans in the detection and evaluation of small pulmonary nodules is published [84, 85]. It is expected that in some years, lung cancer screening by ultra low-dose CT scanners will be performed, aiming to minimize the radiation dose as far as reasonably achievable, with doses far below 1 mSv (annual background radiation exposure 3.5 mSv in The Netherlands). However, future research on the feasibility of using this imaging technique for lung cancer screening has to be awaited.

A major stage shift at lung cancer detection (preferably stage I) is required to achieve a substantial reduction in lung cancer specific mortality. In order to accomplish this, the optimal length of the screen-interval between two subsequent screening examinations should be determined. The risk of stage shift in case of a longer screen interval, and radiation exposure and costs in case of a shorter screen interval should be balanced.

In case the NELSON reveals positive mortality results, cost-effective analyses should determine whether lung cancer screening in the Netherlands can be implemented. One should keep in mind that cancer screening of the lung is different from current screening programs of other malignancies, like breast cancer screening and colorectal cancer screening, in that respect that lung cancer screening only applies for heavy smokers. A general trend among Dutch people is the idea that persons who live unhealthy should pay for the consequences by themselves [86]. Lung cancer caused by smoking is "self-inflicted", which will be in relation to the cost aspects an important subject of the discussion in a national debate.

The final analysis of the difference in lung cancer mortality between the screen group and the control group of the NELSON trial will determine whether LDCT lung cancer screening in a high-risk population, using the volume-based nodule management protocol, led to the intended mortality reduction of at least 25% [70]. Whether or not CT lung cancer screening must be implemented in Europe should depend on these results, and on the results of ongoing sub studies on other aspects of lung cancer screening, e.g. cost-effectiveness [41]. Nevertheless, we need to prepare for lung cancer screening with an integrated smoking policy because this combined approach will save more lives than any other lung cancer intervention in the near future [41].
References


