Lung Cancer Screening:
Optimization through risk stratification

Kevin ten Haaf

Lung Cancer Screening: Optimization through risk stratification

ISBN: 978-94-6332-264-5

Thesis, Erasmus University

Cover illustration: Naomi Cornel

Print: GVO drukkers & vormgevers B.V.

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

Lung Cancer Screening: Optimization through risk stratification

Vroege opsporing van longkanker: optimalisatie door risicostratificatie

Proefschrift

Ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

Op gezag van de rector magnificus

prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 7 december 2017 om 11.30 uur

door

Kevin ten Haaf

geboren te Breda

Erasmus University Rotterdam

Promotiecommissie:

Promotor:	Prof.dr. H.J. de Koning
Overige leden:	Prof.dr. J.G.J.V. Aerts
	Prof.dr. M. Oudkerk
	Prof.dr. K.G.M. Moons

Content

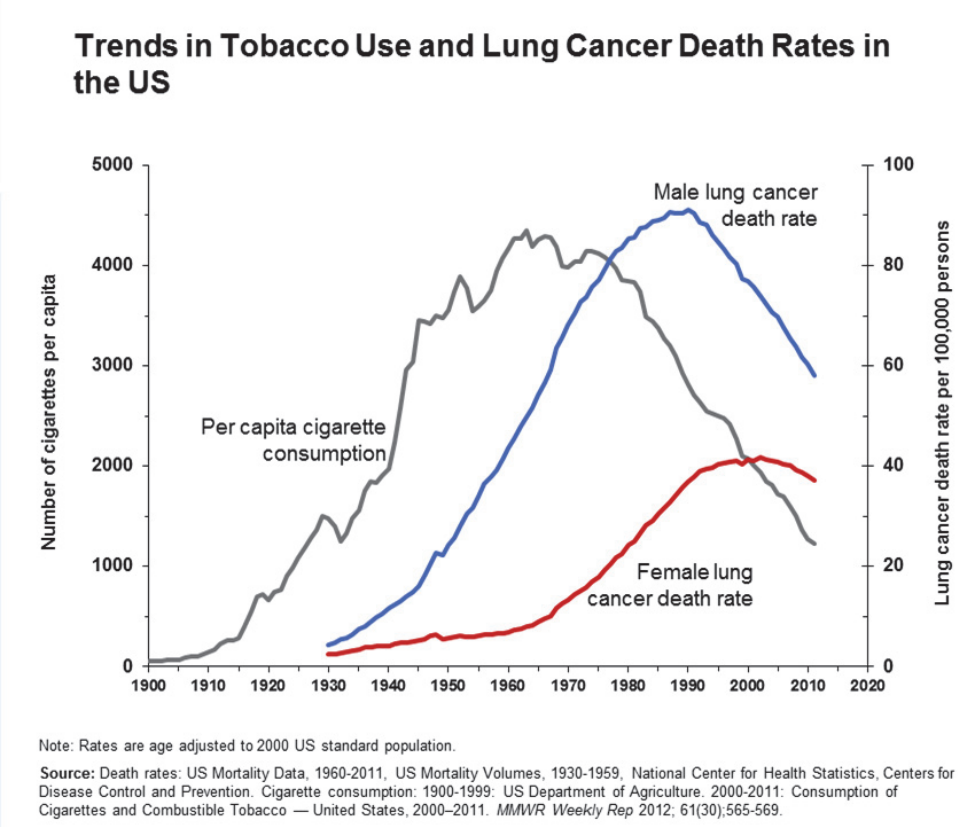
Part 3: Optimization through risk stratification

General introduction

Lung cancer epidemiology

At the start of the 20th century, lung cancer was considered a rare disease by many physicians.[1,2] However, with the increase in the consumption of tobacco, a stark increase in lung cancer incidence and mortality occurred, as shown for the United States (U.S.) in Figure 1.

Figure 1: Trends in cigarette consumption and male and female lung cancer death rates in the U.S.



Source: Cancer Risk Factors and Screening 2016; a presentation from the American Cancer Society, © 2016 American Cancer Society. Reproduced with kind permission from the American Cancer Society.[3]

The causal relation between smoking and lung cancer is well known in this day and age, but this has not always been the case. Initially, there was uncertainty whether there was a true increase in the incidence of lung cancer or whether this increase could be attributed to other factors such as improvements in diagnosis of the disease.[4,5] However, it became increasingly apparent that the rise in incidence of lung cancer could not be accounted for by improvements in diagnostic methods alone, warranting investigation into the cause(s) of this increase.[6,7] While an increasing number of retrospective (case-control) studies

suggested that smoking was associated with lung cancer, the existence of a causal relation between the two was disputed, in part due to the retrospective design of these studies.[8-12] Instead, some suggested tar dust from new roads and air pollution as potential causes.[13,14] However, when the initial results of the British Doctors' Study and the follow-up study of the American Cancer Society, some of the first prospective cohort studies ever performed, were published in 1954, the existence of a causal relation between tobacco smoking and lung cancer started to become more increasingly accepted among researchers.[15,16] In response to these studies, the tobacco industry aggressively promoted the suggestion that there was controversy on the causal relation between tobacco smoking and lung cancer within the scientific community, in order to cast doubts on the harmful effects of smoking.[17] The efforts of the tobacco industry were successful in influencing the perception of the general public: while 90% of respondents of a 1954 U.S. Gallup poll indicated that they had read or heard reports that "*cigarettes may be one of the causes of lung cancer*", only 41% indicated that they believed that "*cigarette smoking is one of the causes of lung cancer*".[18,19]

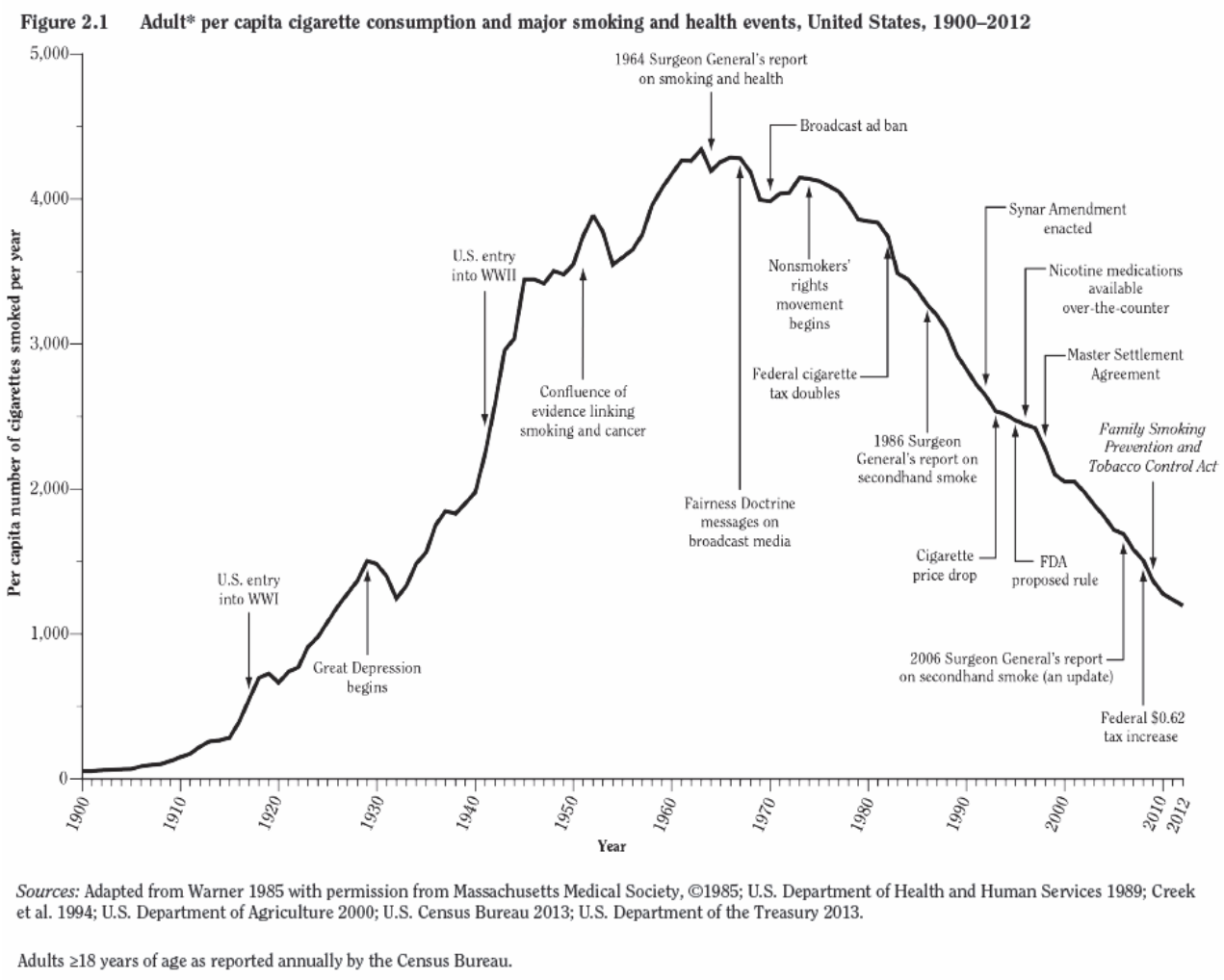
This changed with the release of the U.S. Surgeon General's report of 1964, which stated that "*Cigarette smoking is causally related to lung cancer in men; the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction*".[20] While previous Surgeon Generals and various medical associations worldwide had given similar statements, the 1964 report had a much greater impact.[21-24] The Surgeon General who initiated the report, Luther Terry, noted: "*The report hit the country like a bombshell. It was front page news and a lead story on every radio and television station in the United States and many abroad*".[24]

The U.S. Surgeon General's Report of 2014 reflected on the 50 years of progress since the original report.[23] Since the release of the 1964 report, the U.S. and other countries have implemented various tobacco control policies, such as tax increases on cigarettes, advertising restrictions and smoke-free air laws.[23,25-27] An overview of the implementation of some tobacco control policies and the per capita cigarette consumption in the U.S. is provided in Figure 2.

The 1964 Surgeon General's report and subsequent public health efforts also drastically altered the perception of the general public on the relationship between cigarettes and lung cancer: in a 1969 Gallup poll, 71% of respondents now answered that they believed that "*cigarette smoking is one of the causes of lung cancer*", and this further rose to 92% in 1999.[18,19]

However, despite decades of tobacco control policies and public awareness on the harmful effects of tobacco smoking, lung cancer still remains a major public health problem. In 2012, 1.8 million new cases and 1.6 million lung cancer deaths were estimated to have occurred worldwide.[28,29] Overall, lung cancer accounts for 13% of all cancer cases and 19% of all cancer related deaths, which makes it the leading cause of cancer related mortality.[28,29] Even though smoking prevalence is still decreasing, lung cancer and other smoking-related diseases are expected to remain a major public health problem worldwide for decades to come.[28]

Figure 2: Trends in cigarette consumption and major smoking and health events

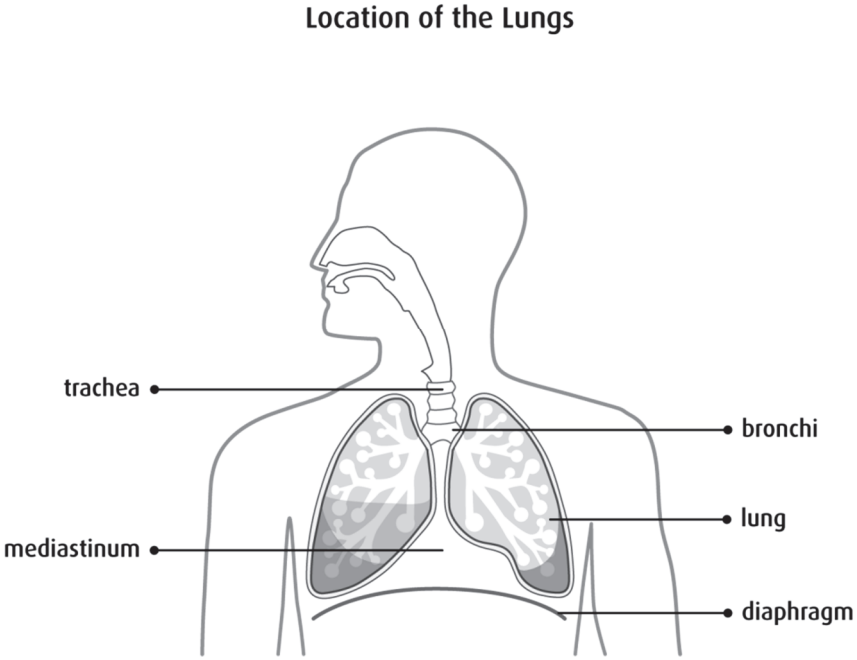


Source: 2014 Surgeon General's Report. Reused with kind permission from the U.S. Department of Health & Human Services.[23]

Etiology of lung cancer

The lungs are part of the body's respiratory system, located in the thoracic cavity of the chest, as shown in Figure 3. The main functions of the lungs are to 1) extract oxygen from inhaled air and transfer it into the bloodstream and 2) to release carbon dioxide from the bloodstream through exhalation.[30] Air is inhaled through the nose and mouth and moves through the trachea located in the neck and chest to the lungs.

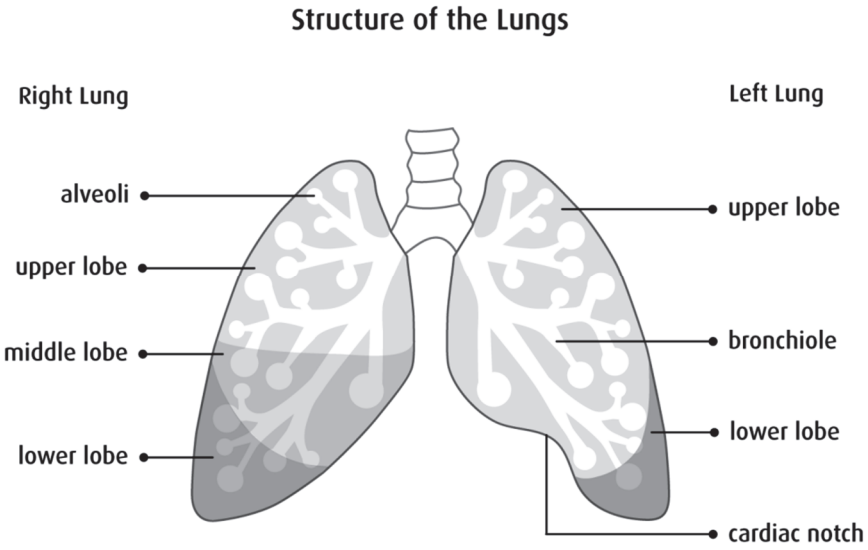
Figure 3: Location of the lungs



© 2017 Canadian Cancer Society. Reused with kind permission from the Canadian Cancer Society.[31]

There, it divides into two branches, the main bronchi, with each going to one of the lungs, as shown in Figure 4. The main bronchi further split out into separate branches, the smaller bronchi, throughout the lungs with each branch ending at a great number of tiny sacs: the alveoli. These alveoli are surrounded by blood vessels, allowing the transfer of oxygen and carbon dioxide between the lungs and the bloodstream. The lungs can be divided into a number of sections: lobes. The left lung can be divided into an upper lobe and a lower lobe; the heart is located next to the lower lobe of the left lobe, in a groove called the cardiac notch. The right lung is larger than the left lung and can be divided into three lobes: an upper, middle and lower lobe. The lungs themselves are covered by the pleura, a membrane which protects the lungs and allows them to move in the chest cavity.

Figure 4: Structure of the lungs



© 2017 Canadian Cancer Society. Reused with kind permission from the Canadian Cancer Society.[31]

Lung cancers are cancerous tumors that can develop within different areas of the lung. Lung cancers can be grouped into two main types: small cell lung cancers and non-small cell lung cancers.

Small cell lung cancers

Small cell lung cancers derive their name from their size compared to other cells in the lung area when observed under a microscope and account for approximately 13% of all lung cancers.[32] This type of lung cancer often occurs in the central area of the lung, near the hilum (where the bronchus enters the lung).[33] Small cell lung cancer is the most aggressive type of lung cancer and at the time of diagnosis, 60-70% of patients present with extensive disease, in which case the disease is too widespread to fit within a radiation field or has spread to other organs.[34]

Non-small cell lung cancers

Within the non-small cell lung cancers, two major groups can be distinguished: squamous cell carcinomas and adenocarcinomas.[35] Squamous cell carcinomas develop in the epithelium, a type of tissue that lines the passages of the lungs and currently account for

approximately 23% of all cases.[32,33] Like small cell lung cancers, this type of cancer generally occurs in the central areas of the lung.

Adenocarcinomas are malignant epithelial tumors which form in the mucus-secreting glands of the lung; some types of adenocarcinomas are mucus-producing.[33] These cancers are often found in the peripheral areas of the lung. Adenocarcinomas are currently the most common type of lung cancer, accounting for approximately 45% of all cases.[32,33,35-38]

Non-small cell lung cancers that cannot be distinguished as either squamous cell carcinomas or adenocarcinomas are composed of a wide range of different types of lung cancers. In this thesis, these cancers are grouped together and referred to as "other non-small cell lung cancers", accounting for the remaining 19% of all cases.[32]

Interventions to reduce lung cancer mortality

Primary prevention

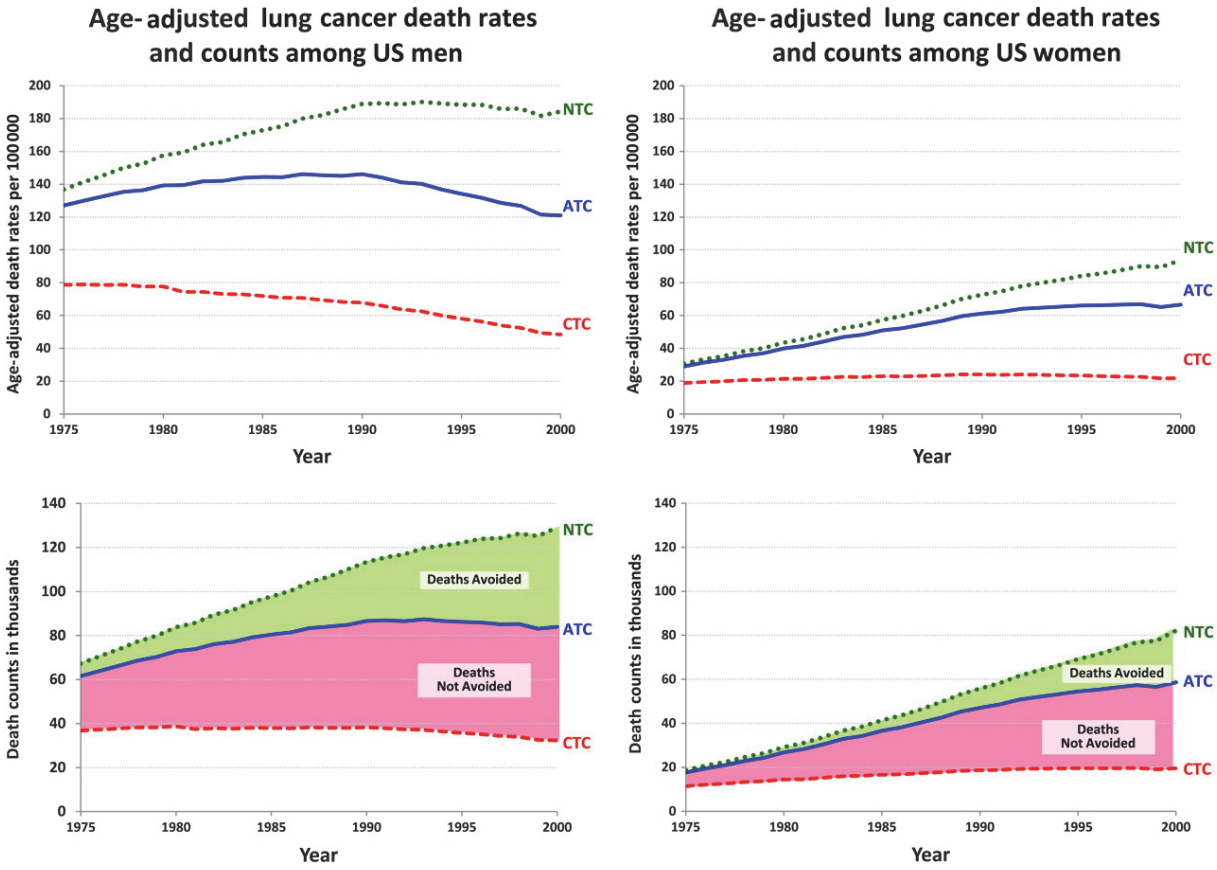
Smoking is estimated to account for 75-90% of all lung cancer cases.[39,40] Therefore, preventing the uptake of smoking in younger individuals would greatly reduce the incidence of lung cancer in the future, as few individuals start smoking after the age of 30.[41,42] For current smokers, smoking cessation is an effective way to reduce their risk for developing lung cancer; cessation at any age reduces the risk of developing lung cancer and death from other diseases.[43-45]

Since the publication of the 1950's Surgeon General's report, a number of smoking cessation policies have been implemented in the U.S. and worldwide, such as smoke-free air laws for work-places, restaurants and bars; smoking cessation counseling and prevention of smoking uptake among youths.[20,23,25-27] In particular, tax increases have been proven to be effective; a 1% price increase is suggested to reduce cigarette consumption by 0.3-0.5%.[46]

Modeling analyses estimate that changes in smoking behaviors in the U.S. (primarily due to tobacco control measures) have averted almost 800,000 lung cancer deaths in the period 1975-2000.[47] However, this only accounts for 32% of the lung cancer deaths in the U.S. in

this period that could have been averted if smoking in the U.S. had ceased in 1965, as shown in Figure 5.

