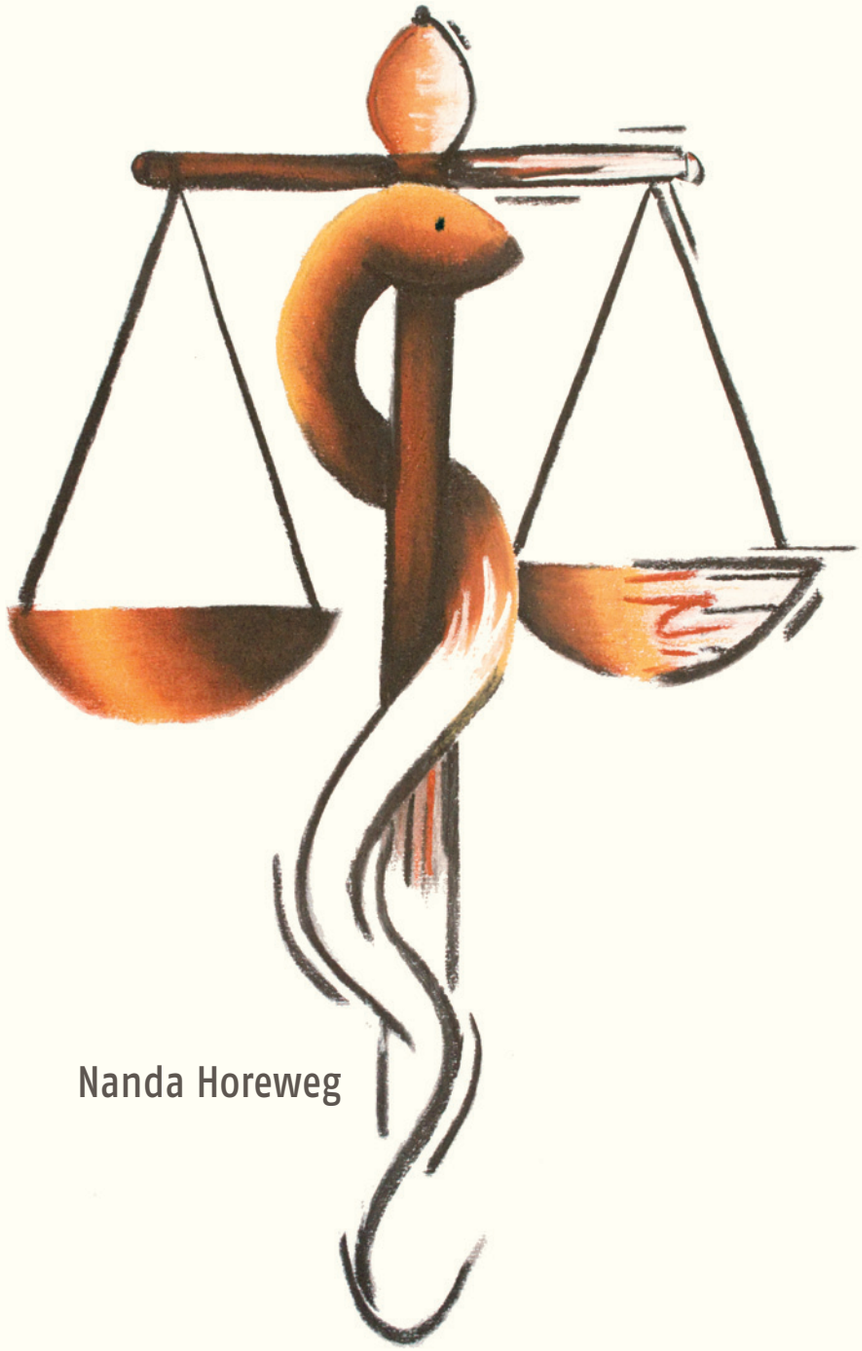
Lung cancer screening in the NELSON trial: balancing harms and benefits

Nanda Horeweg

Lung cancer screening
in the NELSON trial
balancing harms and benefits



Nanda Horeweg

Lung cancer screening in the NELSON trial

balancing harms and benefits

Nanda Horeweg

Lung cancer screening in the NELSON trial:
balancing harms and benefits

Thesis, Erasmus University Rotterdam, Netherlands

Cover illustration: Ivanca Horeweg
Lay-out and print: Optima Grafische Communicatie, Rotterdam, The Netherlands

The studies reported in this thesis were funded by The Netherlands Organisation of Health Research and Development (ZonMw), the Dutch Cancer Society (KWF), and the Health Insurance Innovation Foundation (Innovatiefonds Zorgverzekeraars), Health Insurance Innovation Foundation, Siemens Germany, Roche Diagnostics, G. Ph. Verhagen Stichting, Rotterdam Oncologic Thoracic Study (ROTS) group, Erasmus Trust Fund, Stichting tegen Kanker, Vlaamse Liga tegen Kanker, and LOGO Leuven.

This thesis was financially supported by the Department of Public Health, Erasmus MC, Rotterdam.

Lung cancer screening in the NELSON trial: balancing harms and benefits

Longkankerscreening in de NELSON studie: afweging van voor- en nadelen

PROEFSCHRIFT

Ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van rector magnificus

Prof.dr. H.A.P. Pols

en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 25 november 2014 om 9.30 uur

door

Nanda Horeweg
geboren te Spijkenisse

PROMOTIECOMMISSIE

Promotoren:	Prof.dr. H.J. de Koning	Erasmus MC Rotterdam, Nederland
	Prof.dr. H.C. Hoogsteden	Erasmus MC Rotterdam, Nederland
Leescommissie:	Prof.dr. M.G.M. Hunick	Erasmus MC Rotterdam, Nederland
	Prof.dr. P.E. Postmus	VUMC Amsterdam, Nederland
	Prof.dr. H. van Swieten	Radboud UMC Nijmegen, Nederland

TABLE OF CONTENTS

Part I: Introduction

1. General introduction 11

Part II: Evaluation of findings

2. Predictive value of screening test results 59
'Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial.'
European Respiratory Journal

3. Characteristics of screen-detected lung cancer 81
Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial.
American Journal of Respiratory and Critical Care Medicine

4. Epidemiological evaluation 105
'Detection of lung cancer through low-dose CT screening: analysis of screening test performance and interval cancers.'
Lancet Oncology

5. Radiological evaluation 127
'Computed tomographic characteristics of interval and post-screen carcinomas in lung cancer screening.'
European Radiology

Part III: Optimisation of screening

6. Optimisation of screening protocols 147
'Lung cancer probability in subjects with CT-detected pulmonary nodules: an analysis of data from the NELSON trial of low-dose CT screening.'
Lancet Oncology

7. Evaluation of bronchoscopy 179
'The role of conventional bronchoscopy in the work-up of suspicious CT screen-detected pulmonary nodules.'
Chest

8. Evaluation of surgical procedures 195
'Complications following lung surgery in the Dutch-Belgian randomized lung cancer screening trial.'
European Journal of Cardio-Thoracic Surgery

Part IV: Evaluation of effectiveness

9. Endpoint determination 221
'Uniform and blinded cause of death verification in a lung cancer CT screening trial.'
Lung Cancer

Part V: Implications for implementation

10. State of the art in lung cancer screening 243
'The importance of screening for lung cancer'
Expert Review in Respiratory Medicine

11. General discussion 263

12. Summary 317

13. Samenvatting 325

Part VI: Miscellaneous

14. Dankwoord 335

15. Curriculum vitae 343

16. PhD portfolio 347

17. List of publications 353

Part I

Introduction

Chapter 1

General introduction

In this thesis, the harms and benefits of lung cancer screening using low-dose computed tomography were investigated. Data of the Dutch-Belgian NELSON trial were used to quantify its harms and benefits and develop strategies to improve the balance between them. If the NELSON trial demonstrates that low-dose CT screening is an effective method to reduce mortality from lung cancer, balance between harms and benefits is a perquisite for the implementation of a lung cancer screening program. Background information on relevant aspects of epidemiology, medical ethics, pulmonary medicine, radiology, and pathology are essential for the interpretation of the studies in this thesis. In this chapter an overview of relevant background information is presented, as well as a description of the design of the NELSON trial.

AETIOLOGY

Lung cancer has been studied thoroughly in the past decades, which has given insight in its aetiology. The single most important cause of lung cancer is tobacco smoking.[1-3] Smokers have a 15-fold to 30-fold increased risk of developing lung cancer compared to non-smokers.[4] Other causative factors of lung cancer are: second-hand tobacco smoke exposure,[5,6] ionising radiation,[7] indoor and outdoor air pollution,[8] soot,[9] radon,[10-12] asbestos,[13,14] tar,[15] arsenic,[15] chromium[15] and nickel.[15]

Besides these causative factors, a number of risk indicators have been identified: older age,[13,16] family history of lung cancer,[16,17] acquired lung disease such as COPD,[18-21] HIV infection[22] and occupational exposures such as silica dust.[13] Physical activity and fruit and vegetable intake have consistently shown to be associated with a decreased risk of lung cancer.[3]

EPIDEMIOLOGY

Since tobacco smoking is the predominant causative agent of lung cancer, lung cancer incidence is strongly correlated with patterns of smoking prevalence.[8,23,24] The characteristic long latency period of smoking-induced lung cancer, which is the period from the start of smoking to lung cancer diagnosis, causes a delay in lung cancer incidence of 20 to 30 years.[8] In the United States, Australia, New Zealand and many countries in North-West Europe, lung cancer incidence rapidly increased from the 1930's onwards and peaked in the 1980's, and has been declining since.[8,23] In contrast, in Southern and Eastern European countries, China, and Japan lung cancer incidence still increases or is stable.[8] Moreover, the lung cancer incidence is predicted to increase substantially throughout Asia and Africa in the future, due to the uptake of western smoking habits.[8] Variations in

lung cancer incidence across countries or between males and females are largely reflected in the differences in the stage and degree of the tobacco epidemic.[25]

PUBLIC HEALTH

Currently, lung cancer is the second most common cancer; accounting for 14% of all cancer cases in the U.S. in both men and women.[23] Moreover, lung cancer causes most cancer-related deaths; 28% of the cancer related deaths in men and 26% in women.[23] Lung cancer causes more deaths than prostate cancer, breast cancer, colon cancer and pancreatic cancer combined, which makes lung cancer a major public health problem.[9,23]

DISEASE CHARACTERISTICS

Lung cancer is such a major public health problem because of its high incidence and high case-fatality. The latter is partly caused by the fact that lung cancer often causes no symptoms at early stages of disease. As result, lung cancer is commonly diagnosed at stages wherein disease has advanced to regional (22%) or distant (56%) spread.[23] Hence, only a minority is diagnosed with localised lung cancer, wherein surgical resection of the entire tumour is still feasible.[10] In this group, the chance to be alive five years after diagnosis is 52%.[23] Which is substantially higher than the five-year survival of regionally and distantly metastasised disease; respectively 25% and 4%.[23] At these more advanced stages, surgical resection of the primary tumour is often not curative, and therapy is often only aimed at improving survival and quality of life.[11,12]

CLINICAL CARE

The advances in treatment of lung cancer have been substantial over the past decades and have improved survival of lung cancer patients. For example: several new chemotherapy regimens have been developed, some specifically directed at histological subtype,[26] and the increased use of chemotherapy as adjuvant therapy.[27,28] More recently, targeted therapies at somatic mutations in receptors or signal proteins have become available.[26] Further, advances in radiotherapy, such as stereotactic body radiotherapy, contributed to improved survival.[29,30] The combination of chemotherapy and radiation therapy has evolved from sequential to concomitant, which further improved overall survival.[31-33] Pre-operative patient selection has improved as result of the use of validated comorbidity indices, multidisciplinary decision-making, and more accurate staging.[34-36] Finally, bet-

ter adoption to standard care treatment guidelines, and a greater proportion of patients receiving any treatment, contributed to the survival of lung cancer patients.[37]

Clearly, many improvements in the treatment of lung cancer have been made, but only modest improvement in the survival of lung cancer patients could be observed over the last decades.8 In the United States, the overall five-year relative survival of lung cancer patients has improved from 12% in 1975-1977 to 17% in 2002-2008.[23] The overall five-year survival of lung cancer patients in Canada improved from 15.7% in 1995-1999 to 18.4% in 2005-2007.[38] In Australia, an improvement from 13.9% in 1995-1999 to 17.0% in 2005-2007 was observed.[38] The overall five-year survival in Europe improved from 9% to 11%, on average.[39] In the United Kingdom, overall survival was substantially lower, 7.0% in 1995-1999, as also the improvement in survival; 1.8% to 8.8% in 2005-2007.[38] In North-West Europe, where high-quality registries with national coverage are available, similar small improvements in survival were observed; in the periods from 1995-1999 and 2005-2009 respectively: from 8.0% to 10.9% in Denmark, 11.0% to 14.4% in Norway, 12.7% to 16.3% in Sweden. In the Netherlands, the overall five-year survival increased from 14.8% in 1989-1993 to 17.4% in 2009.[31]

PREVENTION

Improvements in the treatment of lung cancer are continued to be made, and will undoubtedly contribute to an improved survival of lung cancer patients in the future. However, the fact that lung cancer is mostly diagnosed at an incurable, advanced stage limits treatment options to improving survival and reducing morbidity. In contrary, prevention may be able to reduce the burden of lung cancer in a different way. As presented in Box 1, three forms of prevention can be distinguished: primary, secondary and tertiary prevention.

Text box 1. Definitions prevention

Prevention and clinical care are the main methods in medicine to improve health. Prevention can be sub-classified in primary, secondary and tertiary prevention.

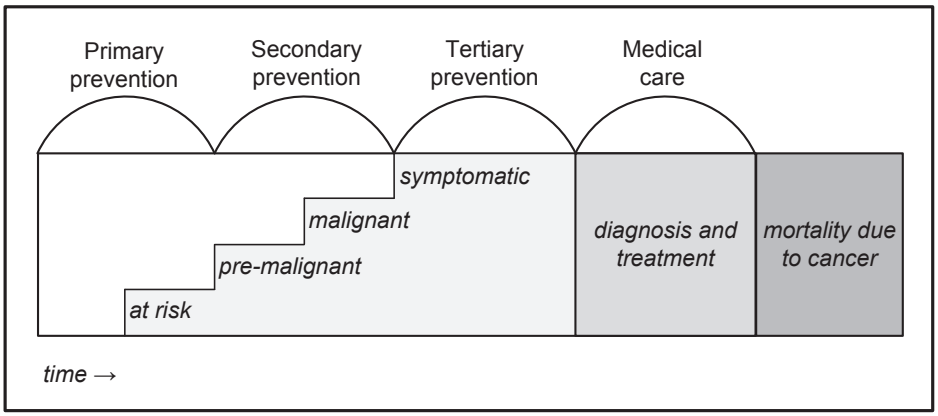
- Primary prevention aims to prevent the occurrence of disease by elimination or reduction of the causes of disease.
- Secondary prevention aims to prevent progression of disease by detecting and treating disease at an early stage.
- Tertiary prevention aims to prevent or limit the unfavorable outcomes of diseases that are already diagnosed.

If these definitions are applied to prevention of lung cancer, the stage of disease plays an import role in the form of prevention applied, which is depicted in Figure 1.

Hence, morbidity and mortality from lung cancer may be reduced by:

I) Primary prevention through reducing the occurrence of lung cancer;
II) Secondary prevention through early detection by screening asymptomatic high-risk subjects;
III) Tertiary prevention through earlier treatment by increasing awareness of the signs and symptoms of lung cancer in the general population.

Figure 1. Prevention and medical care according to cancer stage



In time, cancer develops in an individual at risk, from pre-malignant, to malignant, to symptomatic cancer, and after some delay diagnosis is made and the individual receives medical care until death. Primary, secondary and tertiary prevention apply to different stages of disease, but could be offered to any individual receiving medical care

Primary prevention

Primary prevention may improve public health by reducing mortality and morbidity from lung cancer in two different ways. On the one hand, the occurrence of lung cancer can be reduced by protection from the carcinogenic agents that specifically cause lung cancer. On the other hand, the occurrence of lung cancer can be counteracted by the use of specific agents to reverse, suppress, or prevent the process of carcinogenesis.

The latter is called chemoprevention, and many substances, such as aspirin,[40-42] β-carotene,[43-45] retinyl palmitate,[45-47] 13-cis-retinoic acid,[46,47] vitamin E,[43] N-acetylcysteine[47] and selenium,[48,49] have been tested in clinical trials.[50] None of these trials demonstrated any beneficial effect, while some did show harmful effects.[50] Therefore, to date not one agent is recommended for use in the chemoprevention of lung cancer.[50]

Hence, the reduction of exposure of the general population to the causative agents of lung cancer may be a safer and more effective approach to improve public health. Cur-

rent knowledge on the aetiology of lung cancer may be used to develop such primary prevention interventions. Since tobacco is responsible for 80-90% of the lung cancer diagnoses,[3,51] both through smoking[1-3] and second-hand smoke exposure,[5] most benefit can be expected from interventions directed at prevention of the initiation of smoking and smoking cessation. The adverse health effects of tobacco smoking became widely apparent in the 1950's,[52] and many interventions have been implemented since: anti-smoking campaigns, marketing and sales restrictions, federal cigarettes taxes, smoke-free air laws, smoking cessation treatments.[53,54] These interventions had substantial impact on smoking prevalence, which was reflected in lung cancer incidence and mortality twenty to thirty years later.[23,54] Millions of premature deaths were prevented by tobacco control interventions; a substantial proportion through prevention of lung cancer deaths.[55,56] Despite this success of primary prevention, global smoking prevalence was still as high as 23.7% in 2010.[57] Moreover, it has been estimated that smoking prevalence will only decrease to 22.0% in 2030 if no additional tobacco control policies are applied.[57]

Concluding, primary prevention is inevitable in the fight against lung cancer, and continuous efforts should be made to force back exposure to its causative agents, tobacco smoking in particular. However, primary prevention solely is not expected to be able to reverse the lung cancer epidemic and reduce morbidity and mortality substantially in the next decades.

Tertiary prevention

The aim of tertiary prevention is to improve survival and reduce mortality by early treatment in symptomatic lung cancer patients. To be able to treat lung cancer as early as possible, delays between the onset of symptoms and treatment should be minimised. Three types of delay are recognised in the literature:[58,59]

I) Patient-related delay due to failure to act immediately on suspicious symptoms through fear or lack of knowledge.
II) Doctor-related delay due to misinterpreting symptoms or not referring for diagnostic testing.
III) System-generated delay due to inefficiency or long waiting times of appointments or tests.

Efforts to reduce delay type II and III are embedded in clinical guidelines and performance indicators of health care.[35,60] Delay type I has been recognised as the most important source of delay between onset of symptoms and start of treatment.[61]

Several studies on patient-related delay in seeking a cancer diagnosis have been published, but identified different sets of determinants. Corner et al identified comorbidity, misinterpretation of symptoms, lack of knowledge, and difficulties of recognising ill health in elderly as determinants.[58] While Leydon et al. found that a person's experiences, expectations from health care, family decisions and fear of cancer were important.[62] In

the study of Ristvedt et al. predisposition to seek help and certain personality traits were identified as determinants of delay.[63]

Further, the spontaneous awareness of the symptoms of lung cancer is limited; nearly a quarter of the general population cannot mention any symptoms of lung cancer, and those who can mention breathlessness and coughing.[64] This information is essential for developing tertiary prevention interventions that reduce patient-related delay, which may address to the poorer survival associated with delay.[65] Henceforth, several initiatives to raise awareness of symptoms of lung cancer and to de-stigmatise the disease have been implemented.[66-68] The effectiveness of tertiary prevention has been investigated in a limited number of studies; effects on self-reported awareness,[69-72] intention to seek care,[70,72] health care policy,[69,73] referral rates,[70] disease incidence[70] have been reported. A favourable effect on disease stage at diagnosis was not consistently proven,[70,72] moreover none of the studies evaluated the effect on survival or lung cancer mortality.

Despite the fact that the effectiveness of the aforementioned interventions on lung cancer morbidity and mortality has not been demonstrated, and the disease itself is often asymptomatic in early stages, tertiary prevention should not be disregarded. The observation that clinically-diagnosed lung cancer has often already progressed to an advanced stage at diagnosis[23] might be not exclusively caused by the biology of the disease; a part of this problem might also result from the social context of the disease.[58,74] The general public has low expectations from health care, because lung cancer is considered as an inevitably fatal condition.[58,74] Earlier diagnosis in symptomatic patients is scarcely promoted because there is little expected gain.[58,74] In addition, patient advocacy movements are disabled by the blame of self-infliction of disease and the relatively small proportion of patients that survive the disease.[58,74] The power of tertiary prevention is best demonstrated in breast cancer: by creating awareness, the general public has become educated on the symptoms of the disease and on the benefits of seeking an early diagnosis, and a powerful social movement has arisen.[75]

Concluding, tertiary prevention has not proven to be able to reduce lung cancer morbidity and mortality. Nonetheless, its effects may reach further than just earlier treatment of symptomatic patients; it may influence the public opinion and professional agendas, which contributes to the development and funding of research, screening, clinical care and aftercare.

Secondary prevention

Lung cancer screening is a form of secondary prevention (Box 1), and aims to reduce mortality by cancer detection at an early and curable stage. As this early stage is often not accompanied by any signs or symptoms, screening is applied to apparently healthy, asymptomatic persons.

Harms and benefits

The benefit asymptomatic high-risk subjects have from an effective screening program is: a reduced probability of dying from lung cancer, and a reduced probability to suffer from advanced disease (Table 1).[76] Unfortunately, the subjects who undergo screening are also exposed to several harms. The harms can be related to the screening test itself; as for example radiation-induced cancer or psychological distress awaiting the test result. But harms can also be related to false positive screenings (e.g. complications of subsequent diagnostic tests) and false negative screenings (e.g. delayed diagnosed due to false reassurance). Further, overdiagnosis is considered to be an important harm of screening. A detailed overview of potential harms and benefits of screening is provided in Table 1.

Medical ethics

The harms and benefits listed in Table 1 represent one of the contradictions in screening. On the one hand, screening aims to improve (public) health by reducing morbidity and mortality from lung cancer, on the other hand, screening unintentionally exposes the screened population to a variety of harms. To be able to perform and interpret research in the field of screening, knowledge on the ethical principles is essential.

Table 1. Benefits and harms of cancer screening

Benefits
Less persons dying from lung cancer
Less persons suffering from advanced lung cancer
Less persons receiving intensive or mutilating primary treatment
Possible positive effects on smoking cessation
Harms
Undergoing screening test and awaiting result - psychological distress
Radiation-induced cancers - morbidity and mortality
False positive results - psychological distress, morbidity and mortality due to subsequent diagnostic procedures
False negative results - false reassurance, delayed diagnosis once symptoms occur
Overdiagnosis - psychological distress, morbidity and mortality due to overtreatment
Persons receiving the diagnosis of lung cancer earlier
Possible negative effects on smoking cessation

Ethical principles

The following four ethical principles are considered most relevant for lung cancer screening:

I) Beneficence: this principle signifies that physicians must help their patients and act in their patients' best interest.

II) Non-maleficence: this principle signifies that physicians must not harm their patients.
III) Autonomy: this principle signifies that physicians must respect the right of patients to decide over their own medical interventions and treatments.
IV) Justice: this principle signifies that physicians must treat equal patients equally and must consider fair distribution of health care resources.

As these ethical principles are part of the medical oath, physicians involved in screening have a number of responsibilities. According to the principle of beneficence physicians should propose lung cancer screening to those individuals in whom it is beneficial for their health. According to the principle of non-maleficence, physicians should not offer lung cancer screening to those individuals in whom it is not beneficial. Moreover, this principle also implies that physicians have the responsibility to minimise the harms of screening in whom screening is considered beneficial. According to the principle of autonomy, physicians should respect a person's decision to undergo lung cancer screening or not. As informed decision-making is a prerequisite for participation in screening, physicians also have the responsibility to inform screening candidates on benefits and harms of screening. According to the principle of justice, physicians have the responsibility to treat individuals, who are equal with respect to aspects relevant for screening, equally. This, for example, refers to providing care that is accessible and appropriate for the entire target population, or to the fair distribution of limited health care resources.

Ethical dilemmas

The responsibilities that result from the ethical principles can be conflicting. As mentioned previously, the most prominent ethical dilemma in lung cancer screening is the conflict between beneficence and non-maleficence. The harms a screening program induces should be weighed against the benefits the program yields. Obviously, it is not ethical to implement a screening program that causes more harm than benefit. In lung cancer screening, harms and benefits are not the same for every individual but depends i.e. on age, smoking history and co-morbidity. Therefore, the ethical dilemma between beneficence and non-maleficence plays an important role in defining the target population for lung cancer screening.

The principle of autonomy can conflict with the principles of beneficence and non-maleficence. Well-informed individuals have the right to decide for themselves whether or not to undergo screening. However, this right can conflict with the principles of beneficence and non-maleficence when the benefits of screening do not outweigh the harms. For example, in case of an individual with a negligible risk of lung cancer who demands to undergo LDCT screening. The principle of autonomy could also conflict with the ethical principle of justice. The right of the individual to undergo screening can

conflict with the responsibility to distribute limited health care resources responsibly to preserve the accessibility of health care.

The significance of the considerations surrounding the ethical principles and dilemmas has been recognised for decades. As a result, screening criteria that encounter these ethical issues have been developed to guide decisions on the implementation of screening programs.

Criteria for screening

In 1968, the World Health Organisation (WHO) commissioned a report from Wilson and Jungner on the "Principles and practice of screening for disease".[77] This report contains the 'Wilson and Jungner criteria' for screening (overview provided in Box 2), which have been regarded as the golden standard in decision-making for a long time.[77]

Although the value of the Wilson and Jungner criteria is still widely recognised,[78] many have suggested adaptions and improvements of the criteria.[79-84] In 2008, the WHO published 'Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past years'.[78] In this article, a new set of criteria was presented, based on the suggested improvement of the Wilson and Jungner criteria proposed over the past forty years (Box 3).[78]

Box 2. Criteria for screening by Wilson and Jungner, 1968

I)	The condition sought should be an important health problem.
II)	There should be an accepted treatment for patients with recognized disease.
III)	Facilities for diagnosis and treatment should be available.
IV)	There should be a recognisable latent or early symptomatic stage.
V)	There should be a suitable test or examination.
VI)	The test should be acceptable to the population.
VII)	The natural history of the condition, including development from latent to declared disease, should be adequately understood.
VIII)	There should be an agreed policy on whom to treat as patients.
IX)	The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
X)	Case-finding should be a continuing process and not a "once and for all" project.

Screening for lung cancer

Lung cancer fulfils a number of the criteria for mass screening (Box 3).[78] Criterion I: as the burden of lung cancer is high, an effective screening program responds to a recognized need.[78] Criterion II: the objective of a lung cancer screening program would be to reduce morbidity and mortality from lung cancer.[78] Current knowledge on the aetiology of lung cancer provides the opportunity to define specific target populations for screening, which is a prerequisite to fulfil criterion III.[78] Criteria IV to X (Box 3) do not relate to characteristics of the disease itself, but to the screening program's effectiveness, balance between harms and benefits, and associated costs.[78]

Box 3. Modern screening criteria proposed by the World Health Organisation

I)	The screening programme should respond to a recognised need.
II)	The objectives of screening should be defined at the outset.
III)	There should be a defined target population.
IV)	There should be scientific evidence of screening programme effectiveness.
V)	The programme should integrate education, testing, clinical services and programme management.
VI)	There should be quality assurance, with mechanisms to minimise potential risks of screening.
VII)	The programme should ensure informed choice, confidentiality and respect for autonomy.
VIII)	The programme should promote equity and access to screening for the entire target population.
IX)	Programme evaluation should be planned from the outset.
X)	The overall benefits of screening should outweigh the harm.

Lung cancer screening: cohort studies

Until the 1990s, there has been little role for lung cancer screening because no effective screening test was available.[85] Screening studies using sputum cytology,[86-88] or chest radiography did not show a significant lung cancer mortality reduction.[89] In the 1990s, trials using low-dose computed tomography (LDCT) as a screening test were initiated,[85] and results were encouraging. LDCT appeared to be able to detect more and smaller lung cancers than chest radiography;[90-92] 61% to 93% of lung cancers were diagnosed at stage I.[90-95] Moreover, survival rates in patients with screen-detected lung cancer were startling: five-year survival in the Japanese ALCA trial[95] was 64.9% to 76.2%, and ten-year survival in the U.S. ELCAP trial was even 80-92%.[91,96]

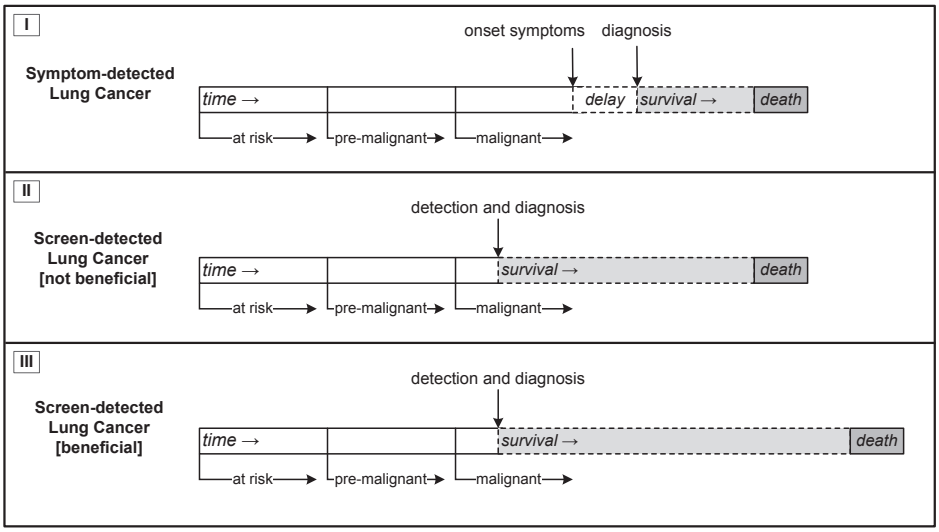
Bias

The survival of patients with screen-detected lung cancer is not the right endpoint to evaluate the effectiveness of a screening program due to three forms of bias:

I) Lead-time bias: by screening asymptomatic individuals, the diagnosis of lung cancer is established earlier than it would have been without screening, which is usually after the onset of symptoms. As survival analyses take the moment of diagnosis as starting point, survival of patients with screen-detected lung cancer will be longer than the survival of patients with clinically diagnosed lung cancer, even when there is no benefit of screening (Figure 2).

II) Length-time bias: as the aforementioned cohort studies analysed survival of patients with screen-detected lung cancer only, this form of bias also plays a role. Lung cancer is a very heterogenic disease, and some subtypes of lung cancer grow slower than other subtypes. The slow-growing cancers have a longer asymptomatic phase than the fast growing cancers. As a result, the likelihood of a slow-growing cancer to be detected by screening is higher than the likelihood of a fast-growing cancer. Hence, screen-detected lung cancers grow slower on average than lung

Figure 2. Schematic depiction of lead-time bias



In panel I, calculation of survival is depicted for symptom-detected lung cancer; starting point is the moment of diagnosis and endpoint is the moment of death. In panel II and III, calculation of survival is depicted for screen-detected lung cancer. In both, survival is substantially longer than for symptom-detected lung cancer, as a result of advancing the moment of diagnosis through detection before the onset of symptoms. However, only in panel III survival is truly prolonged by screening. In panel II, the moment of diagnosis is advanced but the moment of death is at the same moment as when lung cancer was diagnosed through symptoms.

cancers not detected by screening. Since slow-growing cancers are associated with longer survival and lower case-fatality, the survival of screen-detected cancers is better than the survival of lung cancers not detected by screening. As the survival analyses of the cohort studies only included the screen-detected lung cancers, the result is biased. Therefore, the survival of both screen-detected lung cancers and the lung cancers not detected by screening should be included in the analysis. The latter requires the availability of high-quality cancer registries or thorough follow-up of study participants.

III) Overdiagnosis: is inseparable from screening and means the detection of cancers which would have never had led to symptoms or death (Figure 3). At the time the ALCA and ELCAP study were conducted, almost all individuals who were diagnosed with screen-detected lung cancer also underwent surgery. As a result, it is not possible to determine how many individuals had a lung cancer that would not have been fatal if left untreated; the overdiagnosed cancers. Subsequently, it's not possible to determine to what extend overall survival is biased by overdiagnosis.

Figure 3. Schematic depiction of overdiagnosis

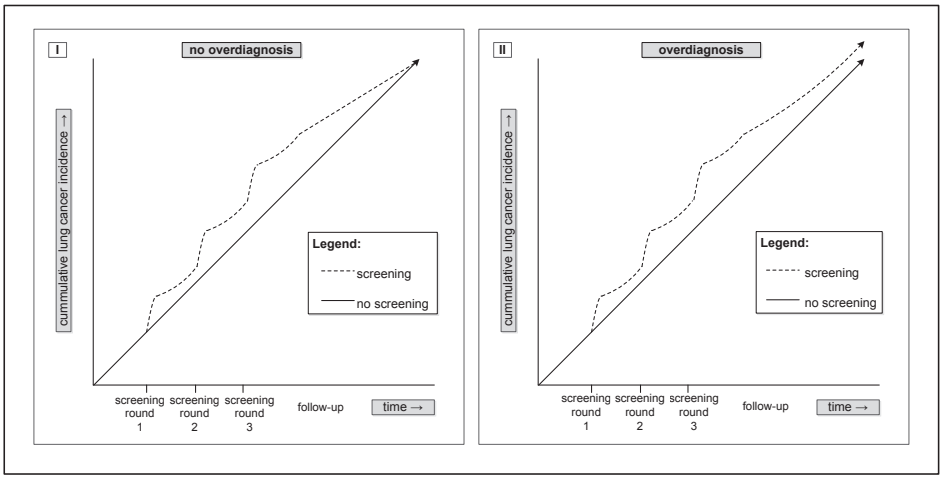


Figure 3 describes a population with a constant lung cancer incidence, that underwent three low-dose computed tomography screening rounds. In panel I, an hypothetical lung cancer screening program that did not lead to any overdiagnosis is depicted. Through earlier detection of lung cancer the cumulative incidence of lung cancer increases faster during screening compared to no screening. However, after screening has stopped and the 'wash-out period' of screening has passed, the cumulative lung cancer incidence is as high as in the situation without screening. Hence, the is no overdiagnosis. Note: in practice, due to concurring mortality, there will always be overdiagnosis. In contrary to panel II, where in the cumulative lung cancer incidence remains higher than in the situation without screening. The difference in lung cancer cases between the two lines are the overdiagnosed lung cancers.

To avoid these three forms of bias, the primary endpoint of screening studies should be disease-specific mortality reduction. Ideally, the effect of screening on lung cancer mortality is determined in a trial wherein participants are randomised between screening and no screening. Next, both groups should be followed up simultaneous for a sufficiently long period of time after screening has stopped. Just one of the aforementioned cohort studies assessed lung cancer mortality reduction; the Mayo Lung Project compared lung cancer mortality between the screened cohort and a historical cohort; analysis showed no significant lung cancer mortality reduction.[94]

Lung cancer screening: randomised trials

The encouraging results of the LDCT cohort studies led to the initiation of several randomised controlled trials (Table 2).[97-104] Although design of the trials varies notably, the primary endpoint of all these trials was lung cancer mortality. Four of the seven randomised trials have currently reported their results. The largest trial, the U.S. National Lung Screening Trial (NLST), reported a statistically significant lung cancer mortality reduction of 20.0% (95% CI 6.8-26.7%) after 6.5 years of follow-up.[76] In the NLST, screen-

Table 2. Characteristics of randomised controlled trials on LDCT screening for lung cancer

Trial	Participants	Initiation	Design	Screenings	Characteristics participants			
	N	Year		N	Sex	Age*	Smoking	Cessation
NLST[97,138]	53,439	2002	LDCT vs. CXR	3	M/F	55-74	≥30 py	<15 yrs
NELSON[108,139]	15,822	2004	LDCT vs. no screening	4	M/F	50-75	≥15/day for 25 yrs or ≥10/day for 30 yrs	≤10 yrs
DLST[99]	4,104	2004	LDCT vs. no screening	5	M/F	50-70	≥20 py	<10 yrs
MILD[100]	4,099	2005	LDCT vs. no screening	5/10	M/F	≥49	≥20 py	<10 yrs
LUSI[101]	4,052	2007	LDCT vs. no screening	4	M/F	50-70	≥15/day for 25 yrs or ≥10/day for 30 yrs	≤10 yrs
UKLS[102,140]	4,000	2011	LDCT vs. no screening	1	M/F	50-75	≥5% risk of lung cancer in 5 yrs	
ITALUNG[103]	3,206	2004	LDCT vs. no screening	4	M/F	55-70	≥20 py	<10 yrs
DANTE[104]	2,472	2001	Initial CXR, followed by LDCT vs. no screening	4	M	60-75	≥20 py	<10 yrs

Definition of abbreviations: LDCT = low-dose computed tomography; CXR = chest x-ray; M = male; F = female; py = pack-years; yrs = years.

** Age range up to, but not including upper limit.*

ing using LDCT was compared to screening using chest radiography[2] which does not affect lung cancer mortality.[89] Moreover, screening using LDCT reduced significantly all-cause mortality with 6.7% (95% CI 1.2-13.6%).[76] When lung cancer mortality was not included in all-cause mortality analysis, the all-cause mortality reduction dropped to 3.2%, and was not statistically significant anymore.[76]

Three smaller trials in Europe, the Danish screening trial, and the Italian DANTE and ITALUNG trials, reported no significant lung cancer or all-cause mortality reduction.[100,105,106] An overview of the outcomes of the trials is presented in Table 3.[107] Pooled estimates of the relative risks of death of the four trials combined were not published. Possibly because the estimates were partly based on interim analyses or absolute number life-years were not provided for the Italian studies, or differences in design of the included trials. Our calculation of the pooled relative risk, based on the published data, suggested that LDCT screening has significantly reduced the risk lung cancer mortality.[53]

Table 3. Effect of LDCT screening on lung cancer and all-cause mortality

Trial*	Quality	Lung cancer deaths			All deaths		
		per 100,000 py[107]		Relative risk[107]	per 100,000 py[107]		Relative risk[107]
		Intervention	Control	RR (95%CI)	Intervention	Control	RR (95%CI)
NLST76	Good	247	309	0.80 (0.73-0.93)	1142	1216	0.93 (0.86-0.99)
DLST105	Fair†	154	112	1.37 (0.63-2.97)	625	429	1.46 (0.99-2.15)
MILD100	Poor‡	216	109	1.99 (0.80-4.96)	558	310	1.80 (1.03-3.13)
DANTE106	Fair¶	527	637	0.83 (0.45-1.54)	1212	1433	0.85 (0.56-1.27)

Definition of abbreviations: py = person-years; RR = relative risk; 95%CI = 95% confidence interval.

* *Trials included with results published before January 2014.*

† *Unclear allocation, differential follow-up.*

‡ *Inadequate randomization, differences in baseline demographic characteristics, differential follow-up.*

¶ *Unclear allocation, differences in baseline demographic characteristics, differential follow-up.*

Concluding, efficacy of LDCT screening for lung cancer has been demonstrated by the NLST.[76] However, the high survival rates in earlier cohort studies created high expectations from LDCT screening,[91,95,96] as a result, the 20% mortality reduction might not be as high as hoped. Nonetheless, there have been no other interventions so far, besides primary prevention, that have proven to be as successful as LDCT screening in reducing lung cancer mortality.

Currently, the Dutch-Belgian (NELSON), German and British lung cancer screening trial are still ongoing.[101,102,108] As soon as enough follow-up time has accrued and data becomes available, final mortality analyses are expected from these studies. Updated pooled analyses also including these trials will provide a definitive conclusion on the effectiveness of LDCT screening.

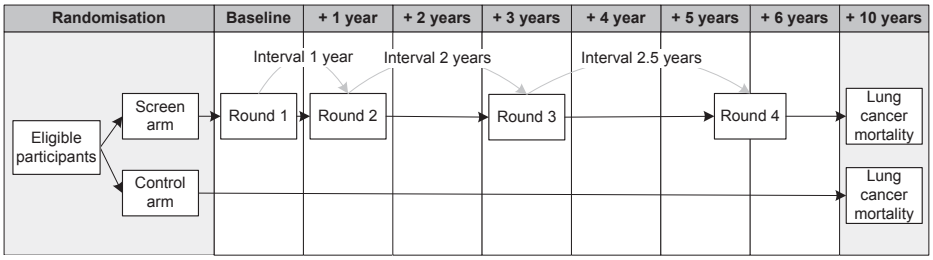
THE NELSON TRIAL

Design

The NELSON trial is a randomised controlled trial on the efficacy of screening using low-dose computed tomography. The trial was conducted in the Netherlands and Belgium. NELSON is the acronym for the Dutch name of the trial: NEderlands-Leuvens longkanker ScreeningsONderzoek.

The NELSON trial was initiated in 2003, after favourable survival of lung cancer patients, who underwent low-dose computed tomography screening, was demonstrated in cohort studies.[95,96] To determine whether LDCT screening yields not only improved survival but also reduced mortality from lung cancer, screening was compared to no screening. As presented in figure 4, study participants were randomised to no screening, or to four rounds of screening using LDCT; at baseline (first screening round), one year later (second screening round), three years later (third screening round), and five and a half years later (fourth screening round). Both groups of participants are followed up, and the difference in lung cancer mortality between the two groups is determined ten years after randomisation.

Figure 4. Design of the NELSON trial



Endpoints

Primary research objectives of the NELSON trial are:

I) To determine whether LDCT screening yields a reduction of ≥25% in lung cancer mortality.

II) To estimate cost-effectiveness of LDCT screening for lung cancer.

III) Secondary research objectives of the NELSON trial are:

IV) To determine whether LDCT screening yields a reduction in all-cause mortality.

V) To determine the effect of LDCT screening on quality of life.

VI) To determine lung cancer incidence and five-year survival rates.

VII) To determine detection rates and stage distribution per screening round.

VIII) To determine the number, stage distribution and time interval since last screening of interval cancers, and the ratio of the screen-detected cancers and the interval cancers.
IX) To determine the screening algorithm's sensitivity, specificity and positive predictive value.
X) To further define best practices in lung imaging and quality assurance.
XI) To define the molecular dynamics of very early lung cancer.
XII) To further define best practice and quality assurance in nodule evaluation and early stage lung cancer management.

Hypothesis

Screening using low-dose computed tomography will yield a lung cancer mortality reduction of ≥25% at ten years of follow-up.

Recruitment

The method of recruitment in the NELSON trial was especially designed to maximise the validity of extrapolation of trial results to the population eligible for lung cancer screening. Recruitment strategies based on media advertisements are known to attract health-concerned individuals, who are eager to participate in health and life style interventions. As the population at high risk for developing lung cancer does not typically consist of health-concerned individuals, such an approach should be avoided. To minimize this so-called 'self-selection bias', a population-based recruitment strategy was chosen for the NELSON trial.

Potential trial participants were identified via population registries and were approached by mail. From the second half of 2003 onwards, more than a half million questionnaires on general health, smoking, alcohol consumption, physical exercise, cancer history, family history of lung cancer, body weight and length, education and opinion on screening programs, were sent to all men and women born between 1928 and 1953 in 7 districts in the Netherlands and 14 municipalities around Leuven in Belgium.[98] This questionnaire was neither accompanied by information on the minimal requirements for participation, such as smoking history, nor by any other information about the trial, to prevent prejudiced answers.

The information obtained with this questionnaire was used to decide whom to invite for the trial. First, the estimated lung cancer mortality risk of the respondents was estimated using data of the US Cancer Prevention Studies.[109,110] Next, the required sample size and the corresponding number of eligible subjects was determined using the same formulas as in the American PLCO (Prostate, Lung, Colorectal and Ovarian) screening trial and the European Randomised Screening Trial on Prostate Cancer.[111,112] For this calculation, a 1:1 randomisation, a power of 80%, a one-sided a significance level of 0.05, 95%

compliance in the screen group, 5% contamination in the control group and 10 years of follow-up after randomisation were assumed. Finally, the required participation rate was determined. The most optimal selection scenario, which required a participation rate as low as possible and a required sample size within the ranges of the capacity, was to invite the following population: 50 to 75-year old current or former smokers who had quit less than 10 years ago with a smoking history of at least 15 cigarettes per day for 25 years or at least 10 cigarettes for 30 years.[98] Hence, to be able to demonstrate a lung cancer mortality reduction of at least 25% in this study population, the estimated required sample size was 17,300 subjects.[98] Therefore, a possible pooling with the Danish lung cancer screening trial was proposed.

In the second phase of recruitment another questionnaire was send, only to the eligible responders, which enclosed questions on smoking habits, smoking cessation, asbestosis exposure and chronic obstructive pulmonary disease, and the trial's information leaflet and the informed consent. However, subjects with: a moderate or bad self-reported health, the inability to climb 2 flights of stairs, a body weight of 140 kg or more, current or past renal cancer, melanoma or breast cancer, or lung cancer diagnosed less than 5 years ago, or a chest CT examination less than 1 year ago, were excluded.

Eligible subjects without any exclusion criteria, who responded to the second questionnaire and provided written informed consent for participation in the NELSON trial were included and randomised. An overview of inclusion and exclusion criteria is provided in Table 4.

Table 4. Inclusion and exclusion criteria of the NELSON trial

Inclusion criteria	
Age	50 - 75 years
Smoking history	≥ 15 cigarettes per day for 25 years ≥ 10 cigarettes per day for 30 years
Smoking cessation	≤ 10 years ago
Exclusion criteria	
Self-reported health	moderate or bad
Ability to climb stairs	≤ 2 flights
Body weight	≥ 140 kg
History of lung cancer	still under treatment diagnosed < 5 years ago
History of other cancer	renal cancer breast cancer melanoma
History of imaging	Computed tomography of the chest < 1 year ago

Equipment and execution of screening examinations

The participants randomised to the screening group were invited by mail to undergo a LDCT examination of the chest at the nearest of the four screening sites. These were in the Netherlands in University Medical Center Groningen, University Medical Center Utrecht, and Kennemer Gasthuis in Haarlem, and in Belgium in University Hospital Leuven.

The CT scans used were all 16-detector MSCT scanners (M×8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA, or Sensation-16, Siemens Medical Solutions, Forchheim, Germany).[113] All scans were realised in about 12 seconds in spiral mode with 16 mm × 0.75 mm collimation and 15 mm table feed per rotation (pitch = 1.5), in a cranial-caudal scan direction, without intravenous contrast in low-dose setting.[113] Depending on the body weight (less than 50 kg, 50 to 80 kg and more than 80 kg) the kVp settings were respectively 80-90 kVp, 120 kVp and 140 kVp.[113] This corresponds with an effective radiation dose of less than 1.6 mSv.[114] To achieve a CTDIvol of respectively 0.8mGy, 1.6mGy and 3.2 mGy, the mAs settings were adjusted for the machine used.[113] Datasets of the thorax were reconstructed at 1.0 mm slice thickness, with 0.7 mm reconstruction increment and soft kernel (Siemens B30 filter, Siemens Medical Solutions, Forchheim, Germany).[114,115] To minimise breathing artefacts, scans were performed in inspiration after appropriate instruction of the participants. Data acquisition and scanning conditions were kept standard across the four screening centres for the duration of the trial.

Image reading and volumetric measurements

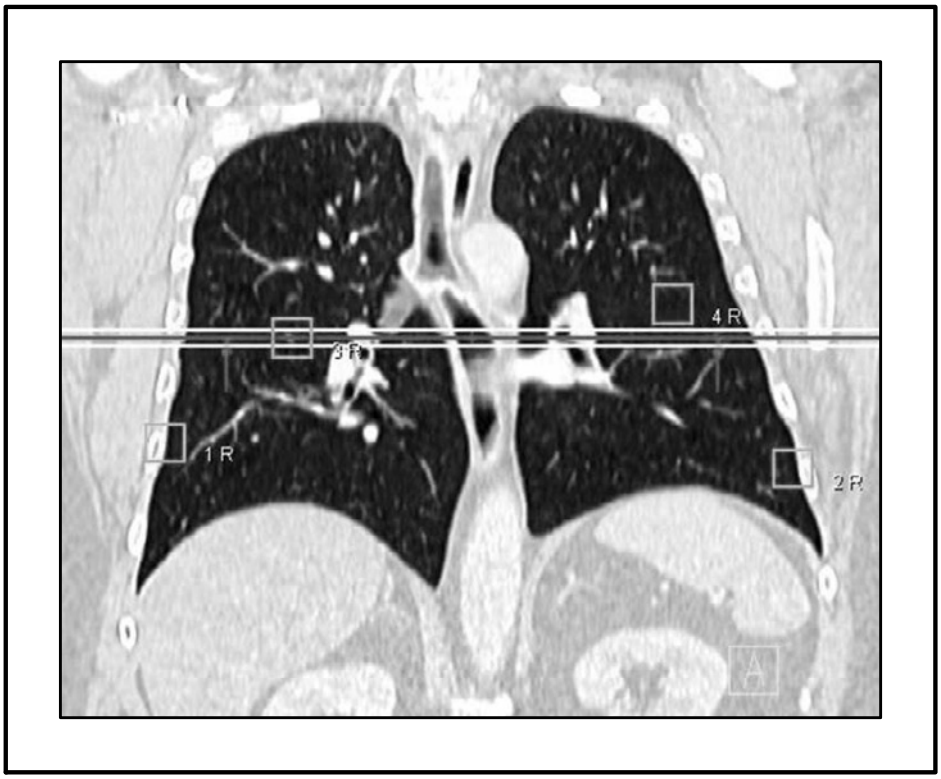
Images were read on digital workstations (Leonardo, Siemens Medical Solutions) using the Syngo Lungcare software package (Version Somaris/5 VB 10A-W) for multi-dimensional image processing and computer viewing. Lung windows were assessed at a width of 1500 to −650 Hounsfield Units.[113]

A nodule was defined as a small approximately spherical, non-linear circumscribed focus of abnormal tissue.[116] Nodules were classified as non-calcified when they did not show a benign pattern of calcification.[116] Transversal, 6 mm thick maximal intensity projections (MIP) reconstructions were used to identify pulmonary nodules. Software to aid radiologist in the detection of pulmonary nodules (Lung-CAD VB10A, Siemens AG Healthcare) was used (Figure 5).[117]

For all non-calcified nodules, the maximum dimensions in x, y and z direction, minimal, maximal and mean diameter, volume, density, location (central versus peripheral, lung segment, slice number) were recorded, as well as nodule surface characteristics (smooth, spiculated or other).[113] The nodule characteristics were uploaded in the NELSON Management System (NMS) immediately after completion of the reading for an unlimited number of evaluated nodules per scan.[113] In case of consecutive CT scans,

nodules were matched with the same nodules documented on previous scans in order to determine changes in volume and to estimate the volume doubling time (VDT).[113] This could be done either automatically, using a matching algorithm in NMS that provides the most probable match of nodules based on the combination of consistency, size and location, or manually.[113]

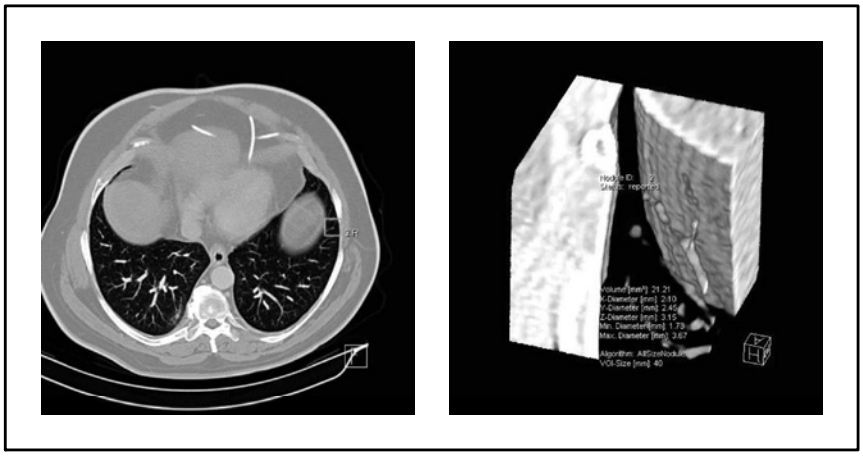
Figure 5. Computer-aided detection of pulmonary nodules in the NELSON trial



For solid nodules and for the solid component of part solid nodules, volume was calculated by three-dimensional volumetric computer assessment (Figure 6).[117]

In case of inappropriate segmentation, the radiologist was able to enter manual measurements that overrule the automatically generated volume calculations. For solid pleural based nodules, the diameter perpendicular to the costal pleura was taken to determine nodule size as the volumetric software used was not accurate enough for pleural-based lesions, due to inappropriate segmentation.[113] For non-solid lesions, nodule size was based on two-dimensional manual measurements, namely the average of length and width.[113] Length was measured in the X-Y-axis on a single CT image that showed the maximum

Figure 6. Volumetric nodule size assessment in the NELSON trial



length.[113] Width was defined as the longest diameter perpendicular to length on the same CT image.[113] For part solid lesions, both the volume of the solid part and overall size of the nodule were recorded.[113]

Throughout the study, the definition of growth was kept constant, and was defined as a percentage volume change (PVC) of 25% or more according to the following formula:

$$PVC = 100 \times ((V2 - V1) / V2)$$

Wherein PVC represents the percentage of the change in volume; V1 represents the volume of the nodule at the first screening examination, and V2 represents the volume of the nodule at the second screening examination.[113] For nodules with a PVC of 25% or more, the volume doubling time (VDT) was estimated using in following formula:

$$VDT = (\ln 2 \times \Delta t) / (\ln (V2 / V1))$$

Wherein VDT represents the volume doubling time in days, Δt represents the time interval between the two screening examinations in days, V1 the volume of the nodule at the first screening examination, and V2 represents the volume of the nodule at the second screening examination.[113] For non-calcified nodules in which only two-dimensional size parameters (dmin or dmean) were available, PVC was not used but volume doubling time was estimated using the following formula:

$$VDT = (\ln 2 \times \Delta t) / (3 \ln (D2 / D1))$$

Wherein VDT represents the volume doubling time in days, Δt represents the time interval between the two screening examinations in days, D1 the two-dimensional measurement of the nodule at the first screening examination, and D2 represents the two-dimensional measurement of the nodule at the second screening examination.[113]

After the initial reading of the screening examination, the images were made available for a second reading. The second radiologist was unaware of the conclusion of the first radiologist and read the images within 3 weeks.[113] After the second reading, discrepancies were identified by the NELSON Management System when no auto-matching was achieved or when the second reader disagreed on nodule number, location or volume.[113] In case of disagreement, an experienced expert radiologist performed a third reading and made the final decision. Finally, the nodule size category and nodule growth category were determined (Table 5).

Table 5. Nodule size and growth categories in the NELSON trial

Nodule size category	**Definition**
NODCAT I	Nodule with benign characteristics such as benign calcification patterns or fat deposition
NODCAT II	Solid nodules with volume < 50 mm^3 Pleural-based solid nodules with minimum diameter < 5 mm Non solid component part solid nodule with average diameter < 8 mm Solid component part solid nodule with volume < 50 mm^3 Non solid nodules with average diameter < 8 mm
NODCAT III	Solid nodules with volume 50 - 500 mm^3 Pleural-based solid nodules with minimum diameter 5 - 10 mm Non solid component part solid nodule with average diameter ≥ 8 mm Solid component part solid nodule with volume 50 - 500 mm^3 Non solid nodules with average diameter ≥ 8 mm
NODCAT IV	Solid nodules with volume > 500 mm^3 Pleural-based solid nodules with minimum diameter > 10 mm Solid component part solid nodule with volume > 500 mm^3
Nodule growth category	**Definition**
GROWCAT A	Percentage volume change ≥ 25% and VDT > 600 days
GROWCAT B	Percentage volume change ≥ 25% and VDT 400 - 600 days
GROWCAT C	Percentage volume change ≥ 25% and VDT < 400 days New solid component in previously non solid nodule

Nodule management protocol

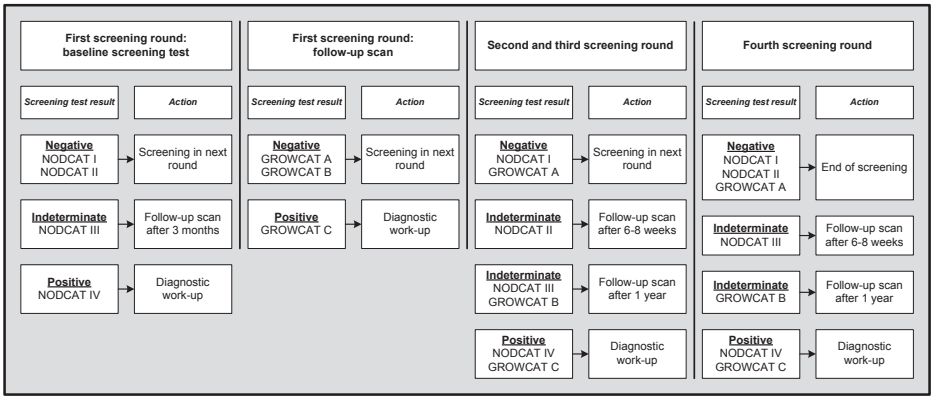
After the nodule size category and growth category are assessed, the screening test result and associated actions to be taken are determined according to the NELSON nodule management protocol. A screening test in the NELSON trial could have three different outcomes:

I) Negative: not suspicious of lung cancer, no additional diagnostic tests warranted. The participant only receives an invitation for the next screening round.
II) Indeterminate: abnormalities are identified at the screening examination; it is unclear whether these represent lung cancer. The participant receives an invitation for a follow-up CT examination to determine nodule growth.
III) Positive: abnormalities suspicious of lung cancer are identified at the screening examination. The participant receives a recommendation to consult a pulmonologist for diagnostic work-up.

An overview of the NELSON nodule management protocol is presented in figure 7. At baseline screening, only the size category of detected nodules can be determined, as only one CT examination is available. All nodules in size category NODCAT I and NODCAT II are classified as a negative screening test result (Table 5). Nodules with size category NODCAT III are classified as an indeterminate screening test result, and nodules with size category NODCAT IV are classified as a positive screening test result (Table 5). At the follow-up CT examination in the participants with indeterminate baseline screening test results, the growth category determines the final screening test result. Hence, nodules with growth category GROWCAT A or GROWCAT B are classified as a negative screening result, and nodules with growth category GROWCAT C are classified as a positive screening result.

From the second screening round onwards, a different classification is used. Nodules with size category NODCAT I or growth category GROWCAT A are classified as a negative screening test result (Table 5). All nodules with size category NODCAT II or NODCAT III or growth category GROWCAT B are classified as an indeterminate screening test result, and nodules with size category NODCAT IV or growth category GROWCAT C as a positive screening test result (Table 5).

Figure 7. Nodule management protocol of the NELSON trial



Diagnostic work-up after positive screening tests

All participants who receive a positive screening result are referred to a pulmonologist via their general practitioner. Most often, participants are referred to one of the pulmonologists involved in the NELSON trial at one of the four screening hospitals (University Medical Center Groningen, University Medical Center Utrecht, and the Kennemer Gasthuis Haarlem in the Netherlands, and University Hospital Gasthuisberg Leuven in Belgium). The nodule detected by screening, which was classified as positive, is considered suspicious for lung cancer, and a diagnostic work-up needs to be performed to diagnose or exclude lung cancer. The NELSON trial provides directives for the diagnostic work-up after a positive screening test result, but did not orchestrate its effectuation. As a result, the work-up was usually performed according to the national guideline.[60]

The diagnostic work-up usually consisted of: personal history, physical examination, regular dose contrast-enhanced CT scan from the supra-clavicular region down to the adrenals, whole-body fluordesoxyglucose (FDG) -positron emission tomography (PET) examination, and conventional white light bronchoscopy (with endobronchial washing and brushing, and biopsy of the nodule or lymph nodes if possible). CT-guided transthoracic biopsy of the suspicious nodule was performed only in a small minority of the diagnostic work-ups. Next, the results of this series of initial diagnostic procedures are discussed in the local multidisciplinary lung oncology team, which usually has members from the following departments: pulmonary medicine, thoracic surgery, radiation oncology, radiology, nuclear medicine and pathology.

In case the initial series of diagnostic procedures did not yield any result that supported the suspicion of lung cancer, or a benign cause of the nodule was identified, the multidisciplinary team would usually decide to end the clinical evaluation and to refer the participant back to the screening programme of the NELSON trial.

In case the initial series of diagnostic procedures did not yield conclusive results, the multidisciplinary team would usually recommend to perform another diagnostic CT examination after three to six months, in accordance with international guidelines.[118,119]

In case the initial series of diagnostic procedures confirmed the suspicion of lung cancer, or yielded a cytological diagnosis of lung cancer, the decision of the multidisciplinary team will depend on the clinical TNM disease stage and the participant's operability.[36,120]

Lung cancer staging

The participants of the NELSON trial who were diagnosed with lung cancer were staged according to the IASLC (International Association for the Study of Lung Cancer) TNM lung cancer staging system. This system uses criteria for the extensiveness of the primary tumour, metastasis in regional lymph nodes and metastasis at distant sites, to classify patients in subgroups with comparable prognoses. As the NELSON trial was initiated in 2003, the sixth edition of the TNM staging system[121] was used. However from 2009

onwards, the seventh edition of the staging system was used.[122] For all studies in this thesis, the lung cancers that have initially been staged according to the sixth edition, were re-staged according to the seventh edition.

To determine the clinical tumour stage (cT), the contrast enhanced CT scan and bronchoscopy are the most important diagnostic procedures. Hence, they are used to measure the size of the primary tumour, to determine the distance of the primary tumour to the lobar bronchus and carina, and to assess the presence of tumour invasion of the pleura or extra-pulmonary structures, separate tumour nodules in the ipsilateral lung, obstructive pneumonitis or atelectasis. Using this information, the T stage can be determined with the criteria presented in table 6a.

To determine the clinical node stage (cN), the contrast enhanced CT scan and the FDG-PET scan are used initially. They are used to determine whether the tumour is adjacent to the mediastinum, whether there are any hilar, mediastinal, infra-clavicular, supra-clavicular or scalene lymph nodes with a short axis diameter of 10 mm or more, or with relevant FDG uptake. If any of the previous is observed, the Dutch guideline recommends to obtain a mediastinal tissue diagnosis using endosonography or surgical mediastinoscopy.[123] Using this information, the N stage can be determined with the criteria presented in table 6b.

To determine the clinical distant metastasis stage (cM), the contrast enhanced CT scan and the FDG-PET scan are used. They are used to determine whether there are any tumour nodules in the contralateral lung, pleural nodules, pleural or pericardial effusions or distant metastasis. Additionally, the Dutch guideline recommends obtaining magnetic resonance imaging (MRI) of the skull to rule out brain metastases in patients with clinical stage III.[123,124] Using this information, the M stage can be determined with the criteria presented in table 6c.

Once the clinical T, N and M stage of the (suspected) lung cancer are known, the disease stage can be determined using the classification presented in table 7. The TNM disease stage is closely correlated with prognosis, and determines which treatment options are feasible.[60,122-125]

Treatment of screen-detected lung cancer

Participants who are diagnosed with lung cancer through the screening program are treated in accordance with the national guidelines for the treatment of non-small cell and small cell lung carcinoma.[60,123,124] The NELSON trial did not provide any directives for the treatment of screen-detected lung cancer.

Small cell lung cancer

The treatment of small cell lung cancer according the national Dutch guideline[124] is divided in two different paths; one for small cell lung cancer diagnosed at limited disease stage,

Table 6a. Criteria for tumour stage 7th edition IASLC staging protocol

T stage (primary tumour)	
Tx	Primary tumour cannot be assessed, or tumour proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumour
Tis	Carcinoma in situ
T1	Tumour ≤ 3 cm in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchusa
T1a	Tumour less ≤ 2 cm in greatest dimension
T1b	Tumour > 2 cm but ≤ 3 cm in greatest dimension
T2	Tumour > 3 cm but ≤ 7 cm or tumour with any of the following features (T2 tumours with these features are classified T2a if ≤ 5 cm):involves main bronchus, ≥ 2 cm distal to the carina, invades visceral pleura, associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung
T2a	Tumour > 3 cm but ≤ 5 cm in greatest dimension
T2b	Tumour > 5 cm but ≤ 7 cm in greatest dimension
T3	Tumour > 7 cm or directly invading any of the following: chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium; or tumour in the main bronchus (< 2 cm distal to the carina[a], but without involvement of the carina; or associated atelectasis or obstructive pneumonitis of the entire lung or separate tumour nodule(s) in the same lobe
T4	Tumour of any size that invades any of the following: mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, oesophagus, vertebral body, carina, separate tumour nodule(s) in a different ipsilateral lobe

[a] *The uncommon superficial spreading tumour of any size with its invasive component limited to the bronchial wall, which may extend proximally to the main bronchus, is also classified as T1a.*

Table 6b. Criteria for node stage 7th edition IASLC staging protocol

N stage (regional lymph nodes)	
Nx	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastases
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)

Table 6c. Criteria for node stage 7th edition IASLC staging protocol

M stage (distant metastasis)	
Mx	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumour nodule(s) in a contralateral lobe; tumour with pleural nodules or malignant pleural (or pericardial) effusion[b]
M1b	Distant metastasis

[b] *Most pleural (and pericardial) effusions with lung cancer are due to tumour. In a few patients, however, multiple cytopathologic examinations of pleural (pericardial) fluid are negative for tumour, and the fluid is non-bloody and is not an exudate. Where these elements and clinical judgement dictate that the effusion is not related to the tumour, the effusion should be excluded as a staging element and the patient should be classified as M0.*

Table 7. Stage groupings 7th edition of the staging protocol

Stage groups	T stage	N stage	M stage
Ia	T1a,b	N0	M0
Ib	T2a	N0	M0
IIa	T1a,b	N1	M0
	T2a	N1	M0
	T2b	N0	M0
IIb	T2b	N1	M0
	T3	N0	M0
IIIa	T1-3	N2	M0
	T3	N1	M0
	T4	N0,1	M0
IIIb	T4	N2	M0
	T1-4	N3	M0
IV	T1-4	N0-3	M1a,b

and another for small cell lung cancer diagnosed at extensive disease stage. Whether lung cancer is diagnosed at limited or advanced disease stage is determined by the possibility to capture all tumour in a single radiation field. In case this is possible, the disease stage is limited, which are usually unilateral tumours with no, hilar or ipsilateral mediastinal lymph node involvement, otherwise the disease stage is extensive.

Patients with small cell lung carcinomas diagnosed at limited disease stage, are recommended to be treated with multimodality therapy. This is usually concomitant chemotherapy, consisting of four cycles of Cisplatin - Etoposide, and radiotherapy, consisting of chest irradiation of 30 fractions of 1.5 Gy. For very early stage small cell lung carcinomas (T1-2N0-1M0) surgical resection may be added to the multimodality treatment. After the initial treatment, prophylactic cranial irradiation (10 fractions of 2.5 Gy) is recommended in patients without disease progression.

Patients with small cell lung carcinomas diagnosed at extensive disease stage, who have a WHO performance score of 0 to 3, are recommended to be treated with chemotherapy, for example 4 to 6 cycles of Cisplatin or Carboplatin - Etoposide. Radiotherapy is only recommended for palliative purposes, such as haemoptysis, superior vena cava syndrome or painful bone metastases. After the initial treatment, prophylactic cranial irradiation (10 fractions of 2.5 Gy) is recommended in patients without disease progression.

As the majority of the patients with limited disease and about all patients with extensive disease will be confronted with recurrence of the cancer after the initial therapy, the recommended treatment for recurrent small cell lung cancer is described as well. Hence, chemotherapy is the only therapeutic option and should be offered to all patients that are not compromised as a result of advanced age, marginal performance status, co-morbidity and complications from the first series of chemotherapy. In case the cancer was sensitive

for the first chemotherapeutic, re-induction therapy or Topotecan may be given. In case the cancer was not sensitive for the first chemotherapeutic, other chemotherapeutics should be chosen as monotherapy or combination therapy.

Non-small cell lung cancer

The treatment of non-small cell lung cancer according the national Dutch guideline[123] depends on the TNM disease stage at diagnosis.[36] Patients diagnosed with resectable or locally advanced non-small cell lung cancer are recommended to undergo surgical resection of the tumour and dissection of the mediastinal lymph nodes. In case it was not possible to establish an histological or cytological diagnosis of lung cancer pre-operatively, an initial limited resection of the tumour should be performed to confirm the diagnosis by frozen section examination.

For confirmed non-small cell lung cancers limited to one lobe, lobectomy with systematic mediastinal lymph node resection is the treatment of choice. If this is not possible due to poor pulmonary function, a more limited resection, such as a segmentectomy or wedge resection can be performed. In such patients without lymph node involvement stereotactic radiotherapy should also be considered. For lung cancers that are not limited to one lobe, complete resection can be achieved by performing a lobectomy of the one lobe and a limited resection of the other lobe, or by performing a bilobectomy or a pneumonectomy. For lung tumours that cannot be resected completely with a lobectomy due to tumour extension up to the ostium of the main bronchus of a lobe or the carina, a sleeve-lobectomy or sleeve-pneumonectomy can be performed. For lung tumours that extended up to or in the parietal pleura or thorax, an 'en bloc' resection of the affected section of the thorax should be performed. For tumours that per-operatively appear to have invaded the intra-pericardial part of the pulmonary artery, a pneumonectomy should be considered. Surgical resection of lung tumours that per-operatively appear to have invaded the superior vena cava, the adventitia of the aortic wall, the pericardium or diaphragm is not excluded. However, lung tumours that have substantially invaded the left atrium or the vertebral column are rarely resectable. Lung tumours that invade the pulmonary trunk, the oesophagus, or through the aortic wall or tumours that have caused pleuritis carcinomatosis are irresectable.

The aforementioned procedures are usually performed via thoracotomy, however video-assisted thoracoscopic procedures are also acceptable in selected patients by experienced surgeons.

Adjuvant radiotherapy is recommended in case of irradical resection and unexpected N2-3 disease. Adjuvant chemotherapy, such as four cycles of Cisplatin combination therapy, is recommended in case of stage II-IIIA disease in patients with a good performance status (WHO 0-1).

Patients with unresectable, locally advanced non-small cell lung cancer (stage III), and a good performance status are recommended to be treated with concomitant chemoradiation therapy. After this initial treatment, the tumour should be re-staged to determine whether complete resection has become an option.

Patients with advanced stage non-small cell lung cancer (stage IV) and performance stage 0-3 are recommended to be treated with combination chemotherapy. For non-squamous cell cancers, a combination therapy of Cisplatin and a third-generation cytostatic (except Gemcitabin) is recommended. However treatment with Carboplatin, Paclitaxel or Bevacizumab can also be considered. For squamous cell carcinomas, Carboplatin combination therapy is recommended (not with Pemetrexed). Only in patients a known activating EGFR-mutation, EGFR-TKIs should be used as initial treatment. EGFR-TKIs can be used second and later treatment lines in patients with known and unknown EGFR status. Pemetrexed can be used as maintenance therapy in progression-free patients after first line chemotherapy, as well as EGFR-Tyrosine-kinase-inhibitors in in patients with an activating EGFR mutation.

Follow-up

After the initial treatment of lung cancer, the Dutch guidelines[123,124] recommend to perform regular follow-up consisting of anamnesis, physical examination and possibly a chest radiograph. Follow-up using imaging, which enables the assessment of disease progression, is only recommended in case an active second or third treatment line can be offered, and in case screening for late side-effects is useful. Follow-up is recommended every three months during the first year, every six months during the second year, and every year for at least five years. The NELSON trial is not actively involved in the follow-up process of the participants who have been treated for lung cancer.

Data collection

To determine the main outcome of the NELSON trial and to be able to perform side-studies, data needs to be collected on the diagnosis, treatment and follow-up of lung cancer. This information is required for both the participants diagnosed with lung cancer through screening and the participants who were diagnosed with lung cancer outside the screening program; e.g. before screening has started, between screenings, after screening has stopped and in the participants randomised to the control group.

The first step of data collection is to identify all participants who were diagnosed with lung cancer. The information on all lung cancer diagnoses is obtained via linkages with the national cancer registries of the Netherlands[126] and Belgium,[127] which have national coverage. The second step is to collect copies of the medical files of all participants diagnosed with lung cancer from the date of the first consultation for (suspected) lung cancer, until the date death or the end of the study. Finally, to obtain medical information on the

last phase of the participants' life, the general practitioner was approached and requested to answer a number of questions concerning the cause of death.

End point verification

Lung cancer-specific mortality is the main outcome measure of the NELSON trial. Therefore, verification of the cause of death of the study participants who were ever diagnosed with lung cancer is crucial. The cause of death could be obtained by using the direct and underlying causes of death reported on the official death certificates of the deceased participants. However, the use of the official certificates for this purpose is debated for several reasons. Firstly, two forms of bias especially affect death certification in screening trials:

I) Sticky-diagnosis bias: CT screening leads to an increased incidence of lung cancer through advanced diagnoses and overdiagnosis. As a result, the prevalence of lung cancer is higher in the screening group than in the control group. Since lung cancer is commonly recognised as a lethal disease, the deaths in the screening group are more likely to be attributed to lung cancer than deaths in the control group.[128]

II) Slippery-linkage bias: deaths as a result of interventions of treatments for lung cancer may be difficult to trace back to screening and could easily be certified as death due to other causes.[128]

Secondly, the merit of death certificates depends on the accuracy of the certifying clinician and nosologist, and the establishment of a correct ante mortem diagnosis.[129,130] Common reasons for misclassification are coinciding malignancies, considerable comorbidity and death after a surgical procedure.[131,132] Finally, the sensitivity and specificity of the death certificate has been reported to range from 84.5 to 99.7% and 91.3 to 99.7%; causing an error that tends to reduce the effect of screening.[132-135]

To overcome these problems, clinical expert committees that review the medical files of the deceased participants to determine the cause of death, are frequently employed in cancer screening trials.[132-137] Assessing the cause of death by such a committee should yield an uniform, objective ad unbiased determination of the trials' main end point. The development of a cause of death review process protocol for the NELSON trial was part of this thesis (Chapter 9).

RESEARCH QUESTIONS

The purpose of this thesis was to evaluate lung cancer screening using low-dose computed tomography in the Dutch-Belgian NELSON trial. Implications for future lung cancer screening programs were identified by assessing the screening strategy's performance and outcomes. The research questions and hypotheses of the studies described in the subsequent chapters of this thesis are described next.

Research question I

Chapter 2. Predictive value of screening test results

Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial.

European Respiratory Journal

Main research question

What was the screening performance of the nodule management protocol of the NELSON trial?

Sub research questions

a) What were the detection rates, test characteristics and numbers needed to screen of the nodule management protocol of the NELSON trial?
b) What was the incidence of invasive diagnostic procedures for false-positive screening test results?
c) What were participant's probabilities of false-positive screening results and lung cancer after baseline and subsequent screening test results?

Research question II

Chapter 3. Characteristics of screen-detected lung cancer

Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial.

American Journal of Respiratory and Critical Care Medicine

Main research question

What was the effect of screening using low-dose computed tomography on the characteristics of screen-detected lung cancer?

Sub research questions

a) What were the tumour characteristics of lung cancers detected by low-dose CT screening?
b) What was the effect of screening round and gender on the characteristics of screen-detected lung cancer?
c) To what extent was screening able to detect lung cancer before the onset of symptoms?

Research question III

Chapter 4. Epidemiological evaluation

Detection of lung cancer through low-dose CT screening: analysis of screening test performance and interval cancers.

Lancet Oncology

Main research question

How can knowledge on the lung cancers not detected by low-dose computed tomography screening be used to improve the performance of the screening strategy?

Sub research questions

a) What were the detection rates and test characteristics of the nodule management protocol of the NELSON trial?
b) Were there any differences in the characteristics between the participants diagnosed with screen-detected lung cancer and the participants diagnosed with an interval cancer?
c) What were the tumour characteristics of the lung cancers not detected by low-dose computed tomography screening?
d) What were the causes of the failure to detect the interval cancers?

Research question IV

Chapter 5. Radiological evaluation

Computed tomographic characteristics of interval and post-screen carcinomas in lung cancer screening.

European Radiology

Main research question

How can knowledge on the radiological characteristics of lung cancers not detected by low-dose CT screening be used to improve the performance of the screening strategy?

Sub research questions

a) What proportion of the lung cancers not diagnosed through screening were, in retrospect, present at the last LDCT screening examination?
b) What were the causes of the failure to detect the missed lung cancers?
c) What were the characteristics of the carcinomas missed on the LDCT screening examination due to radiological detection or interpretation errors?

Research question V

Chapter 6. Optimisation of screening protocols

Lung cancer probability in subjects with CT-detected pulmonary nodules: an analysis of data from the NELSON trial of low-dose CT screening.

Lancet Oncology

Main research question

How should a participant's predicted lung cancer probability, based on size and growth of CT-detected nodules, be used to optimise the nodule management protocol of the NELSON trial?

Sub research questions

a) Was it valid to predict the two-year lung cancer probability of an individual who underwent screening using low-dose computed tomography, using a model based on nodule size and growth rate?

b) What was the probability of lung cancer in an individual who underwent screening using low-dose computed tomography, based on nodule size and growth rate?
c) How should the current thresholds for nodule size and growth rate be adjusted to improve risk stratification, test characteristics and reduce harms?

Research question VI

Chapter 7. Evaluation of bronchoscopy

The role of conventional bronchoscopy in the work-up of suspicious CT screen-detected pulmonary nodules.

Chest

Main research question

What was the value of bronchoscopy for diagnosing lung cancer in screen-detected nodules?

Sub research questions

a) What were the test characteristics of bronchoscopy and its ancillary procedures?
b) What were predictors for a true-positive bronchoscopic procedure?
c) Which diagnoses were made in false-negative bronchoscopic procedures?

Research question VII

Chapter 8. Evaluation of surgical procedures

Complications following lung surgery in the Dutch-Belgian randomized lung cancer screening trial.

European Journal of Cardio-Thoracic Surgery

Main research question

To what extent did adverse events related to thoracic surgery, occur in participants after a positive screening test results?

Sub research questions

a) How often occurred re-thoracotomy, complications, and post-operative mortality in participants who underwent thoracic surgery for a positive screening test result?

b) What was the length of hospital stay for lung resection performed by thoracotomy and video-assisted thoracoscopic surgery?
c) To what extent were surgical procedures performed for benign nodules?

Research question VIII

Chapter 9. Endpoint determination

Uniform and blinded cause of death verification in a lung cancer CT screening trial.

Lung Cancer

Main research question

How should the endpoint verification process of the NELSON trial be designed to ensure uniform, objective and unbiased endpoint determination?

Sub research questions

a) How to develop a cause of death review protocol that ensures uniform, objective and unbiased endpoint determination?
b) How was the performance of the developed cause of death protocol compared to the official death certificates?
c) What were sources of disagreement between users of the developed cause of death protocol?
d) What were the best sources of information for a review of the cause of death of a participant?