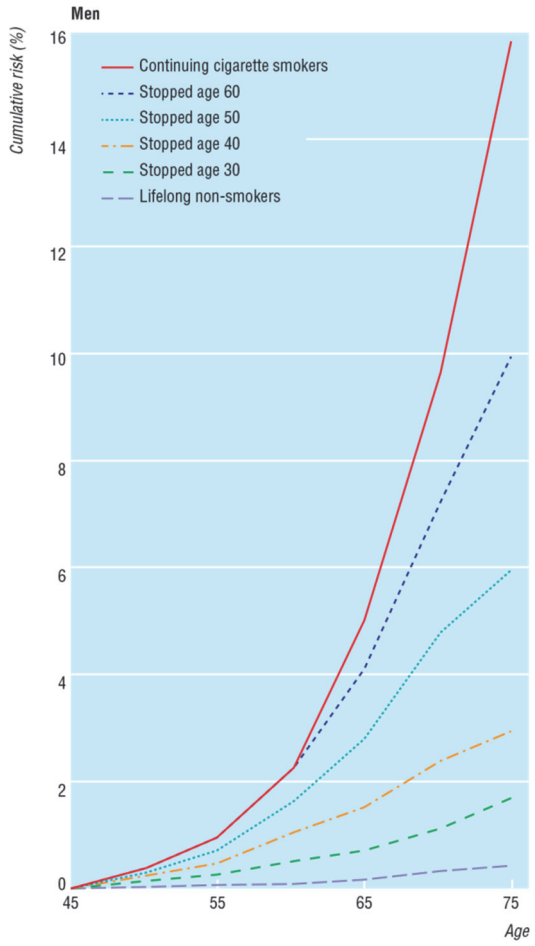
Figure 5: Lung cancer death rates in the United States by different scenarios of tobacco control, assuming: No Tobacco Control (NTC) measures were taken, Actual Tobacco Control (ATC) as occurred in reality and if smoking had been eliminated through Complete Tobacco Control (CTC)



Source: Moolgavkar, Journal of the National Cancer Institute , 2012. Reprinted with kind permission from Oxford University Press.[47]

Thus, smoking cessation and prevention of smoking initiation among youths is effective in preventing lung cancer (as well as other tobacco-related diseases) and further tobacco control measures should be encouraged. However, former smokers still remain at elevated risk for developing lung cancer compared to never-smokers for decades after smoking cessation, as shown in Figure 6.[43-45,48] Therefore, developments in treatment and other forms of prevention are essential to further reduce the burden of lung cancer.

Figure 6: Effects of stopping smoking at various ages on the cumulative risk (%) of death from lung cancer up to age 75, at death rates for men in the United Kingdom in 1990

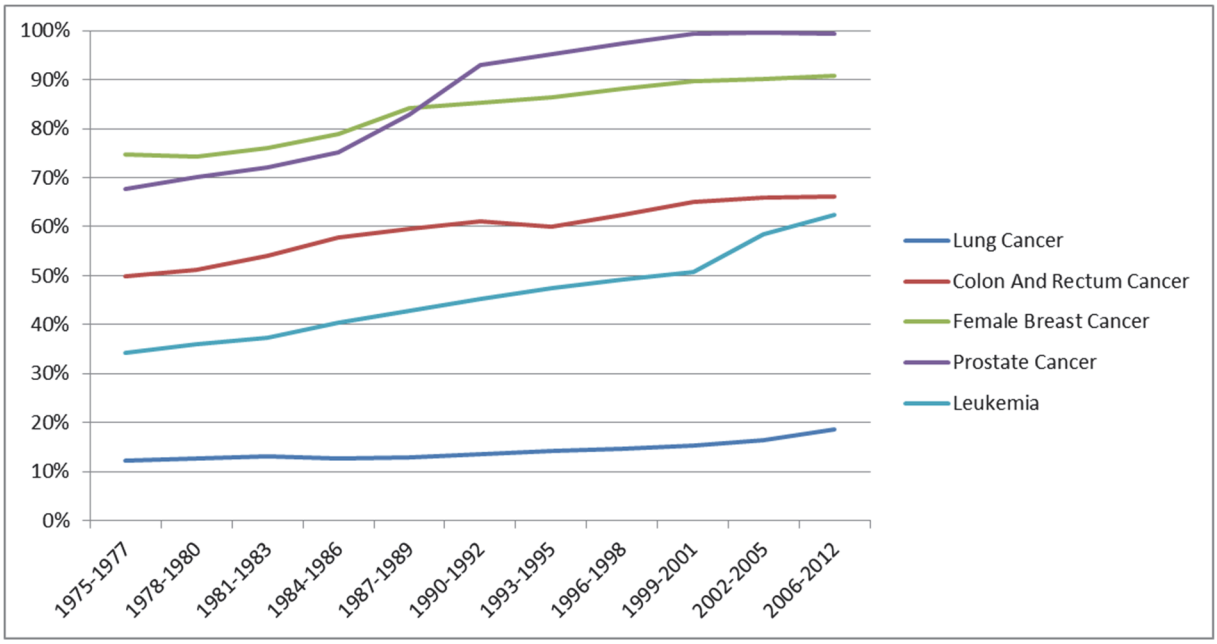


Reproduced from: Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies, Peto et al British Medical Journal, volume 321, issue 7257, pages 323-329, 2000, with kind permission from BMJ Publishing Group Ltd. [44]

Treatment

Oschner, DeBakey and Dixon wrote in 1948 that "*Primary cancer of the lung was considered a relatively rare and hopeless disease until about fourteen years ago. Since that time both of these concepts have been refuted*".[49] Lung cancer is indeed far from a rare disease at this moment in time, however, while the diagnosis of lung cancer may not lead to a hopeless prognosis, it is still often a pessimistic one. Although the overall five-year lung cancer survival rate in the U.S. in 2012 had improved by over 50% compared with 1975, it was still less than 19% [32] In contrast, current five-year survival rates for colorectal cancer and leukemia were over 60%, while those for female breast cancer and prostate cancer were over 90%, as shown in (Figure 7).

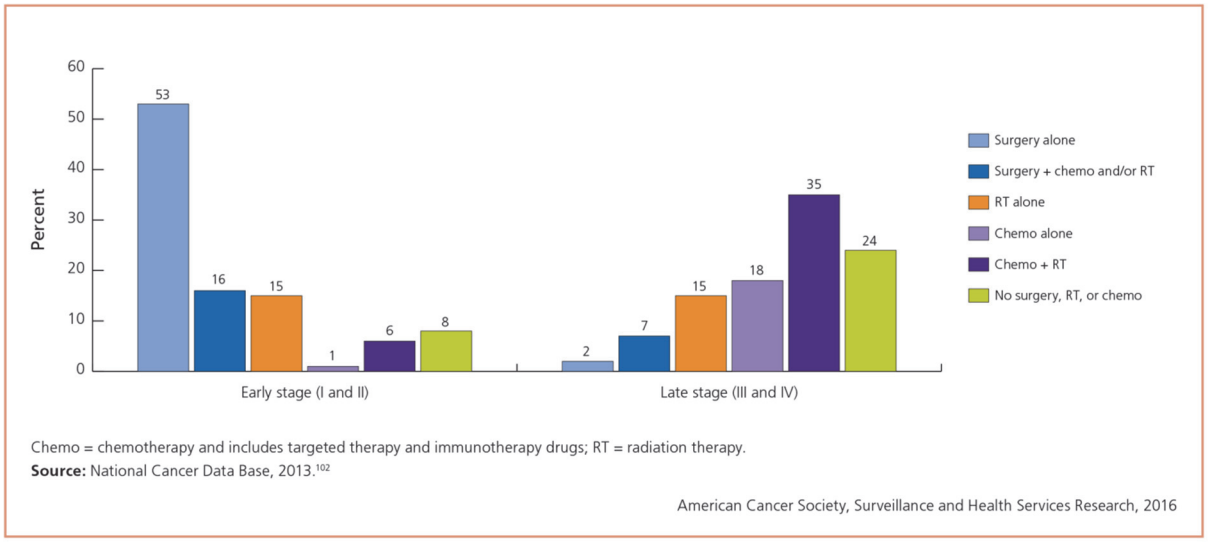
Figure 7: Relative 5-year survival for four major cancer sites in the United States over different time periods



Source: Howlader, 2016. Reproduced with kind permission from the U.S. National Cancer Institute.[32]

Suggested treatment for lung cancer differs between small cell and non-small cell lung cancers. Small-cell lung cancers are generally treated with a combination of radiotherapy and chemotherapy.[34] The treatment of non-small cell lung cancers is largely dependent on the stage of detection, as shown in Figure 8. Early stage (stages I/II) non-small cell lung cancers are generally treated with surgery, while late stage (stages III/IV) lung cancers are treated with chemotherapy and/or radiotherapy.[50]

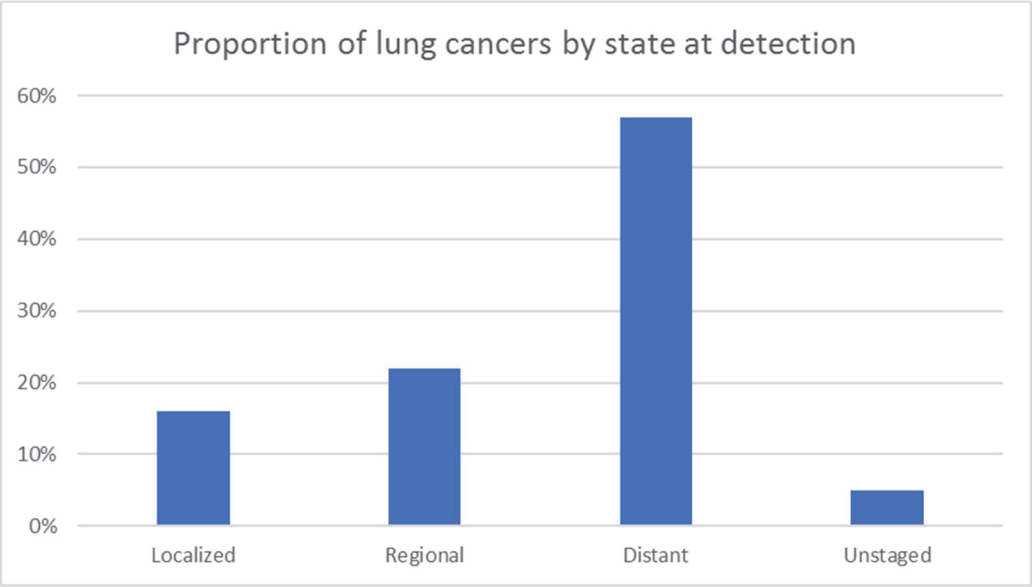
Figure 8: Treatment patterns for non-small cell lung cancer by stage in 2013



Source: American Cancer Society. Cancer Treatment & Survivorship Cancer Facts and Figures 2016-2017. Atlanta: American Cancer Society, Inc. Used with kind permission of the American Cancer Society. [50]

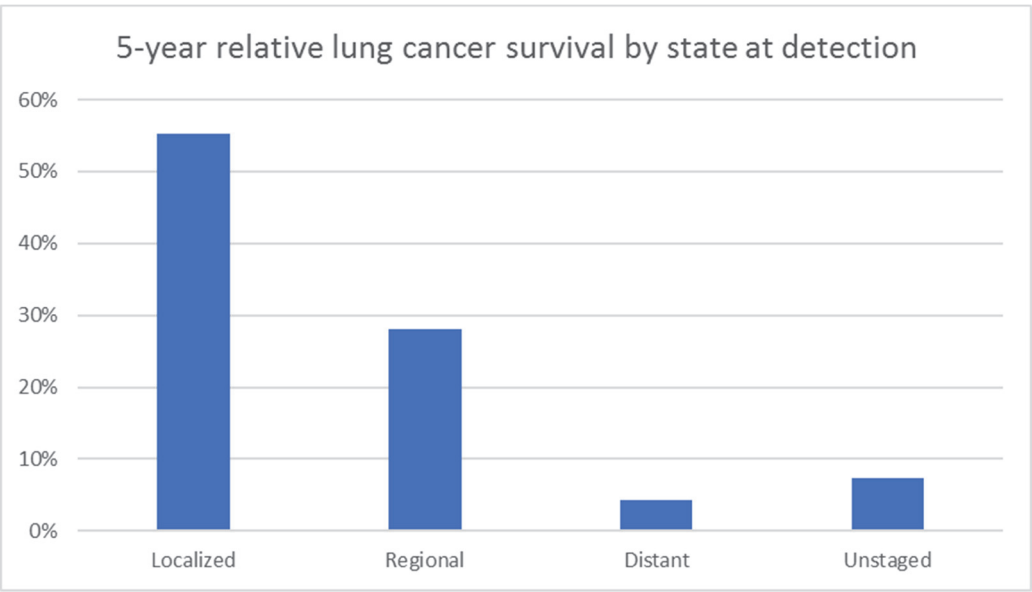
However, while the types of chemotherapy, radiotherapy and surgery have improved, little has changed in general treatment patterns since the 1950's.[51] In fact, Oschner, DeBakey and Dixon already favored surgical resection in 1948: "*The only curative treatment for cancer of the lung is surgical extirpation of the tumor-bearing lung and regional lymph nodes*".[49] Unfortunately, most lung cancers are detected at a late stage (over 50%), in which the cancer has spread to another organ or part of the body (metastasized, denoted as stage IV), as shown in Figure 9. At this point, curative surgery is generally not possible and as a result, the 5-year survival for lung cancers detected in this state is less than 10%, as shown in Figure 10.

Figure 9: Proportion of lung cancers by state at detection. Localized: the cancer is confined to the primary site. Regional: the cancer has spread to regional lymph nodes



Distant: the cancer has metastasized. Sources: Howlader, 2016 and SEER Cancer Stat Facts: Lung and Bronchus Cancer. Reproduced with kind permission from the U.S. National Cancer Institute.[32,52]

Figure 10: 5-year relative lung cancer survival by state at detection. Localized: the cancer is confined to the primary site



Regional: the cancer has spread to regional lymph nodes. Distant: the cancer has metastasized. Sources: Howlader, 2016 and SEER Cancer Stat Facts: Lung and Bronchus Cancer. Reproduced with kind permission from the U.S. National Cancer Institute.[32,52]

Screening

Given that curative treatment (generally surgery) is generally only possible for early stage lung cancers, the detection of lung cancer at an earlier stage could improve the potential for curative treatment. This was already noted by Oschner, DeBakey and Dixon: *"…it seems reasonable to believe that with greater awareness of the problem of pulmonary malignancy, and consequent increase in the proportion of cases that are diagnosed early, there should be considerable improvement in the survival rates of the disease*".[49] The detection of lung cancer at an earlier stage could potentially be achieved through screening: the examination of individuals in order to detect asymptomatic disease at an earlier stage with the aim to improve prognosis.[53]

The first investigations into lung cancer screening started in the 1960's and 1970's and investigated chest radiography screening, sometimes in combination with sputum cytology, but none of the randomized clinical trials from this period showed a benefit for chest radiography screening.[54-61] However, the trials from this period all suffered from methodological shortcomings, which caused uncertainty on the effectiveness of chest radiography screening to remain.[61]

The interest in screening for lung cancer diminished in the 1980's after the trials of the 1960's and 1970's failed to show any benefits. However, interest in screening for lung cancer was rekindled with the advent of computed tomography. A number of single-arm studies on computed tomography screening showed promising results for detecting lung cancer at an early stage, in particular the International Early Lung Cancer Action Project (I-ELCAP), in which 85% of the lung cancers detected through screening were found in stage I, with an estimated 10-year survival rate of 88%.[62-64]

However, improvements in survival rates do not necessarily translate to a mortality reduction, due to lead-time bias and length-time bias.[65] Lead-time bias refers to the case where the time of detection of the disease is advanced due to screening, but the time of death remains unaltered, artificially increasing the survival rates without extending the person's life. Length-time bias occurs due to screening favoring the detection of slower growing and less aggressive cancers, which have a better survival compared to more

aggressive tumors, again artificially increasing survival. Therefore, these single-arm studies could not provide an unbiased answer on the effectiveness of computed tomography screening.

The absence of definite answers on the effectiveness of chest radiography and computed tomography screening led to the inception of a number of randomized controlled trials in the 1990's and the 2000's, which will be discussed in the next section.

Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial

The lung component of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial was meant to provide a definite answer on the effect of chest radiography screening on lung cancer mortality.[66,67] In the lung component of this randomized controlled trial, 154,901 individuals aged 55 through 74 years were randomized to either four annual chest radiography screens (77,445 individuals) or usual care (no screening, 77,456 individuals) across 10 screening centers in the United States, between November 1993 and July 2001. There were no requirements with regards to smoking history to be eligible for participating in the PLCO, however, never-smoker participants randomized after April 1995 were not offered the fourth screen. After thirteen years of follow-up, no evidence of a reduction in lung cancer mortality was found for chest radiography screening compared to no screening.[68]

National Lung Screening Trial

The National Lung Screening Trial (NLST) was a randomized controlled screening trial, which compared lung cancer screening with low-dose computed tomography (CT) to chest radiography screening.[69] The NLST randomized 53,454 individuals between the ages of 55 and 74 to either three annual chest radiography screens (26,732 individuals) or three computed tomography screenings (26,722) from August 2002 through April 2004 across 33 medical centers in the United States. Participants were required to be current or former smokers (who quit less than 15 years) and to have had accumulated a minimum smoking exposure of 30 pack-years to be eligible for participation. The NLST found a significant relative reduction in lung cancer mortality of 20% for CT screening compared to chest

radiography screening, as well as an relative reduction in all-cause mortality of 6.7%.[70] Analyses in a subset of PLCO participants that met the eligibility criteria of the NLST also found no evidence of a reduction in lung cancer mortality for chest radiography screening compared to no screening.[68] Given the similarity between the NLST-eligible PLCO participants and NLST participants, the PLCO investigators suggested that the mortality reduction found for CT screening compared to chest radiography screening in the NLST should approximate the mortality reduction for CT screening compared to no screening.[68]

European lung cancer screening trials

In Europe, a number of lung cancer screening trials are currently ongoing or have recently published (preliminary) results: the Danish Lung Cancer Screening Trial (DLCST), the Multicentric Italian Lung Detection (MILD) study, the Italian Lung (ITALUNG) study, the Italian Detection And screening of early lung cancer with Novel imaging Technology (DANTE) trial, the German Lung Cancer Screening Intervention Trial (LUSI) and the United Kingdom Lung Cancer Screening (UKLS).[71-76] However, the Dutch-Belgian randomized lung cancer screening trial (NELSON) is the only randomized controlled trial in Europe that has sufficient power to show a lung cancer mortality reduction of 25% after ten years of follow-up.[77]

In NELSON, 15,792 individuals aged between 50 and 75 years were randomized to either four CT screens (at baseline, one year after baseline, three years after baseline and five-and-a-half years after baseline, 7,900 individuals) or regular care (no screening, 7,892 individuals) from December 2003 to July 2006 across four screening sites.[78] Participants were required to be current smokers or former smokers who had quit smoking less than 10 years ago and to have had accumulated a smoking history of >15 cigarettes a day for >25 years or >10 cigarettes a day for >30 years. One of the defining features of the NELSON trial is the nodule management protocol, which considers not only the diameter of detected lung nodules, but also their volume.[79] This protocol based on nodule volumetry has yielded a substantially lower proportion of false-positive screens compared to the protocol used in the NLST, which was based on nodule diameter (1.2% in the NELSON trial compared with 23.3% in the NLST).[70,80-82]

Microsimulation modeling of lung cancer screening

Randomized clinical trials are essential to provide information on the efficacy of screening. However, participants in randomized clinical trials may not be representative for the general population, which may limit the trial's generalizability. For example, a comparison between the participants of the NLST and the part of the general U.S. population that met the NLST's age and smoking history eligibility criteria suggested that the NLST participants were younger, higher educated and less likely to be current smokers.[83] Furthermore, within a (series of) randomized controlled trial(s), only a limited number of screening strategies can be considered. Finally, the follow-up duration of randomized clinical trials is often limited, which complicates deriving information on the long-term benefits and harms of screening.

Indeed, although the NLST showed that CT screening can reduce lung cancer mortality, many questions remained on whether and how to implement a lung cancer screening program, such as which persons to invite for screening, and the optimal screening regimen.[70] For example, further analyses of the NLST participants suggested that the 60% of participants at the highest risk for lung cancer mortality accounted for 88% of the prevented lung cancer deaths, while the 20% of participants at the lowest risk accounted for 1% of the prevented lung cancer deaths.[84] This finding suggests the selection of individuals for lung cancer screening could be optimized through risk stratification. Furthermore, whether to recommend the implementation of a lung cancer screening program requires a careful evaluation of the balance between the benefits, harms and costs of such a program. Concerns were raised on the harmful effects of false-positive screening results and the extent of overdiagnosis (the detection of a disease that would never have been detected, if screening had not occurred) resulting from the implementation of a lung cancer screening program, as well as the cost-effectiveness of such a program.[70] The NLST investigators indicated that these questions would be unlikely to be investigated by new randomized trials, but that (microsimulation) modeling could be used to address these questions.[70]

Modeling has proven to be a valuable tool in bridging the gap between published evidence and the information needed to develop clinical guidelines, for example for breast cancer and colorectal cancer screening.[85] The MIcrosimulation SCreening ANalysis (MISCAN) Lung

model is used to address a number of the research questions posed in this thesis. In brief, MISCAN-Lung simulates the life histories of individuals from birth until death, in the presence and absence of screening. The model simulates both ever- and never-smokers; for ever-smokers, a smoking history is generated which influences the probability of developing preclinical lung cancer and the probability of dying from other causes. Through comparing the life histories in the presence and absence of screening, MISCAN-Lung can estimate the costs and effects of screening. This information can be used to identify which groups of individuals should be invited for screening and identify the optimal screening regimen to implement.

MISCAN-Lung is one of the models used by the Lung Working Group of the Cancer Intervention and Surveillance Modeling Network (CISNET, www.cisnet.cancer.gov). CISNET is a consortium of research groups funded by the United States' National Cancer Institute (NCI) which uses statistical modeling to evaluate the impact of cancer control interventions with regards to prevention, screening and treatment. Within CISNET, research groups compare and contrast the results of independently developed models. This provides a framework for comparative modeling, which provides more robust estimates compared to studies based on a single model and is encouraged by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR).[86] Part of the work in this thesis was conducted through the lung cancer working group of the CISNET consortium.

An initial version of the MISCAN-Lung model was used to evaluate the impact of tobacco control on U.S. lung cancer mortality in a joint analysis with other CISNET modelers.[47,87] However, this version of the MISCAN-Lung model did not incorporate the information provided by the NLST and the PLCO. The NLST provides valuable information on the effectiveness of CT screening for lung cancer and the epidemiology of lung cancer in participants with an extensive smoking history.[70] While the PLCO did not show a significant reduction in lung cancer mortality, the trial provides insights in lung cancer epidemiology across participants with a wide variety of smoking histories.[68] In addition, the usual care arm of the PLCO provides information on lung cancer epidemiology in the absence of screening, in contrast to the NLST which did not have an unscreened control arm. Part of this thesis describes how the information provided by the NLST and the PLCO is incorporated in the

MISCAN-Lung model and how this information can be used to evaluate the long-term benefits and harms of lung cancer screening in populations other than those considered in randomized clinical trials.

Research questions and outline of this thesis

In this thesis, the effects of the implementation and optimization of lung cancer screening through risk stratification will be investigated. This thesis is divided into three parts, with each part addressing one of the following research questions:

1: What additional insights can models provide beyond the observed data of randomized controlled trials?

2: What are the long-term benefits and harms of lung cancer screening policies, and the potential barriers for the implementation of these policies?

3: How can risk stratification be used to optimize lung cancer screening policies?

The first research question, will be addressed in Chapters 1-3. Chapter 1 provides an overview of how different models can reproduce the results of randomized controlled trials on lung cancer screening. Chapter 2 indicates how modeling can use data from randomized clinical trials to derive information on the preclinical progression and screen-detectability of lung cancer. Chapter 3 will show how the information derived in the previous chapters can provide additional insights on the occurrence of overdiagnosis in lung cancer screening.

The second part of this thesis consists of Chapters 4-7, which will provide answers to the second research question. Chapter 4 will assess which lung cancer screening policies warrant a detailed investigation of their long-term benefits and harms. Chapter 5 investigates which of the policies identified in Chapter 4 provide the most advantageous balance between their long-term benefits and harms. Chapter 6 describes how a greater emphasis on overdiagnosis affects the preferred lung cancer screening policy. The potential barriers for the implementation of lung cancer screening policies are discussed in Chapter 7.

The final part of this thesis is formed by Chapters 8-12, which will provide insights to the final research question. Chapter 8 assesses the feasibility of lung cancer screening for never-smokers at elevated risk for lung cancer, by evaluating at which level of risk the benefits of lung cancer screening in this group outweigh the harms. Whether lung cancer screening can be implemented in a cost-effective manner is evaluated in Chapter 9. The role of lung cancer risk prediction models to optimize the identification of individuals eligible for lung cancer screening is investigated in Chapter 10. Finally, Chapter 11 investigates how pulmonary nodules detected through CT lung cancer screening can provide information on an individual's risk for developing lung cancer.

This thesis concludes with summary answers to and further discussion of these research questions, as well as directions for future research.

References

1. Adler IA. Classics in oncology: Primary malignant growths of the lung. *CA: A Cancer Journal for Clinicians* 1980; **30**(5): 295-301.
2. Bland-Sutton J. Tumours innocent and malignant: their clinical features and appropriate treatment. Philadelphia: Lea Brothers & Co.; 1893.
3. Cancer Risk Factors and Screening 2016; a presentation from the American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html.
4. Clemmesen J, Busk T. On the Apparent Increase in the Incidence of Lung Cancer in Denmark, 1931-1945. *British Journal of Cancer* 1947; **1**(3): 253-9.
5. Cancer of the Lung. *British Medical Journal* 1942; **1**(4247): 672-3.
6. Clemmesen J, Nielsen A, Jensen E. The Increase in Incidence of Carcinoma of the Lung in Denmark, 1931 to 1950. *British Journal of Cancer* 1953; **7**(1): 1-9.
7. Kennaway EL, Kennaway NM. A Further Study of the Incidence of Cancer of the Lung and Larynx. *British Journal of Cancer* 1947; **1**(3): 260-98.
8. Doll R, Hill AB. Smoking and Carcinoma of the Lung. *British Medical Journal* 1950; **2**(4682): 739.
9. Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma: A study of six hundred and eighty-four proved cases. *Journal of the American Medical Association* 1950; **143**(4): 329-36.
10. Parliament. *The Lancet* 1953; **261**(6761): 645.
11. Shera AG. "The Slandered Fragrant Weed". *British Medical Journal* 1953; **1**(4805): 338-.
12. Cancer of the Lung. *The Lancet* 1952; **260**(6736): 667-8.
13. Tighe HR. Cancer of the Lung and Smoking. *British Medical Journal* 1950; **2**(4670): 107-.
14. QUESTIONS AND COMMENTS. *British Medical Journal* 1948; **2**(4578): 667-8.
15. Doll R, Hill AB. The Mortality of Doctors in Relation to Their Smoking Habits. *British Medical Journal* 1954; **1**(4877): 1451-5.
16. Hammond E, Horn D. The relationship between human smoking habits and death rates: A follow-up study of 187,766 men. *Journal of the American Medical Association* 1954; **155**(15): 1316-28.
17. Brandt AM. Inventing Conflicts of Interest: A History of Tobacco Industry Tactics. *American Journal of Public Health* 2011; **102**(1): 63-71.
18. Saad L. Tobacco and smoking, August 15, 2002; http://www.gallup.com/poll/9910/tobacco-smoking.aspx#4; accessed: January 11, 2017.
19. Saad L. A half-century of polling on tobacco: most don't like smoking but tolerate it. Public Perspective 1998;9(5):1–4.
20. U.S. Department of Health, Education, and Welfare. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service. Washington: U.S. Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, 1964. PHS Publication No. 1103.
21. Royal College of Physicians. Smoking and Health: Summary and Report of the Royal College of Physicians of London on Smoking in Relation to Cancer of the Lung and Other Diseases. London: Pitman Medical Publishing Co., 1962.
22. Burney LE. Smoking and lung cancer: A statement of the public health service. *Journal of the American Medical Association* 1959; **171**(13): 1829-37.
23. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Published January 17, 2014. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, Ga.
24. Parascandola J. The Surgeons General and smoking. *Public Health Reports* 1997; **112**(5): 440-2.
25. World Health Organization. WHO report on the global tobacco epidemic, 2015: raising taxes on tobacco. WHO: Geneva; 2015.

26. Cokkinides V, Bandi P, McMahon C, Jemal A, Glynn T, Ward E. Tobacco Control in the United States—Recent Progress And Opportunities. *CA: A Cancer Journal for Clinicians* 2009; **59**(6): 352-65.
27. Cokkinides V, Bandi P, Ward E, Jemal A, Thun M. Progress and Opportunities in Tobacco Control. *CA: A Cancer Journal for Clinicians* 2006; **56**(3): 135-42.
28. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* 2015; **65**(2): 87-108.
29. Stewart BW, Wild CP. World cancer report 2014. 2014. http://public.eblib.com/choice/publicfullrecord.aspx?p=1978030.
30. Gray's anatomy for students (3rd ed.) Drake, Richard L.; Vogl, Wayne; Mitchell, Adam W.M. (2014). Edinburgh: Churchill Livingstone/Elsevier. pp. 167–174. ISBN 978-0-7020-5131-9.
31. Canadian Cancer Society, http://www.cancer.ca/en/cancer-information/cancer-type/lung/lung-cancer/the-lungs?region=on, accessed: 30-03-2017.
32. Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2013, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2013/, based on November 2015 SEER data submission, posted to the SEER web site, April 2016.
33. Travis WD, Brambilla E, Müller-Hermelink HK, Harris CC. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart. Lyon: IARC, 2004.
34. Jackman DM, Johnson BE. Small-cell lung cancer. *The Lancet*; **366**(9494): 1385-96.
35. Devesa SS, Bray F, Vizcaino AP, Parkin DM. International lung cancer trends by histologic type: Male:Female differences diminishing and adenocarcinoma rates rising. *International Journal of Cancer* 2005; **117**(2): 294-9.
36. Meza R, Meernik C, Jeon J, Cote ML. Lung Cancer Incidence Trends by Gender, Race and Histology in the United States, 1973–2010. *PLOS ONE* 2015; **10**(3): e0121323.
37. Thun MJ, Lally CA, Calle EE, Heath JCW, Flannery JT, Flanders WD. Cigarette Smoking and Changes in the Histopathology of Lung Cancer. *Journal of the National Cancer Institute* 1997; **89**(21): 1580-6.
38. Lewis DR, Check DP, Caporaso NE, Travis WD, Devesa SS. U.S. lung cancer trends by histologic type. *Cancer* 2014; **120**(18): 2883-92.
39. Koo LC, Ho JHC. Worldwide Epidemiological Patterns of Lung Cancer in Nonsmokers. *International Journal of Epidemiology* 1990; **19**(Supplement_1): S14-S23.
40. Parkin DM, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. *CA: A Cancer Journal for Clinicians* 2005; **55**(2): 74-108.
41. Anderson CM, Burns DM, Dodd KW, Feuer EJ. Chapter 2: Birth-Cohort-Specific Estimates of Smoking Behaviors for the U.S. Population. *Risk Analysis* 2012; **32**: S14-S24.
42. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort–Specific Smoking Histories, 1965–2009. *American Journal of Preventive Medicine* 2014; **46**(2): e31-e7.
43. Thun MJ, Carter BD, Feskanich D, et al. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine* 2013; **368**(4): 351-64.
44. Jha P, Ramasundarahettige C, Landsman V, et al. 21st-Century Hazards of Smoking and Benefits of Cessation in the United States. *New England Journal of Medicine* 2013; **368**(4): 341-50.
45. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *British Medical Journal* 2000; **321**(7257): 323.
46. U.S. Department of Health and Human Services. Reducing Tobacco Use: A Report of the Surgeon General. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2000.
47. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of Reduced Tobacco Smoking on Lung Cancer Mortality in the United States During 1975–2000. *Journal of the National Cancer Institute* 2012; **104**(7): 541-8.

48. Pinsky PF, Zhu CS, Kramer BS. Lung cancer risk by years since quitting in 30+ pack-year smokers. *Journal of Medical Screening* 2015; **22**(3): 151-7.
49. Ochsner A, DeBakey M, Dixon L. Carcinoma of the Lung. *Postgraduate Medicine* 1948; **3**(6): 427-40.
50. American Cancer Society. Cancer Treatment & Survivorship Facts & Figures 2016-2017. Atlanta: American Cancer Society; 2016. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures.html.
51. Stephen GS, Gerard AS. One Hundred Years of Lung Cancer. *American Journal of Respiratory and Critical Care Medicine* 2005; **172**(5): 523-9.
52. SEER Cancer Stat Facts: Lung and Bronchus Cancer. National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/statfacts/html/lungb.html. Accessed: 30-03-2017.
53. Wilson JMG, Jungner G. Principles and practices of screening for disease. Geneva, Switzerland: World Health Organization; 1968. Report No.: Public Health Papers No. 34.
54. Brett GZ. The value of lung cancer detection by six-monthly chest radiographs. *Thorax* 1968; **23**(4): 414.
55. Friedman GD, Collen MF, Fireman BH. Multiphasic Health Checkup Evaluation: a 16-year follow-up. *Journa of Chronic Diseases*. 1986;**39**:453-63. [PMID: 3711252].
56. Melamed MR, Flehinger BJ, Zaman MB, Heelan RT, Perchick WA, Martini N. Screening for Early Lung Cancer: Results of the Memorial Sloan-Kettering Study in New York. *Chest* 1984; **86**(1): 44-53.
57. Levin ML, Tockman MS, Frost JK, Ball WC. Lung Cancer Mortality in Males Screened by Chest X-ray and Cytologic Sputum Examination: A Preliminary Report. In: Band PR, ed. Early Detection and Localization of Lung Tumors in High Risk Groups. Berlin, Heidelberg: Springer Berlin Heidelberg; 1982: 138-46.
58. Kubik A, Parkin DM, Khlat M, Erban J, Polak J, Adamec M. Lack of benefit from semi-annual screening for cancer of the lung: Follow-up report of a randomized controlled trial on a population of high-risk males in czechoslovakia. *International Journal of Cancer* 1990; **45**(1): 26-33.
59. Wilde J. A 10 year follow-up of semi-annual screening for early detection of lung cancer in the Erfurt County, GDR. *European Respiratory Journal* 1989; **2**(7): 656.
60. Marcus PM, Bergstralh EJ, Fagerstrom RM, et al. Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up. *Journal of the National Cancer Institute* 2000; **92**(16): 1308-16.
61. Humphrey LL, Teutsch S, Johnson M. Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: An update for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 2004; **140**(9): 740-53.
62. The International Early Lung Cancer Action Program I. Survival of Patients with Stage I Lung Cancer Detected on CT Screening. *New England Journal of Medicine* 2006; **355**(17): 1763-71.
63. Stephen J S, James R J, Jeff A S, et al. Screening for Lung Cancer with Low-Dose Spiral Computed Tomography. *American Journal of Respiratory and Critical Care Medicine* 2002; **165**(4): 508-13.
64. 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. *British Journal of Cancer* 2001; **84**(1): 25-32.
65. Zelen M, Feinleib M. On the Theory of Screening for Chronic Diseases. *Biometrika* 1969; **56**(3): 601-14.
66. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial. *Controlled Clinical Trials* 2000; **21**(6): 273S-309S.
67. Hocking WG, Hu P, Oken MM, et al. Lung Cancer Screening in the Randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *Journal of the National Cancer Institute* 2010; **102**(10): 722-31.
68. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
69. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and Study Design. *Radiology* 2011; **258**(1): 243-53.

70. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
71. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomized Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; **67**(4): 296-301.
72. 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. *European Journal of Cancer Prevention* 2012; **21**(3): 308-15.
73. Pegna AL, Picozzi G, Falaschi F, et al. Four-Year Results of Low-Dose CT Screening and Nodule Management in the ITALUNG Trial. *Journal of Thoracic Oncology* 2013; **8**(7): 866-75.
74. Infante M, Cavuto S, Lutman FR, et al. Long-Term Follow-up Results of the DANTE Trial, a Randomized Study of Lung Cancer Screening with Spiral Computed Tomography. *American Journal of Respiratory and Critical Care Medicine* 2015; **191**(10): 1166-75.
75. Becker N, Motsch E, Gross ML, et al. Randomized Study on Early Detection of Lung Cancer with MSCT in Germany: Results of the First 3 Years of Follow-up After Randomization. *Journal of Thoracic Oncology* 2015; **10**(6): 890-6.
76. Field JK, Duffy SW, Baldwin DR, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax* 2016; **71**(2): 161-70.
77. 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 randomized lung cancer multi-slice CT screening trial (NELSON). *International Journal of Cancer* 2007; **120**(4): 868-74.
78. Horeweg N, Scholten ET, de Jong PA, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. *The Lancet Oncology* 2014; **15**(12): 1342-50.
79. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomized lung cancer screening trial. *Lung Cancer* 2006; **54**(2): 177-84.
80. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of Lung Nodules Detected by Volume CT Scanning. *New England Journal of Medicine* 2009; **361**(23): 2221-9.
81. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *European Respiratory Journal* 2013; **42**(6): 1659.
82. Yousaf-Khan U, van der Aalst C, de Jong PA, et al. Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval. *Thorax* 2017;**72**(1):48-56.
83. Aberle DR, Adams AM, Berg CD, et al. Baseline Characteristics of Participants in the Randomized National Lung Screening Trial. *Journal of the National Cancer Institute* 2010; **102**(23): 1771-9.
84. Kovalchik SA, Tammemägi M, Berg CD, et al. Targeting of Low-Dose CT Screening According to the Risk of Lung-Cancer Death. *New England Journal of Medicine* 2013; **369**(3): 245-54.
85. Habbema JF, Wilt TJ, Etzioni R, et al. Models in the development of clinical practice guidelines. *Annals of Internal Medicine* 2014; **161**(11): 812-8.
86. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices 2014;Modeling Studies. *Value in Health*; **6**(1): 9-17.
87. Schultz FW, Boer R, de Koning HJ. Chapter 7: Description of MISCAN-Lung, the Erasmus MC Lung Cancer Microsimulation Model for Evaluating Cancer Control Interventions. *Risk Analysis* 2012; **32**: S85-S98.

Part 1: Development of a lung cancer screening model

Chapter 1

Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials

Rafael Meza, Kevin ten Haaf, Chung Yin Kong, Ayca Erdogan, William C. Black, Martin C. Tammemägi, Sung Eun Choi, Jihyoun Jeon, Summer S. Han, Vidit Munshi, Joost van Rosmalen, Paul Pinsky, Pamela M. McMahon, Harry J. de Koning, Eric J. Feuer, William D. Hazelton, Sylvia K. Plevritis

Published as:

Meza R, ten Haaf K, Kong CY, et al.

Cancer 2014; **120**(11): 1713-24.

Abstract

BACKGROUND: The National Lung Screening Trial (NLST) demonstrated that low-dose computed tomography screening is an effective way of reducing lung cancer (LC) mortality. However, optimal screening strategies have not been determined to date and it is uncertain whether lighter smokers than those examined in the NLST may also benefit from screening. To address these questions, it is necessary to first develop LC natural history models that can reproduce NLST outcomes and simulate screening programs at the population level.

METHODS: Five independent LC screening models were developed using common inputs and calibration targets derived from the NLST and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO). Imputation of missing information regarding smoking, histology, and stage of disease for a small percentage of individuals and diagnosed LCs in both trials was performed. Models were calibrated to LC incidence, mortality, or both outcomes simultaneously.

RESULTS: Initially, all models were calibrated to the NLST and validated against PLCO. Models were found to validate well against individuals in PLCO who would have been eligible for the NLST. However, all models required further calibration to PLCO to adequately capture LC outcomes in PLCO never-smokers and light smokers. Final versions of all models produced incidence and mortality outcomes in the presence and absence of screening that were consistent with both trials.

CONCLUSIONS: The authors developed 5 distinct LC screening simulation models based on the evidence in the NLST and PLCO. The results of their analyses demonstrated that the NLST and PLCO have produced consistent results. The resulting models can be important tools to generate additional evidence to determine the effectiveness of lung cancer screening strategies using low-dose computed tomography.

Introduction

The National Lung Screening Trial (NLST) found a significant lung cancer (LC) mortality reduction in its low-dose computed tomography (CT) screening arm in comparison with its chest radiography (CXR) screening arm, suggesting that screening heavy smokers with low-dose CT can be effective in the early detection of LC.[1] Meanwhile, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) found no statistical difference in LC mortality when comparing a no-screen control arm versus a CXR screening arm.[2] Consequently, several health policy groups have made recommendations endorsing low-dose CT for LC screening based on the NLST entry criteria and LC screening programs are being established across the United States.[3]

However, there is still uncertainty regarding the optimal screening strategies because the NLST only evaluated the impact of 3 consecutive annual screens among current and former smokers aged 55 years to 74 years at the time of enrollment with an exposure of at least 30 pack-years and with ≤ 15 years since quitting. It is unknown whether current and former smokers with lower levels of exposure would also benefit from screening. Furthermore, screening effectiveness may vary by sex, number of screens, and periodicity. In the absence of results from other randomized control trials (RCTs) evaluating these questions, mathematical modeling of the natural history of LC may be the only approach to integrate available evidence and estimate the effectiveness and cost-effectiveness of different LC screening strategies in the general population.[3,4]

Mathematical models of cancer natural history have been shown to be valuable in assessing and determining optimal cancer prevention and control strategies. Recent examples include analyses of the impact of tobacco control on LC mortality rates, comparative studies assessing the effects of different screening modalities in patients with colorectal cancer, cost-effectiveness analyses of breast cancer screening strategies, and studies evaluating the impact of prostate-specific antigen screening in reducing prostate cancer rates.[5-9] All these examples used a comparative modeling framework by which researchers across institutions can directly compare and contrast results from distinct models.[10-12] The conclusions arising from comparative modeling analyses are more robust and reliable than single-model studies

and this approach has been cited as an example of good modeling practices.[13] To estimate the potential impact of LC screening at the U.S. population level, a consortium of National Cancer Institute (NCI)-sponsored investigators, the Cancer Intervention and Surveillance Modeling Network (CISNET; cisnet.cancer.gov), developed 5 independent natural history models of LC and screening. In the current study, we describe the models' development and calibration approach to the NLST and PLCO, the common shared inputs and calibration targets, and the differences and similarities between models. We compared model predictions versus observed trial outcomes and highlighted the advantages and challenges of developing natural history models based on large-scale RCTs.

Methods

Data

De-identified data from all NLST and PLCO participants were provided to CISNET after obtaining Institutional Review Board approvals from each institution. These data included smoking history variables such as the age at the start of smoking, the average number of cigarettes smoked per day (CPDs), and the age at quitting for former smokers. Screening variables included the individual's age at entry into the study and, for screened individuals, age at each screen, outcomes of each screen, and the follow-up procedures for positive screens. For each individual, the age at death or censoring and (if applicable) the cause of death were available. For individuals diagnosed with LC, the age at diagnosis, LC histology, and LC stage (according the 6th edition of the American Joint Committee on Cancer) were provided, as well as information regarding the screen associated with the LC diagnosis for screen-detected cancers.

NLST

The NLST was a RCT that compared the impact of low-dose CT versus CXR screening on LC mortality. From August 2002 through April 2004, a total of 53,454 individuals aged 55 years to 74 years were recruited; follow-up occurred through December 31, 2009. Entry criteria included a minimum exposure of 30 pack-years and ≤ 15 years since quitting for former smokers. Individuals in both screening arms received up to 3 annual screens. The trial found a 20% LC mortality reduction in the low-dose CT versus the CXR arm.[1] A small percentage of

LC cases (50 cases; 2.4%) had missing histology and/or stage information. To complete the missing data, a multistep imputation procedure based on observed histology and stage distributions, tumor sizes, and expert opinion was conducted. More details are provided in the supplementary material of this Chapter, in the section: "Imputation algorithm for missing data in NLST lung cancer subjects". Final analyses included data from 53,342 individuals, due to the exclusion of 112 subjects who died or were diagnosed with LC before the first screen (110 patients) or those with missing smoking information (age at start and/or time since quitting).

PLCO

The PLCO was a RCT that compared the impact of CXR screening (intervention arm) versus usual care (no-screening control arm) on LC mortality. The trial recruited 154,901 individuals aged 55 years to 74 years between November 1993 and July 2001. Participants were followed through December 31, 2009 or for 13 years from the time of enrollment, whichever came first. No minimum smoking exposure was required to enroll. Individuals in the intervention arm received up to 4 annual CXR screens. The study found no difference in LC mortality between the intervention and control arms.[2] Contamination (CXR screening) in the control arm was limited (11% contamination rate).[2] Additional smoking variables came from a supplemental questionnaire implemented toward the middle of the trial. Missing baseline data regarding the age at the start of smoking or CPDs for ever-smokers were imputed according to the corresponding U.S. distributions by birth cohort and age. Final analyses included data from 148,025 individuals, after the exclusion of individuals with missing baseline smoking status or (if applicable) age at time of quitting. More details are provided in the supplementary material of this Chapter, in the section: "Imputation of missing smoking data for PLCO participants".

Models

Models were developed by investigators at 5 institutions: Erasmus Medical Center (model E), Fred Hutchinson Cancer Research Center (model F), Massachusetts General Hospital (model M), University of Michigan (model U), and Stanford University (model S). The models were developed independently but the groups collaborated to develop common inputs and define standardized analyses. Below we provide a description of the five models. Additional

details are provided in the supplementary material of this Chapter, in the section: “Model Descriptions”.

Smoking dose-response module

All models simulate individual LC natural history and include a dose-response module that translates personal cigarette exposure to LC risk. This smoking dose-response module can be used to simulate age-specific LC outcomes given an individual’s smoking history.[5] Model M uses as its dose-response module a probabilistic LC risk model previously calibrated to Surveillance, Epidemiology, and End Results (SEER) and U.S. LC data and recalibrated to the NLST and PLCO, whereas all other groups use multistage carcinogenesis models.[14-18] Both multistage and probabilistic models have been used extensively to investigate the effects of smoking on LC risk.[5,12,16,17,19-21] Model E uses a multistage model based on the Nurses’ Health Study (NHS) and Health Professionals Follow-Up Study (HPFS).[16] Model S uses a modified version of this model. Model U uses a LC multistage model by histology, also calibrated to the NHS/HPFS. Model F uses a multistage model calibrated to the NLST and PLCO. Three models (models F, M, and U) use histology-specific smoking dose-response modules, and three models (models E, F, and M) recalibrated their smoking dose-response to the NLST and PLCO. More details are given in Table 1 and in the supplementary material of this Chapter, in the section: “Model Descriptions”.