OUTLINE OF THIS THESIS

Part I of this thesis "Introduction" consists of the General introduction (*Chapter 1*). Part II of this thesis "Evaluation of findings" consists of four chapters, covering several aspects of the performance screening algorithms. In the first study (*Chapter 2*) , data on screening test results and screen-detected lung cancer were used to determine positive predictive value and 5.5-year lung cancer probability. In the second study (*Chapter 3*) the tumour characteristics of the lung cancers detected by screening are presented. In the third study (*Chapter 4*), the performance of the screening algorithm of the NELSON trial was estimated, and opportunities to improve the performance were identified. In the fourth study of part II of this thesis (*Chapter 5*), radiological causes of the failure to detect the lung cancers not diagnosed through screening were assessed, and opportunities to improve the performance of the screening algorithm were identified. Part III of this thesis "Optimisation of screening" presents three studies. In the first study, (*Chapter 6*), lung cancer probability of participants was estimated and used to design improved nodule management protocols. In the second study (*Chapter 7*), the value of bronchoscopy in the diagnostic work-up of suspicious CT-detected nodules was determined. In the third study (*Chapter 8*), adverse events related to thoracic surgery, performed in the diagnostic work-up of suspicious CT-detected nodules were assessed. Part IV of this thesis "Evaluation of effectiveness" consists of one study (*Chapter 9*), which presents the design and evaluation of the endpoint verification process of the NELSON trial. Part V of this thesis "Implications for implementation", presents an overview of lung cancer screening and the studies presented in the parts II to IV of this thesis. Firstly, a review of the currently published literature (*Chapter 10*) is performed to determine the state of the art in lung cancer screening. Secondly, a review of the studies of this thesis (*Chapter 11*) is performed to interpret important results, answer the research questions of this thesis, and formulate general conclusions and recommendations. Furthermore, a summary of this thesis in English (*Chapter 12*) and in Dutch (*Chapter 13*) is provided. Finally, part VI of this thesis "Miscellaneous" consists of acknowledgements (*Chapter 14*), curriculum vitae (*Chapter 15*), PhD portfolio of the Erasmus University (*Chapter 16*), and the list of publications (*Chapter 17*).

REFERENCE LIST

1. Doll R, Peto R, Boreham J, et al. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* 2004; 328(7455):1519.
2. Thun MJ, Carter BD, Feskanich D, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med* 2013; 368(4):351-64.
3. Alberg AJ, Brock MV, Ford JG, et al. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(Suppl5):e1S-29S.
4. Sasco AJ, Secretan MB, Straif K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. *Lung Cancer* 2004; 45(Suppl2):S3-9.
5. Besaratinia A, Pfeifer GP. Second-hand smoke and human lung cancer. *Lancet Oncol* 2008; 9(7):657-66.
6. Oberg M, Jaakkola MS, Woodward A, et al. Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. *Lancet* 2011; 377(9760):139-46.
7. Linet MS, Slovis TL, Miller DL, et al. Cancer risks associated with external radiation from diagnostic imaging procedures. *CA Cancer J Clin* 2012; 62(2):75-100.
8. Youlden DR, Cramb SM, Baade PD. The International Epidemiology of Lung Cancer: geographical distribution and secular trends: official publication of the International Association for the Study of Lung Cancer. *J Thorac Oncol* 2008; 3(8):819-31.
9. Baan RA. Carcinogenic hazards from inhaled carbon black, titanium dioxide, and talc not containing asbestos or asbestiform fibers: recent evaluations by an IARC Monographs Working Group. *Inhal Toxicol* 2007; 19(Suppl1):213-28.
10. Mc Laughlin J. An historical overview of radon and its progeny: applications and health effects. *Radiat Prot Dosimetry* 2012; 152(1-3):2-8.
11. Zhang ZL, Sun J, Dong JY, et al. Residential radon and lung cancer risk: an updated meta- analysis of case-control studies. *Asian Pac J Cancer Prev* 2012; 13(6):2459-65.
12. Sethi TK, El-Ghamry MN, Kloecker GH. Radon and lung cancer. *Clin Adv Hematol Oncol* 2012; 10(3):157-64.
13. Berry G, Gibbs GW. An overview of the risk of lung cancer in relation to exposure to asbestos and of taconite miners. *Regul Toxicol Pharmacol* 2008; 52(1 Suppl):S218-22.
14. van der Bij S, Koffijberg H, Lenters V, et al. Lung cancer risk at low cumulative asbestos exposure: meta-regression of the exposure-response relationship. *Cancer Causes Control* 2013; 24(1):1-12.
15. Alberg AJ, Yung R, Strickland PT, et al. Respiratory cancer and exposure to arsenis, chromium, nickel and polycyclic aromatic hydrocarbons. *Clin Occup Environ Med* 2002; 2(4):779-801.
16. Tammemagi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med* 2013; 368(8):728-36.
17. Raji OY, Duffy SW, Agbaje OF, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. *Ann Intern Med* 2012; 157(4):242-50.
18. Wu CY, Hu HY, Pu CY, et al. Pulmonary tuberculosis increases the risk of lung cancer: a population-based cohort study. *Cancer* 2011; 117(3):618-24.
19. Liang HY, Li XL, Yu XS, et al. Facts and fiction of the relationship between preexisting tuberculosis and lung cancer risk: a systematic review. *Int J Cancer* 2009; 125(12):2936-44.
20. Brenner DR, Boffetta P, Duell EJ, et al. Previous lung diseases and lung cancer risk: a pooled analysis from the International Lung Cancer Consortium. *Am J Epidemiol* 2012; 176(7):573-85.

21. Brenner DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. *PloS one* 2011; 6(3):e17479.
22. Shiels MS, Cole SR, Kirk GD, et al. A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals. *J Acquir Immune Defic Syndr* 2009; 52(5):611-22.
23. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin* 2013; 63(1):11-30.
24. Bray FI, Weiderpass E. Lung cancer mortality trends in 36 European countries: secular trends and birth cohort patterns by sex and region 1970-2007. *Int J Cancer* 2010; 126(6):1454-66.
25. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011; 61(2):69-90.
26. Food and Drug Administration. Drug approvals and databases. http://www.fda.gov/Drugs/InformationOnDrugs/default.htm 2013. Date last accessed 15-01-2014.
27. Booth CM, Shepherd FA, Peng Y, et al. Adoption of adjuvant chemotherapy for non-small-cell lung cancer: a population-based outcomes study. *J Clin Oncol* 2010; 28(21):3472-8.
28. Pisters KM, Evans WK, Azzoli CG, et al. Cancer Care Ontario and American Society of Clinical Oncology adjuvant chemotherapy and adjuvant radiation therapy for stages I-IIIA resectable non-small cell lung cancer guideline. *J Clin Oncol* 2007; 25(34):5506-18.
29. Palma D, Visser O, Lagerwaard FJ, et al. Impact of introducing stereotactic lung radiotherapy for elderly patients with stage I non-small-cell lung cancer: a population-based time-trend analysis. *J Clin Oncol* 2010; 28(35):5153-9.
30. Timmerman R, Paulus R, Galvin J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA* 2010; 303(11):1070-6.
31. van der Drift MA, Karim-Kos HE, Siesling S, et al. Progress in standard of care therapy and modest survival benefits in the treatment of non-small cell lung cancer patients in the Netherlands in the last 20 years. *J Thorac Oncol* 2012; 7(2):291-8.
32. Auperin A, Le Pechoux C, Rolland E, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. *J Clin Oncol* 2010; 28(13):2181-90.
33. Pfister DG, Johnson DH, Azzoli CG, et al. American Society of Clinical Oncology treatment of unresectable non-small-cell lung cancer guideline: update 2003. *J Clin Oncol* 2004; 22(2):330-53.
34. Birim O, Maat AP, Kappetein AP, et al. Validation of the Charlson comorbidity index in patients with operated primary non-small cell lung cancer. *Eur J Cardiothorac Surg* 2003; 23(1):30-4.
35. Ost DE, Yeung SC, Tanoue LT, et al. Clinical and organizational factors in the initial evaluation of patients with lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(Suppl5):e121S-41S.
36. Detterbeck FC, Postmus PE, Tanoue LT. The stage classification of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(Suppl5):e191S-210S.
37. Vinod SK, Sidhom MA, Gabriel GS, et al. Why do some lung cancer patients receive no anticancer treatment? *J Thorac Oncol* 2010; 5(7):1025-32.
38. Coleman MP, Forman D, Bryant H, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011; 377(9760):127-38.
39. Karim-Kos HE, de Vries E, Soerjomataram I, et al. Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. *Eur J Cancer* 2008; 44(10):1345-89.
40. Peto R, Gray R, Collins R, et al. Randomised trial of prophylactic daily aspirin in British male doctors. *Br Med J* 1988; 296(6618):313-6.

41. Steering committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989; 321(2):129-35.
42. Cook NR, Lee IM, Gaziano JM, et al. Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. *JAMA* 2005; 294(1):47-55.
43. Blumberg J, Block G. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study in Finland. *Nutr Rev* 1994; 52(7):242-5.
44. Hennekens CH, Buring JE, Manson JE, et al. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med* 1996; 334(18):1145-9.
45. Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med* 1996; 334(18):1150-5.
46. van Zandwijk N, Dalesio O, Pastorino U, et al. EUROSCAN, a randomized trial of vitamin A and N-acetylcysteine in patients with head and neck cancer or lung cancer. For the EUropean Organization for Research and Treatment of Cancer Head and Neck and Lung Cancer Cooperative Groups. *J Natl Cancer Inst* 2000; 92(12):977-86.
47. Lippman SM, Lee JJ, Karp DD, et al. Randomized phase III intergroup trial of isotretinoin to prevent second primary tumors in stage I non-small-cell lung cancer. *Journal of the National Cancer Institute* 2001; 93(8):605-18.
48. Karp DD, Lee SJ, Keller SM, et al. Randomized, double-blind, placebo-controlled, phase III chemoprevention trial of selenium supplementation in patients with resected stage I non-small-cell lung cancer: ECOG 5597. *J Clin Oncol* 2013; 31(33):4179-87.
49. Fritz H, Kennedy D, Fergusson D, et al. Selenium and lung cancer: a systematic review and meta-analysis. *PloS one* 2011; 6(11):e26259.
50. Szabo E, Mao JT, Lam S, et al. Chemoprevention of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(Suppl5):e40S-60S.
51. The health consequences of active smoking: a report of the surgeon general. Atlanta, GA: Department of health and human services, centers for chronic disease prevention and health promotion, office on smoking and health; 2004.
52. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *Br Med J* 1950; 2(4682):739-48.
53. Rothman KJ, Greenland S. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
54. Burns DM, Lee L, Shen LZ, et al. Cigarette smoking behaviour in the United States. In: Shopland DR, editor. Smoking and tobacco control monographs. Monograph 8: Changes in cigarette-related disease risks and their implications for prevention and control: National Cancer Institute; 1977.
55. Holford TR, Meza R, Warner KE, et al. Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964-2012. *JAMA* 2014; 311(2):164-71.
56. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975-2000. *N Engl J Med* 2012; 104(7):541-8.
57. Mendez D, Alshanqeety O, Warner KE. The potential impact of smoking control policies on future global smoking trends. *Tob Control* 2013; 22(1):46-51.
58. Corner J, Hopkinson J, Roffe L. Experience of health changes and reasons for delay in seeking care: a UK study of the months prior to the diagnosis of lung cancer. *Soc Sci Med* 2006; 62(6):1381-91.
59. Burgess C, Hunter MS, Ramirez AJ. A qualitative study of delay among women reporting symptoms of breast cancer. *Br J Gen Pract* 2001; 51(473):967-71.

60. van Meerbeeck JP, Koning CC, Tjan-Heijnen VC, et al. Guideline on 'non-small cell lung carcinoma; staging and treatment'. Richtlijn 'Niet-kleincellig longcarcinoom; stadiëring en behandeling'. *Ned Tijdschr Geneeskd* 2005; 149(2):72-7.
61. Andersen BL, Cacioppo JT, Roberts DC. Delay in Seeking a Cancer-Diagnosis - Delay Stages and Psychophysiological Comparison Processes. *Br J Soc Psychol* 1995; 34:33-52.
62. Leydon GM, Bynoe-Sutherland J, Coleman MP. The journey towards a cancer diagnosis: the experiences of people with cancer, their family and carers. *Eur J Cancer Care* 2003; 12(4):317-26.
63. Ristvedt SL, Trinkaus KM. Psychological factors related to delay in consultation for cancer symptoms. *Psychooncology* 2005; 14(5):339-50.
64. Elgood J. Public awareness of lung cancer symptoms: the GLCC/IPSOS MORI 2013 international consumer poll. Abstract presented at the 15th world conference on lung cancer 2013 http://www.2013worldlungcancer.org/documents/WCLC2013-AbstractBook.pdf Date last accessed 15-01-2014.
65. Myrdal G, Lambe M, Hillerdal G, et al. Effect of delays on prognosis in patients with non-small cell lung cancer. *Thorax* 2004; 59(1):45-9.
66. Lung Cancer Alliance. Make a difference: getting involved. http://www.lungcanceralliance.org/get-involved/help-raise-awareness/lung-cancer-awareness-month.html 2014. Date last accessed 16-01-2014.
67. Public Health England. Be alert to symptoms of lung cancer. National Health Service: Be clear on cancer 2013. http://campaigns.dh.gov.uk/2013/07/04/be-alert-to-symptoms-of-lung-cancer-new-nhs-campaign-urges Date last accessed 16-01-2014.
68. Global Lung Cancer Coalition. Global awareness of lung cancer symptoms. 2013. http://www.lungcancercoalition.org/ Date last accessed 16-01-2014.
69. Baird J. Raising the public profile of lung cancer--report of a National Lung Cancer Awareness campaign in the UK. *Lung Cancer* 2003; 42(1):119-23.
70. Athey VL, Suckling RJ, Tod AM, et al. Early diagnosis of lung cancer: evaluation of a community-based social marketing intervention. *Thorax* 2012; 67(5):412-7.
71. Sawa T, Eguchi K, Iwamoto Y, et al. Clinicians as advocates: raising public awareness of lung cancer - the west Japan onology group experience. Abstract presented at the world lung cancer conference 2013. http://www.2013worldlungcancer.org/documents/WCLC2013-AbstractBook.pdf Date last accessed 16-01-2014.
72. Austoker J, Bankhead C, Forbes LJ, et al. Interventions to promote cancer awareness and early presentation: systematic review. *Br J Cancer* 2009; 101(Suppl2):S31-9.
73. Peake MD. How clinicians and patients can benefit from better data on lung cancer. Abstract presented at the world lung cancer conference. 2013. http://www.2013worldlungcancer.org/documents/WCLC2013-AbstractBook.pdf Date last accessed 16-01-2014.
74. Brown P. Naming and framing: the social construction of diagnosis and illness. *J Health Soc Behav* 1995; SpecNo:34-52.
75. Pink Ribbon International. http://www.pinkribbon.org/ Date last accessed 16-01-2014.
76. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365(5):395-409.
77. Wilson JM, Jungner YG. Principles and practice of mass screening for disease. *Bol Oficina Sanit Panam* 1968; 65(4):281-393.
78. Andermann A, Blancquaert I, Beauchamp S, et al. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ* 2008; 86(4):317-9.
79. Genetic screening: a supplement to the 1993 report by the Nuffield council on bioethics. London; 2006.

80. Andermann A, Blancquaert I, Beauchamp S, et al. Guiding policy decisions for genetic screening: developing a systematic and transparent approach. *Public Health Genomics* 2011; 14(1):9-16.
81. Cornel MC, Rigter T, Weinreich SS, et al. A framework to start the debate on neonatal screening policies in the EU: an Expert Opinion Document. *Eur J Hum Genet* 2014; 22(1):12-7.
82. Hall AE, Chowdhury S, Pashayan N, et al. What ethical and legal principles should guide the genotyping of children as part of a personalised screening programme for common cancer? *J Med Ethics* 2014; 40(3):163-7.
83. Petros M. Revisiting the Wilson-Jungner criteria: how can supplemental criteria guide public health in the era of genetic screening? *Genet Med* 2012; 14(1):129-34.
84. Harris R, Sawaya GF, Moyer VA, et al. Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive services task force. *Epidemiol Rev* 2011; 33(1):20-35.
85. Jett JR, Horeweg N, de Koning HJ. Screening for lung cancer. In: Pass HI, Ball D, Scagliotti GV, editors. *The IASLC Multidisciplinary approach to thoracic oncology*. Aurora, Colorado, USA: IASLC Press; 2014.
86. Frost JK, Ball WC, Jr., Levin ML, et al. Sputum cytopathology: use and potential in monitoring the workplace environment by screening for biological effects of exposure. *J Occup Med* 1986; 28(8):692-703.
87. Melamed MR, Flehinger BJ, Zaman MB, et al. Screening for early lung cancer. Results of the Memorial Sloan-Kettering study in New York. *Chest* 1984; 86(1):44-53.
88. Fontana RS, Sanderson DR, Woolner LB, et al. Lung cancer screening: the Mayo program. *J Occup Med* 1986; 28(8):746-50.
89. Oken MM, Hocking WG, Kvale PA, et al. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial. *JAMA* 2011; 306(17):1865-73.
90. Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology* 1996; 201(3):798-802.
91. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 354(9173):99-105.
92. Sone S, Li F, Yang ZG, Honda T, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* 2001; 84(1):25-32.
93. Pastorino U, Bellomi M, Landoni C, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. *Lancet* 2003; 362(9384):593-7.
94. Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: five-year prospective experience. *Radiology* 2005; 235(1):259-65.
95. Sobue T, Moriyama N, Kaneko M, et al. Screening for lung cancer with low-dose helical computed tomography: anti-lung cancer association project. J Clin Oncol 2002; 20(4):911-20.
96. Henschke CI. Survival of patients with clinical stage I lung cancer diagnosed by computed tomography screening for lung cancer. Clin Cancer Res. 2007; 13(17): 4949-50.
97. National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. *Radiology* 2011; 258(1):243-53.
98. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120(4):868-74.
99. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish randomized lung cancer CT screening trial--overall design and results of the prevalence round. *J Thorac Oncol* 2009; 4(5):608-14.
100. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev* 2012; 21(3):308-15.

101. Becker N, Motsch E, Gross ML, et al. Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. *J Cancer Res Clin Oncol* 2012; 138(9):1475-86.
102. Baldwin DR, Duffy SW, Wald NJ, et al. UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. *Thorax* 2011; 66(4):308-13.
103. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer* 2009; 64(1):34-40.
104. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer* 2008; 59(3):355-63.
105. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; 67(4):296-301.
106. Infante M, Cavuto S, Lutman FR, et al. A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial. *Am J Respir Crit Care Med* 2009; 180(5):445-53.
107. Humphrey LL, Deffebach M, Pappas M, et al. Screening for Lung Cancer With Low-Dose Computed Tomography: A Systematic Review to Update the U.S. Preventive Services Task Force Recommendation. *Ann Intern Med* 2013; 159(6):411-20.
108. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361(23):2221-9.
109. Burns DM, Shanks TG, Choi W, et al. The American Cancer Society Cancer Prevention Study I: 12-year follow up of 1 million men and women. Changes in cigarette-related risks and their implications for prevention and control Smoking and tobacco control monograph No 8. Burns DM, Garfinkel L, Samet J, eds. Bethesda, MD: US Department of Health and Human Services, Public Health Service, National Institute of Health, National Cancer Institute. 1997; Chapter3:113-304.
110. Thun MJ, Meyers DG, Day-Lally C, et al. Age and the exposure-response relationships between cigarette smoking and premature death in Cancer Prevention Study II. Changes in cigarette-related risks and their implications for prevention and control. Smoking and tobacco control monograph No 8. Burns DM, Garfinkel L, Samet J, eds. Bethesda, MD: US Department of Health and Human Services, Public Health Service, National Institute of Health, National Cancer Institute. 1997; (Chapter 5):383-475.
111. Gohagan JK, Prorok PC, Kramer BS, et al. Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute. *J Urol* 1994; 152(5Pt2):1905-9.
112. de Koning HJ, Liem MK, Baan CA, et al. Prostate cancer mortality reduction by screening: power and time frame with complete enrollment in the European Randomised Screening for Prostate Cancer (ERSPC) trial. *Int J Cancer* 2002; 98(2):268-73.
113. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54(2):177-84.
114. Heuvelmans MA, Oudkerk M, de Bock GH, et al. Optimisation of volume-doubling time cutoff for fast-growing lung nodules in CT lung cancer screening reduces false-positive referrals. *Eur Radiol* 2013; 23(7):1836-45.
115. Zhao Y, de Bock GH, Vliegenthart R, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. *Eur Radiol* 2012; 22(10):2076-84.
116. Austin JH, Muller NL, Friedman PJ, et al. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology* 1996; 200(2):327-31.

117. Siemens Healthcare Nederland. syngo LungCARE CT and syngo Lung CAD. 2014. https://www.healthcare.siemens.nl/computed-tomography/options-upgrades/clinical-applications/syngo-lungcare-ct-and-syngo-lung-cad Date last accessed 06-03-2014.
118. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005; 237(2):395-400.
119. Naidich DP, Bankier AA, MacMahon H, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013; 266(1):304-17.
120. Brunelli A, Kim AW, Berger KI, et al. Physiologic evaluation of the patient with lung cancer being considered for resectional surgery: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(5Suppl):e166S-90S.
121. UICC TNM Classification of malignant tumours. In: Sobin LH, Wittekind C, editors. New York; 2002. p.272.
122. Goldstraw P, Crowley J, Chansky K, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol* 2007; 2(8):706-14.
123. Longtumoren LW. Landelijke richtlijn Niet kleincellig longcarcinoom. Richtlijnen oncologische zorg 22-05-2011. http://www.oncoline.nl/niet-kleincellig-longcarcinoom Date last accessed 07-03-2014.
124. Nederlandse Vereniging van Artsen voor Longziekten en Tuberculose. Landelijke richtlijn Kleincellig longcarcinoom. Richtlijnen oncologische zorg 10-05-2011. http://www.oncoline.nl/kleincellig-longcarcinoom Date last accessed 07-03-2014.
125. Shepherd FA, Crowley J, Van Houtte P, et al. The International Association for the Study of Lung Cancer lung cancer staging project: proposals regarding the clinical staging of small cell lung cancer in the forthcoming (seventh) edition of the tumor, node, metastasis classification for lung cancer. *J Thorac Oncol* 2007; 2(12):1067-77.
126. Dutch Cancer Registry. 2011. http://www.cijfersoverkanker.nl/about-the-registration-37.html Date last accessed 13-03-2014.
127. Belgian Cancer Registry. http://www.kankerregister.org/Home Date last accessed 13-03-2014.
128. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst* 2002; 94(3):167-73.
129. Harteloh P, de Bruin K, Kardaun J. The reliability of cause-of-death coding in The Netherlands. *Eur J Epidemiol* 2010; 25(8):531-8.
130. Maudsley G, Williams EM. "Inaccuracy' in death certification--where are we now? *J Public Health Med* 1996; 18(1):59-66.
131. Albertsen PC, Walters S, Hanley JA. A comparison of cause of death determination in men previously diagnosed with prostate cancer who died in 1985 or 1995. *J Urol* 2000; 163(2):519-23.
132. Makinen T, Karhunen P, Aro J, et al. Assessment of causes of death in a prostate cancer screening trial. *Int J Cancer* 2008; 122(2):413-7.
133. Otto SJ, van Leeuwen PJ, Hoekstra JW, et al. Blinded and uniform causes of death verification in cancer screening: a major influence on the outcome of a prostate cancer screening trial? *Eur J Cancer* 2010; 46(17):3061-7.
134. Doria-Rose VP, Marcus PM, Miller AB, et al. Does the source of death information affect cancer screening efficacy results? A study of the use of mortality review versus death certificates in four randomized trials. *Clin Trials* 2010; 7(1):69-77.

135. Doria-Rose VP, Marcus PM. Death certificates provide an adequate source of cause of death information when evaluating lung cancer mortality: an example from the Mayo Lung Project. *Lung Cancer* 2009; 63(2):295-300.
136. Ederer F, Geisser MS, Mongin SJ, et al. Colorectal cancer deaths as determined by expert committee and from death certificate: a comparison. The Minnesota Study. *J Clin Epidemiol* 1999; 52(5):447-52.
137. Chamberlain J, Coleman D, Ellman R, et al. Verification of the cause of death in the trial of early detection of breast cancer. UK Trial of Early Detection of Breast Cancer Group. Trial Co-ordinating Centre. *Br J Cancer* 1991; 64(6):1151-6.
138. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Baseline characteristics of participants in the randomized national lung screening trial. *J Natl Cancer Inst* 2010; 102(23):1771-9.
139. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J* 2013;42(6):1659-67.
140. Field JK, Devaraj A, Baldwin DR, et al. UK lung cancer screening trial (UKLS): baseline data. Poster presented at the world lung cancer conference 2013. http://www.2013worldlungcancer.org/documents/WCLC2013-AbstractBook.pdf Date last accessed 17-01-2014.

Part II

Evaluation of findings

Chapter 2

Predictive value of screening test results

Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial

Horeweg N,
van der Aalst CM,
Vliegenthart R,
Zhao YR, Xie X,
Scholten ET,
Mali W,
Thunnissen E,
Weenink C,
Groen HJM,
Lammers JWJ,
Nackaerts K,
van Rosmalen J,
Oudkerk M,
de Koning HJ

European Respiratory Journal
December 2013

Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial

European Respiratory Journal. 2013 Dec; 42(6):1659-67.

ABSTRACT

Several medical associations recommended lung cancer screening by low-dose computed tomography scanning for high-risk groups. Counselling of the candidates on the potential harms and benefits and their lung cancer risk is a prerequisite for screening.

In the NELSON trial, screenings are considered positive for (part) solid lung nodules with a volume >500 mm3 and for (part) solid or nonsolid nodules with a volume-doubling time <400 days. For this study, the performance of the NELSON strategy in three screening rounds was evaluated and risk calculations were made for a follow-up period of 5.5 years.

458 (6%) of the 7582 participants screened had a positive screen result and 200 (2.6%) were diagnosed with lung cancer. The positive screenings had a predictive value of 40.6% and only 1.2% of all scan results were false-positive. In a period of 5.5 years, the risk of screen-detected lung cancer strongly depends on the result of the first scan: 1.0% after a negative baseline result, 5.7% after an indeterminate baseline and 48.3% after a positive baseline.

The screening strategy yielded few positive and false-positive scans with a reasonable positive predictive value. The 5.5-year lung cancer risk calculations aid clinicians in counselling candidates for lung cancer screening with low-dose computed tomography.

INTRODUCTION

A number of prominent medical associations recently recommended screening for lung cancer in high-risk groups by low-dose computed tomography (LDCT) scanning.[1–4] The recommendation resulted from the efforts that have been made by many researchers over the past decade, especially by the National Lung Screening Trial (NLST) research team.[5] The latest systematic review on computed tomography (CT) screening for lung cancer concluded that there are still substantial uncertainties regarding how to translate the positive recommendation into clinical practice.[6]

Most individuals eligible for screening will not develop lung cancer but are exposed to several potential harms: radiation exposure, psychological distress while awaiting results, and distress, morbidity and mortality in case of false-positive results.[7,8] However, for individuals who actually will develop lung cancer, LDCT screening is often able to detect lung cancer at an early stage.[5,9,10] The NLST has demonstrated that LDCT screening reduces the risk of dying from lung cancer significantly.[5] Nevertheless, the early detection of lung cancer also leads to a prolonged disease course and will not be beneficial in persons who would otherwise never be diagnosed with lung cancer.

Therefore, to be able to counsel individuals adequately on the benefits and harms of LDCT-screening, clinicians should inform the candidates of their risk of true-positive and false-positive screen results.[6] In the NLST, for example, 24.2% of the subjects had a positive screening, but only 3.6% was diagnosed with lung cancer.[5] Furthermore, to be able to make an informed choice on future screenings, high-risk subjects should know how their probability of screen-detected lung cancer changes after their first screening.

In our trial, the Dutch–Belgian lung cancer screening trial (NELSON) solid lung nodules are assessed with three-dimensional measurements (volume). Screening results are considered positive for volumes >500 mm^3 (diameter ~9.8 mm) or volume-doubling times (VDT) <400 days.[9,11] This is considerably more stringent than the NLST policy to refer any nodule with a maximum diameter ≥4 mm.[11,12] The volumetry-based screening strategy of the Danish lung cancer screening trial (DLCST) was adopted from our trial and led to a positive screen result in 2.0% of the participants with 34.8% of these results being true-positive.[10,13]

In this study, we will evaluate the performance of the NELSON screening strategy in the first three screening rounds. We will calculate lung cancer detection rates and positive predictive values and compare our results with other LDCT screening trials. Furthermore, we will calculate the 5.5-year risk of false-positive screen results and screen-detected lung cancer stratified by the result of the first screening scan. This will provide valuable information for clinicians who are confronted with individuals who consider or have already undergone LDCT screening for lung cancer.

METHODS

Details of the design and conduct of the NELSON trial have been reported elsewhere.[11,14] Briefly, subjects aged 50–75 years, who had smoked either 15 cigarettes or more per day for 25 years or 10 cigarettes or more for 30 years and were still smoking or had quit less than 10 years ago met the inclusion criteria. Before inviting the eligible subjects, persons with a moderate or bad self-reported health, the inability to climb two flights of stairs, a body weight of 140 kg or more, current or past renal cancer, melanoma or breast cancer and lung cancer diagnosed less than 5 years ago or still under treatment were excluded.[14]

Ultimately, 15,822 individuals were randomised (1:1) to screening (n=7915) with low-dose CT at baseline (first round), 1 year later (second round) and 3 years later (third round) or no screening (n=7909). The main purpose of the trial is to determine whether CT screening will have reduced mortality from lung cancer by at least 25% at 10 years of follow-up.[14,15]

For this study, all 7915 participants randomised to the screening arm were included. Complete data on interval cancers were not yet available and, consequently, no analyses of screening sensitivity were performed.

Equipment and execution of screening examinations

A detailed description of the equipment and the execution of the screening examinations have previously been published.[11] In short, in each of the four screening sites, 16-detector CT scanners were used in a low-dose setting, without the administration of intravenous contrast media.[11] Datasets were derived from images of the thorax with a slice thickness of 1 mm and a slice interval of 0.7-mm.[11] CT images were analysed using software for semi-automated volume measurements (LungCARE; Siemens AG, Erlangen, Germany).[16–18] In cases where the software was not able to measure nodule volume (e.g. in pleural based or nonsolid nodules), the diameter of the nodule was measured manually by the radiologist.

Nodule management protocol

The management protocol of the NELSON trial has been published previously.[9,11,19] Briefly, screening could lead to three different outcomes: I) a negative screen result (no other action than an invitation for the next screening round); II) an indeterminate result (invitation for a follow-up scan); III) a positive result (referral to a pulmonologist for diagnostic work-up).

For newly detected solid nodules and the solid component of part-solid nodules, the volume determined the screening result as follows: <50 mm^3 was negative, 50–500 mm^3 was indeterminate and >500 mm^3 was positive.

For previously detected and nonsolid nodules, the percentage volume change was calculated: <25% was a negative result and ≥25% led to the assessment of the VDT. The VDT in days was calculated using the following formula:

$$VDT=(\ln 2 \times \Delta t)/(\ln(V2/V1))$$

where V1 represents nodule volume on the first examination and V2 the volume the second examination and Δt the time between the examinations in days.[11] In case the software was not able to measure nodule volume, manually measured diameters were used to calculate VDT in days using the following formula:

$$VDT=(\ln 2 \times \Delta t)/(\ln((MaxDiamXY2 \times PerpDiamXY2 \times MaxDiamZ2)/MaxDiamXY1 \times PerpDiamXY1 \times MaxDiamZ1))$$

where MaxDiamXY is the maximum diameter in the x/y-axis, PerpDiamXY the maximum diameter perpendicular to MaxDiamXY and MaxDiamZ is the maximum diameter in z-axis.[11]

For nodules with VDTs of 400–600 days, the result was indeterminate; for VDTs of <400 days the result was positive. From the second round onwards, participants with a nodule with a VDT of 400–600 days were invited for a 12-month repeat scan.[19] Furthermore, the screening was also positive if a new solid component had emerged in a previously nonsolid nodule. The screening result was negative for all nodules with fat, benign calcification patterns or other benign abnormalities.[11,19]

Referral, diagnostic work-up and diagnoses

After a positive screening, participants were referred for diagnostic work-up via their general practitioner and received usual care according to national and international guidelines.[4,20–23] All data were prospectively collected and histological specimens were reassessed by our chief pathologist (ET).

Definitions and statistics

Screen-detected lung cancers are the lung cancers that are diagnosed by the diagnostic work-up initiated for a positive screening. The lung cancer detection rate is the number of screen-detected lung cancers divided by the number of screened participants. A true-positive test result is a positive scan in a participant who actually has lung cancer. A false-positive test result is a positive scan, when lung cancer is not diagnosed.

The normality of the distribution of the continuous variables (age and pack-years) was evaluated by studying the Q-Q plots. As the variables were not normally distributed, the variables were described by the median and interquartile range. For analysing the difference between the continuous variables across the three screening rounds, the Kruskal–Wallis H test was used. For analysing the difference between the nominal variables (sex and smoking status) across the three screening rounds, the likelihood ratio-based

Chi-squared test was used. To calculate 95% confidence intervals of proportions, bootstrapping was performed based on 1000 samples. For all analysis, $\alpha<0.05$ was considered significant and PASW Statistics, SPSS version 20 (SPSS Inc., Chicago, IL, USA) was used.

Ethics and legal approval

The NELSON trial was approved by the Dutch Minister of Health and the ethics board at each participating centre. The NELSON trial is registered at www.trialregister.nl (number ISRCTN63545820). All participants gave written informed consent for participation and the evaluation of personal data from hospital charts and national registries.

RESULTS

7,582 (95.8%) of the 7,915 participants randomised to the screen-arm of the trial were actually screened. The participation rates remained high across the three screening rounds: 7,557 (95.5%) in round one, 7,295 (92.2%) in round two and 6,922 (87.5%) in round three.

In three screening rounds, 24,354 CT scans were made. 21,773 (89.4%) of the scans were a regular "round scans" and 2,581 (10.6%) were follow-up scans, performed to assess the VDT of indeterminately sized nodules. The scans detected a total of 31 683 nodules: 266 (0.8%) were part-solid and 298 (0.9%) nonsolid.

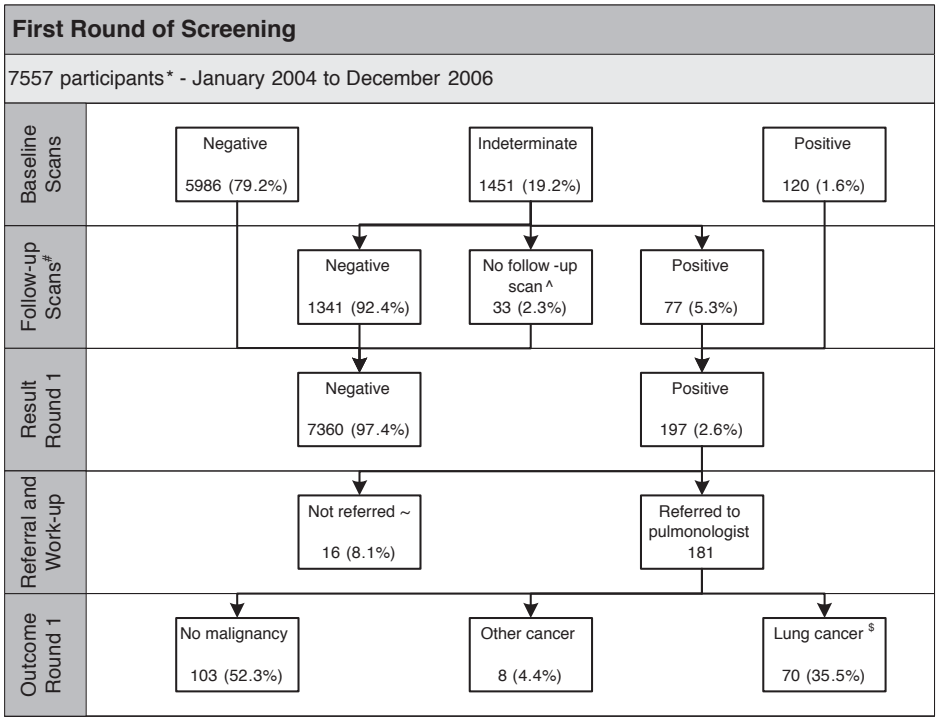
The screening result was negative in 87.2% of all scans (21,232 out of 24,354). The result was indeterminate in 10.8% (2,629 out of 24,354) and positive in 2.0% (493/24,354) of the scans. In the first round, the proportion of indeterminate and positive scan results was relatively higher than in later rounds. A detailed overview of the scan results per screening round is presented in figures 1-3.

The 493 positive screen results led to the diagnosis of lung cancer in 200 participants. 14 (7.0%) of these 200 participants were referred for a part-solid nodule and eight participants (4.0%) for a nonsolid nodule. 40.6% (200 out of 493) of all positive screenings were "true-positive" (95% CI 36.1-45.2%). The positive predictive value slightly increased across the three rounds, from 35.5% (95% CI 28.4-42.1%) in round one to 42.0% in round two (95% CI 34.4-49.6%) to 45.5% (95% CI 37.6-53.3%) in round three.

The cumulative lung cancer detection rate of the three rounds was 200 (2.6%) out of 7582 (95% CI 2.3-3.0%). This detection rate was relatively stable across the three screening rounds: 0.9% (75 of 6,922, 95% CI 0.7-1.2%) in round one, 0.8% (55 of 7,295, 95% CI 0.6-1.0%) in round two and 1.1% (75 of 6,922, 95% CI 0.8-1.3%) in round three.

The 493 positive screen results did not lead to a lung cancer diagnosis in the remaining 293 cases. Hence, 59.4% (293 of 493, 95% CI 54.8-63.9%) of the positive screen results

Figure 1. Results of the first round of screening



* *7,915 participants were randomised to the screen-arm of the trial and invited for screening; 25 (0.3%) participants missed screening in the first round, but were screened in the second round and 333 (4,2%) participants did not respond to the invitation.*

\# *Follow-up scans were performed after 99.6 days (mean; SD 18.3). In 8.3% of the subjects with an indeterminate result the nodule(s) had disappeared.*

^ *Reasons: administrative error (n=15), no show (n=13), refusal (n=3), already receiving treatment from other specialist (n=2).*

~ *Reasons: decision by tumour board (n=10), administrative error (n=3), already receiving treatment from other specialist (n=3).*

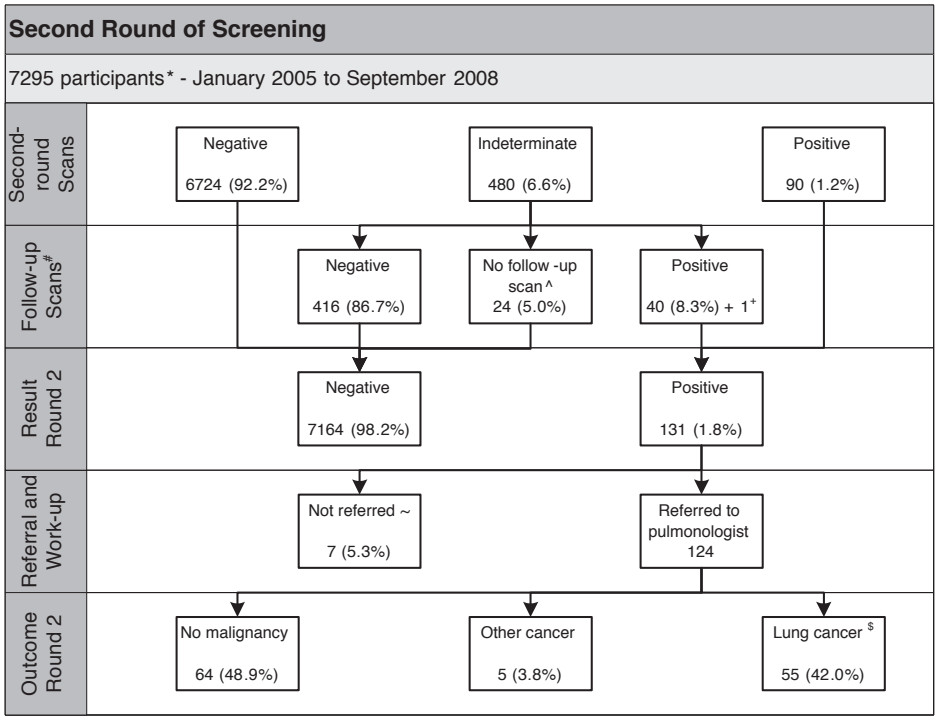
$ *67 of 70 (95.7%) lung cancer diagnoses were confirmed by cytology or histology. Details concerning the basis of the diagnosed the three other cases can be found in the Appendix.*

were actually "false-positive". Overall, 1.2% (293 of 24,354) of the scans performed in three rounds of the NELSON trial had a false-positive result.

The ratio of the overall true-positive and false-positive results (the true-positive/false-positive ratio) was 0.69. The true-positive/false-positive ratio tended to improve over time, from 0.69 in round one to 0.72 in round two, and to 0.83 in round three.

To detect lung cancer in 200 participants, 7,582 individuals underwent three rounds of screening. In the first screening round, 108 (7,557/70) participants were screened to detect one lung cancer. In the second round, 133 (7,295/55) and in the third round 92

Figure 2. Results of the second round of screening



* *287 participants did not undergo a second-round scan (7,557 participants of the first round plus 25 participants who missed screening in round 1, minus 7,295) because of: lung cancer (n=68: two subjects diagnosed with lung cancer did receive a second round scan because of an administrative error), death (n=27), participant declined (n=115), participant unattainable or repeatedly no show (n=47), still in diagnostic work-up round one (n=1), administrative error (n=1), no screening in second round, but screened in third round (n=28).*

\+ *Cave, 1 participant missed the second round scan (therefore only 7,294 second-round scans were performed and only 480 scans were indeterminate), but he received a follow-up scan instead later on, which had a positive result.*

\# *Follow-up scans were performed after 76.5 days (mean; SD=35.4). In 15.5% of the subjects with an indeterminate result the nodule(s) had disappeared.*

^ *Reasons: administrative error (n=12), no show (n=6), already receiving treatment from other specialist (n=5), death (n=1).*

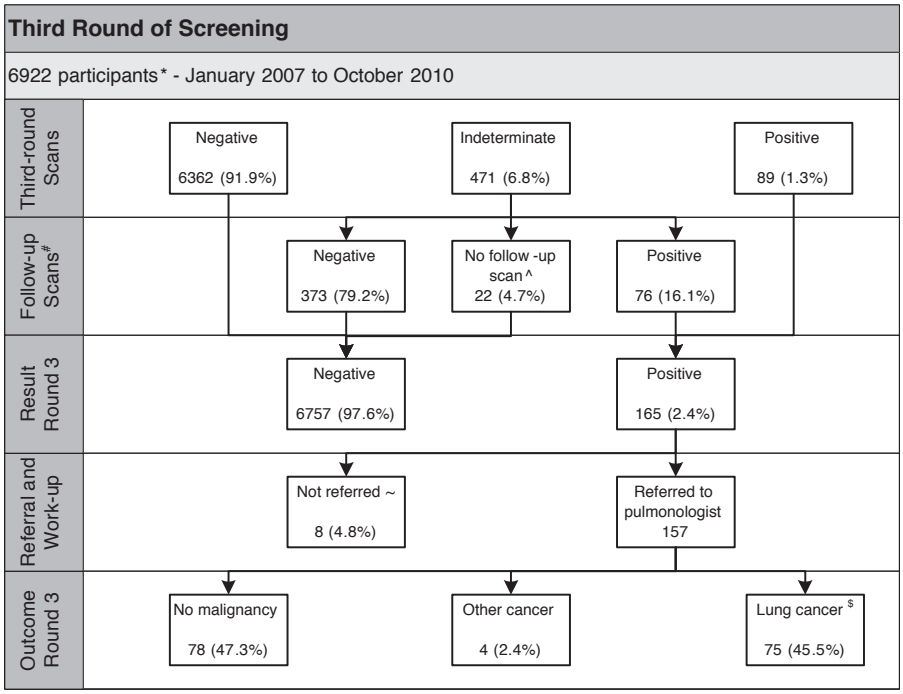
~ *Reasons: administrative error (n=2), already receiving treatment from other specialist (n=5).*

$ *52 of 55 (94.5%) lung cancer diagnoses were confirmed by cytology or histology. Details concerning the basis of the diagnosed the three other cases can be found in the Appendix.*

Cave: mortality data were available only for the Dutch participants until August 14th 2011.

(6,922/75) subjects were screened for the detection of one lung cancer. Cumulatively, to detect one lung cancer 38 participants underwent three screening rounds.

Figure 3. Results of the third round of screening



* *400 participants were not screened in the third round (7,294 participants of the second round plus 28 participants who missed screening in round 2 minus 6,922) because of: lung cancer (n=57), death (n=84), participant declined (n=155), participant unattainable or repeatedly no show (n=98), administrative error (n=3), unknown (n=3).*

\# *Follow-up scans were performed after 60.3 days (mean; SD=61.9). In 13.6% of the subjects with an indeterminate result the nodule(s) had disappeared.*

^ *Reasons: administrative error (n=8), no show (n=6), refusal (n=3), already receiving treatment from other specialist (n=5).*

~ *Reasons: decision tumour board (n=3), refusal (n=1) already receiving treatment from other specialist (n=4)*

$ *68 of 75 (90.7%) lung cancer diagnoses were confirmed by cytology or histology. Details concerning the basis of the diagnosed the seven other cases can be found in the Appendix.*

Cave: mortality data were available only for the Dutch participants until August 14th 2011.

False-positive screenings

6% (458 out of 7,582) of the participants had at least one positive screening result. 31 subjects had two positive screening results and two subjects had three positive screens. As 200 individuals were diagnosed with lung cancer, this implies that the remaining 258 participants had one or more false-positive screening result (244 subjects had one, 12 subjects had two and two subjects had three false-positive results). However, even 15 participants who were diagnosed with lung cancer had a false-positive screening in an

earlier round. Thus, 3.6% of all participants (273 out of 7,582) had a false-positive screening result.

67 (24.5%) out of the 273 participants with one or more false-positive screen result underwent an invasive procedure in the diagnostic work-up. 61 (91.0%) of these invasive procedures were surgeries (three mediastinoscopies, one sternotomy, nine video-assisted thoracoscopies and 48 thoracotomies) and the remaining six procedures were transthoracic biopsies (more details are supplied in the Appendix). Hence, 0.9% (67 out of 7,582) of all screened participants underwent an, in retrospect, "unnecessary" invasive diagnostic procedure.

5.5-year risk calculations

In this part of the study, we present an overview of subsequent screening results and lung cancer diagnoses to visualise the longitudinal character of the 5.5-year risk calculations (Figures 1a-d in Appendix). 70.4% of the screened participants (5,340 out of 7,582) had exclusively negative screen results.

The individuals with a negative first screening had a probability of 86.5% to receive exclusively negative screening results in 5.5 years (Figure 1a in Appendix). Furthermore, their risk of a false-positive screen result in the following 5.5 years was 1.3% (80 out of 5,986 participants) and their 5.5-year risk of lung cancer was only 1.0% (60 out of 5,986 participants).

The participants with an indeterminate result from their first screening had a probability of 72.1% to have exclusively negative screening results in the 5.5 years after the first screening (Figure 1b in Appendix). Their risk of a false-positive follow-up scan in the first screen round was 4.3% (62 out of 1,451). The risk of one or more false-positive scans in round two or three in this subgroup was 4.8% (70 out of 1,451). To summarise, after an indeterminate baseline scan result, the risk of one or more false-positive scan results in 5.5 years was 8.8% (128 out of 1,451). The risk of screen-detected lung cancer after an indeterminate baseline scan was 1.0% (15 out of 1,451) in round one and 4.6% (67 out of 1,451) in rounds two and three. Hence, the 5.5-year lung cancer risk after an indeterminate baseline scan result was 5.7% (82 out of 1,451).

The participants with a positive first screen result had a probability of 30.0% (36 out of 120) to have only negative screening results in the following 5.5 years (Figure 1c in Appendix). Their risk of a false-positive screening was 54.2% (65 out of 120) in the first round and 4.2% (five out of 120) in the second or third round. Furthermore, their risk to be diagnosed with screen-detected lung cancer within 5.5 years was 48.3% (58 out of 120). This was 45.8% (55 out of 120) directly in round one and 2.5% (three out of 120) in rounds two and three. The three individuals with a lung cancer diagnosis in rounds two or three were, in retrospection, referred twice for the same suspicious nodule.

The risk calculations show that the result of the baseline scan divides the screened population in three subgroups with distinct risks of lung cancer. The characteristics of the screened participants and the three subgroups are presented in table 1. When comparing participants with a negative, indeterminate and a positive baseline scan result, a statistically significant increase in age and number of pack-years was observed. However, there was no significant difference in the proportion of females and current smokers (Table 1).

Table 1. Participants' characteristics and comparison stratified by baseline scan result

Characteristics	All screened	Baseline scan result			
	participants n (%)	negative n (%)	indeterminate n (%)	positive n (%)	p- value
Females	1,254 (16.5)	1,016 (17.0)	210 (14.5)	22 (18.3)	0.06
Age - median (IQR)	58.0 (8)	57.0 (8)	59.0 (8)	59.0 (8)	<0.001
Current smoker	4,215 (55.6)	3315 (55.4)	809 (55.8)	809 (55.8)	0.94
Pack-years - median (IQR)	37.8 (19.8)	38.0 (19.8)	38.7 (19.8)	38.7 (19.8)	<0.001
Total	**7,582 (100.0)**	**5,986 (100.0)**	**1,451 (100.0)**	**120 (100.0)**	**NA**

Data are presented as n or n (%), unless otherwise stated.
Definition of abbreviations: IQR = interquartile range; NA = not applicable.

DISCUSSION

In this study, we evaluated the performance of the NELSON screening strategy in the first three screening rounds and we assessed the 5.5-year risk of false-positive screenings and screen-detected lung cancer.

If we compare the performance of the NELSON screening strategy with other LDCT screening trials we find notable differences. The percentage of positive scans in our trial (2.0%) was the same as in a Danish trial,[10,13] but substantially lower than in the NLST (24.2%).[5] Also, the percentage of participants with one or more positive scan was 6.0% in our trial, which is low compared with the 39.1% in the NLST (the percentage in DLCST was not published).[5]

Despite the lower percentage positive screenings, our strategy detected 200 lung cancers in the three screening rounds. As a result, the cumulative lung cancer detection rate (2.6%) was a little higher than in the NLST (2.4%: 649 out of 26,309), but lower than in the DLCST (3.4%: 69 out of 2,047).[5,10] The latter is probably due to the two additional screening rounds that have been completed in the DLCST.

The predictive value of a positive screen result was higher in the NELSON trial (40.6%) than in both the DLCST (34.8%) and the NLST (3.6%).[5,10,13] Hence, the percentage of false-positive results was 59.4% in the NELSON trial, 65.2% in the DLCST and 96.4% in

the NLST. The proportion of false-positive scans out of all scans is 1.2% in the NELSON trial, 1.3% in the DLCST and 23.3% in the NLST.[5,10,13]

In the NELSON trial, we observed that the ratio between the true-positive and false-positive results improved over the rounds (0.69, 0.72 and 0.83 respectively in rounds one, two and three). This is probably the result of the possibility in later rounds to compare current with previous images and to calculate VDTs. In the NLST, the true-positive/false-positive ratios were respectively 0.039 in round one, 0.025 in round two and 0.055 in round three (figures in the DLCST were not published).[5] The improvement in the third round probably results from the fact that only in the third round were stable nodules ≥4 mm in diameter not classified as positive.

Finally, the number needed to screen for the detection of one lung cancer was 92-133 per round in the NELSON trial, which is a little less than in the other trials (97-147 in the NLST and 116-180 in the DLCST).[5,10]

In the three screening rounds, 3.6% of all participants had a false-positive screening result and this led to invasive diagnostic procedures in 0.9% of all participants. Although we are convinced of the need to reduce these numbers, we realise that these "unnecessary" invasive procedures cannot be eliminated because it is sometimes the only way to distinguish lung cancer from other malignancies or benign conditions.

In the second part of this study, we found that participants with a negative, indeterminate or positive baseline scan had very distinct risks of positive screening results and lung cancer. Hence, the risk of a false-positive screening result in the next 5.5 years was respectively 1.3%, 8.8% and 54.2% for the individuals with a negative, indeterminate or positive baseline scan. Moreover, the 5.5-year risk of screen-detected lung cancer was only 1.0% for the individuals with a negative baseline scan result, 5.7% for subjects with an indeterminate baseline result and 48.3% for those with a positive baseline. In other words, after the first screening, the individual's lung cancer risk has either decreased by 62% or increased by 219% or 1858%.

Analyses showed a significant increase in age and number of pack-years when comparing participants with a negative, indeterminate and positive baseline scan, which are all well known risk factors for developing lung cancer.[24]

The presented results could aid clinicians when counselling high-risk subjects who are considering or have already undergone LDCT screening for lung cancer. This study has created the opportunity to personalise counselling and enables the individual at risk to make an informed choice. Moreover, this is the first study that quantifies both the potential benefit of screening (early detection) and a potential harm of screening (false-positive screening results).

The main strengths of this trial are its design (a large, randomised controlled trial), the population-based recruitment and prospective data collection.[14,25] Limitations of the current study are the lack of data on false-negative screenings, the control arm of the trial

and lung cancer mortality. These analyses were not performed because the required data was not yet available.[14]

Future research should focus on confirming the efficacy of LDCT screening for reducing lung cancer mortality. The planned lung cancer mortality analyses of the NELSON trial will be crucial in this part, as our trial is the only other trial (besides the NLST) that is sufficiently powered. Furthermore, efforts should be made to reduce false-positive screen results by optimising the cut-off criteria for nodule volume and VDT.

CONCLUSION

In this study, we evaluated the performance of the NELSON screening strategy in the first three screening rounds. We demonstrated that our strategy yields a low percentage of positive and false-positive scans with a reasonable positive predictive value. Furthermore, we used our experience with lung cancer screening to provide an overview of the 5.5-year risks of lung cancer and false-positive screenings, which aids clinicians in counselling individuals who are considering or have already undergone LDCT screening for lung cancer.

TWITTER

@ERSpublications
5.5-year lung cancer risk calculations aid clinicians in counselling for lung cancer screening with low-dose CT http://ow.ly/p9J3q

SUPPORT STATEMENT

The NELSON trial is supported by: "Zorg Onderzoek Nederland-Medische Wetenschappen" (ZonMw) (grant number 120610015), "KWF Kankerbestrijding" (grant number EMCR 2007-3857) and "Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen" (RvvZ).

Roche Diagnostics provided a grant for the performance of proteomics-research. Siemens Germany provided four digital workstations and LungCARE for the performance of the volumentric measurements of lung nodules.

CONFLICTS OF INTEREST

None declared

ACKNOWLEDGEMENTS

We thank our secretary M. Quak, our data manager R.M. Vernhout and our system controllers R. Faber and F.J.P. Santegoets (all at the Erasmus Medical Center, Rotterdam, the Netherlands) for their contributions and maintenance of the database.

REFERENCES

1. American Lung Association. Guidance on CT Lung Cancer Screening. http://www.lung.org/lung-disease/lungcancer/lung-cancer-screening-guidelines/lung-cancer-screening.pdf Date last updated: April 23, 2012. Date last accessed: October 3, 2013.
2. American Society of Clinical Oncology. The role of CT screening for Lung Cancer in clinical practice. The evidence based practice guideline of the American College of Chest Physicians and the American Society for Clinical Oncology. http://www.asco.org/institute-quality/role-ct-screening-lung-cancer-clinical-practice-evidence-basedpractice-guideline Date last accessed: October 3, 2013.
3. American College of Chest Physicians. ACCP and ASCO Release Joint Systematic Review and Clinical Practice Guideline on the Role of CT Screening for Lung Cancer. www.chestnet.org/News/ACCP-Clinical-News/2012/05/ACCP-and-ASCO-Joint-LC-Screening Date last updated: May 21, 2012. Date last accessed: December 5 2012.
4. Wender R, Fontham ET, Barrera E Jr, et al. American Cancer Society lung cancer screening guidelines. *CA Cancer J Clin* 2013; 63:107-117.
5. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365:395-409.
6. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA* 2012; 307:2418-2429.
7. Linet MS, Slovis TL, Miller DL, et al. Cancer risks associated with external radiation from diagnostic imaging procedures. *CA Cancer J Clin* 2012; 62(2):75-100.
8. van den Bergh KA, Essink-Bot ML, Borsboom GJ, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *Br J Cancer* 2010; 102:27-34.
9. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361:2221-2229.
10. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; 67:296-301.
11. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54:177-184.
12. National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. *Radiology* 2011; 258:243-253.
13. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish randomized lung cancer CT screening trial - overall design and results of the prevalence round. *J Thorac Oncol* 2009; 4:608–614.
14. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120:868-874.
15. Horeweg N, van Klaveren RJ, Groen HJ, et al. Blinded and uniform cause of death verification in a lung cancer CT screening trial. *Lung Cancer* 2012; 77:522-525.
16. Wormanns D, Kohl G, Klotz E, et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. *Eur Radiol* 2004; 14:86-92.
17. Gietema HA, Wang Y, Xu D, et al. Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. *Radiology* 2006; 241:251-257.

18. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. *Radiology* 2008; 248:625-631.
19. Ru Zhao Y, Xie X, de Koning HJ, et al. NELSON lung cancer screening study. *Cancer Imaging* 2011; 11:S79-S84.
20. van Meerbeeck JP, Koning CC, Tjan-Heijnen VC, et al. Richtlijn "Niet-kleincellig longcarcinoom; stadiering en behandeling", Guideline on "non-small cell lung carcinoma; staging and treatment". *Ned Tijdschr Geneeskd* 2005; 149:72-77.
21. Goldstraw P, Crowley J, Chansky K, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol* 2007; 2:706-714.
22. Travis WD, Brambilla E, Müller-Hermelink HK, et al., eds. *Pathology and Genetics*. Tumours of the Lung, Pleura and Heart. Lyon, IARC Press, 2004.
23. van 't Westeinde SC, de Koning HJ, Thunnissen FB, et al. The role of the 18F-fluorodeoxyglucose-positron emission tomography scan in the Nederlands Leuvens Longkanker screenings Onderzoek lung cancer screening trial. *J Thorac Oncol* 2011; 6:1704-1712.
24. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *Br Med J* 1956; 2:1071-1081.
25. van der Aalst CM, van Iersel CA, van Klaveren RJ, et al. Generalisability of the results of the Dutch–Belgian randomised controlled lung cancer CT screening trial (NELSON): does self-selection play a role? *Lung Cancer* 2012; 77:51-57.

APPENDIX

Lung cancer diagnoses not proven by histology

Lung cancer diagnoses in the first three rounds of the NELSON trial were based on histology or cytology in 187 of 200 (93.5%). The basis for the diagnosis in the 13 participants without histology or cytology is:

Round one:

1) Tumour in the right upper lobe, volume 1502mm^3, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to cardiac impairment.
2) Tumour in left lower lobe, volume 2687mm^3, and PET positive, and cT1aN0M0, patient did not undergo thoracic surgery due to COPD stage IV.
3) Tumour in left lower lobe, volume 2792mm^3, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to COPD and renal failure.

Round two:

1) Tumour in right upper lobe, volume 580mm^3, PET-positive, cT1aN0M0, the patient did not undergo thoracic surgery due to metastasized prostate carcinoma.
2) Tumour in right lower lobe, volume 2793mm^3, PET-positive, cT1bN0M0, the patient did not undergo thoracic surgery due to poor pulmonary function.
3) Tumour in right upper lobe, volume 891mm^3, PET indeterminate, cT1aN0M0, the patient died just before intended thoracic surgery due to bowel ischemia.

Round three:

1) Tumour in right lower lobe, volume 731mm^3, PET-positive, cT1aN0M0, the patient did not undergo thoracic surgery due to poor pulmonary function.
2) Tumour in left lower lobe, volume 108mm^3, VDT 125 days, PET-positive, cT1aN0M0, the patient did not undergo thoracic surgery because he also participated in another study and was randomised to the radiotherapy treatment arm.
3) Tumour in right upper lobe, volume 383mm^3, VDT 289 days, PET indeterminate, cT1aN0M0, the patient did not undergo thoracic surgery because he refused, he was treated with stereotactic radiotherapy instead.
4) Tumour in left lower lobe, volume 1108mm^3, PET positive, cT1aN1M0, the patient did not undergo thoracic surgery due to poor pulmonary function.
5) Tumour in left lower lobe, diameter 10mm, PET positive, cT1aN0M0, and the patient did not undergo thoracic surgery due to poor pulmonary function.
6) Tumour in right upper lobe, diameter 13.2x11.6mm, PET positive, cT1aN0M0, the patient did not undergo thoracic surgery due to poor pulmonary function and general condition

7) Tumour in right upper lobe, diameter 19.2x12.7mm, PET positive, cT1bN0M0, the patient did not undergo thoracic surgery due to poor general condition

Figure 1a. Overview of subsequent screening test results and lung cancer diagnoses: participants with a negative baseline scan

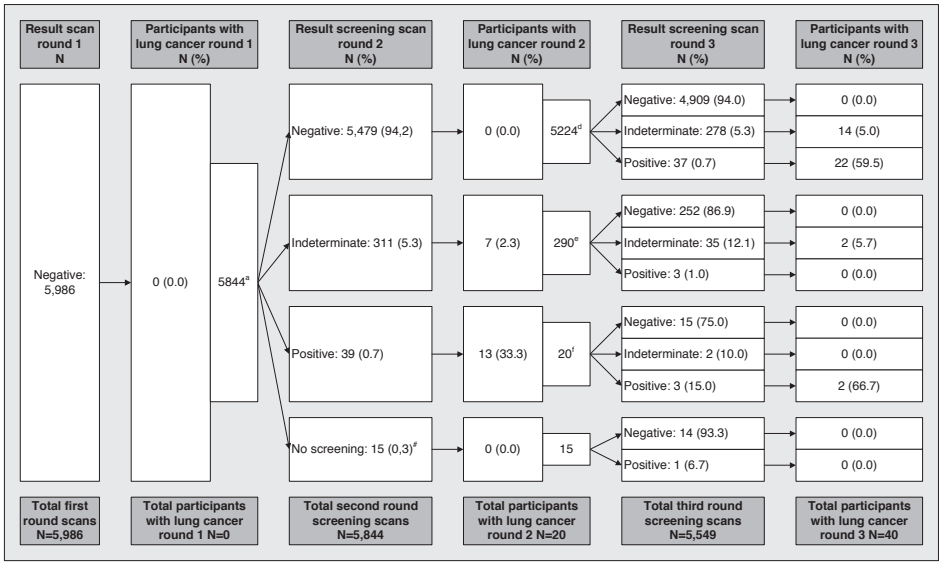


Figure 1b. Overview of subsequent screening test results and lung cancer diagnoses: participants with an indeterminate baseline scan

Figure 1c. Overview of subsequent screening test results and lung cancer diagnoses: participants with a positive baseline scan

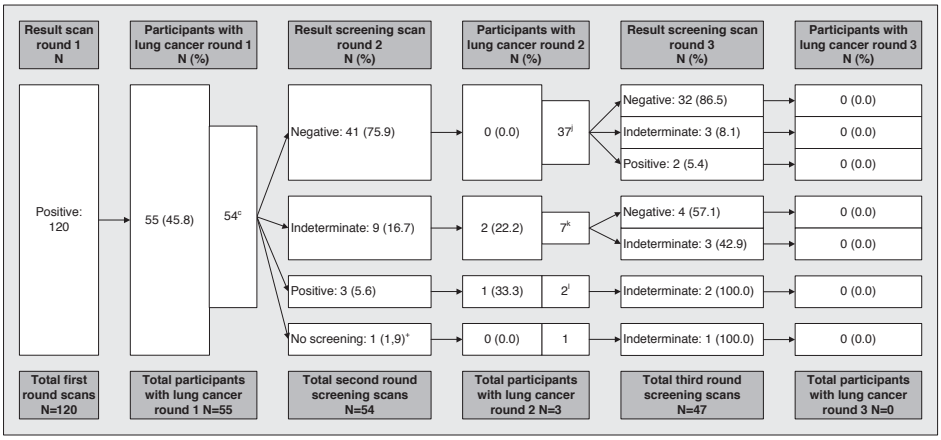
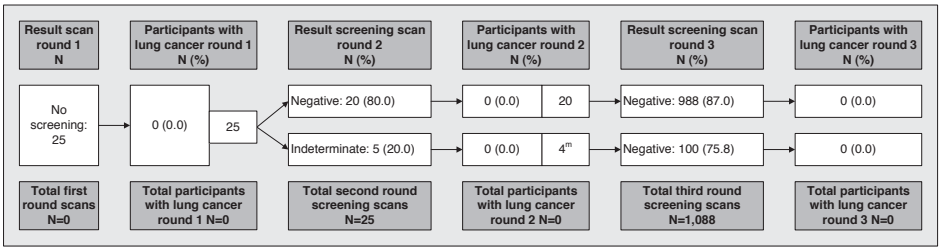


Figure 1d. Overview of subsequent screening test results and lung cancer diagnoses: participants with no baseline scan



* *25 participants had no screening scan in the first round, because there was a delay in the returning of the informed consent.*

\# *15 participants had no screening scan in the second round because: participant declined (n=1), participant unattainable or repeatedly no show (n=14).*

¶ *12 participants had no screening scan in the second round because: still in diagnostic work-up round one (n=2), participant declined (n=2), participant unattainable or repeatedly no show (n=4), administrative error (n=4).*

\+ *1 participant had no screening scan in the second round because: still in diagnostic work-up round one (n=1)*

a *142 participants (5,986 minus 5,829) were not screened in the second and third round. Reasons: death (n=19), participant declined (n=81), participant unattainable or repeatedly no show (n=41), administrative error (n=1).*

b *52 participants (1,451 minus 1,399) were not screened in the second and third round. Reasons: lung cancer (n=13), death (n=7) still in diagnostic work-up round one (n=1), participant declined (n=25), participant unattainable or repeatedly no show (n=6). Cave: the other two subjects with screen-detected lung cancer did receive a second round scan because of an administrative error.*

c *66 participants (120 minus 54) were not screened in the second and third round. Reasons: screen-detected lung cancer (n=55), death (n=1), participant declined (n=10).*

d *255 participants (5,479 minus 5,224) were not screened in the third round. Reasons: death (n=56), participant declined (n=112), participant unattainable or repeatedly no show (n=82), administrative error (n=2), unknown (n=3).*

e *21 participants (311 minus 290) were not screened in the third round. Reasons: lung cancer (n=7), death (n=3), participant declined (n=9), participant unattainable or repeatedly no show (n=2).*

f *19 participants (39 minus 20) were not screened in the third round. Reasons: lung cancer (n=13), death (n=2), participant declined (n=4).*

g *49 participants (1,184 minus 1,135) were not screened in the third round. Reasons: death (n=16), participant declined (n=22), participant unattainable or repeatedly no show (n=10), administrative error (n=1).*

h *23 participants (155 minus 132) were not screened in the third round. Reasons: lung cancer (n=10), death (n=6), participant unattainable or repeatedly no show (n=2), participant declined (n=5). Cave, one participant was already diagnosed with lung cancer in de first round, but received a second round scan because of an administrative error.*

i *25 participants (48 minus 23) were not screened in the third round. Reasons: lung cancer (n=24), participant declined (n=1). Cave, one participant was already diagnosed with lung cancer in de first round, but received a second round scan because of an administrative error.*

j *4 participants (41 minus 37) were not screened in third round. Reasons: death (n=1), participant declined (n=2), participant unattainable or repeatedly no show (n=1).*

k *2 participants (9 minus 7) were not screened in the third round. Reasons: lung cancer (n=2).*

l *1 participant (3 minus 2) were not screened in the third round. Reasons: lung cancer (n=1).*

m *1 participant (5 minus 4) were not screened in the third round. Reasons: participant unattainable or repeatedly no show (n =1).*

Chapter 3

Characteristics of screen-detected lung cancer

Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial

Horeweg N,
van der Aalst CM,
Thunnissen E,
Nackaerts K,
Weenink C,
Groen HJM,
Lammers JWJ,
Aerts JG, Scholten ET,
van Rosmalen J,
Mali W, Oudkerk M,
de Koning HJ

American Journal of Respiratory and Critical Care Medicine
April 15th 2013

Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial

Am J Respir Crit Care Med. 2013 Apr 15;187(8):848-54.

ABSTRACT

The NELSON trial is, with 15,822 participants, the largest European lung cancer computer tomography screening trial. A volumetry-based screening strategy, stringent criteria for a positive screening, and an increasing length of screening interval are particular features of the NELSON trial.

To determine the effect of stringent referral criteria and increasing screening interval on the characteristics of screen-detected lung cancers, and to compare this across screening rounds, between sexes, and with other screening trials.

All NELSON participants with screen-detected lung cancer in the first three rounds were included. Lung cancer stage at diagnosis, histological subtype, and tumour localisation were compared between the screening rounds, the sexes, and with other screening trials.

In the first three screening rounds, 200 participants were diagnosed with 209 lung cancers. Of these lung cancers, 70.8% were diagnosed at stage I and 8.1% at stage IIIB–IV, and 51.2% were adenocarcinomas. There was no significant difference in cancer stage, histology, or tumour localisation across the screening rounds. Women were diagnosed at a significantly more favourable cancer stage than men. Compared with other trials, the screen-detected lung cancers of the NELSON trial were relatively more often diagnosed at stage I and less often at stage IIIB–IV.

Despite stringent criteria for a positive screening, an increasing length of screening interval, and few female participants, the screening strategy of the NELSON trial resulted in a favourable cancer stage distribution at diagnosis, which is essential for the effectiveness of our screening strategy.

INTRODUCTION

Lung cancer is the leading cause of cancer-related death in males and the second in females globally, accounting for 1.4 million lung cancer deaths per year.[1] Despite treatment advances, survival has not improved substantially, mainly because the majority of the patients have distant metastases at the time of diagnosis.[2] Several randomised lung cancer screening trials were conducted with low-dose computer tomography (LDCT) scanning of high-risk groups, aiming to detect lung cancer at an earlier and curable stage.[3–7]

The world's largest randomised CT screening trial, the National Lung Screening Trial (NLST), demonstrated in 2011 that early detection by LDCT scanning has yielded a 20% lung cancer mortality reduction compared with screening by chest radiograph.[8] Sixty-one percent of the LDCT-detected lung cancers were diagnosed at stage I. To accomplish this impressive result, considerable efforts were made. Namely, 26,722 high-risk subjects underwent annual LDCT screening for 3 years. Positive screening results were defined as any non-calcified pulmonary nodule measuring at least 4 mm in any diameter. In the three screening rounds, 39.1% of the individuals had at least one positive result.[8]

Our trial, the Dutch–Belgian Lung Cancer Screening Trial (Nederlands-Leuvens longkanker screeningsonderzoek; the NELSON trial), is the world's second-largest randomised CT screening trial and differs from the NLST by screening interval, referral policy, and a control arm wherein individuals receive no screening.[3] The 7,915 participants randomised to the screening arm of the NELSON trial underwent LDCT screening at baseline, 1 year later, 2 years later, and finally 2.5 years later. Positive screening results were defined as non-calcified nodules with a volume greater than 500 mm^3 (about 9.8 mm in diameter) or volume-doubling time (VDT) less than 400 days.[3,9,10] In the first three screening rounds, 6.0% of the participants had at least one positive screening. Clearly, the differences between the two largest randomised CT screening trials are substantial. Whether the NELSON trial will be able to demonstrate a significant lung cancer mortality reduction must be awaited, because the mortality analyses are planned 10 years after randomisation.[11] However, the characteristics of the screen-detected lung cancers, especially the stage distribution, might give an indication of the effectiveness of our screening strategy.

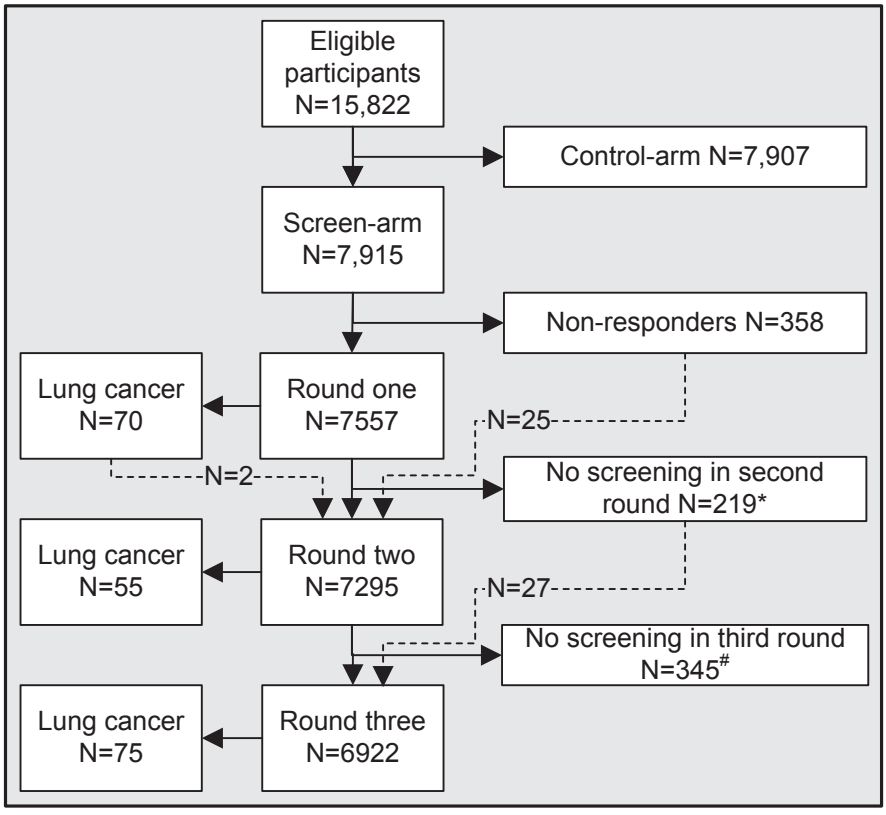
For this study, all participants with screen-detected lung cancer in the first three rounds of the NELSON trial were included. Lung cancer stage at diagnosis, histological subtype, and tumour localisation were compared between the screening rounds, the sexes, and several randomised CT screening trials.

METHODS

NELSON trial

The 15,822 individuals in the NELSON trial were randomised (1:1) to screening (n = 7,915) with LDCT at baseline (first round), 1 year later (second round), 3 years later (third round), and 5.5 years later (fourth round) or no screening (n = 7,907) (Figure 1). The main purpose of the trial is to determine whether LDCT screening will have reduced lung cancer mortality by at least 25% at 10 years of follow-up.[11,12] A more detailed report of the design and conduct was published previously.[9,11]

Figure 1. Participant flowchart in the first three rounds of the NELSON trial



* *219 participants were not screened in the second round because of death (n = 27), participant declined (n = 115), participant unattainable or repeatedly no show (n = 47), administrative error (n = 1), still in diagnostic work-up round 1 (n = 1), no screening in second round, but screened in third round (n = 27).*

\# *345 participants were not screened in the third round because of lung cancer in round 1 (n = 2), death (n = 84), participant declined (n = 155), participant unattainable or repeatedly no show (n = 98), administrative error (n = 3), unknown (n = 3).*

Note: mortality data were available only for the Dutch participants until August 14, 2011.

Participants

Individuals aged 50 to 75 years, who had smoked 15 or more cigarettes per day for 25 years or 10 or more cigarettes for 30 years and were still smoking or had quit less than 10 years ago, met the inclusion criteria. The exclusion criteria and calculation of expected lung cancer mortality were published in 2006.[11] For this study, all participants diagnosed with screen-detected lung cancer in the first three screening rounds were included (Figure 1). Hence, the interval cancers were not included in the analyses.

Equipment and Nodule Management Protocol

In short, 16-detector CT modality was used in a low-dose setting, without intravenous contrast medium.[9] CT images were analysed with software for semi-automated volume measurements (LungCARE, Siemens Healthcare, Erlangen, Germany).[13,14]

Briefly, the screening test 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). The nodule volume determined the screen result for newly detected nodules: less than 50 mm^3 was negative, 50 to 500 mm^3 was indeterminate, and more than 500 mm^3 was positive. The percentage volume change was calculated for previously detected nodules: at least 25% led to the assessment of the VDT. The VDT was calculated according to the formula: VDT(days) = [ln 2 x (time between current scan and baseline screening)]/[ln(nodule volume on current scan/volume on baseline scan)].[9] The screen result was positive for a VDT less than 400 days. A full description of the protocol was published previously.[9,10]

Referral and Diagnostic Work-Up

After a positive screening, the participants were referred for diagnostic work-up via their general practitioner and received usual care according to national and international guidelines.[15-19] All data were prospectively collected and histological specimens were reassessed by our chief pathologist (E.T.).

Statistical Analyses

Continuous variables were tested for normality using the Kolmogorov-Smirnov test for 50 or more samples, and using the Shapiro-Wilk test for fewer than 50 samples. Continuous, normally distributed variables were described by means and standard deviations. The difference between the means of continuous variables was calculated by one-way analysis of variance. Non-normally distributed variables were described by medians and interquartile ranges. The difference between nominal variables was calculated using the chi^2-test and differences between categorical variables were calculated using the Mann-Whitney U test. The difference between more than two samples of a categorical variable was calculated using the Kruskal-Wallis H test. Predictors of cancer stage were tested

using ordinal logistic regression; variables entered multivariate models when the P-value did not exceed 0.05 univariately. P-values less than 0.05 were treated as significant. SPSS Statistics version 20 (IBM, Armonk, NY) was used for all analyses.

Ethics and Legal Approval

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

RESULTS

Participants

Of the 7,915 (95.8%) participants randomised to the screening arm of the trial, 7,582 received at least one screening (Figure 1). Their baseline characteristics are presented in Table 1. The three screening rounds yielded 493 positive screen results and 200 (40.6%) participants were diagnosed with lung cancer. Synchronous double tumours were detected in four participants in round 1, in three participants in round 2, and in two participants in round 3. Thus, 200 participants were diagnosed with a total of 209 lung cancers. The patients with lung cancer were significantly older and had smoked significantly more pack-years than had the subjects not diagnosed with lung cancer (Table 1).

Table 1. Characteristics of the NELSON participants

Characteristics*	All participants n (%)	Participants diagnosed with lung cancer † n (%)	Participants not diagnosed with lung cancer n (%)	p-value ‡
Female gender, n(%)	1,254 (16.5)	34 (17,0)	1,220 (16.5)	0.86
Current smoker, n(%)	4,215 (55.6)	112 (56,0)	4,103 (55.6)	0.91
Pack-years, median (IQR)	38.0 (29.7 - 49.5)	43.7 (32.2 - 75.8)	38.0 (29.7 - 49.5)	< 0.001
BMI, median (IQR)	25.8 (23.9 - 28.1)	25.4 (23.3 - 28.0)	25.8 (23.9 - 28.1)	0.09
Total	**7,582 (100.0)**	**200 (100.0)**	**7,382 (100.0)**	**NA**

Definition of abbreviations: IQR= interquartile range; BMI= body mass index; NA = not applicable.