All models are capable of accommodating detailed individual level smoking histories, including temporal factors such as age at start, age at cessation, and age-specific changes in CPDs. The variability across dose-response modules reflects the modelers’ judgment regarding the best available data and approaches to capture the complex relationship between smoking and LC. The NHS and HPFS are arguably the best prospective cohorts with which to investigate smoking related LC. They have >30 years and >20 years of follow-up, respectively, and collect smoking information every 2 years. However, their LC histology information is much less comprehensive than that of the NLST and PLCO, and staging information was not available. The NLST and PLCO are excellent data sources with thorough information available regarding LC histology and staging, but have more limited follow-up and less extensive smoking data than NHS/HPFS. In addition, the NLST includes only ever-

smokers and individuals in both arms were screened for LC. Approximately one-half of individuals in PLCO were also screened.

Histology distribution

Three models (models F, M, and U) have smoking dose-response modules that are histology specific. In these models, the LC histology distribution is a model outcome that depends on the dose-response module and the participants' smoking histories. Two models (models E and S) have smoking dose-response modules that are not histology-specific, and therefore they calibrated their histology to the NLST and PLCO. Histology categories varied by model (Table 1). Differences in histology categorization across models are due partly to differences in dose-response modules, which are based on different data sets that vary in their LC histology classifications (NHS/HPFS, NLST/PLCO, and SEER). However, they are also due to variations in model structure, and the modelers' judgment regarding the histology detail needed to characterize screening efficacy.

Table 1: Model Comparison

	Model E	Model F	Model U	Model M	Model S
Central smoking dose-response model	Two-stage clonal expansion model (TSCE)	Longitudinal multistage observation by histology	Multistage clonal expansion model by histology	Probabilistic by histology	TSCE
Central dose-response parameter calibration	NHS/HPFS,SEER,NLST, PLCO	NLST, PLCO, PLuSS CT,CARET	NHS/HPFS	SEER,NLST,PLCO	NHS/HPFS
Histological types	Adenocarcinoma, squamous, small cell (SCLC), other non-small cell (ONSCLC)	Adenocarcinoma, large-cell, squamous, BAC, ONSCLC, SCLC	Adenocarcinoma, ONSCLC, SCLC	Adenocarcinoma, BAC, large-cell, squamous, SCLC, and other	Adenocarcinoma, large-cell, squamous, SCLC
LC stages	IA,IB,II,IIIA,IIIB,IV	IA1,IA2,IB,II,IIIA,IIIB,IV	IA1,IA2,IB,II,IIIA,IIIB,IV	IA1,IA2,IB,II,IIIA,IIIB,IV	Early (I–II), Advanced (III–IV)
Stage progression model	Markov state-transition by histology	Based on tumor size and presence of metastasis	Markov state-transition by histology and gender; rates proportional to tumor size	Based on tumor volume and metastatic burden	Based on tumor volume and metastatic burden
LC survival	Based on SEER-17 2004–2008 survival	Based on NLST and PLCO	By gender, histology, stage and age at dx. Based on SEER-17 2004–2008 survival	Calibrated to SEER-17 1973–2008 survival	Based on SEER-17 1988–2003 survival
OC Mortality	U.S. rates (NCI Smoking History Generator)	As observed in NLST and PLCO	Gompertz-model of OCM calibrated to each trial	Cox-model of OCM calibrated to each trial	Gompertz model of OCM based on NLST
Calibration method	Nelder-Mead optimization of likelihood-based deviance criterion	Maximum likelihood approach	MCMC and Nelder-Mead simplex	Simulated annealing based on weighted-sum total deviance	Nelder-Mead simplex for Natural History Model calibration to SEER, and multi-dimensional grid search for calibration to trials
Data sources used for calibration	NLST; PLCO; SEER-17 2004–2008 incidence by age, stage, histology; NHS/HPFS	NLST; PLCO; originally fitted to PLuSS CT, CARET	NLST; NHS/HPFS LC incidence by histology; SEER LC survival by gender, age, histology and stage	NLST; SEER 1990–2000 incidence by age, stage, histology; survival by stage; Mayo CT; LSS	NLST; PLCO; NHS/HPFS LC incidence, SEER 1988–2003 survival by histology and gender
Number of parameters estimated by calibration	110	90	50	53	13 in natural history, 8 for calibration
Screening sensitivity model	By stage and histology	By size (number of cells), histology and gender	By size (number of cells), histology and gender	By size (mm) and location in lung (central/peripheral)	By size (mm) and histology
Screening effectiveness mechanism	Cure model	Combination cure model and stage-shift	Stage-shift model, with adjustments for age	Not stage-shift model	Not stage-shift model
Positive Nodule Follow-up algorithm	Implicit	Implicit based on NLST follow-up rates	Implicit based on NLST follow-up rates	Explicit. Based on size at diagnosis and smoking history. LCs diagnosed on follow-up are categorized as 'non-screen detected'	Explicit

Abbreviations: BAC, bronchioloalveolar carcinoma; CARET, Carotene and Retinol Efficacy Trial; CT, computed tomography; HPFS, Health Professionals' Follow-up Study; LC, lung cancer; LSS, Lung Screening Study; MCMC, Markov chain Monte Carlo; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model S, Stanford University; model U, University of Michigan; NCI, National Cancer Institute; NHS, Nurses' Health Study; NLST, National Lung Screening Trial; OC, other-cause; ONSCLC, other non-small cell lung cancer; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PLuSS, Pittsburgh Lung Screening Study; SCLC, small cell lung cancer; SEER, Surveillance, Epidemiology, and End Results; TSCE, Two-stage clonal expansion model.

Stage progression

All models assume that stage progression rates vary by sex and histology. Models E and U use Markov state transition processes to model stage progression.[22] Model U further assumes that the progression rate at each stage is dependent on tumor size (cell number). Models F, M, and S model stage as a function of tumor size and the presence or absence of metastasis. Variability in stage categorization (Table 1) is due to the underlying data inputs, model structure, and the modelers' criteria regarding the stage detail needed to capture the effects of screening on LC mortality.

LC survival

All models assume that LC survival varies by histology and stage. Models F, M, S, and U also assume that survival varies by sex. Model U further assumes that survival varies by age at diagnosis. Models E, M, and U use LC survival modules calibrated to the SEER-17 (2004-2008) survival. Survival in model S was calibrated to SEER-17 (1988-2003) survival. LC survival in model F was calibrated to the NLST and PLCO.

Other-cause mortality

Model E uses an other-cause mortality (OCM) module based on the NCI's smoking history generator, which produces OCM rates consistent with the U.S. population.[23,24] All other models use OCM based on the NLST and PLCO (Table 1).

Screening and follow-up

Screening sensitivities vary by model. In model E, screen sensitivity varies by modality, stage, and histology. Models F and U have screen sensitivities that also vary by tumor size (cell number). Sensitivities in models M and S depend on screening modality, tumor size (in mm), and lung nodule location (central vs peripheral). Model S also considers histology. The variability in assumptions is primarily due to differences in model structure (e.g., models that do not model tumor size explicitly cannot have size dependent sensitivities). Follow-up examinations are defined as those received after a positive screen but before diagnosis, if it occurred. Algorithms for follow-up of a positive screen are simulated with varying detail; models M and S include detailed algorithms based on nodule size thresholds and risk factors

(explicit), whereas models E, F, and U incorporate a global probability of receiving several follow-up examinations (implicit) based on the observed frequency of imaging examinations per positive screen in the NLST. Because the NLST and PLCO did not specify a follow-up regimen, models M and S specify less aggressive protocols than the Fleischner Society guidelines, to approximate the observed follow-up rate in the NLST.[25]

Trial simulations

Four models (models E, M, S, and U) generate individual LC outcomes using microsimulations.[26] The simulation depends on individual smoking history, sex, age at enrollment, and screening arm. The specific simulation approach depends on the model's structure. Three models (models E, M, and S) simulate age at onset of lung tumors via their smoking dose-response module and then simulate each tumor's natural history, including malignant conversion, stage progression (models E, M, and S), tumor growth (models M and S), and clinical and screen-detection (models E, M, and S). Model U simulates the initiation of tumors via mutations of normal cells, and then the premalignant and malignant tumor cell dynamics (cell division, death, stage progression, and clinical and screen-detection). Model F uses a likelihood-based approach to estimate LC outcomes and death via a longitudinal, multistage, observation model.[18] All models simulate all trial participants and then compare their aggregate modeled outcomes with those of the trials (LC incidence and mortality and OCM by screening arm, sex, histology, and stage).

Screening effectiveness and mortality reduction

All models evaluate screening effectiveness, but based on different assumptions that depend on model structure. Model M assumes that patients with early-stage non-small cell LC (NSCLC) would undergo resection (lobectomy, consistent with practice guidelines), which removes the primary tumor. In model M, therefore, for patients without undetected distant metastases or additional primary LCs in another lobe, resection is curative for LC. In model U, the benefit of screening is due to the early detection of LC, leading to improved cure probabilities and survival times, which depend on histology, stage, sex, and age at diagnosis, but not on detection mode. Model F assumes that screen-detected cancers are treated according to clinical practice guidelines with cure rates that vary by tumor stage and histology. In model E, screen-detected patients experience a reduced risk of LC mortality

versus clinically detected cases. This improved prognosis is represented as a cure fraction (dependent on stage and screening modality for stages IA, IB, and II) calibrated to the trials. Model S estimates probabilities of lethal metastases as function of tumor size, histology, and sex. All advanced stage LCs are, by definition, detected after the onset of lethal metastases. Some early-stage cancers may have occult lethal metastases at the time of detection. For patients with early-stage and late-stage tumors detected after the onset of lethal metastases, LC survival is not affected by screening. However, with screening, patients are more likely to be detected at early stages before the onset of lethal metastases, and therefore are cured of their disease after standard care.

Model Calibration and Validation Approach

Models were first calibrated to the NLST LC incidence and mortality by screening arm, sex, histology, stage, and detection mode. Models were then validated against PLCO by first comparing model predictions and observed LC incidence and mortality by sex and screening arm in the subset of individuals in PLCO who would have been eligible for the NLST (PLCO-NLST–eligible). Model predictions were consistent with the observed outcomes in the PLCO-NLST–eligible group, demonstrating the consistency between the two trials. However, model outcomes did not consistently match against observed outcomes among PLCO participants who were not eligible for the NLST (never-smokers and light smokers). As a result, models were further calibrated to fit the whole PLCO data set to ensure that they could be used with confidence to extrapolate the effects of CT screening to smokers with lower exposure (<30 pack-years). Calibration methods (targets, measures of goodness of fit, and optimization algorithms) varied by model and are described in Table 1 and in the supplementary material of this Chapter, in the section: “Model Descriptions”.

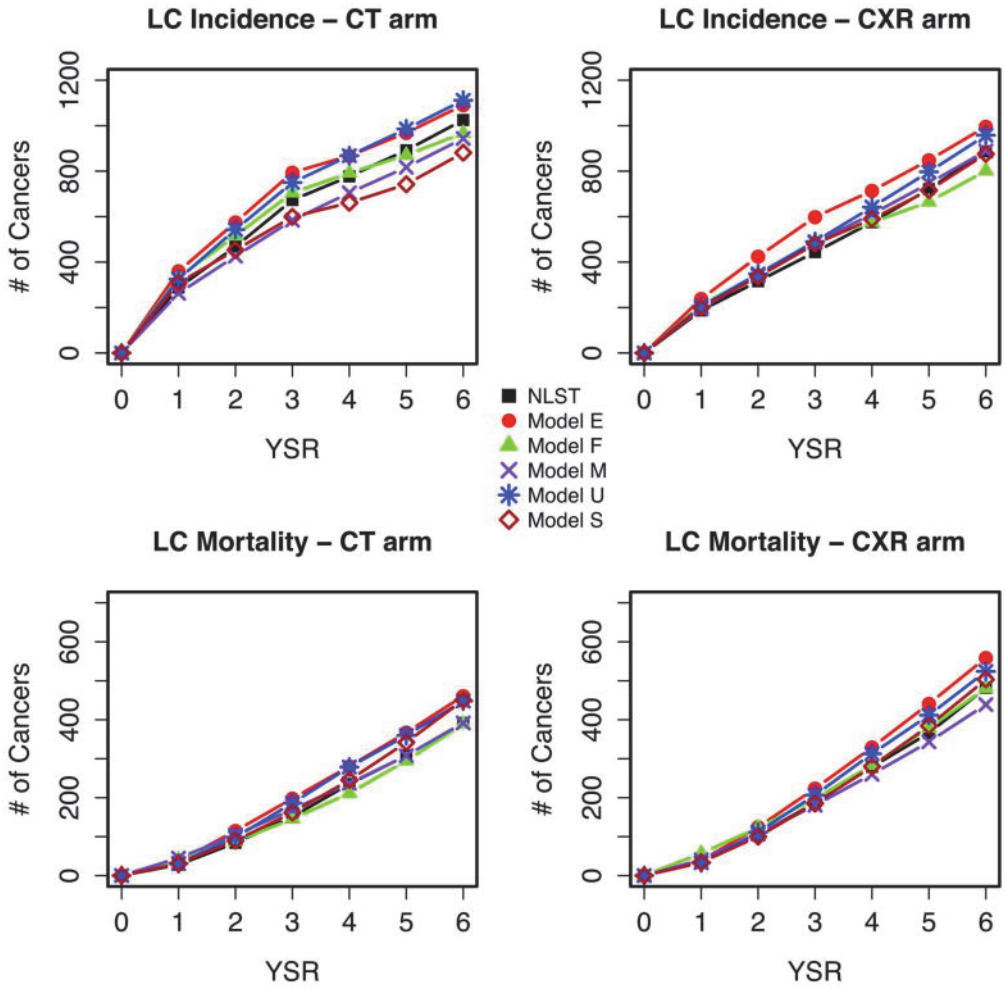
Results

After final calibration, all models produced LC outcomes consistent with both trials (within the confidence intervals of the data). We demonstrated several measures of LC incidence and mortality in the NLST and PLCO for both sexes combined and compared observed and model outcomes. Calibration targets varied by model, and therefore the modeling results shown in each figure include combinations of calibrated outcomes and model

predictions/extrapolations. Modeled outcomes were computed using the "final" version of each model. Figure 1 shows NLST observed and modeled incidence and mortality by screening arm and years since randomization (YSR). The figure shows that as previously reported, the observed cumulative LC incidence was higher in the CT screening arm, whereas the cumulative mortality was higher in the CXR screening arm.[1]

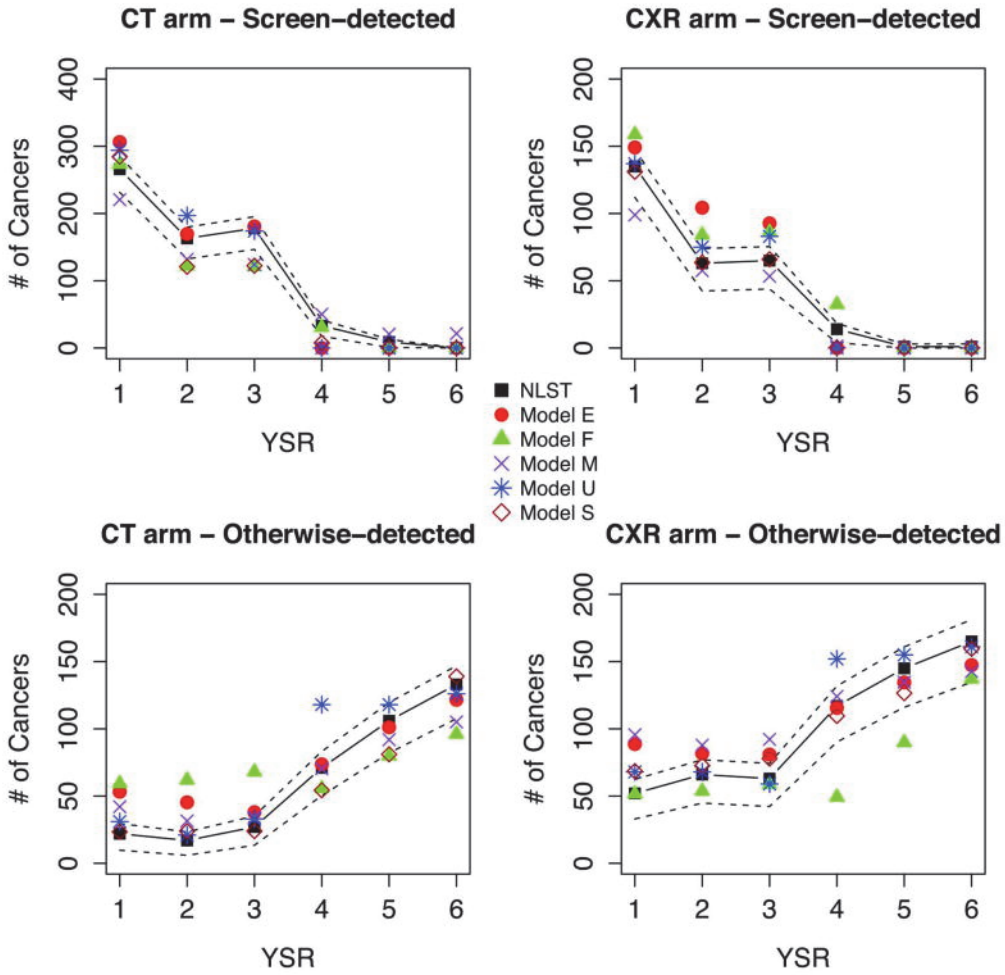
Figures 2 and 3 display observed versus modeled LC cases and deaths in the NLST by detection modality (screen-detected vs non–screen-detected), screening arm, and YSR. The figures show the contrasting pattern between screen-detected and non–screen-detected cancers, with an early increase and peaking by YSR for screen-detected cancers in both screening arms, in contrast with the slow progressive rise for non–screen-detected cancers. The figures indicate that the models reproduce the general patterns of incidence and mortality by screening arm, detection modality, and YSR. Figure S6 in the supplementary material of this Chapter provides additional information on the observed and predicted LC mortality reduction in the NLST by YSR.

Figure 1: National Lung Screening Trial (NLST) observed and modeled incidence and mortality are shown by screening arm and years since randomization (YSR)



Abbreviations: LC, lung cancer; CT, computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.

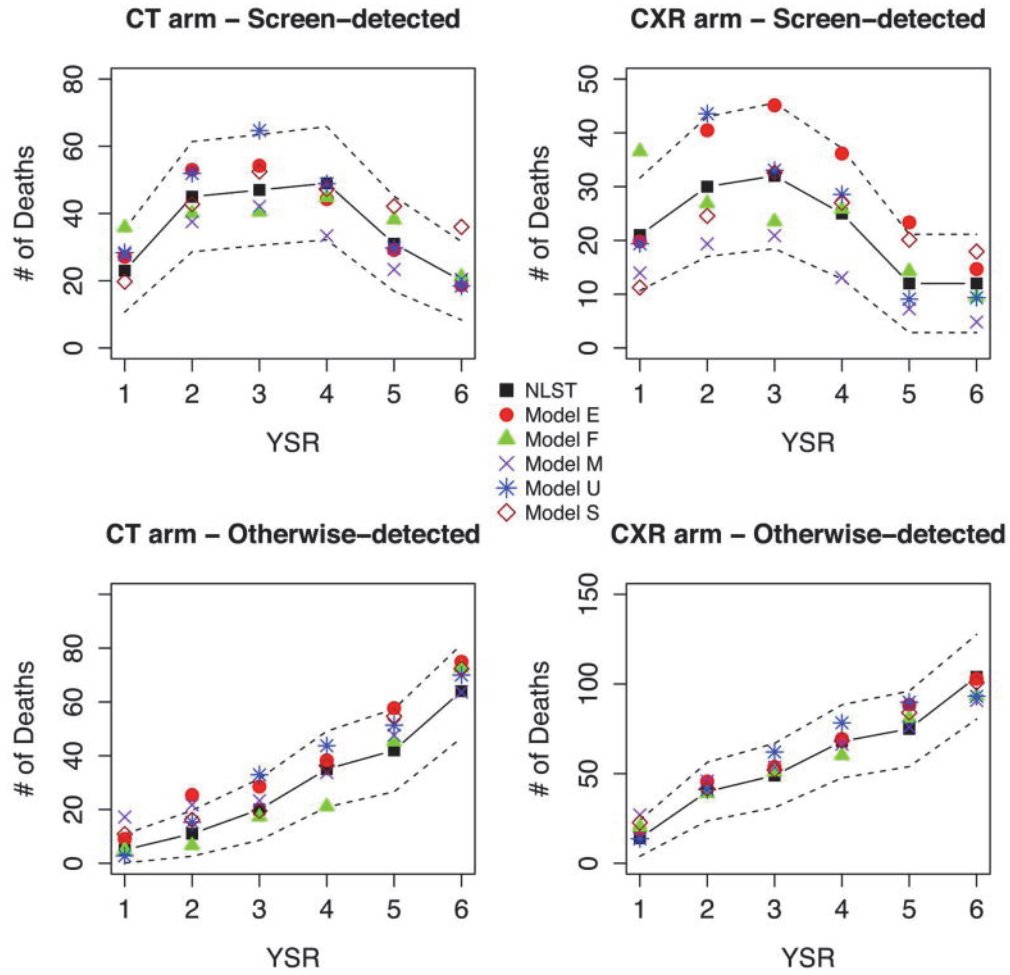
Figure 2: Observed versus modeled lung cancer cases in the National Lung Screening Trial (NLST) are shown by detection modality (screen vs non–screen-detected), arm and years since randomization (YSR)



Abbreviations: CT, computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.

Dashed lines represent 95% binomial confidence intervals for the observed values. Observed screen-detected cancers after year 3 are due to delay in diagnosis after the last screen.

Figure 3: Observed versus modeled lung cancer deaths in the National Lung Screening Trial (NLST) are shown by detection modality (screen vs non–screen-detected), arm and years since randomization (YSR)

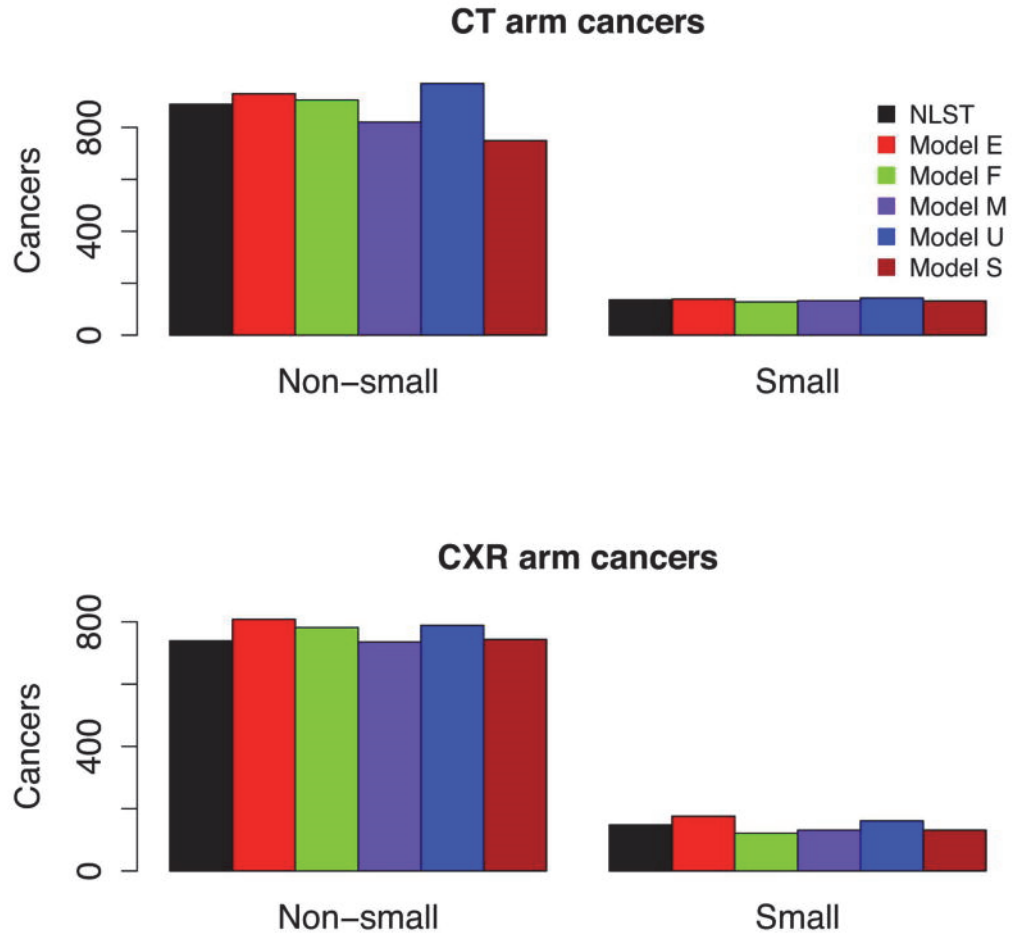


Abbreviations: CT, computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.

Dashed lines represent 95% binomial confidence intervals for the observed values.

Figure 4 shows observed versus model-predicted LCs in the NLST by histology. Because models have varying LC histology categories, we grouped them here as small cell LC (SCLC) and NSCLC. The figure shows that the observed NSCLC incidence was higher in the CT arm, whereas the SCLC incidence was approximately similar in both screening arms. Modeled histology distributions matched well with the observed distributions.

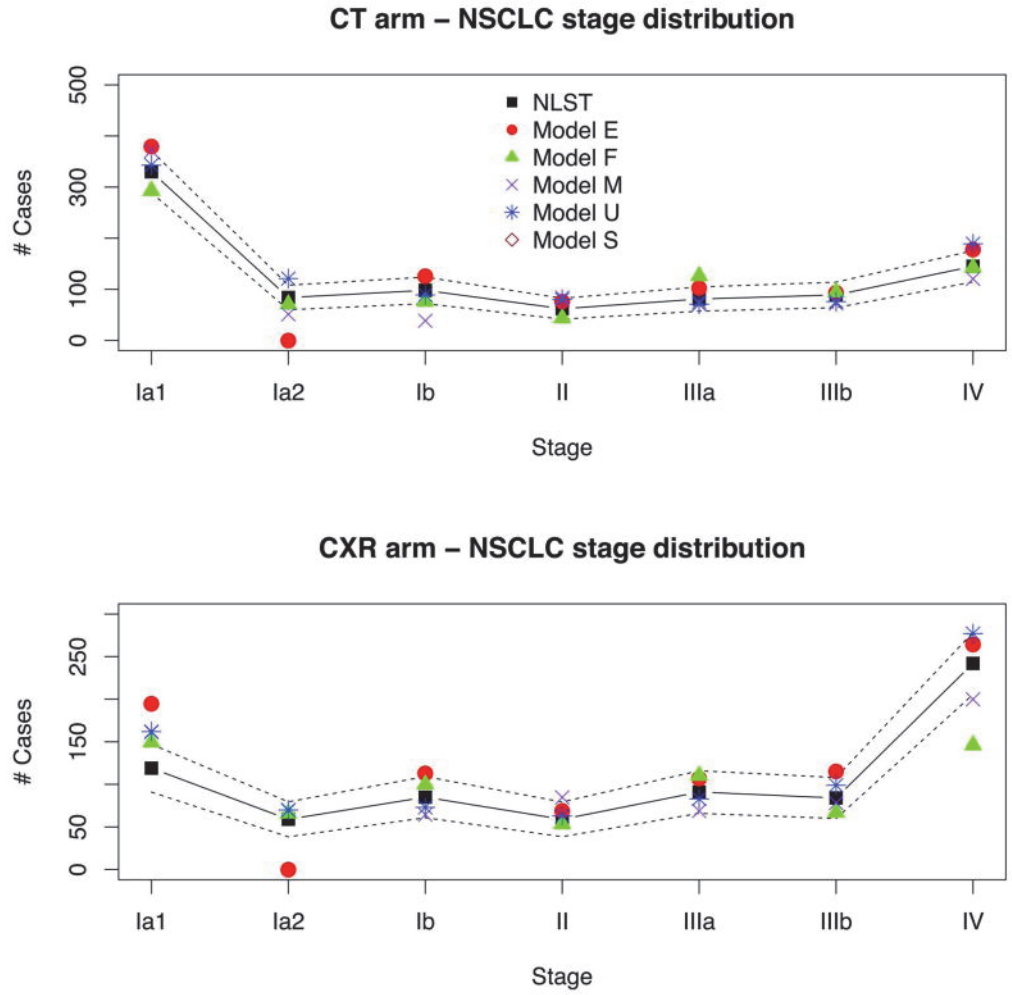
Figure 4: Observed versus model-predicted lung cancers in the National Lung Screening Trial (NLST) are shown by histology



Abbreviations: CT, computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.

Figure 5 shows the NLST observed versus predicted NSCLC incidence by clinical stage and screening arm. The figure demonstrates the shift toward earlier stages in NSCLC incidence in the CT versus the CXR arm.

Figure 5: National Lung Screening Trial (NLST) observed versus predicted non-small cell lung cancer (NSCLC) incidence is shown by clinical stage and screening arm. The figure demonstrates the shift toward earlier stages in NSCLC incidence in the computed tomography (CT) versus the chest radiography (CXR) arm

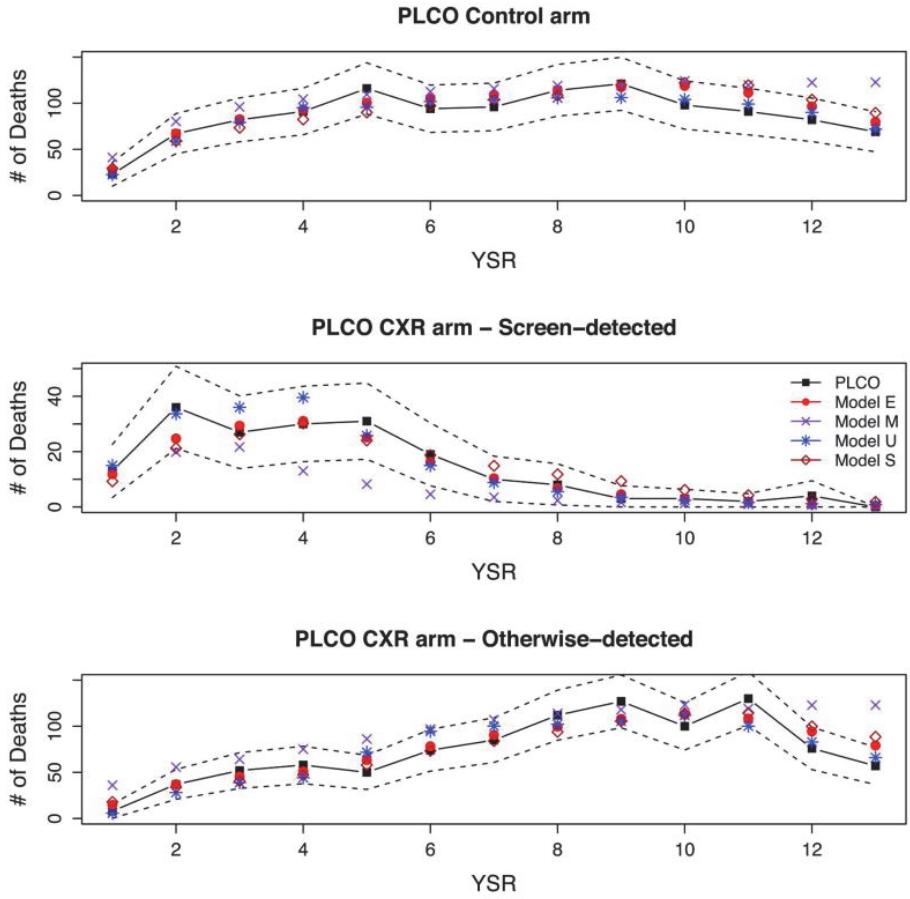


Abbreviations: Model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.
Dashed lines represent 95% multinomial confidence intervals for the observed values. Model E does not model separately Ia1 and Ia2 cancers, so their Ia1 value represents all IA cancers. Model S models early versus late stage cancers.

Figures 6 and 7 show full PLCO and PLCO-NLST-eligible observed and modeled deaths by screening arm, detection mode (CXR arm), and YSR. The figures display the early increase and peaking of screen-detected cancers in the CXR arm by YSR, and the slower increase of otherwise detected cancers in the CXR arm and for all cancers in the control arm. The figures demonstrate a decrease in the non-screen-detected cancers in the CXR and control arms toward the end of the trial, most likely due to the weeding out and loss to follow-up of

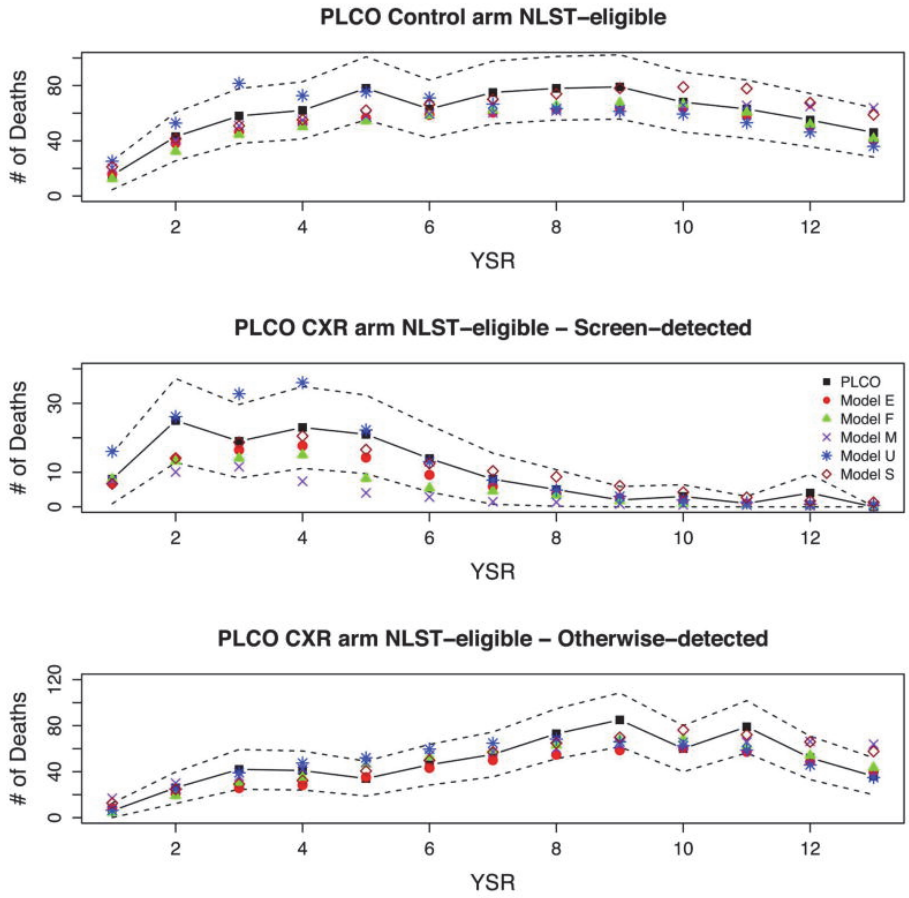
high-risk individuals. All models reproduce the general patterns of incidence and mortality in the PLCO. Figures S6 and S7 in the supplementary material of this Chapter provide additional information on the observed and predicted (cumulative) LC incidence and mortality in the PLCO-NLST-eligible population.

Figure 6: Full Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) observed and modeled deaths are shown by screening arm, detection mode (chest radiography [CXR] arm), and years since randomization (YSR)



Abbreviations: Model E, Erasmus Medical Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University. Dashed lines represent 95% binomial confidence intervals for the observed values.

Figure 7: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)-National Lung Screening Trial (NLST)–eligible observed and modeled deaths are shown by screening arm, detection mode (chest radiography [CXR] arm), and years since randomization (YSR)



Abbreviations: Model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.
Dashed lines represent 95% binomial confidence intervals for the observed values.

Discussion

Main Findings

We derived 5 independent LC and screening natural history models calibrated to the two largest screening trials to date, the NLST and PLCO. The 5 models are diverse in structure, assumptions, and additional data inputs. All models produce outcomes that are generally consistent with the trial results. We found that models calibrated only to the NLST validated well against the PLCO-NLST-eligible population, thereby demonstrating the consistency between the two trials. However calibrating only to the NLST may be insufficient for the

purposes of evaluating screening protocols, allowing for lower smoking exposures, and making projections for the U.S. population. This is particularly true for models that base their smoking dose-response fully on the NLST and also for models with histology distributions based on observed trial data, because the NLST only includes information regarding current and former heavy smokers and it is well documented that the LC risk from smoking varies greatly by histology.[27,28] To derive models that could be used with confidence to extrapolate the impact of low-dose CT screening to smokers with lower exposures (<30 pack-years) and to the U.S. population, it is essential to calibrate such models to data sets with information on LC risk for light and never-smokers, such as NHS/HPFS or PLCO.

Study Limitations and Strengths

The current study has some limitations. First, as in any mathematical modeling approach, our models are simplifications of the biological complexity of lung carcinogenesis and neglect the influence of various endogenous and exogenous LC risk factors such as family history, chronic obstructive pulmonary disease, residential radon, occupational exposures, race, and socioeconomic status. However, it is well known that smoking still accounts for the large majority of LC deaths (≥90%) and our models do capture the complex relation between smoking and LC via their smoking dose-response module.[29] Furthermore, in contrast with the majority of LC risk models in the literature, several of our models do account explicitly for the differential impact of smoking on LC risk by histology. The diversity in model structure, assumptions, and data sources provides additional strength (and an assessment of model uncertainty) to the conclusions of our comparative modeling analysis, as does the long history of collaboration between the CISNET groups.

Another potential limitation of the current study is that the screening mortality reductions predicted by each model are largely dependent on the findings of the NLST and PLCO. To our knowledge the NLST and PLCO are currently the best existing studies of LC screening reporting on the main outcome of LC mortality (reduction), and therefore calibrating models to these trials is the best available option. Some other studies, particularly in Europe, have been underpowered to demonstrate the benefits of low-dose CT screening whereas others are still ongoing.[30] Once data from other trials become available, which is not expected for a

few years, the models could be validated against new trials and, if deemed necessary, calibrated further, particularly if applied to non-US populations. In any case, the models will be helpful to compare trial results and, if needed, to investigate the reasons behind any potential discrepancies. Finally, the current study highlights the benefits of modeling as a way to synthesize information coming from diverse and complex data sources. The models developed use individual data from RCTs (the NLST and PLCO), prospective cohort studies (e.g., NHS/HPFS), and cancer registry data (NCI-SEER). These data sources are extremely valuable on their own, and provide information regarding different aspects of LC. However, it is only through modeling that they can be integrated and jointly inform the biology and epidemiology of LC, as well as the potential benefits of LC screening at the population level.

Implications and Future Research

The results of the current analyses demonstrate that the NLST and PLCO produced consistent results, and suggest that it is critical to use data covering a wide range of smoking histories (never-smokers, light smokers, and heavy smokers) to develop models that can extrapolate the effects of screening to the general population. The 5 models presented herein are currently being used to evaluate the impact of alternative low-dose CT screening protocols on LC mortality in the United States. Specifically, we are assessing the effectiveness of screening programs with varying age eligibility, exposure criteria, and screening frequency.[31] In the near future, we will use the models to predict the potential levels of overdiagnosis due to LC screening and determine optimal screening strategies at both the national and state levels. Using models calibrated to the NLST and PLCO will enhance the validity of effectiveness and cost-effectiveness analyses of LC screening.

References

1. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
2. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
3. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of ct screening for lung cancer: A systematic review. *Journal of the American Medical Association* 2012; **307**(22): 2418-29.
4. Bach PB, Gould MK. When the average applies to no one: Personalized decision making about potential benefits of lung cancer screening. *Annals of Internal Medicine* 2012; **157**(8): 571-3.
5. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of Reduced Tobacco Smoking on Lung Cancer Mortality in the United States During 1975–2000. *Journal of the National Cancer Institute* 2012; **104**(7): 541-8.
6. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. *New England Journal of Medicine* 2012; **366**(8): 687-96.
7. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2009; **151**(10): 738-47.
8. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. *Cancer* 2012; **118**(23): 5955-63.
9. Heijnsdijk EAM, Wever EM, Auvinen A, et al. Quality-of-Life Effects of Prostate-Specific Antigen Screening. *New England Journal of Medicine* 2012; **367**(7): 595-605.
10. Eric JF, Ruth E, Kathleen AC, Angela M. The use of modeling to understand the impact of screening on U.S. mortality: examples from mammography and PSA testing. *Statistical Methods in Medical Research* 2004; **13**(6): 421-42.
11. Kuntz KM, Lansdorp-Vogelaar I, Rutter CM, et al. A Systematic Comparison of Microsimulation Models of Colorectal Cancer. *Medical Decision Making* 2011; **31**(4): 530-9.
12. McMahon PM, Hazelton WD, Kimmel M, Clarke LD. Chapter 13: CISNET lung models: comparison of model assumptions and model structures. *Risk Analysis* 2012; **32 Suppl 1**: S166-78.
13. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value Health* 2003; **6**(1): 9-17.
14. McMahon PM, Kong CY, Johnson BE, et al.Estimating Long-term Effectiveness of Lung Cancer Screening in the Mayo CT Screening Study. *Radiology* 2008; **248**(1): 278-87.
15. McMahon PM, Kong CY, Johnson BE, et al. Chapter 9: The MGH-HMS lung cancer policy model: tobacco control versus screening. *Risk Analysis* 2012; **32 Suppl 1**: S117-24.
16. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.
17. Hazelton WD, Clements MS, Moolgavkar SH. Multistage Carcinogenesis and Lung Cancer Mortality in Three Cohorts. *Cancer Epidemiology Biomarkers & Prevention* 2005; **14**(5): 1171.
18. Hazelton WD, Goodman G, Rom WN, et al. Longitudinal multistage model for lung cancer incidence, mortality, and CT detected indolent and aggressive cancers. *Mathematical Biosciences* 2012; **240**(1): 20-34.
19. Schollnberger H, Manuguerra M, Bijwaard H, et al. Analysis of epidemiological cohort data on smoking effects and lung cancer with a multi-stage cancer model. *Carcinogenesis* 2006; **27**(7): 1432-44.
20. Tammemägi MC, Lam SC, McWilliams AM, Sin DD. Incremental Value of Pulmonary Function and Sputum DNA Image Cytometry in Lung Cancer Risk Prediction. *Cancer Prevention Research* 2011; **4**(4): 552.
21. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *Journal of the National Cancer Institute* 2007; **99**(9): 715-26.

22. Uwe S, Oguzhan A, Ahmed MB, et al. State-Transition Modeling. *Medical Decision Making* 2012; **32**(5): 690-700.
23. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Analysis* 2012; **32 Suppl 1**: S51-68.
24. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.
25. 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.
26. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic Microsimulation Models for Health Outcomes. *Medical Decision Making* 2011; **31**(1): 10-8.
27. Haiman CA, Stram DO, Wilkens LR, et al. Ethnic and Racial Differences in the Smoking-Related Risk of Lung Cancer. *New England Journal of Medicine* 2006; **354**(4): 333-42.
28. Kenfield SA, Wei EK, Stampfer MJ, Rosner BA, Colditz GA. Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco Control* 2008; **17**(3): 198.
29. Centers for Disease C, Prevention. Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. *Morbidity and Mortality Weekly Report* 2008; **57**(45): 1226-8.
30. 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. *Annals of Internal Medicine* 2013; **159**(6): 411-20.
31. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and Harms of Computed Tomography Lung Cancer Screening Programs for High-Risk Populations. AHRQ Pub. No. 13-05196-EF-2. Rockville, MD: Agency for Healthcare Research and Quality 2013.

Chapter 1

Supplementary material

Imputation algorithm for missing data in NLST lung cancer subjects

The purpose of this imputation is to provide missing stage, histology, and lesion size values for subjects who had been diagnosed with lung cancer in NLST. These missing values are needed to estimate: 1) baseline hazards and covariate coefficients in NLST lung cancer survival models; and 2) life expectancy for those alive with lung cancer at the end of the Trial.

Start with 2,058 cases of lung cancer in Phase IB dataset.

307 subjects with 325 missing values for

1. de_stage=11,94,96,98,99 (26 cases)
2. de_type=M (24 cases)
3. lesionsize = >30 with de_stage = 3 (7 cases)
4. lesionsize = M for de_stag = 3 (5 cases)
5. lesionsize=M for de_stag = 3 (263 cases)

Reclassify stage among subjects with de_stag = 3 to 10 as follows:

IA1= de_stag =3 & lesionsize <=20 mm

IA2= de_stag =3 & lesionsize >20, <= 30 mm

IB= de_stag =4

II = de_stag =6 + de_stag =7

III = de_stag =8 + de_stag =9

IV= de_stag =10

Reclassify histology as listed in Table S1:

Table S1: Reclassification of histologies in the NLST

Histology	code	de_type
Small cell carcinoma (sclc)	1	8002,8041,8042,8044,8045,8246
Squamous cell carcinoma	2	8070,8071,8075,8083,8052,8072,8073,8050,8084
Adenocarcinoma	3	8550,8140,8255,8260,8310,8480,8481,8490,8323,8570
Bronchiolo-alveolar carcinoma	4	8250,8252,8253,8254
Large cell carcinoma	5	8012,8013
Adenosquamous carcinoma	6	8560
Pleomorphic/sarcomatoid	7	8022,8033,8032
Carcinoid tumor	8	8240,8249
Unclassified carcinoma	9	8000,8010,8046,8001
Carcinosarcoma	10	8980
Other	11	3070,9120,8021,8830
Missing	.	.M

Multi-step algorithm

1. Impute new stage for de_stage = 11,94,96,98,99 (26 cases) based on modal stage after matching on known rndgroup, can_scr, histology(nejm_hist = sclc or not), treatment (TreatNum = 300 or not), clinical T stage (only if clinical_t = 4-7)
2. Impute histology for de_type = M (24 cases) based on modal hist after matching on known reclassified stage, rndgroup, can_scr, treatment (TreatNum = 300 or not), clinical T stage (only if clinical_t = 4-7)
3. Impute lesionsize for de_stage = 3 & lesionsize >30 (7 cases) or lesionsize missing (5 cases) based on mean lesionsize after matching on known de_stage = 3, rndgroup, can_scr
4. Reassign de_stage = 3 with imputed lesionsize to IA1 and IA2 based on imputed lesionsize.
5. Impute lesion size=M for stages IB through IV (263 cases) based on mean lesionsize after matching on known rndgroup, reclassified stage (IB thru IV) , clinical T stage (only if clinical_t = 4-7).