* *At randomisation.*

† *In the first three screening rounds of the NELSON trial.*

‡ *Comparison participants with versus without lung cancer.*

Lung Cancer Symptoms

Eleven of the 200 participants (5.5%) had symptoms suspicious of lung cancer before they were diagnosed. Five of them had symptoms before the screening scan was made;

however, none of them had symptoms at randomisation. Three subjects had symptoms in the period between the positive scan and the first consultation, and three subjects had symptoms in the period between the first consultation and the diagnosis date. Box plots of the time to screening result, referral, and diagnosis of the 200 participants and a detailed description of the symptoms can be found in the Appendix (Figure 1).

Lung Cancer Characteristics

More than half of the 209 screen-detected lung cancers were adenocarcinomas (51.2%) and a large majority was diagnosed at an early stage (stage I, 70.8%) (Table 2). Adenocarcinomas appeared to be diagnosed at a significantly lower cancer stage (univariate analysis p = 0.045), but in multivariate analysis this was no longer significant (p = 0.56) (Table 1 in Appendix). However, all bronchoalveolar carcinomas (n = 11) and carcinoids (n = 6) were diagnosed at stage Ia (Table 2). Four other histological subtypes were prone

Table 2. Histology and cancer stage of 209 screen-detected lung cancers in 200 participants

Cancer stage*	**Ia**	**Ib**	**IIa**	**IIb**	**IIIa**	**IIIb**	**IV**	**Overall**
Histology†	**n(%)**	**n(%)**	**n(%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Adenocarcinoma	75 (70.1)	9 (8.4)	8 (7.5)	.	9 (8.4)	4 (3.7)	2 (1.9)	107 (51.2)
Bronchoalveolar carcinoma	11 (100.0)	.	.	.	.	.	.	11 (5.3)
Squamous cell carcinoma	21 (61.8)	.	3 (8.8)	.	8 (23.5)	.	2 (5.9)	34 (16.3)
Adenosquamous carcinoma	.	.	.	.	.	.	.	4 (1.9)
Large cell carcinoma	7 (41.2)	1 (5.9)	.	.	6 (35.3)	2 (11.8)	1 (5.9)	17 (8.1)
Large cell neuro-endocrine carcinoma	2 (50.0)	1 (25.0)	.	.	1 (25.0)	.	.	4 (1.9)
Small cell carcinoma	.	.	.	.	5 (62.5)	.	3 (37.5)	8 (3.8)
Small/large cell carcinoma	.	.	.	.	1 (50.0)	.	1 (50.0)	2 (1.0)
Pleiomorph carcinoma	.	.	1 (100.0)	.	.	.	.	1 (0.5)
NSCLC-NOS	.	.	.	.	.	1 (50.0)	1 (50.0)	2 (1.0)
Carcinoid	6 (100.0)	.	.	.	.	.	.	6 (2.9)
No histological diagnosis‡	12 (92.3)	.	1 (7.7)	.	.	.	.	13 (6.2)
Total	**137 (65.6)**	**11 (5.3)**	**14 (6.7)**	**.**	**30 (14.4)**	**7 (3.3)**	**10 (4.8)**	**209 (100)**

Definition of abbreviations: NSCLC = non-small cell lung carcinoma; NOS = not otherwise specified.

* *7th edition of the IASLC TNM staging system (2009).*

† *According to Travis et al. Pathology and Genetics. Tumours of the Lung, Pleura and Heart. IARC Press (2004).*

. *= 0 (0.0).*

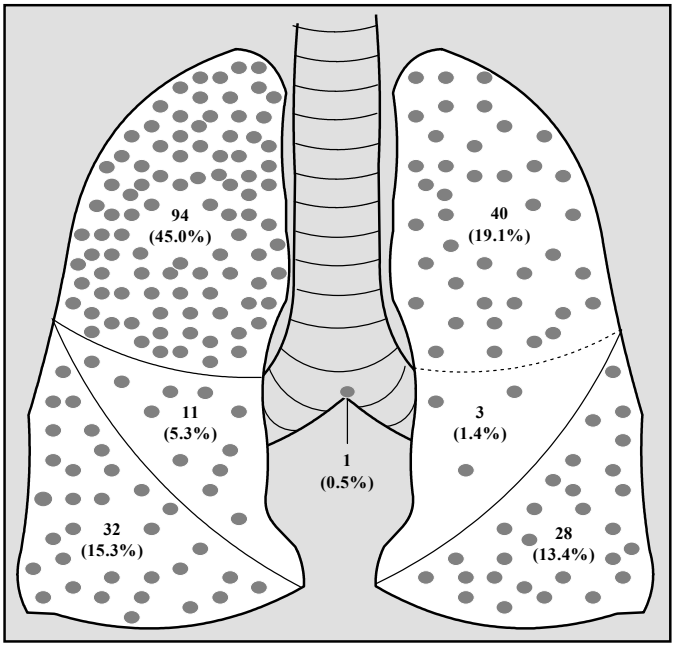
‡ *In 13 participants no histological diagnosis was established because biopsies were unsuccessful or not performed and the patient did not undergo thoracic surgery because of poor pulmonary function (n = 7), poor heart function (n = 1), poor general condition (n = 1), metastasised prostate carcinoma (n = 1), death due to mesenteric ischemia before intended surgery (n = 1), radiotherapy because of participation in other clinical trial (n = 1) and refusal (n = 1).*

In 10 lung resection specimens the pathologist found, besides the lung cancer, a focus of atypical adenomatous hyperplasia.

to be diagnosed at a higher cancer stage; for example, small cell carcinomas (multivariate analysis P < 0.001) (Table 1 in Appendix).

Most lung cancers were localised in the right lung (65.6%) and a large proportion (45.0% of all lung cancers) was localised in the right upper lobe (Figure 2). We also observed that the lung cancers were localised predominantly in the periphery of lungs (Figure 3). Of the nodules, 62.2% were found in the outer one-third of the costal-hilar diameter (Figure 3). In particular, adenocarcinomas were more often detected in the periphery and attached to the pleura than in the middle or central one-third of the lungs (82.2% vs. 17.8%; p = 0.001). But the reverse was not true for squamous cell carcinomas (62.9% peripheral or pleural-attached vs. 37.1% central or middle one-third; p = 0.16).

Figure 2. Localisation across the lobes of the 209 screen-detected lung cancers.

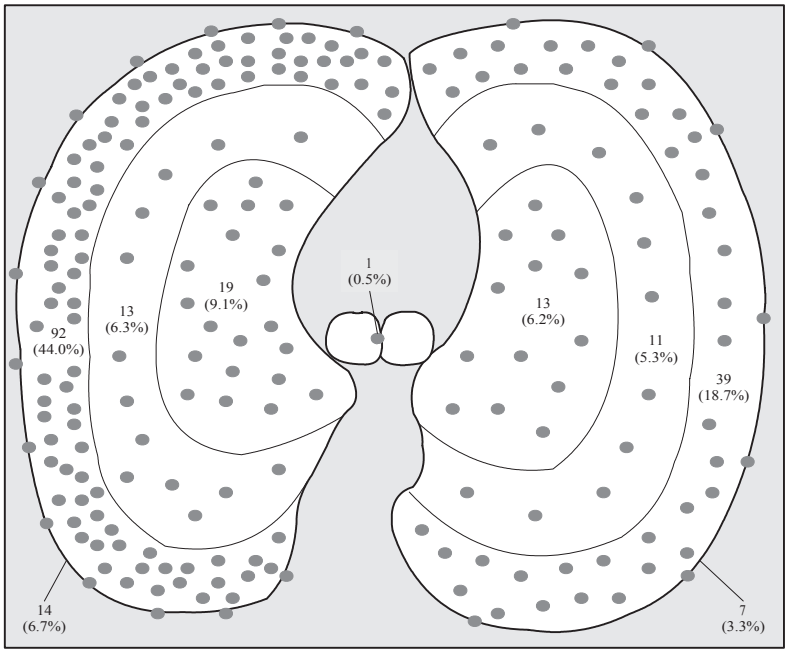


A schematic depiction of the lungs and large airways. The right lung is displayed on the left side and vice versa, as on a chest radiograph. The left upper lobe is divided in the pars superior and the lingula by the dotted line. The lung cancers are depicted as dark grey dots; their localisation corresponds with the lobe where the nodule was detected, not with the exact localisation.

Effect of screen round

The lung cancers detected in round 1 had a slightly higher disease stage (stage Ia 59.5%, stage IV 6.8%) than in later rounds (round 2: stage Ia 74.1%, stage IV 3.4%, and round 3:

Figure 3. Localisation of the 209 screen-detected lung cancers; central versus peripheral.



A schematic depiction of the lungs and main carina in the transversal plane. The right lung is displayed on the left side and vice versa as in computer tomography images.
The lungs are categorised in four sections: central (inner one-third of the hilar costal diameter), middle (middle one-third of the hilar-costal diameter), peripheral (outer one-third of the hilar-costal diameter) and pleural-attached nodules (nodules depicted on the bold outline of the lungs) The lung cancers are depicted as dark grey dots; their localisation corresponds with the section where the nodule was detected, not with the exact localisation. One lung cancer is not represented in the figure because the participant was referred for a fast-growing nodule peripheral in the right upper lobe, but this nodule disappeared during the work-up. However, soon thereafter lung cancer developed in the left lower lobe (transversal localisation unknown) and the patient was treated with radiotherapy.

stage Ia 64.9%, stage IV 3.9%) (Tables 2a-2c in Appendix). But this was not statistically significant between rounds 1 and 2 (p = 0.09) or across the three rounds (p = 0.23).

Also, the proportion adenocarcinomas was not significantly different between rounds 1 and 2 (47.3% vs. 60.3%; p = 0.14) or across the three rounds (round 3: 48.1% adenocarcinomas; p = 0.26) (Tables 2a-2c in Appendix).

Likewise, tumour localisation was not significantly different across the screen rounds: neither for the division over the lobes (p = 0.88) nor for the division over the peripheral versus central lung fields (p = 0.09).

Effect of sex

The women diagnosed with lung cancer were significantly younger (58.0 vs. 62.0 years; p = 0.03), had smoked less (pack-years: 36.0 vs. 43.0; p = 0.03) and had a lower BMI (23.8

vs. 25.9; p = 0.03) than the men diagnosed with lung cancer. The percentage current smokers however, was not lower in females (56.7 vs. 55.9%; p = 0.93).

None of the histological subtypes were unevenly distributed between the sexes (Tables 3a and 3b in Appendix). Also, the localisation of the lung cancers was not significantly different between the sexes: neither for the left lung versus right lung localisation (p = 0.92), nor for peripheral versus central localisation (p = 0.89). However, the cancer stage at diagnosis was significantly lower in women than in men (p = 0.005) (Tables 3a and 3b in Appendix). When correcting for the sex differences in age, number of pack-years and BMI, women still had a statistically significant lower cancer stage than men (p = 0.028) (Table 4 in Appendix).

Coincidentally, we found that a higher body mass index (BMI) (before randomisation) was a significant multivariate predictor (p = 0.004) of a more unfavourable cancer stage at diagnosis in both sexes (Table 4 in Appendix).

Comparison of trials

A total of 1,078 lung cancers were detected by CT screening in 43,983 participants of randomised screening trials (Table 3). On average, 64.7% of the lung cancers were diagnosed at stage I and 10.9% at stage IIIb–IV (Table 3). The stage distribution in the NELSON trial appears to be relatively favourable compared with the other trials. When we compare the

Table 3. Overview of cancer stage at diagnosis of screen-detected lung cancers in randomised CT screening trials

Trial	Participants screen arm n	Screening rounds n	Length screening interval (yrs)	Males – females (% – %)	No. of published CT-detected lung cancers	Stage Ia + Ib lung cancers n (%)	Stage IIIb + IV lung cancers n (%)
NLST[8]	26,722	3	1	59.0 – 41.0	649	400 (61.6)	130 (20.0)
NELSON	7,915	4	1, 2 and 2.5	83.5 – 16.5	209	148 (70.8)	17 (8.1)
DLST[36]	2,052	5	1	54.6 – 45.4	69	47 (68.1)*	11 (15.9)†
ITALUNG[7]	1,613	4	1	64.2 – 35.8	22	11 (50.0)‡	5 (22.7)
DANTE[37]	1,276	4	1	100.0 – 0.0	58	41 (70.7)	4 (6.9)
MILD[38]	1,190	10	1	68.4 – 31.6	29	18 (62.1)	4 (20.0)
	1,186	5	2	68.5 – 31.5	20	14 (70.0)	5 (17.2)
LUSI[39]	2,029	4	1	64.8 – 35.2	22	18 (81.8)	0 (0)
Total	**43,983**	**3 to 10**	**1 to 2.5**	**65.4 – 34.6**	**1078**	**697 (64.7)§**	**118 (10.9)***

Definition of abbreviations: CT = computed tomography; DLST = Danish Lung Cancer Screening Trial; MILD = Multicentric Italian Lung Detection; NELSON = Nederlands-Leuvens Longkanker ScreeningsOnderzoek (Dutch–Belgian Lung Cancer Screening Trial); NSLT = National Lung Screening Trial.

* *This not include two participants diagnosed with limited stage small cell lung carcinoma.*

† *This includes the participant diagnosed with extensive stage small cell lung carcinoma.*

‡ *This not include the three participants diagnosed with limited stage small cell lung carcinoma.*

§ *This does not include the four participants with limited stage small cell lung carcinoma.*

whole range of cancers stages between the two largest trials (NLST and NELSON) we observe that the cancer stage was significantly lower (p = 0.001) in the NELSON trial.

DISCUSSION

In this study, we have presented the characteristics of the lung cancers detected in the first three rounds of the NELSON trial. We investigated whether the screening strategy of the NELSON trial led to detection of lung cancer at a more favourable stage and how this relates to other randomised lung cancer CT screening trials.

In the three screening rounds, 493 participants had a positive screening result and were referred for diagnostic work-up. Ultimately, 200 (40.6%) participants were diagnosed with a total of 209 lung cancers. Eleven (5.5%) of these participants had symptomatic lung cancer before diagnosis; in five subjects the symptoms emerged before the screening scan was made.

More than half of the 209 screen-detected lung cancers were adenocarcinomas (51.2%) and a large majority was diagnosed at stage I (70.8%). Moreover, only 10 lung cancers were diagnosed at stage IV. This favourable stage distribution has created the opportunity for most patients to undergo curative surgery, which hopefully will reduce lung cancer mortality. However, screening detected only a few small cell lung cancers and all were diagnosed at stage III-IV. This finding could imply that LDCT screening is not, or is less, capable of early detection in some fast-growing histological subtypes of lung cancer. Further research should be conducted to investigate whether the lung cancers not detected by screening are predominantly of the same histological subtypes that are detected only in small amounts or high stages by screening.

Most screen-detected lung cancers were localised in the periphery of the lungs, which is probably a result of the large amount of adenocarcinomas that are significantly more often localised peripherally (p = 0.001). We also observed that 45.0% of all lung cancers were localised in the right upper lobe. This is a known phenomenon in patients with non-small cell lung cancer and could be explained by the fact that the airflow at the beginning of the breath is the largest toward the right upper lobe bronchus.[20,21] As a result, the deposition of particles in tobacco smoke and their carcinogenic effects are the largest in the right upper lobe.[22,23]

Further, our analyses showed no significant effect of screening round on cancer stage, histology, or tumour localisation. However, a decrease in advanced-stage lung cancers was observed at the second screening round (stage IV dropped from 6.8 to 3.4%). This was probably not statistically significant because of the low absolute number of advanced-stage lung cancers. In the third screening round, no evidential increase in stage IV lung cancers was observed (3.9%), despite the screening interval of 2 years.

The differences in lung cancer characteristics between men and women have been studied extensively. In general, studies demonstrated that women are diagnosed at an earlier age,[24,25] at a more favourable cancer stage,[25-27] and are more often diagnosed with adenocarcinomas than are men.[24,28,29] NELSON is the first trial to report on these differences in a screening setting. We also found that women were diagnosed at a significantly more favourable cancer stage than men ($p = 0.028$, after correction for confounding). However, the histological subtype and localisation of the lung cancers were not significantly different between the sexes.

In the NELSON trial, the body mass index (BMI) was not significantly higher in the participants diagnosed with lung cancer than in participants who were not diagnosed ($p = 0.09$). However, a higher BMI was a significant multivariate predictor of a more unfavourable cancer stage at diagnosis ($p = 0.004$). This finding is in line with one other study.[30] However, most studies demonstrated a negative association between BMI and lung cancer risk and prognosis.[31-33] This discrepancy could be explained in the first place by reversed causation: BMI is usually measured at diagnosis, at that time weight loss has often occurred, especially patients with a higher cancer stage. In the NELSON trial, BMI was measured just before randomisation and because none of the participants had symptomatic lung cancer at that time, the BMI was not influenced by lung cancer itself. In the second place, the discrepancy could be explained by the strong confounding effect of smoking in many trials: smokers have a lower mean BMI than non-smokers[34] and smoking is major risk factor for lung cancer mortality.[32] This bias is probably limited in the NELSON trial because we included only (ex-)smokers.[11]

In this article, we have presented an overview of the disease stage of the LDCT-detected lung cancers of the randomised screening trials. The cancer stage distribution in the NELSON trial appeared favourable relative to the other trials and was significantly lower ($p < 0.001$) than in the NLST. This last finding should be interpreted with caution because the NELSON trial used the 7th edition and the NLST the 6th edition of the TNM staging system[16,35] Classification according to the 7th edition results more often in a lower cancer stage than in a higher stage compared with classification according to the 6th edition.[16,35] Consequently, this might have contributed to the lower cancer stage in the NELSON trial. Nonetheless, the NELSON trial has a number of features that could cause a higher cancer stage: firstly, relatively few female participants (16.5% vs. 41% in NLST), who are diagnosed at a lower stage; secondly, larger nodules at referral, due to relatively stringent referral criteria (nodule volume > 500 mm^3 or nodule VDT < 400 days vs. nodule diameter > 4 mm in NLST); and thirdly, a longer screening interval (1, 2, and 2.5 years vs. annual screening in NLST). All things considered, it seems that the NELSON strategy is at least as capable as the NLST strategy to diagnose lung cancer at a more favourable stage.

Naturally, this result rises a question concerning what the difference in cancer stage between the two trials would be if all lung cancers in screened participants were compared. Analysis showed no significant difference ($p = 0.21$), despite the shorter interval between screen rounds in the NLST.[8]

Strengths of this study are the robust design (a large, randomised controlled trial) and prospective data collection. Limitations of this study are the lack of data for the control arm of the trial and lung cancer mortality. We have planned to perform analyses with those data 10 years after randomisation, in accordance with the main purpose of our trial.[11]

CONCLUSION

Despite stringent referral criteria, an increasing length of screening interval, and a small proportion of female participants, the screening strategy of the NELSON trial resulted in a favourable cancer stage distribution at diagnosis, which is a pre-requisite for the effectiveness of our screening strategy.

SUPPORT STATEMENT

Supported by Zorg Onderzoek Nederland-Medische Wetenschappen (ZonMw), KWF Kankerbestrijding, and Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen (RvvZ). 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.

CONFLICTS OF INTEREST

None declared

ACKNOWLEDGEMENTS

The authors thank their secretary M. Quak, data manager R. M. Vernhout, and system controllers R. Faber and F. Santegoets for contributions and for maintenance of the database. Also, the authors thank Dr. H. Stam (lung physiologist, Erasmus Medical Center) for useful comments.

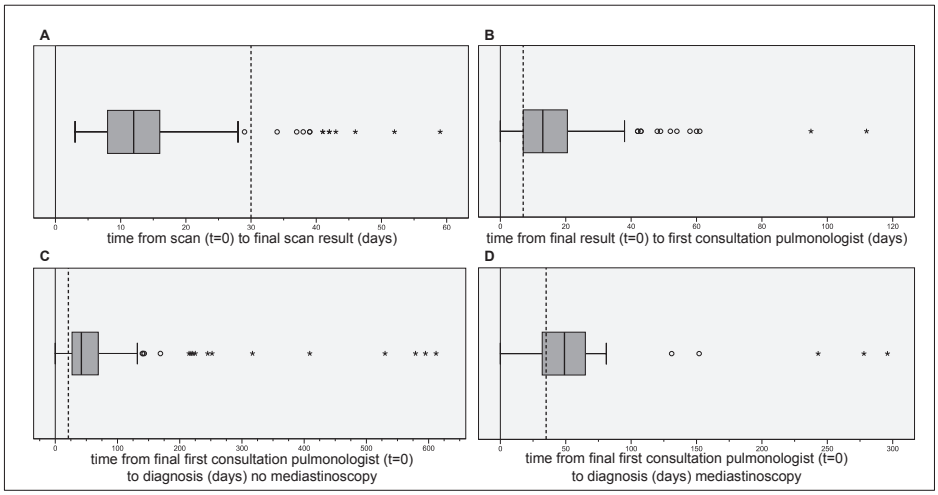
REFERENCES

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011; 61:69-90.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012; 62:10-29.
3. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361:2221-2229.
4. National Lung Screening Trial Research Team. The National Lung Screening Trial: overview and study design. *Radiology* 2011; 258:243–253.
5. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish randomized lung cancer CT screening trial - overall design and results of the prevalence round. *J Thorac Oncol* 2009; 4:608-614.
6. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer* 2008; 59:355-363.
7. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer* 2009; 64:34-40.
8. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365:395-409.
9. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54:177-184.
10. Ru Zhao Y, Xie X, de Koning HJ, et al. NELSON lung cancer screening study. Cancer Imaging 2011; 11:S79-S84.
11. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch–Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120:868-874.
12. Horeweg N, van Klaveren RJ, Groen HJ, et al. Blinded and uniform cause of death verification in a lung cancer CT screening trial. *Lung Cancer* 2012; 77:522-525.
13. Gietema HA, Wang Y, Xu D, et al. Pulmonary nodules detected at lung cancer screening: interobserver variability of semiautomated volume measurements. *Radiology* 2006; 241:251-257.
14. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. *Radiology* 2008; 248:625–631.
15. van Meerbeeck JP, Koning CC, Tjan-Heijnen VC, et al. Guideline on "non–small cell lung carcinoma; staging and treatment". *Ned Tijdschr Geneeskd* 2005; 149:72-77.
16. Goldstraw P, Crowley J, Chansky K, et al. International Association for the Study of Lung Cancer International Staging Committee. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. *J Thorac Oncol* 2007; 2:706-714.
17. Travis WD, Brambilla E, Müller-Hermelink HK, et al. World Health Organization classification of tumours: pathology and genetics of tumours of the lung, pleura, thymus and heart. Lyon, France: *IARC Press*; 2004.
18. van't Westeinde SC, de Koning HJ, Thunnissen FB, et al. The role of the ^{1}Fluorodeoxyglucose-positron emission tomography scan in the Nederlands Leuvens Longkanker Screenings Onderzoek Lung Cancer Screening Trial. *J Thorac Oncol* 2011; 6:1704-1712.
19. van'tWesteinde SC, Horeweg N, Vernhout RM, et al. The role of conventional bronchoscopy in the work-up of suspicious CT screen detected pulmonary nodules. *Chest* 2012; 142:377-384.

20. Parkash O. Lung cancer: a statistical study based on autopsy data from 1928 to 1972. *Respiration* 1977; 34:295-304.
21. Lince L, Lulu DJ. Carcinoma of the lung: a comparative series of 687 cases. *Arch Surg* 1971; 102:103-107.
22. Subramaniam RP, Asgharian B, Freijer JI, et al. Analysis of lobar differences in particle deposition in the human lung. *Inhal Toxicol* 2003; 15:1-21.
23. Churg A, Stevens B. Association of lung cancer and airway particle concentration. *Environ Res* 1988; 45:58-63.
24. Radzikowska E, Glaz P, Roszkowski K. Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20,561 cases. *Ann Oncol* 2002; 13:1087-1093.
25. Sagerup CM, Smastuen M, Johannesen TB, et al. Sex-specific trends in lung cancer incidence and survival: a population study of 40,118 cases. *Thorax* 2011; 66:301-307.
26. Ries L, Eisner M, Kosary C. SEER cancer statistics review 1975–2002; *National Cancer Institute* [accessed 2012 Sep 5]. Available from: http://seer.cancer.gov/csr/1975_2002/results_merged/sect_15_lung_bronchus.pdf
27. Schwartz KL, Crossley-May H, Vigneau FD, et al. Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes Control* 2003; 14:761-766.
28. Caldarella A, Crocetti E, Comin CE, et al. Gender differences in non–small cell lung cancer: a population-based study. *Eur J Surg Oncol* 2007; 33:763-768.
29. Ringer G, Smith JM, Engel AM, et al. Influence of sex on lung cancer histology, stage, and survival in a Midwestern United States tumor registry. *Clin Lung Cancer* 2005; 7:180-182.
30. Ramos Chaves M, Boleo-Tome C, Monteiro-Grillo I, et al. The diversity of nutritional status in cancer: new insights. *Oncologist* 2010; 15:523-530.
31. Parr CL, Batty GD, Lam TH, et al. Body-mass index and cancer mortality in the Asia–Pacific Cohort Studies Collaboration: pooled analyses of 424,519 participants. *Lancet Oncol* 2010; 11:741-752.
32. Renehan AG, Tyson M, Egger M, et al. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008; 371:569-578.
33. Kabat GC, Wynder EL. Body mass index and lung cancer risk. *Am J Epidemiol* 1992; 135:769-774.
34. Canoy D, Wareham N, Luben R, et al. Cigarette smoking and fat distribution in 21,828 British men and women: a population-based study. *Obes Res* 2005; 13:1466-1475.
35. American Joint Committee on Cancer. *AJCC cancer staging manual*, 6th ed. Chicago, IL: AJCC; 2002. pp. 167-177.
36. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; 67:296-301.
37. Infante M, Chiesa G, Solomon D, et al. Surgical procedures in the DANTE trial, a randomized study of lung cancer early detection with spiral computed tomography: comparative analysis in the screening and control arm. *J Thorac Oncol* 2011; 6:327-335.
38. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev* 2012; 21:308-315.
39. Becker N, Motsch E, Gross ML, et al. Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. *J Cancer Res Clin Oncol* 2012; 138:1475-1486.

APPENDIX

Figure 1. Time from positive screen to first consultation and to lung cancer diagnosis



Time to screening result, referral and diagnosis

The dotted line represents the deadline according to the NELSON protocol in the first box plot and the deadline according to the national Dutch guideline in the other three box plots.

Panel A: The median time to the final screening result for the 200 participants with screen-detected lung cancer was twelve days (interquartile range (IQR): 8 - 16) and 87.5% (175/200) waited ≤3 weeks.

Panel B: Thereafter, the median time to the first consultation by a pulmonologist was 13 days (IQR: 7 - 20.75) and 27.5% (55/200) had their first consultation in ≤5 work-days.

Panel C: In subjects who did not undergo a mediastinoscopy as part of the diagnostic work-up (n = 162), was the median time from the first consultation to the lung cancer diagnosis 42 days (IQR: 26.75 - 69), 17.9% was diagnosed in 3 weeks.

Panel D: When mediastinoscopy was performed (n = 38) the median time to diagnosis was 49 days (IQR: 31.5 - 66) and 28.9% was diagnosed in 5 weeks.

Analyses showed that neither a delayed final scan result ($p = 0.39$) nor a delayed first consultation ($p = 0.19$) was related to a more unfavourable cancer stage at diagnosis. Moreover, the participants with a delayed lung cancer diagnosis had a significantly lower disease stage than the persons without a delay ($p < 0.001$).

Outliers

Three participants had an extremely long lead-time between the positive scan and the first consultation (marked with a * in panel A). One participant (134 days) refused to go to the pulmonologist before a planned stay abroad. The two other participants (106 and 100 days) were delayed because of an administrative error.

Fifteen subjects had an extremely long lead-time between the first consultation and the diagnosis (marked with a * in panel B). Reasons were: watchful waiting approach by the pulmonologist (n = 9), delay caused by the participant (n = 2), comorbidity that required immediate treatment (n = 2), malignant nodule missed by wedge-resection; requiring a second procedure to perform a lobectomy (n = 1) and treatment of another benign nodule first (n = 1).

In total, eleven of the 200 (5.5%) participants had symptoms suspicious of lung cancer before they were diagnosed.

Symptomatic participants

Five participants had already symptoms suspicious of lung cancer before the screening scan was made:

1) 74 days before the third round scan: dyspnoea and cough
2) 287 days after the pre-randomisation questionnaire and 15 days before the baseline scan: dyspnoea, cough and thoracic pain.
3) 172 days before the second round scan: start weight loss >10%
4) 88 days before the baseline scan: start weight loss >10% and thoracic pain
5) 224 days after the pre-randomisation questionnaire and 124 days before the baseline scan: fatigue

Three participants got their first symptoms suspicious of lung cancer in the interval between the positive scan and the first consultation:

1) weight loss >10% and fatigue (interval was 15 days)
2) haemoptysis and thoracic pain (interval was 4 days)
3) cough (interval was 10 days)

Three other participants developed symptoms suspicious of lung cancer in the interval between the first consultation and the diagnosis date:

1) cough (interval was 278 days, delay due to cardiac valve replacement that had to be performed before lung surgery)
2) weight loss >10% (interval was 131 days, delay due to a false-positive N3 on the PET-scan, which required CT-guided punction and mediastinoscopy, that were both negative) haemoptysis (interval was 30 days, no delay)
3) cough (interval was 10 days)

Table 1. Predictive value of histological subtype for cancer stage at diagnosis

Thresholds for significant histological subtypes	Adenocarcinoma	Large cell carcinoma	Small cell carcinoma	Mixed LCSC carcinoma	NSCLC-NOS	All significant histological subtypes
Stage Ia to Ib	0.35	0.74	0.75	0.67	0.67	1.08
Stage Ib to IIa	0.59	0.99	1.02	0.92	0.92	1.37
Stage IIa to IIIa	0.95	1.36	1.41	1.28	1.28	1.82
Stage IIIa to IIIb	2.16	2.58	2.75	2.51	2.55	3.40
Stage IIIb to IV	2.73	3.15	3.36	3.09	3.17	4.11
Histological	**Univariate log regression analyses**			**Multivariate log regression analysis**		
subtypes	**estimate**	**95% CI**	**p-value**	**estimate**	**95% CI**	**p-value**
Adenocarcinoma	-0.57	-1.13--0.01	0.045	0.20	-0.46-0.86	0.56
Bronchoalveolar carcinoma	*	*	*			
Squamous cell carcinoma	0.19	-0.54-0.92	0.61			
Adenosquamous carcinoma	-0.60	-2.94-1.74	0.61			
Large cell carcinoma	1.16	0.25-2.08	0.013	1.68	0.67-2.70	0.001
Large cell NE carcinoma	0.32	-1.61-2.24	0.75			
Small cell carcinoma	2.93	1.57-4.28	<0.001	3.54	2.09-5.00	<0.001
Mixed LCSC carcinoma	3.08	0.45-5.72	0.022	3.97	1.29-6.65	0.004
Pleiomorph carcinoma	1.07	-2.48-4.63	0.55			
NSCLC- NOS	3.67	0.84-6.50	0.011	4.59	1.70-7.47	0.002
Carcinoid	*	*	*			
No histological diagnosis	-1.95	-4.04-0.13	0.07			

The thresholds in the columns "adenocarcinomas", "large cell carcinoma", "small cell carcinoma", "mixed LCSC" and "NSCLC-NOS" are parameters for the effect of the variable on the cancer stage (like; the distance between the stages) in four separate univariate logistic categorical regression analysis. There is no threshold for stage IIA to stage IIB because none of the participants were diagnosed with stage IIB lung carcinoma.

The threshold column "all significant histological subtypes" represents the threshold for the multivariate (adeno, large cell, small cell, mixed and NSCLC-NOS) logistic categorical regression model.

Definition of abbreviations: Mixed LCSL carcinoma = mixed large cell and small cell lung carcinom; NSCLC-NOS = non-small cell lung carcinoma; not otherwise specified; 95% CI = 95% confidence interval of the estimate; Large cell NE carcinoma = large cell neuro-endocrine carcinoma.

* *Both the bronchoalveolar carcinomas and the carcinoids were all diagnosed in stage IA, which caused separation, therefore no estimate or p-value could be calculated.*

Table 2a. Histology and disease stage of the 74 screen-detected lung cancers in round one

Disease stage*	Ia	Ib	IIa	IIb	IIIa	IIIb	IV	Overall
Histology†	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Adenocarcinoma	21 (60.0)	2 (5.7)	4 (11.4)	.	4 (11.4)	3 (8.6)	1 (2.9)	35 (47.3)
Bronchoalveolar carcinoma	2 (100.0)	.	.	.	.	.	.	2 (2.7)
Squamous cell carcinoma	9 (60.0)	.	2 (13.3)	.	3 (20.0)	.	1 (6.7)	15 (20.3)
Adenosquamous carcinoma	2 (100.0)	.	.	.	.	.	.	2 (2.7)
Large cell carcinoma	2 (40.0)	1 (20.0)	.	.	2 (40.0)	.	.	5 (6.8)
Large cell NE carcinoma	1 (33.3)	1 (33.3)	.	.	1 (33.3)	.	.	3 (4.1)
Small cell carcinoma	.	.	.	.	.	.	1 (100.0)	1 (1.4)
Mixed LCSC carcinoma	.	.	.	.	.	.	1 (100.0)	1 (1.4)
Pleiomorph carcinoma	.	.	1 (100.0)	.	.	.	.	1 (1.4)
NSCLC-NOS	.	.	.	.	.	1 (50.0)	1 (50.0)	2 (2.7)
Carcinoïd	4 (100.0)	.	.	.	.	.	.	4 (5.4)
No histological diagnosis§	3 (100.0)	.	.	.	.	.	.	3 (4.1)
Total	**44 (59.5)**	**4 (5.4)**	**7 (9.5)**	**.**	**10 (13.5)**	**4 (5.4)**	**5 (6.8)**	**74 (100.0)**

Definition of abbreviations: large cell NE carcinoma = large cell neuro-endocrine carcinoma; mixed LCSC carcinoma = mixed large cell/small cell carcinoma; NSCLC-NOS = non-small cell lung carcinoma, not otherwise specified; . = 0.0.

* *7th edition TNM staging system (2009).*

† *According to IARC Tumours of the Lung, Pleura and Heart (2004).*

§ ***In three participants no histological diagnosis was established because biopsies were unsuccessful or not performed and the patient did not undergo thoracic surgery because of poor pulmonary function (N = 2) and poor heart function (N = 1).***

Table 2b. Histology and disease stage of the 58 screen-detected lung cancers in round two

Disease stage*	Ia	Ib	IIa	IIb	IIIa	IIIb	IV	Overall
Histology†	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n(%)**
Adenocarcinoma	29 (82.9)	1 (2.9)	3 (8.6)	.	2 (5.7)	.	.	35 (60.3)
Bronchoalveolar carcinoma	3 (100.0)	.		.	.	.	.	3 (5.2)
Squamous cell carcinoma	2 (66.7)	.	.	.	1 (33.3)	.	.	3 (5.2)
Adenosquamous carcinoma	1 (50.0)	.	1 (50.0)	.	.	.	.	2 (3.4)
Large cell carcinoma	5 (50.0)	.	.	.	2 (20.0)	2 (20.0)	1 (10.0)	10 (17.2)
Large cell NE carcinoma	.	.	.	.	.	.	.	.
Small cell carcinoma	.	.	.	.	1 (50.0)	.	1 (50.0)	2 (3.4)
Mixed LCSC carcinoma	.	.	.	.	.	.	.	.
Pleiomorph carcinoma	.	.	.	.	.	.	.	.
NSCLC-NOS	.	.	.	.	.	.	.	.
Carcinoïd	.	.	.	.	.	.	.	.
No histological diagnosis§	3 (100.0)	.	.	.	.	.	.	3 (5.2)
Total	**43 (74.1)**	**1 (1.7)**	**4 (6.9)**	**.**	**6 (10.3)**	**2 (3.4)**	**2 (3.4)**	**58 (100.0)**

Def*inition of abbreviations: . = 0 (0.0); large cell NE carcinoma = large cell neuro-endocrine carcinoma; mixed LCSC carcinoma = mixed large cell/small cell carcinoma; NSCLC-NOS = non-small cell lung carcinoma, not otherwise specified.*

* *7th edition TNM staging system (2009).*

† *According to IARC Tumours of the Lung, Pleura and Heart (2004).*

§ *In 3 participants no histological diagnosis was established because biopsies were unsuccessful or not performed and the patient did not undergo thoracic surgery because of poor pulmonary function (n = 1), metastasised prostate carcinoma (n = 1) and death due to mesenteric ischemia before intended surgery (n= 1).*

Table 2c. Histology and disease stage of the 77 screen-detected lung cancers in round three

Disease stage*	Ia	Ib	IIa	IIb	IIIa	IIIb	IV	Overall
Histology†	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Adenocarcinoma	25 (67.6)	6 (16.2)	1 (2.7)	.	3 (8.1)	1 (2.7)	1 (2.7)	37 (48.1)
Bronchoalveolar carcinoma	6 (100.0)	.	.	.	.	.	.	6 (7.8)
Squamous cell carcinoma	10 (62.5)	.	1 (6.3)	.	4 (25.0)	.	1 (6.3)	16 (20.8)
Adenosquamous carcinoma	.	.	.	.	.	.	.	.
Large cell carcinoma	.	.	.	.	2 (100.0)	.	.	2 (2.6)
Large cell NE carcinoma	1 (100.0)	.	.	.	.	.	.	1 (1.3)
Small cell carcinoma	.	.	.	.	4 (80.0)	.	1 (20.0)	5 (6.5)
Mixed LCSC carcinoma	.	.	.	.	1 (100.0)	.	.	1 (1.3)
Pleiomorph carcinoma	.	.	.	.	.	.	.	.
NSCLC-NOS	.	.	.	.	.	.	.	.
Carcinoïd	2 (100.0)	.	.	.	.	.	.	2 (2.6)
No histological diagnosis§	6 (85.7)	.	1 (14.3)	.	.	.	.	7 (9.1)
Total	**50 (64.9)**	**6 (7.8)**	**3 (3.9)**	**.**	**14 (18.2)**	**1 (1.3)**	**3 (3.9)**	**77 (100.0)**

Definition of abbreviations: . = 0 (0.0); large cell NE carcinoma = large cell neuro-endocrine carcinoma; mixed LCSC carcinoma = mixed large cell/small cell carcinoma; NSCLC-NOS = non-small cell lung carcinoma, not otherwise specified.

* *7th edition TNM staging system (2009).*

† *According to IARC Tumours of the Lung, Pleura and Heart (2004).*

§ *In 7 participants no histological diagnosis was established because biopsies were unsuccessful or not performed and the patient did not undergo thoracic surgery because of poor pulmonary function (n= 4), poor general condition (n = 1), radiotherapy because of participation in other clinical trial (n = 1) and refusal (n = 1).*

Table 3a. Histology and disease stage of the 175 screen-detected lung cancers in 166 men

Disease stage*	Ia	Ib	IIa	IIb	IIIa	IIIb	IV	Overall
Histology†	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Adenocarcinoma	58 (66.7)	8 (9.2)	8 (9.2)	.	7 (8.0)	4 (4.6)	2 (2.3)	87 (49.7)
Bronchoalveolar carcinoma	8 (100.0)	.	.	.	.	.	.	8 (4.6)
Squamous cell carcinoma	17 (56.7)	.	3 (10.0)	.	8 (26.7)	.	2 (6.7)	30 (17.1)
Adenosquamous carcinoma	1 (50.0)	.	1 (50.0)	.	.	.	.	2 (1.1)
Large cell carcinoma	6 (40.0)	1 (6.7)	.	.	5 (33.3)	2 (13.3)	1 (6.7)	15 (8.6)
Large cell NE carcinoma	2 (66.7)	1 (33.3)	.	.	.	.	.	3 (1.7)
Small cell carcinoma	.	.	.	.	5 (62.5)	.	3 (37.5)	8 (4.6)
Mixed LCSC carcinoma		.	.	.	1 (50.0)	.	1 (50.0)	2 (1.1)
Pleiomorph carcinoma	.	.	1 (100.0)	.	.	.	.	1 (0.6)
NSCLC-NOS	.	.	.	.	.	1 (50.0)	1 (50.0)	2 (1.1)
Carcinoïd	4 (100.0)	.	.	.	.	.	.	4 (2.3)
No histological diagnosis§	12 (92.3)	.	1 (7.7)	.	.	.	.	13 (7.4)
Total	**108 (61.7)**	**10 (5.7)**	**14 (8.0)**	**.**	**26 (14.9)**	**7 (4.0)**	**10 (5.7)**	**175 (100.0)**

Definition of abbreviations: . = 0 (0.0); large cell NE carcinoma = large cell neuro-endocrine carcinoma; mixed LCSC carcinoma = mixed large cell/small cell carcinoma; NSCLC-NOS = non-small cell lung carcinoma, not otherwise specified.

* *7th edition TNM staging system (2009).*

† *According to IARC Tumours of the Lung, Pleura and Heart (2004).*

§ ***In 13 participants no histological diagnosis was established because biopsies were unsuccessful or not performed and the patient did not undergo thoracic surgery because of poor pulmonary function (n = 7), poor heart function (n= 1), poor general condition (n = 1), metastasized prostate carcinoma (n = 1), death due to mesenteric ischemia before intended surgery (n = 1), radiotherapy because of participation in other clinical trial (n = 1) and refusal (n = 1).***

Table 3b. Histology and disease stage of the 34 screen-detected lung cancers in 34 women

Disease stage*	Ia	Ib	IIa	IIb	IIIa	IIIb	IV	Overall
Histology†	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Adenocarcinoma	17 (85.0)	1 (5.0)	.	.	2 (10.0)	.	.	20 (58.8)
Bronchoalveolar carcinoma	3 (100.0)	.	.	.	.	.	.	3 (8.8)
Squamous cell carcinoma	4 (100.0)	.	.	.	.	.	.	4 (11.8)
Adenosquamous carcinoma	2 (100.0)	.	.	.	.	.	.	2 (5.9)
Large cell carcinoma	1 (50.0)	.	.	.	1 (50.0)	.	.	2 (5.9)
Large cell NE carcinoma	.	.	.	.	1 (100.0)	.	.	1 (2.9)
Small cell carcinoma	.	.	.	.	.	.	.	.
Mixed LCSC carcinoma	.	.	.	.	.	.	.	.
Pleiomorph carcinoma	.	.	.	.	.	.	.	.
NSCLC-NOS	.	.	.	.	.	.	.	.
Carcinoïd	2 (100.0)	.	.	.	.	.	.	2 (5.9)
No histological diagnosis	.	.	.	.	.	.	.	.
Total	**29 (85.3)**	**1 (2.9)**	**.**	**.**	**4 (11.8)**	**.**	**.**	**34 (100.0)**

Definition of abbreviations: . = 0 (0.0); large cell NE carcinoma = large cell neuro-endocrine carcinoma; mixed LCSC carcinoma = mixed large cell/small cell carcinoma; NSCLC-NOS = non-small cell lung carcinoma, not otherwise specified.

* *7th edition TNM staging system (2009).*

† *According to IARC Tumours of the Lung, Pleura and Heart (2004).*

Table 4. Predictive value of histological subtype for cancer stage at diagnosis

Thresholds	Gender	Age	BMI	Pack-years	All
Stage Ia to Ib	0.39	-0.28	4.11	1.20	3.90
Stage Ib to IIa	0.65	-0.03	4.38	1.45	4.18
Stage IIa to IIIa	1.01	0.33	4.77	1.82	4.58
Stage IIIa to IIIb	2.23	1.53	6.01	3.04	5.85
Stage IIIb to IV	2.80	2.09	6.64	3.61	6.48

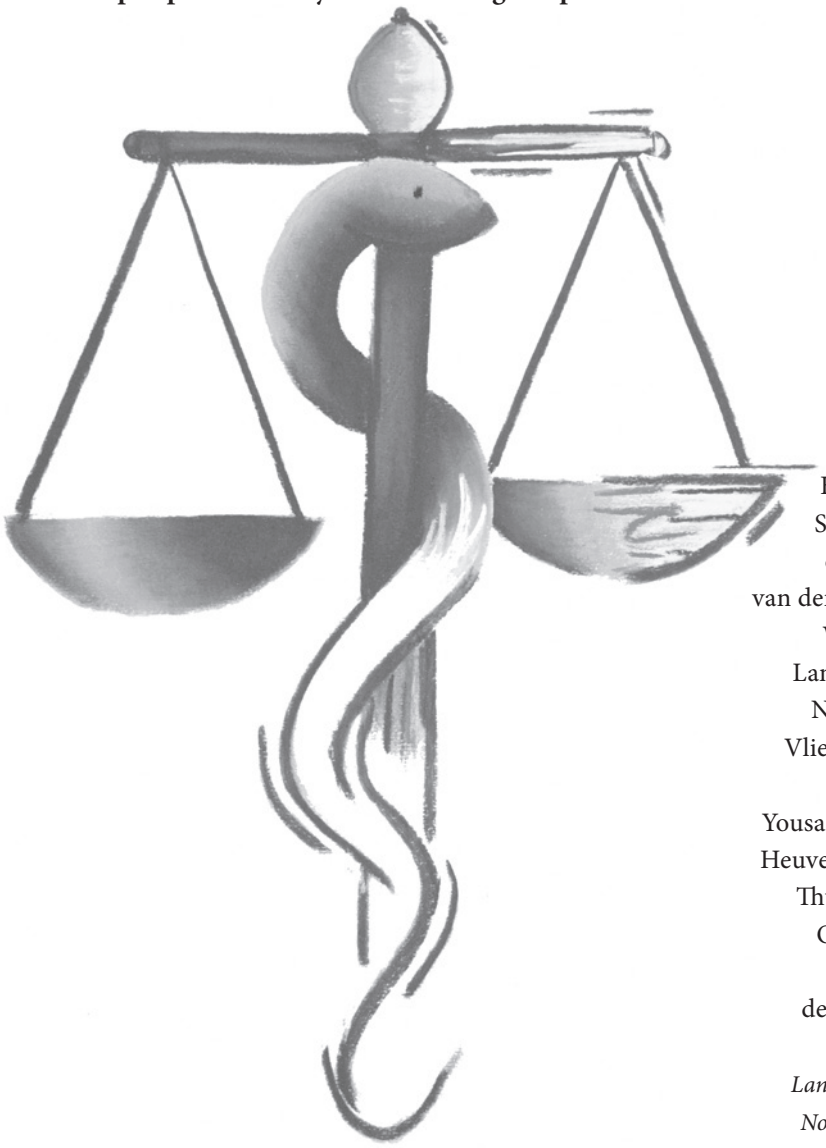
	Univariate log regression analyses			Multivariate log regression analysis		
Parameter estimates	**Estimate**	**95% CI**	**p-value**	**Estimate**	**95% CI**	**p-value**
Gender	-1.36	-2,35--0.37	0.007	-1.13	-2.13--0.12	0.028
Age	-0.01	-0.06-0.03	0.54			
BMI	0.14	0.06-0.21	0.001	0.12	0.038-0.20	0.004
Pack-years	0.01	-0.000-0.03	0.047	0.008	-0.005-0.021	0.22

Definitions of abbreviations: BMI = body-mass index; 95% CI = 95% confidence interval of the estimate
The thresholds in the columns "gender" (male as reference), "age", "BMI" and "pack-years" are parameters for the effect of the variable on the cancer stage (like; the distance between the stages) in four separate univariate logistic categorical regression analysis. There is no threshold for stage IIa to stage IIb because none of the participants were diagnosed with stage IIb lung carcinoma. The threshold column "all" represents the threshold for the multivariate (gender and BMI) logistic categorical regression model.

Chapter 4

Epidemiological evaluation

Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers



Horeweg N,
Scholten ET,
de Jong PA,
van der Aalst CM,
Weenink C,
Lammers JWJ,
Nackaerts K,
Vliegenthart R,
ten Haaf K,
Yousaf-Khan AU,
Heuvelmans MA,
Thunnissen E,
Oudkerk M,
Mali W,
de Koning HJ

Lancet Oncology
November 2014

Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers

Lancet Oncol. 2014 Nov;15(11).

ABSTRACT

Low-dose CT screening is recommended for individuals at high risk of developing lung cancer. However, CT screening does not detect all lung cancers: some might be missed at screening, and others can develop in the interval between screens. The NELSON trial is a randomised trial to assess the effect of screening with increasing screening intervals on lung cancer mortality. In this study, we aimed to assess screening test performance, and the epidemiological, radiological, and clinical characteristics of interval cancers in NELSON trial participants assigned to the screening group.

Eligible participants in the NELSON trial were those aged 50–75 years, who had smoked 15 or more cigarettes per day for more than 25 years or ten or more cigarettes for more than 30 years, and were still smoking or had quit less than 10 years ago. We included all participants assigned to the screening group who had attended at least one round of screening. Screening test results were based on volumetry using a two-step approach. Initially, screening test results were classified as negative, indeterminate, or positive based on nodule presence and volume. Subsequently, participants with an initial indeterminate result underwent follow-up screening at short notice to classify their final screening test result as negative or positive, based on nodule volume doubling time. We obtained information about all lung cancer diagnoses made during the first three rounds of screening, plus an additional 2 years of follow-up from the national cancer registry. We determined epidemiological, radiological, participant, and tumour characteristics by reassessing medical files, screening CTs, and clinical CTs. The NELSON trial is registered at www.trialregister.nl, number ISRCTN63545820.

15,822 participants were enrolled in the NELSON trial, of whom 7,915 were assigned to low-dose CT screening with increasing interval between screens, and 7,909 to no screening. We included 7,155 participants in our study, with median follow-up of 8.16 years (IQR 7.56-8.56). 187 (3%) of 7,155 screened participants were diagnosed with 196 screen-detected lung cancers, and another 34 (<1%, 19 [56%] in the first year of the interval, and 15 [44%] in the second year) were diagnosed with 35 interval cancers. The overall (three rounds of screening, and 2 years' follow-up) sensitivity was 84.6% (95% CI 79.6-89.2), specificity was 98.6% (98.5-98.8), positive predictive value was 40.4% (35.9-44.7), and negative predictive value was 99.8% (99.8-99.9). Retrospective assessment of

CT examinations showed that 12 (35%) of the 35 interval cancers were not visible at the last screening CT. The remaining cancers were visible when retrospectively assessed, but were not diagnosed because of radiological detection and interpretation errors (17 [50%]), misclassification by the protocol (two [6%]), participant noncompliance (two [6%]), and non-adherence to protocol (one [3%]). Compared with screen-detected cancers, interval cancers were diagnosed at more advanced stages (29 [83%] of 35 interval cancers vs 44 [22%] of 196 screen-detected cancers diagnosed in stage III or IV; $p < 0.001$, were more often small cell carcinomas (seven [20%] vs. eight [4%], $p = 0.003$) and less often adenocarcinomas (nine [26%] vs. 102 [52%], $p = 0.005$).

Lung cancer screening in the NELSON trial yielded high specificity and sensitivity, with only a small number of interval cancers. The results of this study could be used to improve screening algorithms, and reduce the number of missed cancers.

INTRODUCTION

Until the 1990s, no effective screening test for lung cancer was available. Screening studies using sputum cytology or chest radiography did not show a significant reduction in lung cancer mortality. In the 1990s, cohort studies using low-dose CT as a lung cancer screening test were initiated.[1,2,3] Low-dose CT seemed to be able to detect more and smaller lung cancers than chest radiography, with most being diagnosed at stage I.[4-6] Moreover, survival in patients with screen-detected lung cancer was impressive. In 2011, the National Lung Screening Trial (NLST) showed a 20% reduction in lung cancer mortality using low-dose CT compared with screening using chest radiography.[7]

The CISNET lung cancer working group modelled and assessed hundreds of screening scenarios using data from NLST; 26 selected efficient screening scenarios led to reductions in lung cancer mortality of between 4.6% to 21.2%.[8] As a result, the US Preventative Services Task Force and several medical societies recommended annual low-dose CT screening in the USA for subjects at high risk of developing lung cancer.[8-10] However, no reduction in lung cancer mortality with an annual low-dose CT screening strategy has been reported in three smaller European trials,[11-13] and results of several other European trials are still awaited. In many European countries, the outcome of the NELSON trial or pooled analyses is awaited before a decision about implementation of a national service lung cancer screening programme is made.

Efficacy and acceptance of low-dose CT screening for lung cancer depends on the sensitivity of the screening test (i.e., the risk of not detecting a lung cancer through screening). Lung cancers not detected by screening but diagnosed during the screening interval, known as interval cancers, might have been missed at screening or might have developed between screening and detection. Few studies about the incidence and characteristics of

interval cancers in lung cancer screening have been reported.[14-16] None of these studies assessed causes of interval cancers, or whether improvements to the screening algorithm were possible.

The NELSON trial is a randomised 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.[20] In this retrospective analysis, we aimed to assess the performance of the screening test to detect interval cancers, and provide insights into the incidence, histopathology, and causes for failed detection of these cancers.

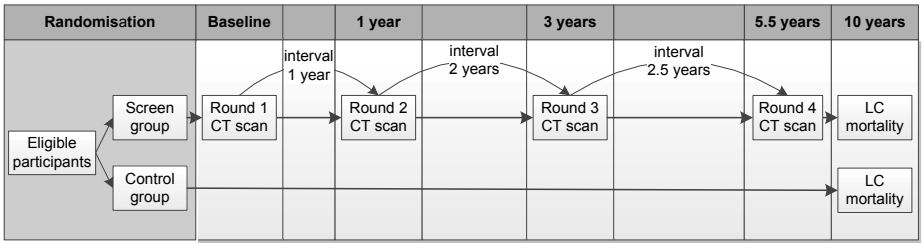
METHODS

Study design and participants

Individuals from four centres 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, inability to climb two flights of stairs, bodyweight of 140 kg or more, current or past renal cancer, melanoma, or breast cancer, lung cancer diagnosed less than five years ago, or a chest CT examination less than one year ago, were excluded. Design of the NELSON trial is presented in Figure 1.

In this study to assess the performance of the screening and the causes of the failure to detect the interval cancers, we included all Dutch participants who were randomly assigned to the screening group and received at least one screening in the first three

Figure 1. Design NELSON trial



Analyses of this study include data from the first screening round up to two years of follow-up after the third round scan. Median length of follow-up, from randomisation to end of follow-up at 31-12-2011 was 8.16 years (interquartile range: 7.56 to 8.56 years). Note, 32 of 7155 (0.4%) participants had their third round scan after 2009; as a result, their two-year follow-up period is not completely covered by the data of the national cancer registry.

screening rounds at baseline, one year later, and three years later. We did not include Belgian participants from the NELSON trial in this study because no data were available from the Belgian Cancer Registry.

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

Procedures

Screening was done using 16-detector CT scanners in a low-dose setting[22] (effective radiation dose <0.4 mSv to <1.6 mSv depending on bodyweight).[18] Datasets were derived from images of the thorax (slice thickness 1 mm, interval 0.7 mm) and volumes of nodules were measured using semi-automatic volumetric software (LungCARE, Siemens, Somaris/5 VB 10A-W). Volume doubling time was calculated for all nodules (no selection was made based on any characteristics suspicious for malignancy) with at least two measurement of its size using the formula:

$$\mathrm{VDT} = \frac{\ln(2)\Delta t}{\ln(V_2) - \ln(V_1)}$$

in which Δt represents time in days between scans, V1 the volume of the nodule at baseline, and V2 the volume of the nodule at the current CT examination.[18]

The screening test results were determined by the presence, size and growth rate of pulmonary nodules. Screening test results were defined to be negative: in the absence of nodules, for all nodules with fat, benign calcification patterns or other benign abnormalities,[18,23] and for non-calcified nodules with a volume <50mm³, a percentage volume-change (PVC) <25%, or a percentage volume change ≥25% combined with a volume doubling-time ≥600days.[18] Screening test results were defined to be positive for non-calcified nodules with a volume >500mm³, or with a PVC of ≥25% combined with a VDT <400days, Moreover, screening test results were also positive if a new solid component had emerged in a previously non-solid nodule.[18,23] Screening test results were defined to be indeterminate for nodules with a volume 50-500mm³, or a PVC ≥25% combined with a VDT of 400-600days.[18] Subjects with an indeterminate screening test result were invited for one additional LDCT examination at the screening center at short notice to determine whether their final screening test result was positive or negative using the aforementioned criteria.

After a negative final screening test result, participants did not undergo any additional diagnostic procedures, but only received an invitation for the next screening round. After a positive final screening result, participants were referred to a pulmonologist via their general practitioner for diagnostic work-up to exclude or diagnose lung cancer.[24-26]

To assess interval cancers, we obtained data for all lung cancers diagnosed since the first screening in round one to the last screening in round three, plus an additional two years of follow-up from the Dutch Cancer Registry.[27] For every patient diagnosed with lung cancer outside of screening (i.e., diagnosed with interval cancer), we collected medical and radiological files. Two radiologists (with 10 [PAdJ] and >30 years [EThS] of experience with chest CT) reviewed the last screening CT from the study and the clinical CT used for diagnosis of lung cancer, and reached a consensus on whether or not the lung cancer could retrospectively be identified on the screening CT. We determined epidemiological, radiological, participant, and tumour characteristics by reassessing medical files, screening CTs, and clinical CTs.

Outcomes

The primary endpoint of the NELSON trial is reduction of lung cancer mortality by 25% or more at ten years after randomisation.[20,21] The primary aim of this analysis was to assess the frequency of interval lung cancers, and to determine the sensitivity, specificity, positive predictive value, and negative predictive value of the screening protocol. The secondary aim was to assess, and compare, the histopathological type and stage of screen-detected and interval lung cancers, and to assess the causes of the failure to detect the interval cancers.

Screen-detected cancers were defined as lung cancers diagnosed by diagnostic work-up initiated for a positive screening test result. We defined interval cancers as: lung cancers diagnosed after a negative screening test; lung cancers diagnosed after an indeterminate screening test, but without any follow-up low-dose CT examination or diagnostic work-up being done in the screening programme; or lung cancers diagnosed after a positive screening result if the diagnostic work-up initiated for the positive screening result did not yield a diagnosis of lung cancer, and the diagnosis was made later because symptoms had triggered diagnostic assessment that eventually yielded diagnosis of lung cancer. Diagnostic work-up was defined not to have yielded diagnosis of lung cancer if a pulmonologist had concluded that the suspicious nodule was not lung cancer and dismissed the participant from any further diagnostic procedures or follow-up, or if diagnostic workup did not yield diagnosis of lung cancer and was still ongoing after two years of follow-up.

A true-positive test result was a positive result in a participant who actually was diagnosed with lung cancer by diagnostic work-up. A false-positive test result was a positive result in the absence of lung cancer. A true-negative test result was a negative scan in the absence of lung cancer, and a false-negative test result was a negative scan followed by diagnosis of interval cancer.

Statistical analysis

All screening test characteristics were estimated with the detection method, as done in the NLST.[14] 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 participants with a true-positive screening by all participants with positive screening. We estimated negative predictive value by dividing all participants with a true-negative screening by all participants with negative screening. To calculate 95% CIs, we did bootstrapping based on 5,000 samples. Continuous variables were tested for normality; the significance of the differences was assessed using one-way analysis of variance. For nominal variables we used Fisher's exact test or likelihood-based χ^2 test. For categorical variables we used the Mann-Whitney U test. All analyses were done with PASW Statistics, IBM SPSS (version 20).

RESULTS

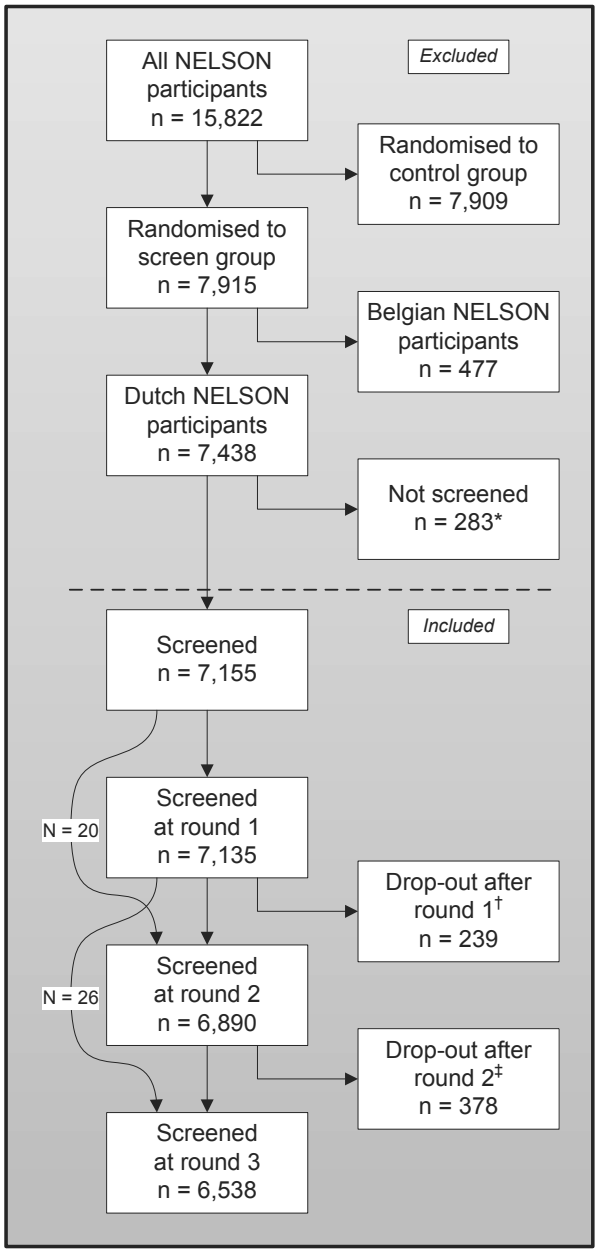
Epidemiological characteristics

Between Dec 23, 2003, and July 6, 2006, 15,822 individuals from four centres in the Netherlands and Belgium were enrolled and randomly assigned to receive low-dose CT screening (n =7,915) or no screening (n = 7,907). For this analysis, we excluded the 7,909 participants randomly assigned to the no screening group, the 477 participants from Belgium (because no data were yet available from the Belgian Cancer Registry), and 283 participants who did not attend their screening examinations (because no screening test characteristics could be calculated in the absence of screening). Thus, we included 7,155 participants in our analysis (Figure 2). Median length of follow-up was 8.16 years (IQR 7.56–8.56).

There was no significant difference between included and excluded participants' baseline characteristics (Table 1). 7,135 (96%) of 7,438 participants in the screening group attended the first round of screening, 6,890 (93%) attended the second round, and 6,538 (88%) attended the third round. The final positive screening test results led to the diagnosis of lung cancer in 187 participants (3%, Table 2). Additionally, 34 (<1%) participants were diagnosed with lung cancer between screening rounds (19 in the first year since screening and 15 in the second year, Table 2).

Test characteristics for each separate screening round are provided in table 2. For the three screening rounds combined, sensitivity was 84.6% (95% CI 79.6–89.2%), specificity was 98.6% (95% CI 98.5–98.8%), positive predictive value was 40.4% (95% CI 35.9-44.7%), and negative predictive value was 99.8% (95% CI 99.8–99.9%). When only the first year of the screening interval was considered, five (26%) of the 19 participants with interval cancers were identified during the interval. This finding implies that maximum sensitivity (assuming not a single lung cancer was missed) of an annual screening programme would be 97.4% (95% CI 94.8-99.5%), and of a two-yearly screening programme (with an initial annual screening round) would be 94.0% (95% CI 90.5-97.0%).

Figure 2. Flowchart of included participants



* *No response despite repeated invitations for first screening round.*

† *Reasons: screen-detected lung cancer (n = 61), death (n = 25), participant declined (n = 110), participant unattainable or repeatedly no show (n = 42), and still in diagnostic work-up round one (n = 1).*

‡ *Reasons: screen-detected lung cancer (n = 54), death (n = 79), participant declined (n = 145), participant unattainable or repeatedly no show (n = 94), administrative error (n = 3), and unknown (n = 3).*

Table 1. Characteristics of included and excluded participants

Characteristics	All participants n (%)	Included n (%)	Excluded n (%)	p-value*
Female gender	2,597 (16.5)	1,156 (16,2)	1,441 (16.7)	0.34
Current smoker	8,768 (55.4)	3,959 (55,3)	4,809 (55.5)	0.82
Age, median (IQR)	58.0 (54.0 - 62.0)	58.0 (54.0 - 62.0)	58.0 (54.0 - 62.0)	0.39
Pack-years, median (IQR)	38.0 (29.7 - 49.5)	38.0 (29.7 - 49.5)	38.0 (29.7 - 49.5)	0.53
Total	15,822 (100.0)	7,155 (45.2)	8,667 (54.8)	NA

Definition of abbreviations: IQR= interquartile range; NA = not applicable.

* *Difference between included and excluded participants.*

Table 2. Epidemiological characteristics round one to three

Epidemiological characteristics	Round 1	Round 2		Round 3		Total round 1-3	
	year 1	year 2	year 3	year 4	year 5	1-year follow-up	2-year follow-up
Screened participants	7,135	6,890		6,538		7,155	7,155
Negative test result	6,951	6,769		6,380		20,100	20,100
- true negative	6,946	6,762	6,750	6,373	6,370	20,081	20,066
- false negative	5	7	7 + 12	7	7 + 3	19	34
Positive test result	184	121		158		463	463
- true positive	62	53	53	72	72	187	187
- false positive	122	68	68	86	86	276	276
Total no. of detected cancers	62	53		72		187	187
- per 1000 screened	8.69	7.69		11.0		26.1	26.1
Total no. of interval cancers	5	7	19	7	10	19	34
- per 1000 screened	0.70	1.02	2.76	1.07	1.53	2.7	4.8
Ratio detected : interval	12.4:1	7.6:1	2.8:1	10.3:1	7.2:1	9.8:1	5.5:1
Sensitivity	92.5	88.3	73.6	91.1	87.8	90.8	84.6
Specificity	98.3	99.0	99.0	98.7	98.7	98.7	98.6
Positive predictive value	33.7	43.8	43.8	45.6	45.6	40.4	40.4
Negative predictive value	99.9	99.9	99.7	99.9	99.8	99.9	99.8

Radiological characteristics

Reassessment of CT examinations of the 34 participants with an interval cancer suggested that no lung cancer was present at the last screening examination in 12 cases (35%, Table 3). In the remaining 22 (65%) cases, we retrospectively identified a suspicious abnormality on the screening CT examination. In most cases, the suspicious abnormality was missed. Causes of the failure to detect these lung cancers were human error (two [6%]), interpretation error (two [6%]), and detection error due to various causes (13 [38%]). The 13 lung

Table 3. Radiological characteristics interval lung cancers

Radiological characteristics	Round 1	Round 2		Round 3		Total round 1-3	
	year 1	year 2	year 3	year 4	year 5	1-year follow-up	2-year follow-up
Causes	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Normal screening CT	1 (20.0)	3 (42.9)	7 (58.3)	1 (14.3)	.	5 (26.3)	12 (35.3)
Non-compliance participant	.	.	.	.	2 (66.7)	.	2 (5.9)
Non-adherence to protocol	.	.	.	1 (14.3)	.	1 (5.3)	1 (2.9)
Inadequacy protocol	1 (20.0)	.	.	1 (14.3)	.	2 (10.5)	2 (5.9)
Detection error*	3 (60.0)	2 (28.6)	4 (33.3)	3 (42.9)	1 (33.3)	8 (42.1)	13 (38.2)
Interpretation error	.	1 (14.3)	1 (8.3)	.	.	1 (5.3)	2 (5.9)
Human error	.	1 (14.3)	.	1 (14.3)	.	2 (10.5)	2 (5.9)
Total no. of interval cancers	**5**	**7**	**12**	**7**	**3**	**19**	**34**

Definition of abbreviation: . = 0 (0.0)

cancers missed because of these detection errors were: intra-bronchial localised lesions (five [15%]); pleural-attached lesions (two [6%]); lesion adjoining bullous structure (one [3%]); lesion surrounded by extensive honeycombing (one [3%]); and four cases where an intrapulmonary lesion was not visible but signs of lung cancer metastasis were missed (three [9%] cases of mediastinal lymphadenopathy, and one [3%] case of pleural effusion). In the remaining five cases, the abnormality was detected, but lung cancer was not diagnosed because of participant non-compliance (two [6%]), or the abnormality was not classified as suspicious by the protocol (one [3%]), or the abnormality was manually classified as not suspicious by the radiologist because of a negative diagnostic work-up in a previous screening round (two [6%]).

We calculated test sensitivity using the results of this retrospective radiological assessment, using only those interval cancers that were due to test failures. Assuming a 1-year screening interval, test sensitivity would have been 93.9% (95% CI 87.9-98.5%) in round one, 93.0% (95% CI 86.0-98.2%) in round two, and 92.3% (95% CI 85.9-97.4%) in round three.

Clinical characteristics

Participants diagnosed with lung cancer (both detected and interval cancers) were significantly older than were participants without lung cancer (Table 4). Only participants with an interval cancer (but not those with detected cancer) were significantly more likely to be current smokers than were participants with no cancer (Table 4).

The 187 participants with screen-detected lung cancer had a total of 196 tumours, and the 34 participants with interval cancer had a total of 35 tumours; nine participants with screen-detected cancers and one participant with interval cancer were diagnosed with synchronous double tumours. Interval cancers were diagnosed at a significantly higher disease stage ($p < 0.001$) than were screen-detected lung cancers (Table 5a).

Table 4. Characteristics of 7,155 included participants

Characteristic	No lung cancer n (%)	Detected cancer n (%)	P-value*	Interval cancer n (%)	P-value*
Male	5,817 (83.9)	154 (82.4)	0.57	28 (82.4)	0.81
Age - median (IQR)	58.0 (54.0 - 62.0)	61.0 (57.0 - 66.0)	<0.001	61.0 (58.0-66.3)	0.03
Current smoker	3,827 (55.3)	104 (55.6)	0.93	28 (82.4)	0.002
Pack-years - median (IQR)	38.0 (30.0 - 49.0)	44.0 (32.0 - 55.0)	0.19	39.0 (34.0-59.3)	0.51
Total	**6,934**	**187**	**NA**	**34**	**NA**

Definition of abbreviations: IQR = interquartile range; NA = not applicable.

* *Compared to included participants without lung cancer.*

Table 5a. Clinical characteristics detected and interval cancers - disease stage

	Disease stage							
Lung cancer	**Ia**	**Ib**	**IIa**	**IIb**	**IIIa**	**IIIb**	**IV**	**Total**
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Round 1								
Detected cancer	41 (62.1)	3 (4.5)	5 (7.6)	.	10 (15.2)	3 (4.5)	4 (6.1)	66 (100.0)
Interval cancer	.	.	.	.	.	1 (20.0)	4 (80.0)	5 (100.0)
Round 2								
Detected cancer	41 (73.2)	1 (1.8)	4 (7.1)	.	6 (10.7)	2 (3.6)	2 (3.6)	56 (100.0)
Interval cancer first year	.	.	.	.	.	.	7 (100.0)	7 (100.0)
Interval cancer second year	1 (8.3)	.	.	1 (8.3)	3 (25.0)	.	7 (58.3)	12 (100.0)
Round 3								
Detected cancer	48 (64.9)	6 (8.1)	3 (4.1)	.	13 (17.6)	1 (1.4)	3 (4.1)	74 (100.0)
Interval cancer first year	.	.	.	1 (14.3)	.	1 (14.3)	5 (71.4)	7 (100.0)
Interval cancer second year	2 (50.0)	.	.	1 (25.0)	.	.	1 (25.0)	4 (100.0)
Total								
All detected cancers	130 (66.3)	10 (5.1)	12 (6.1)	.	29 (14.8)	6 (3.1)	9 (4.6)	196 (100.0)
Interval cancers first year	.	.	.	1 (5.3)	.	2 (10.5)	16 (84.2)	19 (100.0)
Interval cancers second year	3 (18.8)	.	.	2 (12.5)	3 (18.8)	.	8 (50.0)	16 (100.0)
All interval cancers	3 (8.6)	.	.	3 (8.6)	3 (8.6)	2 (5.7)	24 (68.6)	35 (100.0)
All lung cancers	133 (57.6)	10 (4.3)	12 (5.2)	3 (1.3)	32 (13.9)	8 (3.5)	33 (14.3)	231 (100.0)

Definition of abbreviation: . = 0 (0.0)

* *According to the 7th edition of the TNM staging system for lung cancer.*

The numbers of lung cancers presented are not equal to the number of participants as nine participants with screen-detected lung cancer (round 1 n = 4; round 2 n = 3; round 3 n = 2), and 1 participants with an interval cancer (second year round 3) were diagnosed with synchronous double tumours.

Diagnosis at stage T1N0M0 occurred only in three (9%) of 35 interval cancers, whereas 130 (66%) of 196 screen-detected lung cancers were diagnosed at that stage. The disease stage of interval cancers diagnosed in the first year since screening was significantly higher than the stage of those diagnosed in the second year (p = 0.02). Interval cancers

were significantly more often small-cell carcinomas (p = 0.003) and less often adenocarcinomas (p = 0.005) than were screen-detected cancers (Table 5b). There was no significant difference between other histological subtypes between screen-detected and interval cancers. The localisation of tumours across the lungs did not significantly differ between interval and screen-detected lung cancers (data not shown).

Table 5b. Clinical characteristics detected and interval cancers - histological subtype

	Histological subtype*							
Lung cancer	**Adeno**	**BAC**	**Squamous**	**Large cell**	**Small cell**	**Other†**	**Unknown‡**	**Total**
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Round 1								
Detected cancer	32 (48.5)	2 (3.0)	11 (16.7)	5 (7.6)	1 (1.5)	12 (18.2)	3 (4.5)	66 (100.0)
Interval cancer	1 (20.0)	.	.	3 (60.0)	.	1 (20.0)	.	5 (100.0)
Round 2								
Detected cancer	34 (60.7)	3 (5.4)	3 (5.4)	10 (17.9)	2 (3.6)	1 (1.8)	3 (5.4)	56 (100.0)
Interval cancer first year	2 (28.6)	.	.	1 (14.3)	4 (57.1)	.	.	7 (100.0)
Interval cancer second year	2 (16.7)	.	5 (41.7)	.	3 (25.0)	1 (8.3)	1 (8.3)	12 (100.0)
Round 3								
Detected cancer	36 (48.6)	5 (6.8)	15 (20.3)	2 (2.7)	5 (6.8)	4 (5.4)	7 (9.5)	74 (100.0)
Interval cancer first year	1 (14.3)	.	1 (14.3)	2 (28.6)	.	2 (28.6)	1 (14.3)	7 (100.0)
Interval cancer second year	3 (75.0)	.	.	.	.	1 (25.0)	.	4 (100.0)
Total								
All detected cancers	102 (52.0)	10 (5.1)	29 (14.8)	17 (8.7)	8 (4.1)	17 (8.7)	13 (6.6)	196 (100.0)
Interval cancers first year	4 (21.1)	.	1 (5.3)	6 (31.6)	4 (21.1)	3 (15.8)	1 (5.3)	19 (100.0)
Interval cancers second year	5 (31.3)	.	5 (31.3)	.	3 (18.8)	2 (12.5)	1 (6.3)	16 (100.0)
All interval cancers	9 (25.7)	.	6 (17.1)	6 (17.1)	7 (20.0)	5 (14.3)	2 (5.7)	35 (100.0)
All lung cancers	107 (46.3)	10 (4.3)	35 (15.2)	23 (10.0)	15 (6.5)	22 (9.5)	15 (6.5)	231 (100.0)

Definition of abbreviations: . = 0 (0.0); Adeno = adenocarcinoma; BAC = bronchoalveolar carcinoma.

* *According to the 7th edition of the TNM staging system for lung cancer.*

† *Other histological sybtypes of lung cancer were: adenosquamous carcinoma; mixed large cell small cell carcinoma; large cell neuroendocrine carcinoma; carcinoid; mucinous carcinoma; pleiomorph carcinoma; non-small cell lung carcinoma, not otherwise specified.*

‡ *In 15 participants no histological diagnosis was established because biopsies were unsuccessful or not performed and the patient did not undergo thoracic surgery because of poor pulmonary function (n=7), poor heart function (n=1), poor general condition (n=3), metastasized prostate carcinoma (n=1), death due to mesenteric ischemia before intended surgery (n=1), radiotherapy because of participation in other clinical trial (n=1), and refusal (n=1).*

The numbers of lung cancers presented are not equal to the number of participants as nine participants with screen-detected lung cancer (round 1 n = 4; round 2 n = 3; round 3 n = 2), and 1 participants with an interval cancer (second year round 3) were diagnosed with synchronous double tumours.

DISCUSSION

In this study, we assessed the epidemiological, radiological, and clinical characteristics of screen-detected and interval lung cancers in the NELSON trial. 187 (3%) of the 7,155 participants studied were diagnosed with lung cancer detected by screening, and another 34 (<1%) participants were diagnosed with interval lung cancer. Overall, sensitivity was about 85%, specificity about 99%, positive predictive value about 40%, and negative predictive value greater than 99%. Retrospectively, about a third of the interval cancers were not visible on the last screening CT; the remaining cancers were retrospectively visible, but were not diagnosed. Interval cancers were diagnosed at more advanced stages, and were more often small cell carcinomas and less often adenocarcinomas than were screen-detected cancers.

Screening test results were based on nodule volume and growth rate, measured by volumetry. Because indeterminate screening test results were not communicated to study participants as being suspicious for lung cancer, and follow-up low-dose CT examinations were done in the context of the screening trial, indeterminate screening results should not be regarded as positive screening test results for the purposes of comparison with other screening trials.

Because all participants received a final screening result that was either positive or negative, calculated test characteristics can be compared with those from other screening trials.

Almost all participants received negative final results from the screening test; only 2.6% had positive final results, and needed diagnostic procedures to exclude or diagnose lung cancer. In other trials of low-dose CT lung cancer screening, the proportion of positive screening tests was higher: 15% (annual screening group) and 14% (biennial screening group) in the Italian MILD trial,[13] 24% in the US NLST,[7] 26% in the ITALUNG trial,[28] 27% in the German LUSI trial,[16] and 27% in the Italian DANTE trial.[12] These differences were probably caused by differences in the criteria for screening test results and the applied screening technique. In the NELSON trial, relatively stringent criteria for a positive screening result (nodule volume >500 mm^3 or volume doubling time <400 days)[18] and volumetry were used, which might have increased measurement accuracy, and reduced false-positive screening results.[17,29]

The predictive value of the positive screening test results was 40.4% (95% CI 35.9-44.7) in the NELSON trial. Although this figure implies that more than half of the participants were referred for false-positive results, this positive predictive value was high compared with positive predictive values of other trials: 3.8% (95% CI 3.4-4.3%) in the NLST,[7] 4.1% in the LUSI trial,[16] and 12.7% (biennial screening group) and 16.4% (annual screening group) in the MILD trial.[13]

Sensitivity in the first (annual) screening round was 92.5% (95% CI 85.5-98.4%), which is similar to that in other screening trials using annual screening: 93.8% (90.6-96.3%) in the NLST,[14,30] and 85.3% in the annual screening group of the MILD trial.[13] However, specificity was higher in the NELSON trial (98.3% [95% CI 98.8-99.2%]) than in either the NLST (73.4% [95% CI 72.8-73.9%]) or the MILD trial (86.8%). Sensitivity in the second screening round (biennial screening) was 73.6% (95% CI 62.5-83.6%) and in the third screening round (the first 2 years of the screening interval) was 87.8% (79.5-92.8%), which is probably similar to the overall sensitivity of 80.0% in the biennial screening group of the MILD trial.[13]

No appropriate comparison of screening test characteristics between annual and biennial screening trials can be made, because differences in performance between screening trials are not only caused by differences in the length of screening interval, but also by differences in the length of follow-up, criteria for a positive screening result, and lung cancer risk of the study population. Some comparisons of annual versus biennial screening were made in two modelling studies.[8,31] These findings suggested that biennial screening is less effective in absolute terms,[8,31] but induces substantially fewer harms (i.e. radiation-related lung cancer deaths, false-positive screening test results, number of screening examinations required per subject, overdiagnosis) than does annual screening[8], and might be similarly cost effective.[31] However, because only data from annual screening trials was used for these two modelling studies,[8,31] estimates of effectiveness and harms of biennial screening were based on extrapolations and thus these data may not be accurate.[8,31]

Whether the NELSON trial will show effectiveness with its increasing length of screening intervals can only be established by mortality analyses, which are planned at 10 years after randomisation. Nonetheless, both sensitivity and specificity noted in the current study are promising for cost-effectiveness. However, ratios between detected and missed lung cancers might be affected by overdiagnosis. The amount of overdiagnosis in the NELSON trial is still unknown because required data are not yet available, although overdiagnosis in lung cancer screening was estimated to be small in a modelling study using data from the NLST.[8]

Reassessment of clinical CT and last screening CT examination showed the causes of the failure to detect interval cancers. Two-thirds of the interval cancers were, retrospectively, visible at the last screening CT examination. Detection errors, interpretation errors, and human errors were identified as the main causes of failure in half of the interval cancers. Increased attention of screening radiologists for lung cancer presenting as endobronchial lesions, pleural-attached lesions, and bulla wall thickenings, and increased attention for extra-pulmonary signs of lung cancer, could help to reduce detection failures. Additionally, one of the interval cancers was not diagnosed through screening due to manual adjustment of the screening test result by the radiologist from positive to negative, because a diagnostic work-up done in an earlier round did not yield the diagnosis of lung cancer.