The algorithm was unable to impute missing lesion size in 13 subjects because of multiple missing values (lesion size plus stage or histology). The algorithm was unable to impute missing stage or histology in 1 subject because both values were missing. Thus, all relevant missing values were imputed for 293 of 307 subjects so that 2,044 of 2,058 lung cancer subjects have known or imputed values related to lung cancer survival.

Imputation of missing smoking data for PLCO participants

Ever smokers who had an unknown age at start were eligible for smoking history imputation, provided that other crucial information was known, namely: a known birth year, gender, number of cigarettes per day, current smoking status and (if applicable) age at quit. Ever smokers who had a missing number of cigarettes per day were eligible for imputation if they had a known birth year, gender, age at start, current smoking status and (if applicable) age at quit.

239 persons had a missing age at start, while 221 persons had a missing number of cigarettes per day. 105 persons who had a missing age at start lacked information on their number of cigarettes per day, current smoking status and/or (if applicable) age at quit. 58 persons who had a missing number of cigarettes per day lacked complete information on age at start, current smoking status and/or (if applicable) age at quit. Therefore necessary information required for imputation was known for 134 ever smokers (58 current and 76 former smokers) with a missing age at start and the necessary information for the imputation of the number of cigarettes per day was known for 163 ever smokers (41 current and 122 former smokers).

For each person, 3,000 smoking histories were generated based on birth year and gender, using the Smoking History Generator developed by the National Cancer Institute.[1,2] Smoking histories in which the person had died before the age of entry into PLCO were excluded.

The age at start of the first history that matched the following criteria was used to substitute the missing age at start:

- The generated smoking history matches the smoking status of the participant at the age of entry into PLCO.
- The generated smoking history matches the smoking quintile (based on the number of cigarettes smoked per day and birth year) of the participant.
- The generated smoking history matches the age of cessation of the participant (if applicable).

For participants with a missing number of cigarettes per day, the number of cigarettes per day (at age 30) of the first history that matched the following criteria substituted the missing number of cigarettes per day:

- The generated smoking history matches the smoking status of the participant at the age of entry into PLCO.
- The generated smoking history matches the age at start of the participant.
- The generated smoking history matches the age of cessation of the participant (if applicable).

In case none of the generated smoking histories met these requirements, a new set of 3,000 smoking histories was generated. These steps were repeated until a matching smoking history was found.

Model Descriptions

Erasmus (E)

Model structure

The MIcrosimulation SCreening Analysis (MISCAN) model was developed for the evaluation of screening for diseases, such as breast and prostate cancer.[3,4] The MISCAN-lung model focuses on screening for lung cancer. While detailed descriptions of this model have been published previously, an extended version of the model was used to calibrate to the National Lung Screening Trial (NLST), Prostate, Lung, Colorectal and Ovarian cancer screening trial (PLCO) and SEER-17 2004-2008 incidence data.[5,6] In brief, MISCAN-Lung is a microsimulation model that simulates a population of individual life histories, which consists of individual smoking histories, development of preclinical and clinical lung cancers, survival of clinically detected lung cancers, death due to lung cancer and death due to other causes.

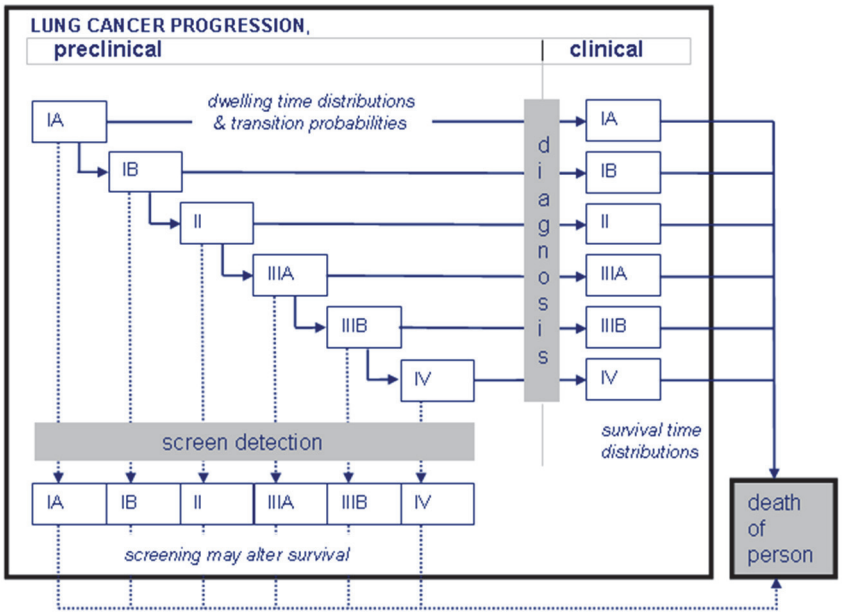
For each individual, a smoking history is sampled from the data, which influences the probability of developing preclinical lung cancer as well as the probability of dying from other causes. Life histories are generated in the presence and absence of screening. If a person dies of lung cancer before he or she dies from other causes, his or her death age is adjusted accordingly. If the screening component is activated, preclinical lung cancers may be detected by screening, which may alter a person's life history, as shown in Figure S1.

NLST Calibration targets

Several calibration targets were chosen, which were divided into various categories. To create these categories, the NLST data was stratified by: gender, age-groups (ages 55-59, 60-64, 65-69 and 70-74 at randomization), smoking status at randomization (current / former), screening arm (computed tomography (CT) / chest radiography), screening round (round 1, 2, or 3) and year of diagnosis (1, 2, 3, 4, 5 or 6 years after randomization). This stratification was used to calibrate the model to the following calibration targets:

a) Number of screen-detected lung cancers and relative distribution of histologies/stages (by gender, trial arm, age-group, smoking status and screening round)
b) Number of clinically detected lung cancers and relative distribution of histologies/stages (by gender, trial arm, age-group, smoking status and year of diagnosis)
c) Lung cancer mortality rate of patients with screen-detected lung cancer (by gender, trial arm, stage and year of death (1, 2, 3, 4, 5 or 6 years after randomization))

Figure S1: An overview of the natural history and screening parts of the MISCAN-lung model



PLCO Calibration targets

The calibration targets for the PLCO population were similar to those chosen for the NLST population, though they were divided into different categories. The PLCO data was stratified by: gender, age-groups (ages 55-59, 60-64, 65-69 and 70-74 at randomization), smoking status at randomization (ever/never-smoker), trial arm (usual care / chest radiography), screening round for the intervention group (round 1, 2, 3 or 4) and year of diagnosis (1 to 13 years after randomization). This stratification was used to calibrate the model to the following calibration targets:

a) Number of screen-detected lung cancers and relative distribution of histologies/stages (by gender, trial arm (intervention arm only), age-group, smoking status and screening round)
b) Number of clinically detected lung cancers and relative distribution of histologies/stages (by gender, trial arm, age-group, smoking status and year of diagnosis)
c) Lung cancer mortality rate of patients with screen-detected lung cancer (by gender, trial arm, stage and year of death (1 to 13 years after randomization))

SEER-17 Calibration targets

Additionally, the model was used to simulate a population, in a situation without lung cancer screening, based on demographics from SEER-17 (years 2004-2008) and the United States population. This population was used to calibrate the MISCAN model to the:

d) SEER-17 2004-2008 lung cancer incidence data (by 5-year age-groups for ages 25 to 84, histological type and stage)[7]

Incidence

Incidence is modeled through a multistep procedure. Once a person's life history and smoke exposure characteristics, based on data from the NLST, PLCO and the U.S. population have been generated, the two-stage clonal expansion (TSCE) model is used to determine whether carcinogenesis occurs.[8]

The majority of the parameter values of the TSCE were based on estimates on data from the Nurses' Health Study (NHS) and Health Professionals Follow-up Study (HPFS).[9] The exceptions to this are the gender-specific parameters for malignant transformation. These were calibrated in conjunction with the mean sojourn times per stage and histological type to match NLST, PLCO and SEER-17 2004-2008 incidence data, stratified by histological type and stage.

Lung cancers can either be clinically detected (due to symptoms) or by CT or chest radiography screening. The incidence of clinically detected lung cancers depends on the sojourn times and the probability of clinical detection, which both vary by stage and histology (and by gender for the sojourn times). If the screening component of the model has not been activated, lung cancers can only be clinically detected.

Histology

MISCAN-Lung distinguishes four histological types of lung cancers, based on ICD-O-3 codes obtained from the SEER website, which can be separated into two categories: non-small cell lung cancers and small-cell lung cancers.[10] The first category contains the histological types: squamous cell carcinoma, adenocarcinoma/large (which consists of the types: adenocarcinoma, large-cell carcinoma and bronchioloalveolar carcinoma) and the histological type "other" (which consists of the remaining non-small cell carcinoma). The second category contains all small-cell carcinoma, which are grouped in the histological type small-cell carcinoma. Each

histological type progresses through six tumor stages, based on the AJCC Cancer Staging Manual, 6th edition: IA, IB, II, IIIA, IIIB and stage IV.[11]

Screening participation

Each lung cancer free (i.e. no clinical lung cancer) participant in the modeled screening arms receives an invitation for each screening. The probability that a participant attends a screening is dependent on age-group, gender, screening arm and whether the participant has attended the previous screening.

The sensitivity of each test is based on screening arm, tumor stage, histology and screening round (for NLST only). The screening round specific parameters were incorporated in the model as, according to NLST investigators, participants with abnormalities in the third round may have been followed more aggressively compared to the other rounds.[12]

Mortality

Upon detection of lung cancer by screening, a person's life history may be altered. Detection by screening may cure a patient, allowing him to resume his normal (lung cancer free) life history. The probability of cure differs by the stage of detection and between computed tomography and chest radiography for stages IA, IB and II. This distinction was made to account for the large difference in mortality for these stages between the two screening methods. The probability of a history change is estimated by estimating the history change parameters in conjunction with the sensitivity of the screening tests and the natural history parameters to match the mortality rate of patients with screen-detected lung cancer.

A person may also die from causes other than lung cancer. MISCAN-Lung incorporates the Smoking History Generator (SHG) application from the National Cancer Institute (NCI), which uses data on smoking habits in the U.S. population to provide probabilities for death from other causes, depending on gender, smoking history and year of birth.[1]

Survival

After clinical detection or screen-detection (without cure) the patient's duration of survival follows a histology and stage specific survival function, which is piecewise uniformly distributed. These distributions were obtained from SEER 2004-2008 lung cancer mortality data.[13]

Optimization procedure

The MISCAN-Lung model was simultaneously calibrated to the NLST, PLCO and SEER-17 populations. For each combination of calibration target and category, model expectations and observed data were compared with a likelihood based goodness of fit criterion. For each outcome measure (e.g. lung cancer incidence, lung cancer mortality) and for each combination of the categories of the stratification variables (i.e. age-group, smoking status, screening arm, and gender), the model expectation and the number of observed cases were compared

using a likelihood-based deviance criterion. Deviances were calculated by assuming a Poisson likelihood for incidence and mortality data, while a multinomial likelihood was assumed for histology and stage distribution data. The goodness-of-fit criterion for the complete NLST, PLCO and SEER-17 data was computed as the sum of these deviance criteria over all outcome measures and all combinations of the categories of the stratification variables. The goodness-of-fit criterion for the entire data set was minimized using the Nelder-Mead optimization algorithm, which simultaneously adjusts model parameters (shown in Table S2) until the best agreement with the observed data is reached.[14] A combination of visual assessment and the improvement in the goodness-of-fit per iteration was used as a stopping criterion. When the improvement in the goodness-of-fit per iteration was very small, the algorithm was manually stopped and the fit to the calibration data was then visually assessed. This acceptance criteria does not allow for multiple parameter sets to be accepted.

Table S2: Estimated parameters of the MISCAN-Lung model

Natural history	Sensitivities	History change
Mean duration of each stage (by histology and gender) and shapes of duration distributions	Sensitivity parameters for each stage and histology (by screening modality)	Probability of cure by stage and screening method
Probability of clinical detection (by histology and stage)	Sensitivity parameters for screening rounds	
Relative distribution of histologies (by gender)		
Malignant transformation parameters (by gender)		

Nodule follow-up approach/Extrapolation of CT screens to CT scans

The MISCAN-Lung model does not include a nodule follow-up approach. Instead, the probability that a malignant nodule is diagnosed due to detection by screening during a screening round is modeled as a mathematical function of the screening modality, histology, tumor stage and screening round (NLST only). The screening round specific parameters were incorporated in the model as, according to NLST investigators, participants with abnormalities in the third round may have been followed more aggressively compared to the other rounds.[12] To extrapolate the number of CT screens to CT scans, it is assumed that for each CT screen a number of follow-up CT scans are performed (based on data from NLST follow-up procedures).

FHCRC (F)

Investigators from the Fred Hutchinson Cancer Research Center (FHCRC) used maximum-likelihood methods to calibrate a longitudinal multistage observation (LMO) model to individual level smoking histories and longitudinal outcomes data in NLST, including separate pathways representing six histological subtypes of lung cancer.[15] Parameters for each of the six pathways were estimated separately by gender. Following calibration of the model to NLST data, the model was validated to independent data from the PLCO trial.

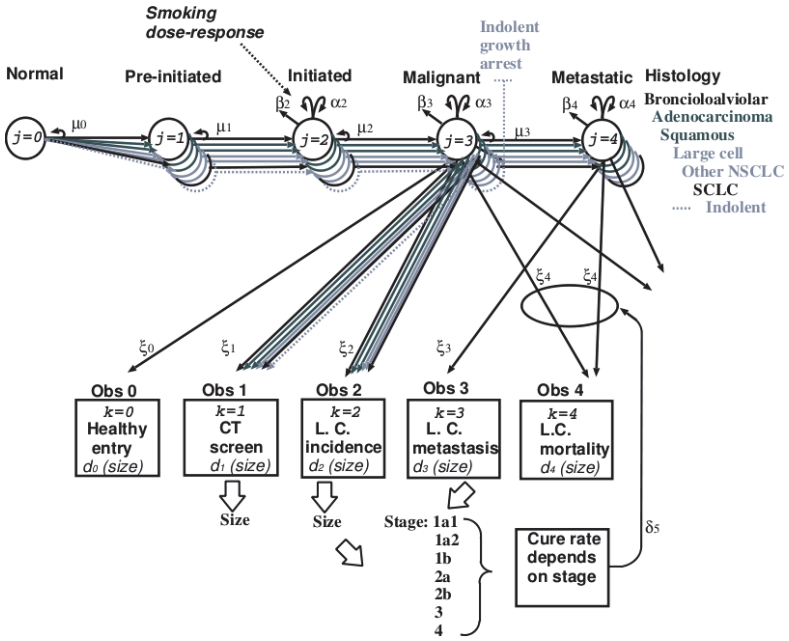
The LMO model assumes each histological subtype of lung cancer may have differing response to cigarette smoke and separate cell division, death, and mutation rates that define the multistage pathways. Normal pulmonary stem cells may undergo two pathway-specific sequential mutational or epigenetic events to generate a pre-malignant cell. Pre-malignant cells may divide and undergo apoptosis or differentiation, leading to clonal expansion, or frequently to extinction of the clone due to the stochastic nature of the clonal expansion process. Pre-malignant cells in the developing clone may also mutate to generate a malignant cell. Malignant cells also divide and die, typically at faster rates causing increased rates of clonal expansion. If the malignant clone does not become extinct, it may grow large enough to be detected through screening, or other observation processes. The model also allows estimation of a rate for malignant cells to become metastatic, and clonal expansion rates of metastatic cells. We compared different mechanisms of smoking exposure-response, finding that the best estimates suggest that smoking influences primarily pre-malignant cells growth or promotion.

As a cancer grows, it may contribute to different longitudinal outcomes, which are modeled using stochastic observation processes. Possible longitudinal outcomes include negative screens, screen-based detection and diagnosis or symptomatic diagnosis, and subsequent outcomes that include the possibility of cure, lung cancer death, censoring, or other-cause death. Each individual outcome, such as a negative or positive screen, conditions the probability for subsequent outcomes. Thus, for example, a negative CT screen indicates that the multistage process does not include large tumors at the time of screening, so the probability of symptomatic detection immediately following the screen is low, but increases with time as small tumors not detected by the screen grow larger.

The probability for a positive outcome from an observation depends on the tumor size and the sensitivity of the observation. We used likelihood methods to estimate all of the cell kinetic rates and separate sensitivities for CT and CXR detection and diagnosis of the growing cancers. If a cancer is not detected through screening, it may grow to a larger size that causes symptomatic detection, modeled as an observation process with lower sensitivity. Calibration to the stage distribution at diagnosis and subsequent survival was based on malignant tumor size and the contribution from metastatic cells. Following diagnosis, the model assumes there is some probability, depending on histology, that the cancer is cured, but if not, the cancer undergoes further growth and metastatic spread, ultimately causing death. The model also allows for the possibility of over-diagnosis

due to growth of indolent cancers that may be detected through screening, but which are unlikely to undergo further growth leading to symptomatic diagnosis or death.

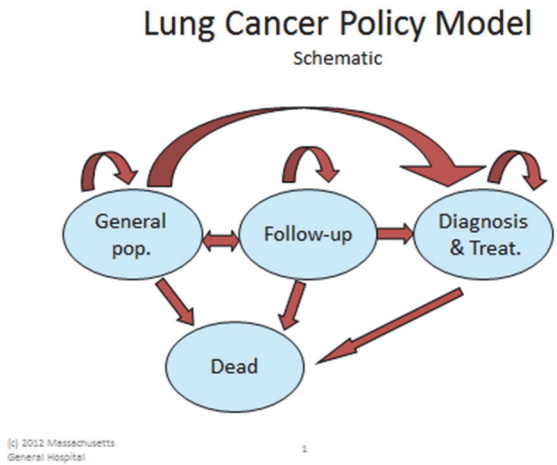
Figure S2: Fred Hutchinson Lung Cancer Model



MGH-HMS (M)

The Massachusetts General Hospital and Harvard Medical School (MGH-HMS) group's model is a state-transition model which simulates an individual patient's lung cancer development, progression, detection, treatment, and survival. The MGH-HMS model was developed to evaluate the clinical effectiveness and cost-effectiveness of low-dose computed tomography (CT) screening for lung cancer.[16-18] The model has also been used to estimate the reduction in lung cancer-specific mortality and the numbers of radiation induced cancers from annual lung cancer screening using low-dose CT.[19]

Figure S3: MGH-HMS Lung Cancer Policy Model



Each hypothetical individual in the MGH-HMS model can develop up to three cancers from any of five lung cancer cell types (adenocarcinoma, large cell, squamous cell, small cell, and other, with bronchioloalveolar carcinoma (BAC) considered a subtype of adenocarcinoma). The natural history parameters related to unobservable events (i.e. the initiation of the first cancer cell) were estimated by calibrating the model using SEER registry data (cancer incidence by cell type, stage distribution at diagnosis, and stage-specific survival) and published cohort studies. The model was previously validated, and a detailed description of the previous version of the MGH-HMS model can be found online (http://cisnet.cancer.gov/profiles/).

Prior to the calibration of NLST results, the MGH-HMS model was modified accordingly to account for the detailed data provided by the trial. The model modifications can be broadly categorized into three types: 1) modeling of disease natural history, 2) patient experience in NLST, and 3) calibration targets. For the model inputs of disease natural history, we updated the model parameter inputs of gender-specific lung cancer stage-IV median survival times from SEER-9 (1990-2000) to SEER-17 (1973-2008). The other-cause mortality rates derived from the NLST data were incorporated into the model, which reflects the lower other-cause mortality rates observed in NLST compared to the U.S. population. The NLST other-cause mortality rates were fitted using a multivariable Cox proportional hazards model stratified by gender and five-year age-groups. To capture the patient experience in NLST, the observed screening adherence and the smoking histories of the patients in the NLST were used as the inputs of the model. The MGH-HMS group also replaced their previous follow-up algorithm for surveillance of lung nodules detected at screening based on the Mayo CT Screening Study with a follow-up protocol similar but less aggressive than the Fleischner Society Guidelines for pulmonary nodules.[20] Finally, we replaced our original model calibration targets with the targets derived from the two arms of the NLST. The new calibration targets include the numbers of trial screen-detected and interval cancers, the histological distribution of the detected cancers, lung cancer specific deaths, and the five-year survival of diagnosed lung cancers. Additionally, we verified the other-cause mortality rates used as model inputs were in agreement with the simulation results calculated using model outputs. The discrepancy between the model outputs and calibration targets was calculated using "deviance" originally proposed by Cressie and Read.[21] To simultaneously fit multiple targets, the deviances of all targets were combined into a single total deviance by a weighted-sum approach.[22] Then simulated annealing optimization algorithm was applied to search through the natural history parameters to minimize the total deviance. The final natural history parameter set with the lowest total deviance was used as model inputs to generate the MGH-HMS results shown in following sections.

The MGH-HMS group also used a subset of patient's data in the PLCO's no-screening and CXR arm for additional model calibration. As described in the last paragraph, the MGH-HMS group initially calibrated their model to the results from the NLST, which consists of patients with heavier smoking history. The PLCO data provide an opportunity for the MGH group to further refine their model's parameters to predict the lung cancer risks of light and non-smokers. The final best-fitting natural history parameter set was used as model inputs to generate the MGH-HMS results shown in this article.

Michigan (U)

The Michigan lung cancer and screening model is an extension of the two-stage clonal expansion model previously developed based on the Nurses' Health and Health Professionals' Follow-up studies.[9] This extension assumes that two mutation (rate-limiting) events are required for the initiation of premalignant lung tumors, and explicitly models the dynamics of premalignant and malignant tumors. Malignant tumors evolve through a stage progression model and can be detected clinically or through screening. The model was first fitted to the lung cancer incidence in the NHS and HPFS using a likelihood-based approach, and then calibrated further to the lung cancer incidence in the NLST via micro-simulations of the carcinogenesis process.