In view of the magnitude and importance of radiological causes, a second study on this topic was done.[32] For this study, CT examinations of interval cancers and post-screening cancers (diagnosed ≥2 years since last attended screening) were reviewed to determine causes of these errors, and to provide recommendations specifically for radiologists.[32]

Failure of the screening protocol to classify cancerous nodules as suspicious was rare. Only two of 34 participants with interval cancers were not diagnosed because the cancerous nodule shrunk or had a volume doubling time greater than 400 days, suggesting that the relatively stringent criteria for a positive result in the NELSON trial did not lead to notable numbers of missed cancers. This finding is encouraging for future screening programmes that aim to limit harms and costs.[33] Moreover, two of 34 participants with interval cancers were actually identified, but diagnosis was not made through screening because participants did not comply with the screening protocol. Instead of undergoing receiving follow-up low-dose CT screening three months after their indeterminate screening test result, they directly underwent diagnostic resection of the nodule, which yielded the diagnosis of lung cancer. Arguably, these interval cancers might have been detected by screening if the participants had complied with the protocol. However, we decided not to classify these cancers as detected by screening because of uncertainty about whether the nodules would have shown malignant growth at follow-up CT screening, and whether diagnostic work-up would have yielded a diagnosis of lung cancer. Finally, a third of interval cancers were, also in retrospect, not visible at the last screening examination, and thus were not missed, but arose during the interval.

All participants of this study were at substantial risk of developing lung cancer because of the enrolment requirements. Even within this population, older age and being a current smoker were still significant risk factors for development of lung cancer. Notably, only interval cancers were significantly associated with being a current smoker, which might be because continued smoking promotes the development of lung cancer subtypes that grow faster and are less perceptible by low-dose CT screening (e.g., small-cell carcinomas).[34] This finding reinforces urgency of smoking cessation in individuals receiving lung cancer screening.

Our findings showed that screening-detected lung cancers differed significantly to interval cancers with regards to stage of diagnosis, and histopathology. Differences in tumour characteristics are probably caused by both earlier diagnosis of screen-detected lung cancer as a result of screening asymptomatic individuals, and by the aggressive nature of interval lung cancers compared with detected cancers. In this study, all cancers that developed during the interval (i.e., were not missed at screening) were diagnosed at stage III or IV. Hence, these cancers grew from undetectable to incurable cancers in less than 1 year (five [36%] of 19) or 2 years (seven [47%] of 15), suggesting an enormous growth and metastatic potential. This observation is consistent with the finding that these

cancers were significantly more often small cell carcinomas than were interval cancers that did not arise during the screening interval.

In this study, 62% of all lung cancers were diagnosed at stage I, and only 18% were diagnosed at stage IIIB or IV. In the NLST, 59% of lung cancers were diagnosed at stage I, and 23% at stage IIIB or IV, which did not significantly differ from the NELSON trial ($p = 0.20$). Thus, despite longer screening intervals, slightly lower sensitivity, and fewer female participants (in whom CT screening appeared to detect lung cancer earlier than in males[35]) in the NELSON trial, lung cancer was diagnosed as early as in the NLST.[7] This finding is encouraging for effectiveness of lung cancer screening regimens using 2-yearly screening after an initial annual screening round.

CONCLUSION

In conclusion, our findings show that using low-dose CT screening with increasing intervals and stringent diagnostic criteria for a positive result to detect lung cancer gives high specificity and a high sensitivity. The results of this study could be used to improve screening algorithms and reduce the number of missed cancers.

RESEARCH IN CONTEXT

Systematic Review

As part of planning for this trial a systematic review was conducted in PubMed database. To identify all relevant articles on the performance of lung cancer screening test performance and interval cancers, the following search terms were used: "Lung Neoplasms"[Mesh], "Tomography, X-Ray Computed"[Mesh], "Mass screening"[Mesh]", "Epidemiologic Study Characteristics as Topic"[Mesh]. In addition, Pubmed was searched for articles on all randomised controlled trials on lung cancer screening by searching for the trial's acronyms. Limits used for all searches: humans, adults; published in English, in core clinical journals or MEDLINE. Titles and abstracts of articles that were identified using these search terms were scanned to select articles relevant for this study. Reference lists of relevant articles were checked to identify more relevant articles.

Interpretation

Compared to the literature, the screening strategy of the NELSON trial performed well. Hence, screening test sensitivity was comparable other studies or slightly lower, the specificity was very high, negative predictive value was as high as in other studies and the positive predictive value was substantially higher. Moreover, lung cancer was as early diagnosed in the NELSON trial as in the NLST, which is a prerequisite for effectiveness.

Only a limited number of studies report on interval cancers in lung cancer CT screening, probably due to low incidence for interval cancers combined with limited sample size of most studies. Our study is the only that reports on radiological characteristics of interval cancers and the causes of the failure to detect interval cancers. In both our study and the literature, observations were made which suggests that interval cancers have different histopathology and are more aggressive than screening-detected lung cancers.

DECLARATION OF INTERESTS

NH, HJdK, CMvdA, KtH, and UAY-K report grants from ZonMw, grants from KWF, grants from Roche diagnostics, and non-financial support from Siemens Germany during the conduct of the study. Additionally, HJdK and KtH report grants from National Cancer Institute or National Institutes of Health and grants from The Agency for Healthcare Research and Quality during the conduct of the study, and grants from Sunnybrook Health Sciences and grants from Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen outside the submitted work. J-WJL reports grants from EU and grants from TiPharma outside the submitted work. KN reports grants from the Belgian Foundation against Cancer and grants from Flemish Cancer Ligue during the conduct of the study. All other authors declare no competing interests.

ACKNOWLEDGEMENTS

We thank M Quak, R Faber, F Santegoets, L van Dongen, and A de Bruijn (Erasmus University Medical Center Rotterdam, Netherlands), H Ziengs (University Medical Center Groningen, Netherlands), A Hamersma and S van Amelsvoort-van de Vorst (University Medical Center Utrecht, Netherlands), and 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 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

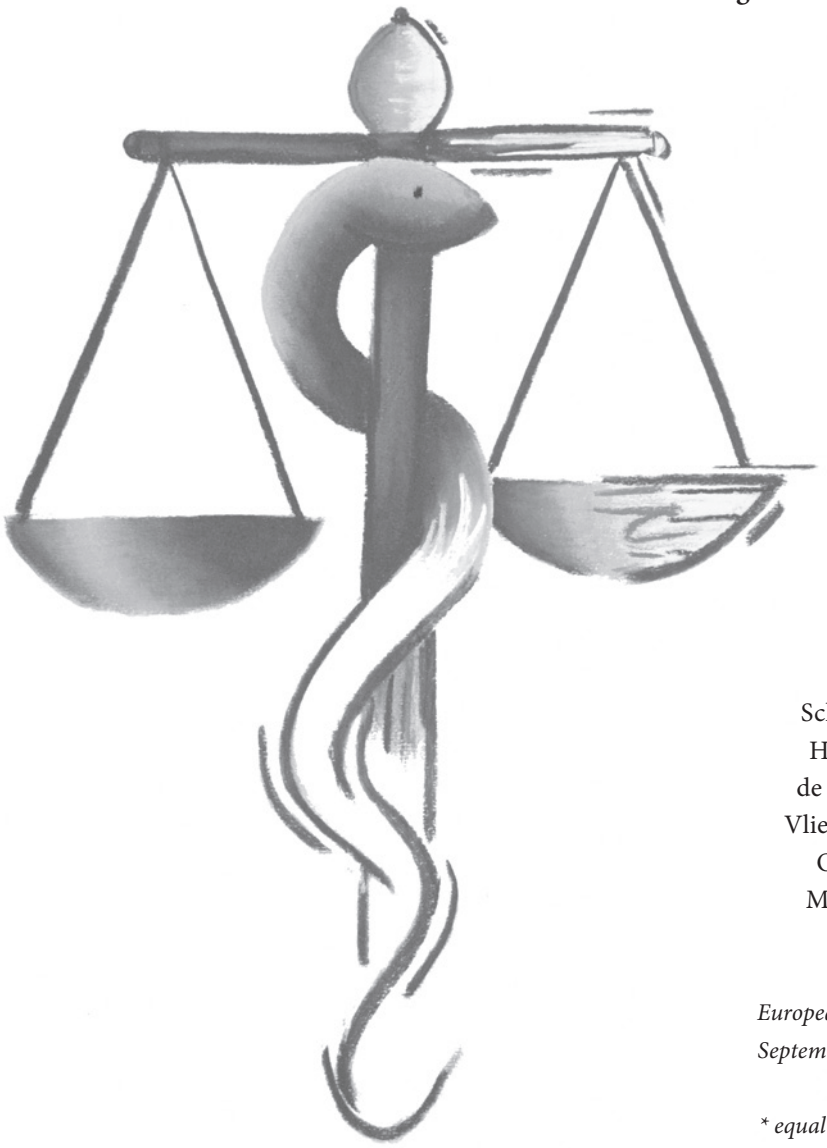
1. Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: screening and detection with low-dose CT versus radiography. *Radiology* 1996; 201(3):798-802.
2. Sone S, Takashina S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet* 1998; 351(9111):1242-5.
3. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 354(9173):99-105.
4. Pastorino U, Bellomi M, Landoni C, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. *Lancet* 2003; 362(9384):593-7.
5. Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: five-year prospective experience. *Radiology* 2005; 235(1):259-65.
6. Sone S, Li F, Yang ZG, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* 2001; 84(1):25-32.
7. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365(5):395-409.
8. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and Harms of Computed Tomography Lung Cancer Screening Strategies: A Comparative Modeling Study for the U.S. Preventive Services Task Force. *Ann Intern Med* 2013; 159(6):411-20.
9. Detterbeck FC, Mazzone PJ, Naidich DP, Bach PB. Screening for Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013; 143(5):e78S-92S.
10. Humphrey LL, Deffebach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med* 2013; 159(6):411-20.
11. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; 67(4):296-301.
12. Infante M, Chiesa G, Solomon D, et al. Surgical procedures in the DANTE trial, a randomized study of lung cancer early detection with spiral computed tomography: comparative analysis in the screening and control arm. *J Thorac Oncol* 2011; 6(2):327-35.
13. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev* 2012; 21:308-15.
14. National Lung Screening Trial Research Team, Church TR, Black WC, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 2013; 368(21):1980-91.
15. Aberle DR, DeMello S, Berg CD, et al. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med* 2013; 369(10):920-31.
16. Becker N, Motsch E, Gross ML, et al. Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. *J Cancer Res Clin Oncol* 2012; 138(9):1475-86.
17. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361(23):2221-9.
18. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54(2):177-84.
19. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J* 2013; 42(6):1659-67.

20. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120(4):868-74.
21. Horeweg N, van Klaveren RJ, Groen HJ, et al. Blinded and uniform cause of death verification in a lung cancer CT screening trial. *Lung Cancer* 2012; 77(3):522-5.
22. Bankier AA, Kressel HY. Through the Looking Glass revisited: the need for more meaning and less drama in the reporting of dose and dose reduction in CT. *Radiology* 2012; 265(1):4-8.
23. Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. *Cancer Imaging* 2011; 11(A):S79-84.
24. van Meerbeeck JP, Koning CC, Tjan-Heijnen VC, Boekema AG, Kaandorp CJ, Burgers JS. Guideline on 'non-small cell lung carcinoma; staging and treatment'. *Ned Tijdschr Geneeskd* 2005; 149(2):72-7.
25. van 't Westeinde SC, Horeweg N, Vernhout RM, et al. The role of conventional bronchoscopy in the workup of suspicious CT scan screen-detected pulmonary nodules. *Chest* 2012; 142(2):377-84.
26. van't Westeinde SC, de Koning HJ, Thunnissen FB, et al. The role of the (18-fluorodeoxyglucose)-positron emission tomography scan in the Nederlands Leuvens Longkanker screenings Onderzoek lung cancer screening trial. *J Thorac Oncol* 2011; 6(10):1704-12.
27. Nederland IKC. The Dutch Cancer Registry. 2011. http://www.cijfersoverkanker.nl/about-cancer-registry-37.html Date last accessed 2 September 2014.
28. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer* 2009; 64(1):34-40.
29. Xie X, Zhao Y, Snijder RA, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. *Eur Radiol* 2013; 23(1):139-47.
30. Horeweg N, Nackaerts K, Oudkerk M, de Koning HJ. Low-dose computed tomography screening for lung cancer: results of the first screening round. *J Comp Eff Res* 2013; 2(5):433-6.
31. Duffy SW, Field JK, Allgood PC, Seigneurin A. Translation of research results to simple estimates of the likely effects of a lung cancer screening programme in the United Kingdom. *Br J Cancer* 2014; 110:1834-40.
32. Scholten ET, Horeweg N, de Koning HJ, et al. Computed tomographic characteristics of interval and post screen carcinomas in lung cancer screening *Eur Radiol* 2014 Sep 4. [Epub ahead of print].
33. Henschke CI, Yip R, Yankelevitz DF, Smith JP, International Early Lung Cancer Action Program Investigators. Definition of a positive test result in computed tomography screening for lung cancer: a cohort study. *Ann Intern Med* 2013; 158(4):246-52.
34. Horeweg N, van der Aalst CM, Thunnissen E, et al. Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *Am J Respir Crit Care Med* 2013; 187(8):848-54.
35. Pinsky PF, Church TR, Izmirlian G, Kramer BS. The National Lung Screening Trial: results stratified by demographics, smoking history, and lung cancer histology. *Cancer*

Chapter 5

Radiological evaluation

Computed tomographic characteristics of interval and post screen carcinomas in lung cancer screening



Scholten ET*,
Horeweg N*,
de Koning HJ,
Vliegenthart R,
Oudkerk M,
Mali WPTM,
de Jong PA

European Radiology
September 14th 2014

** equal contribution*

Computed tomographic characteristics of interval and post-screen cancers in lung cancer screening

Eur Radiol. 2014 Sep 4 [Epub ahead of print].

ABSTRACT

Objective of this study was to analyse computed tomography (CT) findings of interval and post-screen cancers in lung cancer screening.

Consecutive interval and post-screen cancers from the Dutch-Belgium lung cancer screening trial were included. The prior screening and the diagnostic chest CT were reviewed by two experienced radiologists in consensus with knowledge of the tumour location on the diagnostic CT.

Sixty-one participants (53 men) were diagnosed with an interval or post-screen cancer. Twenty-two (36%) were in retrospect visible on the prior screening CT. Detection error occurred in 20 and interpretation error in 2 cancers. Errors involved intrabronchial tumour (n=5), bulla with wall thickening (n=5), lymphadenopathy (n=3), pleural effusion (n=1), and intra parenchymal solid nodules (n=8). These were missed due to a broad pleural attachment (n=4), extensive reticulation surrounding a nodule (n=1) and extensive scarring (n=1). No definite explanation other than human error was found in two cases (n=2). None of the interval or post-screen cancers involved a sub-solid nodule.

Interval or post-screen cancers that were visible in retrospect were mostly due to detection errors of solid nodules, bulla wall thickening or endobronchial lesions Interval or post-screen cancers without explanation other than human errors are rare.

INTRODUCTION

Early detection of lung cancer by low dose computed tomography (CT) scanning in asymptomatic smokers at high risk for developing lung cancer is a promising strategy to reduce lung cancer mortality. Several randomised lung cancer screening trials were conducted using low-dose CT scanning of high-risk groups, with the aim to detect lung cancer at an early and curable stage.[1-6] The National Lung Screening Trial (NLST) reported in 2011 a 20.0% decrease in lung cancer mortality when comparing CT screening with chest radiography screening.[7]

Fast growing tumours, protocol inadequacies and protocol violations and missed cancers on CT may result in an interval cancer. Interval cancers are cancers diagnosed between screening rounds after a negative or indeterminate screening result (defined as no recommendation for referral) or after a positive screen in which diagnostic work-up did not yield the diagnosis of cancer. Interval cancers may be missed or may arise during the screening interval. Missed cancers may be caused by detection and interpretation errors. In detection errors, the lesion is not mentioned in the report but can be seen in retrospect on the last CT. While in interpretation errors, the lesion was noted but considered a benign lesion. Post-screen cancers are lung cancers diagnosed after the last scheduled screening CT of the participant. In this study, interval cancers were distinguished from post-screen cancers; both are subdivided in radiological detection errors, interpretation errors and other causes (e.g. normal screening examination, or non-compliance participant).

Missed cancers in CT-based lung cancer screening trials have received limited attention in radiological literature.[8] In 1999, Kakinuma et al[9] concluded on a study of seven interval cancers, that minute nodules may be missed at spiral CT exams with a slice thickness of 10 mm. Further, Li et al[10] reported in 2002 a study of 32 missed lung cancers in a CT screening setting (using 10-mm slice thickness) that the missed cancers were very subtle, appeared as small faint nodules, and 92% of their 20 detection errors involved sub-solid nodules. Henceforth, many studies were published on Computer Aided Diagnosis (CAD) systems for detection of pulmonary nodules and CT equipment improved substantially.[11-18]

Purpose of the present study was to analyse CT findings in post-screen and interval cancers of the NELSON trial, focussing on CT findings in cases with radiological detection and interpretation errors. This is the first study reporting on missed lung cancers in a lung cancer screening program using multi-detector CT equipment and thin-slice reconstruction.

METHODS

This is an ancillary study of NELSON trial[1], which was approved by the Dutch Ministry of Health and ethical boards of participating hospitals. Written informed consent was obtained from each participant. Screening was initiated in 2004. Study population comprised of current and former smokers aged 50 to 75 years, with a smoking history of 15 or more cigarettes per day during more than 25 years, or 10 or more cigarettes per day during more than 30 years. Former smokers were included only if they quit smoking less than 10 years before start of the study. Exclusion criteria were: self-reported moderate or poor health status, inability to climb two flights of stairs, a chest CT within the last twelve months, body weight 140 kg or more, history of lung cancer in the last five years, history of melanoma, breast cancer or hypernephroma, and a previous pneumonectomy.

CT scanning and reading protocol

In participants randomised to the screening group, CT screening was performed at baseline, 1 year and 3 years and 5.5 years after baseline, plus additional follow-up CT exams in case indeterminate nodules were detected.[19] Multi-detector scanners (Somatom Sensation 16, Siemens Medical Solutions, Mx8000 IDT or Brilliance-16, Philips Medical Systems, Cleveland, OH) were used with 16x0.75mm collimation and 1.3 pitch. Unenhanced full inspiration CTs were acquired using 30mAs at 120kVp for subjects weighing 80kg or less, and 30mAs at 140kVp for those weighing more than 80kg. Axial 1.0mm images were reconstructed at 0.7mm increment using a 512x512 matrix, with a moderately soft kernel and the smallest field of view that included both lungs.

All CTs were analysed for non-calcified nodules. Detected nodules were characterised as solid nodule or sub-solid nodule, the latter being either pure or part-solid. At each site, CT data were analysed by the local radiologist with 1 year to more than 20 years of experience with thoracic CT. Subsequently, CT data were independently analyses by a second central reader with more than 6 years of experience. One type of digital workstation (Leonardo, Siemens Medical Solutions) with software for nodule identification and semi-automated volume measurements (LungCare, version Somaris/5 VA70C-W, Siemens Medical Solutions) was used; its use for nodule detection was not obligatory. After the radiologist marked a potential nodule with a mouse click, the program defines a volume of interest around the nodule which can further be analysed by volume rendering displays or multi-planar reformations. Once a potential nodule was approved, a second mouse-click initiated the automatic volume measurement program. In case of discrepancy, the radiologists tried to reach a consensus about the reading. If no consensus was reached, a third reading was performed by an expert radiologist with over twenty years of experience, who made the final decision.

Participants who were referred to a pulmonologist underwent diagnostic work-up which included a standard dose CT with intravenous contrast, bronchoscopy and/or biopsy. Based on results of these exams, the pulmonologist decided whether resection of the suspicious nodule was appropriate.

Study population

For the present study, all 7,155 participants (1254 females, 16.5%) randomised to the CT screening group of the participating Dutch screening centres (University Medical Center Groningen, University Medical Center Utrecht and Kennemer Gasthuis, Haarlem, The Netherlands) were included. Belgian participants (n=935) had to be excluded as no data on interval cancers were available yet. Median age at baseline of included subjects was 58.0 years (interquartile range (IQR) 54.0-62.0), median number of pack-years was 38.0 (IQR 29.7-49.5), and 4215 participants (55.6%) were current smokers.

Interval and post-screen cancers

Some participants developed lung cancer a considerable time, up to six years, after their last attended screening examination. Since conclusions may be drawn from these cases, they were included in this study. Interval and post-screen cancers were identified through linkages with the Dutch Cancer Registry, which has complete national coverage.

In the first three screening rounds, 187 of the 7,155 (2.6%) included subjects were diagnosed with screen-detected lung cancer. Between or after screening examinations in the NELSON trial, 61 of 7,155 participants (0.85%) were diagnosed with interval or post-screen cancer; 53 men and eight women. Hence, a total of 248 screen-detected-, interval, - and post-screen cancers were diagnosed. Median age of these participants at the time of the diagnosis was 64 years (IQR 6 years).

Of the 61 participants with interval or post-screen cancer, clinical and radiological files were retrieved from the various hospitals where diagnosis of lung cancer was established. Also, their last available screening CT examination was reviewed and compared to the clinical CT at the time of the diagnosis. Two radiologists, one chest radiologist with 10 years of experience, and one general radiologist with over 30 years of experience with chest CT decided in consensus whether or not lung cancer or CT evidence of metastatic disease (such as mediastinal or bone metastases) could in retrospect be identified on the screening CT, and whether it was not mentioned or misinterpreted in the original report in the trial database. Furthermore, significant other pathology that might has influenced the original reading was noted as well. Depending on the findings noted in the trial database, missed cancers were classified as either a detection error or an interpretation error. An error was considered a detection error if no mention of the lesion was found in the trial database and an interpretation error if the lesion was mentioned but the potentially

malignant character not recognized. An attempt was made to formulate reasons why the abnormality was not detected or misinterpreted by the screening radiologists.

Data analysis

Descriptive statistics were used to analyse and present the data.

RESULTS

Based on consensus reading, 26 of 61 cases (42.6%) had a normal last screening CT examination and the screening protocol was not violated. In 11 of these 26 cases (42.3%) the lung cancer was considered an interval cancer as it was diagnosed before the next scheduled screening CT examination, the remaining 15 cases (57.7%) were considered a post-screen cancer as they were diagnosed after the screening program was finished (Table 1).

Table 1. CT findings of interval and post screen cancers: causes and length of delay

Findings	Interval cancer n (%)	Days since last CT at diagnosis median (range)	Post-screen cancer n (%)	Days since last CT at diagnosis median (range)
Screening CT examination normal:				
Protocol followed	11 (18.0)	425 (169-676)	15 (24.6)	817 (202-2037)
Cancer not treated due to:				
Protocol inadequate	1	257	1	349
False-negative work-up	0	NA	3	815 (68-1140)
Non-compliance participant	2	(373-461)	6	1805 (1319-2179)
Cancer not detected due to detection error:				
Intrabronchial localisation	5	367 (232-646)	0	NA
Adjoining bullous structure	1	358	3	890 (735-1290)
Lymphadenopathy	3	310 (217-436)	0	NA
Pleural effusion	1	177	0	NA
Extensive fibrotic changes	0	NA	1	311
Small pleural attachment	1	581	1	1515
Large pleural attachment	1	234	1	1089
Probably human error	0	NA	1	323
Cancer not detected due to interpretation error:				
Large nodule classified as scarring	1	192	0	NA
Adjoining bullous structure	1	562	0	NA
Total	**29 (47.5)**		**32 (52.5)**	

Definition of abbreviations: CT = computed tomography; NA = not applicable.

Table 2. Characteristics of participants with 22 missed lung cancers

Participant characteristics	**n (%)**
Age at diagnosis - median (range)	64.0 yrs (56-76 yrs)
Male gender	21 (95.5)
Current smoker	15 (68.1)
Pack-years - median (range)	49.5 (28.0-123.5)
Computed tomography characteristics	**n (%)**
Tumour size at diagnosis	
>5 cm	11 (50.0)
<5 cm	10 (45.5)
Not measurable	1 (4.5)
Tumour localisation	
Left upper lobe	6 (27.3)
Left lower lobe	4 (18.2)
Right upper lobe	8 (36.4)
Middle lobe	1 (4.5)
Right lower lobe	3 (13.6)
Tumour type	
Solid	17 (77.3)
Non-solid	0 (0.0)
Bulla wall thickening	5 (22.7)
Underlying lung disease	
Fibrosis	1 (4.5)

Definition of abbreviations: yrs = years.

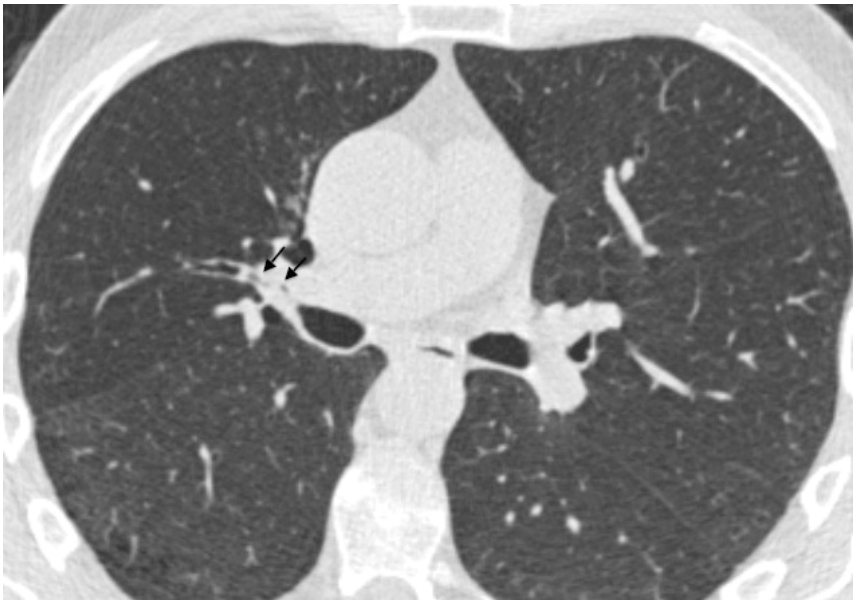
In 13 of 61 cases (21.3%), lung cancer was not diagnosed through screening due to a variety of reasons: participant drop-out (n=8, 61.5%); two (15.4%) were interval cancers and six (46.2%) post-screen cancers. Lung cancer was not diagnosed through screening after a previous false-negative work-up by the pulmonologist (n=3, 23.1%); both were post-screen cancers. In the remaining two of 13 cases (15.4%), the protocol was considered inadequate as it was adhered to but the malignant nodules were not classified as positive. Hence, one 13mm nodule failed to show growth on at follow-up scanning after 3 months (later diagnosed as interval cancer), and another 12 mm nodule was considered stable over a period of three years, but was later on diagnosed as a post-screen cancer.

The remaining 22 of 61 cases (36.1%) were 15 interval cancers and 7 post-screen cancers; the radiological abnormality was either not detected (in 20 cases) or misinterpreted (in 2 cases). These 22 cases were 0.31% of the total study population of 7,155 participants, and 8.9% of 248 lung cancers.

Missed endobronchial abnormalities

In 5 of the 22 (22.7%) cases wherein the abnormality was not detected or misinterpreted, a central intra-bronchial tumour was overlooked on the screening examination. All were small, although difficult to measure, estimated to be about 5 mm (Figure 1). Four of the endobronchial tumours were right-sided: two localised in the pectoral segmental bronchus, one in the lateral segmental bronchus, and one in the right upper lobe bronchus. One endobronchial tumour was localised left at the lingular bronchus. In one of the cases, a note was made that lymphadenopathy was present, but no further action was taken. All five cancers were classified as interval cancers. Median number of days since the last screening CT at the time of diagnosis was 367 days (range 232-646).

Figure 1. Example of missed endobronchial lung cancer



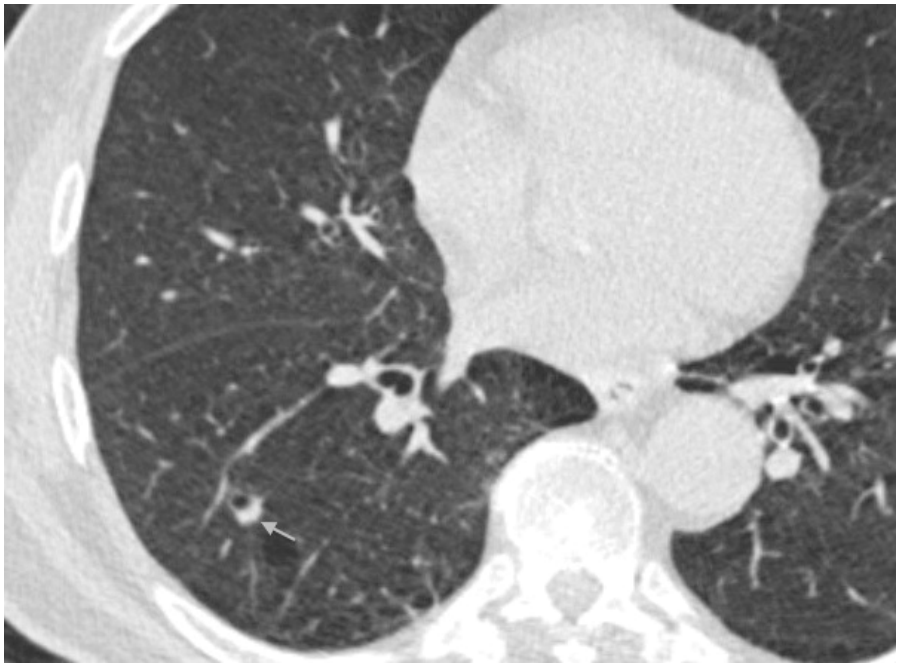
Narrowing of right upper lobe segmental bronchus (arrows).

Missed focal bulla wall thickenings

In 5 of 22 cases (22.7%), a bulla with a thickening of the wall was noted during the consensus meeting that was not reported in the database. In four of these cases, the thickening of the wall was already visible at the first screening CT examination. In one case, the bulla with wall thickening developed in a previously normal lung. The five lesions were evenly distributed over the lungs: two in the right upper lobe, and one in the left upper lobe, right lower lobe and left lower lobe. In two cases the wall thickening was focal; in one of these cases a 7 mm nodule was noted in the trial database, so this was considered an

interpretation error as no mention was made about the adjoining bulla (Figure 2). Of the five lung cancers, two were classified as interval cancers and three as post-screen cancers; the latter were all detected more than two years after the last screening CT.

Figure 2. Example of missed lung cancer in bulla wall



Small (7 mm) elliptical nodule in bulla wall (arrow).

Missed lymphadenopathy

In 3 of 22 cases (13.6%), lymphadenopathy was missed: two were localised in a slightly enlarged right hilum, in which it was without intra venous contrast inseparable from the right pulmonary artery. One was mainly localised in the aortopulmonary window, maximal diameter was 22 mm. All three cases were classified as interval cancers. Time since the last screening CT at the time of diagnosis was 217, 310 and 436 days.

Missed pleural effusions

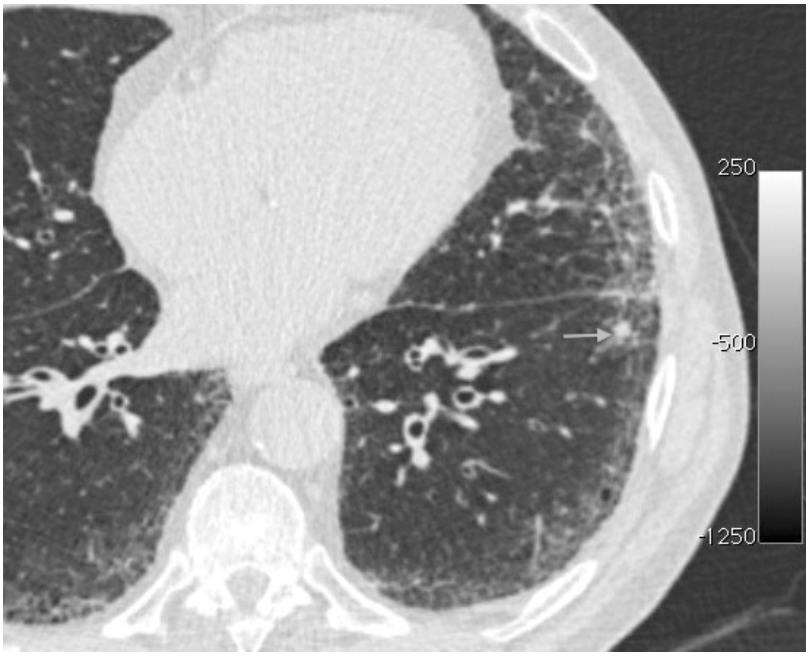
In one of 22 cases (4.5%), right-sided pleural fluid remained unnoticed on the screening CT examination. An interval cancer was diagnosed 177 days after the last CT, presenting as a large carcinoma with massive pleural effusions.

Missed cancers due to other reasons

The remaining 8 of 22 cases (36.4%) were not detected due to various causes. Small nodule size might have played a role in three cases, as small nodules of 7, 7 and 5 mm were not detected. However, two of these nodules were also attached to the pleura, which may also have played a role (delay of diagnosis were 581 days for the interval cancer, and 1515 days for the post-screen cancer). One nodule was surrounded by extensive reticulation (Figure 3), which probably caused the detection error (delays of diagnosis of this post-screen cancer was 311 days).

Further, five larger (>1 cm) nodules were also not detected: two in the left lower lobe, and one in the left upper lobe, right upper lobe and right lower lobe. Three of these large nodules were classified as an interval cancer, two as a post-screen cancer. One of these five nodules was a 22 mm-large pleural-attached nodule that was interpreted by the screening radiologists as scarring (Figure 4); consequently this lung cancer was considered missed due to interpretation error. Two of five lesions were broadly pleural-based (Figure 5). Two other of the five larger nodules were not attached to the pleura. As no obvious reason for not detecting these lesions was found, these detection failures were attributed to human error. Median number of days since the last CT at the time of diagnosis was 234 days (range 96-1089) for the five larger nodules.

Figure 3. Example of missed lung cancer due to distractive other pathology



Small nodule in left under lobe hidden in reticulation (arrow).

Figure 4. Example of missed lung cancer due to interpretation error

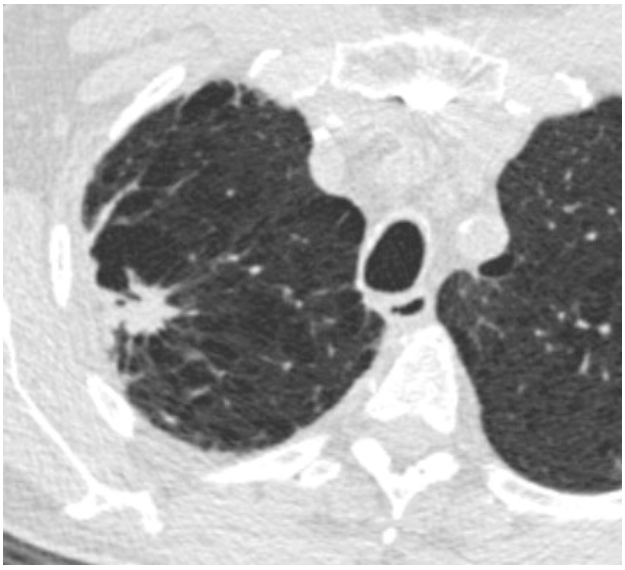
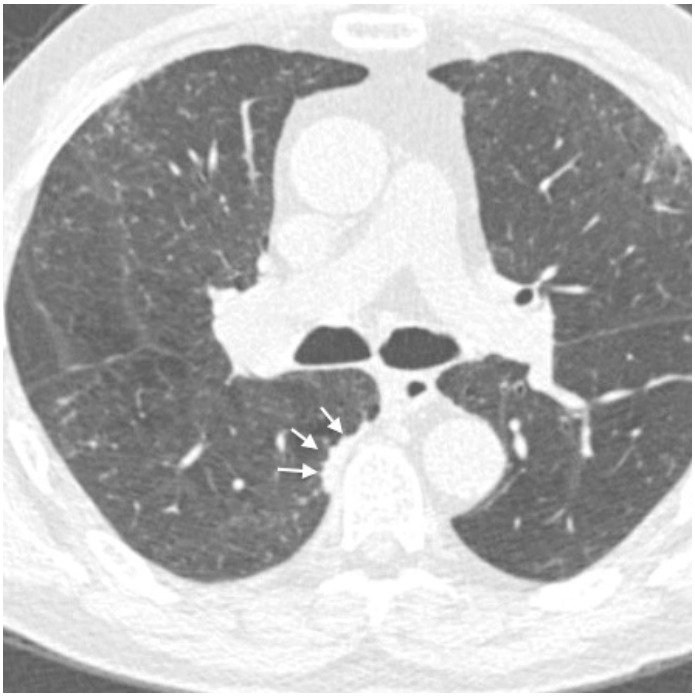


Image suggestive of lung cancer, but interpreted as fibrotic scarring.

Figure 5. Example of missed lung cancer attached to the pleura



Prevertebral broad-based tumour on the right (arrows).

DISCUSSION

In this study, the radiological characteristics and causes of the failure to detect interval cancers and post-screen cancers in the NELSON trial were investigated. The majority of the 61 (n=39, 64%) interval and post-screen cancers were not due to radiological detection or interpretation errors. A minority of 22 (36%) cancers was visible in retrospect at the last screening examination and missed. Most missed cancers were due to detection errors of a nodule either localised in a bronchus, attached to a bulla or sub-pleural in the lung parenchyma. According to the protocol, these nodules (with possibly exception of the bulla wall thickening) should have been followed up by either a repeat CT scan after three months or by referral to a pulmonologist for further evaluation. Interpretation errors seemed to have played a minor role in missed lung cancers. Findings of this study may aid improving lung cancer detection in lung cancer screening, although the predictive value of some findings, especially bulla wall thickenings, need to be determined yet.

Limited number of studies have been published on interval, - and post-screen cancers in a CT-based lung cancer screening setting. In contrast to the findings of Li et al.[10], who reported in 2002 that 92% of their missed cancers were non-solid, none of the missed nodules were part or pure non-solid. This can at least partly be explained by the considerable difference in spatial resolution of the 1 mm slice thickness used in the NELSON study, compared to the 10 mm slice thickness in the study by Li et al.[10] However, since sub-solid nodules may grow very slowly, it cannot be excluded that an interval cancer arising from a sub-solid nodule will manifest in a longer follow-up period.

In this study, the most common detection error was missing endobronchial lesions. This is probably because endobronchial nodules are far less common in a screening population than intraparenchymal nodules. As a result, attention of screening radiologists was probably primarily focused on the lungs and not the bronchi. This is not compensated by the CAD-system its search for nodules does not include the bronchi. In 1996, White et al.[20] reported on 14 primary lung cancers overlooked on CT in a clinical setting; 67% were at a central endobronchial location.[20] White et al. gave a similar explanation, not focusing on central airways, for detection error in their series. Another important factor at the time of their study was the use of 5 mm or even thicker sections. Computer Aided Diagnosis of lung cancer in the bronchi has had considerable attention in the literature[21] and as a training tool.[22] Extension of lung cancer detection CAD systems to the bronchi may prove helpful in reducing these detection errors. However, until this extension is realised extra focus on the bronchial tree is warranted.

Another common characteristic of missed lung cancers was a thickened bulla wall. This entity was not recognised as an important abnormality at the start of the NELSON screening trial in 2004. Therefore, bulla wall thickenings were not a pre-defined abnormality in the nodule management system. In 2010, Keneda et al.[23] described clinical features of

primary lung cancer adjoining bullae. In their retrospective study of 545 clinical cases who underwent surgery for lung cancer, they identified an adjoining bulla in 19 cases (3.5%)[23], which suggests that this finding is not uncommon. Keneda et al. also state that the association of bullae and lung cancer is not well recognised. This was confirmed in this study, as it was identified as one of the main causes for detection errors. Hence, in one of five cases with nodular thickening in the wall of a bulla the abnormality was noted and its nodule features were described in the database, but no mention was made of the adjoining bulla. In 2012, Farooqi et al. reported on lung cancers associated with cystic airspaces in the Early Lung Cancer Action Program.[24] They found that in their baseline and annual screening series respectively 25% and 12% of lung cancers were associated with cystic airspaces. They concluded that the finding of an isolated cystic air space with increased wall thickness at annual repeat CT screening is suspicious for lung cancer. Since no data on the prevalence of bulla wall thickening in the NELSON population was collected, no positive predictive value of this finding can be estimated. However, findings of this study justify increased attention to focal and diffuse bulla wall thickenings in lung cancer screening.

Intraparenchymal nodules were, with 8 cases (36%), the most common cause for a missed cancer. Two smaller and two larger lesions were probably missed by the screening radiologist due to pleural-attachment or a broad shape. One case of these cases was interpreted as fibrotic scarring, which was classified as an interpretation error. White et al.[20] reported in a clinical series of 14 primary lung cancers overlooked on CT that 6 of 14 cases (43%) were due to major distractive findings elsewhere in the chest, such as aortic aneurysm or large oesophageal tumour. In the current study, similar cases were not common. Only one lesion was missed due to extensive reticulation in its immediate surroundings. In two cases of a missed large intrapulmonary nodule no plausible reason other than human error could be found.

Lymphadenopathy is more difficult to detect on low-dose screening CTs without intravenous contrast than on clinical contrast enhanced CT's. Moreover, the screening radiologist's focus is primarily on the lungs, and significant lymphadenopathy uncommon compared to in a clinical setting. Concluding, missed lymphadenopathy, which was responsible for 13.6% of missed lung cancers, is probably difficult to prevent.

The most important limitation of this study was the lack of data on the prevalence of abnormalities such as bulla wall thickening. As a result, it is not possible to determine the positive predictive value of the characteristics of missed lung cancers. This problem may be resolved as screening programs check the CT for abnormalities such as bulla wall thickenings, and report them as a separate item. Another limitation was that the total number of cancers found in the NELSON study is not known at present, so this number cannot be related to the number of interval cancers. However, this study suggests that interval-, or post-screen cancers due to human errors were rare, as it only concerned 0.31%

of the total screen population. A third limitation is the inherent focus on cancers not detected by screening. Performing the same analysis of cancers found in during screening may learn whether and how screen-detected cancers can be detected in earlier screening rounds. Finally, limited experience of some screening radiologists with thoracic CT could have been a limitation of the study. However, this is not supported by a previous study on the benefit of consensus double reading in screening for lung cancer.[25]

In conclusion, interval-, and post-screen cancers in the NELSON trial that were visible in retrospect, were mostly due to detection errors of solid nodules. Thickening of a bulla wall should be looked at with suspicion, at least until more of the natural course of such lesions is known. Detection of endobronchial lesions might improve with extension of CAD systems to the bronchi.

SUPPORT STATEMENT

The NELSON study has received funding by: "Zorg onderzoek Nederland - Medische wetenschappen" (Zon-Mw), "KWF Kankerbestrijding", "Stichting Centraal Fonds van Reserves voor Voormalig Vrijwillige Ziekenfondsverzekeringen"(RvvZ), "G. Ph. Verhagen Foundation", "Rotterdam Oncologic Thoracic Study Group" (ROTS) and "Erasmus Trust Fund", "Stichting tegen Kanker", "Vlaamse Liga tegen Kanker", and LOGO Leuven en Hageland".

CONFLICTS OF INTEREST

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

REFERENCES

1. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120(4):868-74.
2. National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. *Radiology* 2011; 258(1):243-53.
3. Pedersen JH, Ashraf H, Dirksen A et al. The Danish randomized lung cancer CT screening trial--overall design and results of the prevalence round. *J Thorac Oncol* 2009; 4(5):608-14.
4. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer* 2008; 59(3):355-63.
5. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer* 2009; 64(1):34-40.
6. Diederich S, Wormanns D, Semik M, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. *Radiology* 2002; 222(3):773-81.
7. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365(5):395-409.
8. Fardanesh M, White C. Missed lung cancer on chest radiography and computed tomography. *Semin Ultrasound CT MR* 2012; 33(4):280-7.
9. Kakinuma R, Ohmatsu H, Kaneko M, et al. Detection failures in spiral CT screening for lung cancer: analysis of CT findings. *Radiology* 1999; 212(1):61-6.
10. Li F, Sone S, Abe H, MacMahon H, et al. Lung cancers missed at low-dose helical CT screening in a general population: comparison of clinical, histopathologic, and imaging findings. *Radiology* 2002; 225(3):673-83.
11. Awai K, Murao K, Ozawa A, et al. Pulmonary nodules: Estimation of malignancy at thin-section helical CT - Effect of computer-aided diagnosis on performance of radiologists. *Radiology* 2006; 239:276-284.
12. Lee Y, Hara T, Fujita H, et al. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Trans Med Imaging* 2001; 20:595-604.
13. Armato SG 3rd, Roy AS, Macmahon H, et al. Evaluation of automated lung nodule detection on low-dose computed tomography scans from a lung cancer screening program. *Acad Radiol* 2005; 12:337-346.
14. Giger ML, Bae KT, MacMahon H. Computerized detection of pulmonary nodules in computed tomography images. *Invest Radiol* 1994; 29:459-465.
15. Li Q, Li F, Doi K. Computerized detection of lung nodules in thinsection CT images by use of selective enhancement filters and an automated rule-based classifier. *Acad Radiol* 2008; 15:165-175.
16. Godoy MC, Cooperberg PL, Maizlin ZV, et al. Detection sensitivity of a commercial lung nodule CAD system in a series of pathologically proven lung cancers. *J Thorac Imaging* 2008; 23:1-6.
17. Yuan R, Vos PM, Cooperberg PL. Computer-aided detection in screening CT for pulmonary nodules. *Am J Roentgenol* 2006; 186:1280-1287.
18. Goldin JG, Brown MS, Petkovska I. Computer-aided diagnosis in lung nodule assessment. *J Thorac Imaging* 2008; 23:97-104.
19. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J* 2013; 42(6):1659-67.
20. White CS, Romney BM, Mason AC, et al. Primary carcinoma of the lung overlooked at CT: analysis of findings in 14 patients. *Radiology* 1996; 199(1):109-15.

21. Hoppe H, Dinkel HP, Thoeny H, et al. Virtuelle Endoskopie der oberen, zentralen und peripheren Atemwege mit Mehrzeilen-Spiral-CT. *Radiologe* 2002; 42(9):703-11.
22. Colt HG, Crawford SW, Galbraith O 3rd. Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest* 2001; 120(4):1333-9.
23. Kaneda M, Tarukawa T, Watanabe F, et al. Clinical features of primary lung cancer adjoining pulmonary bulla. *Interact Cardiovasc Thorac Surg* 2010; 10(6):940-4.
24. Farooqi , Cham M, Zhang L, et al. Lung cancer associated with cystic airspaces. *Am J Roentgenol* 2012; 199(4):781-6.
25. Wang Y, van Klaveren RJ, de Bock GH, et al. No benefit for consensus double reading at baseline screening for lung cancer with the use of semiautomated volumetry software. *Radiology* 2012; 262(1):320-6.

Part III

Optimisation of screening

Chapter 6

Optimisation of screening protocols

Lung cancer probability in subjects with CT-detected pulmonary nodules: an analysis of data from the NELSON trial of low-dose CT screening

Horeweg N*,
van Rosmalen J*,
Heuvelmans MA,
van der Aalst CM,
Vliegenthart R,
Scholten ET,
ten Haaf K,
Nackaerts K,
Lammers JWJ,
Weenink C,
Groen HJM,
van Ooijen P,
de Jong PA,
de Bock GH,
Mali W,
de Koning HJ*,
Oudkerk M*

Lancet Oncology
November 2014

** equal contribution*

Lung cancer probability in subjects with CT-detected pulmonary nodules: an analysis of data from the NELSON trial of low-dose CT screening

Lancet Oncol. 2014 Nov;15(11).

ABSTRACT

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.

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.

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% [CI 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 mm³ 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]).

Small nodules (those with a volume <100 mm³ or diameter <5 mm) are not predictive for lung cancer. Immediate diagnostic evaluation is necessary for large nodules (≥300 mm³ or ≥10 mm). Volume doubling time assessment is advocated only for intermediate-sized nodules (with a volume ranging between 100–300 mm³ 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 subjects 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 either have lung cancer.[3,4] Validated guidelines to determine optimum patient management strategies based on characteristics of detected nodules are urgently needed.

At first, the accepted standard of practice was to regard all non-calcified pulmonary nodules detected at CT 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 or smaller in diameter in high-risk people 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 ionising 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 probability of individual participants' 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 volume of 8 mm^3 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 randomised 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.[15] 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.[8,13]

METHODS

Study design and participants

Details about the design and conduct of the NELSON trial have been reported previously.[15,16] Briefly, participants from four centres 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, 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 registries 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 baseline and one year later. We excluded Belgian participants because data about interval cancers were not yet available from the Belgian cancer registry; interval cancers from Dutch participants were identified with use of the Dutch Cancer Registry.

Procedures

The protocol describing how CT screening was done in the NELSON trial has been previously published,[16] and is summarised 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).[16] Datasets were derived from images of the thorax (slice thickness 1 mm, interval 0.7 mm) and analysed with software for semi-automated volume measurements (LungCARE, Somaris/5 VB 10A-W, Siemens).[16]

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

$$V = \frac{1}{6}\pi D^3$$

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:

$$\mathrm{VDT} = \frac{\ln(2)\Delta t}{\ln(V_2) - \ln(V_1)}$$

in which Δt 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:

$$VDT = \frac{\ln(2)\Delta t}{3 * \ln(MaxDiamXY_2 / MaxDiamXY_1)}$$

in which Δt represents time in days between scans, and MaxDiamXY1 and MaxDiamXY2 are maximum diameters on the X-Y axis at first and second assessment.[16] 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.[8,13] We predicted lung cancer risk in the two years following each screening round using regression analysis 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 undergoing intensified CT surveillance[8] besides screening. Participants with a significantly increased lung cancer probability (but less than about 5%; adopted from ACCP guideline[13]) 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).[13] Furthermore, the ACCP management protocol[8,13] (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,[12] 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.[15,17] The primary aim of this study was to quantify the probability of developing lung cancer within two years after the screening round, 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 the number of cases with cancer by the total number of cases per stratum. Differences between lung cancer probabilities were tested using with Fisher's exact test; 95% CIs were calculated using the Agresti-Coull method.

To predict lung cancer risk in the two 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.[18] We used a closed-test procedure to control the family-wise type I error rate in a situation with multiple testing.[19] 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 (details provided in the 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 (more details provided in the 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).

RESULTS

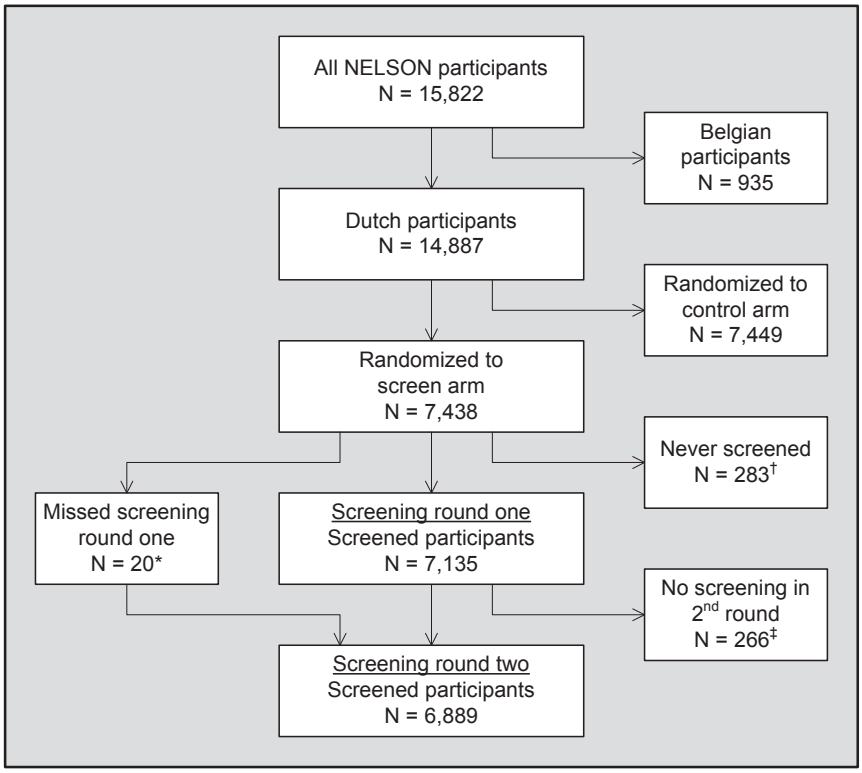
Participants

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 Jan, 2004 to Dec, 2006, and screening round two from Jan, 2005 to Sept, 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 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, 4165 (56%) were current smokers, and their median number of pack-years at randomisation was 38 (IQR 19-57).

Figure 1. Participant flowchart



Screening round one was conducted from January 2004 to December 2006 and screening round two from January 2005 to September 2008.

* *20 Dutch participants missed the baseline CT due to late returning of their informed consent.*

† *283 Dutch participants were randomised 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).*

Quantifying lung cancer probability

Two-year lung cancer probability for all included participants was 1.3% (95% CI 1.2-1.5, Table 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% (95% CI 0.3-0.6). In all participants with CT-detected nodules, lung cancer probability was 2.5% (95% CI 2.1-2.9%), but individuals' probabilities depended strongly on nodule volume, diameter and volume doubling time (Table 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 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% (95% CI 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 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%] vs. 0.4% [0.3-0.6], 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 at being at intermediate and high risk of developing lung cancer, respectively.

The probability of being diagnosed with lung cancer within two years after CT scan according to nodule volume doubling time for the participants whose largest nodule measured 50-500 mm^3 is presented in Table 1. Participants with slowly-growing (volume doubling time ≥600 days), stable, shrunken, or resolved 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% [95% CI 0.4-1.7], p = 0.06. Lung cancer probability was significantly increased for participants with nodule volume doubling times of times 400-600 days (4.0% [95% CI 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% [95% CI 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 (such as stable, shrinking, and resolving nodules) were done,

Table 1a. Probability of lung cancer diagnosis within two years after a screening test, by volume of largest nodule

Volume of largest nodule in mm³†	Round 1		Round 2		Rounds 1 and 2		Lung cancer probability+	
	Cases n	All n	Cases n	All n	Cases n	All n	percentage (95%CI)	p value
≥1000	36	137	26	104	62	241	25.7 (20.6-31.6)	<0.0001
750 - 1000	8	33	4	30	12	63	19.0 (11.1-30.6)	<0.0001
500 - 750	8	63	4	47	12	110	10.9 (6.2-18.2)	<0.0001
300 - 500	12	101	6	102	18	203	8.9 (5.6-13.7)	<0.0001
200 - 300	9	127	5	116	14	243	5.8 (3.4-9.5)	<0.0001
100 - 200	6	428	7	440	13	868	1.5 (0.9-2.6)	0.0002
50 - 100	6	800	6	843	12	1,643	0.7 (0.4-1.3)	0.07
25 - 50	6	961	4	1,008	10	1,969	0.5 (0.3-0.9)	0.44
<25	3	539	2	515	5	1,054	0.5 (0.2-1.1)	0.61
No nodule detected	15	3,946	15	3,684	30	7,630	0.4 (0.3-0.6)	ref
All participants	**109**	**7,135**	**79**	**6,889**	**188**	**14,024**	**1.3 (1.2-1.5)**	**<0.0001**

Definition of abbreviation: 95%CI = 95% confidence interval; ref = reference value.

† *Volume of the largest non-calcified nodule in a participant in mm³, the interval includes the lower limit, not the upper limit.*

+ *Probability of malignancy within two years after a CT scan. The difference in lung cancer risk with subjects without nodules was evaluated using Fisher's exact test.*

Table 1b. Probability of lung cancer diagnosis within two years after a screening test, by volumetry-based diameter of largest non-calcified nodule

Max. diameter of largest nodule†	Round 1		Round 2		Rounds 1 and 2		Lung cancer probability+	
	Cases n	All n	Cases n	All n	Cases n	All n	percentage (95%CI)	p value
≥30	3	10	3	9	6	19	31.6 (15.2-54.2)	<0.0001
20 - 30	13	52	9	36	22	88	25.0 (17.1-35.0)	<0.0001
15 - 20	22	84	7	64	29	148	19.6 (14.0-26.8)	<0.0001
10 - 15	28	229	21	213	49	442	11.1 (8.5-14.4)	<0.0001
8 - 10	7	260	9	296	16	556	2.9 (1.7-4.7)	<0.0001
7 - 8	8	327	4	328	12	655	1.8 (1.0-3.2)	<0.0001
6 - 7	1	371	2	331	3	702	0.4 (0.1-1.3)	0.75
5 - 6	7	628	5	721	12	1,349	0.9 (0.5-1.6)	0.03
4 - 5	3	799	1	776	4	1,575	0.3 (0.1-0.7)	0.50
<4	2	429	3	431	5	860	0.6 (0.2-1.4)	0.40
No nodule detected	15	3,946	15	3,684	30	7,630	0.4 (0.3-0.6)	ref
All participants	**109**	**7,135**	**79**	**6,889**	**188**	**14,024**	**1.3 (1.2-1.5)**	**<0.0001**

Definition of abbreviation: 95%CI = 95% confidence interval; ref = reference value.

† *Maximum diameter of the largest nodule in a participant in mm, the interval includes the lower limit, not the upper limit. Estimates based on diameters assessed using semi-automated volumetry. Manually measured diameters are less accurate and will overestimate nodule size, which corresponds with lower lung cancer probabilities than presented in this table.*

+ *Probability of lung cancer within two years after a CT scan. The difference in lung cancer risk with subjects without nodules was evaluated using Fisher's exact test.*

Table 1c. Probability of lung cancer within two years by VDT of fastest growing nodule

VDT of fastest growing nodule in days†	Round 1		Round 2		Rounds 1 and 2		Lung cancer probability+	
	Cases n	All n	Cases n	All n	Cases n	All n	percentage (95%CI)	p value
<100	7	24	2	10	9	34	26.5 (14.4-43.3)	<0.0001
100 - 200	3	40	3	16	6	56	10.7 (4.7-21.8)	<0.0001
200 - 400	5	130	7	52	12	182	6.6 (3.7-11.3)	<0.0001
400 - 600	3	92	4	81	7	173	4.0 (1.8-8.3)	<0.0001
600 - 800	0	56	0	74	0	130	0.0 (0.0-3.4)	1.00
800 - 1000	0	45	1	63	1	108	0.9 (0.0-5.6)	0.35
≥1000	5	171	2	542	7	713	1.0 (0.4-2.1)	0.03
Smaller or equal volume on 2nd CT	3	476	3	430	6	906	0.7 (0.3-1.5)	0.27
Resolved on 2nd CT	0	135	0	70	0	205	0.0 (0.0-2.2)	1.00
No follow-up CT, not referred	4	281	0	155	4	436	0.9 (0.3-2.4)	0.11
No follow-up CT, directly referred	3	5	2	6	5	11	45.5 (21.3-72.0)	<0.0001
All participants with largest nodule 50-500mm³	**33**	**1,455**	**24**	**1,499**	**57**	**2,954**	**1.9 (1.5-2.5)**	**<0.0001**

Definition of abbreviations: VDT = volume doubling time; 95%CI = 95% confidence interval; ref = reference value

† *Maximum VDT in subjects whose largest nodule measured 50-500 mm³, the interval for VDT includes the lower limit, not the upper limit.*

+ *Probability of lung cancer within two years after a CT scan. The difference in lung cancer risk with subjects without nodules was evaluated using Fisher's exact test.*

but did not significantly differ from the findings above. Lung cancer probability according to few categories of nodule volume and VDT is provided in Table 1 of the Appendix.

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 (Table 2 in Appendix), and multinodularity (Table 3a-b in Appendix) were also significant multivariate predictors. However, the relationship between multinodularity and lung cancer risk was ambiguous: for those participants whose nodules were growing and measured 50-500 mm³, the relative proportion of participants with lung cancer decreased as the numbers of nodules per participant increased (Figure 1a in Appendix). However, in the total study population, the proportion of lung cancers varied as the amount of nodules per participant increased (Figure 1b in Appendix). 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 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). Figure 2 shows

Table 2. Performance evaluation of simulated nodule management protocols for CT-detected nodules at the first screening round

Screening result[+]	**Management protocol using volume**	**Management protocol using diameter***	**Management protocol of the ACCP***
Positive	volume ≥300 mm^3	diameter ≥10 mm	diameter ≥8 mm
Indeterminate	volume ≥100 to ≤300 mm^3[‡]	diameter ≥5 to <10 mm[‡]	diameter >4 to <8 mm[l]
Negative	volume <100 mm^3	diameter <5 mm	diameter ≤4 mm
Screening test results	**Percentage (n/n)**	**Percentage (n/n)**	**Percentage (n/n)**
Direct referral due to positive result	4.7 (334/7,135)	5.3 (375/7,135)	8.9 (635/7,135)
Follow-up examination due to indeterminate result	7.8 (555/7,135)	22.2 (1586/7,135)	29.8 (2125/7,135)
- positive result after follow-up examination	1.2 (84/7,135)	5.5 (394/7,135)	4.7 (333/7,135)
- negative result after follow-up examination	6.6 (471/7,135)	16.7 (1192/7,135)	25.1 (1792/7,135)
Detected lung cancers	90.9 (60/66)	92.4 (61/66)	90.9 (60/66)
Screen test parameters	**Percentage (95%CI)**	**Percentage (95%CI)**	**Percentage (95%CI)**
Sensitivity	90.9 (81.2 - 96.1)	92.4 (83.1 - 97.1)	90.9 (81.2 - 96.1)
Specificity	94.9 (94.4 - 95.4)	90.0 (89.3 - 90.7)	87.2 (86.4 - 87.9)
Positive predictive value	14.4 (11.3 - 18.1)	7.9 (6.2 - 10.1)	6.2 (4.8 - 7.9)
Negative predictive value	99.9 (99.8 - 100.0)	99.9 (99.8 - 100.0)	99.9 (99.8 - 99.9)

Definition of abbreviation: 95%CI = 95% confidence interval (calculated using the Agresti-Coull method).

[+] *In case of multiple nodules, the size of the largest nodule determines the screening result.*

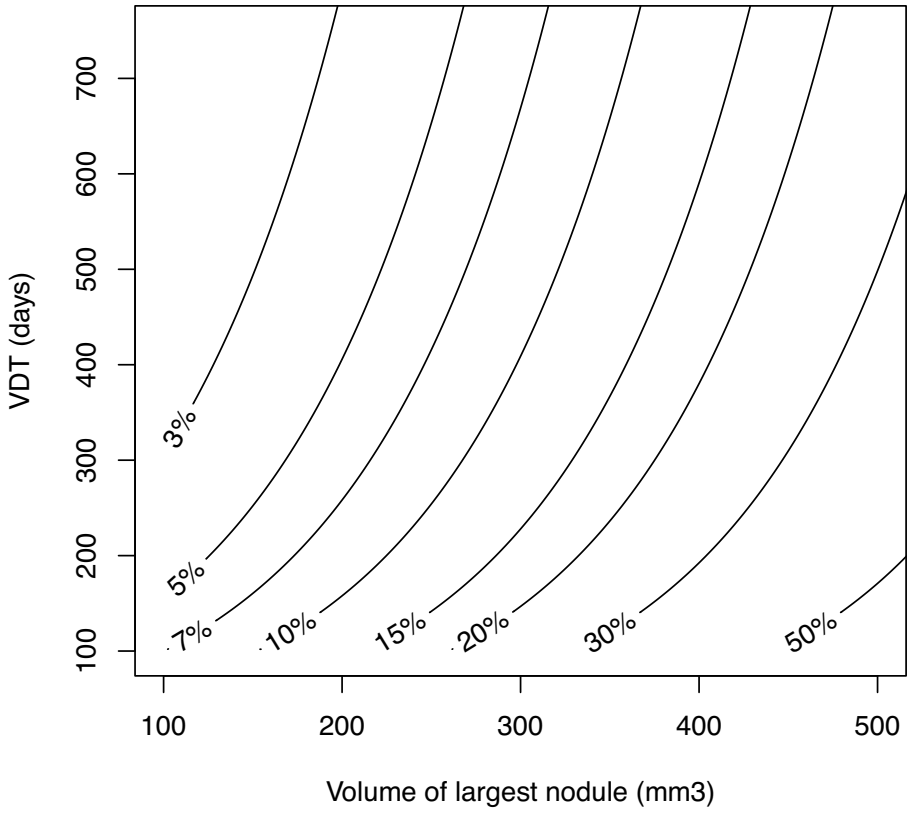
* *Estimates based on diameters assessed using semi-automated volumetry. Manually measured diameters are less accurate and will overestimate nodule size. As a result, the performance of the presented nodule protocol using diameter will be worse when manually measured diameters are used to calculate nodule 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.*

[l] *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 (2013).*

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).

Figure 2. Contour plot of the effect of the combined effect of nodule volume and VDT on the two-year lung cancer probability



Abbreviations: VDT = volume-doubling time.
The risk isolines represent the percentage of NELSON participants that will be diagnosed with lung cancer within two years according to the volume of their largest nodule and VDT of the fastest growing nodule in the 50-500 mm³ range.

that in participants with nodules of 300 mm³ in size or larger, the lung cancer probability remained substantial (from 5.9% to >50%), even in case of slow nodule growth. In participants with nodules sized 100-300 mm³, lung cancer probability ranged from less than 3% to 20%, dependent on the volume doubling time.

Evaluating management protocols for CT-detected nodules

Two nodule volume or diameter thresholds based on lung cancer probability, or using the simulated ACCP management protocol, are presented in Table 2. After the first screening round (for a one 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%] of 7,135) 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 (1,586 [22%] of 7,135), and additional diagnostic procedures (769 [11%] of 7,135), 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 and additional diagnostic procedures (968 [14%] of 7,135). Performance of the lung cancer probability-based volume and diameter protocols in the second screening round with the same thresholds is provided in Table 4 of the Appendix.

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.

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%.[7,13,21-25] 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.[8,13,26] 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,[14] and confusion, distress and frustration as reported by Wiener and colleagues[27]), and associated pressure on financial resources.[28,29]

Participants whose nodules measured 100-300 mm³ (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,[8,14,26] 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%),[23] 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 ≥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[8,21,24,30] and 8.9% (95% CI 5.6-13.7) or higher for volumes 300 mm³ 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 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 ≥300 mm³ or ≥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,[25] 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 management protocols.

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 volume doubling 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.[3,5,13]

The protocol that used lung cancer probability-based diameter thresholds was more sensitive than the simulated ACCP protocol[8,15] 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.[20] 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,[13] 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.[20,31-34] 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.[9,10,24,25] 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.[25] 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 ≤100 mm^3 or ≤5 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.

DECLARATION OF INTEREST STATEMENT

NH, HJdK, CMvdA, and KtH report grants from ZonMw, grants from KWF, grants from Roche diagnostics, and non-financial support from Siemens Germany during the conduct of the study. Additionally, HJdK and KtH report grants from National Cancer Institute/National Institutes of Health and grants from The Agency for Healthcare Research and Quality during the conduct of the study; grants from Sunnybrook Health Sciences and grants from Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen (RVVZ) outside the submitted work. J-WJL reports grants from the EU and grants from TiPharma outside the submitted work. KN reports grants from Belgian Foundation against Cancer and grants from Flemish Cancer League during the conduct of the study.

REFERENCES

1. Detterbeck FC, Mazzone PJ, Naidich DP, Bach PB. Screening for Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013; 143:e78S-92S.
2. Wender R, Fontham ET, Barrera E, Jr., et al. American Cancer Society lung cancer screening guidelines. *CA Cancer J Clin* 2013; 63:107-17.
3. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361:2221-9.
4. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365:395-409.
5. Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. *N Engl J Med* 2003; 348:2535-42.
6. Tan BB, Flaherty KR, Kzaerooni EA, Iannettoni MD. The solitary pulmonary nodule. *Chest* 2004; 123(suppl1):89S-96S.
7. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 354:99-105.
8. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005; 237:395-400.
9. Henschke CI, Yip R, Yankelevitz DF, International Early Lung Cancer Action Program Investigators. Definition of a positive test result in computed tomography screening for lung cancer: a cohort study. *Ann Intern Med* 2013; 158:246-52.
10. Yip R, Henschke CI, Yankelevitz DF, Smith JP. CT screening for lung cancer: alternative definitions of positive test result based on the National Lung Screening Trial and International Early Lung Cancer Action Program databases. *Radiology* 2014 Jun 19; doi:132950.
11. Linet MS, Slovis TL, Miller DL, et al. Cancer risks associated with external radiation from diagnostic imaging procedures. *CA Cancer J Clin* 2012; 62:75-100.
12. Van't Westeinde SC, Horeweg N, De Leyn P, et al. Complications following lung surgery in the Dutch-Belgian randomized lung cancer screening trial. *Eur J Cardiothorac Surg* 2012; 42:420-9.
13. Gould MK, Donington J, Lynch WR, et al. Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer?: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013; 143:e93S-e120S.
14. Wiener RS, Gould MK, Woloshin S, Schwartz LM, Clark JA. "What do you mean, a spot?": A qualitative analysis of patients' reactions to discussions with their doctors about pulmonary nodules. *Chest* 2013; 143(3):672-7.
15. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120:868-74.
16. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54:177-84.
17. Horeweg N, van Klaveren RJ, Groen HJ, et al. Blinded and uniform cause of death verification in a lung cancer screening trial. *Lung Cancer* 2012; 77(3):522-5.
18. Royston P, Altman D. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *J R Stat Soc* 1994; 43:429-67.

19. Royston P, Sauerbrei W. Multivariable Model-building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables. Wiley, Hoboken, New-York, 2008.
20. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology* 2004; 231(2):453-8.
21. Midthun DE, Swensen SJ, Jett JR, Hartman TE. Evaluation of nodules detected by screening for lung cancer with low-dose spiral computed tomography. *Lung Cancer* 2003; 41(supplement 2):O-127.
22. Henschke CI, Yankelevitz DF, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. *Radiology* 2004; 231:164-8.
23. Swensen SJ, Jett JR, Hartman TE, et al. Lung cancer screening with CT: Mayo Clinic experience. *Radiology* 2003; 226:756-61.
24. National Lung Screening Trial Research Team, Church TR, Black WC, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 2013; 368:1980-91.
25. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013; 369:910-9.
26. Naidich DP, Bankier AA, MacMahon H, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013; 266:304-17.
27. van den Bergh KA, Essink-Bot ML, Borsboom GJ, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *Br J Cancer* 2010; 102:27-34.
28. Pyenson BS, Sander MS, Jiang Y, Kahn H, Mulshine JL. An actuarial analysis shows that offering lung cancer screening as an insurance benefit would save lives at relatively low cost. *Health Aff* 2012; 31(4):770-9.
29. Villanti AC, Jiang Y, Abrams DB, Pyenson BS. A cost-utility analyses of lung cancer screening and the additional benefits of incorporating smoking cessation interventions. *PLoS One* 2013; 8(8):e71379.
30. Gohagan J, Marcus P, Fagerstrom R, et al. Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute. *Chest* 2004; 126:114-21.
31. Gavrielides MA, Kinnard LM, Myers KJ, Petrick N. Noncalcified lung nodules: volumetric assessment with thoracic CT. *Radiology* 2009; 251:26-37.
32. Reeves AP, Biancardi AM, Apanasovich TV, et al. The Lung Image Database Consortium (LIDC): a comparison of different size metrics for pulmonary nodule measurements. *Acad Radiol* 2007; 14:1475-85.
33. Xie X, Zhao Y, Snijder RA, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. *Eur Radiol* 2013; 23:139-47.
34. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *Eur Radiol* 2006; 16:781-90.

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

NODCAT 3:

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

NODCAT 4:

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

- VDT <400 days

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: <50mm^3 was negative, 50-500mm^3 was indeterminate, and >500mm^3 was positive. For previously detected nodules, VDT was calculated and determined the screening result: >600days was negative, 400-600days was indeterminate and <400days 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 benignancy (e.g. benign calcification patterns).

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:).

$$V = \frac{1}{6}\pi D^3$$

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:

I) For the evaluation of the simulated ACCP algorithm: participants with at least one indeterminate nodule with a VDT ≤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 ≤600 days are referred; participants in whom all indeterminate nodules have VDT>600 days are not referred.

II) 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 ≥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:

I) The diagnostic work-up, which was initiated for a positive 'final' screening result, led to the diagnosis of lung cancer (true positive screening results).

II) 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:

I) 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).

II) 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.

III) The positive predictive value was estimated by dividing all subjects with a true positive screening by all subjects with positive screening.

IV) 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:

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

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

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

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

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

VI) Tumour in right upper lobe, volume 891 mm^3, PET indeterminate, cT1aN0M0, and patient died due to bowel ischemia just before intended thoracic surgery.

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

VIII) Tumour 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 randomised to the radiotherapy treatment arm.

IX) Tumour 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.

X) Tumour in left lower lobe, volume 1,108 mm^3, PET positive, cT1aN1M0, patient did not undergo thoracic surgery due to poor pulmonary function.

XI) Tumour in left lower lobe, diameter 10 mm, PET positive, cT1aN0M0, and patient did not undergo thoracic surgery due to poor pulmonary function.

XII) Tumour in right upper lobe, diameter 13.2 x 11.6 mm, PET positive, cT1aN0M0, patient did not undergo thoracic surgery due to poor pulmonary function and general condition

XIII) Tumour in right upper lobe, diameter 19.2 x 12.7 mm, PET positive, cT1bN0M0, patient did not undergo thoracic surgery due to poor general condition

Figure 1a. Relationship multi-nodularity and lung cancer probability in all subjects with nodules

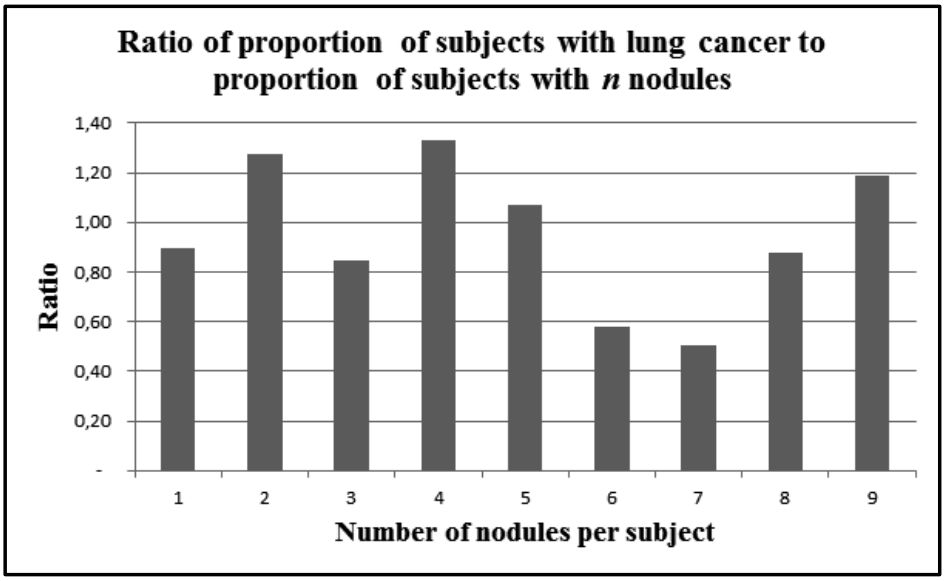


Figure 1b. Relationship multi-nodularity and lung cancer probability in subjects whose largest measure 50-500mm³ and have a VDT>0

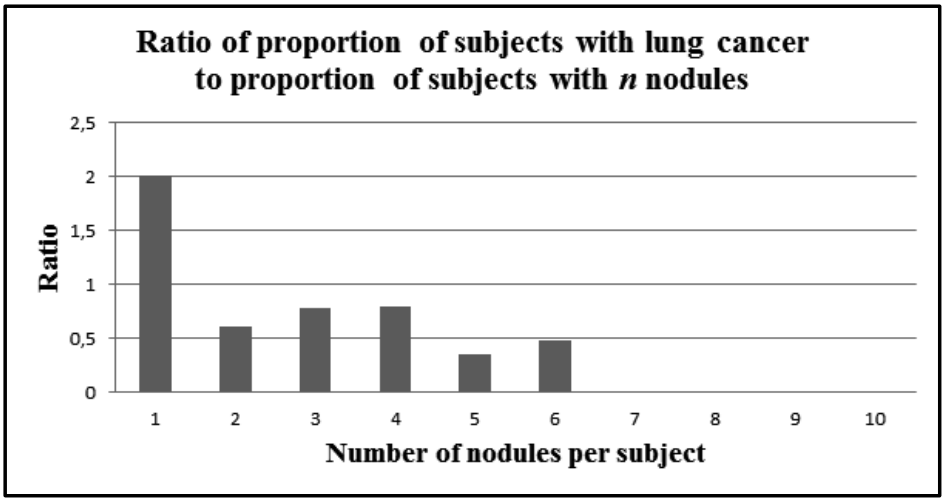


Table 1. Two-year lung cancer probability by nodule volume and volume doubling-time

	Nodule volume doubling-time		
Nodule volume	**<600 days**	**≥600 days**	**shrunken or resolved**
< 100 mm^3	7/199 (3.5%)	2/521 (0.4%)	0/4,803 (0.0%)
≥100 to <300 mm^3	19/207 (9.2%)	3/379 (0.8%)	0/525 (0.0%)
≥300 mm^3	99/464 (21.3%)	3/51 (5.9%)	0/102 (0.0%)

Table 2. Multivariable logistic regression model for the probability to be diagnosed with lung cancer

Variable	**Odds ratio (95% CI)**
Nodule volume^	2.12 (1.64-2.75)*
Nodule VDT^{+}	0.45 (0.35-0.60)*
Constant	1.35 (0.24-7.79)

In this model, only the participants in whom the largest detected nodule had a volume of ≥50 mm^3 and <500 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 = 0.7.

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 mm^3 divided by 100.*

$^{+}$ *Logarithmic effect: nodule VDT was defined as the natural logarithm of VDT in days.*

* *p-value < 0.001.*

Table 3a. Multivariable logistic regression model for the probability to be diagnosed with lung cancer

Variable	Odds ratio (95% CI)	p-value
Nodule volume^	2.19 (1.69-2.84)	<0.001
Nodule VDT+	0.43 (0.32-0.57)	<0.001
Multi-nodularity*	0.68 (0.55-0.85)	0.001
Constant	5.00 (0.74-33.79)	0.099

In this model, only the participants in whom the largest detected nodule had a volume of ≥50 mm³ and <500 mm³ 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, multi-nodularity, and a constant term. Hosmer-Lemeshow goodness-of-fit test: p = 0.8.

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 mm³ divided by 100.*

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

* *Linear effect: multi-nodularity was defined as the number of nodule present at the scan.*

Table 3b. Multivariable logistic regression model for the probability to be diagnosed with lung cancer

Variable	Odds ratio (95% CI)	P value
Nodule volume^	2.20 (1.69-2.88)	<0.001
Nodule VDT+	0.44 (0.33-0.59)	<0.001
Multi-nodularity*	0.20 (0.10-0.41)	<0.001
Constant	3.60 (0.55-33.41)	0.18

In this model, only the participants in whom the largest detected nodule had a volume of ≥50 mm³ and <500 mm³ 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, multinodularity, and a constant term. Hosmer-Lemeshow goodness-of-fit test: p = 0.8.

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 mm³ divided by 100.*

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

* *Linear effect: multinodularity as binary variable (0 = 1 nodule, 1 = ≥ 2 nodules).*

Table 4. Performance evaluation of simulated management algorithms for CT-detected nodules at the second screening round

Screening result[+]	Management protocol based on volumetry	Management protocol based on diameter*	Management protocol of the ACCP*
Positive	volume ≥300 mm^3	diameter ≥10 mm	diameter ≥8 mm
Indeterminate	volume ≥100 to ≤300 mm^3[‡]	diameter ≥5 to <10 mm[‡]	diameter >4 to <8 mm[l]
Negative	volume <100 mm^3	diameter <5 mm	diameter ≤4 mm
Screening test results	**Percentage (n/n)**	**Percentage (n/n)**	**Percentage (n/n)**
Direct referral due to positive result	4.1 (283/6,889)	4.7 (322/6,889)	9.0 (618/6,889)
Follow-up examination due to indeterminate result	8.1 (556/6,889)	24.3 (1676/6,889)	31.3 (2,156/6,889)
- positive result after follow-up examination	1.0 (72/6,889)	5.4 (370/6,889)	3.1 (314/6,889)
- negative result after follow-up examination	7.0 (484/6,889)	19.0 (1,306/6,889)	28.2 (1,942/6,889)
Detected lung cancers	83.1 (49/59)	86.4 (51/59)	88.1 (52/59)
Screen test parameters	**Percentage (95%CI)**	**Percentage (95%CI)**	**Percentage (95%CI)**
Sensitivity	83.1 (71.3 - 90.7)	86.4 (75.2 - 93.2)	88.1 (77.2 - 94.4)
Specificity	95.5 (95.0 - 96.0)	90.6 (89.9 - 91.3)	88.6 (87.8 - 89.3)
Positive predictive value	13.8 (10.6 - 17.8)	7.4 (5.6 - 9.6)	6.3 (4.8 - 8.1)
Negative predictive value	99.8 (99.7 - 99.9)	99.9 (99.7 - 99.9)	99.9 (99.8 - 99.9)

Definition of abbreviation: 95%CI = 95% confidence interval (calculated using the Agresti-Coull method).

[+] *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.*

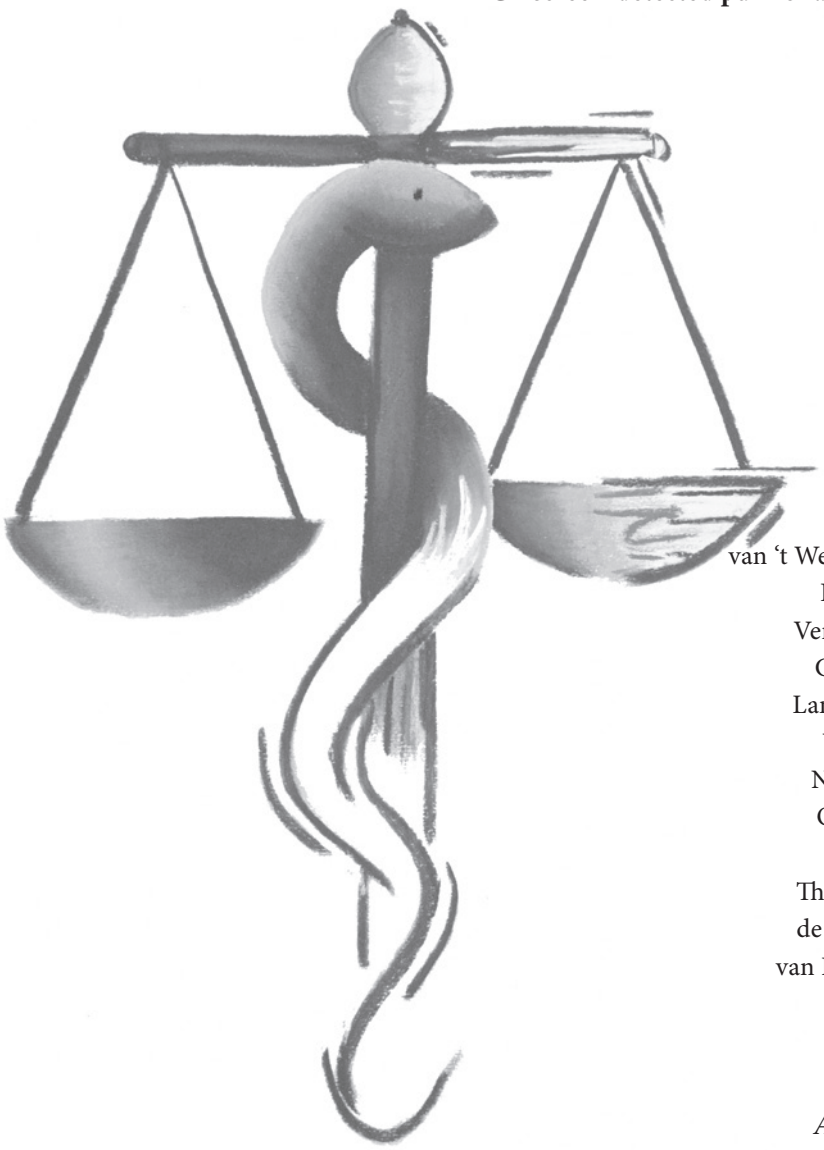
[l] *Subjects with an indeterminate screening result should undergo a follow-up scan after three months to assess the VDT; a VDT <400 days is a positive screening result and leads to referral for diagnostic work-up, according to the ACCP guideline (2013).*

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 data supplement).

Chapter 7

Evaluation of bronchoscopy

The role of conventional bronchoscopy in the work-up of suspicious CT screen-detected pulmonary nodules



van 't Westeinde SC,
Horeweg N,
Vernhout RM,
Groen HJM,
Lammers JWJ,
Weenink C,
Nackaerts K,
Oudkerk M,
Mali W,
Thunnissen E,
de Koning HJ,
van Klaveren RJ

Chest

August 2012

The role of conventional bronchoscopy in the work-up of suspicious CT screen-detected pulmonary nodules

Chest. 2012 Aug;142(2):377-84.

ABSTRACT

Up to 50% of the participants in CT scan lung cancer screening trials have at least one pulmonary nodule. To date, the role of conventional bronchoscopy in the workup of suspicious screen-detected pulmonary nodules is unknown. If a bronchoscopic evaluation could be eliminated, the cost-effectiveness of a screening program could be enhanced and the potential harms of bronchoscopy avoided.

All consecutive participants with a positive result on a CT scan lung cancer screening between April 2004 and December 2008 were enrolled. The diagnostic sensitivity and negative predictive value were calculated at the level of the suspicious nodules. In 95% of the nodules, the gold standard for the outcome of the bronchoscopy was based on surgical resection specimens.

A total of 318 suspicious lesions were evaluated by bronchoscopy in 308 participants. The mean SD diameter of the nodules was 14.6 8.7 mm, whereas only 2.8% of nodules were <30 mm in diameter. The sensitivity of bronchoscopy was 13.5% (95% CI, 9.0%-19.6%); the specificity, 100%; the positive predictive value, 100%; and the negative predictive value, 47.6% (95% CI, 41.8%-53.5%). Of all cancers detected, 1% were detected by bronchoscopy only and were retrospectively invisible on both low-dose CT scan and CT scan with IV contrast.

Conventional white-light bronchoscopy should not be routinely recommended for patients with positive test results in a lung cancer screening program.

INTRODUCTION

Depending on the geographic region, 26% to 51% of participants in multi-detector computed tomography (CT) lung cancer screening trials showed at least one non-calcified pulmonary nodule on their CT scan.[1-4] The likelihood of these nodules being malignant depends on their size.[1,5] The Fleischner Society guideline recommends a recall CT scan, PET scan, or biopsy for nodules >8 mm detected on a CT scan 5 but not by bronchoscopy. The American College of Chest Physicians (ACCP) guideline recommends only evaluation by bronchoscopy under the condition that an air bronchogram is present on CT scan or in centres with expertise in newer techniques.[6,7] Literature on the role of newer techniques, such as ultrathin bronchoscopy, autofluorescence bronchoscopy, and CT scan-guided bronchoscopy in lung cancer screening settings is sparse. To our knowledge, a study by McWilliams et al[8] is the only one to report on the role of autofluorescence bronchoscopy in a lung cancer screening trial. The diagnostic yield of bronchoscopy to evaluate solitary pulmonary nodules outside a CT scan screening program varies from 51% to 76%[9-14] and highly depends on the size and location of the nodule.[9,10,13-15]

The nodule management strategy of the Dutch-Belgian Randomised Lung Cancer Screening Trial (NELSON) is based on the size and the volume-doubling time (VDT) of nodules detected by CT scan, without the use of fine-needle aspiration, PET scan, or evaluation after antibiotics.[1] Subjects with positive test results were referred for work-up of suspicious nodules, which included a physical examination, a standard CT scan with contrast, and bronchoscopy.[5,6,16,17]

Recently, a the U.S. National Lung Screening Trial (NLST) demonstrated a 20% mortality reduction with low-dose CT screening.[20] In the low-dose CT group, 320 subjects (1.8% of all subjects with a positive test result) underwent bronchoscopy without biopsy or cytological testing, whereas 391 subjects (2.2% of all subjects with a positive test result) underwent bronchoscopy with biopsy or cytological testing. The investigators did not report on the diagnostic performance of bronchoscopy in their study. In Pan-Canadian Lung Cancer Screening Trial, all participants were offered an auto-fluorescence bronchoscopy to detect central airway lesions; 67% (378 of 561) actually underwent this procedure.[21] Ideally, all subjects should have undergone bronchoscopy for this purpose. Four of 22 subjects (18%) bronchoscopy yielded a diagnosis of radiological occult lung cancer. In the Canadian trial, the purpose of bronchoscopy appears to have been inspection of the central airways.[21] In about 45% (320 of 711) of cases, cytology or histology specimens were obtained. It is unclear to what extent the ACCP guidelines were followed. In both the Canadian trial and the NLST, no nodule criteria were specified before the decision to perform bronchoscopy.

So far, lung cancer screening trials have not provided specific recommendations on the use of bronchoscopy to evaluate suspicious CT-detected nodules,[2,18,19,22] nonetheless

a significant number of bronchoscopies have been performed.[20] Screening detects more early-stage lung cancers, whereas advanced-stage lung cancers that are present as interval cancers amenable to bronchoscopy are excluded from analyses.[1] Our hypothesis was that the diagnostic value of bronchoscopy in this workup process might be low as suspicious nodules are usually small and peripherally located.[1,18,19] If this is true, bronchoscopic evaluation may be eliminated from the standard work-up of suspicious CT-detected nodules; which would enhance the cost-effectiveness of a lung cancer screening program and avoid the harms of bronchoscopy. Therefore, our objective was to prospectively investigate the diagnostic value of bronchoscopy in the NELSON trial and to evaluate the diagnostic yield of the various diagnostic techniques used during bronchoscopy.

METHODS

Study Population

The nodule management strategy of the NELSON trial has been published previously.[16,23] Briefly, 15,822 individuals with at high risk for developing lung cancer were randomised to either four low-dose CT examinations (n=7,915) at baseline (first round), one year later (second round), three years later (third round), and five and a half years later, or no screening (n=7,907). All consecutive participants with a positive test result at the first, second, and third screening round from April 2004 to December 2008 were included in this study.

A test result was considered positive for pulmonary nodules with a volume of >500 mm^3(about 9.8 mm in diameter) and nodule with a volume doubling-time (VDT) of <400 days.[1,16,21] For nodules with a solid component measuring ≥50 to ≤500 mm^3 the test result was indeterminate, and a repeat scan was performed to assess the VDT of the nodule. When the VDT was <400 days at repeat scanning, the final test result was considered positive, otherwise, it was considered negative.[1,16]

The NELSON trial was approved by the ethics committees of all participating centres, and all participants provided written informed consent (approval number IRB00001838).

Bronchoscopy

Conventional white-light bronchoscopies were performed by experienced pulmonologists working at the four screening sites in The Netherlands (Utrecht, Groningen, and Haarlem) and Belgium (Leuven).[16] During bronchoscopy, bronchial washings were performed for cytology and culture; bronchial brushings and biopsy specimens were taken (52C-1 forceps) in the case of central lesions. In less than 1% of the cases, biopsy was performed under fluoroscopic guidance. The bronchoscopists did not use CT scan fluoroscopic guidance or ultrathin bronchoscopes. A flexible Pentax video bronchoscope

(Pentax Medical Company) was used in Utrecht, whereas Groningen and Haarlem used the Olympus flexible video bronchoscope (Olympus America Inc), and in Leuven both types were used.

Endobronchial abnormalities were classified as visible tumour, constriction, or compression of the airways. Nodules within the inner third, middle, and outer third of the hilar-costal diameter on CT scan were classified as respectively central, intermediate, or peripheral.

If the bronchoscopy revealed cancer, the outcome of the procedure was considered positive; otherwise, it was considered negative. The gold standard for the outcome of the bronchoscopy was the pathology result of the surgical resection specimen of the suspicious lesion. If no surgical resection was performed, the presence or absence of cancer during a follow-up of at least two years after the first and second screening rounds and at least one year after the third round was used as the gold standard. Nodules with a VDT of >400 days at follow-up CT examinations were considered benign.[1]

Data Analysis

Statistical analyses were performed using SPSS, version 17.0 (SPSS, Inc.) software. The sensitivity was defined as the ratio between the number of positive bronchoscopy results and the number of positive results according to the gold standard. The diagnostic sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated at the level of the suspicious nodules. Suspicious nodules that were not approached during bronchoscopy were excluded from the analysis.

Mann-Whitney U test was used for continuous variables, and chi^2-test was used for binomial and categorical data. Binary logistic regression was used to determine the effect of the individual nodule characteristics on the diagnostic yield of bronchoscopy. In the multivariate analysis, the characteristics with a p-value ≤ 0.10 were included for a stepwise-forward procedure. P-values <0.05 were considered statistically significant.

RESULTS

Of the 415 participants with positive test results, 74.2% (308 of 415) underwent bronchoscopy to evaluate 318 suspicious pulmonary lesions; 25.8% (107 of 415) did not undergo bronchoscopy for several reasons (Fig 1). In 2.4% (10 of 415) of the cases, referral was based on non-nodular lesions on CT. Six bronchoscopies were performed (Fig 1) on these participants. No significant differences were found in participant characteristics and nodule characteristics between those who did and those who did not undergo bronchoscopy (Table 1); except for a sex difference and a difference in cancer detection rate of 22.4%

Table 1. Participant and nodule characteristics of participants who underwent and who did not underwent bronchoscopy

Characteristics	Bronchoscopy n (%)	No bronchoscopy n (%)	p-value
Female gender	60 (19.5)	8 (7.5)	<0.001
Age - mean(range)	62 (50-75)	62 (51-74)	0.44
Pack-years - mean(range)	47 (21-160)	43 (22-108)	0.12
Nodule			
Diameter - mean(SD)	14.6 (8.7)	14.6 (9.1)	0.78
VDT <400 days	132 (90.4)	39 (84.8)	0.29
Localisation, central*	38 (12.8)	9 (9.2)	0.33
Lobe, upper	152 (51.0)	48(48.0)	0.60
Screen round			
First	144 (46.8)	44 (41.4)	0.31
Second	91 (24.5)	29 (27.1)	0.63
Third	73 (23.7)	34 (31.8)	0.10
Scan type			
Regular scan	202 (65.6)	67 (62.6)	0.58
Repeat scan	106 (34.4)	40 (37.4)	0.58
Total	**308 (100.0)**	**107 (100.0)**	**NA**

Definition of abbreviations: NA = not applicable, VDT = volume doubling-time, n=number, SD= standard deviation.

* *Central indicates a localisation within in the inner one-third of the costal-hilar diameter on computed* ***tomo****graphy.*

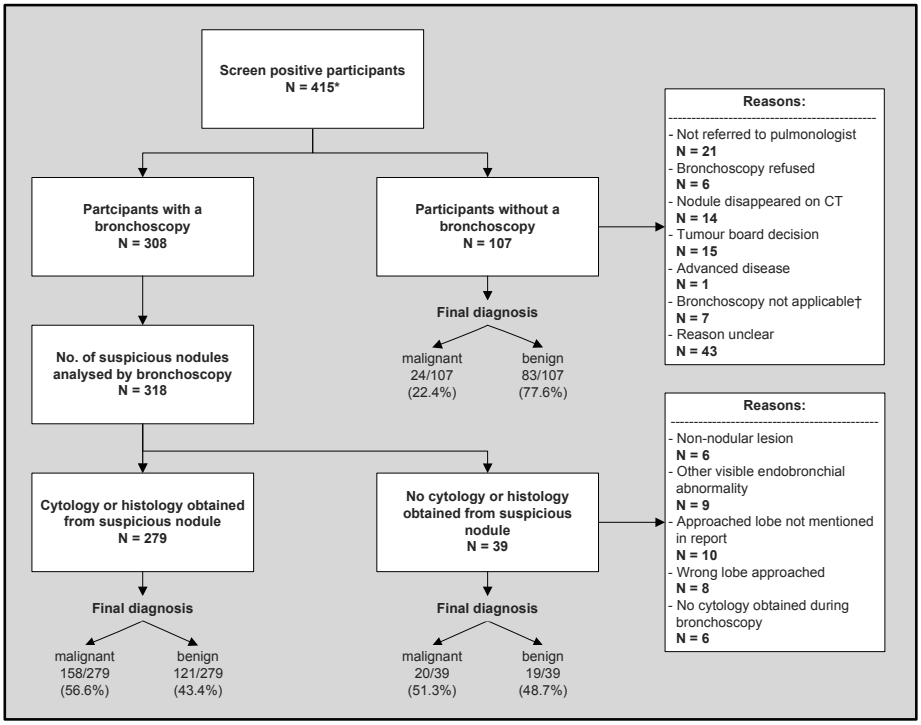
(24 of 107) in the group without bronchoscopy compared with 57.8% (178 of 308) in the bronchoscopy group (p<0.001) (Fig 1).

The average maximum diameter of the suspicious nodules was 14.6 mm; the maximum diameter varied from 5 to 68 mm (SD 8.7 mm), only 2.8% of the lesions measured >30 mm. Cancer was diagnosed by bronchoscopy in only 24 of 318 suspicious lesions. Suspicious nodules identified as cancer by bronchoscopy were significantly larger (odds ratio (OR) 1.07, 95% CI 1.02-1.13), and were more often visible during bronchoscopy (OR 87.6, 95% CI 4.9-564.9) compared to the cancer cases missed by bronchoscopy (Table 2).

Based on the gold standard used in this study, a total of 178 cancer cases were detected among the 318 lesions, including 167 lung cancer cases. In 77% (137 of 178) of cases the gold standard was based on surgical resection specimens, in 18.5% (33 of 178) of cases it was based on surgical biopsy specimens (mediastinoscopy, true-cut biopsy performed during surgery), and in 4.5% (eight of 178) of the cases it was based on the combination of a new and growing PET-positive lesion on CT scan.

In 24 of the 178 subjects with cancer, bronchoscopy yielded the diagnosis of cancer. Hence, the sensitivity of bronchoscopy to detect cancer was 13.5% (95% CI 9.0-19.6%),

Figure 1. Bronchoscopies performed for the evaluation of positive screening test results in the NELSON trial



Outcome of the bronchoscopies performed for the evaluation of 308 participants with positive screening results of the Dutch-Belgian randomised lung cancer screening trial.

* *Includes 10 participants referred for non-nodular lesions detected on screening CT scan (interstitial lung disease, pneumonia, lobar atelectasis, pleural effusion, pleural thickening, or mediastinal mass).*

and the NPV was 47.6% (140 of 294; 95% CI 41.8-53.5%). Accordingly, 48.4% (154 of 318) of the bronchoscopic findings were false-negative (Table 3). As no false-positive diagnoses were made by bronchoscopy specificity and PPV were both 100%.

In 7.5% (23 of 308) of all bronchoscopies, an endobronchial abnormality was found, and in 47.8% (11 of 23) of the cases, the tumour was endobronchially visible. When an endobronchial tumour was visible, the sensitivity of bronchoscopy to detect cancer was 81.8% (95% CI 47.8-96.8%). In 2.6% of the 308 bronchoscopies, an endobronchial tumour was detected, which was not visible on CT scan, also in retrospect. This accounts for 4.5% (eight of 178) of all cancer cases detected by CT scan screening in this period. Of these eight additional cancer cases, only three were stage I, the remaining five were stage III or IV. When the diagnostic performance of bronchoscopy was limited to the suspicious nodules visible on CT scan, the sensitivity to detect cancer was 8.3% (14 of 168, 95% CI 4.8-13.9%), and the NPV was 47.6% (140 of 294, 95% CI 41.8-53.5%).

Table 2. Univariate and multivariate analyses of bronchoscopies in participants with cancer

	Cancer diagnosed by bronchoscopy (n=24)	Final diagnosis of cancer (n=178)	Univariate		Multivariate	
Characteristic	**n (%)**	**n (%)**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Nodule*						
Diameter in mm - mean (SD)	26.0 (3.6)	17.7 (0.7)	1.08 (1.04-1.13)	<0.001		
VDT >400 days	0/5 (0.0)	8/86 (9.3)	ref			
VDT <400 days	5/5 (100.0)	78/86 (90.7)				
Localisation[a]						
Peripheral	11/22 (50.0)	146/176 (83.0)	ref			
Central[b]	11/22 (50.0)	30/176 (17.0)	7.11 (2.71-18.63)	<0.001	0.44 (0.04-5.16)	0.52
Middle and lower lobe	5/21 (23.8)	64/175 (36.6)	ref			
Upper lobe	16/21 (76.2)	111/175 (63.4)	1.99 (0.69-5.71)	0.20		
Screen round						
First	12/24 (50.0)	74/178 (41.6)	ref			
Second	6/24 (25.0)	51/178 (28.7)	0.69 (0.24-1.97)	0.49		
Third	6/24 (25.0)	53/178 (29.8)	0.66 (0.23-1.89)	0.44		
Scan type						
Regular scan	17/24 (70.8)	12/178 (69.1)	ref			
Repeat scan	7/24 (29.2)	55/178 (30.9)	0.91 (0.35-2.34)	0.84		
Bronchoscopy						
Normal	11/24 (45.8)	155/178 (87.1)	ref			
Bronchial compression	4/24 (16.7)	12/178 (6.7)	6.54 (1.70-25.19)	<0.001	4.06 (0.56-29.21)	0.16
Visible tumour	9/24 (37.5)	11/178 (6.2)	58.91 (11.31-306.8)	<0.001	87.61 (4.90-564.9)	<0.001

Definition of abbreviations: OR = odds ratio; 95% CI = 95% confidence interval; SD = standard deviation; VDT = volume doubling-time.

[a] *The denominator is not always equal to 24 or 178 because it was not possible to classify all nodules.*

[b] *Central indicates the inner one-third of the costal-hilar diameter on CT scan.*

The sensitivities of the various diagnostic techniques used during bronchoscopy ranged from 7.9% for brush to 45.8% for endobronchial biopsy (Table 4).

During the bronchoscopies, minor complications (nose bleeding and mild bleeding after the biopsy) occurred in only 0.6% (2 of 308) of participants. There were no major complications.

Table 3. Overview of false-negative diagnoses by bronchoscopy

Cytological or histological diagnosis	No. (%)
No abnormality	101 (65.6)
Infectious	
Aspergillus fuminatus	1 (0.6)
Aspecific inflammation	42 (27.3)
Preinvasive lesion	
Atypia	1 (0.6)
Benign, other	
Fibrosis	2 (1.3)
Resolving haemorrhage	1 (0.6)
Metaplasia	6 (3.9)
Total false-negative diagnoses	**154 (100.0)**

Table 4. Diagnostic value of ancillary procedures during bronchoscopy

	Malignancy		Diagnostic performance			
Diagnostic technique	**Bronchoscopy**	**Final diagnosis**	**Sensitivity**	**95% CI**	**NPV**	**95% CI**
Wash	17/322 (5.3)	182/322 (56.5)	9.3	5.7-14.8	45.9	40.2-51.7
Brush	6/125 (4.8)	76/125 (60.8)	7.9	3.3-17.0	41.2	32.4-50.6
TBNA	2/6 (33.3)	6/6 (100.0)	33.3	6.0-75.9	0.0	0.0-6.0
TBB	1/12 (8.3)	8/12 (66.7)	16.7	0.9-63.5	44.4	15.3-77.3
EBB	11/40 (27.5)	24/40 (60.0)	45.8	26.2-66.8	55.2	36.0-73.0
Overall	**37/505 (7.3)**	**296/505 (58.6)**	**12.5**	**9.1-16.4**	**44.7**	**40.1-49.3**

Definition of abbreviations: 95% CI = 95% confidence interval; NPV = negative predictive value; TBNA = transbronchial needle aspiration; TBB = transbronchial biopsy; EBB = endobronchial biopsy.
No false-positive diagnoses were made by any of the ancillary procedures during bronchoscopy, therefore all specificities and positive predictive values were 100%.

DISCUSSION

In this study, the diagnostic value of conventional white-light bronchoscopy in the NELSON trial was prospectively evaluated.

The overall sensitivity was 13.5%, and the NPV was 47.6%. The sensitivity was only 8.3% when limited to CT-detected suspicious nodules. Of all cancers detected within the time frame of this study, 4.5% were identified by bronchoscopy only and were not visible on CT.

In non-screening studies, the sensitivity of conventional bronchoscopy varied from 51% to 76%,[9-14] which is much higher than the 13.5% in the NELSON trial. This can be

explained by the fact that in the present study, only 2.8% of the nodules were >30 mm, whereas in non-screening studies, lesion size ranged from 48 to 72 mm.[9,10,13,14] Further, fewer lesions were endobronchially visible in the current study compared to the literature (7.3% vs. 8-64%).[10-12] Both nodule size and endobronchial visibility were independent predictors for high diagnostic yield in the present study.

As far as we know, only Kanemoto et al[24] retrospectively evaluated the diagnostic value of bronchoscopy in a selected study population in which 108 suspicious pulmonary nodules had been detected by mass screening (chest radiography or CT scan). All nodules were ≤20 mm and 42% of the nodules were malignant; based on fluoroscopy-guided bronchoscopy or lung biopsy specimens. The drawback of that study is the selection bias of the study population and the absence of a gold standard for the outcome of the bronchoscopy. As a result, the investigators were unable to provide data on the diagnostic performance of bronchoscopy in this screening program.

According to current guidelines, bronchoscopy is recommended only for the evaluation of nodules with an air bronchogram[6,25,26] without a standard position in the routine workup of suspicious pulmonary nodules. Although we did not evaluate whether the presence of an air bronchogram increased the diagnostic yield of bronchoscopy, the results clearly demonstrate that bronchoscopy is not justified for the evaluation of CT-detected pulmonary nodules, because of its very low sensitivity and NPV.

The use of more advanced bronchoscopic techniques is not yet recommended by the ACCP.[6] Nevertheless, we believe that electromagnetic-navigated or peripheral endobronchial ultrasound-guided bronchoscopy may play a role in the evaluation of small, peripheral CT-detected nodules, since their sensitivity is respectively 59-74%[27-30] and 49-80%[27,31-34]. In addition, ultrathin bronchoscopy may play a role in the future for diagnostic evaluation of peripheral pulmonary nodules.[35,36]

Because of the poor diagnostic performance, we do not recommend the routine use of conventional bronchoscopy for patients with suspicious CT-detected nodules, even though bronchoscopy yielded the detection of eight cancers that were not visible on CT. Only one-third of these cancers was early stage and may be treated with curative intent. It is arguable what percentage missed lung cancers is acceptable. We believe that this depends on the setting, lung cancer screening, or daily practice. In lung cancer screening trials, vast numbers of subjects are exposed to invasive procedures, which are accompanied by morbidity, anxiety and costs. Therefore, we consider the benefit of bronchoscopy too small. Moreover, only 38% of these radiologically occult cancers were stage I. The reason that the eight cancers detected by bronchoscopy were not visible on CT may be the use of low-dose CT examinations, without the use of intravenous contrast. The lesions were, in retrospect, visible on the standard-dose CT with intravenous contrast performed in the workup of these participants. Other investigators reported that 1%[37] to 5%[6] of CT

occult tumours was detected by white-light bronchoscopy, whereas 18% was detected with autofluorescence bronchoscopy.[21]

Additional diagnostic techniques, such as brush and biopsy, should only be used to evaluate visible endobronchial tumours. In the present study, we found sensitivity of >80% for brush and biopsy.

The strength of the present study lies in the fact that it was prospectively conducted and that for the majority of the suspicious nodules, histologic conformation was obtained by either surgical resection or biopsy specimen. So far, the diagnostic value of a conventional bronchoscopy has not previously been evaluated properly in a CT screening trial which included asymptomatic, high-risk participants from the general population. In addition, we were able to evaluate the diagnostic value of bronchoscopy based on individual nodule level. Despite this, our study also has its limitations. The proportion of women and the cancer detection rate in the group that did not undergo bronchoscopy was lower than in those who underwent bronchoscopy. Because sex was not associated with a higher diagnostic yield in the study, this did not introduce selection bias. The cancer detection rate in the group without bronchoscopy was lower because 20% (21 of 107) of cases were referred for non-nodular lesions or because the suspicious nodule had disappeared on the diagnostic CT scan with contrast. If all participants with positive test results had undergone bronchoscopy, the sensitivity may have been even lower, which further strengthens the present conclusions. In the workup protocol, bronchoscopy only was performed on participants with positive test results. To date, it is not known what the performance of conventional bronchoscopy is, when it is used as a screening tool, instead of as diagnostic tool. This is currently under investigation in the Pan-Canadian lung cancer screening trial.[38]

CONCLUSION

In conclusion, routine use of conventional bronchoscopy in patients with suspicious CT-detected pulmonary nodules in a lung cancer screening program is not recommended.

SUPPORT STATEMENT

We thank Siemens Germany for providing four digital workstations, Roche Diagnostics for providing an unrestricted research grant, and Tom and Josephine Rijke for their legacy gift.

CONFLICTS OF INTEREST

The authors have reported to CHEST that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article. The unrestricted financial support of the sponsors provided for the costs of CT screening. The sponsors had no role in analysing or interpreting the data or in preparation of the manuscript.

ACKNOWLEDGEMENTS

We thank Roel Faber and Frank Santegoets for their assistance as information and communication technology scan managers as well as the data managers of the four screening sites: Henk Pruiksma (Haarlem), Liesbet Peeters (Leuven), Saskia van Amelsvoort-van de Vorst (Utrecht), and Ria Ziengs (Groningen).

REFERENCES

1. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361(23):2221-2229.
2. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish randomized lung cancer CT screening trial: overall design and results of the prevalence round. *J Thorac Oncol* 2009; 4(5):608-614.
3. Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002; 165(4):508-513.
4. Nawa T Nakagawa T Kusano S, et al. Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies. *Chest*. 2002; 122(1):15-20.
5. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005; 237(2):395-400.
6. Gould MK, Fletcher J, Iannettoni MD, et al. American College of Chest Physicians. Evaluation of patients with pulmonary nodules: when is it lung cancer? ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007; 132(suppl3):108S-130S.
7. Rivera MP, Mehta AC. American College of Chest Physicians. Initial diagnosis of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007; 132(suppl3):131S-148S.
8. McWilliams A, Mayo J, MacDonald S, et al. Lung cancer screening: a different paradigm. *Am J Respir Crit Care Med* 2003; 168(10):1167-1173.
9. Ost D, Shah R, Anasco E, et al. A randomized trial of CT fluoroscopic-guided bronchoscopy vs conventional bronchoscopy in patients with suspected lung cancer. *Chest* 2008; 134(3):507-513.
10. Roth K, Hardie JA, Andreassen AH, et al. Predictors of diagnostic yield in bronchoscopy: a retrospective cohort study comparing different combinations of sampling techniques. *BMC Pulm Med* 2008; 8:2.
11. Vincent BD, Fraig M, Silvestri GA. A pilot study of narrowband imaging compared to white light bronchoscopy for evaluation of normal airways and premalignant and malignant airways disease. *Chest* 2007; 131(6):1794-1799.
12. Chhajed PN, Bernasconi M, Gambazzi F, et al. Combining bronchoscopy and positron emission tomography for the diagnosis of the small pulmonary nodule ≤3 cm. *Chest* 2005; 128(5):3558-3564.
13. Reichenberger F, Weber J, Tamm M, et al. The value of transbronchial needle aspiration in the diagnosis of peripheral pulmonary lesions. *Chest* 1999 ; 116(3):704-708.
14. Baaklini WA, Reinoso MA, Gorin AB, et al. Diagnostic yield of fiberoptic bronchoscopy in evaluating solitary pulmonary nodules. *Chest* 2000; 117(4):1049-1054.
15. Khan A. ACR appropriateness criteria on solitary pulmonary nodule. *J Am Coll Radiol* 2007; 4(3):152-155.
16. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54(2):177-184.
17. Association of Integrated Cancer Centres. Guideline Non-small Cell Lung Cancer: Staging and Treatment. Alphen aan den Rijn, The Netherlands: Van Zuiden Communications. 2004 .
18. Henschke CI, Yankelevitz DF, International Early Lung Cancer Action Program Investigators. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006; 355(17):1763-1771.
19. Swensen SJ, Jett JR, Hartman TE, et al. Lung cancer screening with CT: Mayo Clinic experience. *Radiology* 2003; 226(3):756-761.
20. Aberle DR, Adams AM, National Lung Screening Trial Research Team. Reduced lung cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365(5):395-409.
21. McWilliams AM, Mayo JR, Ahn MI, et al. Lung cancer screening using multi-slice thin-section computed tomography and autofluorescence bronchoscopy. *J Thorac Oncol* 2006; 1(1):61-68.

22. Hocking WG, Hu P, PLCO Project Team. Lung cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *J Natl Cancer Inst* 2010; 102(10):722-731.
23. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120(4):868-874.
24. Kanemoto K, Satoh H, Ishikawa H, et al. Diagnostic procedures for small pulmonary nodules detected by mass-screening. *Anticancer Res* 2008; 28(5B):3153-3155.
25. Beigelman-Aubry C, Hill C, Grenier PA. Management of an incidentally discovered pulmonary nodule. *Eur Radiol* 2007; 17(2):449-466.
26. Ost D, Fein AM, Feinsilver SH. Clinical practice. The solitary pulmonary nodule. *N Engl J Med* 2003; 348(25):2535-2542.
27. Eberhardt R, Anantham D, Ernst A, et al. Multimodality bronchoscopic diagnosis of peripheral lung lesions: a randomized controlled trial. *Am J Respir Crit Care Med* 2007; 176(1):36-41.
28. Bertoletti L, Robert A, Cottier M, et al. Accuracy and feasibility of electromagnetic navigated bronchoscopy under nitrous oxide sedation for pulmonary peripheral opacities: an outpatient study. *Respiration* 2009; 78(3):293-300.
29. Gildea TR, Mazzone PJ, Karnak D, et al. Electromagnetic navigation diagnostic bronchoscopy: a prospective study. *Am J Respir Crit Care Med* 2006; 174(9):982-989.
30. Eberhardt R, Anantham D, Herth F, et al. Electromagnetic navigation diagnostic bronchoscopy in peripheral lung lesions. *Chest* 2007; 131(6):1800-1805.
31. Chung YH, Lie CH, Chao TY, et al. Endobronchial ultrasonography with distance for peripheral pulmonary lesions. *Respir Med* 2007; 101(4):738-745.
32. Paone G, Nicastri E, Lucantoni G, et al. Endobronchial ultrasound-driven biopsy in the diagnosis of peripheral lung lesions. *Chest* 2005; 128(5):3551-3557.
33. Kurimoto N, Miyazawa T, Okimasa S, et al. Endobronchial ultrasonography using a guide sheath increases the ability to diagnose peripheral pulmonary lesions endoscopically. *Chest* 2004; 126(3):959-965.
34. Herth FJ, Ernst A, Becker HD. Endobronchial ultrasound-guided transbronchial lung biopsy in solitary pulmonary nodules and peripheral lesions. *Eur Respir J* 2002; 20(4):972-974.
35. Oki M, Saka H, Kitagawa C, et al. Novel thin bronchoscope with a 1.7-mm working channel for peripheral pulmonary lesions. *Eur Respir J* 2008; 32(2):465-471.
36. Rooney CP, Wolf K, McLennan G. Ultrathin bronchoscopy as an adjunct to standard bronchoscopy in the diagnosis of peripheral lung lesions. A preliminary report. *Respiration*2002; 69(1):63-68.
37. Cazzadori A, Di Pewrri G, Marocco S, et al. Staging bronchoscopy in the preoperative assessment of a solitary nodule. *Chest* 1994; 106(2):653.
38. Early detection of lung cancer: a Pan-Canadian study. The Terry Fox Research Institute website. http://www.tfri.ca/programs/institute_2008-01.asp. Date last accessed 07-07-2011.

Chapter 8

Evaluation of surgical procedures

Complications following lung surgery in the Dutch-Belgian randomized lung cancer screening trial

van't Westeinde SC,
Horeweg N,
de Leyn P,
Groen HJM,
Lammers JWJ,
Weenink C,
Nackaerts K,
van Klaveren RJ

European Journal of Cardio-Thoracic Surgery
September 2012

Complications following lung surgery in the Dutch-Belgian randomized lung screening trial

Eur J Cardiothorac Surg. 2012 Sep;42(3):420-9.

ABSTRACT

To assess the complication rate in participants of the screen group of the NELSON lung cancer screening trial who underwent surgical resection and to investigate, based on a literature review, whether the complication rate, length of hospital stay, re-thoracotomy and mortality rates after a surgical procedure were different from those of the non-screening series, taking co-morbidity into account.

Between April 2004 and December 2008, 198 subjects underwent thoracic surgery. Comorbid conditions were retrieved from the medical records. Postoperative complications were classified as minor and major.

In total, 182 thoracotomies, 5 thoracotomies after video-assisted thoracoscopic surgery (VATS), and 11 VATS procedures were performed. 36% of the participants had chronic obstructive lung disease, 16% coronary artery disease, 14% diabetes mellitus and 11% peripheral vascular disease. Following thoracotomy, 47% (88/187) had ≥1 minor (7-57% in literature) and 10% (18/187) ≥1 major complication (2-26% in literature); following VATS, 38% (6/16) had ≥1 minor complication, but no major complications. Seventeen percent (3/18) of major complications and 21% (20/96) of minor complications were observed in subjects operated for benign disease. The re-thoracotomy rate was 3%, there was no 30-day mortality after thoracotomy or VATS (0-8.3% in literature). The mortality rate of 0% after surgical procedures is low compared to non-screening series (0-8.3%); the rate of complications (53%) was within the range of published non-screening series (8.5-58%).

In conclusion, mortality rates after surgical procedures were lower in the NELSON lung cancer screening trial than in non-screening series. The rate of complications was within the same range as in non-screening series.

INTRODUCTION

It has been shown that lung cancer screening by low-dose multi-detector computer tomography (CT) can detect lung cancer at an early stage.[1] Before considering implementation of CT screening, a reduction in lung cancer mortality has to be demonstrated by randomised clinical trials, and the balance between the benefits and harms of screening has to be evaluated thoroughly. Important aspects to be taken into account are the effects of CT screening on health-related quality of life, and the occurrence of complications associated with the work-up and treatment of participants with a positive test result.

Patient-related factors, such as a poor general health status, age and co-morbidity, contribute to the risk of postoperative pulmonary complications.[2] Screening populations usually consist of heavy current and former smokers at an advanced age and at high risk for co-morbid disease. In several studies, it has been demonstrated that co-morbidity is predictive of morbidity and mortality related to surgical procedures.[3] Hence, to be able to make a fair comparison with the mortality and complication rates reported in non-lung cancer screening series, the co-morbidity of the screened population should be assessed.

Our objective was to assess the complication rate in participants in the screen group of the Dutch-Belgian lung cancer screening trial (NELSON) who underwent a surgical resection, and to investigate, based on a literature review, whether the complication rate, length of stay and re-thoracotomy and mortality rates after a surgical procedure were different from those in non-screening series.

METHODS

Inclusion criteria and work-up

NELSON trial participants were current and former smokers at high risk for developing lung cancer. Detailed information on the inclusion and exclusion criteria have been reported previously.[4] Briefly, current and former smokers aged 50–75 with a smoking history of >15 cigarettes per day during >25 years or >10 cigarettes per day during >30 years (quit ≤10 years ago) were invited. Subjects with a moderate or bad self-reported health, subjects who were unable to climb two flights of stairs and persons with a body weight ≥140 kg were excluded, as were those with a history of cancer.

The prospective screening study was approved by the Ministry of Health and by the Medical Ethical Boards of each of the four participating hospitals. Written informed consent was obtained from all participants.

In the NELSON trial, 7,557 subjects underwent a CT scan at baseline, the second screening round (1 year after baseline) and the third screening round (2 years after the second round).[1] Subjects with a positive test result were referred for work-up to a pul-

monologist and, depending on the outcome of this work-up, a resection of the suspicious lesion was performed. The standard non-invasive work-up included a physical exam, pulmonary function test, bronchoscopy, FDG-PET-scan and a standard-dose CT scan with intravenous contrast of the chest and upper abdomen.

CT data acquisition and image reading

Data acquisition and image reading were as described previously.[5] In brief, all four participating screening sites used 16-detector CT scanners (Sensation-16, Siemens Medical Solutions, Forchheim, Germany Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA). Scan data were obtained in a spiral mode, with 16 × 0.75 mm collimation and 1.5 pitch. No contrast was administered. Data acquisition and scanning conditions were standardized and equal for baseline and repeat screening. Digital workstations (Leonardo®, Siemens Medical Solutions, Erlangen, Germany) were used in all screening sites with commercially available software for semi-automated volume measurements (LungCare®, Siemens Medical Solutions, version Somaris/5:
VA70C-W).

Nodule management and diagnostic work-up

At baseline, a scan was considered positive if any non-calcified nodule had a solid component >500 mm^3 (about >9.8 mm in diameter) and indeterminate if the volume of the largest solid nodule or the solid component of a partially solid nodule was 50-500 mm^3 (about 4.6-9.8 mm in diameter), or >8 mm in diameter for non-solid nodules.[5] Subjects with an indeterminate result underwent a follow-up scan after three months to assess nodule growth. Significant growth was defined as a change in volume between the first and second scan of ≥25%. Subjects with positive screening tests were referred to a chest physician for work-up and diagnosis.[1] If lung cancer was diagnosed, the participant was treated for the disease and no longer underwent screening; if no lung cancer was found the regular second-round CT scan was scheduled twelve months after baseline scan. For participants with one or more new nodules on the second-round scan, the result (positive or negative) was based on the size of the nodule, as for round one; in the case of an indeterminate result, a follow-up scan was performed 6 weeks later.[5] For participants with previously existing nodules, the second-round result was based on the volume doubling-time (VDT). If there was no growth, or if the VDT was >600 days, the scan was classified negative.[1] If the VDT was <400 days, or if a new solid component had emerged in a previously non-solid nodule, the scan was considered positive. When the VDT was 400-600 days, the test was classified indeterminate and follow-up scanning was performed one year after the second round. For nodules with a VDT of <400 days, the final result was considered to be positive. If both new and existing nodules were present, the nodule with the largest volume or fastest growth determined the screening test result.

All participants with a negative second-round result were invited to undergo the third screening two years after the second round.

Work-up and staging were standardized for all screening sites according to national and international guidelines and included a physical exam, a standard CT scan with contrast of the chest and upper abdomen, FDG-PET scan and a bronchoscopy.[5] Subjects with a negative non-surgical work-up were referred for surgery to obtain histology of the suspicious nodule. Bronchoscopies were done in accordance with Dutch national guidelines in order to evaluate the central airways and (if possible) to diagnose lung cancer or benign disease. Pulmonologists and thoracic surgeons were not blinded for the result of the positron emission tomography (PET) examination. All subjects with suspected lung cancer were discussed in multidisciplinary tumour boards, which included a thoracic surgeon, before progressing to surgery; all imaging studies were available during these meetings. National and international pathology review panels evaluated all cytological and histological specimens.

Operative details

All resections were performed at one of the four screening centres, of which three were academic institutions and one a peripheral hospital. In Groningen, three experienced thoracic surgeons were involved, in Haarlem two, in Louvain three and in Utrecht eleven. Participants with a benign diagnosis after non-surgical work-up were scheduled for the next screening round. In the remaining test-positive subjects, the suspicious nodules were removed either by VATS or thoracotomy with wedge resection and frozen section. A preoperative tissue biopsy was not routine. Lobectomies were performed only for central nodules that could not be approached by wedge resection, meaning limited resections were performed for benign lesions. If lung cancer was diagnosed by VATS, the procedure was converted to an open thoracotomy with sampling of lobar, interlobar, hilar and mediastinal lymph nodes. This is because VATS resection for lung cancer was not yet fully implemented in daily practice in the Netherlands at the time of the present study. A mediastinoscopy was performed before proceeding to VATS or thoracotomy in subjects with mediastinal lymph nodes larger than 10 mm in the short-axis and/or FDG-PET positive mediastinal lymph nodes. No specific strategies were employed to prevent prolonged air leak, such as reinforced staple lines. The chest tube was removed if there was no air leak and the fluid production was 200 ml or less per 24 hours.

Data collection and co-morbidity scoring

The date, nature, number and outcome of all adverse events related to all diagnostic and treatment procedures between April 2004 and 31 December 2008 were entered into an web-based database 'the NELSON Management System' by investigators at the four screening sites after completion of the diagnostic work-up and therapeutic procedures.

In addition, a hard copy of the medical records of all subjects referred for work-up and treatment was centrally stored at the data centre of Erasmus MC Rotterdam in order to review for complications.

Co-morbid conditions were retrieved from the medical records based on the medical history at the time of referral because of a positive screening test result. Subjects were defined as having chronic obstructive pulmonary disease (COPD) when the forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) ratio was <0.70 and/or the medical history mentioned COPD and the participant used inhaled steroids and/or bronchodilators. Coronary artery disease included a history of myocardial infarction, coronary artery bypass graft, percutaneous transluminal coronary angioplasty or angina pectoris. Peripheral vascular disease included a history of intermittent claudication, abdominal aneurysm, percutaneous transluminal angioplasty or bypass grafting of the peripheral arteries.

Literature search

A review of the literature was performed using a Pubmed search up to February 2011. The search string consisted of a combination of medical subject headings [MeSH] and keywords including 'Lung Neoplasms', 'Postoperative Complications', 'Co-morbidity',

'Mortality', 'Thoracotomy', 'Thoracic Surgery, Video-Assisted' and related synonyms. We summarised the main results of the literature study and the current study with regard to co-morbidity and adverse events following thoracotomy in forest plots. This was not done for VATS procedures in view of the low number of VATS procedures in the current study. Lack or incomplete reporting of comorbidity was not used as an exclusion criterion. Studies in which all participants had previously received chemotherapy and/or radiotherapy were excluded. For studies without classification of complications, complications were scored according to our definitions of minor and major complications, provided that a complete overview of all complications was reported. In addition, in the case of studies that graded complications based on the Common Terminology Criteria for Adverse Events, grade 4–5 events were considered major complications.

Definitions of complications

Postoperative mortality was defined as death within 30 days after the operation or within the same hospital admission. According to EuroSCORE[6] and Birim et al.[7] major complications included bleeding requiring re-operation, empyema, pneumonia (Center for Disease Control and Prevention definition of nosocomial pneumonia)[2], myocardial infarction, renal failure requiring temporary or permanent dialysis, postoperative stroke, critical arrhythmia (ventricular fibrillation, ventricular tachycardia) and pulmonary embolism. Additional major complications included respiratory failure requiring ventilator support for >48 hours[8] and postoperative heart failure with pulmonary oedema.[9] We

classified a chylothorax, haemothorax and gastro-intestinal complications requiring operative re-intervention (re-thoracotomy) or laparotomy as major complications. Non-life threatening complications were classified as minor complications. All minor and major complications were scored for each VATS and thoracotomy procedure.

Statistical analysis

Data were analysed using SPSS (version 17.0, SPSS, Inc., Chicago, IL, USA). A two-tailed Mann–Whitney U-test was used to analyse continuous data in the absence of normal distribution. Chi²-test was used for binomial or categorical data and Fisher's exact test for small groups. Statistical significance was defined as a p-value <0.05. Asymmetric confidence intervals (CI) were calculated for the literature study data presented in Figures 1 and 2 using log-linear regression, where we estimated the observation as the log of a β; a weighted standard error (SE) was calculated for this β and subsequently the CI was obtained.

RESULTS

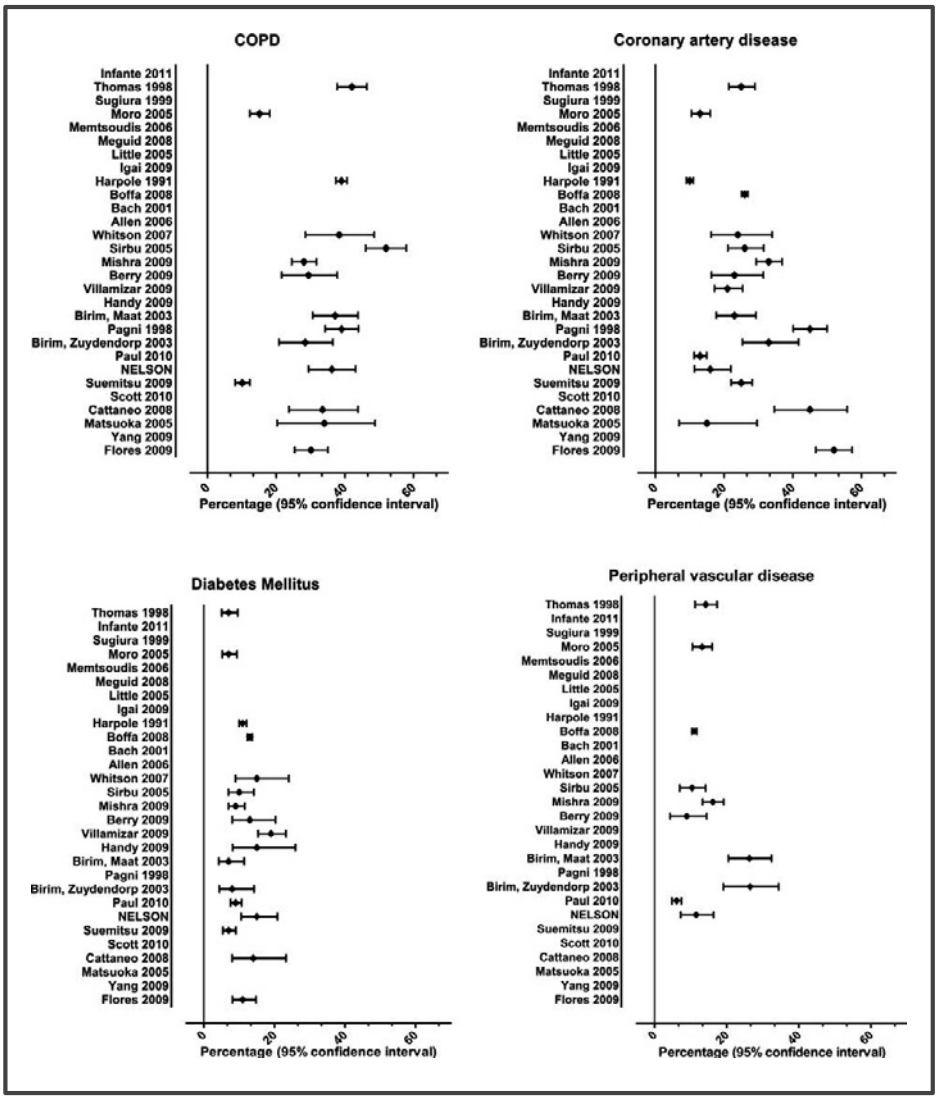
Background and treatment characteristics

A total of 415 subjects had a positive test result following CT screening between April 2004 and December 2008. The role of FDG-PET in the work-up of these test-positive participants has been described elsewhere. In seventeen of the participants surgical procedures consisted of a mediastinoscopy only; fifteen were subsequently diagnosed with lung cancer, which was at an early stage in two, who were inoperable because of co-morbidity (Figure 3). In 178 participants, the final benign diagnosis was based on FDG-PET, CT with intravenous contrast or biopsies. Transthoracic biopsies were only performed in 5% (22/415) of test-positive participants.

In twenty-two participants cancer was diagnosed but the subjects did not undergo resection because of: advanced stage disease (n=13) or co-morbidity (n=9), the latter group was treated with stereotactic radiotherapy. In the twenty-two subjects, diagnosis was based on biopsy (n=15) cases, or imaging studies (n=7). In 198 participants, non-surgical work-up showed lung cancer or was inconclusive. These subjects underwent a resection either via thoracotomy (n=182), VATS converted to thoracotomy (n=5), or wedge resection by VATS (n=11) (Figure 3).

The characteristics of the subjects who underwent a resection are presented in Tables 1 and 2. The most frequent comorbid conditions were COPD (36%), coronary artery disease (16%), diabetes mellitus (14%) and peripheral vascular disease (11%) (Table 2). Table 1 shows the clinical and pathological lung cancer stages. Three subjects with clinical stage III (T4N0M0) were operated and a microscopic complete resection could

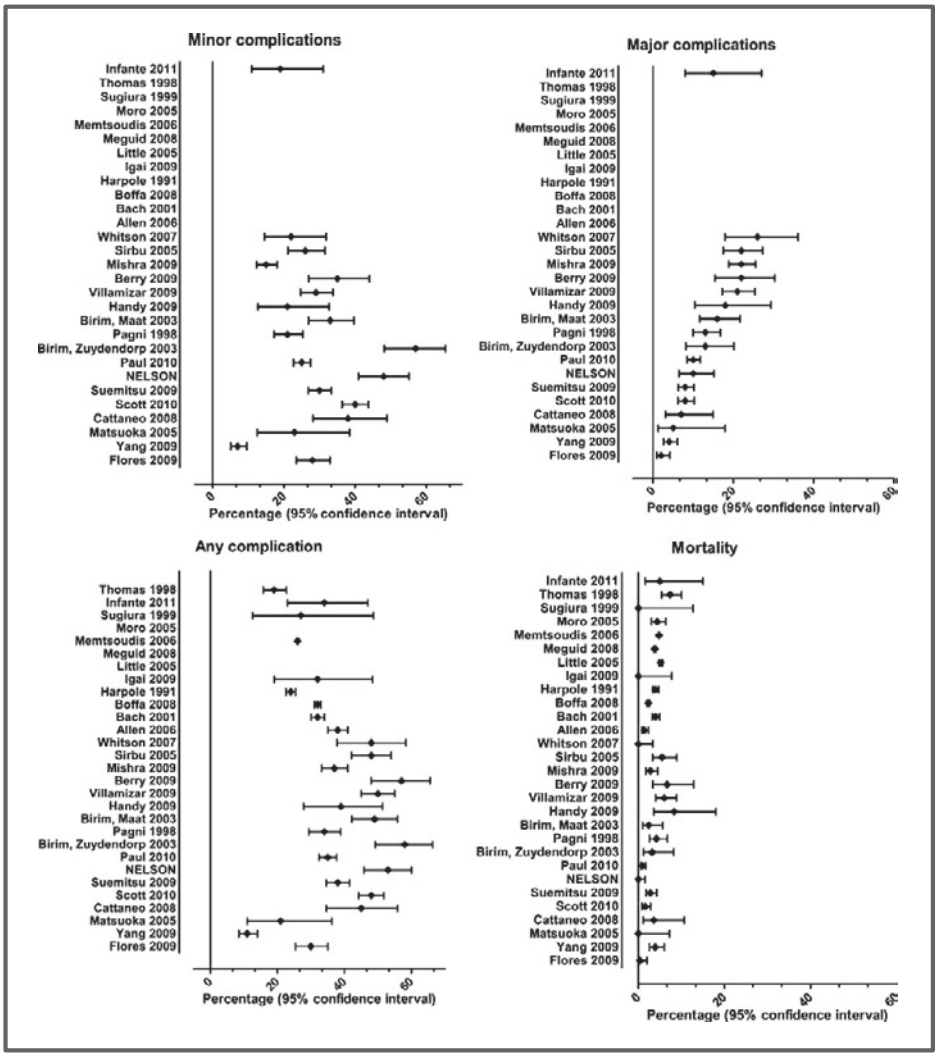
Figure 1. Prevalence of comorbidity in subjects undergoing thoracotomy



Prevalence of chronic obstructive pulmonary disease, coronary artery disease, diabetes mellitus and peripheral vascular disease in NELSON lung cancer screening participants who underwent a thoracotomy in comparison with non-screening series from the literature.

be achieved. Five subjects had pathological stage IV lung cancer after surgery. Two of them had an indeterminate preoperative FDG-PET result, which in retrospect appeared to be metastatic lesions. In two patients the preoperative FDG-PET was false-negative for distant metastasis. In one subject, no preoperative FDG-PET was made due to an

Figure 2. Prevalence of complications and mortality in subjects undergoing thoracotomy



Complication and mortality rates after thoracotomy in the NELSON lung cancer screening trial in comparison with non-screening series from the literature.

administrative error; the postoperative FDG-PET scan showed distant metastasis. Of eight patients with pathological stage III disease, this was due to unforeseen N2 disease in seven patients and due to a bronchoalveolar carcinoma in the middle lobe, which was resected in one patient; a second suspicious upper lobe nodule could not be found during surgery. The clinical stage at that time was cT1N0M1. One month later the upper lobe

nodule showed rapid growth and mediastinal lymphadenopathy was noted (clinical stage T1N2M0); mediastinoscopy showed metastasis of a large cell carcinoma.

Table 1. Characteristics of participants who underwent surgery after a positive screening test

Characteristics	Lung surgery n (%)
Female gender	35 (18)
Age - mean(range)	61 (50-74)
Pack-years - mean(range)	46 (21-133)
COPD	71 (36)
GOLD I	38 (19)
GOLD II	20 (10)
GOLD III	6 (3)
GOLD stage unknown	7 (4)
Type of surgery	
Pneumonectomy	4 (2)
(Bi)lobectomy	137 (70)
True cut biopsy, segment/wedge resection	56 (26)
Sternotomy	1 (0.5)
Diagnosis	
Lung cancer	139 (70)
Other cancer	12 (6)
Benign abnormalities	47 (24)
Clinical lung cancer stage[a]	
I	117 (84)
II	18 (13)
III	3 (2)
IV	1 (1)
Pathological lung cancer stage	
I	112 (81)
II	11 (8)
III	11 (8)
IV	5 (4)
Total	**198 (100)**

Definition of abbreviations: COPD = Chronic Obstructive Pulmonary Disease, GOLD = Global initiative for chronic Obstructive Lung Disease.

[a] *Sixth edition of TNM classification for lung cancer.*

Table 2. Comorbidity of participants who underwent surgery after a positive screening test

Comorbidity	Lung surgery n (%)
Any primary tumour	16 (8)
Chronic obstructive pulmonary disease	71 (36)
Congestive heart failure	5 (3)
Coronary artery disease	31 (16)
Cerebrovascular disease	5 (3)
Peripheral vascular disease	22 (11)
Diabetes mellitus	28 (14)
Chronic kidney disease	2 (1)
Connective tissue disease	4 (2)
Peptic ulcer disease	7 (4)
Total	**198 (100)**

Complications after surgery

Tables 3 and 4 present all complications observed. Following thoracotomy, 47% (88/187) had at least one minor and 10% (18/187) at least one major complication. Thirty-eight percent (6/16) of the VATS procedures was complicated by at least one minor complication, but no major complications have been observed. As 5% had both minor and major complications, the proportion of participants with any complication was 53%. Seventeen percent (3/18) of major complications and 21% (20/96) of minor complications were seen in subjects operated for benign disease.

The overall median length of hospital stay was 13 days (2-51 days) after thoracotomy and 8 days (4-12 days) after VATS. In subjects with minor complications, this was 15 days (6-51 days) and 9 days (7-12 days), respectively. In the case of major complications following thoracotomy, the median length of stay was 21 days (range 8-51 days). The re-thoracotomy rate was 3% after thoracotomy and 0% after VATS. Re-admissions occurred in 5% of those who underwent a thoracotomy (eight after minor complications and one after a major complication), but were absent after VATS. There was no 30-day mortality after thoracotomy or VATS in the NELSON trial.

Table 5 shows that a higher rate of minor complications was seen in the case of more extensive resections. Limited resections (true-cut biopsies, and wedge and segment resections) had lower rates of minor complications (OR 0.51, 95% CI 0.26-1.03, p-value 0.06) compared to bilobectomy, lobectomy and pneumonectomy. No significant correlation could be established between type of resection and risk of major complications.

Five subjects were re-admitted because of minor complications: three subjects with chest pain and one with dyspnoea a pulmonary embolism could be excluded; in one

Table 3. Minor complications following thoracotomy and VATS procedures

Minor complication	Thoracotomy n (%)	VATS n (%)
Air-leakage >5 days	42 (23)	5 (31)
Supraventricular tachycardia	17 (9)	0
Infection	16 (9)	1 (6)
Diaphragm paralysis	10 (5)	0
Chest tube >5 days	8 (4)	0
Atelectasis	8 (4)	1 (6)
Drop hand	3 (2)	0
Wound infection	3 (2)	0
Delirium	3 (2)	0
Chest pain	3 (2)	0
COPD exacerbation	2 (2)	2 (13)
Blood transfusion	2 (2)	0
Urinary retention	2 (2)	0
Haemoptysis	2 (2)	0
Persistant ptosis	1 (1)	0
Paralysis serratus anterior muscle	1 (1)	0
Deep venous thrombosis	1 (1)	0
Ileus	1 (1)	0
Pleuritic effusion	1 (1)	0
Dyspnoea	1 (1)	0

Definition of abbreviations: VATS = video-assisted thoracoscopic surgery; COPD = Chronic Obstructive Pulmonary Disease.

Table 4. Major complications following thoracotomy

Major complication	n (%)
Pneumonia	10 (5)
Empyema	2 (1)
Bleeding, re-operation	1 (0.5)
Chylothorax, re-operation	1 (0.5)
Pulmonary embolism	1 (0.5)
Respiratory failure	1 (0.5)
Myocardial infarction	1 (0.5)
Congestive heart failure	1 (0.5)
Ventricular tachycardia	1 (0.5)
Bowel perforation	1 (0.5)

subject with pleural effusion an empyema was excluded; no repeat chest tube placement or thoracocentesis was necessary (Table 3). Atelectasis was diagnosed by chest radiograph in nine subjects; in five subjects bronchoscopy was performed.

Table 5. Complications according to type of surgery

Thoracotomy		Minor complications				Major complications				VATS	Minor complications				
Type of surgery	**n (%)**	**1**	**2**	**Any**	**Total**	**1**	**2**	**Any**	**Total**	**n (%)**	**1**	**2**	**3**	**Any**	**Total**
True cut biopsy	6 (3)	2	0	2 (33)	2	0	0	0 (0)	0	0 (0)	0	0	0	0 (0)	0
Wedge resection	35 (19)	8	3	11 (31)	14	3	0	3 (9)	3	16 (100)	4	1	1	6 (38)	9
Segmentectomy	4 (2)	1	0	1 (25)	1	1	0	1 (25)	1	0 (0)	0	0	0	0 (0)	0
Lobectomy	131 (70)	40	26	69 (53)	101	11	2	13 (10)	15	0 (0)	0	0	0	0 (0)	0
Bilobectomy	5 (3)	1	2	4 (80)	8	1	0	1 (20)	1	0 (0)	0	0	0	0 (0)	0
Sleeve resection	1 (1)	0	0	0 (0)	0	0	0	0 (0)	0	0 (0)	0	0	0	0 (0)	0
Pneumonectomy	4 (3)	1	0	1 (25)	1	0	0	0 (0)	0	0 (0)	0	0	0	0 (0)	0
Sternotomy	1 (1)	0	0	0 (0)	0	0	0	0 (0)	0	0 (0)	0	0	0	0 (0)	0
Total	**187 (100)**	**53**	**31**	**88 (47)**	**127**	**16**	**2**	**18 (10)**	**20**	**16 (100)**	**4**	**1**	**1**	**6 (38)**	**9**

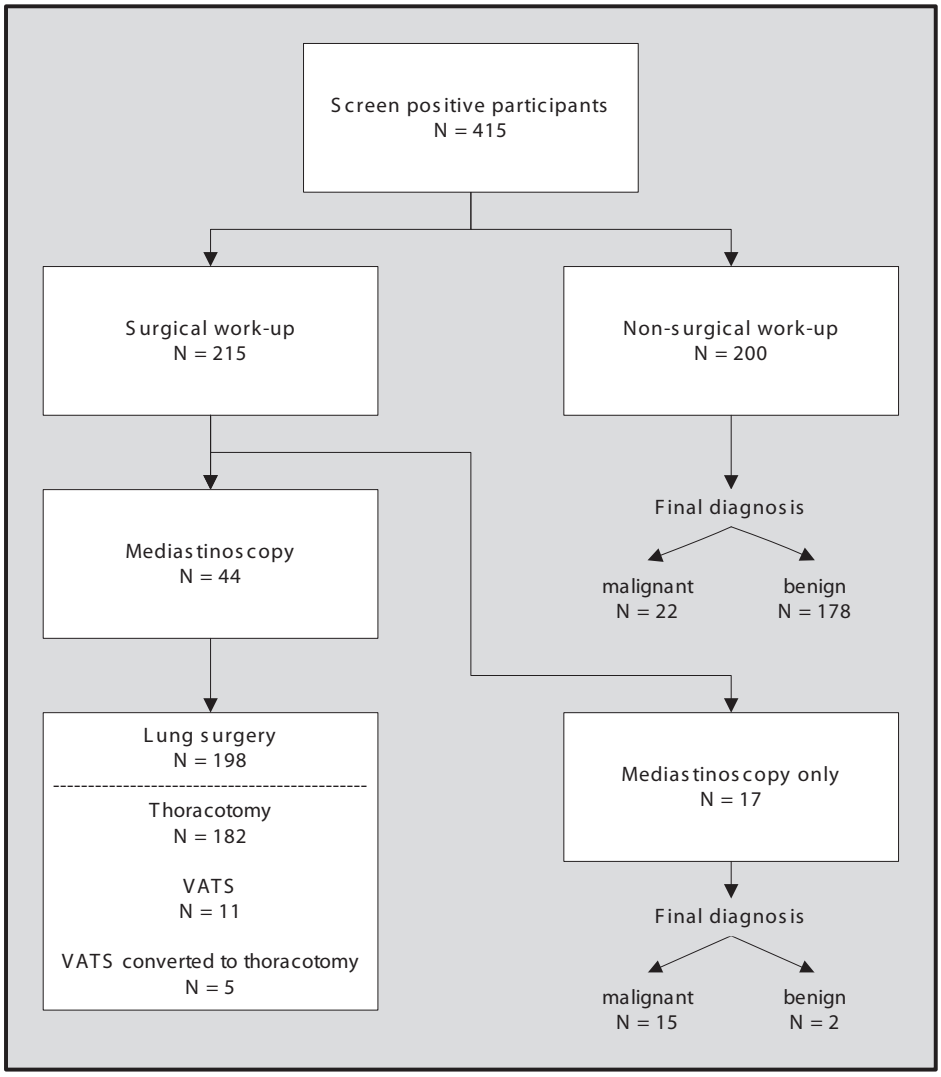
Data from the literature review

Literature search revealed sixteen studies on thoracotomy and twelve studies on both thoracotomy and VATS which met the selection criteria (Appendix Table 1). The prevalence of comorbidity in the literature on thoracotomy ranged from 43% to 80%[3,7] (Appendix Table 1). Figure 1 shows the prevalence of co-morbidities in subjects who underwent a thoracotomy in non-lung cancer screening studies. The most frequently reported co-morbid conditions were COPD (10–52%), coronary artery disease (10-52%), diabetes mellitus (7-19%) and peripheral vascular disease (6-26%). The number of lobectomies performed during thoracotomy procedures varied from 40% to 100%.[22] (Suemitsu et al. 2009), while pneumonectomies were performed in 0-27%[10,14] (Pagni et al. 1998). The percentage of stage I disease in the thoracotomy group was 69.0% (median, range 31-100%).[10,11] Mortality rates reported after thoracotomy varied from 0% to 8% (Figure 2). The National Emphysema Treatment Trial (NETT) found a 90-day mortality of 5% following lung volume reduction surgery in subjects with severe emphysema (mean FEV1 0.7 l; 26% of predicted).[12] Figure 2 shows major and minor complications after thoracotomy, ranges varied respectively from 4-26% and 7-57%. The median length of hospital stay after thoracotomy reported in the literature was 5-22 days[13] (Boffa et al, 2008). The reported re-thoracotomy rates after a thoracotomy varied from 0% to 9%.[7,13]

In the majority of the studies on VATS, lobectomies were performed. Complications after VATS were reported in 9-51% (Kim et al. 2010, Petersen et al. 2010) and major

complications in only 0-12%[17] (Jaklitsch et al. 1996, Petersen et al. 2010). The median length of stay after a VATS reported in the literature was 4-23 days[13] (Villamizar et al. 2009). The reported re-operation rate after a VATS varied from 1% to 5%[13] (Paul et al. 2010), with a mortality rate of 0-4%[21] (Handy et al. 2009).

Figure 3. Flowchart of surgical procedures and outcomes



Surgical procedures and outcomes in 415 screen positives of the NELSON randomised lung cancer screening trial.

DISCUSSION

Our study compared the complications rates, length of hospital stay, rethoracotomy and mortality rates of participants of the NELSON trial who underwent thoracic surgery with data from non-screening series. The comparison with non-screening series could be made because we demonstrated that the age range and co-morbidity level of the NELSON trial participants who underwent a surgical resection was the same as those in the non-screening series.

Literature review and complications

The studies included in our literature review displayed a large heterogeneity with respect to the definition, classification and way in which data on complications have been collected so far. For example, prolonged air leak has been defined as >5 days[7] and >7 days.[13] In addition, chest-tube management with regard to output differs in studies or is not defined. Few authors make a distinction between minor and major complications, and complication data are collected by reviewing individual patient charts, based on ICD-9 codes[14] or on claims in Medicare files.[3] The latter methods may lead to underreporting of complications, especially for minor complications. Probably because we screened all individual patient files, our minor complication rates are in the higher range of what has been reported before. The most important observation was the relatively low rate of major complications and the absence of postoperative mortality after the thoracotomy and VATS procedures performed in the screen group of the NELSON trial. This could probably be explained by the fact that screening participants were asymptomatic individuals, screen-detected tumours are usually smaller[15,16] and that pneumonectomies were less often required in the NELSON compared to published studies, wherein more complex resections with a higher expected complication rate were performed. Nevertheless, the proportion of stage I disease was in the same range of what has been reported in our literature review of the non-screening series.

Lung cancer-screening studies and complications

In a recent study, Infante et al.[17] report on the outcome of surgical procedures in the DANTE trial. A total of 59 subjects underwent a thoracotomy procedure. Three died following the thoracotomy and a total of twenty complications were noted, which were major complications in nine subjects. No major complications or postoperative deaths were seen in subjects diagnosed with benign disease. Fifteen subjects underwent a VATS procedure; no postoperative deaths or major complications were noted in this subset of patients. The postoperative mortality rate in the DANTE study was higher than expected. All subjects had central tumours of stage IIA or higher, two had co-morbid conditions and two had undergone a pneumonectomy. Veronesi et al.[18] reported that 25% of subjects

developed complications following thoracotomy and VATS procedures, which were serious in 6% and required re-operation in 2%. No postoperative complications or mortality was noted in the subjects with benign disease. While only two subjects underwent a pneumonectomy in this study,[18] pneumonectomies were performed on four subjects in our study. Infante et al.[17] performed a relatively high number of pneumonectomies, in seven subjects in total, which may explain the higher mortality rate. The rate of major complications in lung cancer screening studies is at the lower limit of the range published in the literature. Mortality rates are also at the lower limit[18], however with more extensive resections the rate may be the same as in the literature.[17] An important observation to make is that no major complications and no deaths were seen in subjects operated for benign disease.[17,18] However, in our study 17% (3/18) of major complications and 21% (20/96) of minor complications were observed in subjects operated for benign disease.

Length of stay after VATS and thoracotomy

Despite these observations, the length of stay (LOS) after thoracotomy and VATS procedures was not shorter for NELSON participants than the average. This can be explained by the fact that patients in the Netherlands and Belgium usually stay in the surgery or pulmonary medicine department and do not routinely go to a short-stay facility after surgery. It has been shown that LOS decreases when the use of skilled nursing facilities increases.[19] Another possible explanation may be that in the Netherlands and Belgium it is socially much less accepted to discharge patients home after three or four days. None of the participants in the NELSON study went to a long-term nursing facility. Prolonged air leak has been described as the most important factor for prolonged hospital stay[20], this was not the case in the current study, presumably because of less severe emphysema.

Type of resection

In the NELSON trial, VATS procedures were only performed for wedge resections, whereas in the majority of studies VATS was used to perform lobectomies, which is a major difference. This is due to the fact that VATS lobectomy had only recently been introduced in the Netherlands at the time of the study.[21] The proportion of lobectomies in the thoracotomy group was comparable with the literature. Therefore, and because of the low number of VATS procedures in the current study, the comparison we made between the VATS results in the NELSON screening trial and the non-screening series from the literature should be interpreted with caution. There is general consensus in the literature that morbidity and mortality rates after VATS are lower than after thoracotomy, and that patients have a better postoperative physical functioning and a shorter postoperative length of stay.[22] In addition, the oncological validity of VATS resections for lung cancer has been proved as 5-year survival rates are similar to those after thoracotomy.[23] We therefore believe that lung cancer screening sites should be equipped to perform VATS

procedures, especially in view of the substantial risk of false-positive test results and resections for benign disease.[1]

SUMMARY

In the NELSON lung cancer screening trial, the rate of minor complications after thoracotomy and VATS was in the upper range of what has been reported for the non-screening series, while the rate of major complications was in the lower range. The postoperative length of stay was not shorter than in the literature. The re-thoracotomy rate for complications such as a haemothorax requiring re-intervention in the NELSON trial was in the range reported in the literature, but no re-thoracotomies were performed after VATS. No postoperative deaths were observed after the thoracotomy and VATS procedures.

To our knowledge, this is the first report on the prevalence of co-morbidity and of complications in a lung cancer screening population. Veronesi et al. presented data on complications as an abstract without information on co-morbidity.[18] Their results support our encouraging data, which demonstrate that participants are at low risk of major complications or postoperative death following thoracotomy or VATS lung cancer screening. Nonetheless, the high rate of resection for benign disease and associated morbidity continues to be a concern. Seventeen percent of the major complications and 21% of the minor complications were seen in subjects operated for benign disease. The use of FDG-PET[24] and combination of FDG-PET and VDT[25] may help to reduce the number of resections for benign disease.

In conclusion, mortality rates after surgical procedures were lower in the NELSON lung cancer screening trial than in non-screening series. The rate of minor and major complications is within the range of non-screening series.

SUPPORT STATEMENT

The NELSON trial is financially supported by Zorg Onderzoek Nederland-Medische Wetenschappen (ZonMw), KWF Kankerbestrijding, Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen (RvvZ), G. Ph. Verhagen Foundation, Rotterdam Oncologic Thoracic Study Group (ROTS), Erasmus Trust Fund, Stichting tegen Kanker (Belgium), Vlaamse Liga tegen Kanker and LOGO Leuven and Hageland. We also wish to thank Tom and Josephine De Rijke for their legacy gift.

CONFLICTS OF INTEREST

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.

ACKNOWLEDGEMENTS

We thank the surgeons at the four screening sites who performed the surgeries on participants in the NELSON study. Groningen: D.J. Drenth, T.J. Klinkenberg and Y.N. Drijver; Haarlem: H. Rijna and H.L.F. Brom. Louvain: G. Decker and E. Internullo and Utrecht: J. Kluin, P.F.A. Bakker-de Wekker, R.C.A. Meijer and F.Z. Ramjankhan. We thank Harry de Koning (Department of Public Health, Erasmus MC Rotterdam), Matthijs Oudkerk (Department of Radiology, UMC Groningen), Willem Mali (Department of Radiology, UMC Utrecht) and Frederik Thunnissen (Department of Pathology, VUMC Amsterdam) for their critical review of the manuscript and their comments, which have been appreciated. Our thanks to René Vernhout (Department of Pulmonology, Erasmus MC Rotterdam) for developing the database used to register complications and data management support. We thank Roel Faber, ICT manager, for his assistance, and Linda van Dongen for her support in data management; and also the local data managers: Henk Pruiksma (Haarlem), Liesbet Peeters (Louvain), Saskia van Amelsvoort-van de Vorst (Utrecht) and Ria Ziengs (Groningen). Finally, our thanks to Caspar Looman for his assistance in statistics.