Smoking Dose-Response

Figure S4 shows a schematic representation of the Michigan model. The model assumes that normal stem cells become pre-initiated according to a Poisson process with intensity $\mu_0 X$, where X is the number of susceptible stem cells. Pre-initiated cells then become initiated with rate μ_1. Initiated cells (premalignant stage) expand clonally (promotion) via a linear birth and death process with rates (α_{pm},β_{pm}). This means that each time that an initiated cell divides, it can produce two initiated cells (with birth rate α_{pm}) or die/differentiate (with death rate β_{pm}). Initiated cells can also divide into one initiated and one malignant cell (with rate μ_{pm}). Malignant (preclinical stage) cells grow according to a linear birth death process with rates (α_{pc},β_{pc}). Each of the normal, pre-initiated and premalignant parameters in the model depend on the cigarette per day dose. The malignant compartment rates are assumed to be independent of smoking. Let $D_i(t)$ denote the cigarette consumption in cigarettes per day of an individual at age t. We assume that each of the smoking-dependent parameters in the model has a dose–response given by:

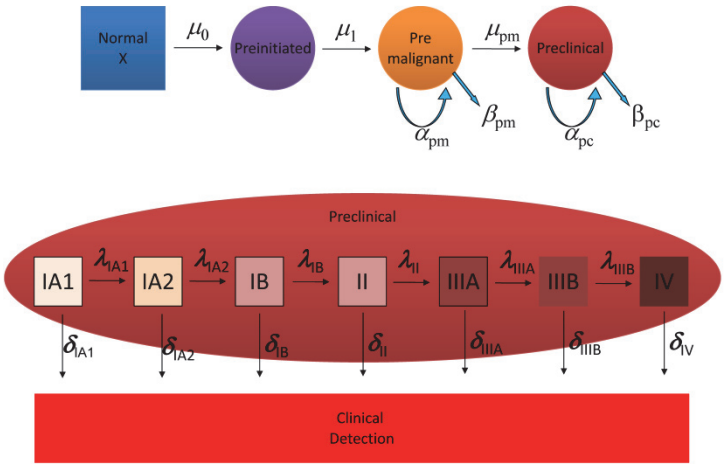
$$\theta_{iob}\ (D_i(t)) = \theta_0 (1 + \theta_c D_i(t)^{\theta_e}),$$

where ϑ_0 is the background parameter, and ϑ_c and ϑ_e are the dose–response coefficients.[9,23,24] Parameters vary by gender and histology (adenocarcinoma, other non-small cell, small cell).

Stage progression model

Once a tumor becomes malignant (preclinical) it goes through a Markov-state transition model. The malignant stages are based on the AJCC classification (IA1, IA2, IB, II, IIIA, IIIB and IV). Transition rates per stage λ_s are proportional to the number of cells in the preclinical compartment. At each stage the cancer can be detected with rate (per cell per year) δ_s.

Figure S4: Michigan LC screening model. Parameters vary by gender and histology



Simulation of the carcinogenesis process

For each individual the whole carcinogenesis process is simulated forward using a tau-leaping approach.[25,26] In particular, at each time step, we generate the number of mutations, cell divisions, cell deaths, of all cell types and tumors. We also generate the transitions of the stage progression model and the clinical detection of preclinical cancers.

Screen-detection

The screen-detection probability is proportional to the total number of premalignant and malignant cells in a tumor. The proportionality constants σ_s vary by screening modality and histology and were calibrated to the NLST trial.

Survival

Once a cancer is detected, the model generates a time to lung cancer death based on a survival function with cure by stage, histology, gender, and age at detection. The survival functions with cure (logistic or lognormal) were estimated using the NCI Cansuv program and the SEER-17 2004-2008 lung cancer survival.[13,27]

Other-Cause Mortality

Death from other causes is generated for each simulated individual based on a Gompertz model of mortality by age, gender, age at randomization, smoking status at baseline and pack-years at baseline:

$$G(a) = A_0 \exp\{G \cdot a\} \exp\{b_1 \cdot status + b_2 \log(pkyr)\}$$

Gompertz model parameters were estimated using the other-cause mortality in each trial.

Calibration

To calibrate the model we simulate the lung cancer natural history (carcinogenesis) process of each individual in the NLST cohort and then compute the number of lung cancers by arm, gender, histology stage, year since randomization and mode of detection. We then compute a chi-square statistic using the difference of the

observed minus expected (model predicted) cases in each category.[22] Optimization of the chi-square statistic is performed using the Nelder-Mead Simplex algorithm.[14] Further calibration of lung cancer initiation rates and survival parameters to match with PLCO observed LC incidence and mortality was performed.

Stanford (S)

Stanford developed a microsimulation model to estimate the impact of screening on lung cancer incidence and mortality, in addition to its impact on lung cancer stage-shift, survival and overdiagnosis. The model operates by generating a set of lung-cancer specific events at the level of individual patients then aggregates these results to produce cohort-level effects. Lung-cancer specific events are modeled both for presence and absence of screening scenarios. For each individual lung cancer patient, given their screening attendance, the model generates the patient's age at detection, histological type, mode of detection (screening vs. symptoms-based), tumor stage, survival time, cause of death (lung cancer vs. other causes) and an indicator as to whether or not the patient would have been symptomatically detected with lung cancer in the absence of screening. At the core of the model is a sub-model of the natural history of lung cancer, which is used to determine the likelihood of cure by screening and estimate the lead-time, length-time and overdiagnosis biases introduced by screening on survival rates. Following is a brief description of the model components:

Lung Cancer Incidence Component

In Model S, the lung cancer incidence component determines whether a simulated individual would become symptomatically detected in the absence of screening, in the absence of death from other causes. LC risk is modeled based on a published Two-Stage Clonal Expansion (TSCE) Model, which was calibrated to the Nurses' Health Study/ Health Professionals Follow-up Study (NHS/HPFS).[9] In this work, individual's smoking histories, sex and age at randomization from NLST population are used to determine the annual hazard of lung cancer using the TSCE model. For each simulated individual, the age at symptomatic detection (age at LC detection in the absence of screening) is determined using these annual hazard rates. The underlying histology proportions in the absence of screening are imputed to the model based on the histology distributions in SEER-17 cancer registries between years 1988 and 2003 for each gender. The histology distributions are then calibrated to observed histology distributions during calibration to trials. Age at death from other causes is determined using Michigan/FH Other-Cause Mortality Model fit to NLST, which is described above.

Lung Cancer Natural History Component

In Model S, the natural history component models the progression of lung cancer in the absence of screening, as illustrated in Figure S5.[28] The size of the primary tumor is modeled starting from $1mm^3$ in volume until it reaches a size (D) that prompts symptomatic detection. We assume that the primary tumor grows exponentially with tumor volume doubling time (TVDT). If patients are detected before the tumor reaches the "onset of lethal metastatic disease" (C), they are cured; otherwise, the lethal metastatic burden starts at C and grows exponentially in proportion to the tumor growth and becomes the cause of lung cancer death at the maximal metastatic tolerance level (B). Lethal metastatic burden excludes subclinical metastasis that may be

eradicated or controlled by surgery or systemic treatment; instead, it is disease that is or has become resistant to treatment and will lead to death from lung cancer.

In the absence of screening, the patient is detected due to the size of the primary tumor or metastasis, dependent on which prompts symptoms first. The metastatic burden becomes observable when it reaches a fraction (c_1) of lethal metastatic burden. The patients detected after this moment are considered to have advanced stage (AJCC Stage III-IV) lung cancer; otherwise, they are considered to have early stage disease, with only those who are detected and treated before the onset of lethal metastases as cured. The model is calibrated to the survival of the detected cases in SEER-17 registries between years 1988 and 2003. The parameters of the natural history model are estimated based on the survival by gender, stage, and size with maximum likelihood estimation using Nelder-Mead optimization procedure.[14] The actuarial survival curves stratified by subtype (adenocarcinoma, squamous cell carcinoma, large/other cell, and small cell), gender, stage (i.e. early: I-II, advanced: III-IV) and by tumor size(i.e. size <2 cm, 2-2.9 cm, 3-3.9 cm, 4-4.9 cm, 5-6.9 cm, ≥7 cm) are generated by NCI SEER*Stat software.

Lung Cancer Screening Component

When a screening schedule is introduced, it is modeled as being superimposed onto the individuals' natural histories of lung cancer. The size of tumor at each screen time is determined backwards from the size at symptomatic detection using the sampled tumor growth rate for each individual. Each simulated person is also assigned a randomly generated screen-detection threshold. These thresholds are determined separately for CT and CXR arms through calibration. At the time of each screen, among the tumors that have not been symptomatically detected, the ones below this threshold are missed and the ones above this threshold are screen-detected. Once a tumor is screen-detected, it is followed up by additional CT exams depending on the size of tumor at detection and the stability of tumor (i.e. TVDT>1 year). In the Stanford simulation model, the recommendations of Fleischner Society are considered as the follow-up protocol but the guidelines are further calibrated to best reflect the less aggressive follow-up regimen observed in NLST.[20]

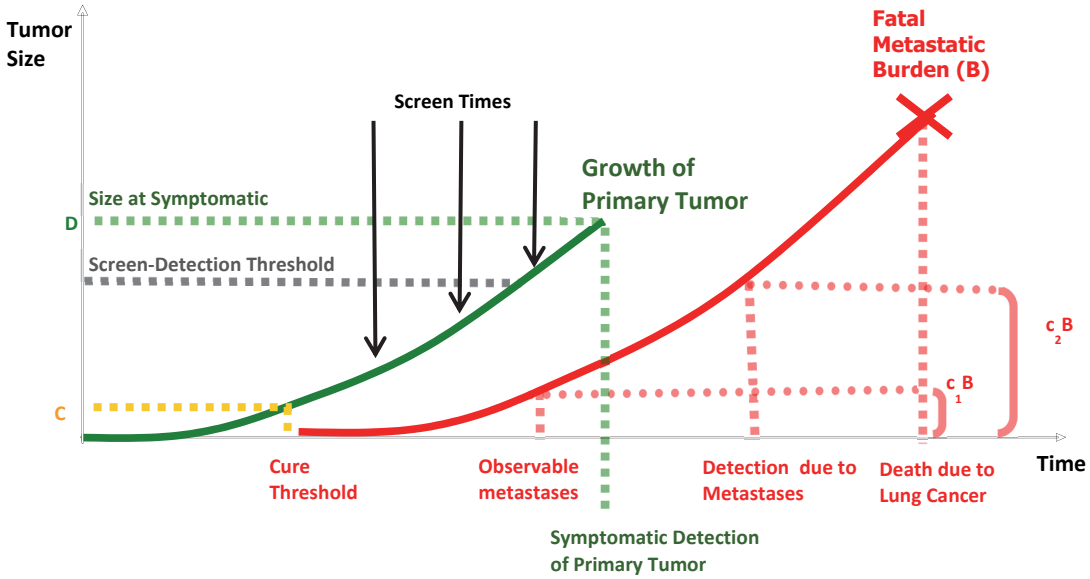
The dissemination and compliance to screening is derived from the observed data or prescribed for a hypothetical analysis. The interval from the moment the tumor is first screen-detectable to the moment it reaches the onset of lethal metastases is defined as the "cure opportunity window". If the patient is screened during the cure opportunity window, then the patient is, by definition, screen-detected and screening has led to cure from lung cancer.

Survival and Dissemination and Efficacy of Treatment Component

If a patient is clinically detected before the onset of lethal metastases, it is assumed that he/she will not die of their disease. If the patient would be clinically detected after the onset of lethal metastases but is screen-detected during the cure opportunity window, then the patient is assumed to be cured. If the patient is detected after the onset of lethal metastases, then the patient will eventually die of lung cancer unless they

die of other causes first. Survival time from clinical detection for each individual is obtained as an output of the natural history model calibrated to SEER survival conditioned on gender, histology and tumor size at detection. The age of death from lung cancer is determined by adding survival duration. The dissemination and efficacy of therapy (resection and adjuvant therapy) are assumed to be captured in the SEER survival curves that were used to estimate the parameters of the natural history model. As treatment for lung cancer improves, the model parameters will need to be re-estimated, as the onset of lethal metastases is likely to change.

Figure S5: Stanford natural history model of lung cancer



Black arrows superimpose screening on to the natural history model in the absence of screening. The cancer is detected with screening if the tumor size at screen is larger than the screening detection threshold.

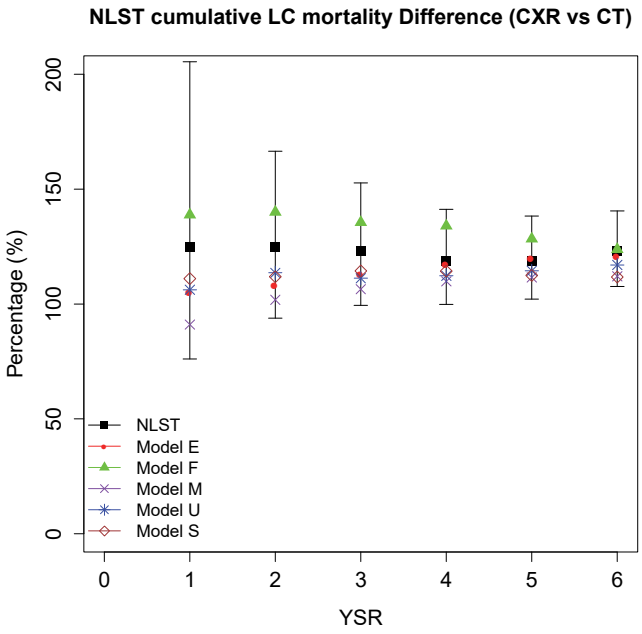
Model Fitting, Parameter Estimation and Calibration

All parameters of the natural history model are estimated from lung cancer survival data from SEER based on size and stage at detection (for men and women, and 4 subtypes separately). Additional parameters, regarding lag time between first malignant cell and clinical detection, maximum screen-detectability window, CT and CXR screen-detection thresholds are determined through calibration to NLST incidence and mortality outcomes via multi-dimensional grid search. The natural history parameter c1, which represents the fraction of lethal metastasis when it becomes observable, is originally estimated by using SEER survival but further calibrated to NLST stage distribution to reflect the possible advancements in imaging technologies to detect metastasis earlier. The Stanford Model uses the NLST eligible population in PLCO for validation. To simulate the entire PLCO population, the underlying histology proportions are modified considering the effect of never and lighter smokers in PLCO.

Comparison with other literature

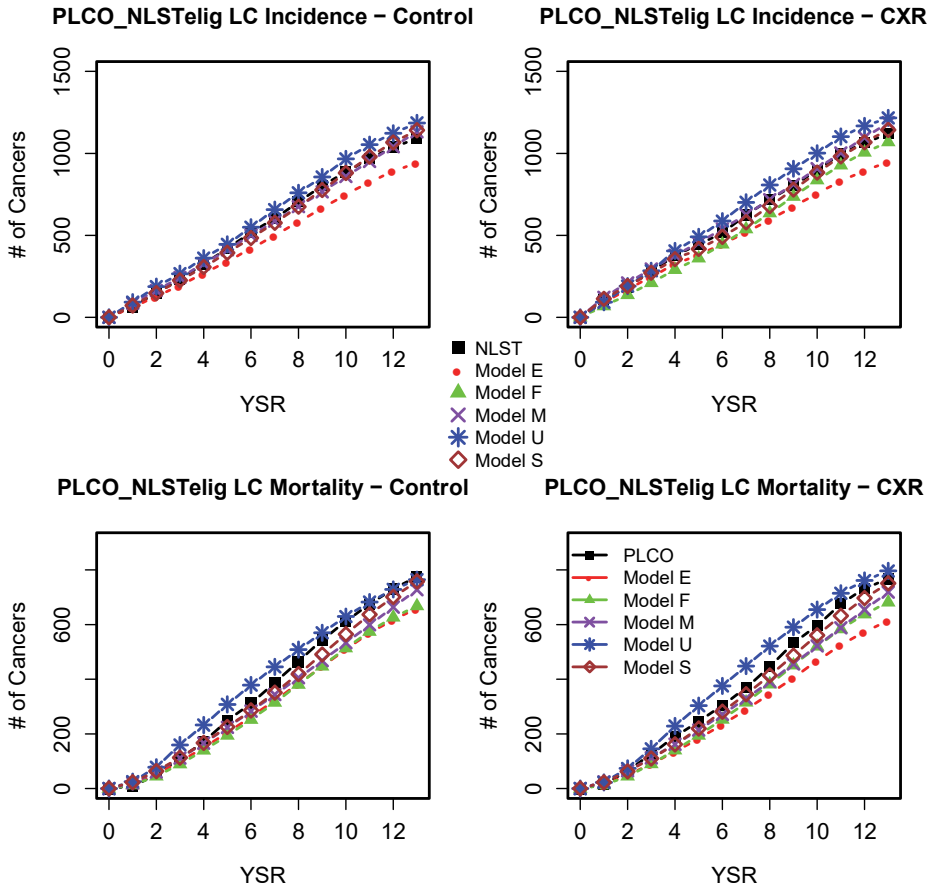
There are only a handful of existing lung cancer screening natural history models in the literature; many of them were developed as earlier versions of the models described in this paper. In particular McMahon et al derived an earlier version of the MGH-HMS model (M) using data from the Mayo CT and the Mayo Lung Project (MLP) studies. This model has been used to assess the effectiveness and cost-effectiveness of low-dose CT screening and to estimate the reduction in lung-cancer mortality and the numbers of radiation induced cancers from annual low-dose CT LC screening.[16-19] Lin et. al. developed an earlier version of the Stanford Model (S) based on SEER lung cancer incidence and validated against the MLP study, and used it to assess the potential benefit of screening in reducing LC mortality at the population level.[28] Goldwasser et. al. developed a lung cancer screening model calibrated to the initial MLP data to estimate the excess risk due to X-rays radiation in the intervention arm, and recently updated the model using data from the Mayo CT study to estimate the tumor diameter at cure threshold among aggressive non-small cancers.[29,30] Schultz et. al. developed an earlier version of the Erasmus-MISCAN model (E) based on SEER, and the MLP.[6] The models of McMahon et. al., Lin et. al., Goldwasser et. al. and Schultz et. al. were recently used simultaneously to model the Mayo CT screening trial and estimated jointly a 20-30% mortality reduction relative to no screening at 5 years after study enrollment (unpublished). Hazelton et. al. developed an earlier version of the Fred Hutchinson lung cancer-screening model (F) using data from the PLuSS CT and CARET studies. This model was used to assess the aggressiveness of lung cancers detected by CT screening and the fraction of indolent cancers found by screening.[15] Foy et. al. used a recently developed lung cancer risk model to evaluate the benefit of CT screening in the NY-ELCAP study.[31]

Figure S6: Observed and predicted LC mortality reduction in the NLST by Years Since Randomization (YSR)



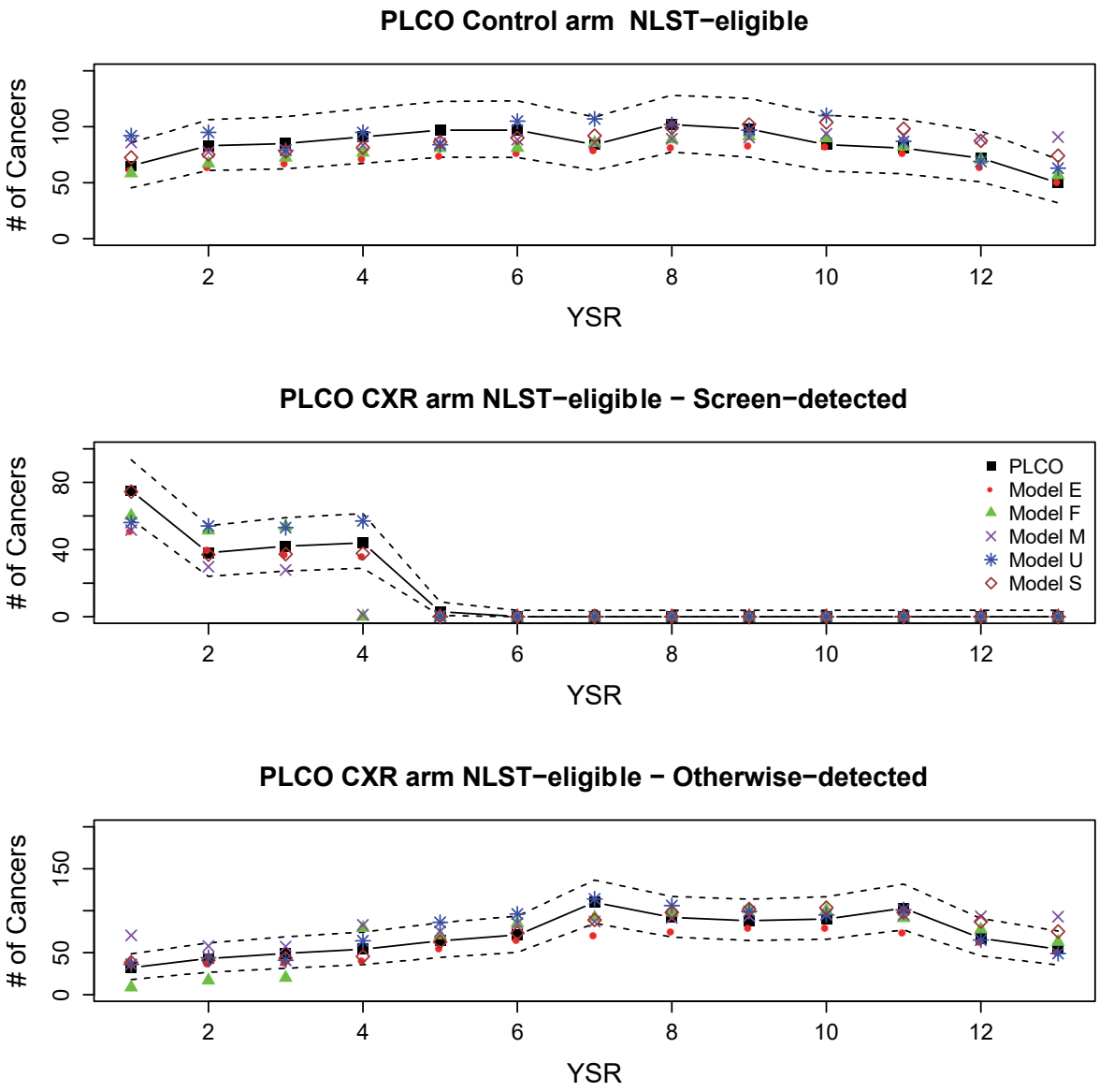
Bars denote 95% confidence intervals.

Figure S7: PLCO-NLST-eligible observed vs. model predicted LC cases and deaths by arm and Years Since Randomization (YSR)



Dashed lines represent 95% binomial confidence intervals for the observed values.

Figure S8: PLCO-NLST-eligible observed versus model predicted LC cases by arm, mode of detection and Years Since Randomization (YSR)



Dashed lines represent 95% binomial confidence intervals for the observed values.