REFERENCES

1. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 23:2221-9.
2. Arozullah AM, Khuri SF, Henderson WG, et al. Participants in the National Veterans Affairs Surgical Quality Improvement P. Development and validation of a multifactorial risk index for predicting postoperative pneumonia after major non-cardiac surgery. *Ann Intern Med* 2001; 135:847-57.
3. Bach PB, Cramer LD, Schrag D, et al. The influence of hospital volume on survival after resection for lung cancer. *N Engl J Med* 2001; 345:181-8.
4. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120:868-74.
5. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54:177-84.
6. Hirose H, Inaba H, Noguchi C, et al. EuroSCORE predicts postoperative mortality, certain morbidities, and recovery time. *Interact CardioVasc Thorac Surg* 2009; 9:613-7.
7. Birim O, Zuydendorp HM, Maat AP, et al. Lung resection for non-small-cell lung cancer in patients older than 70: mortality, morbidity, and late survival compared with the general population. *Ann Thorac Surg* 2003; 76:1796-801.
8. Johnson RG, Arozullah AM, Neumayer L, et al. Multivariable predictors of postoperative respiratory failure after general and vascular surgery: results from the patient safety in surgery study. *J Am Coll Surg* 2007; 204:1188-98.
9. Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major non-cardiac surgery. *Circulation* 1999; 100:1043-9.
10. Yang X, Wang S, Qu J. Video-assisted thoracic surgery (VATS) compares favorably with thoracotomy for the treatment of lung cancer: a five-year outcome comparison. *World J Surg* 2009; 33:1857-61.
11. Moro-Sibilot D, Aubert A, Diab S, et al. Comorbidities and Charlson score in resected stage I nonsmall cell lung cancer. *Eur Respir J* 2005; 26:480-6.
12. Sanchez PG, Kucharczuk JC, Su S, et al. National Emphysema Treatment Trial redux: accentuating the positive. *J Thorac Cardiovasc Surg* 2010; 140:564-72.
13. Sugiura H, Morikawa T, Kaji M, et al. Long-term benefits for the quality of life after video-assisted thoracoscopic lobectomy in patients with lung cancer. *Surg Laparosc Endosc Percutan Tech* 1999; 9:403-8.
14. Memtsoudis SG, Besculides MC, Zellos L, et al. Trends in lung surgery: United States 1988 to 2002. *Chest* 2006; 130:1462-70.
15. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 354:99-105.
16. Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002; 165:508-13.
17. Infante M, Chiesa G, Solomon D, et al. Surgical procedures in the DANTE trial, a randomized study of lung cancer early detection with spiral computed tomography: comparative analysis in the screening and control arm. *J Thorac Oncol* 2011; 6:327-35.
18. Veronesi G, Spaggiari L, Preda P, et al. Morbidity and mortality after surgical treatment of screening detected lung cancer. *J Clin Oncol* 2008; 26(Supp.No.15S):7531.
19. Bohmer RM, Newell J, Torchiana DF. The effect of decreasing length of stay on discharge destination and readmission after coronary bypass operation. *Surgery* 2002; 132:10-5.

20. Bardell T, Petsikas D. What keeps postpulmonary resection patients in hospital? *Can Respir J* 2003; 10:86-9.
21. Belgers EH, Siebenga J, Bosch AM, et al. Complete video-assisted thoracoscopic surgery lobectomy and its learning curve. A single center study introducing the technique in Netherlands. *Interact CardioVasc Thorac Surg* 2010; 10:176-80.
22. Cattaneo SM, Park BJ, Wilton AS, et al. Use of video-assisted thoracic surgery for lobectomy in the elderly results in fewer complications. *Ann Thorac Surg* 2008; 85:231-5.
23. Walker WS, Codispoti M, Soon SY, et al. Long-term outcomes following VATS lobectomy for non-small cell bronchogenic carcinoma. *Eur J Cardiothorac Surg* 2003; 23:397-402.

APPENDIX

Table 1. Outline studies on thoracotomy and studies on thoracotomy and VATS

Study	Period	Participants	Age*
Pagni	1971-1996	385	75 (70-96)
Thomas	1975-1996	500	74 ± 3
Moro	1979-2003	588	63 (33-86)
Igai†	1982-2008	37	82 ± 2.3
Bach	1985-1996	2,118	≥ 65
Sirbu	1986-2001	273	73 (70-88)
Memtsoudis	1988-2002	512,758	62 (1-91)
Birim	1989-2001	125	74 (70-82)
Harpole	1991-1995	3,516	64 (22-91)
Birim	1996-2001	205	64 (29-82)
Suemitsu	1996-2006	756	20-90
Yang†	1996-2003	508	52 (23-79)
Matsuoka	1997-2004	40	82 (80-88)
Sugiura†	1997-1998	22	61 ± 9
Meguid	1998-2003	26,310	66
Handy†	1998-2007	64	64
Whitson†	1998-2005	88	65
Allen	1999-2004	1,023	68 (23-89)
Berry†	1999-2007	119	76 ± 0.2
Boffa	1999-2006	9,033	67 (20-94)
Scott†	1999-2004	686	68 ± 9
Villamizar†	1999-2008	382	64 ± 11
Little	2001	11,668	67
Mishra	2001-2005	597	69 (63-74)
Flores†	2002-2007	343	67 (35-89)
Cattaneo†	2002-2005	82	76 (70-89)
Paul†	2002-2007	1,281	65 ± 12
NELSON	2004-2008	187	62 (50-74)
Infante†	2001-2009	59	64 (64.0-64.7)

Definition of abbreviations: VATS = video-assisted thoracoscopic surgery.

* *Median (range) or, mean ± standard deviation.*

† *Thoracotomy arm.*

Table 2. Outline of studies on VATS and studies VATS and thoracotomy

Study	Period	Participants n	Age*
Igai†	1982-2008	58	83 ± 2.4
Jaklitsch	1991-1994	307	65-90
McKenna	1992-2004	1,100	71 (16-94)
Walker	1992-2001	178	66 (43-85)
Congregado	1993-2006	237	61 (12-79)
Yang†	1996- 2003	113	54 (9-77)
Sugiura†	1997-1998	22	62 ± 12
Handy†	1998-2007	49	63
Whitson†	1998-2005	59	67
Berry†	1999-2007	219	76 ± 0.2
Scott†	1999-2004	66	71 ± 9.7
Villamizar†	1999-2008	697	67± 10
Nakanishi	2000-2006	58	70 (52-90)
Cattaneo†	2002-2005	82	76 (70-88)
Flores	2002-2007	328	67 (36-90)
Paul†	2002-2007	1,281	65 ± 12.1
Kim	2003-2008	704	57 (12-86)
Petersen	2005-2008	197	65 (44-85)
Belgers	2006-2008	70	66 (41-85)
Infante†	2001-2009	15	64 (64.0-64.7)
NELSON	2004-2008	16	61(52-72)

Definition of abbreviations: VATS = video-assisted thoracoscopic surgery.

* *Median (range) or, mean ± standard deviation.*

† *VATS arm.*

Part IV

Evaluation of effectiveness

Chapter 9

Endpoint determination

Uniform and blinded cause of death verification in a lung cancer CT screening trial.

Horeweg N,
van Klaveren RJ,
Groen HJM,
Lammers JWJ,
Weenink C,
Nackaerts K,
Mali W,
Oudkerk M,
de Koning HJ

Lung Cancer
September 2012

Uniform and blinded cause of death verification in a lung cancer CT screening trial

Lung Cancer. 2012 Sep; 77(3):522-5.

ABSTRACT

Disease-specific mortality is the final outcome of a lung cancer screening trial, therefore cause of death verification is crucial. The use of death certificates for this purpose is debated because of bias, inaccurate completion and incorrect ante mortem diagnoses. A cause of death evaluation process was designed to ensure a uniform and unbiased determination of the graduation of certainty that lung cancer was the underlying cause of death. An independent clinical expert committee will review the medical files of all deceased participants once diagnosed with lung cancer and will make use of a flow chart and predetermined criteria. A pilot study of fifty cases was conducted to determine the performance of this process and to compare the outcome with the official death certificates. The independent review has shown an agreement of 90% (kappa 0.65), which demonstrates a uniform classification. The sensitivity and specificity of the death certificates for lung cancer specific mortality were 95.2 and 62.5%. This demonstrates a limited distinctive character of the death certification process in lung cancer patients. Our results imply that the final outcome of a lung cancer screening trial cannot reliably be established without predetermined criteria and an independent review of blinded cases.

INTRODUCTION

Lung cancer is the first cause of cancer-related death in males and the second in females globally, accounting for 1.4 million deaths per year.[1] Despite treatment advances, survival has not improved substantially over the past 30 years, mainly because the majority of the patients have distant metastasis at the time of diagnosis.[2] The early detection of lung cancer by screening asymptomatic smokers with low dose computer tomography (CT) scanning is a promising strategy to reduce lung cancer mortality, since the results of the National Lung Screening Trial (NLST) were published.[3,4]

Disease-specific mortality is the outcome of lung cancer screening; therefore, cause of death (CoD) verification is crucial. The use of death certificates for this purpose is debated for several reasons. Firstly, two forms of bias especially affect death certification in screening trials. Sticky-diagnosis bias; because lung cancer is more likely to be diagnosed in the screen arm, deaths are more likely to be attributed to lung cancer compared to the usual care arm.[5] Slippery-linkage bias; deaths as a result of interventions for lung cancer may be difficult to trace back to screening and could easily be certified as death due to other causes.[5] Secondly, the merit of death certificates depends on the accuracy of the certifying clinician and nosologist and the establishment of a correct ante mortem diagnosis.[6,7] Common reasons for misclassification are coinciding malignancies, considerable comorbidity and death after a surgical procedure.[8,9] Finally, the sensitivity and specificity of the death certificate has been reported to range from 84.5 to 99.7% of screening[9–12] and 91.3 to 99.7%; causing an error that tends to reduce the effect of screening.[9–12]

To overcome these problems clinical expert committees (CEC), reviewing the medical files of the deceased participants to determine the cause of death, are frequently employed in cancer screening trials.[9–14] The additional value of a CEC depends on the use of predetermined criteria and a thorough and independent evaluation of all cases with lung cancer blind towards each arm, to prevent an unbalanced outcome between the study arms.

We hypothesized that a clinical expert committee cannot reliably establish the outcome of a lung cancer screening trial, unless they are independent and review the medical files blinded and with predetermined criteria and flowcharts. The aim of this study is to develop a CoD review process protocol that will be used in the Dutch-Belgian lung cancer CT screening trial (NELSON). The performance of the protocol has been tested in a pilot and the outcomes will be compared with the official death certificates.

METHODS

Study design and subjects for the NELSON trial

Details of the design and conduct of the Dutch-Belgian lung cancer screening trial have been reported elsewhere.[15,16] Briefly, randomly assigned eligible participants underwent CT screening at baseline (first round), 1 year later (second round), 3 years later (third round) and 5.5-year later (fourth round) or no screening. The purpose of the trial is to determine whether at 10 years after randomisation, CT screening will have reduced mortality from lung cancer by at least 25%.[16] The trial was approved by the Dutch Minister of Health and the ethics board at each participating centre.[4] All participants provided written informed consent for the evaluation of personal data from hospital charts and national registers. The CoD evaluation process of the NELSON trial was designed to ensure a uniform and unbiased determination of the primary cause of death in participants with lung cancer.

Identification of subjects for the CoD review and data collection

The causes of death of all participants of the NELSON trial that are diagnosed with lung cancer (during their lifetime or at autopsy) are subject of the 'review process' to ensure a valid determination of the primary outcome measure of the screening trial. The lung cancer cases are identified by linkages with the national cancer registries of the Netherlands and Belgium and by checking all official death certificates for the diagnosis lung cancer, which are obtained from Statistics Netherlands and the Flemish Agency for Care and Health. For all identified cases, the diagnosis of lung cancer is verified by a pathology panel[17] or clinical experts for cases without cytology or histology. This verification process of the lung cancer diagnosis was performed separately from the CoD-verification process in the NELSON-trial and will not be addressed in this manuscript.

After the identification of the subjects, all relevant medical information will be collected and blinded for the participant's identity and study arm by an individual who is not otherwise involved in the trial. The medical files include: information provided by the general practitioner, discharge, outpatient visit letters, reports of radiology, nuclear medicine, pathology and microbiology, laboratory results, and autopsy reports.

Formation of the clinical expert committee

All cases will be reviewed and classified separately by the three members of the CEC, who are no employees of the screening trial. The committee is formed by a pulmonologist–oncologist and pathologist specialised in lung oncology and a clinical epidemiologist specialised in screening. For a random sample of 10%, cases with disagreement and all intervention related deaths the committee will meet. An international committee will be consulted in case no consensus is reached.

The cause of death evaluation process protocol

The evaluation process performed by the experts will be guided by the use a flowchart (Figure 1a-d in Appendix) and a detailed list of criteria (Table 1 in Appendix). The product of the evaluation is the classification of the cause of death of the participant in one of the six categories which define graduation of certainty that lung cancer was the primary cause of death (Table 1).

Table 1. Classification of the cause of death

Cause of death	Definition
Definitely lung cancer death	Death certainly as a direct result of (second primary) lung cancer, a paraneoplastic syndrome or a diagnostic or therapeutic intervention, including euthanasia and palliative sedation. No clear other cause of death is present.
Probable lung cancer death	Participants with (second primary) lung cancer with evidence of loco-regional or distant disease progression or a paraneoplastic syndrome. It is uncertain whether this is the final direct cause of death. No clear other cause of death is present.
Possible lung cancer death	Participants with (second primary) lung cancer with evidence of loco-regional or distant disease progression or a paraneoplastic syndrome and one or more coinciding malignancies. It is not possible to determine which malignancy was the primary cause of death.
Unlikely lung cancer death	Participants with (second primary) lung cancer, but without evidence of loco-regional or distant disease progression, a paraneoplastic syndrome or death as a result of an intervention for lung cancer. No clear other cause of death is present.
Definitely no lung cancer death	The cause of death is definitely not a direct or indirect result from (second primary) lung cancer, a paraneoplastic syndrome or an intervention for lung cancer. Another cause of death is present.
Intercurrent death with lung cancer as contributing factor	Only use this option when the cause of death cannot be classified as listed above. The cause of death is definitely not a direct result from (second primary) lung cancer. Another cause of death is present and lung cancer contributed to the death of the patient.

Design and subjects of the CoD pilot

Before the implementation of the protocol we decided to perform a pilot study by ourselves with a limited number of cases to test its user-friendliness and performance compared with the official death certificates. Therefore, we included the first fifty consecutive deceased participants diagnosed with lung cancer. In contrary to the CEC of externals to be formed for the review of all lung cancer deaths, a medical doctor (N.H.) and a clinical epidemiologist (H.J.d.K), internals of the NELSON-trial formed the committee for the pilot study. The collection and blinding of the medical files and the review process itself was performed as described. After the completion of the evaluation of the cases by

both reviewers separately, the reviewers met and discussed the cases with disagreement. Two of the pulmonologist–oncologists of the NELSON trial (H.J.M.G. and J.-W.J.L.) were consulted in case of persistent disagreement. After that, the final outcome of the pilot study was compared with the primary cause of death on the official death certificate.

Analysis

The primary cause of death is defined as 'the disease that initiated the chain of morbid events directly leading to death'. Lung cancer mortality, the primary endpoint of the study, is defined as "definitely" or "probable lung cancer death" (Table 1). "Possible", "unlikely" and "definitely no lung cancer death" and "intercurrent death with lung cancer as a contributing factor" are considered as death due to other causes (Table 1).

The agreement between the two reviewers of the CoD pilot is assessed by means of kappa statistics. A kappa of 1 and 0, respectively indicates a perfect agreement and no agreement.

The cause of death, as assigned by the review committee of the pilot after consensus meeting, is considered as the gold standard. The sensitivity and specificity of the official death certificates were defined as the proportion of lung cancer deaths assigned by both sources and as death due to other causes.

Because it is not yet allowed to analyse the data by study arm, no absolute numbers of lung cancer deaths per arm are disclosed. Therefore, it is not possible to determine if the CoD review process enhances or attenuates the effect of screening.

RESULTS

The baseline characteristics, base for the diagnosis of lung cancer and the disease stage of the fifty subjects that were included in the pilot are displayed in Table 2. The separate classification of the cause of death by the reviewers is shown in Table 3. In thirty-eight of the fifty participants (76%) the reviewers reached a concordant conclusion. The twelve remaining cases with disagreement had; significant comorbidity (n=3), multiple malignancies (n=2), death after an intervention (n=3) and death indirectly caused by lung cancer (n=4), such as death due to post-obstruction pneumonia or paraneoplastic pulmonary embolism. However, when clustering all "definitely" and "probable" lung cancer deaths into one group and "possible", "unlikely" and "definitely not" lung cancer death and "intercurrent death" into another, the differences were minimal; agreement in 45 cases (90%) resulting in a kappa of 0.65.

The comparison between the results of the CoD review, after consensus meeting, and the primary cause of death on the official certificates is displayed in Table 4. The sensitivity and specificity of the death certificates are 95.2% (95% confidence interval:

84.2-98.7%) and 62.5% (95% confidence interval: 30.6–86.3%), respectively. Disagreement was observed in 10% (5 of 50 individuals) with the following causes of death: adult respiratory distress syndrome after lobectomy, rupture of an abdominal aneurysm during chemotherapy, another malignancy besides lung cancer in two cases (breast carcinoma and acute myeloid leukaemia) and small cell lung carcinoma diagnosed after the person's death by autopsy.

Autopsy was performed in 3 (6%) of the cases. Five of the 41 (12%) lung cancer deaths involved euthanasia or palliative sedation. The place of death was in the hospital in 48%, in a hospice or nursing home in 10% and at home in 42% of the subjects. In 65% of the

Table 2. Characteristics of the fifty subjects of the pilot study

Age[a]	Mean: 62.6 years Range: 51-73 years
Gender	Male: 42/50 (84%) Female: 8/50 (16%)
Base for the diagnosis lung cancer	Surgical resection of primary tumour: 16/50 (32%) Histology or cytology of primary tumour: 15/50 (30%) Histology or cytology of lymph node metastasis: 6/50 (12%) Histology or cytology of distant metastasis: 8/50 (16%) Autopsy: 1/50 (2%) Clinical picture and imaging techniques: 4/50 (8%)
Disease stage at diagnosis[b]	Ia: 12/50 (24%) IIa: 2/50 (4%) IIb: 1/50 (2%) IIIa: 6/50 (12%) IIIb: 3/50 (6%) IV: 26/50 (52%)

[a] *Age at the inclusion in the NELSON trial.*

[b] *TNM staging system for lung cancer 7th edition.*

Table 3. Outcome of the separate review of the cause of death

Lung cancer death	Review 1	Review 2	Level of agreement
	n of 50 (%)	n of 50 (%)	kappa
Definitely or probable	41 (82)	42 (84)	0.65
- definitely	33 (66)	41 (82)	0.60
- probable	8 (16)	1 (2)	0.19
Possible	1 (2)	0 (0)	-
Unlikely	1 (2)	0 (0)	-
Definitely not	3 (6)	7 (14)	0.56
Contributory to other CoD	4 (8)	1 (2)	0.38

Definition of abbreviation: CoD = cause of death.

Table 4. The causes of death by the reviewers and the official certificates

CoD review[a]	Death certificates		
	LC death n (%)	Other CoD n (%)	Total n (%)
LC death	40 (80)	2 (4)	42 (84)
Other CoD	3 (6)	5 (10)	8 (16)
Total	43 (86)	7 (14)	50 (100)

Definition of abbreviations: CoD = cause of death; LC = lung cancer.

[a] *Cause of death after consensus meeting of the reviewers.*

cases, the reviewers indicated the letters of the pulmonologist as the most valuable source of information.

DISCUSSION

In this pilot study, we have presented the principles of the CoD review process that will be used in the NELSON trial. The pilot study of fifty cases has shown an agreement of 90% (kappa 0.65) between the two reviewers, which demonstrates a reasonable classification. We expect an increase of the level of agreement for the actual review process, performed by clinical experts, with the number of cases they evaluate; the so-called 'learning-effect'.

When comparing to the CoD process, the sensitivity and specificity of the official death certificates for lung cancer specific mortality were 95.2 and 62.5%, respectively. Despite the lack of a 'gold standard' for the cause of death of lung cancer participants, this still demonstrates, in our opinion, a limited distinctive character of the official cause of death certification in lung cancer patients for scientific purposes.

Potential limitations of the present study relate to the sample size and the selection of subjects of the pilot study. We have taken the first fifty consecutive deceased participants that were diagnosed with lung cancer. This has introduced a selection bias of individuals with a high lung cancer disease stage at diagnosis (Table 2) compared with the screen-arm of the trial.[4] In the pilot study, most deaths were due to lung cancer. It is plausible that death due to other causes than lung cancer plays a bigger part when the files of all NELSON participants will be reviewed. Hence, the figures demonstrated in the pilot could differ from those of the entire study.

No other lung cancer CT screening trial has published results of their methodology of CoD evaluation yet, to our knowledge. In the chest X-ray screening trials, such as the Mayo Lung Project, Hopkins and Sloan-Kettering Lung Trials and the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial, an expert review panel determined the CoD.[11,18–20] Lung cancer mortality was 5–6% overestimated in the intervention

arm and 2% underestimated in the usual-care arm by the death certificates in these trials.[11,18] In this initial pilot, the misestimate is 10%.

CONCLUSION

Our and other studies' results imply that the outcome of a lung cancer screening trial cannot reliably be established without a concordance analysis between vital statistics and a CoD review of blinded cases. Moreover, the principles and flowcharts presented here aim to provide one of the essential tools to make data pooling with other CT screening trials in the future possible.

CONFLICTS OF INTEREST STATEMENT

Roche diagnostics provided a grant for the performance of proteomics-research. Siemens Germany provided four digital workstations and LungCARE for the performance of 3Dmeasurements.

ACKNOWLEDGEMENTS

We thank the Central Bureau for Genealogy, Statistics Netherlands, the Flemisch Agency for Care and Health, the Dutch Cancer Registry, the Belgian Cancer Registry and the co-operating general practitioners in the Netherlands for providing the required data. Furthermore, we thank our data manager R.M. Vernhout and we thank S.J. Schop (medical student Maastricht University, Netherlands) for blanking and scanning the medical documents.

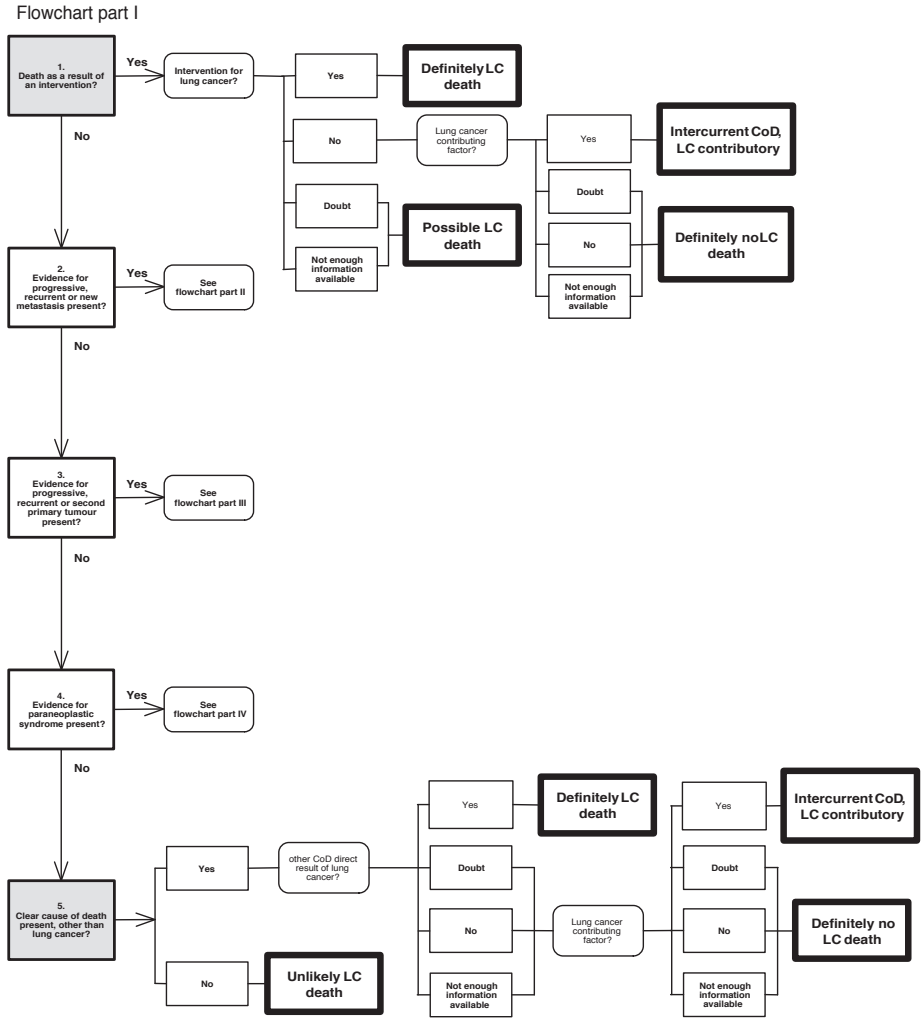
The NELSON trial is supported by: "Zorg Onderzoek Nederland-Medische Wetenschappen" (ZonMw), "KWF Kankerbestrijding", "Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen" (RvvZ), "G. Ph. Verhagen Foundation", "Rotterdam Oncologic Thoracic Study Group" (ROTS) and "Erasmus Trust Fund", "Stichting tegen Kanker", "Vlaamse Liga tegen Kanker" and "LOGO Leuven and Hageland".

REFERENCES

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011; 61(2):69–90.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012; 62(1):10–29.
3. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365(5):395–409.
4. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361(23):2221–9.
5. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst* 2002; 94(3):167–73.
6. Harteloh P, de Bruin K, Kardaun J. The reliability of cause-of-death coding in The Netherlands. *Eur J Epidemiol* 2010; 25(8):531–8.
7. Maudsley G, Williams EM. "Inaccuracy" in death certification—where are we now? *J Public Health Med* 1996; 18(1):59–66.
8. Albertsen PC, Walters S, Hanley JA. A comparison of cause of death determination in men previously diagnosed with prostate cancer who died in 1985 or 1995. *J Urol* 2000; 163(2):519–23.
9. Makinen T, Karhunen P, Aro J, et al. Assessment of causes of death in a prostate cancer screening trial. *Int J Cancer* 2008; 122(2):413–7.
10. Otto SJ, van Leeuwen PJ, Hoekstra JW, et al. Blinded and uniform causes of death verification in cancer screening: a major influence on the outcome of a prostate cancer screening trial? *Eur J Cancer* 2010; 46(17):3061–7.
11. Doria-Rose VP, Marcus PM, Miller AB, et al. Does the source of death information affect cancer screening efficacy results? A study of the use of mortality review versus death certificates in four randomized trials. *Clin Trials* 2010; 7(1):69–77.
12. Doria-Rose VP, Marcus PM. Death certificates provide an adequate source of cause of death information when evaluating lung cancer mortality: an example from the Mayo Lung Project. *Lung Cancer* 2009; 63(2):295–300.
13. Ederer F, Geisser MS, Mongin SJ, et al. Colorectal cancer deaths as determined by expert committee and from death certificate: a comparison. The Minnesota Study. *J Clin Epidemiol* 1999; 52(5):447–52.
14. Chamberlain J, Coleman D, Ellman R, et al. Verification of the cause of death in the trial of early detection of breast cancer. UK Trial of Early Detection of Breast Cancer Group. Trial Co-ordinating Centre. *Br J Cancer* 1991; 64(6):1151–6.
15. Xu DM, Gietema H, de Koning HJ, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54(2):177–84.
16. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120(4):868–74.
17. Thunnissen FB, Kerr KM, Brambilla E, et al. EU-USA pathology panel for uniform diagnosis in randomised controlled trials for HRCT screening in lung cancer. *Eur Respir J* 2006; 28(6):1186–9.
18. Marcus PM, Bergstralh EJ, Fagerstrom RM, et al. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst* 2000; 92(16):1308–16.
19. Doria-Rose VP, Marcus PM, Szabo E, et al. Randomized controlled trials of the efficacy of lung cancer screening by sputum cytology revisited: a combined mortality analysis from the Johns Hopkins Lung Project and the Memorial Sloan-Kettering Lung Study. *Cancer* 2009; 115(21):5007–17.
20. Miller AB, Yurgalevitch S, Weissfeld JL, et al. Death review process in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control Clin Trials* 2000; 21(6 Suppl.):400S–6S.

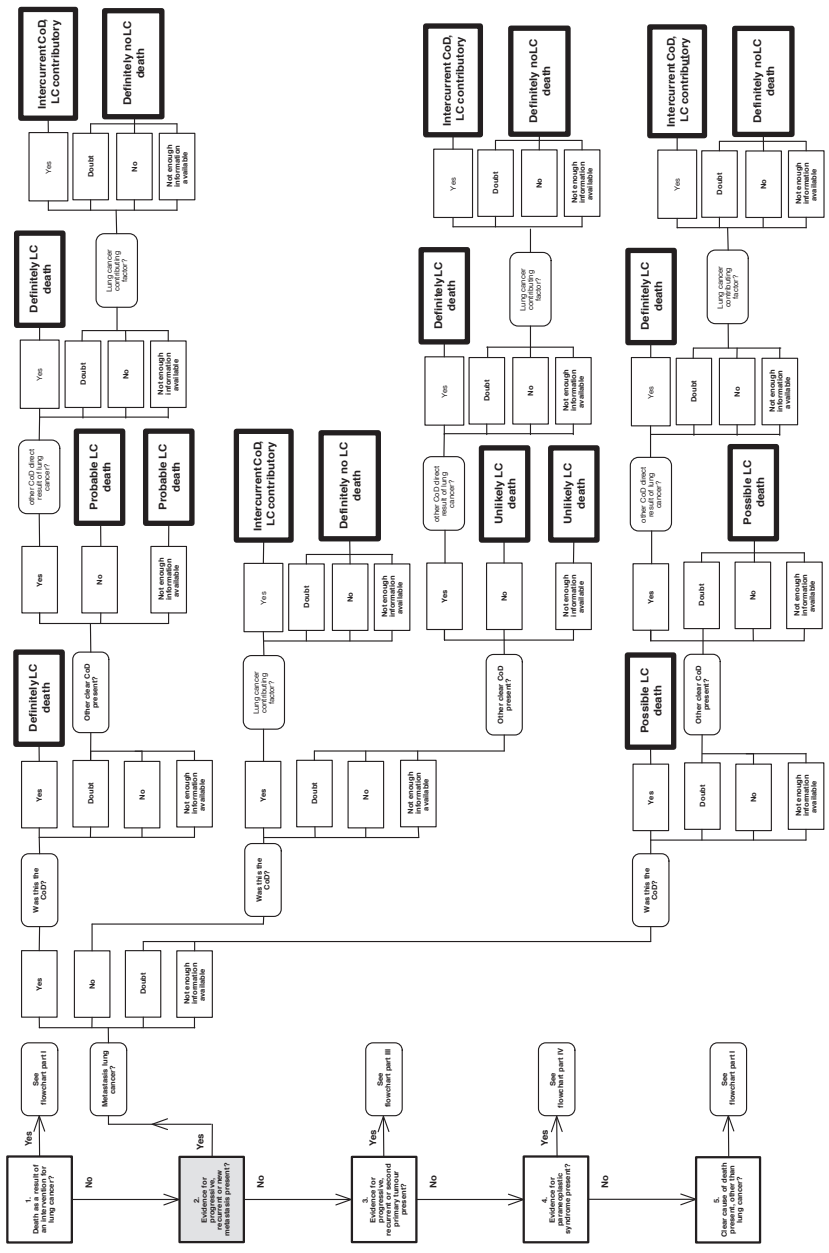
APPENDIX

Figure 1a. Cause of death evaluation flowchart part I



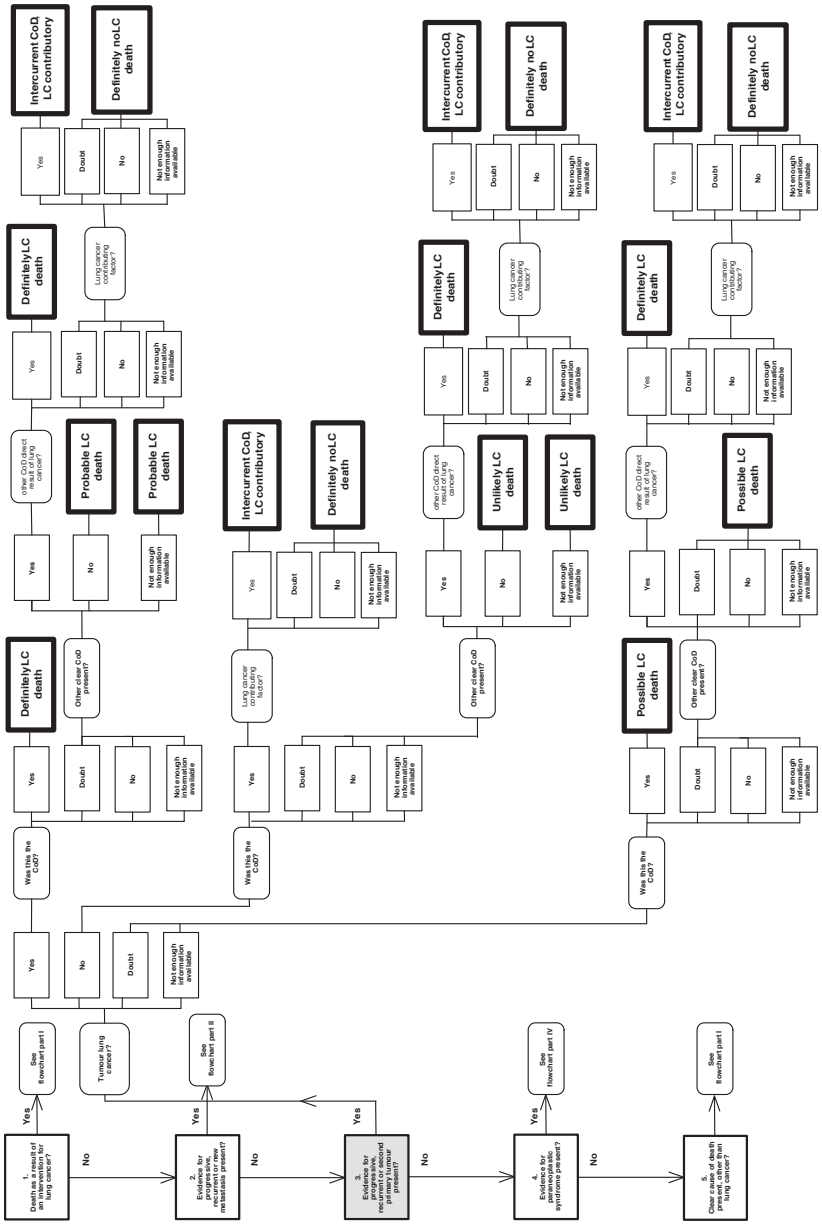
Definition of abbreviations: LC = lung cancer; CoD = cause of death.

Figure 1b. Cause of death evaluation flowchart part II



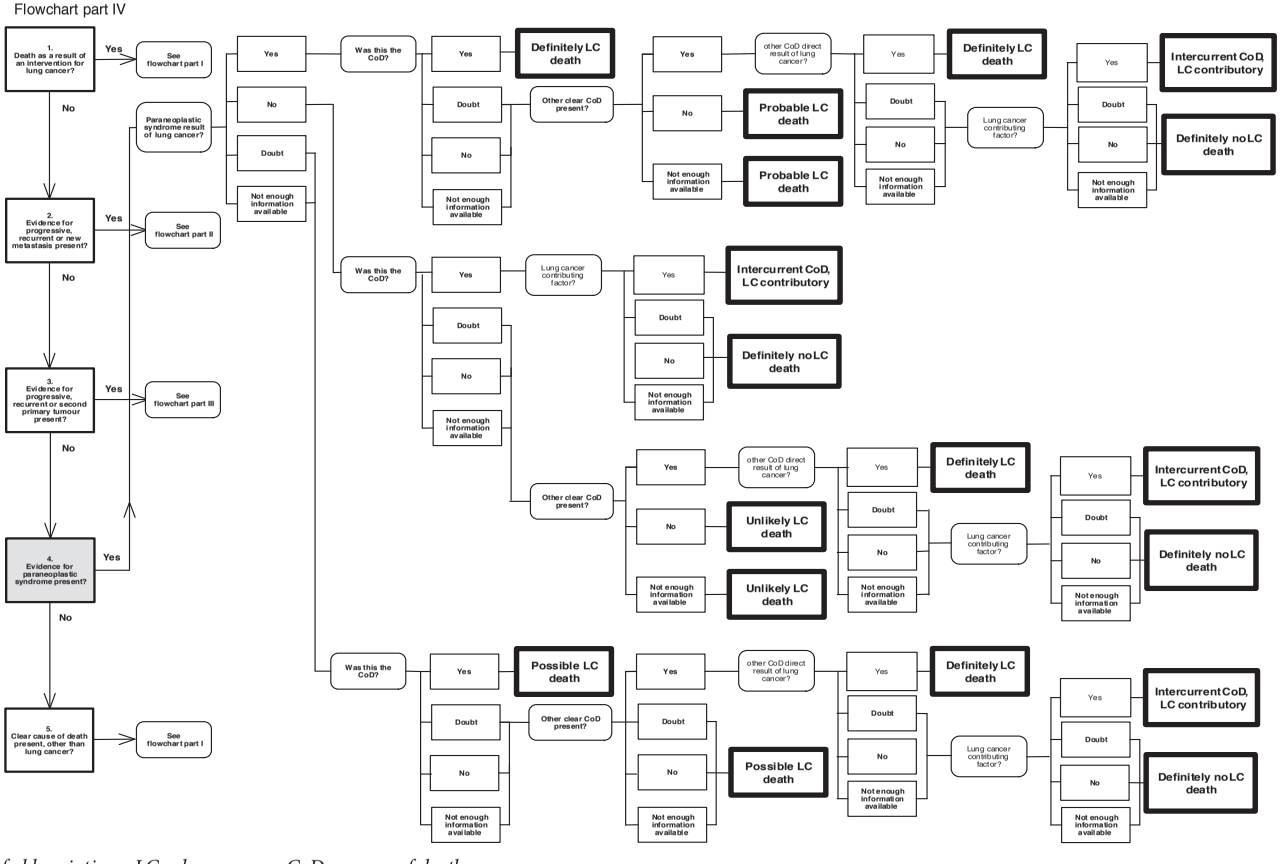
Definition of abbreviations: LC = lung cancer; CoD = cause of death.

Figure 1c. Cause of death evaluation flowchart part III



Definition of abbreviations: LC = lung cancer; CoD = cause of death.

Figure 1d. Cause of death evaluation flowchart part IV



Definition of abbreviations: LC = lung cancer; CoD = cause of death.

Table 1. The cause of death evaluation process protocol

1. Death as a result of an intervention of lung cancer? (Figure1a in appendix)

Death certainly as a direct result of a diagnostic intervention (for example: intravenous contrast or radiopharmacon (CT, Magnetic Resonance Imaging (MRI) or Fluorodeoxyglucose-Positron Emission Tomography (FDG-PET)), endoscopic interventions (transbronchial biopsy, endoscopic ultrasound (EUS) or endobronchial ultrasound (EBUS)), transthoracic punction or biopsy, pleural punction or drainage, punction of distant metastases or mediastinoscopy) or a therapeutic intervention (such as: surgery (in-hospital mortality and 30-day mortality), chemotherapy, radiotherapy, combined modality treatment, pleural drainage, endobronchial and endoesophageal interventions and brachytherapy, medication, euthanasia, palliative sedation) performed for (second primary) lung cancer or a paraneoplastic syndrome.

2. Evidence for progressive, recurrent or new metastasis present? (Figure 1b in appendix)

3. Evidence for progressive, recurrent or second primary tumour present? (Figure 1c in appendix)

To answer these two questions the following criteria are used:

I. Pathology
   Proof of relapse, progression or a second primary lung cancer:
   - Histological/cytological proof of lung cancer relapse, progression or 2nd primary

II. Radiology (X-ray, CT scan, MRI)[17]
   Proof of relapse, progression or a second primary lung cancer:
   - Growth of primary tumour (≥20% increase in largest diameter or unequivocal progression if not measurable)
   - Recurrence (short axis >10mm) or progressive pathologically enlarged lymph nodes (≥20% increase in short axis or unequivocal progression if not measurable)
   - Recurrence or progressive growth of previously existing intrapulmonary tumours (≥20% increase in largest diameter or unequivocal progression if not measurable)
   - Recurrence or progressive growth of previously existing distant metastases (≥20% increase in diameter or unequivocal progression if not measurable)
   - The appearance of any new malignant lesion (intrapulmonary or at distant sites) or pathologically enlarged lymph nodes (short axis >10mm)
   - Increase of pleural or pericardial effusions from 'trace' to 'large'
   - Increase of lymfangitis carcinomatosa from 'localised' to 'widespread'

III. Nuclear scans (FDG-PET scan, bone scintigraphy)
   Proof of relapse, progression or a second primary lung cancer:
   - New positive lesion(s) in case of a previously negative scan

IV. Bronchoscopy
   Proof of relapse, progression or a second primary lung cancer:
   - New, recurrent or progressive visible endobronchial tumour
   - New, recurrent or progressive compression of the airways by tumour mass

V. Laboratory
   Findings suggestive of relapse, progression or a second primary lung tumour, which must be confirmed by additional testing:
   - Hypercalcaemia
   - Progressive liver chemistry abnormalities
   - Increase in alkaline phosphatase
   - Increase in tumour markers (carcinoembryonic antigen, neuron-specific enolase, cytokeratin 19 fragments (CYFRA))

VI. Clinical picture

Proof of relapse, progression or a second primary lung tumour:
   - Physical exam: vena cava superior syndrome, pathologically enlarged lymph nodes ≥10mm.[17]

Findings suggestive of relapse, progression or a second primary lung tumour, which must be confirmed by additional testing:
   - Physical exam: brain metastases, abdominal masses or organomegaly, pleura or pericardial effusions, ascites, skin metastases.
   - Anamnesis: decline in WHO-performance status, weight loss >10% in past 3 months, progressive dyspnoea, bone pain

4. Evidence for paraneoplastic syndrome present? (Figure1d in appendix)

To answer this questions the following criteria are used:

I. Angiography

Proof of a paraneoplastic syndrome:

- Deep venous thrombosis (DVT): contrast venography (intraluminal filling defect in two or an abrupt cut-off of a deep vein)[18]
- Pulmonary embolism (PE): pulmonary angiography (intraluminal filling defect in two views and or an occluded pulmonary artery with or without a trailing edge)[18-19]

II. Radiology

Proof of a paraneoplastic syndrome:

- DVT: ultrasound (compression technique, duplex or colour flow imaging)[18-19]
- DVT: MRI (intravascular filling defect or occlusion of a vessel)[18]
- PE: spiral CT scanning (intravascular filling defect or occlusion of a vessel)[18-19]
- PE: MRI (intravascular filling defect or occlusion of a vessel with a 'trailing embolus' sign)[18]
- Nonbacterial thrombotic endocarditis: thoracic or transoesophageal echocardiography (evident valve vegetation)

Findings suggestive of a paraneoplastic syndrome, which must be confirmed by additional testing:

- PE: chest X-ray (no abnormalities, atelectasis, pleural effusion, pulmonary infiltrates, elevation of a hemidiaphragm, Hampton's hump, Westermark's sign)[18]
- PE: thoracic or transoesophageal echocardiography (emboli in the main, right or left pulmonary artery, right ventricular dysfunction)[18-19]
- PE using transthoracic ultrasound (peripheral wedge-shaped opacities)[19]

III. Nuclear scans

Proof of a paraneoplastic syndrome:

- PE: (ventilation-)perfusion scanning ((sub)segmental perfusion defect)[18-19]

IV. Electrocardiography (ECG)

Findings suggestive of a paraneoplastic syndrome, which must be confirmed by additional testing:

- PE: P-wave pulmonale, axis deviation, right bundle branch block, S1 Q3 T3 pattern, ST segment abnormalities, T-wave changes.[18]

V. Laboratory

Proof of a paraneoplastic syndrome:

- Hypercalcaemia of malignancy (elevated total serum calcium corrected for albumin and low intact parathyroid hormone(PTH) and elevated PTH-related protein)
- Syndrome of inappropriate Antidiurectic Hormone (ADH) secretion (low serum sodium and elevated ADH and elevated urine osmolality)
- Ectopic Adrenocorticotropic Hormone (ACTH) secretion causing Cushing's syndrome (elevated ACTH (>20pg/ml) and a negative corticotrophin-releasing hormone simulation or dexamethasone suppression test and no central step-up at inferior petral sinus sampling)
- Neurologic paraneoplastic syndrome; always in combination with corresponding clinical picture (antibodies: anti-Hu, anti-Ri, anti-Tr, anti-Crossveinless-2/anti-Collapsin response mediator protein-5, anti-Ma1, anti-Ma2, anti-amphiphysin, anti-Zic 4, anti-neuronal nuclear antibody-3, purkinje cell antibody-2, anti-Voltage-gated calcium channel, anti-Nicotinic acetylcholine receptors)
- Disseminated intravascular coagulation: microangiopathic changes on the peripheral blood smear and increased fibrinolysis (e.g. elevated fibrinogen-fibrin degradation products and D-dimer)
- Thrombotic microangiopathy: haemolytic anaemia, thrombocytopenia, increased turnover of platelets, normal level of coagulation components and little or no prolongation of prothrombine time or activated partial thromboplastin time.

Findings suggestive of a paraneoplastic syndrome, which must be confirmed by additional testing:

- PE (hypoxemia in arterial blood gas analysis and positive D-dimer)[18-19]

VI. Clinical picture

Findings suggestive of paraneoplastic syndrome, which must be confirmed by additional testing:

- DVT: limb pain, tenderness or swelling, Homans' sign.[18]
- PE: unexplained dyspnoea, pleuritic chest pain, haemoptysis, tachypnea, tachycardia, syncope, hypoxemia. [18]

- Hypercalcaemia: constipation, fatigue, polyuria, polydipsia, dehydration, anorexia, nausea, muscle weakness.
- Syndrome of inappropriate ADH secretion: fatigue, headache, oedema, nausea, vomiting, altered mental status, coma.
- Cushing's syndrome: muscle weakness, weight loss, hypertension, hirsutism.
- Severe paraneoplastic syndrome of the central nervous system (e.g. Lambert-Eaton myasthenic syndrome, cerebellar degeneration, pandysautonomia): focal neurological signs, autonomic dysfunction.
- Disseminated intravascular coagulation: signs of bleeding, acute organ failure (renal, liver, lungs), shock, thromboembolism, central nervous system dysfunction.
- Thrombotic microangiopathy: signs of anaemia, bleeding, acute renal failure, central neurologic abnormalities.
- Nonbacterial thrombotic endocarditis: acute ischemic cerebrovascular accident or acute peripheral arterial thromboembolism.

5. Clear cause of death present, other than lung cancer? (Figure 1a in appendix)

In the last step of the flowchart (Fig. 1a), no evidence for death related to interventions, (second) primary or metastatic lung cancer or a paraneoplastic syndrome is present. In case no other clear cause of death is known, the case is categorised as "unlikely lung cancer death".

Part V

Implications for implementation

Chapter 10

State of the art in lung cancer screening

The importance of screening for lung cancer

Horeweg N,
de Koning HJ

Expert review in Respiratory Medicine
August 27th 2014

THE IMPORTANCE OF SCREENING FOR LUNG CANCER

Expert Rev Respir Med. 2014 Aug 27; 1-18.

ABSTRACT

Lung cancer is a major public health problem since it causes most cancer-related deaths worldwide. As the disease often causes no symptoms at early stages, diagnosis at advanced stages, wherein cure is no longer possible, is common.

Improvements in lung cancer treatment have been made, but yielded only modest improvement in survival over the last decades. Continuous efforts should be made to force back exposure to causative agents of lung cancer, tobacco smoking in particular. However, this is not expected to reverse the lung cancer epidemic in the next decades.

The U.S. National Lung Cancer Screening Trial has demonstrated that lung cancer screening using low-dose computed tomography can reduce morbidity and mortality by detecting lung cancer at an early and curable stage. Effectiveness of a screening program is a prerequisite for implementation. In addition, the benefits of a screening program should outweigh the harms the program induces.

The currently available literature on all relevant aspects of LDCT screening for lung cancer was reviewed to determine whether the benefits of LDCT screening outweigh the harms. Next, it was determined whether LDCT screening meets the World Health Organisation criteria for screening.

Initial estimates of many harms and benefits of screening have been made, suggesting that the benefits of LDCT screening outweigh the harms. Nonetheless, the success of an implemented screening program will be determined by the benefit it yields for public health.

LUNG CANCER SCREENING: STATE OF THE ART

On November 4th 2010, the U.S. National Cancer Institute announced that lung cancer screening using low-dose computed tomography (LDCT) reduced mortality from lung cancer by 20%, compared to screening using chest radiography.[1] This impressive result was achieved in the U.S. National Lung Screening Trial (NLST), which is the world's largest randomised lung cancer screening trial.[2] As lung cancer is the leading cause of cancer-related death in the U.S. and beyond,[3,4] the impact of a significant lung cancer mortality reduction was directly recognised.

In this review, the World Health Organisation (WHO) criteria for screening (Box 1) were used as a guide for the literature search through all relevant aspects of LDCT screening for lung cancer. Hence, the ten WHO criteria for screening will be discussed, followed by a series of general conclusions on LDCT screening for lung cancer. Ultimately, the author's current and five year view on lung cancer screening will be presented.

Box 1. Modern screening criteria proposed by the World Health Organisation

I)	The screening programme should respond to a recognized need.
II)	The objectives of screening should be defined at the outset.
III)	There should be a defined target population.
IV)	There should be scientific evidence of screening programme effectiveness.
V)	The programme should integrate education, testing, clinical services and programme management.
VI)	There should be quality assurance, with mechanisms to minimize potential risks of screening.
VII)	The programme should ensure informed choice, confidentiality and respect for autonomy.
VIII)	The programme should promote equity and access to screening for the entire target population.
IX)	Programme evaluation should be planned from the outset.
X)	The overall benefits of screening should outweigh the harm.

I) The screening programme should respond to a recognised need

A lung cancer screening programme responds to a recognised need, since lung cancer is a major public health problem. Lung cancer is currently the most prevalent cause of cancer-related death.[3,4] Considering the ongoing global tobacco epidemic, lung cancer is expected to stay an important cause of death over the next decades.[5] Screening could be

a valuable addition to clinical care, primary and tertiary prevention in the fight against lung cancer.

II) The objectives of screening should be identified at the outset

The objective of screening should be reducing morbidity and mortality from lung cancer. Earlier detection of cancer should not be the objective of screening, since a prolonged disease course without a reduced burden of disease or a lower risk of lung cancer death is not beneficial (Table 1). Increasing survival should also not be the objective of screening, since estimates of survival in a screening setting are distorted by lead-time bias, length-time bias and overdiagnosis.

Table 1. Benefits and harms of cancer screening

Benefits
Less persons dying from lung cancer
Less persons suffering from advanced lung cancer
Less persons receiving intensive or mutilating primary treatment
Possible positive effects on smoking cessation
Harms
Undergoing screening test and awaiting result - psychological distress
Radiation-induced cancers - morbidity and mortality
False positive results - psychological distress, morbidity and mortality due to subsequent diagnostic procedures
False negative results - false reassurance, delayed diagnosis once symptoms occur
Overdiagnosis - psychological distress, morbidity and mortality due to overtreatment
Persons receiving the diagnosis of lung cancer earlier
Possible negative effects on smoking cessation

III) There should be a defined target population

The significant mortality reduction in the NLST made many medical societies recommend the NLST inclusion criteria as the definition of the optimal target population for screening (Table 2).[6-9] The National Cancer Comprehensive Network (NCCN) and the American Association of Thoracic Surgery (AATS) recommended an extended or modified versions of the NLST inclusion criteria.[10,11] Remarkably, two New England Journal of Medicine publications by the NLST pointed out that the trial's inclusion criteria were suboptimal.[12,13]

The first article demonstrated that the use of a lung cancer prediction model, (including age, race, education, body-mass index, chronic obstructive pulmonary disease, personal and family history of cancer, smoking status, intensity duration and quit time) yielded a higher sensitivity and an equal specificity for lung cancer compared to the NLST inclu-

sion criteria.[12] The second article demonstrated that screening was most effective and least harmful in the NLST participants at the highest risk of lung cancer mortality.[13] In the participants at a relatively low 5-year risk of lung cancer death (lowest risk quintile 0.15-0.55%), the benefit of screening (1% of prevented lung cancer deaths) was small compared to the numbers needed to screen (5276:1) and the false-positive screenings (97.0%).[13]

Table 2. Inclusion criteria of randomised controlled trials on LDCT screening for lung cancer

Trial	Inclusion criteria			
	Sex	Age*	Cigarette smoking	Cessation
NLST[2,21]	male or female	55-74	≥30 pack-years	<15 yrs
NELSON[29,33]	male or female	50-75	≥15 per day for 25 years or ≥10 per day for 30 years	≤10 yrs
DLST[34]	male or female	50-70	≥20 pack-years	<10 yrs
MILD[40]	male or female	≥49	≥20 pack-years	<10 yrs
LUSI[38]	male or female	50-70	≥15 per day for 25 years or ≥10 per day for 30 years	≤10 yrs
UKLS[44,55]	male or female	50-75	≥5% risk of lung cancer in 5 years	
ITALUNG[56]	male or female	55-70	≥20 pack-years	<10 yrs
DANTE[57]	male	60-75	≥20 pack-years	<10 yrs

** Age range up to, but not including upper limit.*

The CISNET lung cancer working group assessed benefits and harms of LDCT screening for lung cancer for the U.S. Preventive Services Task Force (USPSTF).[14] Five modelling groups independently evaluated 576 screening scenarios to determine the optimal target population.[14] Eligibility criteria were: age at begin and end of screening, minimum number of pack-years, and maximum number of years since smoking cessation. Their analyses identified a range of possible 'optimal' target populations, including the set of eligibility criteria for screening which were adopted by the USPSTF (Table 3).[14]

The latter study provided solid evidence for optimal target populations for lung cancer screening. Future studies will provide more insight in the value of prediction models

Table 3. Eligibility criteria for lung cancer screening adopted by the USPSTF

Eligibility criteria	Estimated optimum
Age of start screening	55
Age of end screening	80
Pack-years of smoking	≥ 30
Years since cessation	<15

for selecting eligible subjects for screening, and optimal inclusion criteria for Asian and other specific populations. Since all prominent medical associations have published their guidelines on lung cancer screening in the past two years, none of them incorporates all currently available evidence on this subject. Possibly, the next generation of guidelines will recommend definitive criteria for the optimal target population for lung cancer screening.

IV) There should be scientific evidence of screening programme effectiveness

The NLST has demonstrated 20% lung cancer mortality reduction by screening using LDCT compared to screening using chest radiograph (Table 4).[1] When the CISNET lung cancer working group evaluated hundreds of screening scenarios, mortality reductions were also estimated.[14] Estimates for twenty-six selected efficient screening scenarios ranged from 4.6-21.2%, which demonstrates the strong correlation between the benefit of a screening programmes and its design and target population. The USPSTF eligibility criteria (Table 3) with annual screening was estimated to yield a 14.0% lung cancer mortality reduction, which corresponds with 497 lung cancer deaths averted and 5250 life-years gained per the 100,000-member cohort.[14] Lung cancer mortality reductions in several European trials are still awaited (Table 4). Since the outcomes of European trials combined have enough statistical power to affect the significant mortality reduction of the NLST, no definitive conclusion can be drawn on the magnitude of the mortality reduction to date.

Table 4. Effect LDCT screening on lung cancer and all-cause mortality

Trial*	Lung cancer deaths			All deaths		
	per 100,000 py[20]		Relative risk[20]	per 100,000 py[20]		Relative risk[20]
	Intervention	Control	RR (95%CI)	Intervention	Control	RR (95%CI)
NLST[1]	247	309	0.80 (0.73-0.93)	1142	1216	0.93 (0.86-0.99)
DLST[41]	154	112	1.37 (0.63-2.97)	625	429	1.46 (0.99-2.15)
MILD[40]	216	109	1.99 (0.80-4.96)	558	310	1.80 (1.03-3.13)
DANTE[58]	527	637	0.83 (0.45-1.54)	1212	1433	0.85 (0.56-1.27)

Definition of abbreviations: py = person-years; RR = relative risk; 95%CI = 95% confidence interval.
** Trials included with results published before July 2014.*

V) The programme should integrate education, testing, clinical services and programme management

Since LDCT screening has been recommended by several U.S. medical associations,[6-8] these implementation-related aspects of screening have become relevant. So far, little has been published about lung cancer screening programme management, clinical services,

education, and testing. Nonetheless, recommendations on these aspects have been made by several medical associations.

I) Multi-disciplinary approach: is recommended by all guidelines.[6-11,15,16] Hence, radiologists, pulmonologists and thoracic surgeons should have regular meetings wherein screening cases are discussed. However, close cooperation with other specialties is also warranted: nuclear medicine experts and pathologists for the assessment of respectively nuclear scans and small biopsies of CT-detected nodules, and medical and radiation oncologists for the treatment of screen-detected lung cancer.

II) Process management: all subsequent steps from the candidate's first attendance to the screening clinic to the treatment of screen-detected lung cancer should be orchestrated before screening is implemented. Two steps of this process have been described in lung cancer screening guidelines; a defined algorithm for scan interpretation,[6-8,15-17] and a diagnostic algorithm for suspicious CT-detected nodules.[6-8,10,11,15,16] Clinical guidelines on the management of incidentally-detected pulmonary nodules may be complementary, since many recommendations are based on data from lung cancer screening studies. Furthermore, logistics of the screening programme should be coordinated to prevent drop-outs and limit waiting times.

III) Facilities: the availability and quality of radiological, surgical, and other facilities is essential for a screening programme's effectiveness and safety. A number of screening guidelines emphasise that screening should be performed in 'centres similar to those wherein the NLST was conducted'.[6-8] This suggest that screening should be performed in large centres which comply with the NLST minimum equipment standards,[2] and have specialised thoracic radiologists and board-certified thoracic surgeons on staff.

Detailed descriptions of radiological requirements are provided in the AATS guideline and are expected from the International Association for the Study of Lung Cancer (IASLC) Radiology Working Group,[15,18] and the Radiological Society of North America (RSNA) / American College of Radiology (ACR) collaboration.[17] The use of volumetry to assess nodule size and growth is currently only recommended by IASLC15 and AATS guidelines.[11]

Specific recommendations for surgical management of suspicious nodules are only provided by the IASLC and the AATS. Both recommend lobectomy with systematic lymph node sampling as preferred procedure for suspected or confirmed early stage lung cancer.[15,16] Segmentectomies with sampling of N1 and N2 lymph node stations are only recommended by the IASLC for sub-solid nodules smaller than 2cm, and in individuals with limited pulmonary reserve or multiple lesions.[15] Wedge resections should only be used for diagnostic purposes according to the IASLC,[15] however according to the AATS, wedge resections could also be used as therapeutic procedure for lung cancers appearing

as sub-solid nodules on CT.[16] The preferred approach to perform these procedures is video-assisted thoracoscopic surgery (VATS), because of the lower post-operative mortality compared to thoracotomy.[8,11,15,16]

The integration of a smoking cessation programme along with screening is recommended by all guidelines.[6-8,10,11,15,16] Besides the ethical necessity of this recommendation, smoking cessation will also increase the benefits of a screening programme.

VI) There should be quality assurance, with mechanisms to minimise potential risks of screening

Screening bears the risk of several harms (Table 1), which can be minimised by adequate quality assurance. This aspect of lung cancer screening is still under development. Many useful lessons could be learned from successfully implemented screening programmes, such as breast cancer screening. However, the implementation of lung cancer screening will give rise to new challenges. Besides screening trials, screening demonstration projects[7,8,15] are also useful for the development of quality metrics and minimum standards for lung cancer screening.[7,8,15,16] CT quality controls are currently already recommended.[6-8,10,11,15,16] Moreover, a LDCT-screening quality standards act and independent quality assurance units that collect and collate data about the performance and outcomes of screening programmes and organise quality assurance visits would enforce quality assurance.[19]

VII) The programme should ensure informed choice, confidentiality and respect for autonomy

Persons, who consider undergoing LDCT screening, should be informed on harms and benefits of screening (Table 1). Currently, many harms have been identified and some risk estimates of these harms have been published.[7,14,20] Education of screening candidates is recommended in almost all guidelines.[6-8,10,16] However, minimum requirements for the harms and benefits that should be discussed, and the level of knowledge of counsellors have not been published. Naturally, the individual's choice on participation after education on the harms and benefits should be respected. By no means, fear of cancer should be used to convince subjects to undergo screening. Once eligible screening candidates are informed and voluntarily agree to undergo LDCT screening, written informed consent should be obtained. In case data from their screening will be used for research purposes, explicit permission should be requested.

VIII) The programme should promote equity and access to screening for the entire target population

The benefit of lung cancer screening for public health depends on the participation rate in the target population. The applied recruitment method influences the population that

will attend to screening. Recruitment through the media attracts younger individuals who are more often ex-smokers with better education and higher social economic status, compared to the entire target population.[21,22] The latter is not desirable[13] since subjects at high risk for developing lung cancer are most predominantly represented in lower social-economic groups.[23]

Henceforth, recruitment should effectively reach out to the lower educated proportion of the population. The population participating in the Dutch-Belgian lung cancer screening trial was slightly lower educated, smoked more heavily, had a worse general health and a higher prevalence of malignancies, but had the same age, percentage current smokers, and BMI as the general population.[24] Recruitment for this trial was performed by determining eligibility for screening first (via an initial mailing which did not contain any information on the trial), followed by a second mailing only to eligible subjects, with information and the invitation to participate in lung cancer screening. Although lower educated groups can be recruited with this method, it will be difficult to obtain information on eligibility from the target population without informing them on lung cancer screening outside a clinical trial nowadays. Unfortunately, no other screening trial, that used a population-based recruitment strategy wherein anyone in the target age range received a mailing with an invitation for screening and a questionnaire to determine eligibility, also investigated participation bias especially by educational level.

Besides difficulties with recruitment of lower educated groups, reaching out to minorities and ethnic groups other than Caucasians is also challenging. The NLST institutions were encouraged to identify regional minorities and develop plan for targeted recruitment.[25] Seven institutions were selected and received funding to implemented their proposals, which were diverse; advertising in minority-specific media, distribution of culturally adapted and translated brochures, outreach programs via general practitioners, face-to-face interaction and word-of-mouth dissemination.[25] The success of the aforementioned strategies varied, and no single strategy was successful among all institutions.[25] Nonetheless, recruitment of local minorities increased from 9.3% to 15.2% in the seven institutes using any strategy.[25] Knowledge on local cultural and ethnic diversity and cooperation with local stakeholders and minority organisations are probably vital for developing suitable recruitment strategies and successful implementation.

Concluding, recruiting the higher educated, health-concerned part of the population for lung cancer screening will probably be successful. However, making the lower educated and minority groups aware of the availability and advantages of lung cancer screening will be more difficult. It will require social responsibility of health policy makers to put effort and resources in the recruitment of these subgroups. Since the subjects at high risk for developing lung cancer are predominantly represented in these subgroups, the success of lung cancer screening for public health will partly depend on it.

IX) Programme evaluation should be planned from the outset

As for all health care programmes, evaluation and feedback with the objective to detect deficiencies and improve care should be planned from the outset. Current guidelines on lung cancer screening do not explicitly mention programme evaluation, but registry/data collection is recommended.[7,8,11,15,16] Collected data could be used to monitor and improve performance, as well as collate with quality standards, as described at criterion VI.

X) The overall benefits should outweigh the harm

One of the most important criteria for screening. Determining whether the benefits of a screening programme outweigh the harms is complex. The benefits of screening can be expressed as the number of quality-adjusted life-years (QALYs) gained, which should be estimated using the number of deaths prevented and the number of advanced stage disease prevented by screening. The number of life-years gained can be estimated with data from randomised controlled trials and micro-simulation modelling. Next, the life-years gained should be corrected for quality of life using utility estimates.[26] Further, the same should be done for the harms of screening; estimating the number of QALYs lost by screening using data from trials, implemented programmes, registries and modelling studies. Once this has been established, it can be determined whether the benefits of lung cancer screening outweigh the harms. An example of such a calculation has recently been published for prostate cancer screening.[27]

Currently, only the gain in quality-adjusted life-years for NLST participants in whom death was prevented has been estimated: 21.7 QALYs per 1000 screened individuals.[28] The gain QALYs for those in whom advanced stage disease was prevented by LDCT screening should be added to obtain an estimate for the total benefit of screening. Moreover, the estimate would become more reliable if data from other screening trials, such as the largest European trial,[29] was added. The CISNET lung cancer working group estimated that per 100,000-person cohort screened with the recommended regimen (Table 4) 497 subjects would not die from lung cancer and 550 no longer needed treatment for advanced lung cancer.

Estimates for numbers of QALYs lost due to screening have not been published yet. However, some estimates of the magnitude of many harms have been published to date, briefly:

I) Undergoing screening test: could be accompanied by psychological distress, however neither clinically relevant, nor statistically significant negative effects on physical health, mental health, self-reported health, generic anxiety, lung cancer-specific distress and on the impact of event scale were demonstrated.[30] Further, undergoing the screening test leads to exposure to ionising radiation.[31] The harmful effect of radiation related to both the screening examination and the subsequent diagnostic procedures for positive screenings have been estimated for

the participants of the NLST; it was estimated that cancer for every 2,500 screened subjects, one subject would die from radiation-induced cancer.[7,8] The CISNET lung cancer working group estimated the number of radiation-related lung cancer deaths for a range of screening scenarios; per a 100,000-person cohort followed from ages 45 to 90, 24 deaths were expected both with the screening strategy most similar to the NLST eligibility criteria (Table 2), and the screening strategy adopted by the USPSTF (Table 3).

II) Awaiting screening test result: led to discomfort in 46.0-51.3% of the screened subjects of the NELSON trial.[32] Awaiting the follow-up scan after an initial indeterminate screening result led to a clinically relevant increase in lung cancer-specific distress; which recovered in case of a negative follow-up scan and remained in case of a positive follow-up scan.[30]

III) False-positive screenings: a positive screening in subjects without lung cancer. The proportion of all screening results that is false-positive depends on the definition of a positive screening (i.e. threshold for nodule size or growth), and to a lesser extent to the technique that is used to review the screening examination (i.e. manual measurements semi-automated volumetry). In the NLST, a relatively low threshold of ≥4 mm for manually measured nodule diameter was applied, and a high percentage of false-positive screenings (23.3%) was observed.[1] In comparison, in the NELSON trial a higher threshold of 500 mm^3 for semi-automatically assessed nodule volume (about 9.8 mm in diameter) was applied, and a relatively low percentage of false-positive screenings was observed (1.2%).[33] False-positive screenings lead to unnecessary diagnostic procedures and psychological distress in subjects without lung cancer. The magnitudes of the psychological distress and their long-term effects have not been investigated in any of the randomised lung cancer screening trials. Invasive diagnostic procedures for false-positive screenings were performed in 0.4-1.3% of the participants of lung cancer screening trials.[33-38] The CISNET lung cancer working group modelled that 67,550 persons would have a false-positive result and 910 would undergo invasive diagnostic procedures for benign nodules per 100,000-person cohort screened annually from age 55 to 80.[14] Data on the incidence of morbidity and mortality as a result of complications of the invasive procedures for false-positive screenings were not published.

IV) False-negatives: a negative screening result in a subject with (early) lung cancer could lead to delayed diagnosis through false reassurance once symptoms emerge. In the randomised lung cancer screening trials, 0.0-6.3% of the lung cancers was not detected through screening.[29,35,38-41] Estimates of the delay to diagnosis caused by the false negative screening result were not provided.

V) Overdiagnosis: estimates of the amount of overdiagnosis for a range of hypothetical screening programmes were made by the CISNET lung cancer working group in a

comparative modelling study.[14] The screening strategy as applied in the NLST was estimated to have led to 8.7% overdiagnosis.[14] The slightly adjusted screening algorithm that was recommended from this study after weighing several harms and benefits of screening (Table 3) would yield 9.9% overdiagnosis.[14] This corresponds with 190 persons with an overdiagnosed lung cancer per 100,000-person cohort.[14] The associated harms consist of the physical harms induced by diagnostic procedures and overtreatment, and mental harms caused by distress due to the aforementioned interventions and anxiety of having a cancer diagnosis; none of these have been quantified to date.

VI) Prolonged disease course: effective screening programmes advance diagnosis of lung cancer. For some screenees, this will lead to cure from lung cancer, however in most screenees screening will not prevent lung cancer death.[1] In these subjects the course of disease was longer because of the earlier diagnosis, but there was no health benefit; which is a harmful side-effect of screening. It was estimated by that 1,970 persons per 100,000-person cohort would receive the diagnosis of lung cancer earlier when annual screening from the age of 55 to 80 is implemented.[14] Estimates of the harms and duration of the lead time in these persons have not been published.

VII) Negative impact on life-style: another potential harm of lung cancer screening is that the screened population considers the LDCT examination as a substitute for smoking cessation; the so-called health certificate effect. In two screening trials, this was investigated and no differences were found in the cessation rate and the number of quit attempts between subjects who received screening and those who received no screening.[42,43] However, in one study smoking abstinence was significantly higher for participants receiving no screening compared to participants who received LDCT examinations.[42] The difference in smoking abstinence between the screened group and the control group was 4.6% (OR 1.40, 95% CI 1.01-1.92).[42] Henceforth, it cannot be precluded that LDCT screening does not have a negative impact on life-style. Effective smoking cessation programmes implemented along with LDCT screening may compensate for any harmful effect of screening on life-style.

CONCLUSIONS ON LDCT SCREENING FOR LUNG CANCER

I) The screening programme should respond to a recognised need

Lung cancer screening responds to a recognised need as lung cancer is a major public health problem.

II) The objectives of screening should be defined at the outset

The objectives of lung cancer screening are reducing morbidity and mortality from lung cancer.

III) There should be a defined target population

Although evidence for an optimal target population for lung cancer screening has been published, discussions are ongoing.

IV) There should be scientific evidence of screening programme effectiveness

Effectiveness of lung cancer screening has been demonstrated in one study, future results of a pooled analysis of all lung cancer screening trials will provide definite conclusions.

V) The programme should integrate education, testing, clinical services and programme management

Current lung cancer screening guidelines provide some useful recommendations on the integration of education, testing, clinical services and management in lung cancer screening programmes, however substantial gaps remain.

VI) There should be quality assurance, with mechanisms to minimise potential risks of screening

Since LDCT screening is already (recommended to be) implemented and adequate quality assurance is still under development, the population undergoing screening is exposed to several risks.

VII) The programme should ensure informed choice, confidentiality and respect for autonomy

Although any lung cancer screening guideline recommends informing screening candidates on the harms and benefits of screening and respecting their autonomy, no minimum requirements on screening information and knowledge of the counsellor are defined.

VIII) The programme should promote equity and access to screening for the entire target population

Successful implementation of LDCT screening depends on the benefit the programme yields for public health, therefore effective methods need to be developed to reach out to the lower educated and minority groups.

IX) Programme evaluation should be planned from the outset
Data collection from lung cancer screening programmes is recommended by several guidelines, and should at least be used for evaluation and quality assurance, which is not recommended yet.

X) The overall benefits of screening should outweigh the harm
Initial estimates of many harms and benefits of screening have been made, suggesting that the benefits of LDCT screening outweigh the harms.

AUTHOR'S VIEW ON LUNG CANCER SCREENING

Screening a population at high risk of developing lung cancer using low-dose computed tomography examinations is a promising technique to reduce morbidity and mortality from lung cancer. A significant reduction in lung cancer mortality by screening using LDCT examination has been demonstrated in one large, high-quality randomised trial.[1] Future pooled analysis of multiple randomised lung cancer screening trials will provide definitive evidence for the effectiveness of LDCT screening.[29]

Initial estimates of many harms and benefits of screening have been made using data from the U.S. lung cancer screening trials.[14] It is unknown to what extent these estimates apply to other screening strategies, such as the volumetry-based nodule protocols of the European trials.[29,34,38,44] Available evidence suggests that LDCT screening can be beneficial if applied as in the NLST.[7,20] However, it is uncertain whether LDCT screening for lung cancer screening is beneficial when implemented in different settings without established safety and quality assurances.

The success of any screening programme is determined by the benefits it yields for public health. Ma et al. estimated that if the screening regimen adopted in the NLST was fully implemented among the 8.6 million screening-eligible U.S. population, 12,250 lung cancer deaths could be averted per year.[45] If the optimal screening scenario according to the USPSTF (Table 4) is implemented in the U.S., 10.5 million individuals will be eligible for screening.[14] The yield, assuming 100% adherence to screening, is estimated at 18,000 deaths avoided per year, which corresponds with a 25% lung cancer mortality reduction in the eligible population and 14% overall lung cancer mortality reduction.[14] Clearly, the projected benefits of LDCT screening for public health are significant. Implementation of LDCT screening programmes that reach out to the entire target population will be essential in realising the potential benefit for public health.

Continued efforts and advances in lung cancer treatment, primary prevention, and tertiary prevention are also expected to reduce the burden of lung cancer in the future. However, each of these methods solely will not be able to reverse the lung cancer epi-

demic. Therefore, screening using LDCT should be regarded as a valuable new tool in the fight against lung cancer.

AUTHOR'S FIVE-YEAR VIEW ON LUNG CANCER SCREENING

Implementation of LDCT screening for lung cancer has only recently become at issue. Hence, the next five years will be important for its success. Many remaining uncertainties and discussions are likely to be addressed to within five years.

Hence, final results of all randomised lung cancer screening trials have become available.[29,38,44] Meta-analyses of lung cancer and all-cause mortality reduction will provide definitive conclusions on the efficacy of LDCT screening. Further, data from lung cancer screening trials and implemented screening programmes will give insight in the magnitude and impact of the harms associated with LDCT screening. With this data, more comprehensive calculations of the balance between harms and benefits of screening can be made. This will eventually also facilitate the performance of more integral cost-effectiveness analyses.

In the next five years, some lingering questions surrounding LDCT screening will probably be answered. Among these are probably: the value of lung cancer prediction models for the selection of eligible subjects;[44] the optimal screening method, including insight in the added value of volumetry and the use of imaging biomarkers for the interpretation and accuracy of screening examinations;[29,46] and the optimal diagnostic work-up from positive screening result to diagnosis for any type of screen-detected pulmonary nodule.[47,48]

Finally, a number of issues will become more relevant in the next five years than they have been so far, as a result of the recommended implementation of LDCT screening. Examples are: recruitment of elderly, lower educated and minority groups for screening,[25] shared decision-making,[47,49,50] management systems for screening programmes,[51] quality assessment and performance indicators,[8,19] non-surgical treatments for screen-detected lung cancer,[52] follow-up regimens for curatively treated screen-detected lung cancers,[53] and patient-centred research.[50,54]

REFERENCES

1. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365(5):395-409.
2. National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. *Radiology* 2011; 258(1):243-53.
3. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin* 2013; 63(1):11-30.
4. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011; 61(2):69-90.
5. Mendez D, Alshanqeety O, Warner KE. The potential impact of smoking control policies on future global smoking trends. *Tob Control* 2013; 22(1):46-51.
6. Wender R, Fontham ET, Barrera E, Jr., et al. American Cancer Society lung cancer screening guidelines. *CA Cancer J Clin* 2013; 63(2):107-17.
7. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and Harms of CT Screening for Lung Cancer: A Systematic Review Benefits and Harms of CT Screening for Lung Cancer. *JAMA* 2012: 307(22):2418-29.
8. Detterbeck FC, Mazzone PJ, Naidich DP, et al. Screening for Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013; 143(5 Suppl):e78S-92S.
9. American Lung Association. Guidance on CT Lung Cancer Screening 2012. http://www.lung.org/about-us/our-impact/top-stories/guidance-on-ct-lung-cancer.html Date last accessed 17-01-2014.
10. Wood DE, Eapen GA, Ettinger DS, Hou L, Jackman D, Kazerooni E, et al. Lung cancer screening. J Natl Compr Canc Netw. 2012; 10(2): 240-65.
11. Jacobson FL, Austin JH, Field JK, et al. Development of The American Association for Thoracic Surgery guidelines for low-dose computed tomography scans to screen for lung cancer in North America: recommendations of The American Association for Thoracic Surgery Task Force for Lung Cancer Screening and Surveillance. *J Thorac Cardiovasc Surg* 2012; 144(1):25-32.
12. Tammemagi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med* 2013; 368(8):728-36.
13. Kovalchik SA, Tammemagi M, Berg CD, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med* 2013; 369(3):245-54.
14. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and Harms of Computed Tomography Lung Cancer Screening Strategies: A Comparative Modeling Study for the U.S. Preventive Services Task Force. *Ann Intern Med* 2014; 160(5):311-20.
15. Field JK, Smith RA, Aberle DR, et al. International Association for the Study of Lung Cancer Computed Tomography Screening Workshop 2011 report. *J Thorac Oncol* 2012; 7(1):10-9.
16. Rocco G, Allen MS, Altorki NK, et al. Clinical statement on the role of the surgeon and surgical issues relating to computed tomography screening programs for lung cancer. *Ann Thorac Surg* 2013; 96(1):357-60.
17. Radiologic Society of North America. RSNA/ACR Joint Statement on Lung Cancer Screening. http://www2.rsna.org/timssnet/media/pressreleases/PDF/pressreleasePDF.cfm?ID=693 2013. Date last accessed 28-01-2014.
18. Field JK, Aberle DR, Altorki N, et al. The International Association Study Lung Cancer (IASLC) Strategic Screening Advisory Committee (SSAC) Response to the USPSTF Recommendations. *J Thorac Oncol* 2014; 9(2):141-3.
19. Public Health England. Quality assurance within the NHS Breast Screening Programme. http://www.cancerscreening.nhs.uk/breastscreen/quality-assurance.html 2013 Date last accessed 31-01-2014.

20. Humphrey LL, Deffebach M, Pappas M, et al. Screening for Lung Cancer With Low-Dose Computed Tomography: A Systematic Review to Update the U.S. Preventive Services Task Force Recommendation. *Ann Intern Med* 2013. 159(6):411-20.
21. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Baseline characteristics of participants in the randomized national lung screening trial. *J Natl Cancer Inst* 2010; 102(23):1771-9.
22. Hestbech MS, Siersma V, Dirksen A, et al. Participation bias in a randomised trial of screening for lung cancer. *Lung Cancer* 2011; 73(3):325-31.
23. Mackenbach JP, Huisman M, Andersen O, et al. Inequalities in lung cancer mortality by the educational level in 10 European populations. *Eur J Cancer* 2004; 40(1):126-35.
24. van der Aalst CM, van Iersel CA, van Klaveren RJ, et al. Generalisability of the results of the Dutch-Belgian randomised controlled lung cancer CT screening trial (NELSON): does self-selection play a role? *Lung Cancer* 2012; 77(1):51-7.
25. Duda C, Mahon I, Chen MH, et al. Impact and costs of targeted recruitment of minorities to the National Lung Screening Trial. *Clin Trials* 2011; 8(2):214-23.
26. Institute for clinical research and health policy studies. Cost-effectiveness analysis registry. http://www.cearegistry.org 2013. Date last accessed 31-01-2014.
27. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med* 2012; 367(7):595-605.
28. Berg C, NLST CEA writing team. Cost-effectiveness of CT screening. Abstract presented at the world lung cancer conference 2013 http://www.2013worldlungcancer.org/documents/WCLC2013-Abstract-Book.pdf Date last accessed 26-01-2014.
29. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361(23):2221-9.
30. van den Bergh KA, Essink-Bot ML, Borsboom GJ, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *Br J Cancer* 2010; 102(1):27-34.
31. Linet MS, Slovis TL, Miller DL, et al. Cancer risks associated with external radiation from diagnostic imaging procedures. *CA Cancer J Clin* 2012; 62(2):75-100.
32. van den Bergh KA, Essink-Bot ML, Bunge EM, et al. Impact of computed tomography screening for lung cancer on participants in a randomized controlled trial (NELSON trial). *Cancer* 2008; 113(2):396-404.
33. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J* 2013; 42(6):1659-67.
34. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish randomized lung cancer CT screening trial--overall design and results of the prevalence round. *J Thorac Oncol* 2009; 4(5):608-14.
35. National Lung Screening Trial Research Team, Church TR, Black WC, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 2013; 368(21):1980-91.
36. Horeweg N, Nackaerts K, Oudkerk M, et al. Low-dose computed tomography screening for lung cancer: results of the first screening round. *J Comp Eff Res* 2013; 2(5):433-6.
37. Infante M, Chiesa G, Solomon D, et al. Surgical procedures in the DANTE trial, a randomized study of lung cancer early detection with spiral computed tomography: comparative analysis in the screening and control arm. *J Thorac Oncol* 2011; 6(2):327-35.
38. Becker N, Motsch E, Gross ML, et al. Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. *J Cancer Res Clin Oncol* 2012; 138(9):1475-86.
39. Aberle DR, DeMello S, Berg CD, et al. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med* 2013; 369(10):920-31.

40. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev* 2012; 21(3):308-15.
41. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; 67(4):296-301.
42. van der Aalst CM, van den Bergh KA, Willemsen MC, et al. Lung cancer screening and smoking abstinence: 2 year follow-up data from the Dutch-Belgian randomised controlled lung cancer screening trial. *Thorax* 2010; 65(7):600-5.
43. Ashraf H, Tonnesen P, Holst Pedersen J, et al. Effect of CT screening on smoking habits at 1-year follow-up in the Danish Lung Cancer Screening Trial (DLCST). *Thorax* 2009; 64(5):388-92.
44. Baldwin DR, Duffy SW, Wald NJ, et al. UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. *Thorax* 2011; 66(4):308-13.
45. Ma J, Ward EM, Smith R, et al. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer* 2013; 119(7):1381-5.
46. Xie X, Zhao Y, Snijder RA, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. *Eur Radiol* 2013; 23(1):139-47.
47. Gould MK, Donington J, Lynch WR, et al. Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer?: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013; 143(5 Suppl):e93S-e120S.
48. Naidich DP, Bankier AA, MacMahon H, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013; 266(1):304-17.
49. Slatore CG, Press N, Au DH, et al. What the heck is a "nodule"? A qualitative study of veterans with pulmonary nodules. *Ann Am Thorac Soc* 2013; 10(4):330-5.
50. Wiener RS, Slatore CG. Framing discussions about CT scan screening for lung cancer so that patients see the whole picture. *Chest* 2013; 144(6):1749-50.
51. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54(2):177-84.
52. Timmerman R, Paulus R, Galvin J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA* 2010; 303(11):1070-6.
53. Colt HG, Murgu SD, Korst RJ, et al. Follow-up and surveillance of the patient with lung cancer after curative-intent therapy: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(5 Suppl):e437S-54S.
54. Wiener RS, Gould MK, Woloshin S, et al. What do you mean, a spot?: A qualitative analysis of patients' reactions to discussions with their physicians about pulmonary nodules. *Chest* 2013; 143(3):672-7.
55. Field JK, Devaraj A, Baldwin DR, et al. UK lung cancer screening trial (UKLS): baseline data. Poster presented at the world lung cancer conference 2013 http://www.2013worldlungcancer.org/documents/WCLC2013-AbstractBook.pdf Date last accessed 17-01-2014.
56. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer* 2009; 64(1):34-40.
57. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer* 2008; 59(3):355-63.

58. Infante M, Cavuto S, Lutman FR, et al. A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial. *Am J Respir Crit Care Med* 2009; 180(5):445-53.

Chapter 11

General discussion

In this thesis, the harms and benefits of lung cancer screening using low-dose computed tomography were investigated. Data of the Dutch-Belgian NELSON trial were used to quantify its harms and benefits and develop strategies to improve the balance between them. If the NELSON trial demonstrates that low-dose CT screening is an effective method to reduce mortality from lung cancer, balance between harms and benefits is a perquisite for the implementation of a lung cancer screening program.

In the General Discussion, this thesis is evaluated to determine its implications. Firstly, the research objectives of this thesis are summarised. Secondly, background information necessary for the interpretation of the results is provided. Thirdly, the main and sub research questions of this thesis are evaluated for each chapter consecutively. The sub research questions are used to present and interpret the main results in each chapter. All sections on the interpretation of results conclude by answering a sub research question. The main research question of each chapter is answered in the conclusion section. Fourthly, a list of general conclusions based on the conclusions of each chapter are presented. Finally, general recommendations for further research and clinical practice are presented.

THIS THESIS

As the design of an effective and feasible screening algorithm is crucial for the implementation of screening, the evaluation of screening trials may provide valuable knowledge. Topic of this thesis is the evaluation of the screening algorithm of the Dutch-Belgian lung cancer screening trial; the NELSON trial. Hence, aims of this thesis were to estimate test characteristics (sensitivity, specificity and predictive values), number of required follow-up CT examinations and additional diagnostic tests, to determine whether improvements to the screening algorithm were possible by identifying failures and unnecessary procedures, and to estimate the performance of improved hypothetical screening algorithms. Ideally, the evaluation of the performance of the screening algorithm also considers effectiveness, which is reduction in lung cancer mortality. However, these analyses are planned at ten years after randomisation, which is outside the scope of this thesis. Nonetheless, this thesis contributed to these mortality analyses by developing the protocol for endpoint verification.

BACKGROUND

A screening algorithm is the management protocol of a screening programme or trial. The algorithm should define each possible screening test result and recommend an ap-

propriate management strategy. Since low-dose computed tomography examination of the chest (LDCT) is used as a screening test, there are numerous options to define the screening test result. Therefore, many different screening algorithms have been designed and applied in lung cancer screening trials.

Although the interpretation of the LDCT examination differs substantially between lung cancer screening trials, three screening test outcomes have been defined quite consistently. Firstly, in case the screening LDCT examination did not show any abnormality, the screening test result is usually defined as 'normal' or 'negative' and the recommended management is not to perform any additional diagnostic tests. Secondly, in case the screening LDCT examination showed an abnormality that required immediate medical care, for example a large aortic aneurysm, screening is usually put on hold and adequate medical care is arranged. After this, an assessment is made to determine whether continuation of screening and consequential diagnostic testing is still appropriate. Finally, in case abnormalities suspicious for advanced lung cancer are detected on the screening LDCT examination, the screening test result is usually defined as 'positive', and the recommended management is to perform a work-up for diagnosis and staging followed by treatment according to (inter)national guidelines. In all other cases (thus cases wherein abnormalities not suspicious for advanced lung cancer were visible and no acute medical care was indicated) lung cancer screening trials use various definitions for screening test results and recommend opposing management strategies.

Although a variety of different abnormalities may be detected on the LDCT screening examination, the screening test result is determined by those abnormalities that are suspicious for (pre)cancerous lesions, which are lung nodules or masses. Pulmonary nodules are defined by the Fleischner Society as rounded or irregular opacities, that are well or poorly defined and measure up to 3 cm in diameter.[1] Masses are defined as pulmonary, pleural, or mediastinal lesions that are seen as an opacity greater than 3 cm in diameter, and are usually solid or partly solid.1 Traditionally, the visual characteristics and size of the lung nodules or masses detected on the LDCT screening examination determines the screening test result.

In the first lung cancer screening cohort studies, such as the U.S. Early Lung Cancer Action Project (ELCAP study), all detected lung nodules which did not show a benign calcification pattern, were classified as a suspicious for lung cancer and the screening test result as 'positive'. Usually, the recommended management strategy was to perform series follow-up CT scans for a period of two years for small nodules (<6 mm in diameter) and invasive diagnostic procedures, such as transthoracic biopsy, for larger nodules and lung masses. From this first generation of lung cancer CT screening algorithms we learned that the majority of these small non-calcified pulmonary nodules were not malignant.[2] As a result, the predictive value of a 'positive' screening test results was very low.[2] This led to vast amounts of follow-up CT examinations and diagnostic procedures for benign

nodules. Another important lesson we learned from these studies is that the size of the nodule is highly correlated with the probability of malignancy.[3]

The next generation of lung cancer screening studies, such as the U.S. National Lung Screening Trial (NLST) used the knowledge from previous studies. Hence, these studies applied screening algorithms that did not classify all non-calcified pulmonary nodules as suspicious for lung cancer, but used the size of the nodule to determine whether the screening test result was 'positive' or not.[4] A threshold for nodule size that was commonly used to define 'suspicion of lung cancer' was a nodule diameter of 4 mm.[5] Hence, subjects with nodules smaller than 4 mm would not receive follow-up CT scans or diagnostic tests. While subjects with nodules of 4 mm or larger were considered as suspicious for lung cancer and were recommended to undergo series of follow-up CT examinations or additional diagnostic tests (often depending on clinical judgement of the referring physician).[4] From these studies we learned that the sensitivity for lung cancer of such screening algorithms is high, which means that it is safe to perform no additional CTs for nodules smaller than 4mm in diameter.[6,7] Nonetheless, the predictive value of a 'positive' screening test result was still low due to a moderate specificity of these screening algorithms.[6,7]

The next generation of lung cancer screening studies, such as the NELSON trial, aimed to achieve both a high sensitivity for lung cancer and a high specificity. To achieve this, another nodule feature that is highly predictive of malignancy was included in the screening algorithms: nodule growth.[8-11] Unfortunately, the traditional manual or visual assessment of nodule diameter was not accurate enough to determine the growth rate of small non-calcified nodules at short notice. Therefore, these studies used semi-automatic volumetric software to assess the differences in nodule size on subsequent screening examinations.[8-11] Hence, the screening protocols used a two-step approach to determine which of the detected non-calcified nodules were suspicious of lung cancer. First, nodules under a certain size threshold (smaller than 50 mm^3 in the NELSON trial) were classified as not suspicious or 'negative', and nodules larger than a certain size threshold (larger than 500 mm^3 in the NELSON trial) were directly classified as suspicious for lung cancer or 'positive'. Next, the nodules with a size between these thresholds were classified as 'indeterminate' and were scheduled for another LDCT screening examination at short notice to determine nodule growth. Only the indeterminate-sized nodules that demonstrated malignant growth were classified as suspicious for lung cancer or 'positive' (in the NELSON trial defined as a percentage volume change of ≥25% combined with a volume-doubling time shorter than 400 days). Summarising, only nodules above a certain size threshold and intermediate-sized nodules growing faster than a certain threshold are classified as suspicious for lung cancer or 'positive' and additional diagnostic testing is recommended. Hypothetically, the use of follow-up CT scans in these screening algorithms will lead to more targeted and economical use of additional diagnostic tests, which will improve specificity and will still yield a high sensitivity for lung cancer.

The series of diagnostic tests performed after a positive screening test result are called the diagnostic work-up, and have the objective to either diagnose or rule out lung cancer. In general, the diagnostic work-up is not coordinated by the screening trial. Referring physicians use (inter)national guidelines for the management of pulmonary nodules to assist them with decision-making in the diagnostic work-up for suspicious screen-detected nodules. Despite the fact that this process is not incorporated in most screening trials, the performance evaluation of a screening protocol should also consider the diagnostic work-up. This is because performing and interpreting a screening test only identifies persons who are suspected of having lung cancer. The work-up is the next essential step in diagnosing lung cancer. Medical tests, other than LDCT examinations, are used to distinguish persons who actually have lung cancer from those who had a false-positive screening test result. The sensitivity of a screening programme will be affected when the diagnostic work-up does not effectively pick out all subjects with lung cancer.

Once the diagnosis of lung cancer has been made, the next essential step is the treatment of lung cancer. This might seem trivial, but diagnosing lung cancer earlier by LDCT screening does not affect lung cancer mortality. It is only a prerequisite that enables the early treatment of lung cancers. Only in case the treatment of early diagnosed lung cancers has a higher cure rate than in lung cancers diagnosed through symptoms, mortality from lung cancer can be reduced by screening. Therefore, the performance evaluation of a screening algorithm should also consider the lung cancer mortality reduction. To determine the mortality reduction of a screening program, a randomised controlled trial is indispensable. The NELSON trial is the largest randomised trial wherein screening using low-dose CT is compared to no screening.

RESEARCH QUESTIONS

Research question I

Chapter 2. Predictive value of screening test results

Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial.

European Respiratory Journal

Main research question

What was the screening performance of the nodule management protocol of the NELSON trial?

Sub research questions

a) What were the detection rates, test characteristics and numbers needed to screen of the nodule management protocol of the NELSON trial?
b) What was the incidence of invasive diagnostic procedures for false-positive screening test results?
c) What were participant's probabilities of false-positive screening results and lung cancer after baseline and subsequent screening test results?

Main results

a) In the first three rounds of the NELSON trial a total of 24,354 CT scans were performed. 89.4% of the scans were a regular scans and 10.6% performed to assess the VDT of indeterminately sized nodules, the so-called 'follow-up scans'. The screening test result was negative in 87.2%, indeterminate in 10.8% and positive in 2.0% of the scans. The positive scans eventually led to the diagnosis of lung cancer in 200 persons (cumulative lung cancer detection rate: 2.6%). Hence, the predictive value of a positive screening test result was 40.6% and 59.4% of the positive screening test results was false-positive. Overall, 1.2% of all 24,354 CT scans had a false-positive result. Finally, the number needed to screen for the diagnosis of one lung cancer was 92-133 per round.
b) Across the three screening rounds, 6.0% of the participants received one or more positive screening result. As 59.4% of the positive screening test screening results was false-positive, 3.6% of all participants had one or more false-positive screening result. 24.5% of the subjects with a false-positive screening result underwent an invasive diagnostic procedure. Hence, invasive diagnostic procedures for false-positive screening results were performed in 0.9% of all participants.
c) The probabilities of receiving a false-positive screening result or being diagnosed with lung cancer depend on the result of the baseline and subsequent screening tests. The estimated risk of a false-positive screening result within the next 5.5 years was respectively 1.3%, 8.8% and 54.2% for the individuals with a negative, indeterminate or positive baseline scan. Moreover, the estimated 5.5-year risk of screen-detected lung cancer was only 1.0% for the individuals with a negative baseline scan result, 5.7% for subjects with an indeterminate baseline result and 48.3% for those with a positive baseline.

Interpretation of results

a) The majority of the LDCT screening examinations in the NELSON trial had a negative result. Only in about one in ten screening tests an additional follow-up LDCT examination had to be performed. Moreover, only 2% of the LDCT screening examinations were positive. These results are comparable to the results of the

Danish lung cancer screening trial (DLCST), which also has a 'third generation' volumetry-based screening algorithm.[9,12] However, the number of positive screening tests results was substantially lower than the 24.2% positive screening test results in the NLST, wherein a second-generation screening algorithm using a single nodule size criterion of 4mm is used to define suspicion for lung cancer.[13]

Despite the lower percentage of positive screenings, the cumulative lung cancer detection rate of 2.6% was slightly higher than in the NLST (2.4%),[13] but lower than in the DLCST (3.4%).[9,12] The latter is probably due to the two additional screening rounds that have been completed in the DLCST.

The predictive value of a positive screen result was higher in the NELSON trial (40.6%) than in both the DLCST (34.8%) and the NLST (3.6%).[9,12,13] As a result of this and the low percentage positive screening test result, the proportion of false-positive scans out of all scans is slightly lower in the NELSON trial (1.2%) compared to the DLCST (1.3%), substantially better than in the NLST (23.3%).[9,12,13]

Finally, the number needed to screen for the detection of one lung cancer was 92-133 per round in the NELSON trial, which is a little less than in the other trials (97-147 in the NLST and 116-180 in the DLCST).[9,12,13]

Concluding, the detection rate, positive predictive value and number needed to screen of the screening algorithm of the NELSON trial compare favourably to other lung cancer screening trials. Nonetheless, before an eventual implementation of lung cancer screening, efforts should be made to reduce the proportion of indeterminate screening test results without loss of efficacy.

b) The percentage of participants with one or more positive scan (6.0%) was low in our trial compared to the NLST (39.1%).[13] Therefore, also the percentage of participants that had one or more false-positive screening tests was lower in the NELSON trial (3.6%) compared to the NLST (25.8% in round one, 27.2% in round two and 15.9% in round three).[6,7]

Despite this, more invasive diagnostic procedures for false-positive screening results were performed in the NELSON trial (0.9%) than in the NLST (0.6%). Apparently, a more cautious approach to positive screening test results was applied in the United States than in the Netherlands and Belgium. Probably, suspicious screen-detected nodules were more often followed up using serial CT examinations, instead of directly proceeding to invasive procedures.[14]

Concluding, invasive procedures for false positive screening test results cannot be eliminated because biopsy or surgery is sometimes the only way to distinguish lung cancer from a benign nodule. Nonetheless, this study showed that the incidence of these 'unnecessary' procedures could be lower than observed in the NELSON trial, as it is lower in the NLST (corrected for the difference in percentage positive screening tests). Therefore, efforts should be made to investigate what the causes of

the slightly higher rate of invasive procedures in the NELSON trial were. Examples are: insufficient guidance of referring clinicians, lack of a national guideline on the management pulmonary nodules, unavailability of tools to estimate lung cancer probability of screen-detected nodules or inexperience with CT-detected nodules. Once the causes have been identified, targeted interventions should be developed, evaluated and if successful implemented before lung cancer screening is implemented.

c) This study demonstrated that participants with a negative, indeterminate or positive first screening test had very distinct risks of false-positive screening results and lung cancer. Analyses showed a significant increase in age and number of pack-years in participants with respectively negative, indeterminate and positive screening test result, which are all well known risk factors for developing lung cancer.[15,16] The presented results could aid clinicians when counselling individuals comparable to NELSON study participants when LDCT examinations are interpreted in the same fashion as in the NELSON trial.

Concluding, the result of the LDCT screening test adequately stratifies participants by their risk of lung cancer. Moreover, the predictive value of screening results lasts up to 5.5 years. As the screening test result is based on the size and growth rate of pulmonary nodules, these variables are probably very strong predictors of lung cancer risk. Hence, future studies should assess whether it is possible to build a reliable lung cancer prediction model that uses nodule size and growth rate in addition to individual characteristics such as age, gender, smoking status and smoking history. Once a reliable lung cancer prediction model has been built and validated, it should be investigated how the model can be made useful for clinicians who are confronted with the management of CT-detected pulmonary nodules. Such a lung cancer prediction tool may help identifying the malignant nodules and may reduce unnecessary diagnostic procedures for benign nodules.

Conclusion

The screening algorithm of the NELSON trial adequately stratified participants according to their lung cancer risk. The NELSON screening algorithm yielded a limited number of follow-up LDCT scans for indeterminate screening test results and a low number of diagnostic work-ups for positive screening test results. Although the predictive value of screening test results compared favourably to other studies, perhaps too many invasive diagnostic procedures for benign nodules were performed.

Research question II

Chapter 3. Characteristics of screen-detected lung cancer

Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial.

American Journal of Respiratory and Critical Care Medicine

Main research question

What was the effect of screening using low-dose computed tomography on the characteristics of screen-detected lung cancer?

Sub research questions

a) What were the tumour characteristics of lung cancers detected by low-dose CT screening?
b) What was the effect of screening round and gender on the characteristics of screen-detected lung cancer?
c) To what extent was screening able to detect lung cancer before the onset of symptoms?

Main results

a) In the first three screening rounds of the NELSON trial, a total of 209 lung cancers in 200 persons were detected via screening. The most common histological type was adenocarcinoma (51.2%), followed by squamous cell carcinoma (16.3%), large cell carcinoma (8.1%), bronchoalveolar carcinoma (5.3%) and small cell carcinoma (3.8%). The majority of the screen-detected lung cancers were detected at a limited stage (70.8% at stage I), a minority was detected at a locally advanced stage (21.1% at stage II and IIIA) and few at an advanced disease stage (8.1% at stage IIIB and IV). Most screen-detected lung cancers were localised in the right lung (65.6% whereof 45.0% in the right upper lobe). The cancers were also predominantly localised in the periphery of lungs; 62.2% in the outer one-third of the costal-hilar diameter; adenocarcinomas in particular (82.2% vs. 17.8%, p=0.001).
b) The lung cancers detected in round 1 had a slightly higher disease stage (stage IA 59.5%, stage IV 6.8%) than in later rounds (round 2: stage IA 74.1%, stage IV 3.4%, and round 3: stage IA 64.9%, stage IV 3.9%), but this was not statistically significant. Also, the proportion histological type and localisation of the screen-detected lung cancers was not significantly different across the screen rounds.

None of the histological subtypes of lung cancer were unevenly distributed between the sexes. Also, the localisation of the lung cancers was not significantly different between the sexes: neither for the left lung versus right lung localisation nor for peripheral versus central localisation. However, cancer stage at diagnosis was significantly lower in women than in men (p=0.005). After correction for the sex differences in age, number of pack-years and BMI, women still had a statistically significant lower cancer stage than men (p=0.028).

c) 189 of the 200 (94.5%) participants who were diagnosed with lung cancer through screening had no symptoms suspicious of lung cancer. The remaining eleven participants (5.5%) had symptoms suspicious of lung cancer before they were diagnosed. In five of them symptoms emerged before the screening scan was made; however, none of them had symptoms at randomisation. Three subjects experienced suspicious symptoms for the first time in the period between the positive scan and the first consultation, and three subjects had reported the symptoms in the period between the first consultation and the diagnosis date.

Interpretation of results

a) The stage distribution of the screen-detected lung cancers in the NELSON trial (70.8% at stage I and 8.1% at stage IIIB/IV) is considerably more favourable than the stage distribution of clinically diagnosed lung cancers (28% at stage I and 18% at stage IIIB/IV).[89] Moreover, the stage distribution appears also to be relatively favourable compared with the other lung cancer screening trials (on average: 64.7% at stage I and 10.9% at IIIB–IV).[11-13,51,86,87]

This study demonstrated that the disease stage at diagnosis strongly correlated with the histological subtype of the lung cancer. On the one hand, LDCT screening detected many relatively slow growing, peripherally localised adenocarcinomas an early stage. On the other hand, LDCT screening detected only a few small cell lung cancers and all were diagnosed at stage III-IV. This finding could imply that LDCT screening is not, or is less, capable of early detection in some fast-growing histological subtypes of lung cancer. Therefore, future studies should include all lung cancers in screened participants, not only the screen-detected lung cancers, to complete the picture of lung cancer characteristics.

This study also demonstrated that most screen-detected lung cancers were localised in the periphery of the lungs, which is probably a result of the detection of many adenocarcinomas and the use of low-dose, unenhanced CT scans which can have limited image quality in hilar regions and the mediastinum.[19] The finding that 45.0% of all lung cancers were localised in the right upper lobe is known from patients with non-small cell lung cancer. This phenomenon may be explained by the fact that the airflow at the beginning of the breath is the largest toward the right

upper lobe bronchus.[90,91] As a result, the deposition of particles in tobacco smoke and their carcinogenic effects are the largest in the right upper lobe.[92,93]

Concluding, the tumour characteristics of the lung cancers detected by low-dose CT screening in the NELSON trial are typically described as early stage adenocarcinomas localised in the periphery of the right upper lobe. Although the stage distribution of the screen-detected lung cancers suggests a stage shift as a result of LDCT screening, a comparison of all lung cancers diagnosed in the screening group with all lung cancers diagnosed in the control group is the only valid method to determine this.

b) Analyses showed no significant effect of screening round on cancer stage, histology, or tumour localisation. However, a decrease in advanced-stage lung cancers was observed at the second screening round (stage IV dropped from 6.8 to 3.4%). This was probably not statistically significant because of the low absolute number of advanced-stage lung cancers. In the third screening round, no evident increase in stage IV lung cancers was observed (3.9%), despite the screening interval of 2 years.

The differences in lung cancer characteristics between men and women have been studied extensively. In general, studies found that women are, in general, diagnosed at an earlier age,[94,95] at a more favourable cancer stage,[95-97] and are more often diagnosed with adenocarcinomas than are men.[94,98,99] This study was the first trial to report on sex differences in lung cancer in a screening setting and also demonstrated that women were diagnosed at a significantly more favourable cancer stage compared to men. However, the histological subtype and localisation of the lung cancers were not significantly different between the sexes. In 2013, a post hoc analyses using data from the NLST was published. The investigators did not specifically analyse sex differences in lung cancer characteristics, however they showed that women benefitted more from LDCT screening than men.[20] This is in line with the more favourable stage distribution in women as demonstrated in the current study of the NELSON trial.

Concluding, no effect of screening round on lung cancer characteristics could be demonstrated in the NELSON trial. The gender of the participants only affected the disease stage at diagnosis of the screen-detected lung cancer, but no effect on histological subtype or localisation has been observed. Because of the more favourable in stage distribution and higher effectiveness of LDCT screening in women, post hoc analyses of the effectiveness of LDCT screening in the NELSON trial stratified by gender should be performed in the future.

c) Almost all (94.5%) screen-detected lung cancers were diagnosed before the onset of symptoms. In 5.5% of the participants, lung cancer was diagnosed after the first symptoms emerged. In 3.0% the participants reported that the symptoms had started in the interval between the LDCT screening test and the date of diagnosis.

It cannot be excluded that these symptoms were actually present for a longer time, but only recognised as serious by the participant after a suspicion of lung cancer was raised by the screening examination and subsequent tests.

Concluding, LDCT screening in the NELSON trial was able to detect lung cancer before the onset of symptoms in the large majority of the participants.

Conclusion

This study suggests that screening using low-dose computed tomography leads to a stage shift towards earlier diagnosis, more in women than in men, and a shift in histology towards slower growing and more peripherally localised subtypes of lung cancer.

Research question III

Chapter 4. Epidemiological evaluation

Detection of lung cancer through low-dose CT screening: analysis of screening test performance and interval cancers.

Lancet Oncology

Main research question

How can knowledge on the lung cancers not detected by low-dose computed tomography screening be used to improve the performance of the screening strategy?

Sub research questions

a) What were the detection rates and test characteristics of the nodule management protocol of the NELSON trial?
b) Were there any differences in the characteristics between the participants diagnosed with screen-detected lung cancer and the participants diagnosed with an interval cancer?
c) What were the tumour characteristics of the lung cancers not detected by low-dose computed tomography screening?
d) What were causes of the failure to detect the interval cancers?

Main findings

For this study, all Dutch participants who received at least one screening in the first three rounds (n=7,155) were included. Data on all lung cancers diagnosed from the first screening round to the last screening in round three, plus an additional two years of follow-up

were obtained from the Dutch Cancer Registry. The Belgian participants (n=935) were excluded as a linkage with the Belgian Cancer Registry was not yet possible.

a) In the first three screening rounds, a total of 187 participants were diagnosed with lung cancer through screening. Another 34 subjects were diagnosed with lung cancer between screening rounds. Hence, per 1000 screened subjects, 26.1 lung cancers were detected by screening, and 4.8 lung cancers were not (ratio 5.5:1). Across the three screening rounds, the ratio between the detected and missed lung cancers decreased from the first to the second round from 12.4:1 to 2.9:1, and increased from round two to round three to 7.2:1.

   The test characteristics for the three screening rounds combined were: sensitivity 84.6% (95% confidence interval (CI) 79.9-89.3%), specificity 98.6% (95% CI 98.5-98.8%), PPV 40.4% (95% CI 35.9-44.7%), and NPV 99.8% (95% CI 99.8-99.9%). When only the first year of the interval between the screening rounds was considered, the performance was: sensitivity 90.8% (95% CI 86.4-94.5%), specificity 98.7% (95% CI 98.5-98.8%), positive and negative predictive values respectively, 40.4% (95% CI 35.9-44.7%) and 99.9% (95% CI 99.9-99.9%). Across the three screening rounds, the sensitivity and specificity were respectively: 92.5% and 98.3% in the first round, 73.6% and 99.0% in the second round, and 87.8% and 98.7% in the third screening round.

b) The participants diagnosed with either screen-detected lung cancer or interval lung cancer were significantly older than the subjects without lung cancer, however no differences were observed in the gender, or number of pack-years smoked. Only participants diagnosed with an interval cancer were significantly more often current smokers than those without lung cancer. Analyses between the participants with interval cancer and participants with screen-detected lung cancer showed that there was only a significant difference in smoking status.

c) The cancer stage at diagnosis of the lung cancers not detected by screening was as follows: 8.6% was diagnosed at stage IA, 8.6% at stage IIB, 8.6% at stage IIIA, 5.7% at stage IIIB, and 68.6% at stage IV. This disease stage distribution was significantly less favourable than the stage distribution of the screen-detected lung cancers. The interval cancers diagnosed within the first year of the screening interval had a significantly higher disease stage than in the interval cancers diagnosed in the second year of the interval.

   The distribution over the histological subtypes of lung cancer was as follows for the interval cancers: 25.7% adenocarcinomas, 20.0% small cell carcinomas, 17.1% squamous cell carcinomas, 17.1% large cell carcinomas, the remaining 20.0% were other rarer subtypes and lung cancers of unknown histopathological subtype. The interval cancers were significantly more often small cell carcinomas, and signifi-

cantly less often adenocarcinomas compared to screen-detected cancers. The other histological subtypes were equally distributed.

Finally, the localisation of both the interval cancers and screen-detected cancer was equally distributed across the lungs.

d) Re-evaluation of the CT examinations of the 34 participants with an interval cancer learned that no lung cancer was present at the last screening examination in 35.3%. In the remaining 64.7% an abnormality suspicious for lung cancer could, in retrospect, be identified on the screening CT examination. In the majority of these cases, the suspicious abnormality was missed. The causes of the failure to detect these lung cancers were: detection errors (38.2%), interpretation errors (5.9%) and human error (5.9%). In the remaining cases, the abnormality was actually detected, but lung cancer was not diagnosed because of: participant non-compliance (5.9%), not classified as suspicious by the protocol (2.9%), and manually classified as not suspicious by the radiologist due to a negative diagnostic work-up at a previous screening round (5.9%).

Interpretation of results

a) This study demonstrated that the detection rate was high at 26.1 per 1000 screened subjects for three screening rounds. Since the NELSON trial has screening intervals of more than one year from the second round onwards, only the detection rates of the first round and the first year of the second round may be compared to other screening trials with a one-year screening interval. Henceforth, the detection rate for the first screening round was 8.69 per 1000 screened in the NELSON trial, which was lower than in the U.S. National Lung Screening Trial (NLST): 10.3 per 1000 screened.[6] This difference may be explained by the lower lung cancer risk of the NELSON participants compared to the NLST participants,[17,18] and the slightly lower sensitivity in the NELSON trial (92.5%) than in the NLST (93.8%).[6] The incidence of interval cancers between the first and second round was comparable in the two trials (0.70 per 1000 screened in NELSON and 0.68 per 1000 in NLST).[6] This might indicate that the majority of lung cancers not detected at baseline in the NELSON trial, did not become symptomatic in the one-year interval and were diagnosed through screening in the second round. This explanation is supported by the observation that the lung cancer detection rate in the second screening round was higher in the NELSON trial (7.69 per 1000) than in the NLST (6.80 per 1000).[7] In the first year of the screening interval after the second round scan, substantially more interval cancers were diagnosed in the NELSON trial (1.02 per 1000) than in the NLST (0.40 per 1000).[7] This probably results from the difference in sensitivity in the second round (NELSON 88.3% versus NLST 94.4%).[7]

The sensitivity and specificity of NELSON screening algorithm were respectively 92.5% and 98.3% in round one, 73.6% and 99.0% in round two, and 87.8% and 98.7% in round three. Whether the sensitivity is sufficient to obtain a significant lung cancer mortality reduction can only be determined by the final mortality analyses of the NELSON trial, which are planned ten years after randomisation. It is likely that the specificity is sufficient to obtain cost-effectiveness.

Since the NELSON trial has screening intervals of more than one year from the second round onwards, only the detection rates of the first round and the first year of the second round may be compared to other screening trials with a one-year screening interval. Henceforth, the sensitivity of 92.5% in the first round and 88.3% in the second round is slightly lower than in the NLST (93.8% in the first round and 94.4% in the second round).[6,7] The specificity of 98.3% in the first round and 99.0% in the second round is substantially higher than in the NLST (73.4% in the first round and 72.6% in the second round).[6,7]

Concluding, the detection rate for the first three screening rounds of NELSON trial was 26.1 per 1000 participants. Simultaneously, 4.8 lung cancers per 100 participants were not detected by screening. The test characteristics for the first three screening rounds combined were: sensitivity 84.6%, specificity 98.6%, PPV 40.4%, and NPV 99.8%. These detection rates and test characteristics are promising for the cost-effectiveness of the NELSON screening trial. Nonetheless, the results of the mortality analysis and subsequent cost-effectiveness analysis should be awaited.

b) All participants of this study were at substantial risk of developing lung cancer as they were at least 50 years old and had smoked ten or more cigarettes a day for over 30 years, or fifteen or more cigarettes a day for over 25 years. Even within this population, older age and being a current smoker were still significant risk factors for developing lung cancer. Remarkably, being a current smoker is only associated with an increased risk of being diagnosed with an interval lung cancer. This may be because continued smoking promotes the development of lung cancer subtypes that grow faster and are less perceptible by LDCT screening, such as small cell carcinomas.[19] This finding reinforces the urgency of smoking cessation in individuals undergoing lung cancer screening.

   Concluding, subjects diagnosed with an interval cancer were significantly more often current smokers than the participants with screen-detected lung cancer; no differences in age, gender and number of pack-years smoked were observed.

c) This study demonstrated, not surprisingly, that screen-detected lung cancers are diagnosed at a notable more favourable cancer stage than interval cancers. More noticeable was the finding that the disease stage of the interval cancers diagnosed in the first year since screening was significantly higher than the stage of those diagnosed in the second year.

This study also demonstrated that interval cancers are different in histopathology compared to screen-detected lung cancers. Interval cancers were significantly less often adenocarcinomas and not a single interval cancer was a bronchoalveolar carcinoma. Interval cancers were slightly more often large cell carcinomas and squamous cell carcinomas, and significantly more often small cell carcinomas than screen-detected lung cancers.

The differences in tumour characteristics are both caused by the earlier diagnosis of screen-detected lung cancer as a result of screening asymptomatic individuals, and by the more aggressive nature of interval lung cancers compared to detected cancers. This study revealed that 35.3% of the interval cancers newly developed during the screening interval, all these interval cancers were diagnosed at stage III/IV. Hence, these cancers grew from undetectable to incurable cancers in less than one or two years. This observation suggests an enormous growth and metastatic potential. This suits with the finding that these interval cancers were significantly more often small cell carcinomas than the interval cancers that did not arise during the screening interval (41.7% versus 8.7%, p=0.03).

This study is the first to present the cancer stage distribution of both the screen-detected and the interval cancers of the NELSON trial. Hence, 61.9% of all lung cancers were diagnosed at stage I, and only 17.8% at stage IIIB/IV. In the NLST, 59.0% of the lung cancers was diagnosed at stage I, and 22.9% at stage IIIB/IV, which is not significantly different (p=0.20). Thus, despite longer screening intervals, slightly lower sensitivity, and fewer female participants[20] in the NELSON trial, lung cancer was diagnosed as early as in the NLST.[13] This finding is encouraging for the effectiveness of lung cancer screening regimens using biannual screening after an initial annual screening round.

Concluding, the lung cancers not detected by CT screening were characterised by a higher cancer stage at diagnosis; more than 70% of the interval cancers were diagnosed at an incurable stage (IIIB/IV). Further, the interval cancers were most frequently of the adenocarcinoma subtype, followed by small cell, squamous cell carcinomas, and large cell subtype.

d) Re-evaluation of the clinical CT and last screening CT examination revealed the causes of the failure to detect interval cancers. Surprisingly, 64.7% of the interval cancers were, in retrospect, visible at the last screening CT examination. Detection, interpretation and human errors were identified as the main cause of failure in 50.0% of the interval cancers. In addition, 2.9% of the interval cancers were not diagnosed through screening because the screening test result was adjusted manually by the radiologist from positive to negative because a diagnostic work-up performed in an earlier round did not yield the diagnosis of lung cancer.

Another remarkable finding of this study was that failure of the screening protocol to classify cancerous nodules as suspicious was rare. Only 5.9% of the interval cancers were not diagnosed because the cancerous nodule shrunk or had a volume doubling-time of more than 400 days. This suggests that the relatively stringent criteria for a positive result in the NELSON trial, did not lead to notable numbers of missed cancers. This finding is encouraging for future screening programmes that pursue limited harms and costs, as more stringent criteria contribute to this.[21]

Further, 5.9% of the interval cancers were actually detected, but the diagnosis was not made through screening because participants refused to comply with the screening protocol. Instead of undergoing a follow-up CT at three months after their indeterminate screening test result, they directly underwent a diagnostic resection of the nodule, which yielded the diagnosis of lung cancer. Arguably, these interval cancers may have been screen-detected lung cancer in case the participants had complied with the protocol.

Finally, 35.3% of the interval cancers were, also in retrospect, not visible at the last screening examination. Hence, these interval cancers were not missed but arose during the interval.

Concluding, radiological errors were the most important cause of the failure to detect the interval cancers. Failures by the protocol or non-adherence to the protocol were infrequent causes of detection failure. More than one third of the interval cancers could not be prevented as they were not present at the last screening CT examination, but arose during the screening interval.

Conclusion

The detection rates and sensitivity of the NELSON screening protocol were sufficient to diagnose lung cancer as early as in the NLST, which demonstrated to have an effective screening protocol.[13] Moreover, the NELSON screening protocol yielded a very high specificity, which is a prerequisite for cost-effectiveness. Nonetheless, the performance of the screening protocol may be improved by co-implementation of CT screening with an effective smoking cessation program, and training of the screening radiologists to reduce the number of detection and interpretation errors.

Research question IV

Chapter 5. Radiological evaluation

Computed tomographic characteristics of interval and post-screen cancers in lung cancer screening.

European Radiology

Main research question

How can knowledge on the radiological characteristics of lung cancers not detected by low-dose CT screening be used to improve the performance of the screening strategy?

Sub research questions

a) What proportion of the lung cancers not diagnosed through screening was, in retrospect, present at the last LDCT screening examination?
b) What were the causes of the failure to detect the missed lung cancers?
c) What were the characteristics of the carcinomas missed on the LDCT screening examination due to radiological detection or interpretation errors?

Main findings

For this study all 7,155 Dutch participants randomised to the CT screening arm of the NELSON trial were included. The participants with lung cancer diagnosed between screening rounds (interval cancers) and the participants with lung cancer diagnosed after screening (post-screen cancers) were identified via linkages with the Dutch Cancer Registry. The Belgian participants (n=935) were excluded as a linkage with the Belgian Cancer Registry was not yet possible.

In the first three rounds of the NELSON trial, LDCT screening detected lung cancer in 187 of the 7,155 (2.6%) participants. In another 61 of the 7,155 (0.85%) participants, lung cancer was diagnosed between screening rounds or after the subject's last attended LDCT screening examination. Of these 61 participants clinical and radiological files were retrieved from the various hospitals in which the diagnosis was established. The clinical CT examination made at the time of the diagnosis was compared to the last screening LDCT examination by two experienced radiologists to determine whether any CT evidence of lung cancer could be identified in retrospect on the screening CT. In case of any abnormalities on the LDCT screening examination, the radiologists determined whether abnormalities were missed or misinterpreted by comparing their reading report by the original reading report in the trial database. The missed lung cancers could be caused by detection errors and interpretation errors. In detection errors the lesion was not mentioned in the report but can be seen in retrospect on the last CT, while in interpretation errors the lesion was noted but considered a benign lesion. Finally, the radiologists searched for other abnormalities on the LDCT screening examination, which may influence detection or interpretation.

a) In 26 of the 61 (42.6%) participants with a lung cancer not detected through screening no abnormalities suspicious of lung cancer were visible on their last attended LDCT screening examination. In eleven of these 26 subjects (42.3%) the lung cancer

was an interval cancer as it was diagnosed before their next scheduled LDCT screening examination. In 15 of the 26 subjects (57.7%) the lung cancer was diagnosed after the participant ceased with the screening program; a 'post-screening cancer'. In the remaining 35 of the 61 (57.4%) participants with a lung cancer not detected through screening, the radiological re-evaluation showed that the lung cancer was actually present at the last LDCT screening examination.

b) In 35 of the 61 participants with a lung cancer not diagnosed through screening a suspicious abnormality was visible on the last LDCT screening examination. Twenty of these 35 lung cancers were interval cancers and fifteen were post-screening cancers. Radiological re-evaluation of these 35 cases showed that there were various reasons for the failure to diagnose the lung cancers.

In twenty cases, the lesions suspicious of lung cancer were not found on the LDCT examination by the radiologist due to detection error or human error. In another two cases the lesions suspicious for lung cancer were detected, but the lung cancers were not diagnosed due to interpretation errors by the radiologist. In thirteen cases the lesions suspicious of lung cancer were detected, but the lung cancers were not diagnosed due to:

I) Failure of the protocol: two (5.7%) lung cancers were not diagnosed via the NELSON trial because the protocol did not classify the nodules as suspicious for lung cancer. In these cases the protocol was adhered to and no positive screening result was issued as the nodules did not show growth.

II) Non-compliance with the protocol by the participant: eight (22.9%) lung cancers were not diagnosed via the NELSON trial because the participant refused to comply with the study's recommendations based on the LDCT screening examination.

III) Non-compliance with the protocol by the radiologist: three (8.6%) lung cancers were not diagnosed via the NELSON trial because the radiologist had replaced the actual screening result by a negative screening result because a previous diagnostic work-up had not yielded the diagnosis of lung cancer.

c) In the 22 of the 61 participants a suspicious abnormality was in retrospect visible on the last LDCT screening examination, but lung cancer was not diagnosed due to interpretation detection or radiological errors. Two lung cancers were detected but not diagnosed due to interpretation error. In one case, a lesion of 7 mm in diameter was noted, but no further action was undertaken because it was interpreted as benign bulla wall thickening. In the other case, a lesion attached to the pleura of 22 mm in diameter was detected, but not referred for diagnostic work-up because it was interpreted as scarring. In the remaining twenty participants, lung cancer was present but was not detected at the LDCT screening examination. The

characteristics of these twenty lung cancers may give an indication of the causes of the detection errors:

I) Endobronchial localisation: Five lung cancers were visible on CT as small central endobronchial tumours. Two were localised in the right pectoral segmental bronchus, one in the right lateral segmental bronchus, one in the right upper lobe bronchus and one in the lingular bronchus.

II) Bulla wall thickening: Four lung cancers were visible on CT as a thickening of the wall of a bulla. These lung cancers were localised in the right upper lobe (n = 2), left upper lobe and the left lower lobe. In one case the wall thickening was focal.

III) Pleural attachment: Four lung cancers were visible on CT as nodules attached to the pleura. Three of these nodules were smaller than 1 cm (5, 7 and 7 mm) and one was larger than 1 cm.

IV) Pleural effusion: One lung cancer was not visible as a lung nodule or lung mass on CT, but as a pleural effusion on the right side. 177 days after the screening examination, the diagnosis of lung cancer with massive malignant pleural effusion was made.

V) Lymphadenopathy: Three lung cancers were not visible as a lung nodule or lung mass on CT, but as lymphadenopathy. In two cases the lymphadenopathy was located in the right hilum, inseparable from the pulmonary artery due to the lack of intravenous contrast. In the third case, lymphadenopathy measured 22 mm was mainly located in the aortopulmonary window.

VI) Fibrosis: One lung cancer was visible on CT as a nodule smaller than 1 cm in diameter surrounded by extensive reticulation.

VII) Human error: Two lung cancers were visible on CT as nodules larger than 1 cm localised in the parenchyma of the lung. As radiological evaluation did not reveal any explanation for the failure to detect these two lung cancers, human error is considered to be the most plausible cause.

Interpretation of results

a) Sixty-one (0.85% of 7,155) Dutch participants randomised to the screening group of the NELSON trial, were diagnosed with a lung cancer that was not detected by screening. In 42.6% of the participants no abnormality suspicious for lung cancer could be identified on the last screening CT examination. Hence, these lung cancers arose after the last screening CT was made. In the remaining 57.4% of the participants, an abnormality suspicious of lung cancer was visible on the LDCT screening examination.

Missed carcinomas in CT-based lung cancer screening trials have received only limited attention in the radiological literature.[22] Moreover, not a single lung cancer

screening study re-evaluated screening CT examinations of the subjects with an interval or post-screen lung cancer to determine which lung cancers were missed and in which subjects the lung cancers developed after the last screening CT examination. As a result, the finding in this study that 57.4% of the interval and post-screen lung cancers were missed cannot be compared to any other study.

The results of this study may be extrapolated to populations with a risk of developing lung cancer that is comparable to the lung cancer risk of the NELSON population. As a previous study of the NELSON trial demonstrated that higher age and current smoking status of the participants were associated with a significantly increased risk of being diagnosed with an interval cancer,[23] the incidence of missed lung cancers may depend on these variables in different populations.

Concluding, about half of the lung cancers (57.4%) not diagnosed through screening was, in retrospect, present at the last LDCT screening examination.

b) In the 35 participants with an interval or post-screen lung cancer lung cancer an abnormality suspicious for lung cancer was visible on the last LDCT screening examination. Radiological re-evaluation showed that 57.2% of the lesions suspicious of lung cancer were not found on the LDCT examination by the radiologist due to detection error or human error. In 5.7%, the lesions suspicious for lung cancer were detected, but the lung cancers were not diagnosed due to interpretation errors by the radiologist. In 37.1%, the lesions suspicious of lung cancer were detected, but the lung cancers were not diagnosed due to: failure of the protocol (5.7%), non-compliance with the protocol by the participant (5.7%), and non-compliance with the protocol by the radiologist (8.6%).

In the literature, detection errors or interpretation errors by the radiologist are reported as common causes of missed lung cancers.[22,24-28] In contrary, no studies indicating the incidence of missed lung cancers on CT due to nodule management protocol failure or non-compliance to the protocol by either the radiologist or the patient have been published. Despite this, the incidence of missed lung cancers due to protocol failure might be somewhat lower in clinical practice. This is because the most commonly used guidelines (based on the criteria of the Fleischner Society) recommend serial follow-up CT examinations for any non-calcified pulmonary nodule of 4mm or more in diameter. Hence, more nodules are followed up and probably less lung cancers are missed. However, the adherence to guidelines has been proven to be moderate,[29] which may result in more missed lung cancers in clinical practice than in a clinical trial.

Concluding, detection errors were the most common cause of the failure to detect interval and post-screen lung cancers that were, in retrospect, visible at the last screening CT examination. Less common causes of missed lung cancers were

interpretation errors, failure of the screening protocol, and non-adherence to the protocol by either the radiologist or the participant.

c) In the 22 of the 61 participants a suspicious abnormality was in retrospect visible on the last LDCT screening examination, but lung cancer was not diagnosed due to interpretation detection or radiological errors. There were a number of typical lung cancer characteristics that may have caused or contributed to the failure to detect them.

The most common characteristic was endobronchial tumour localisation (22.7%, n=5 all detection errors), which was also the most common cause in a study by White et al.[24] There are two possible explanations for this: firstly, endobronchial nodule localisation is far less common than intraparenchymal localisation. As a result, the attention of the radiologist is primarily focused on the lungs and not the bronchi which facilitates the occurrence of detection errors. Secondly, the computed-aided detection system that is used to find nodules the radiologist misses, cannot search for nodules in the bronchial tree.[30] To reduce the occurrence of missed lung cancers, it is recommended to screening radiologists to force themselves to check the bronchial tree visually after the regular reading using the computed-aided detection system.

The other most common characteristic was bulla wall thickening (22.7%, one interpretation error and four detection errors). Bulla wall thickening was not classified as an important abnormality in the protocol of the NELSON trial. Hence, this characteristic was not reported unless the radiologist recognised it as suspicious for lung cancer. As a result, it is unknown what the incidence of bulla wall thickening is in the whole screened population and in the participants with screen-detected lung cancer in particular. Nonetheless, it is possible that the incidence of a localisation in a bulla wall is higher in missed lung cancers than in detected lung cancer as no specific attention was paid to bulla walls in the NELSON trial. There are no published studies that report on the incidence of this characteristic in lung cancers missed at CT examinations. However, bulla wall thickening was reported to be not uncommon in lung cancers that were detected by CT: in the Early Lung Cancer Action Project, 2% of the lung cancers detected at the first screening round and 12% of those detected in the second screening round were associated with cystic airspaces[31]; and in a clinical series of 545 lung cancer patients 3.5% of the cancers were associated with a bulla.[25] Since no data on the prevalence of bulla wall thickening in the screened population is available, the positive predictive value of this characteristic for lung cancer is unknown. Nonetheless, the prevalence of this characteristic of 22.7% (of the lung cancers missed due to radiological errors) justifies the recommendation to pay more attention to focal or diffuse bulla wall thickening in lung cancer screening.

The other most common characteristic of missed lung cancer due to detection and interpretation errors was pleural attachment of the nodule (22.7%, n=5, four detection errors and one interpretation error). Malignant nodules that are pleural-attached may be more difficult to detect and interpret than intraparenchymal localised malignant lung cancers as the computer aided-detection system is less sensitive for pleural attached nodules, and the size of pleural-attached nodules cannot be assessed by the volumetric software used in the NELSON trial.[8] Moreover, the NELSON publication by Xu et al. wherein in a sub-selection of 891 nodules was determined that none of the screen-detected lung cancers were attached to the pleura[32], may have induced a decreased alertness of the NELSON radiologists to pleural-attached lesions. Missed lung cancer in pleural-attached lesions (22.7%) was relatively more common in the NELSON trial than in to other studies: 6.3%[26], 6.7%24, 11.1%28 and 14.3%.[27] This may indicate that screening radiologists should consider pleural-attached nodules as potentially malignant despite the fact that no association between this characteristic and lung cancer could be established in the study by Xu et al.[32]

A less common presentation of missed lung cancers was lymphadenopathy without visible lung lesions (13.6%, n=3). This type of missed lung cancer is probably difficult to prevent as the radiological evaluation of the screening CTs is focussed on the lungs. Moreover, the low-dose scans are performed without the administration of intravenous contrast, which is not optimal for imaging the mediastinum.

One other missed lung cancer was also not visible as a lung lesion, but presented as pleural effusion at the screening CT examination (4.5%). Also in this case the effusion was probably overlooked because the screening radiologists focussed on intrapulmonary abnormalities.

Detection errors due to other pathology were uncommon in this study, only one lung cancer was not detected because it was surrounded by extensive reticulation (4.5%). Missed lung cancers due to other distractive pathology on the CT is also reported in several clinical series.[22,26,27]

No plausible explanation for the failure to detect the two remaining missed lung cancers of this study could be identified. As both nodules were larger than 1 cm in diameter and no specific characteristics or distractions were present, these nodules should have been detected. Nonetheless, human error will be difficult to prevent.

Some specific nodule characteristics that were reported to be associated with missed lung cancers were not observed in the NELSON trial; smaller nodule size,[22,26,28] peripheral nodule localisation[22], sub-solid nodule type[26] and attachment to a vessel.[27]

Concluding, this study identified several radiological characteristics that were related to the failure to detect or interpret interval and post-screening lung cancers.

Some of the causes of the detection and interpretation errors were probably not preventable, such as lung cancer presenting as extra-pulmonary abnormalities, distractive other pathology and human error. However, three different causes may presented opportunities to reduce the number of missed lung cancers: endobronchial tumours, tumours arising from thickenings in bulla walls and pleural-attached nodules. The design of this study does not allow for the calculation of the positive predictive value of these characteristics. To determine this, the incidence of these characteristics should be determined both in the detected lung cancers and representative sample of the screened population without lung cancer. If such a study would demonstrate that endobronchial tumours, tumours arising from thickenings in bulla walls and pleural-attached nodules are relevant risk-factors for (interval) lung cancer, the recommendation to check the bronchial tree visually, pay more attention to bulla wall thickening and consider also pleural-attached nodules as potentially malignant should be included in guidelines for lung cancer screening.

Conclusion

The majority of the lung cancers not detected by low-dose CT screening were not preventable. The performance of the screening strategy may be improved by reducing the number of detection and interpretation errors. This may achieved by increasing the radiologist's attention for endobronchial lesions, bulla wall thickenings, and pleural-attached lesions.

Research question V

Chapter 6. Optimisation of screening protocols

Lung cancer probability in subjects with CT-detected pulmonary nodules: an analysis of data from the NELSON trial of low-dose CT screening.

Lancet Oncology

Main research question

How should a participant's predicted lung cancer probability, based on size and growth of CT-detected nodules, be used to optimise the nodule management protocol of the NELSON trial?

Sub research questions

a) Was it valid to predict the two-year lung cancer probability of an individual who underwent screening using low-dose computed tomography, using a model based on nodule size and growth rate?

b) What was the probability of lung cancer in an individual who underwent screening using low-dose computed tomography, based on nodule size and growth rate?
c) How should the current thresholds for nodule size and growth rate be adjusted to improve risk stratification, test characteristics and reduce harms?

Main findings

For this study all 7,155 Dutch participants randomised to the CT screening arm of the NELSON trial were included. The participants with lung cancer diagnosed between screening rounds (interval cancers) and the participants with lung cancer diagnosed after screening (post-screen cancers) were identified via linkages with the Dutch Cancer Registry. The Belgian participants (n=935) were excluded as a linkage with the Belgian Cancer Registry was not yet possible.

a) This study aimed to improve the management of CT-detected pulmonary nodules by designing improved management protocol based on lung cancer probability. As nodule size and growth rate are reported to be the most important predictors of lung cancer probability, these determinants were chosen as the base for the new protocols. The first step in the design of the new protocols was to determine whether nodule size (volume or diameter) and nodule growth rate (volume-doubling time) were valid predictors of lung cancer probability in our dataset.

   For this, logistic regression analysis was performed to predict lung cancer risk in the two years following each screening round, using diameter, volume and VDT as potential predictor variables. The model only included participants whose largest nodule measured 50-500 mm^3 and who had ≥1 growing nodule in this volume range, because the VDT was available only for this subgroup.

   The model estimated that nodule volume, nodule diameter and nodule VDT were significant lung cancer predictors (all $p<0.001$). However, nodule volume was a stronger predictor of lung cancer than nodule diameter; if nodule volume was included in the model, nodule diameter was not a significant predictor anymore.

b) The two-year lung cancer probability for all included participants was 1.3% (95% CI 1.2-1.5%). On the CT examinations of 54.4% of the participants no pulmonary nodules were detected. Their probability to be diagnosed with lung cancer over the next two years was only 0.4%.

   The probability to be diagnosed with lung cancer in the two years following the screening examination was low for subjects with small nodules: for nodule with a volume <100 mm^3 0.5-0.7%, and for nodule with a diameter <5 mm 0.3-0.6%. Moreover, these probabilities were not significantly different from the lung cancer probability of subjects without nodules.

   The two-year lung cancer probability was intermediate for subjects whose nodules had a volume of 100-300 mm^3 or a diameter of 5-10 mm, as the associated risks were

respectively 1.5-5.8% and 0.9-2.9%; which was significantly increased compared to the probability of subjects without nodules.

Lung cancer probability in the two years following the screening examination was high for subjects whose nodules measured ≥300 mm^3 or >10 mm; 8.9-25.7% and 11.1-31.6% respectively. These probabilities were also significantly higher than the probability of subjects without any nodules.

The lung cancer probability according to nodule volume doubling-time (VDT) was calculated for the subjects whose largest nodule measured 50-500mm^3. Subjects with slowly-growing nodules (VDT ≥600 days), or nodules that were stable in size, or nodules that had shrunken or resolved, had a low probability of lung cancer (0.0-1.0%). Lung cancer probability was significantly increased for subjects with nodules with a VDT <600 days. Hence, subjects whose nodules had a VDT of 400-600 days were at intermediate risk (4.1% in two years), and subjects whose nodules had a VDT <400 days were at high risk (6.7-25.0% in two years).

Finally, both nodule volume and nodule VDT were used to estimate the two-year lung cancer probability. In subjects with large nodules of ≥300 mm^3, lung cancer probability remained substantial (from 5·9% to >50%) even in case of slow nodule growth. In subjects with intermediate-sized nodules (volume 50-300 mm^3), the lung cancer probability ranged from low (<1·5%) to high (30%), for VDTs ranging from <50 days to 600 days.

c) The current guideline[33] for the management of CT-detected pulmonary nodules is based on the criteria of the Fleischner Society.[5] To be able to estimate the performance of the current guideline, it was simulated as follows:

I) Subjects whose nodules measured ≤4 mm were classified as negative. In these subjects the next screening examination was made after one year, which is in accordance with the guideline which recommends follow-up CT at twelve months.[33]

II) Subjects whose nodules measured 4-8 mm were classified as indeterminate. In these subjects a follow-up CT is made after three months, while the guideline recommends performing a follow-up CT at 6-12 and 18-24 months. The final result is classified as positive for VDTs <400 days, and negative for VDTs ≥400 days, in accordance with the guideline.[33]

III) Subjects whose nodules measured >8 mm were classified as positive. In these subjects additional diagnostic procedures were performed, which is in accordance with the guideline.[33]

The simulated ACCP protocol yielded a sensitivity of 90.9% (95% CI 81.2-96.1%), and a specificity of 87.2% (95% CI 86.4-87.9%). The predictive value of a positive test result (PPV) was 6.2% (95% CI 4.8-7.9%), which led to follow-up CT examinations in 29.8% of the screened subjects, and additional diagnostic procedures in 13.6%.

Next, a nodule management protocol was designed using thresholds for nodule diameter that were based on the lung cancer probability:

I) Subjects whose nodules measured ≤5 mm were classified as negative. In these subjects the next screening examination was made after one year.

II) Subjects whose nodules measured 5-10 mm were classified as indeterminate. In these subjects a follow-up CT was made after three months, to assess nodule VDT. The final result was classified as positive for VDTs <600 days, and negative for VDTs ≥600 days.

III) Subjects whose nodules measured >10 mm were classified as positive. In these subjects additional diagnostic procedures were performed.

This protocol yielded a sensitivity of 92.4% (95% CI 83.1-97.1%) and a specificity of 90·0% (95% CI 89.3-90.7%). The PPV of this protocol was 7.9% (95% CI 6.2-10.1%), which led to follow-up CT examinations in 22.2% of the screened individuals and to additional diagnostic procedures in 10.8%.

Finally, a nodule management protocol was designed using thresholds for nodule volume that were based on the lung cancer probability:

I) Subjects whose nodules measured ≤100 mm^3 were classified as negative. In these subjects the next screening examination was made after one year.

II) Subjects whose nodules measured 100-300 mm^3 were classified as indeterminate. In these subjects a follow-up CT was made after three months, to assess nodule VDT. The final result was classified as positive for VDTs <600 days, and negative for VDTs ≥600 days.

III) Subjects whose nodules measured >300 mm^3 were classified as positive. In these subjects additional diagnostic procedures were performed.

This protocol yielded a sensitivity of 90.9% (95% CI 81.2-96.1%) and a specificity of 94.9% (95% CI 94.4-95.4%). The PPV of this protocol was 14.4% (95% CI 11.3-18.1%), which led to follow-up CT examinations in 7.8% of the screened individuals and to additional diagnostic procedures in 5.9%.

Interpretation of results

a) A logistic regression model was used to determine whether nodule size (volume or diameter) and nodule growth rate (volume-doubling time) were valid predictors of lung cancer probability in our dataset. Analyses showed that nodule volume, nodule diameter and nodule VDT were significant lung cancer predictors. However, nodule volume was a stronger predictor of lung cancer than nodule diameter.

Concluding, nodule volume, diameter and volume-doubling time are strong predictors of the lung cancer probability of subjects with CT-detected nodules. Therefore, it is valid to use these determinants for designing new nodule management protocols. The performance of protocols using nodule diameter is expected to be

worse than the performance of protocols using nodule volume, as the relationship between nodule diameter and lung cancer probability is weaker than the relationship between nodule volume and lung cancer probability.

b) In more than half of the included participants (54.4%) no pulmonary nodules were detected. Their two-year lung cancer probability of only 0.4% suggests that it may be safe to apply a screening interval of at least two years in these individuals. In the subjects with CT-detected nodules, lung cancer probability depended strongly on nodule volume and VDT.

This study confirms that the lung cancer probability of small nodules (volume <50 mm^3 or diameter <4 mm) is low; at ≤0.6% compared to <1% reported in the literature.2, 6, 33-36 Moreover, the lung cancer probability in subjects whose nodules measure 50-100 mm^3 or 4-5 mm, is also low (0.3-0.7%) and not significantly different from that in subjects without nodules. Currently, guidelines recommend two to four follow-up scans for such nodules.[5,33,37] Omitting these CT surveillance schedules should be considered, as the risk of malignancy does not justify the harms of ionizing radiation, psychological distress, and the associated costs.[38-40]

Next, subjects with intermediate-sized nodules (volume 100-300 mm^3 or diameter 5-10 mm) have a significantly higher two-year lung cancer risk (0.9-5.8%) compared to subjects without nodules. This justifies additional CT examinations, which is in accordance with current guidelines.[5,33,37] Since the lung cancer risk of subjects with nodules between 5 and 8 mm is comparable (0.9-1.8%),[6] a uniform CT surveillance schedule may be applied.

For intermediate-sized nodules, the VDT assessed at CT surveillance should be used to re-assess the lung cancer probability. Subjects with slowly-growing (VDT ≥600 days), stable, shrunken and resolved nodules are at low risk of lung cancer (0.0-1.0%) and could withdraw from intensified CT surveillance and return to regular screening. By contrast, subjects whose nodules have a VDT <600 days have a significantly increased risk of lung cancer (4.1-25.0%), which justifies intensified CT surveillance and additional diagnostic procedures. Subjects whose nodules have a VDT of 400-600 days may be regarded as at intermediate risk, since their lung cancer probability is 4.1% in two years. Hence, a follow-up CT examination at short notice to re-assess nodule size and growth may be preferred over more invasive or expensive diagnostic procedures.

Finally, this study confirmed the high lung cancer probability in subjects with large nodules (volume ≥300 mm^3 or diameter ≥10 mm): >10% in the literature[5,6,34,41] and 8.9-31.6% in this study. A remarkable finding was that the risk of these large nodules is also high (5.9-50%) when they grow slowly. Therefore, a follow-up CT examination to assess growth for nodules ≥300 mm^3 or ≥10 mm provides little

additional information, but may delay lung cancer diagnosis. Hence, immediate diagnostic work-up is suggested instead.

Concluding, the probability of lung cancer in subjects who underwent CT screening depends on the presence, size and growth rate of CT-detected pulmonary nodules. Subjects without nodules, subjects with small nodules (volume ≤100 mm^3 or diameter ≤5 mm), or with slow growing (VDT ≥600 days), stable, shrinking, resolving nodules that are smaller than 300 mm^3 or 10 mm, have a low lung cancer risk. Individuals with intermediate-sized nodules (volume 100-300 mm^3 or diameter 5-10 mm), or nodules smaller than 300 mm^3 or 10 mm with a VDT of 400-600 days, have an intermediate lung cancer risk. Subjects with large nodules (volume >300 mm^3 or diameter >10 mm), or with fast-growing nodules (VDT <400 days) are at high risk of developing lung cancer within two years.

c) In this study, three nodule management protocols were simulated and their performance evaluated; the ACCP guideline for nodule management[33] based on the Fleischner criteria,[5] a new lung cancer probability-based protocol using nodule diameter, and a new lung cancer probability-based protocol using nodule volume.

Comparing the new lung cancer probability-based protocol using diameter to the simulated ACCP protocol, the new protocol yielded a higher sensitivity (92.4% instead of 90.9%), fewer follow-up CT examinations (-7.6%), and additional diagnostic procedures (-2.8%), than the ACCP protocol. These results imply that the ACCP nodule management protocol performed well. However, with small adjustments of the thresholds for nodule diameter (raising the lower threshold from 4 mm to 5 mm, and raising the upper threshold from 8 mm to 10 mm) and the criterion for malignant nodule growth (VDT <600 days instead of VDT <400 days), the performance could be improved.

The new lung cancer probability-based protocol using volume classified no nodules and nodules <100 mm^3 as negative, nodules 100-300 mm^3 as indeterminate (final result positive if VDT <600 days, otherwise negative), and nodules >300 mm^3 as positive. This protocol demonstrated a very high specificity (94.9%), which yielded substantially less follow-up CT examinations (-22.0%), and additional diagnostic procedures (-7.7%) compared to the ACCP protocol. Moreover, the sensitivity of this volume protocol is as high as the ACCP protocol's sensitivity: 90.9%. Although this sensitivity is slightly lower than the sensitivity of the new protocol using diameter, the advantages of the very high specificity of the volume-based protocol outweighs the disadvantage of a slightly lower sensitivity. Therefore, the use of the new lung cancer probability-based protocol using nodule volume has the best performance of these three nodule management protocols.

Concluding, the current guideline for the management of CT-detected nodules[33] may be optimised by adjusting the thresholds for nodule size and growth. Hence,

the threshold that is used to classify nodules as at very low risk of lung cancer should be raised from 4 mm to 5 mm. The threshold that is used to classify nodule as at substantial risk of lung cancer, requiring more invasive diagnostic procedures, should be raised from 8 mm to 10 mm. Furthermore, for intermediate-sized nodules, more invasive procedures should only be performed for nodule demonstrating growth with VDT of <600 days, instead of visual growth or VDT <400 days. A more efficient alternative for these diameter protocols is the new lung cancer probability-based protocol using nodule volume. In this protocol, no nodules and nodules <100 mm^3 were classified as negative, nodules 100-300 mm^3 as indeterminate (final result positive if VDT<600 days, otherwise negative), and nodules >300 mm^3 were classified as positive. Prospective studies or micro-simulation studies using these new protocols are required to determine the effect on lung cancer mortality.

Conclusion

The size and growth of CT-detected nodules are valid predictors of the screened individual's two-year lung cancer probability. Subjects with nodules ≤100 mm^3 or ≤5 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 the assessment of VDT is appropriate (<600 days warrants follow-up evaluation). The risk of subjects with nodules >300 mm^3 or >10 mm demands immediate diagnostic evaluation. New management protocols for CT-detected nodules using these thresholds for nodule size and VDT were estimated to perform better than the current guideline.

Research question VI

Chapter 7. Evaluation of bronchoscopy

The role of conventional bronchoscopy in the work-up of suspicious CT screen-detected pulmonary nodules.

Chest

Main research question

What was the value of bronchoscopy for diagnosing lung cancer in screen-detected nodules?

Sub research questions

a) What were the test characteristics of bronchoscopy and its ancillary procedures?

b) What were predictors for a true-positive bronchoscopic procedure?
c) Which diagnoses were made in false-negative bronchoscopic procedures?

Main results

In a series of 415 participants with a positive screening result in the NELSON trial, 308 (74.2%) underwent a conventional white-light bronchoscopy as part of the diagnostic work-up. In these 308 persons, a total of 318 suspicious pulmonary nodules or masses were analysed by bronchoscopy with the objective to diagnose or exclude lung cancer. According to the gold standard, which was a histological confirmation on histological or cytological samples in 95.5%, 178 of the 318 (56.0%) suspicious lesions were malignant.

a) Cancer was diagnosed by bronchoscopy in only 24 of the 318 suspicious lesions. The overall sensitivity of bronchoscopy to detect cancer was 13.5% (24 of 178; 95% confidence interval (95% CI) 9.0-19.6%), and the negative predictive value was 47.6% (140 of 294, 95% CI 41.8-53.5%). As no false-positive diagnoses were made, the specificity and positive predictive value were both 100%. The sensitivity and negative predictive value of the ancillary procedure bronchial washing were respectively 9.3% (95% CI 5.7-14.8) and 45.9% (95% CI 40.2-51.7%), for bronchial brushing respectively 7.9% (95% CI 3.3-17.0%) and 41.2% (95% CI 32.4-50.6%). The sensitivity and negative predictive value for the ancillary procedures transbronchial needle aspiration (TBNA), transbronchial biopsy (TBB) and endobronchial biopsy (EBB) were respectively TBNA: 33.3% (95% CI 6.0-75.9%) and 0% (95% CI 0-6.0%); TBB: 16.7% (95% CI 0.9-63.5%) and 44.4% (95% CI 15.3-77.3%); EBB: 45.8% (95% CI 26.2-66.8%) and 55.2% (95% CI 36.0-73.0%).
b) Multivariate regression analyses were performed on the subset of 178 cases with cancer to identify predictors of a successful bronchoscopic procedure. Nodule size, defined as nodule diameter in mm, was a statistically significant positive predictor (odds ratio 1.07 (95% CI 1.02-1.13). Further, the visibility of the lesion during bronchoscopy was also a statistically significant positive predictor (87.61 (95% CI 4.90-564.88). The following characteristics were significantly associated with a successful bronchoscopic procedure: nodule volume doubling time, central versus peripheral localisation, upper lobe localisation versus lower or middle lobe localisation, screening round, scan type, presence or absence of bronchial compression.
c) In 24 of the 178 bronchoscopies of malignant lesions, the result of the bronchoscopy was positive. Hence, in 154 bronchoscopic evaluations the result was false-negative, which corresponds with a negative predictive value of 47.6%. The diagnoses made based on bronchoscopy in these false-negative procedures were: aspecific inflammation (27.3%), metaplasia (3.9%), fibrosis (1.3%), Aspergillus Fumigatus infection (0.6%), atypia (0.6%) and resolving haemorrhage (0.6%), in the remaining cases (65.6%) no abnormalities were detected.

Interpretation of results

a) The overall sensitivity and negative predictive value of bronchoscopy in subjects with suspicious CT-detected pulmonary nodules was respectively 13.5% and 47.6%. As no false-positive diagnoses of cancer were made by bronchoscopy, the sensitivity and positive predictive value were both 100%.

To date, the have been no other publications on the diagnostic performance of conventional white light bronchoscopy for diagnosing lung cancer in subjects with CT-detected nodules. This is probably because bronchoscopy is not routinely performed in other lung cancer screening trials.

In non-screening studies, the published sensitivity of bronchoscopy varied from 51% to 76%.[42-47] Differences between these and the current study were: a smaller size of the lesions (2.8% >30 mm versus 48-72 mm)[42,43,46,47] and a lower incidence of endobronchial abnormalities (7.3% versus 8-64%).[43-45] The lower sensitivity in the current study can be explained by these differences as both nodule size and endobronchial visibility are independent predictors for a successful bronchoscopy procedure.[48]

According to the guidelines that were available at the time this study was conducted, conventional white light bronchoscopy was only recommended for suspicious lesions with an air bronchogram on CT.[49,50] According to the currently available guideline,[33] conventional white light bronchoscopy is not recommended anymore. Only more advanced techniques such as radial endobronchial ultrasound, electromagnetic navigational bronchoscopy and virtual bronchoscopy navigation techniques are recommended for individuals who are poor candidates for transthoracic biopsy in case the lesion is located in proximity to a patent bronchus.

Concluding, this study demonstrates that the routine use of conventional bronchoscopy in the diagnostic work-up of CT-detected pulmonary nodules is not justified. It is highly unlikely that the yield outweighs the harms (distress and risk of complications) and costs (health care facilities, personnel and resources) associated with routine use of bronchoscopy.[33] However, to determine this cost-effectiveness analyses including all these aspects should be performed. The use of conventional white light bronchoscopy in selected cases (larger nodules located in proximity to a patent bronchus) will result in better test characteristics, which will yield a higher cost-effectiveness. Nonetheless, this can also not be recommended as conventional white light bronchoscopy has become an outdated technique for diagnosis lung cancer in suspicious CT-detected nodules.

b) The size and visibility of the suspicious lesions were statistically significant predictors of a true-positive result of the bronchoscopic procedure. It was estimated that for every millimetre increase in nodule diameter, the probability of a true-positive procedure will increase with 7%. Further, lesions which were visible during bron-

choscopy had a significantly higher probability of a successful procedure compare to lesions not visible during bronchoscopy.

Concluding, size and visibility of suspicious CT-detected lesions are predictive of a true-positive bronchoscopic procedure.

c) A variety of diagnoses is made in false-negative bronchoscopic procedures. Instead of lung cancer, Aspergillus infection, aspecific inflammation, atypia, metaplasia, fibrosis and resolving haemorrhage are found in the histological or cytological samples obtained by bronchoscopy. Some of these findings, as infections or fibrosis, may also present as an approximately spherical opacity on CT. This demonstrates that lung cancer can be missed by bronchoscopy in the presence of benign abnormalities that could also explain the suspicious lesion opacity on CT. Such a situation bares the risk of a missed or delayed lung cancer diagnosis. This information, combined with the estimated negative predictive value of 47.6%, proofs that bronchoscopy is not an appropriate technique to exclude a diagnosis of lung cancer in subjects referred for screen-detected nodules.

Concluding, deceitful benign diagnoses can be made by bronchoscopy in persons wherein the suspicious lesion is actually lung cancer. Therefore, the use of bronchoscopy to exclude a diagnosis of lung cancer is not recommended in a lung cancer screening program.

Conclusion

The performance of white light bronchoscopy is not sufficient to justify routine use in subjects with suspicious pulmonary nodules detected in a lung cancer screening programme.

Research question VII

Chapter 8. Evaluation of surgical procedures

Complications following lung surgery in the Dutch-Belgian randomized lung cancer screening trial.

European Journal of Cardio-Thoracic Surgery

Main research question

To what extent did adverse events related to thoracic surgery, occur in participants after a positive screening test results?

Sub research questions

a) How often occurred re-thoracotomy, complications, and post-operative mortality in participants who underwent thoracic surgery for a positive screening test result?
b) What was the length of hospital stay for lung resection performed by thoracotomy and video-assisted thoracoscopic surgery?
c) To what extent were surgical procedures performed for benign nodules?

Main results

In a series of 415 participants with a positive screening result in the NELSON trial, 215 (51.8%) underwent a surgical procedure. Seventeen of these 215 participants (7.9%) only underwent a mediastinoscopy. The remaining 198 participants underwent lung surgery; in 44 (22.2%) lung surgery was preceded by a mediastinoscopy. The majority (n = 182; 91.9%) of the lung surgeries were resections performed via a thoracotomy. 5.6% (n = 11) of the procedures were wedge resections performed via a video-assisted thoracoscopic (VATS) procedure. The remaining 2.5% (n = 5) of the procedures were initiated as VATS procedures, but were converted to a thoracotomy. Summarising, in 198 subjects 187 thoracotomies and 16 VATS procedures were performed.

a) 47% (n = 88) of the thoracotomies were complicated by at least one non-life threatening condition and 10% (n = 18) was complicated by at least one life threatening condition. In 38% (n = 6) of the VATS procedures at least one non-life threatening complication occurred, but no life threatening complications have been observed. As 5% had both minor and major complications, the proportion of participants with any complication was 53%. The complications caused by thoracotomy necessitated a re-thoracotomy in 3% and re-admission to the hospital after discharge in 5%. After VATS procedures, no re-thoracotomies or re-admissions occurred. There was no mortality within the first 30 days after thoracotomy or VATS in the NELSON trial.
b) The median length of hospital stay after thoracotomy was 13 days (range 2 to 51 days). After a VATS procedure the median length of hospital stay was 8 days (range 4 to 12 days). In subjects with non-life threatening complications, the length of hospital stay after thoracotomy was a median 15 days, ranging from 6 to 51 days and after VATS 9 days (range 7 to 12 days). In the case of life threatening complications following thoracotomy, the median length of hospital stay was 21 days, ranging from 8 to 51 days.
c) The resection specimens obtained by the 198 surgical procedures yielded the diagnosis of lung cancer in 151 (76.3%) cases. Hence, in the remaining 47 cases (23.7%) benign abnormalities were resected. Twenty of the 47 subjects (42.6%) who underwent surgery for benign disease experienced non-life threatening complications and three of the 47 subjects (6.4%) had life threatening complications.

Interpretation of results

a) In this study, the incidence of re-thoracotomy, complications, and mortality after lung surgery in the NELSON lung cancer screening trial was assessed. A few other lung cancer screening trials published their adverse events. One or more complications after surgery occurred in the NELSON trial in 53.% of the participants, which is higher than in all other screening trials: 45.0% of the participants of the U.S. National Lung Screening Trial (NLST)[13]; 33.9% of the participants of the Italian DANTE trial[51]; 25% of the participants of the Italian COSMOS lung cancer screening cohort study[52]; and 0% of the participants of the Danish lung cancer screening trial.[9] However, in the incidence of major complications,10% in the NELSON trial, was within the range of the other screening trials: 15.3% in the DANTE trial[51]; 11.9% in the NLST[13]; 6% in the COSMOS trial[52]; and 0% in the Danish trial.[9] Post-operative mortality was 0% in the NELSON trial, which is comparable to the NLST (0.01% within 60 days)[13], the Danish trial (0%)[9] and the COSMOS trial (0%)[52], only in the DANTE trial a higher mortality rate (5.1%) was published.[51]

The incidence of non-life threatening complications after thoracotomy in NELSON (47%) was also high in the range of the incidences of published non-screening series (7% to 57%).[51,53-68] However, the incidence of life threatening complications after thoracotomy in the NELSON trial (10%) was low in the range (4% to 26%) of non-screening studies.[51,53-68] The incidence of non-life threatening complications after VATS procedures was 38% in the NELSON trial, which is high compared to the range in the literature (9% to 51%).[69,70] No life threatening complications have been observed after VATS in the NELSON trial, which is at the lower range of the reported incidence in the literature (0% to 12%).[51,70,71] In the NELSON trial, complications after thoracotomy necessitated a re-thoracotomy in 3%. The reported re-thoracotomy rates after a thoracotomy varied from 0 to 9%.[59,72] No re-thoracotomies after VATS were performed in the NELSON trial, while the reported re-operation rate after VATS varied between 1 and 5%.[62,72] Finally, no post-operative mortality after respectively thoracotomy and VATS were observed in the NELSON trial, compared to mortality rates of respectively 0-8%[51,53-61,63-68,72-82] and 0-4%[58,83] after thoracotomy and VATS in other studies.

The aforementioned comparisons with other screening studies and clinical series can be made as this study also demonstrated that the age range and co-morbidity level were comparable.[84] Nonetheless, the studies were quite heterogenic with respect to the definition, classification and methods of data collection on complications. Not in all studies, a distinction was made between life threatening and non-life threatening complications. Moreover, some studies collected the data by reviewing individual patient charts, and others based on ICD-9 codes or on claims in Medicare files. The latter two methods result in an underestimation of complica-

tions, especially for minor complications. Probably because all individual patient files were carefully evaluated in the current study, the minor complication rate was relatively high. As reporting major complications and mortality occurs more accurate than reporting minor complications, the comparisons concerning major complications and post-operative mortality between this study and the literature is more reliable. Hence, both the incidence of major complications and post-operative mortality were relatively low compared to clinical series, and comparable to other screening trials. This could probably be explained by the fact that screen-detected cancers, in general, are diagnosed earlier than symptom-detected lung cancers, and as a result the required resection of screen-detected lung cancer are more often less extensive.[13,19,85] This is supported by the observation that pneumonectomies were rarely performed in the NELSON trial, while more complex resections with higher expected complication rates were performed in the published clinical studies.

Concluding, post-operative minor complications (47%) were more frequent in the NELSON than reported in the literature. The incidence of major complications (10%), re-thoracotomy (3%) and post-operative mortality (0%) were at the lower range of the reported incidences in other studies. This suggests that lung surgery for lung cancer detected by low-dose computed tomography is at least as safe as lung surgery for clinically detected lung cancer. Although the design of the current study does not allow drawing conclusions on comparisons between thoracotomy and VATS; the incidence of complications, re-operations and post-operative mortality were lower after VATS procedures. As this was also observed in other published studies, VATS may be a safer method for lung resections for screen-detected lung cancers, and should probably be the preferred method for those cases wherein both VATS and thoracotomy are appropriate.

b) The median length of hospital stay in after thoracotomy in the NELSON trial (13 days) was in the middle of the range reported in the literature (5 to 22 days).[72,80] In subjects with non-life threatening complications, the length of hospital stay after thoracotomy was a slightly longer (median 15 days), and in the case of life threatening complications substantially longer (median 21 days). After a VATS procedure, the median length of hospital stay in the NELSON trial was 8 days, which is in the lower range of lengths of stay reported in the literature 4 to 23 days.[57,72] In subjects with non-life threatening complications, the length of hospital stay after VATS was slightly longer (median 9 days).

Concluding, the length of hospital stay (median 13 days after thoracotomy and 8 days after VATS) in the NELSON trial was comparable to the literature. Although the design of the current study does not allow drawing conclusions on comparisons between thoracotomy and VATS; the length of hospital stay was lower after VATS procedures, both in this study and in other published studies. Therefore, VATS may

be considered as the preferred method for lung resections for screen-detected lung cancers, for those cases wherein both methods are appropriate.

c) The 198 surgical procedures included in this study yielded the diagnosis of lung cancer in 151 (76.3%) cases. Hence, in the remaining 47 cases (23.7%) benign abnormalities were resected. Most other lung cancer screening trials reported on the percentage of surgeries for benign disease and the incidence in NELSON is at the higher range: NLST 32.2% surgeries in the absence of lung cancer,[13] in the DANTE trial 22% of the resected nodules was benign,[51] in the Danish trial this was 18.2%,[9] in the Italian MILD trial 9%[86] and in the ITALUNG trial only 5.5%.[87] The high percentage of surgeries of benign nodules in the NLST may be partly due to the vast proportion of the participants (39.1%) who received one or more positive screening result,[13] due to the low specificity (73.4-83.9%) of the NLST screening algorithm.[7,14] In the NELSON trial however, the specificity of the screening algorithm is much higher (98.6%)[23] and the proportion of subjects with a positive screening result much lower (6.0%).[88] This suggests that the relatively high number of surgeries for benign nodules are not caused by the screening algorithm, but by decisions made during the diagnostic work-up for positive screening results. One aspect that might play a role is the fact that the diagnostic work-up in the NELSON trial usually only consisted of imaging and bronchoscopy. As a result, proof of the suspicion of lung cancer by biopsy was rarely obtained before surgery. The lower numbers of surgeries for benign nodules in the other screening trials suggests that they applied a more cautious approach towards suspicious CT-detected nodules. Since these studies did not publish on the counter side of this approach (number of additional diagnostic tests, months of follow-up and cancerous nodules unjustly not resected), it is not possible to determine whether their approach is recommendable. Nonetheless, efforts should be made to investigate whether it is possible to safely reduce the number of surgeries for benign nodules.

Another reason for this is the occurrence of complications in this group. Hence, 20 of the 47 subjects (42.6%) who underwent surgery for benign disease, non-life threatening complications occurred and in three of the 47 subjects (6.4%) life threatening complications occurred. In other lung cancer screening trials that published on such adverse events, the incidence of complications was lower: NLST 20.5% (from which 2.4% was minor, 7.9% was intermediate and 5.5% were major complications),[13] DANTE trial 0%[51] and COSMOS trial 0%.[52] Post-operative mortality did not occur after surgery for benign nodules in the NELSON trial, DANTE trial,[51] COSMOS trial,[52] and was rare in the NLST 1.2%.[13]

Concluding, 23.7% of the lung surgeries in the NELSON trial were performed for benign nodules. This percentage, as well as the incidence of complications (42.6%) is

relatively high compared to other screening trials. Future studies should investigate how to safely reduce the number of surgeries for benign nodules.

Conclusion

This study demonstrated that adverse events after thoracic surgery for positive screening test results were common. The incidence of minor complications was relatively high, while in the incidence of major complications, re-operations and post-operative mortality was relatively low. Finally, a relatively high percentage of the surgeries was performed for benign nodules.

Research question VIII

Chapter 9. Endpoint determination

Uniform and blinded cause of death verification in a lung cancer CT screening trial.

Lung Cancer

Main research question

How should the endpoint verification process of the NELSON trial be designed to ensure uniform, objective and unbiased endpoint determination?

Sub research questions

a) How to develop a cause of death review protocol that ensures uniform, objective and unbiased endpoint determination?
b) How was the performance of the developed cause of death protocol compared to the official death certificates?
c) What were sources of disagreement between users of the developed cause of death protocol?
d) What were the best sources of information for a review of the cause of death of a participant?

Main results

a) The primary endpoint of the NELSON trial is lung cancer-specific mortality. Information on the cause of death of the NELSON participants can be obtained from the death certificates, which are available from Statistics Netherlands and the Flemish Agency for Care and Health. Therefore, the first step in the design the endpoint verification process of the NELSON trial, was to perform a literature study on the

reliability of the use of official death certificates for endpoint verification in screening trials.

This initial study learned that the use of death certificates for this purpose is debated for several reasons: sticky-diagnosis bias and slippery linkage bias,[100] inaccurate form completion and errors in encoding,[101] and incorrect ante mortem diagnoses.[102] Further, the sensitivity and specificity of the official death certificates for (lung) cancer death have been reported to range respectively from 84.5% to 99.7% and from 91.3% to 99.7%.[103-106] Moreover, the errors introduced by all these aforementioned inaccuracies are biased towards a reduction in the efficacy of screening.[103-106]

To overcome these problems clinical expert committees, reviewing the medical files of the deceased participants to determine the cause of death, are frequently employed in cancer screening trials.[103-108] The additional value of such a clinical expert committee depends on its independence from the screening trial and the quality and uniformity of the review process. Therefore, predetermined criteria and flowcharts are often used for the evaluation of the medical files, which should be blinded for the participants' identity and study group.

The next part of this study was to define the principles of the cause of death review process that will be used in the NELSON trial. Firstly, the definition of the primary cause of death to be used was adopted from the definition of the World Health Organisation. Next, a classification system that defines the graduation of certainty that lung cancer was the primary cause of death was adopted from the European Randomised Study of Screening for Prostate Cancer.[104]

Secondly, the target population of the end point verification process was defined to be all participants of the NELSON trial that have ever been diagnosed with lung cancer. The lung cancer cases will be identified by linkages with the national cancer registries of the Netherlands and Belgium and by checking all official death certificates for the diagnosis lung cancer, which are obtained from Statistics Netherlands and the Flemish Agency for Care and Health. For all identified cases, the diagnosis of lung cancer will be verified in a separate verification process.

Thirdly, the data required for the endpoint verification process was defined to be the complete medical file from the first consultation or diagnostic test for (suspected) lung cancer, until death, including autopsy report if available.

Fourthly, the requirements for the clinical expert committee were defined to be: three independent experts; a pulmonologist–oncologist, a pathologist specialised in lung oncology and a clinical epidemiologist, who have never been employees of the NELSON trial.

Finally, the tools that the expert committee should use to classify the cause of death were designed; a flowchart and detailed list of criteria which classify the cause

of the death into one of the six categories defining the graduation of certainty that lung cancer was the primary cause of death.

b) To determine the performance of the aforementioned newly developed endpoint verification process compared to the official death certificates, a pilot study of fifty cases was conducted. When classifying the outcome of the cause of death review process as golden standard, the sensitivity and specificity of the death certificates were respectively 95.2% (95% confidence interval: 84.2-98.7%) and 62.5% (95% confidence interval: 30.6–86.3%). Disagreement was observed in 10% (5 of 50 individuals) with the following causes of death: adult respiratory distress syndrome after lobectomy, rupture of an abdominal aneurysm during chemotherapy, another malignancy besides lung cancer in two cases (breast carcinoma and acute myeloid leukaemia) and small cell lung carcinoma diagnosed after the person's death by autopsy.

c) The agreement between the uses of the newly developed endpoint verification process was also investigated in the pilot study. Hence, in 76% of the cases the reviewers reached a concordant conclusion. In the remaining cases, the sources of disagreement were: significant comorbidity, multiple coinciding malignancies, death after an intervention and death indirectly caused by lung cancer, such as death due to post-obstruction pneumonia or paraneoplastic pulmonary embolism. When clustering all 'definitely' and 'probable' lung cancer deaths into one group and all 'possible', 'unlikely' and 'definitely not' lung cancer deaths and 'intercurrent deaths' into another, the differences were minimal; agreement in 90% (kappa of 0.65).

d) Finally, the pilot study learned that the letters of the pulmonologist were the best source of information for the review of the cause of death in 65% of the cases.

Interpretation of results

a) Endpoint verification of a cancer screening trial should not solely be based on the official death certificate because of biases, inaccuracies in diagnosing, filling in forms and encoding, and suboptimal sensitivity and specificity for cancer-specific primary cause of death. Instead, endpoint verification should be based on a cause of death verification process that provides a validated method and tools used by a committee of independent experts. The following aspects should be defined in the protocol:
   I) the definition of the primary cause of death
   II) a classification system that defines the grade of certainty that the primary endpoint was the primary cause of death
   III) the target population and methods to identify them within the study population
   IV) the data required for the cause of death review process

V) requirements for the clinical expert committee
VI) tools to be used by the clinical expert committee
VII) method to evaluate or compare the cause of death verification process to the official death certificates

Concluding, a cause of death review protocol that ensures uniform, objective and unbiased endpoint determination should employ a committee of independent experts that use a protocol that defines the aforementioned criteria.

b) The sensitivity and specificity of the official death certificates for lung cancer specific mortality were 95.2% and 62.5%, respectively. Despite the lack of a 'gold standard' for the cause of death of lung cancer participants, this still demonstrates the limitations of the official cause of death certification in lung cancer patients for scientific purposes.

Concluding, the official death certificates probably have insufficient distinctive character for lung cancer-specific death for determining the primary endpoint of a cancer screening trial.

c) The agreement between the two users of the cause of death verification protocol was reasonable. Cases that resulted in disagreement between the two users in the pilot study had: significant comorbidity, multiple coinciding malignancies, intervention-related death or death indirectly caused by lung cancer. Cases with significant comorbidity or coinciding malignancies are well-known sources of disagreement,[103,109] and will probably often be discussed in the expert committee to reach consensus. The other sources of disagreement between these two users indicate a lack of knowledge on complications from lung cancer treatments (such as surgery and chemotherapy) and indirect causes of lung cancer death (such as post-obstruction pneumonia and paraneoplastic syndromes). This illustrates the necessity of the employment of experts in the committee.

Further, the pilot study learned that the voluntary use of a flowchart and pre-specified criteria as an aid in the decision-making process will not always result in the use of these tools. Therefore, it is recommended to make the use of the flowchart obligatory in the decision-making process. In the NELSON trial, this will be accomplished by applying an electronic questionnaire with mandatory questions which are directly derived from the flowchart. Once the cause of death is classified using this electronic questionnaire, the expert has the opportunity to indicate whether he agrees or disagrees with the conclusion and whether he wants to discuss the case with the other experts or not.

Concluding, when the developed cause of death protocol is used by a clinical expert committee the patients with significant comorbidity and multiple coinciding malignancies will probably yield disagreement. In such cases, meetings of the

experts should be conducted to facilitate the discussion of these cases and to reach consensus.

d) The item that was regarded as the best source of information for the cause of death review in the medical file was the letters of the pulmonologist. In the Netherlands and Belgium, the pulmonologist is the main caretaker of lung cancer patients. Therefore, the pulmonologist regularly writes letters to the other involved caretakers and the general practitioner of the patient, to inform them on the clinical findings, results from diagnostic procedures and recommendations from multidisciplinary meetings. Moreover, about half of the lung cancer patients die at the hospital, which is usually at the pulmonology department, which will result in an accurate documentation of the death of death of the patient in the letter of the pulmonologist. As autopsies are not commonly performed in the Netherlands and Belgium, this valuable report will rarely be available for the cause of death review process.

Concluding, the best source of information for the cause of death review process of a lung cancer screening trial are, in the Netherlands and Belgium, the letters of the pulmonologist.

Conclusion

To ensure uniform, objective and unbiased endpoint determination in the NELSON trial an independent committee of experts should perform a cause of death verification process. For this process, the medical files of all deceased study participants diagnosed with lung cancer should be blinded for the participant's identity and study group. These files should be reviewed using a flowchart and detailed criteria to determine the grade of certainty that lung cancer was the primary cause of death.

GENERAL CONCLUSIONS FROM THIS THESIS

I) The screening algorithm of the NELSON trial adequately stratified participants according to their lung cancer risk.

II) The screening algorithm of the NELSON trial yielded a relatively low number of diagnostic work-ups for positive screening test results.

III) The screening algorithm of the NELSON trial yielded a relatively limited number of follow-up LCDT scans for indeterminate screening test results.

IV) The positive predictive value of screening test results in the NELSON trial compared favourably to other studies, nonetheless false-positive screening test results are one of the most common harms of the NELSON screening algorithm.

V) The negative predictive value of screening test results was very high in the NELSON trial, as in other lung cancer screening studies.

VI) The sensitivity in the NELSON trial was slightly lower than in other studies, but the lung cancers in the screening group were diagnosed as early as in other studies.

VII) The specificity in the NELSON trial was substantially higher than in other studies, which is a prerequisite for cost-effectiveness.

VIII) The majority of the lung cancers not detected by low-dose CT screening were not preventable: some lung cancers were not missed but arose during the screening interval, and other lung cancers were missed due to causes that can never completely be eliminated, such as human error and non-compliance by participants. Preventable causes of detection failures were radiological detection and interpretation errors.

IX) The performance of the screening strategy may be improved by reducing the number of detection and interpretation errors by increasing the radiologist's attention for endobronchial lesions, bulla wall thickenings, and pleural-attached lesions.

X) The performance of the NELSON screening strategy may be improved by an effective smoking cessation program, as current smokers are at increased risk of being diagnosed with a lung cancer not detectable by screening.

XI) The performance of the NELSON screening strategy may be improved by using a nodule management protocols with thresholds for nodule size and growth based on the lung cancer probability of the screened individuals.

XII) The use of nodule volume in lung cancer probability-based nodule management protocols yields higher efficiency and less harm than the use of nodule diameter.

XIII) In the NELSON lung cancer screening trial, a relatively high number of invasive diagnostic procedures for benign nodules were performed.

XIV) The performance of white light bronchoscopy is not sufficient to justify routine use in subjects with suspicious pulmonary nodules detected in a lung cancer screening programme.

XV) In the NELSON lung cancer screening trial, minor complications after surgical procedures was common, while major complications, re-operations and post-operative mortality were relatively uncommon.

XVI) In the NELSON lung cancer screening trial, a relatively high percentage of the surgical procedures was performed for benign nodules.

XVII) Screening for lung cancer using low-dose computed tomography probably leads to a stage shift towards earlier diagnosis; this effect is stronger in women than in men.

XVIII) The screening strategy of the NELSON trial was capable of detecting lung cancer as early as in another screening trial that demonstrated a significant lung cancer mortality reduction.

XIX) Screening for lung cancer using low-dose computed tomography probably leads to a shift in histology towards detecting slower growing and more peripherally localised subtypes of lung cancer.

XX) The lung cancers not detected by screening had a different histopathology, with a higher growth rate and metastatic potential, than the lung cancers that were detected by screening.

XXI) The endpoint determination of lung cancer screening trials should encounter the grade of certainty that lung cancer was the primary cause of death of the participants.

XXII) The endpoint determination of lung cancer screening trials should be performed by an independent committee of experts who review the blinded medical file of all deceased study participants diagnosed with lung cancer by using a flowchart and detailed criteria.

GENERAL RECOMMENDATIONS BASED ON THIS THESIS

I) Future studies should investigate whether the harms induced by false-positive screening results can be reduced by: optimising the nodule management protocol and the use of additional determinants to determine the screening test result, such as participant characteristics, radiological and other biomarkers.

II) An effective smoking cessation program should be co-implemented with CT screening, as current smokers have demonstrated to be at increased risk of being diagnosed with a lung cancer not detectable by screening.

III) Methods to increase radiologist's attention for endobronchial lesions, bulla wall thickenings, and pleural-attached lesions should be developed, as this may help reducing the number of missed lung cancers.

IV) Lung cancer screening programs should use a nodule management protocol with thresholds for nodule size and growth based on lung cancer probability.

V) Individuals without pulmonary nodules at CT screening may undergo their next screening after a screening interval of two years.

VI) Subjects with pulmonary nodules measuring ≤100 mm^3 or ≤5 mm should not undergo serial follow-up CT examinations, but regular CT screening with annual or biannual intervals.

VII) In subjects with pulmonary nodules measuring 100-300 mm^3 or 5-10 mm the assessment of nodule volume-doubling time by follow-up CT at short notice is appropriate; volume doubling times <600 days warrant diagnostic evaluation.

VIII) Subjects with pulmonary nodules measuring >300 mm^3 or >10 mm should undergo immediate diagnostic evaluation to diagnose or exclude lung cancer.

IX) Lung cancer screening programs should use nodule volume for estimating nodule size and growth since it yields higher screening efficiency and less harms than the use of nodule diameter.

X) Future studies should assess the causes of the high rate of invasive procedures for benign nodules in the NELSON trial, and targeted interventions should be developed, evaluated and implemented.

XI) The lung cancer prediction model using nodule size and growth rate should be extended with individual characteristics to investigate whether lung cancer prediction can be improved. A validated and reliable lung cancer prediction tool may help identifying malignant nodules and may reduce unnecessary diagnostic procedures for benign nodules.

XII) Conventional white light bronchoscopy should not routinely be used in subjects with suspicious CT-detected pulmonary nodules, as the diagnostic yield is insufficient to outweigh harms and costs.

XIII) Lung cancer screening should be performed in hospitals offering minimal invasive thoracic surgery, to limit complications and post-operative mortality both in individuals undergoing surgery for lung cancer and for individuals undergoing surgery for benign nodules.

XIV) Future studies should compare all lung cancers diagnosed in the screening group with all lung cancers diagnosed in the control group to determine whether LDCT screening led to a shift in stage or histopathology.

XV) Analysis of the difference in lung cancer mortality between the screening group and the control group of the NELSON trial will determine whether the observed favourable stage distribution of the lung cancers in the screening group yielded a significant lung cancer mortality reduction.

XVI) A post hoc analysis of the effectiveness of LDCT screening stratified by gender should be performed, as women with screen-detected lung cancer have demonstrated to be diagnosed at more favourable cancer stages than men.

XVII) Endpoint determination in lung cancer screening trials should encounter the determination of the grade of certainty that lung cancer was the primary cause of death of the study participants.

XVIII) Endpoint determination in lung cancer screening trials should be performed by an independent committee of experts who perform a blinded review of the medical file, of all deceased study participants with lung cancer by using a flowchart and detailed criteria.

REFERENCES

1. Hansell DM, Bankier AA, MacMahon H, et al. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; 246(3):697-722.
2. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 354(9173):99-105.
3. Henschke CI, Naidich DP, Yankelevitz DF, et al. Early lung cancer action project: initial findings on repeat screenings. *Cancer* 2001; 92(1):153-9.
4. National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. *Radiology* 2011; 258(1):243-53.
5. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005; 237(2):395-400.
6. National Lung Screening Trial Research Team, Church TR, Black WC, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 2013; 368(21):1980-91.
7. Aberle DR, DeMello S, Berg CD, et al. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med* 2013; 369(10):920-31.
8. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer* 2006; 54(2):177-84.
9. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish Randomized Lung Cancer CT Screening Trial-Overall Design and Results of the Prevalence Round. *J Thorac Oncol* 2009; 4(5):608-14.
10. Baldwin DR, Duffy SW, Wald NJ, et al. UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. *Thorax* 2011; 66(4):308-13.
11. Becker N, Motsch E, Gross ML, et al. Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. *J Cancer Res Clin Oncol* 2012; 138(9):1475-86.
12. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; 67(4):296-301.
13. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365(5):395-409.
14. Horeweg N, Nackaerts K, Oudkerk M, et al. Low-dose computed tomography screening for lung cancer: results of the first screening round. *J Comp Eff Res* 2013; 2(5):433-6.
15. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *Br Med J* 1950; 2(4682):739-48.
16. Alberg AJ, Brock MV, Ford JG, et al. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(Suppl5):e1S-29S.
17. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer* 2007; 120(4):868-74.
18. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Baseline characteristics of participants in the randomized national lung screening trial. *J Natl Cancer Inst* 2010; 102(23):1771-9.

19. Horeweg N, van der Aalst CM, Thunnissen E, et al. Characteristics of Lung Cancers Detected by Computer Tomography Screening in the Randomized NELSON Trial. *Am J Respir Crit Care Med* 2013; 187(8):848-54.
20. Pinsky PF, Church TR, Izmirlian G, et al. The National Lung Screening Trial: results stratified by demographics, smoking history, and lung cancer histology. *Cancer* 2013; 119(22):3976-83.
21. Henschke CI, Yip R, International Early Lung Cancer Action Program Investigators. Definition of a positive test result in computed tomography screening for lung cancer: a cohort study. *Ann Intern Med* 2013; 158(4):246-52.
22. Fardanesh M, White C. Missed lung cancer on chest radiography and computed tomography. *Semin Ultrasound CT MR* 2012; 33(4):280-7.
23. Horeweg N, Scholten ETS, de Jong PA, et al. Lung cancer detection in the NELSON trial: screening test performance and opportunities for improvement. *Submitted* 2014.
24. White CS, Romney BM, Mason AC, et al. Primary carcinoma of the lung overlooked at CT: analysis of findings in 14 patients. *Radiology* 1996; 199(1):109-15.
25. Kaneda M, Tarukawa T, Watanabe F, et al. Clinical features of primary lung cancer adjoining pulmonary bulla. *Interact Cardiovasc Thorac Surg* 2010; 10(6):940-4.
26. Li F, Sone S, Abe H, et al. Lung cancers missed at low-dose helical CT screening in a general population: comparison of clinical, histopathologic, and imaging findings. *Radiology* 2002; 225(3):673-83.
27. Kakinuma R, Ohmatsu H, Kaneko M, et al. Detection failures in spiral CT screening for lung cancer: analysis of CT findings. *Radiology* 1999; 212(1):61-6.
28. Gurney JW. Missed lung cancer at CT: imaging findings in nine patients. *Radiology* 1996; 199(1):117-22.
29. Lacson R, Prevedello LM, Andriole KP, et al. Factors associated with radiologists' adherence to Fleischner Society guidelines for management of pulmonary nodules. *J Am Coll Radiol* 2012; 9(7):468-73.
30. Yuan R, Vos PM, Cooperberg PL. Computer-aided detection in screening CT for pulmonary nodules. *AJR Am J Roentgenol* 2006; 186(5):1280-7.
31. Farooqi AO, Cham M, Zhang L, et al. Lung cancer associated with cystic airspaces. *AJR Am J Roentgenol* 2012; 199(4):781-6.
32. Xu DM, van der Zaag-Loonen HJ, Oudkerk M, et al. Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. *Radiology* 2009; 250(1):264-72.
33. Gould MK, Donington J, Lynch WR, et al. Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer?: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2013; 143(Suppl5):e93S-e120S.
34. Midthun DE, Swensen SJ, Jett JR, et al. Evaluation of nodules detected by screening for lung cancer with low-dose spiral computed tomography. *Lung Cancer* 2003; 41(suppl2):S40.
35. Henschke CI, Yankelevitz DF, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. *Radiology* 2004; 231(1):164-8.
36. Swensen SJ, Jett JR, Hartman TE, et al. Lung cancer screening with CT: Mayo Clinic experience. *Radiology* 2003; 226(3):756-61.
37. Naidich DP, Bankier AA, MacMahon H, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013; 266(1):304-17.
38. Linet MS, Slovis TL, Miller DL, et al. Cancer risks associated with external radiation from diagnostic imaging procedures. *CA Cancer J Clin* 2012; 62(2):75-100.

39. Wiener RS, Gould MK, Woloshin S, et al. What do you mean, a spot?: A qualitative analysis of patients' reactions to discussions with their physicians about pulmonary nodules. *Chest* 2013; 143(3):672-7.
40. van den Bergh KA, Essink-Bot ML, Borsboom GJ, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *Br J Cancer* 2010; 102(1):27-34.
41. Gohagan J, Marcus P, Fagerstrom R, et al. Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute. *Chest* 2004; 126(1):114-21.
42. Ost D, Shah R, Anasco E, et al. A randomized trial of CT fluoroscopic-guided bronchoscopy vs conventional bronchoscopy in patients with suspected lung cancer. *Chest* 2008; 134(3):507-13.
43. Roth K, Hardie JA, Andreassen AH, et al. Predictors of diagnostic yield in bronchoscopy: a retrospective cohort study comparing different combinations of sampling techniques. *BMC Pulm Med* 2008; 8:2.
44. Vincent BD, Fraig M, Silvestri GA. A pilot study of narrow-band imaging compared to white light bronchoscopy for evaluation of normal airways and premalignant and malignant airways disease. *Chest* 2007; 131(6):1794-9.
45. Chhajed PN, Bernasconi M, Gambazzi F, et al. Combining bronchoscopy and positron emission tomography for the diagnosis of the small pulmonary nodule < or = 3 cm. *Chest* 2005; 128(5):3558-64.
46. Reichenberger F, Weber J, Tamm M, et al. The value of transbronchial needle aspiration in the diagnosis of peripheral pulmonary lesions. *Chest* 1999; 116(3):704-8.
47. Baaklini WA, Reinoso MA, Gorin AB, et al. Diagnostic yield of fiberoptic bronchoscopy in evaluating solitary pulmonary nodules. *Chest* 2000; 117(4):1049-54.
48. van 't Westeinde SC, Horeweg N, Vernhout RM, et al. The role of conventional bronchoscopy in the workup of suspicious CT scan screen-detected pulmonary nodules. *Chest* 2012; 142(2):377-84.
49. Gould MK, Fletcher J, Iannettoni MD, et al. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* 2007; 132(Suppl3):108S-30S.
50. Beigelman-Aubry C, Hill C, Grenier PA. Management of an incidentally discovered pulmonary nodule. *Eur Radiol* 2007; 17(2):449-66.
51. Infante M, Chiesa G, Solomon D, et al. Surgical procedures in the DANTE trial, a randomized study of lung cancer early detection with spiral computed tomography: comparative analysis in the screening and control arm. *J Thorac Oncol* 2011; 6(2):327-35.
52. Veronesi G, Spaggiari L, Preda P, et al. Morbidity and mortality after surgical treatment of screening detected lung cancer. *J Clin Oncol* 2008; 26(Suppl.No.15S):7531.
53. Whitson BA, Andrade RS, Boettcher A, et al. Video-assisted thoracoscopic surgery is more favorable than thoracotomy for resection of clinical stage I non-small cell lung cancer. *Ann Thorac Surg* 2007; 83(6):1965-70.
54. Sirbu H, Schreiner W, Dalichau H, et al. Surgery for non-small cell carcinoma in geriatric patients: 15-year experience. *Asian Cardiovasc Thorac Ann* 2005; 13(4):330-6.
55. Mishra PK, Pandey R, Shackcloth MJ, et al. Cardiac comorbidity is not a risk factor for mortality and morbidity following surgery for primary non-small cell lung cancer. *Eur J Cardiothorac Surg* 2009; 35(3):439-43.
56. Berry MF, Hanna J, Tong BC, et al. Risk factors for morbidity after lobectomy for lung cancer in elderly patients. *Ann Thorac Surg* 2009; 88(4):1093-9.
57. Villamizar NR, Darrabie MD, Burfeind WR, et al. Thoracoscopic lobectomy is associated with lower morbidity compared with thoracotomy. *J Thorac Cardiovasc Surg* 2009; 138(2):419-25.

58. Handy JR, Jr., Asaph JW, Douville EC, et al. Does video-assisted thoracoscopic lobectomy for lung cancer provide improved functional outcomes compared with open lobectomy? *Eur J Cardiothorac Surg* 2010; 37(2):451-5.
59. Birim O, Zuydendorp HM, Maat AP, et al. Lung resection for non-small-cell lung cancer in patients older than 70: mortality, morbidity, and late survival compared with the general population. *Ann Thorac Surg* 2003; 76(6):1796-801.
60. Birim O, Maat AP, Kappetein AP, et al. Validation of the Charlson comorbidity index in patients with operated primary non-small cell lung cancer. *Eur J Cardiothorac Surg* 2003; 23(1):30-4.
61. Pagni S, McKelvey A, Riordan C, et al. Pulmonary resection for malignancy in the elderly: is age still a risk factor? *Eur J Cardiothorac Surg* 1998; 14(1):40-4.
62. Paul S, Altorki NK, Sheng S, et al. Thoracoscopic lobectomy is associated with lower morbidity than open lobectomy: a propensity-matched analysis from the STS database. *J Thorac Cardiovasc Surg* 2010; 139(2):366-78.
63. Suemitsu R, Takeo S, Hamatake M, et al. The perioperative complications for elderly patients with lung cancer associated with a pulmonary resection under general anesthesia. *J Thorac Oncol* 2009; 4(2):193-7.
64. Scott WJ, Allen MS, Darling G, et al. Video-assisted thoracic surgery versus open lobectomy for lung cancer: a secondary analysis of data from the American College of Surgeons Oncology Group Z0030 randomized clinical trial. *J Thorac Cardiovasc Surg* 2010; 139(4):976-83.
65. Cattaneo SM, Park BJ, Wilton AS, et al. Use of video-assisted thoracic surgery for lobectomy in the elderly results in fewer complications. *Ann Thorac Surg* 2008; 85(1):231-6.
66. Matsuoka H, Okada M, Sakamoto T, et al. Complications and outcomes after pulmonary resection for cancer in patients 80 to 89 years of age. *Eur J Cardiothorac Surg* 2005; 28(3):380-3.
67. Yang X, Wang S, Qu J. Video-assisted thoracic surgery (VATS) compares favorably with thoracotomy for the treatment of lung cancer: a five-year outcome comparison. *World J Surg* 2009; 33(9):1857-61.
68. Flores RM, Park BJ, Dycoco J, et al. Lobectomy by video-assisted thoracic surgery (VATS) versus thoracotomy for lung cancer. *J Thorac Cardiovasc Surg* 2009; 138(1):11-8.
69. Kim K, Kim HK, Park JS, et al. Video-assisted thoracic surgery lobectomy: single institutional experience with 704 cases. *Ann Thorac Surg* 2010; 89(6):S2118-22.
70. Petersen RH, Hansen HJ. Learning thoracoscopic lobectomy. *Eur J Cardiothorac Surg* 2010; 37(3):516-20.
71. Jaklitsch MT, Bueno R, Swanson SJ, et al. New surgical options for elderly lung cancer patients. *Chest* 1999; 116(Suppl6):480S-5S.
72. Sugiura H, Morikawa T, Kaji M, et al. Long-term benefits for the quality of life after video-assisted thoracoscopic lobectomy in patients with lung cancer. *Surg Lapsc Endosc Perc Tech* 1999; 9(6):403-8.
73. Thomas P, Piraux M, Jacques LF, et al. Clinical patterns and trends of outcome of elderly patients with bronchogenic carcinoma. *Eur J Cardiothorac Surg* 1998; 13(3):266-74.
74. Moro-Sibilot D, Aubert A, Diab S, et al. Comorbidities and Charlson score in resected stage I nonsmall cell lung cancer. *Eur Respir J* 2005; 26(3):480-6.
75. Memtsoudis SG, Besculides MC, Zellos L, et al. Trends in lung surgery: United States 1988 to 2002. *Chest* 2006; 130(5):1462-70.
76. Meguid RA, Brooke BS, Chang DC, et al. Are surgical outcomes for lung cancer resections improved at teaching hospitals? *Ann Thorac Surg* 2008; 85(3):1015-25.
77. Little AG, Rusch VW, Bonner JA, et al. Patterns of surgical care of lung cancer patients. *Ann Thorac Surg* 2005; 80(6):2051-6.

78. Igai H, Takahashi M, Ohata K, et al. Surgical treatment for non-small cell lung cancer in octogenarians--the usefulness of video-assisted thoracic surgery. *Interact Cardiovasc Thorac Surg* 2009; 9(2):274-7.
79. Harpole DH, Liptay MJ, DeCamp MM, et al. Prospective analysis of pneumonectomy: risk factors for major morbidity and cardiac dysrhythmias. *Ann Thorac Surg* 1996; 61(3):977-82.
80. Boffa DJ, Allen MS, Grab JD, et al. Data from The Society of Thoracic Surgeons General Thoracic Surgery database: the surgical management of primary lung tumors. *J Thorac Cardiovasc Surg* 2008; 135(2):247-54.
81. Bach PB, Cramer LD, Schrag D, et al. The influence of hospital volume on survival after resection for lung cancer. *N Engl J Med* 2001; 345(3):181-8.
82. Allen MS, Darling GE, Pechet TT, et al. Morbidity and mortality of major pulmonary resections in patients with early-stage lung cancer: initial results of the randomized, prospective ACOSOG Z0030 trial. *Ann Thorac Surg* 2006; 81(3):1013-20.
83. Belgers EH, Siebenga J, Bosch AM, et al. Complete video-assisted thoracoscopic surgery lobectomy and its learning curve. A single center study introducing the technique in The Netherlands. *Interact Cardiovasc Thorac Surg* 2010; 10(2):176-80.
84. Van't Westeinde SC, Horeweg N, De Leyn P, et al. Complications following lung surgery in the Dutch-Belgian randomized lung cancer screening trial. *Eur J Cardiothorac Surg* 2012; 42(3):420-9.
85. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; 361(23):2221-9.
86. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev* 2012; 21(3):308-15.
87. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer* 2009; 64(1):34-40.
88. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J* 2013; 42(6):1659-67.
89. Detterbeck FC, Postmus PE, Tanoue LT. The stage classification of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143(Suppl5):e191S-210S.
90. Parkash O. Lung cancer. A statistical study based on autopsy data from 1928 to 1972. *Respiration* 1977; 34(5):295-304.
91. Lince L, Lulu DJ. Carcinoma of the lung. A comparative series of 687 cases. *Arch Surg* 1971; 102(2):103-7.
92. Subramaniam RP, Asgharian B, Freijer JI, et al. Analysis of lobar differences in particle deposition in the human lung. *Inhal Toxicol* 2003; 15(1):1-21.
93. Churg A, Stevens B. Association of lung cancer and airway particle concentration. *Environ Res* 1988; 45(1):58-63.
94. Radzikowska E, Glaz P, Roszkowski K. Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases. *Ann Oncol* 2002; 13(7):1087-93.
95. Sagerup CM, Smastuen M, Johannesen TB, et al. Sex-specific trends in lung cancer incidence and survival: a population study of 40,118 cases. *Thorax* 2011; 66(4):301-7.
96. Ries L, Eisner, M., Kosary, C. . SEER Cancer Statistics Review 1975-2002. National Cancer Institute. 2005. http://seercancergov/csr/1975_2002/results_merged/sect_15_lung_bronchuspdf Date last accessed 05-09-2012.
97. Schwartz KL, Crossley-May H, Vigneau FD, et al. Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes Control* 2003; 14(8):761-6.

98. Caldarella A, Crocetti E, Comin CE, et al. Gender differences in non-small cell lung cancer: a population-based study. *Eur J Surg Oncol* 2007; 33(6):763-8.
99. Ringer G, Smith JM, Engel AM, et al. Influence of sex on lung cancer histology, stage, and survival in a midwestern United States tumor registry. *Clin Lung Cancer* 2005; 7(3):180-2.
100. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst* 2002; 94(3):167-73.
101. Harteloh P, de Bruin K, Kardaun J. The reliability of cause-of-death coding in The Netherlands. *Eur J Epidemiol* 2010; 25(8):531-8.
102. Maudsley G, Williams EM. "Inaccuracy' in death certification--where are we now? *J Public Health Med* 1996; 18(1):59-66.
103. Makinen T, Karhunen P, Aro J, et al. Assessment of causes of death in a prostate cancer screening trial. *Int J Cancer* 2008; 122(2):413-7.
104. Otto SJ, van Leeuwen PJ, Hoekstra JW, et al. Blinded and uniform causes of death verification in cancer screening: a major influence on the outcome of a prostate cancer screening trial? *Eur J Cancer* 2010; 46(17):3061-7.
105. Doria-Rose VP, Marcus PM. Death certificates provide an adequate source of cause of death information when evaluating lung cancer mortality: an example from the Mayo Lung Project. *Lung Cancer* 2009; 63(2):295-300.
106. Doria-Rose VP, Marcus PM, Miller AB, et al. Does the source of death information affect cancer screening efficacy results? A study of the use of mortality review versus death certificates in four randomized trials. *Clin Trials* 2010; 7(1):69-77.
107. Ederer F, Geisser MS, Mongin SJ, et al. Colorectal cancer deaths as determined by expert committee and from death certificate: a comparison. The Minnesota Study. *J Clin Epidemiol* 1999; 52(5):447-52.
108. Chamberlain J, Coleman D, Ellman R, et al. Verification of the cause of death in the trial of early detection of breast cancer. UK Trial of Early Detection of Breast Cancer Group. Trial Co-ordinating Centre. *Br J Cancer* 1991; 64(6):1151-6.
109. Albertsen PC, Walters S, Hanley JA. A comparison of cause of death determination in men previously diagnosed with prostate cancer who died in 1985 or 1995. *J Urol* 2000; 163(2):519-23.

Chapter 12

Summary

Part I: Introduction

Lung cancer is a major public health problem since it causes most cancer-related deaths worldwide. As the disease often causes no symptoms at early stages, diagnosis at advanced stages, wherein cure is no longer possible, is common. Improvements in lung cancer treatment have been made, but yielded only modest improvement in survival over the last decades. Continuous efforts should be made to force back exposure to causative agents of lung cancer, tobacco smoking in particular. However, this is not expected to reverse the lung cancer epidemic in the next decades. Lung cancer screening can reduce morbidity and mortality from lung cancer by detecting the disease at an early and curable stage. As this early stage is often not accompanied by any signs or symptoms, screening has to be applied to apparently healthy, asymptomatic persons. Unfortunately, screening exposes these persons to several harms: some related to the screening test itself, such as exposure to ionising radiation, others related to false-positive, false-negative screening test results, or overdiagnosis. Therefore, only lung cancer screening programs wherein benefits outweigh harms should be implemented. This thesis aimed to identify opportunities to improve the balance between benefits and harms of a screening program.

Part II: Evaluation of findings

In *Chapter 2*, data on screening test results and screen-detected lung cancer were used to determine positive predictive value and 5.5-year lung cancer probability. This study demonstrated that the screening algorithm of the NELSON trial adequately stratified participants according to their lung cancer risk. Further, the screening algorithm yielded a limited number of follow-up low-dose computed tomography (LDCT) and additional diagnostic procedures for positive screening test results. Although the predictive value of screening test results in the NELSON trial compared favourably to other studies, too many invasive diagnostic procedures were performed for false-positive screening test results.

In *Chapter 3,* the tumour characteristics of the lung cancers detected by screening were analysed. Analyses showed that screening yielded a stage shift towards earlier diagnosis, more in women than in men; and a shift in histology towards slower growing and more peripherally localised subtypes of lung cancer.

In *Chapter 4*, the performance of the screening algorithm of the NELSON trial was estimated, and opportunities to improve its performance were identified. Detection rates and sensitivity of the NELSON screening protocol appeared to be sufficient for diagnosing lung cancer as early as in a lung cancer trial that reduced lung cancer mortality significantly. Moreover, the NELSON screening protocol yielded a very high specificity, which is a prerequisite for cost-effectiveness. Nonetheless, performance of the screening protocol may be improved by co-implementation of CT screening with an effective

smoking cessation program, and training of screening radiologists to reduce the number of detection and interpretation errors.

In *Chapter 5*, radiological causes of the failure to detect lung cancers by screening were investigated, and opportunities to improve performance of the screening algorithm were identified. This evaluation learned that the majority of the lung cancers not detected by low-dose CT screening were not preventable. Performance of the screening strategy may be improved by reducing the number of detection and interpretation errors, which may be achieved by increasing the radiologist's attention for endobronchial lesions, bulla wall thickenings, and pleural-attached lesions.

Part III: Optimisation of screening

In *Chapter 6*, lung cancer probability of participants was estimated and used to design and evaluated nodule management protocols. The current guideline on the management of nodules classifies nodules <4 mm as not suspicious for lung cancer; nodules of 4 to 8 mm as indeterminate (for which growth assessment is required: nodules with a volume doubling-time <400 days are subsequently classified as suspicious for lung cancer); and nodules ≥8 mm as suspicious for lung cancer. Analyses showed that the guideline performed well, but also that improvements were possible. Raising nodule size diameter threshold from 4 mm to 5mm and from 8 mm to 10 mm, and nodule volume doubling-time threshold from 400 days to 600 days was estimated to yield both a higher sensitivity and a higher specificity. Further, a nodule management protocol using nodule volume thresholds of 100 mm^3 and 300 mm^3, and a nodule volume doubling-time threshold of 600 days was evaluated. This protocol was estimated to yield the same sensitivity as the current guideline, but a substantially higher specificity. Results of this study imply that use of volumetry and lung cancer probability-based thresholds for nodule size and growth can improve lung cancer detection and reduce unnecessary follow-up CT examinations and invasive diagnostic procedures.

In *Chapter 7*, the value of white light bronchoscopy in the diagnostic work-up of suspicious CT-detected nodules was determined. This study demonstrated that bronchoscopy could not diagnose lung cancer effectively due to insufficient sensitivity, and could not exclude lung cancer reliably due to deceitful benign diagnoses. Therefore, routine use of bronchoscopy in subjects with suspicious pulmonary nodules detected in a lung cancer screening programme is not recommended.

In *Chapter 8*, adverse events related to thoracic surgery, performed in the diagnostic work-up of suspicious CT-detected nodules were assessed. This study demonstrated that adverse events after thoracic surgery for positive screening test results were common. Incidence of minor complications was relatively high, while incidence of major complications, re-operations and post-operative mortality was relatively low. Finally, a relatively high percentage of surgeries were performed for benign nodules.

Part IV: Evaluation of effectiveness

In *Chapter 9*, design and evaluation of the endpoint verification process of the NELSON trial was presented. This study demonstrated that an independent committee of experts should perform a cause of death verification process to ensure uniform, objective and unbiased endpoint determination. For this process, medical files of all deceased study participants diagnosed with lung cancer should be blinded for participant identity and study group. Subsequently, these files should be reviewed using a flowchart and criteria to determine grade of certainty that lung cancer was the primary cause of death. A pilot study demonstrated that this method is preferred over use of official death certificates, which have insufficient distinctive character for lung cancer-specific death.

Part V: Implications for implementation

In *Chapter 10*, currently available literature on all relevant aspects of LDCT screening for lung cancer was reviewed to determine whether benefits of LDCT screening outweigh its harms. Next, it was determined whether LDCT screening meets the World Health Organisation criteria for screening. This review learned that initial estimates of several harms and benefits of screening have been made, but substantial gaps in knowledge remain. Currently available evidence suggests that benefits of LDCT screening outweigh its harms.

In *Chapter 11*, the 'General Discussion', results of this thesis were summarised and discussed. The screening algorithm of the NELSON trial yielded a favourable balance between negative, indeterminate and positive screening test results, which led to a limited number of follow-up LDCT scans and diagnostic work-ups. The screening algorithm had a high sensitivity, which is promising for mortality analysis which is planned at ten years of follow-up. The screening algorithm of the NELSON trial also yielded a very high specificity, which is promising for the planned cost-effectiveness analysis. Lung cancers detected through screening were diagnosed at early stages. Moreover, the cancer stage distribution of lung cancers detected and missed by screening combined was also favourable. The majority of lung cancers missed by screening were not preventable. Nonetheless, performance of the screening algorithm may be improved by reducing detection and interpretation errors, which may be achieved by increasing the radiologist's attention for endobronchial lesions, bulla wall thickenings, and pleural-attached lesions. Further, performance of the screening algorithm may be improved by slight adjustments of the thresholds for nodule size and growth rate that determine the screening test result. Additionally, co-implementation of CT screening with an effective smoking cessation program will also contribute to improved performance of the screening algorithm. Harms and costs of the screening algorithm can safely be reduced by eliminating routine use of bronchoscopy for suspicious screen-detected nodules from diagnostic work-up. Harms induced by adverse events after thoracic surgery both for lung cancer and for benign

nodules may be reduced by routine use of video-assisted thoracoscopic surgery. Finally, effectiveness of the screening algorithm of the NELSON trial will be determined using the endpoint determination procedure presented in this thesis.

Chapter 13

Samenvatting

Deel I: Introductie

Longkanker is een groot maatschappelijk gezondheidsprobleem doordat het wereldwijd één van de meest voorkomende vormen van kanker is en de meeste kanker-gerelateerde sterfgevallen veroorzaakt. Longkanker veroorzaakt vaak pas in een gevorderd stadium klachten, dit geeft een vertraging in het tijdstip waarop de diagnose wordt gesteld en meestal is dan geen genezing meer mogelijk is. Hoewel de behandeling van longkanker in de afgelopen decennia is verbeterd, heeft dit slechts geresulteerd in een minimale verbetering in de overleving van longkankerpatiënten. Het blijft noodzakelijk om het gebruik van en de blootstelling aan stoffen die longkanker veroorzaken te blijven terugdringen, dit geldt in het bijzonder voor tabaksrook. Toch is het de verwachting dat preventieve maatregelen op zichzelf niet afdoende zullen zijn om de longkankerepidemie in de komende decennia terug te dringen. Screening op longkanker kan de gezondheidsschade en sterfte veroorzaakt door longkanker verminderen door de ziekte in een vroeg en behandelbaar ziektestadium op te sporen. Aangezien het vroege stadium van longkanker vaak geen klachten veroorzaakt, vindt de screening plaats op schijnbaar gezonde personen. Deze personen, van wie slechts een deel een vroeg stadium van longkanker onder de leden heeft, lopen door de screening echter ook risico's. De risico's van screening zijn onder andere gerelateerd aan de screening test zelf, zoals blootstelling aan ioniserende straling. Daarnaast is er een kans op een fout-positieve of een fout-negatieve screeningtestuitslag en de kans op overdiagnose. Vanwege deze schadelijke neveneffecten screening is het wenselijk om alleen longkankerscreeningsprogramma's in te voeren waarvan de voordelen opwegen tegen de nadelen. Het doel van dit promotieonderzoek was mogelijkheden identificeren die de balans tussen de voordelen en nadelen van een longkankerscreeningsprogramma kunnen verbeteren.

Deel II: Evaluatie van de bevindingen

In *Hoofdstuk 2*, wordt beschreven hoe de screeningtestuitslagen en de door screening gedetecteerde longkankers werden gebruikt om de positief voorspellende waarde van de screeningtest te bepalen. Verder werd in dit hoofdstuk de kans op longkanker in de komende 5,5 jaar op basis van de screeningtestuitslag geschat. Dit onderzoek toonde aan dat het screeningprotocol van de NELSON studie goed in staat was om de studiedeelnemers in te delen naar hun risico op longkanker. Screening in de NELSON studie leidde bovendien tot een beperkt aantal vervolgscans en aanvullende diagnostische onderzoeken. In vergelijking met andere studies was de voorspellende waarde van een positieve screeningtestuitslag hoog in de NELSON studie. Echter, er werden wel relatief meer invasieve diagnostische onderzoeken gedaan voor fout-positieve screeningtestuitslagen.

In *Hoofdstuk 3* werden de tumorkarakteristieken van de door screening gedetecteerde longkankers gepresenteerd. Analyses toonden aan dat screening heeft geleid tot de diagnose van longkanker in een gunstiger ziektestadium, dit effect was sterker bij vrouwen

dan bij mannen. Daarnaast werd aangetoond dat screening frequent leidt tot de detectie van traag-groeiende longkankers die zich in de buitenranden van de longen bevinden.

In *Hoofdstuk 4* werden de testkarakteristieken van het screeningsprotocol van de NELSON studie geschat en werden mogelijkheden om het screeningsprotocol te verbeteren geïdentificeerd. Het detectievermogen en de sensitiviteit van het screeningsprotocol bleken in staat om longkanker in een net zo'n vroeg stadium op te sporen als gepresenteerd in een ander longkankerscreeningsstudie waarvan de effectiviteit al is aangetoond. Bovendien bleek dat het screeningsprotocol van de NELSON studie een zeer hoge specificiteit heeft, wat een voorwaarde is voor een kosteneffectief screeningsprogramma. Toch kunnen de uitkomsten van het screeningsprotocol mogelijk verbeterd worden door gelijktijdige implementatie met een effectief stoppen-met-roken-programma en een trainingsprogramma voor screeningsradiologen dat het aantal detectie-, en interpretatiefouten verminderd.

In *Hoofdstuk 5* werd onderzocht of er radiologische oorzaken ten grondslag lagen aan het missen van longkankers in de studie en of er mogelijkheden om het screeningsprotocol te verbeteren. Dit onderzoek toonde aan dat de meerderheid van de gemiste longkankers niet te voorkomen was. Echter, winst valt te behalen door het verminderen van het aantal detectie-, en interpretatiefouten door radiologen. Deze fouten kunnen mogelijk verminderd worden door de aandacht van de screeningsradioloog te verhogen voor endobronchiale laesies, wandverdikkingen in longblazen en laesies die vastzitten aan de longvliezen.

Deel III: Optimalisatie van screening

In *Hoofdstuk 6* werd het risico op longkanker van de studiedeelnemers geschat. Deze risico-inschattingen werden gebruikt om protocollen voor management van nodules te ontwerpen en te evalueren. De huidige richtlijn voor het management van longnodules classificeert nodules kleiner dan 4 mm als niet verdacht voor longkanker; nodules van 4 tot 8 mm als onduidelijk (waarvoor bepaling van de groeisnelheid geïndiceerd is: nodules met een volume-verdubbelingstijd korter dan 400 dagen worden vervolgens als verdacht voor longkanker geclassificeerd); en nodules van 8 mm en groter als verdacht voor longkanker. Dit onderzoek toonde aan dat deze richtlijn voldoet, maar ook dat er mogelijkheid tot verbetering is. Zo zullen de sensitiviteit en specificiteit toenemen door het verhogen van de afkapwaarden voor nodulegrootte van 4 naar 5 mm en van 8 naar 10 mm, en de afkapwaarde voor nodulegroei van 400 naar 600 dagen. Daarnaast is een nodulemanagementprotocol ontworpen en geëvalueerd dat gebruik maakt van volumetrie. De afkapwaarden van dit protocol waren; volumes van 100mm^3 en 300mm^3 voor nodulegrootte, en een volume verdubbelingstijd van 600 dagen. Dit protocol heeft dezelfde sensitiviteit als de huidige richtlijn, maar een substantieel hogere specificiteit. Concluderend, het gebruik van volumetrie en het gebruik van afkapwaarden voor nodulegrootte en nodulegroei die

gebaseerd zijn op longkankerrisico kunnen longkankerdetectie verbeteren en het aantal onnodige CT scans en invasieve diagnostische onderzoeken verminderen.

In *Hoofdstuk 7* is de waarde van bronchoscopie als diagnostisch instrument in de work-up van verdachte longnodules vastgesteld. Dit onderzoek toonde aan dat longkanker niet effectief vastgesteld kon worden met bronchoscopie door onvoldoende sensitiviteit. Daarnaast kon longkanker niet betrouwbaar uitgesloten worden met bronchoscopie omdat bronchoscopie tot misleidende goedaardige diagnoses kon leiden in patiënten met longkanker. Kortom, het routinematig gebruik van bronchoscopie in een longkankerscreeningsprogramma wordt op basis van dit onderzoek afgeraden.

In *Hoofdstuk 8* werden de complicaties in kaart gebracht die kunnen optreden na chirurgie van de long die plaats vond naar aanleiding van verdachte nodules op de CT-scan. Deze studie toonde aan dat complicaties na longchirurgie vaak voorkwamen. Vergeleken met andere studies kwamen milde complicaties relatief vaak voor in de NELSON studie, maar ernstige complicaties, zoals een her-operatie en postoperatieve sterfte kwamen relatief weinig voor. Daarnaast werd in deze studie gevonden dat er in de NELSON studie relatief vaak geopereerd is voor goedaardige nodules.

Deel IV: Evaluatie van effectiviteit

In *Hoofdstuk 9* werd het eindpuntverificatieproces van de NELSON studie ontworpen en geëvalueerd. Deze studie toonde aan dat een onafhankelijke commissie van experts een geprotocolleerd review proces zouden moeten uitvoeren. Dit leidt tot een eenduidige, objectieve en betrouwbare bepaling van de eindpunten. Om de betrouwbaarheid te verhogen is het is belangrijk dat de medische status van alle studiedeelnemers die ooit gediagnosticeerd zijn met longkanker geblindeerd worden voor de identiteit en studiegroep van de deelnemer. En om vast te stellen of de deelnemer al dan niet overleden is aan longkanker moeten deze medische statussen gereviewd worden aan de hand van een stroomdiagram en vooraf vastgestelde criteria. De pilotstudie toonde aan dat deze methode te verkiezen is boven het gebruik van de officiële overlijdenscertificaten. De laatst genoemde heeft als nadeel dat dit onvoldoende onderscheidend vermogen voor longkanker-specifieke sterfte hebben.

Deel V: Implicaties voor implementatie

In *Hoofdstuk 10* werd een studie gepresenteerd waarin gepubliceerde literatuur over alle relevante aspecten van longkankerscreening in ogenschouw werd genomen om te bepalen of de voordelen van longkankerscreening met CT scans opwegen tegen de nadelen. Vervolgens werd bepaald of longkankerscreening met CT voldoet aan de screeningscriteria van de Wereldgezondheidsorganisatie. Dit onderzoek toonde aan dat er voorlopige schattingen zijn gemaakt van de voordelen en nadelen van longkankerscreening, maar dat er ook nog veel onduidelijkheden zijn. Op basis van deze onvolledige informatie kan

slechts geconcludeerd worden dat de literatuur suggereert dat de voordelen van longkankerscreening opwegen tegen de nadelen.

In *Hoofdstuk 11*, de 'Algemene Discussie' werden de resultaten van dit promotieonderzoek samengevat en besproken. Het screeningsprotocol van de NELSON studie leverde een gunstige balans op tussen het aantal negatieve, twijfelachtige en positieve testuitslagen, waardoor het aantal vervolgonderzoeken beperkt was. Het screeningsprotocol had een hoge sensitiviteit, wat veelbelovend is voor de eindanalyses naar het effect van screening op longkankersterfte. Het screeningsprotocol had ook een zeer hoge specificiteit, wat veelbelovend is voor de analyses naar de kosteneffectiviteit van longkankerscreening. De longkankers gedetecteerd door de screening werden in een vroeg ziektestadium gediagnosticeerd. Bovendien is ook de gecombineerde stadiumverdeling van de longkankers die gedetecteerd en gemist zijn door screening gunstig. De meerderheid van de oorzaken van het missen van longkankers in de screeningsstudie waren niet te voorkomen. Toch zou het screeningsprotocol verbeterd kunnen worden door een vermindering van het aantal detectie-, en interpretatiefouten, hetgeen bereikt zou kunnen worden door meer aandacht van de radiologen voor endobronchiale laesies, wandverdikkingen in longblazen en laesies die vastzitten aan de longvliezen. Daarnaast kan het screeningsprotocol mogelijk verbeterd worden door kleine aanpassingen van de afkapwaarden voor nodulegrootte en de snelheid van nodulegroei die het screeningsresultaat bepalen. Implementatie van longkankerscreening gecombineerd met een effectief stoppen-met-roken-programma zal de balans tussen de voordelen en nadelen van longkanker screening verbeteren. Tijdens longkankerscreening lijkt geen plaats te zijn voor het routinematig toepassen van bronchoscopie om longkanker in een nodule uit te sluiten. Door deze vorm van diagnostiek niet op routinebasis toe te passen zullen de kosten en de kans op nadelige effecten van de screening afnemen. Tenslotte zal de effectiviteit van longkankerscreening in de NELSON studie worden vastgesteld met het eindpuntverificatieproces dat gepresenteerd werd in dit proefschrift.

Part VI

Miscellaneous

Chapter 14

Dankwoord

Er zijn een groot aantal mensen die ik wil bedanken omdat zij direct of indirect hebben bijgedragen aan de totstandkoming van dit proefschrift.

Prof.dr. H.J. de Koning, geachte promotor, beste Harry, je was zowel mijn promotor als mijn directe begeleider en daarmee de belangrijkste persoon tijdens mijn promotieonderzoek. Ik ben blij dat ik bij jou, in mijn ogen dé expert in de wereld op het gebied van screening, mijn onderzoek heb kunnen doen. Je hebt me de vrijheid gegeven om mijn eigen ideeën uit te werken en me vele mogelijkheden geboden om naar interessante workshops, congressen en cursussen te gaan. Bedankt voor het vertrouwen dat je in me hebt gehad en alles wat ik van je heb geleerd.

Prof.dr. H.C. Hoogsteden, geachte promotor, beste professor Hoogsteden, hartelijk dank voor uw steun tijdens mijn promotie en de mogelijkheden die u me geboden heeft.

Prof.dr. M.G.M. Hunink, prof.dr. P.E. Postmus, prof.dr. H. van Swieten, prof. I.D. de Beaufort, geachte leden van de promotiecommissie, hartelijk dank voor uw bereidheid om mijn proefschrift te beoordelen en met mij hierover van gedachten te wisselen op 25 november 2014.

Prof.dr. M. Oudkerk, geacht lid van de promotiecommissie, beste professor Oudkerk, hartelijk dank dat u zitting wilde nemen in de promotiecommissie en mijn proefschrift wilde beoordelen. Daarnaast wil ik u bedanken voor de tijd en energie die u in mijn artikelen heeft gestoken; met uw kritische commentaren heeft u telkens weer geprobeerd om het beste naar boven te halen.

Prof.dr. K. Nackaerts, geacht lid van de promotiecommissie, beste professor Nackaerts, allereerst wil ik u bedanken dat u zitting wilde nemen in de promotiecommissie. Tevens wil ik u bedanken voor de prettige samenwerking, uw waardevolle commentaren op mijn manuscripten, de uitnodiging om deel te nemen aan het Leuvens Longkanker Symposium en de uitnodiging om samen het 'priority paper evaluation'-artikel te schrijven.

Drs. S. Silva en M.F. Jonker MSc, lieve Sonja en Marcel, ik ben blij dat jullie mij als paranimfen tijdens de verdediging van mijn proefschrift willen bijstaan. Als kamer-(108)-genootjes hebben jullie me al vanaf het begin van mijn onderzoek bijgestaan; met verstandige adviezen, maar vooral ook met humor en een borrel op z'n tijd. Jullie hebben mijn promotietijd onvergetelijk gemaakt, 108 rules! ;)

Dr. E.Th. Scholten, beste Ernst, ik wil je heel erg bedanken voor je enorme bijdrage aan de twee artikelen over de intervalkankers (hoofdstuk 4 en 5), zonder jou waren ze er niet geweest. Ik vond het heel prettig om met je samen te werken en heb veel gehad aan je adviezen en steun.

Dr. J. van Rosmalen, beste Joost, heel erg bedankt voor je langdurige inzet voor het 'lung cancer probability'-artikel (hoofdstuk 6). Dankzij jouw bewonderenswaardige kennis van de biostatistiek en je flexibiliteit om dit te gebruiken om medische richtlijnen te

optimaliseren is ons artikel 'Lancet Oncology'-waardig geworden. Ik wens je veel succes met je nu al indrukwekkende loopbaan.

Dr. R.J. van Klaveren, beste Rob, ik wil je heel erg bedanken voor je begeleiding bij het 'bronchoscopie'-artikel (hoofdstuk 7) en het 'cause of death'-artikel (hoofdstuk 9). Je hebt me op een gedegen manier geleerd hoe onderzoek opgezet en artikelen geschreven moet worden. Dankzij het vertrouwen dat jij in me had, heb ik de mogelijkheid gekregen om dit promotieonderzoek te gaan doen. Ik vind het erg jammer dat onze samenwerking niet langer heeft mogen duren.

Dr. S.C. van 't Westeinde, lieve Susan, ten eerste wil ik je bedanken voor je begeleiding en hulp bij mijn allereerste artikel (hoofdstuk 7). Ik heb veel van je geleerd, zowel over het onderzoek als over andere dingen die belangrijk zijn voor een jonge dokter. Ik ben blij dat we als 'NELSON ladies' contact zullen houden.

Dr. C.M. van der Aalst, lieve Carlijn, toen ik met mijn promotieonderzoek begon was ik erg blij dat je me op sleeptouw nam en me wegwijs maakte op de afdeling. Bedankt voor je hulp bij mijn artikelen en de zaken er om heen. Ik vond het erg gezellig met je op de kamer, en we zeker zullen als 'NELSON ladies' contact houden.

K. ten Haaf MSc, lieve Kevin, je bent een bijzondere collega voor me geweest. Je bent een van de meest oprechte persoon die ik ken en ik heb je humor leren waarderen. Je gave om me op te vrolijken hebben me door teleurstellingen en frustraties, die bij het promoveren horen, heen geholpen.

Drs. A.U. Yousaf - Khan, beste Uraujh, wat leuk dat je het NELSON team bent komen versterken. Ik vond het gezellig om met je samen te werken en ik heb er alle vertrouwen in dat je een straks een mooi proefschrift zult afleveren. Veel succes met je verdere loopbaan.

M.A. Quak, lieve Marianne, heel erg bedankt voor al je inspanningen voor de NELSON studie en de hulp bij mijn projecten.

R. Faber en F. Santegoeds, beste Roel en Frank, heel erg bedankt voor jullie inspanningen om de datastromen van de NELSON studie in goede banen te leiden en voor jullie adviezen en hulp bij mijn projecten.

A. de Bruijn, lieve Arry, bedankt voor jouw hulp en ondersteuning die je direct en indirect aan de NELSON studie en mij hebt geboden. Harry boft maar met jou als secretaresse.

Dr. P.A. de Jong, beste Pim, bedankt voor je nuttige commentaren op mijn artikelen en je belangrijke bijdrage aan de twee intervalkanker artikelen (hoofdstuk 4 en 5).

Prof.dr. J.-W.J. Lammers en prof.dr. H.J.M. Groen, hartelijk dank voor uw waardevolle bijdragen aan al mijn artikelen en in het bijzonder voor uw rol als expert bij de pilot-studie naar het evalueren van de doodsoorzaken voor het 'cause of death'-artikel (hoofdstuk 9).

Drs. C. Weenink, Dr. E. Thunnissen, Dr. R. Vliegenthart, veel dank voor jullie bijdragen als co-auteurs aan mijn artikelen.

Dr M.A. den Bakker, Dr. J.A.M. Aerts, prof.dr. J.-W. Coeberg, beste Joachim, Michael en professor Coebergh, hartelijk dank dat jullie wilden plaatsnemen in de expertcommissie van de NELSON studie. Dankzij jullie wordt mijn eindpuntverificatieprotocol (hoofdstuk 9) toegepast en kan het zijn nut bewijzen.

Alle overige coauteurs die nog niet bij naam zijn genoemd, wil ik ook bij deze nogmaals bedanken voor hun bijdrage aan mijn artikelen.

René Vernhout, Emile Gras en Linda van Dongen, heel erg bedankt voor jullie inspanningen voor de NELSON studie en bedankt voor de gastvrijheid op het trialbureau in de Daniel den Hoed.

Ton de Jongh, dank voor het ontwerpen en onderhouden van het NELSON management systeem waar ik veel gebruik van heb gemaakt tijdens mijn onderzoek.

Dames van de ondersteuning en het secretariaat en heren van de ICT van de Afdeling Maatschappelijke Gezondheidszorg, bedankt dat jullie het mede mogelijk hebben gemaakt dat ik mijn werk kon doen.

Saskia van Amelsvoort - van de Vorst, Anneke Hamersma, Henk Pruiksma, Ria Ziengs, Liesbet Peeters, Beatrijs Anrijs, bedankt voor jullie belangrijke bijdrage aan de NELSON studie en jullie hulp bij de dataverzameling voor mijn artikelen.

Daarnaast wil ik alle iedereen bedanken die heeft bij gedragen aan het opzetten en uitvoeren van de NELSON studie, het beoordelen en behandelen van de patiënten die doorverwezen zijn door de NELSON studie, en iedereen heeft bijgedragen aan de dataverzameling.

Verder wil ik het Nederlands Kanker Register (in het bijzonder Reini Bretveld), het Centraal Bureau voor de Genealogie (in het bijzonder Martijn Spruit), Centraal Bureau voor de Statistiek (in het bijzonder Jan Kardaun en Kim de Bruin) en het Belgisch Kanker Register (in het bijzonder Liesbet Van Eycken en Karen Vos) bedanken voor hun belangrijke bijdragen aan het medisch-wetenschappelijk onderzoek in Nederland en België.

Tenslotte wil ik de deelnemers van de NELSON studie hartelijk danken voor hun ongelooflijk belangrijke bijdrage aan het wetenschappelijke onderzoek naar de vroegopsporing van longkanker.

Lieve Astrid, Elise, Kirsten, Nikki, en Jitske, ofwel girls van de koffieclub, jullie hebben mijn tijd bij MGZ zo veel leuker gemaakt door onze koffiebreaks waarin we alles over onze promotietrajecten en ons leven daar buiten met elkaar deelden. Dus speciaal voor jullie mijn twaalfde stelling: “Met een koffieclub overleef je elk promotieonderzoek.” ;)

Ik wil ook mijn andere collega’s van de Afdeling Maatschappelijke Gezondheidszorg bedanken voor de leuke tijd en veel succes wensen met hun promotieonderzoek: Domino, Kerstin, Fenna, Suzette, Britt, David en Katja, en iedereen die ik nog vergeet te noemen.

Katinka, Claudia en Jessica en andere vriendinnen en vrienden bedankt voor jullie interesse in mijn onderzoek en afleiding die jullie me in deze periode gegeven hebben.

Aad en Margot, ik kan me geen betere schoonouders wensen, bedankt voor jullie vertrouwen en steun.

Lieve Mam, Elien, en Tim, bedankt dat jullie er voor me waren en dat jullie vertrouwen in me hebben gehad.

Lieve Ivanca, bedankt voor de prachtige illustratie voor de cover van dit proefschrift en de gezellige afleiding in mijn promotietijd.

Lieve Wouter, ik weet niet beter dan dat je er altijd voor me bent, me steunt, aanmoedigt, en zorgt voor afleiding als ik te veel met werk bezig ben. Met jou kan ik de wereld aan.

Chapter 15

Curriculum vitae

Nanda Horeweg was born on May 14th 1986 in Spijkenisse, the Netherlands. In 2004, she completed secondary school at 'De Ring van Putten' in Spijkenisse. Subsequently, she started studying Medicine at the 'Erasmus University Rotterdam'. She wrote her graduation thesis on the value of bronchoscopy in the diagnostic work-up of suspicious pulmonary nodules detected by CT screening (chapter 8 of this thesis) under supervision of Dr. S.C. van 't Westeinde and Dr. R.J. van Klaveren. In 2010, she graduated with honours and obtained her medical degree. Thereafter, she started working as resident at the department of Pulmonary Medicine at Erasmus University Medical Center. In 2011, she was appointed as research-physician for the Dutch-Belgian lung cancer screening trial (NELSON trial) at the department of Public Health and the department of Pulmonary Medicine at Erasmus University Medical Center. Henceforth, she evaluated several aspects of lung cancer screening in the NELSON trial, under supervision of prof.dr. H.J. de Koning. Resultant research findings are presented in this thesis.

Nanda Horeweg werd geboren op 14 mei 1986 in Spijkenisse. Zij behaalde haar VWO diploma in 2004 aan 'De Ring van Putten' te Spijkenisse. Vervolgens ging zij Geneeskunde studeren aan de Erasmus Universiteit Rotterdam. Haar afstudeeronderzoek ging over de waarde van bronchoscopie in de diagnostische opwerking van longnodules gedetecteerd middels CT screening (hoofdstuk 8 van dit proefschrift), onder begeleiding van dr. S.C. van 't Westeinde en dr. R.J. van Klaveren. In 2010 studeerde ze cum laude af aan de Erasmus Universiteit en behaalde ze haar artsentitel. Aansluitend ging zij werken als arts-assistent op de afdeling Longziekten van het Erasmus Medisch Centrum. Vervolgens werkte zij vanaf 2011 als arts-onderzoeker voor het Nederlands-Leuvens longkankerscreeningsonderzoek (NELSON studie) op de afdeling Maatschappelijke Gezondheidszorg en de afdeling Longziekten van het Erasmus Medisch Centrum. Binnen de NELSON studie evalueerde zij verschillende aspecten van screening onder begeleiding van prof.dr. H.J. de Koning. Resultaten van dit onderzoek worden gepresenteerd in dit proefschrift.

Chapter 16

PhD portfolio

PHD PORTFOLIO

Summary of PhD training and teaching	
PhD student	Nanda Horeweg MD
Erasmus Medical Center Department	Public Health Pulmonary Medicine
PhD period	01-02-2012 to 01-08-2014
Promotors	H.J. de Koning MD PhD H.C. Hoogsteden MD PhD

1. PhD training

	Year	Workload
General courses		
Planning and evaluation of screening, NIHES, Rotterdam, Netherlands	2011	1.4 ECTS
Certificate of English course, Embassy CES, New York, USA	2011	60 hours
Biostatistical Methods II: classical regression models, NIHES, Rotterdam, Netherlands	2012	4.3 ECTS
Courses for the quantitative researcher, NIHES, Rotterdam, Netherlands	2012	1.4 ECTS
Repeated measurements, NIHES, Rotterdam, Netherlands	2012	1.4 ECTS
Missing values in clinical research, NIHES, Rotterdam, Netherlands	2012	0.7 ECTS
Analysis of growth data, NIHES, Rotterdam, Netherlands	2012	0.6 ECTS
Absolute risk prediction, Netherlands Cancer Institute, Amsterdam, Netherlands	2012	0.3 ECTS
Study design, NIHES, Rotterdam, Netherlands	2013	4.3 ECTS
Specific courses		
Methodologie van patiëntgebonden-onderzoek en voorbereiding subsidieaanvragen	2012	8 hours
Teach the teacher: Vaardigheidsonderwijs geven, Desiderius school, Erasmus University, Rotterdam, Netherlands	2013	12 hours
BROK course, Erasmus MC, Rotterdam, Netherlands	2013	20 hours
Scientific integrity, Erasmus MC, Rotterdam, Netherlands	2014	8 hours
Seminars and workshops		
PhD day, Erasmus MC, Rotterdam, Netherlands	2012	6 hours

Presentations		
14th World Conference on Lung Cancer, Amsterdam, Netherlands: poster presentation "The role of conventional bronchoscopy in the work-up of suspicious CT screen detected pulmonary nodules"	2011	1 ECTS
NELSON lung cancer screening symposium, Rotterdam, Netherlands: oral presentation "Blinded and uniform cause of death verification in a lung cancer CT screening trial"	2011	1 ECTS
WEON Congres, Rotterdam, Netherlands: oral presentation "Predictive value of scan results" and poster presentation "Blinded and uniform cause of death verification in a lung cancer CT screening trial"	2012	2 ECTS
European meeting lung cancer screening trials, Rotterdam, Netherlands: oral presentation "Cause of death verification in lung cancer screening trials"	2012	1 ECTS
Lung cancer screening symposium, Leuven, Belgium: oral presentation "Screening conditions"	2013	1 ECTS
American Thoracic Society International Conference, Philadelphia, USA: oral presentation "Characteristics of CT-detected lung cancers" and "Outcomes three rounds of the NELSON trial" and poster session facilitator	2013	2 ECTS
European Lung Cancer Conference, Lugano, Switserland: oral presentation "Volumetric screening for lung cancer"	2013	1 ECTS
Meeting design SCAPIS study, Stockholm, Sweden: "Management of CT-detected nodules"	2013	1 ECTS
World Conference on Lung Cancer, Sydney Australia: oral presentation at plenary presidential symposium "Lung cancer probability of subjects with CT-detected nodules"	2013	1 ECTS
Research meeting department of Public Health Erasmus University Medical Center, oral presentation "NELSON study"	2013	1 ECTS
(Inter)national conferences		
14th World Conference on Lung Cancer, Amsterdam, Netherlands	2011	24 hours
NELSON lung cancer screening symposium, Rotterdam, Netherlands	2011	8 hours
WEON Congres, Rotterdam, Netherlands	2012	16 hours
European meeting lung cancer screening trials	2012	12 hours
Lung cancer screening symposium, Leuven, Belgium	2013	24 hours
European Lung Cancer Conference, Lugano, Switserland	2013	24 hours

American Thoracic Society International Conference, Philadelphia, USA	2013	40 hours
World Conference on Lung Cancer, Sydney, Australia	2013	40 hours
Other		
Peer reviews for international medical journals		
International Journal of Cancer	2012	6 hours
Lung cancer	2012	6 hours
JAMA internal medicine	2013	4 hours
Health expectations	2013	2 hours
Thorax	2013	2 hours
Journal of Thoracic Oncology	2013	14 hours
Lung Cancer	2013	4 hours
Respiration	2013	6 hours
Journal of Thoracic Oncology	2014	8 hours
Expert review of Respiratory Medicine	2014	4 hours
Lung Cancer Management	2014	2 hours
2. Teaching		
'Medication safety' 3rd year medical students, Erasmus University, Rotterdam, Netherlands	2012	8 hours
Checking bachelor essays, 3rd year medical students, Erasmus University, Rotterdam, Netherlands	2012	60 hours
'Primary prevention in doctor's practice' 3rd year medical students, Erasmus University, Rotterdam, Netherlands	2013	12 hours
Checking bachelor essays, 3rd year medical students, Erasmus University, Rotterdam, Netherlands	2013	80 hours
Checking bachelor essays, 3rd year medical students, Erasmus University, Rotterdam, Netherlands	2014	80 hours
Total	**2011-2014**	**46.5 ECTS***

* 1 ECTS = 28 hours

Chapter 17

List of publications

- Scholten ET, de Jong PA, Jacobs C, van Ginneken B, van Riel S, Willemink MJ, Vliegenthart R, Oudkerk M, de Koning HJ, **Horeweg N**, Prokop M, Mali WPTM, Gietema HA. Interscan variation of semi-automated volumetry of subsolid pulmonary nodules. [submitted]
- Scholten ET, de Jong PA, de Hoop B, van Klaveren RJ, van Amelsvoort-van der Vorst S, Oudkerk M, Vliegenthart R, de Koning HJ, van der Aalst CM, Vernhout RM, Groen HJM, Lammers JWJ, van Ginneken B, Jacobs C, Mali WPTM, **Horeweg N**, Weenink C, Thunnissen E, Prokop M, Gietema HA. Towards a close CT monitoring approach for screen detected subsolid pulmonary nodules? *European Respiratory Journal;* [accepted].
- **Horeweg N**, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers JWJ, Nackaerts K, Vliegenthart R, ten Haaf K, Yousaf-Khan AU, Heuvelmans MA, Thunnissen E, Oudkerk M, Mali W, de Koning HJ. Lung cancer detection in the randomised NELSON trial: screening test performance and interval cancers. *Lancet Oncology* 2014 Nov 1;15(11).
- **Horeweg N***, van Rosmalen J*, Heuvelmans MA, van der Aalst CM, Vliegenthart R, Scholten ET, ten Haaf K, Nackaerts K, Lammers JWJ, Weenink C, Groen HJM, van Ooijen P, de Jong PA, de Bock GH, Mali W, de Koning HJ*, Oudkerk M*. HJ. Lung cancer probability in subjects with CT-detected pulmonary nodules: an analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncology* 2014 Nov 1;15(11).
- Scholten ET, Jacobs C, van Ginneken B, van Riel S, Vliegenthart R, Oudkerk M, de Koning HJ, **Horeweg N**, Prokop M, Gietema HA, Mali WPTM, de Jong PA. Detection and quantification of the solid component in pulmonary subsolid nodules by semiautomatic segmentation. *European Radiology* 2014; [accepted].
- Scholten ET*, **Horeweg N***, de Koning HJ, Vliegenthart R, Oudkerk M, Mali WPTM, de Jong PA. Computed tomographic characteristics of interval and post screen carcinomas in lung cancer screening. *European Radiology* 2014 Sep 4; Epub ahead of print.
- **Horeweg N**, de Koning HJ. The importance of screening for lung cancer. *Expert review of respiratory medicine* 2014 Aug 27; 1-18.
- Jett JR, **Horeweg N**, de Koning HJ. Early Detection and Radiologic Screening. Book chapter in: *Thoracic Oncology: The IASLC Multidisciplinary Approach* 2014 May.
- **Horeweg N**, van der Aalst CM, Vliegenthart R, Zhao YR, Xie X, Scholten ET, Mali W, Thunnissen E, Weenink C, Groen HJM, Lammers JWJ, Nackaerts K, van Rosmalen J, Oudkerk M, de Koning HJ. Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial. *European Respiratory Journal* 2013 Dec;42(6):1659-67.

- **Horeweg N**, de Koning HJ. Reply: Stage distribution of lung cancers detected by computed tomography screening in the NELSON trial. *American Journal of Respiratory and Critical Care Medicine* 2013 Oct 15;188(8):1035-6.
- **Horeweg N**, Nackaerts K, Oudkerk M, de Koning HJ. Low-dose computed tomography screening for lung cancer: results of the first screening round. *Journal of Comparative Effectiveness Research* 2013 Sep;2(5):433-6.
- **Horeweg N**, van der Aalst, Thunnissen E, Nackaerts K, Weenink C, Groen HJM, Lammers JWJ, Aerts JG, Scholten ET, van Rosmalen J, Mali W, Oudkerk M, de Koning HJ. Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *American Journal of Respiratory and Critical Care Medicine* 2013 Apr 15;187(8):848-54.
- **Horeweg N**, van Klaveren RJ, Groen HJM, Lammers JWJ, Weenink C, Nackaerts K, Mali W, Oudkerk M, de Koning HJ. Uniform and blinded cause of death verification in a lung cancer CT screening trial. *Lung Cancer* 2012 Sep;77(3):522-5.
- van't Westeinde SC, **Horeweg N**, de Leyn P, Groen HJM, Lammers JWJ, Weenink C, Nackaerts K, van Klaveren RJ. Complications following lung surgery in the Dutch-Belgian randomized lung cancer screening trial. *European Journal of Cardio-Thoracic Surgery* 2012 Sep;42(3):420-9.
- van 't Westeinde SC, **Horeweg N**, Vernhout RM, Groen HJM, Lammers JWJ, Weenink C, Nackaerts K, Oudkerk M, Mali W, Thunnissen E, de Koning HJ, van Klaveren RJ. The role of conventional bronchoscopy in the work-up of suspicious CT screen-detected pulmonary nodules. *Chest* 2012 Aug;142(2):377-84.

* equal contribution