References

1. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Analysis* 2012; **32 Suppl 1**: S51-68.
2. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.
3. van Ravesteyn NT, Miglioretti DL, Stout NK, et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: A comparative modeling study of risk. *Annals of Internal Medicine* 2012; **156**(9): 609-17.
4. Wever EM, Draisma G, Heijnsdijk EAM, et al. Prostate-Specific Antigen Screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer–Rotterdam. *Journal of the National Cancer Institute* 2010; **102**(5): 352-5.
5. CISNET Lung Cancer Model Profiles, Erasmus Medical Center. Available from URL: http://cisnet.cancer.gov/lung/profiles.html [accessed 07/12, 2012].
6. Schultz FW, Boer R, de Koning HJ. Chapter 7: Description of MISCAN-Lung, the Erasmus MC Lung Cancer Microsimulation Model for Evaluating Cancer Control Interventions. *Risk Analysis* 2012; **32**: S85-S98.
7. Surveillance, epidemiology, and end results (SEER) program (www.seer.cancer.gov) SEER*Stat database: Incidence - SEER-17 regs research data + hurricane katrina impacted louisiana cases, nov 2010 sub (2000-2008) <Single ages to 85+, Katrina/Rita population adjustment> - linked to county attributes - total U.S., 1969-2009 counties, national cancer institute, DCCPS, surveillance research program, cancer statistics branch, released april 2011 (updated 10/28/2011), based on the november 2010 submission.
8. Heidenreich WF, Luebeck EG, Moolgavkar SH. Some Properties of the Hazard Function of the Two-Mutation Clonal Expansion Model. *Risk Analysis* 1997; **17**(3): 391-9.
9. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.
10. ICD-O-3 SEER Site/Histology Validation List - 12/04/2009 - updated from 2/9/2001. Available from URL: http://seer.cancer.gov/icd-o-3/ .
11. Greene FL, Page DL, Fleming ID, Fritz AG, Balch CM, Haller DG, et al. eds, for the American joint committee on cancer. AJCC Cancer Staging Manual. 6th ed. New York, NY: Springer-Verlag; 2002.
12. Aberle DR, DeMello S, Berg CD, et al. Results of the Two Incidence Screenings in the National Lung Screening Trial. *New England Journal of Medicine* 2013; **369**(10): 920-31.
13. Surveillance, epidemiology, and end results (SEER) program (www.seer.cancer.gov) SEER*Stat database: Incidence - SEER-17 regs research data + hurricane katrina impacted louisiana cases, nov 2010 sub (1973-2008 varying) - linked to county attributes - total U.S., 1969-2009 counties, national cancer institute, DCCPS, surveillance research program, cancer statistics branch, released april 2011 (updated 10/28/2011), based on the november 2010 submission. .
14. Barton RR, Ivey JS. Nelder-Mead Simplex Modifications for Simulation Optimization. *Management Science* 1996; **42**(7): 954-73.
15. Hazelton WD, Goodman G, Rom WN, et al. Longitudinal multistage model for lung cancer incidence, mortality, and CT detected indolent and aggressive cancers. *Mathematical Biosciences* 2012; **240**(1): 20-34.
16. McMahon PM, Kong CY, Bouzan C, et al. Cost-Effectiveness of Computed Tomography Screening for Lung Cancer in the United States. *Journal of Thoracic Oncology* 2011; **6**(11): 1841-8.
17. McMahon PM, Kong CY, Johnson BE, et al. Estimating long-term effectiveness of lung cancer screening in the Mayo CT screening study. *Radiology* 2008; **248**(1): 278-87.
18. McMahon PM, Kong CY, Weinstein MC, et al. Adopting helical CT screening for lung cancer: potential health consequences during a 15-year period. *Cancer* 2008; **113**(12): 3440-9.
19. Kong CY, Lee JM, McMahon PM, et al. Using radiation risk models in cancer screening simulations: important assumptions and effects on outcome projections. *Radiology* 2012; **262**(3): 977-84.

20. 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.
21. Cressie N, Timothy RCR. Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society Series B (Methodological)* 1984; **46**(3): 440-64.
22. Kong CY, McMahon PM, Gazelle GS. Calibration of disease simulation model using an engineering approach. *Value Health* 2009; **12**(4): 521-9.
23. Hazelton WD, Clements MS, Moolgavkar SH. Multistage Carcinogenesis and Lung Cancer Mortality in Three Cohorts. *Cancer Epidemiology Biomarkers & Prevention* 2005; **14**(5): 1171.
24. Hazelton WD, Luebeck EG, Heidenreich WF, Moolgavkar SH. Analysis of a Historical Cohort of Chinese Tin Miners with Arsenic, Radon, Cigarette Smoke, and Pipe Smoke Exposures Using the Biologically Based Two-Stage Clonal Expansion Model. *Radiation Research* 2001; **156**(1): 78-94.
25. Daniel TG, Linda RP. Improved leap-size selection for accelerated stochastic simulation. *The Journal of Chemical Physics* 2003; **119**(16): 8229-34.
26. Desmond JH. Modeling and Simulating Chemical Reactions. *SIAM Review* 2008; **50**(2): 347-68.
27. Statistical Methodology and Applications Branch, Data Modeling Branch, National Cancer Institute. Cansurv 1.1. May 2012.
28. Lin RS, Plevritis SK. Comparing the benefits of screening for breast cancer and lung cancer using a novel natural history model. *Cancer Causes & Control* 2012; **23**(1): 175-85.
29. Goldwasser DL, Kimmel M. Modeling excess lung cancer risk among screened arm participants in the Mayo Lung Project. *Cancer* 2010; **116**(1): 122-31.
30. Goldwasser DL, Kimmel M. Small median tumor diameter at cure threshold (<20 mm) among aggressive non-small cell lung cancers in male smokers predicts both chest X-ray and CT screening outcomes in a novel simulation framework. *International Journal of Cancer* 2013; **132**(1): 189-97.
31. Foy M, Yip R, Chen X, Kimmel M, Gorlova OY, Henschke CI. Modeling the mortality reduction due to computed tomography screening for lung cancer. *Cancer* 2011; **117**(12): 2703-8.

Chapter 2

Lung cancer detectability by test, histology, stage, and gender: estimates from the NLST and the PLCO trials

Kevin ten Haaf, Joost van Rosmalen, Harry J. de Koning

Published as:

ten Haaf K, van Rosmalen J, de Koning HJ.

Cancer Epidemiology Biomarkers & Prevention 2015; **24**(1): 154-61.

Abstract

Background: Implementing optimal lung cancer screening programs requires knowledge of the natural history and detectability of lung cancer. This information can be derived from the results of clinical trials with the aid of microsimulation models.

Methods: Data from the Surveillance, Epidemiology, and End Results (SEER) program and individual-level data from the National Lung Screening Trial (NLST) and the Prostate, Lung, Colon, and Ovarian Cancer Screening trial (PLCO) were used to investigate the sensitivity (by histology and stage) of CT and chest radiography (CXR) and the mean preclinical sojourn time (MPST) of lung cancer (by gender, histology, and stage). The MISCAN-Lung model was used to reproduce the lung cancer incidence by method of detection (clinically or screen-detected), gender, histology, and stage in both trials and SEER, by calibrating CT and CXR sensitivity and natural history parameters.

Results: CT sensitivity ranges from 8.83% to 99.35% and CXR sensitivity from 2.51% to 97.31%, depending on histology and stage. CT sensitivity for stage IA is more than 3-fold higher compared with CXR, for all histologies. The total MPST estimates for lung cancer progressing through preclinical stages IA to IV range from 3.09 to 5.32 years for men and 3.35 to 6.01 years for women. The largest difference in total MPST between genders was estimated for adenocarcinoma.

Conclusions: We estimate longer MPSTs for lung cancer compared with previous research, suggesting a greater window of opportunity for lung cancer screening.

Impact: This study provides detailed insights into the natural history of lung cancer and CT screening effectiveness.

Introduction

Lung cancer causes 1.4 million deaths per year globally, accounting for 18% of the total number of cancer deaths.[1] Randomized controlled trials, which evaluated the benefits of lung cancer screening with chest radiography (CXR), such as the Prostate, Lung, Colorectal and Ovarian Cancer Screening trial (PLCO), did not show a significant reduction in lung cancer mortality.[2] In contrast, the results of the National Lung Screening Trial (NLST) show that lung cancer mortality may be reduced by screening with CT.[3] However, the knowledge of the natural history and detectability of lung cancer remains limited. Information on the natural history and detectability of lung cancer cannot be derived directly from observed data, as the onset and progression of preclinical disease is unobservable. This information can only be obtained by modeling disease onset, progression, and detection. With the aid of microsimulation models, results of clinical trials can be used to derive information on the natural history and detectability of lung cancer. Estimates of the mean preclinical sojourn times (MPST) and sensitivities of screening modalities for lung cancer are essential for the extrapolation of clinical trials. These extrapolations allow the investigation of screening policies that deviate from those investigated in the trials, which may aid policy makers in developing optimal screening policies, as was shown recently.[4] Furthermore, studies on other cancers have shown that both the MPST of the disease and the sensitivity of the screening test influence the efficiency of screening policies.[5]

Previous studies obtained estimates for the MPST and detectability of lung cancer based on data from epidemiologic studies and clinical trials, using simulation or analytical modeling.[6-9] However, these studies did not investigate differences in lung cancer MPST by gender, histology, and stage or investigate the detectability by histology and stage. Because of the large number of participants and the use of CT and CXR as screening modalities, the NLST and PLCO provide detailed data, which allow in-depth analyses on the natural history and detectability of lung cancer. In this study, we estimate the MPST by gender, histology and stage, and CT and CXR sensitivity for lung cancer by histology and stage using the MISCAN-Lung model, based on data from the Surveillance, Epidemiology, and End Results (SEER) Program and individual-level data from the NLST and PLCO.

Methods

Data

NLST

The NLST was a randomized controlled screening trial, which compared lung cancer screening with low-dose CT to CXR.[10] 53,454 current or former smokers (who had quit within the previous 15 years) ages 55 to 74 years with a smoking history of at least 30 pack-years were randomized to three annual screenings with CT or CXR. The NLST showed a relative reduction in lung cancer mortality of 20% by screening with CT compared with CXR, after six years of follow-up.[3] NLST data were available on the individual level.

PLCO

The PLCO was a randomized controlled screening trial which investigated screening for prostate, lung, colorectal, and ovarian cancer.[11] The lung component of the trial randomized 154,901 participants ages 55 to 74 years to either receive four annual CXR screenings or usual care (i.e., no screening). The data from the PLCO control arm provide information on the natural history of lung cancer when screening does not occur. A combination of detailed data from screened and nonscreened populations is essential to accurately assess the lead-time achieved by screening and the potential for overdiagnosis. The PLCO had no eligibility criteria with regard to smoking history; however, participants who had never smoked and were randomized after April 1995 were not invited to the fourth screening round.[2] The PLCO did not show a significant reduction in lung cancer mortality by screening with CXR compared with usual care after 13 years of follow-up.[2] Data from the lung component of the PLCO were available on the individual level.

SEER

SEER is a source for cancer statistics from various U.S. states, providing information on the incidence of lung cancer in the absence of screening. Data on lung cancer incidence by 5-year age-groups for ages 25 to 84, stage, histology, and gender were extracted from the SEER-17 database for years 2004 to 2008.[12]

MISCAN-Lung

The MISCAN-Lung model used in this analysis was developed within the Cancer Intervention and Surveillance Modeling Network (CISNET) and is an extension of the model used to evaluate the impact of tobacco control on U.S. lung cancer mortality.[13] The model was used to simulate NLST and PLCO, using individual-level data on smoking behavior, for six and 13 years of follow-up, respectively.[14] The SEER-17 population was simulated for years 2004 to 2008, using the National Cancer Institute's Smoking History Generator to generate smoking behavior.[15] These data were used to modify the population parameters to reflect the populations of the NLST and PLCO trials and SEER-17. The calibration targets for NLST, PLCO, and SEER-17 have been described previously and can be found in Chapter 1 of this thesis.[14] All model parameters were calibrated simultaneously using the Nelder-Mead optimization algorithm.[16] The MISCAN-Lung model was first calibrated to the NLST and then validated on the NLST-eligible individuals in the PLCO, as demonstrated in Figure 7 in Meza and colleagues (Chapter 1 of this thesis).[14] However, additional calibration was required to adequately replicate the outcomes in never smokers and light smokers in the PLCO.[14] The recent analyses of the CISNET-Lung working group incorporated the version of the MISCAN-Lung model described in this investigation.[4,14,17]

Lung carcinogenesis

Lung carcinogenesis is modeled using the two-stage clonal expansion model (TSCE) as described by Heidenreich, in contrast to an earlier version of MISCAN-Lung, which implemented an approximation of this model.[13,18,19] The TSCE estimates a person's risk of lung cancer, as a function of age and smoking history. The TSCE parameters were estimated by Meza, using data from the Nurses' Health Study and the Health Professionals Follow-up Study.[19] However, the gender-specific parameters for malignant transformation were recalibrated to NLST, PLCO, and SEER-17.[14] In SEER-17 simulations, we accounted for the lower age-adjusted incidence in the SEER-17 states compared with the U.S. population (6.1% for women and 10.9% for men.[20,21]

Histologies

The model distinguishes four histologies (compared with three histologies in the previous version of MISCAN-Lung), based on the International Classification of Diseases for Oncology,

Third Edition (ICD-O-3): adenocarcinoma, squamous cell carcinoma, other non–small cell carcinoma, and small cell carcinoma.[22] An overview of the histology classifications by ICD-O-3 codes is given in the supplementary material of this Chapter, in Table S1. Large cell carcinomas were included in the adenocarcinoma histology as "*a significant proportion of large cell carcinoma has immunohistochemical, cytogenetic, mutational, and gene expression profiles that overlap with adenocarcinomas, which may reflect a common cell of origin*".[23] Recent guidelines reclassify bronchioalveolar carcinoma (BAC) into adenocarcinoma in situ (AIS), invasive adenocarcinoma, and minimally invasive adenocarcinoma (MIA).[24] Therefore, BACs were included in the adenocarcinoma histology. 147 cancers were denoted as BACs in NLST, 45 stage IB or higher, and thus invasive adenocarcinoma. Of 102 persons with stage IA BACs (79.41% screen-detected), 101 received treatment and eight died from lung cancer. 188 cancers were denoted as BACs in PLCO, 109 stage 1B or higher. Of 79 persons with stage IA BACs (24.05% screen-detected), 75 were known to have received treatment and 11 died from lung cancer. Most stage IA BACs found in PLCO were detected clinically, which suggests most of these cases were invasive adenocarcinoma. Although most IA BACs in NLST were detected by screening, it is unknown what proportion may have been AIS or MIA, as only a subset of BACs in NLST were reanalyzed using the recent guidelines.[25] Therefore, it is unclear which stage IA BACs in NLST and PLCO would have been classified as AIS or MIA, which precludes these categories from separate analysis.

Stage progression

Cancers are assumed to progress sequentially from less advanced to more advanced preclinical stages, as shown in the supplementary material of this Chapter, in Figure S1. Six stages are distinguished (compared with two stages in the previous version of MISCAN-Lung), based on the American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 6th edition: IA, IB, II, IIIA, IIIB, and IV.[26] Stages IA1 and IA2 were not considered in this edition. Stage IA was not divided into stages IA1 and IA2 because data for these stages were available for NLST, but not for PLCO or the extracted SEER-17 data. The probability that a cancer progresses to a more advanced preclinical stage or is diagnosed clinically (e.g., diagnosed due to symptoms) is detailed by histology and stage in the supplementary material of this Chapter, in Table S2. Only lung cancers with known stages were taken into

account, with the exception of small cell carcinoma in PLCO. Small cell carcinomas in PLCO were distinguished as either limited or extensive instead of stages based on the AJCC criteria. Therefore, we did not distinguish small cell carcinoma by stage for PLCO.

Mean preclinical sojourn times

The preclinical duration (in the absence of screening) of each stage (by histology) is assumed to follow a Weibull distribution, for which the mean and shape parameters are estimated. The MPST for each stage and histology is modeled as a multiplicative function of stage, histology, and gender-specific duration parameters. The duration parameters for men and other non–small cell carcinoma serve as the baseline durations. For example, the mean duration of stage IA adenocarcinoma for men is modeled as:

$$Duration\_AD\_IA\_M = DurationAD*DurationIA$$

whereas the mean duration for women is modeled as:

$$Duration\_AD\_IA\_F = DurationAD*DurationIA*DurationWomenAD$$

where *DurationWomenAD* measures the relative difference in MPST between women and men for adenocarcinoma. The shape parameters and the dependence between the durations of the preclinical states within an individual were assumed to be similar for all stages, histologies, and both genders; the values of these parameters are shown in the supplementary material of this Chapter, in Table S3.

Sensitivities

The probability that a lung cancer is detected by screening (including additional diagnostic evaluations) is modeled as a mathematical function of the screening modality, histology, stage, and screening round. Screening round was included as a dichotomous predictor variable in this model to account for the fact that participants with abnormalities detected in the third round of NLST may have been evaluated more aggressively compared with previous rounds.[27] These parameters were assumed not to differ by histology, stage, or

gender. The sensitivity of each screening modality is restricted to be higher for more advanced stages, i.e., the CT sensitivity for stage IB adenocarcinoma is higher than stage IA adenocarcinoma. This is accomplished by modeling the sensitivity indices as additive functions of model parameters. For example, the stage-specific CT sensitivity index for stage IB adenocarcinoma is modeled as the sum of the parameters for the CT sensitivity of stages IB adenocarcinoma and IA adenocarcinoma:

$$SensADIB\_CT = Sensitivity\_AD\_IB\_CT + Sensitivity\_AD\_IA\_CT$$

The sensitivity is further restricted between 0 and 1 by incorporating the stage-specific sensitivity parameters in inverse-logit (expit) functions of the sensitivity indices. Thus, the CT sensitivity for stage IB adenocarcinoma in round 2 is modeled as:

$$\frac{\exp(SensADIB\_CT)}{1 + \exp(SensADIB\_CT)}$$

Assessment of statistical uncertainty

Profile likelihood confidence intervals were calculated using an approach similar to Wever's.[28]

Results

The sensitivity estimates, shown in Table 1 (for the four screening rounds of the PLCO and the first two rounds of the NLST), indicate that CT sensitivity ranges from 8.83% to 99.35% and CXR sensitivity from 2.51% to 97.31%. The profile likelihood confidence intervals suggest that the plausible range of the CT sensitivities ranges from 2.32% lower to 4.06% higher compared with the sensitivities shown in Table 1 (e.g., the estimated CT sensitivity of stage IA adenocarcinoma is 56.63%, but may vary between 54.31% and 60.69%). The profile likelihood confidence intervals suggest that the plausible range of the CXR sensitivities ranges from 0.48% lower to 1.79% higher compared with the sensitivities shown in Table 1 (e.g., the estimated CXR sensitivity of stage IA adenocarcinoma is 16.91%, but may vary between 16.43% and 18.70%). CT and CXR sensitivity varies greatly by histology, particularly

for the early stages. For example, CT sensitivity for stage IA squamous cell carcinoma is 30.95% compared with 20.78% for stage IA other non–small cell carcinoma, whereas CXR sensitivity for stage IA adenocarcinoma is 16.91% compared with 2.51% for stage IA small cell carcinoma. However, CT is more sensitive for each stage and histology compared with CXR. This is especially notable for stage IA, where CT sensitivity is more than 3-fold higher compared with CXR, for all histologies. Furthermore, even at advanced stages such as IIIA and IIIB, a 23% to 34% difference in sensitivity remains. There is a notable difference in CXR sensitivity for stage IV other non–small cell carcinoma compared with other histologies, which may reflect an atypical presentation that may be difficult to detect using CXR. The sensitivity for the third round of NLST is estimated to have been up to 14.77% and 0.10% higher for CT and CXR respectively compared with previous rounds, as shown the supplementary material of this Chapter, in Table S4.

Table 1: Sensitivity estimates by screening modality

CT				
	AD	SQ	SM	OTH
IA	16.91%	9.72%	2.51%	6.27%
IB	27.13%	28.90%	4.25%	7.57%
II	27.26%	30.02%	6.64%	7.57%
IIIA	48.11%	46.31%	14.74%	29.78%
IIIB	49.29%	47.96%	53.18%	34.40%
IV	96.31%	78.62%	97.31%	36.94%
CXR				
	AD	SQ	SM	OTH
IA	56.63%	30.95%	8.83%	20.78%
IB	64.12%	38.05%	10.28%	24.75%
II	64.48%	39.19%	11.19%	24.78%
IIIA	75.93%	69.67%	41.58%	60.40%
IIIB	80.21%	79.39%	87.06%	68.27%
IV	98.88%	97.66%	99.35%	95.67%

Abbreviations: AD, adenocarcinoma; SQ, squamous cell carcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Table 2 shows the MPST by gender, histology, and stage. The MPST estimates in Table 2 should be interpreted as follows: the time for an adenocarcinoma cancer to progress from preclinical stage IA to preclinical stage II (or be clinically detected in stage IB) in a male is on average 2.46 (1.82 + 0.64) years, of which 1.82 years is spent in the preclinical state of stage

IA and 0.64 years is spent in the preclinical state of stage IB. The profile likelihood confidence intervals suggest that the plausible range of the MPST estimates ranges from 1.95% shorter to 10.60% longer compared with the estimates shown in Table 2 (e.g., the estimated MPST for stage IA adenocarcinoma in men is 1.82 years, but may vary between 1.78 and 2.01 years). The total MPST estimates of lung cancer progressing through preclinical stages IA to IV range from 3.09 to 5.32 years for men and 3.35 to 6.01 years for women. Small cell carcinomas are estimated to have the shortest MPST for both genders: the MPST in the preclinical stages IA to IV was estimated to be 3.09 years for men compared with 3.35 for women. The largest difference between genders was estimated for adenocarcinoma, for which the MPST in the preclinical stages IA to IV was estimated to be 4.48 years for men compared with 6.01 years for women. In contrast, there is little difference between genders for squamous cell carcinoma. The MPST for all histologies is estimated to be the longest in stage IA followed by stage IV. The MPST ranges for stage IA are estimated to be 1.25 to 2.16 years for men and 1.36 to 2.44 years for women, depending on the histology. An overview of the estimates of the histology and gender-specific duration parameters can be found in the supplementary material of this Chapter, in Table S3.

Table 2: MPST estimates (in years) of preclinical stages by gender*

Men				
	AD	SQ	SM	OTH
IA	1.82	2.16	1.25	1.96
IB	0.64	0.76	0.44	0.69
II	0.46	0.55	0.32	0.5
IIIA	0.46	0.55	0.32	0.5
IIIB	0.36	0.42	0.25	0.39
IV	0.74	0.88	0.51	0.8
Total mean preclinical duration**	4.48	5.32	3.09	4.84

Women				
	AD	SQ	SM	OTH
IA	2.44	2.15	1.36	2.31
IB	0.86	0.76	0.48	0.81
II	0.62	0.55	0.34	0.59
IIIA	0.62	0.55	0.35	0.59
IIIB	0.48	0.42	0.27	0.45
IV	0.99	0.88	0.55	0.94
Total mean preclinical duration**	6.01	5.31	3.35	5.69

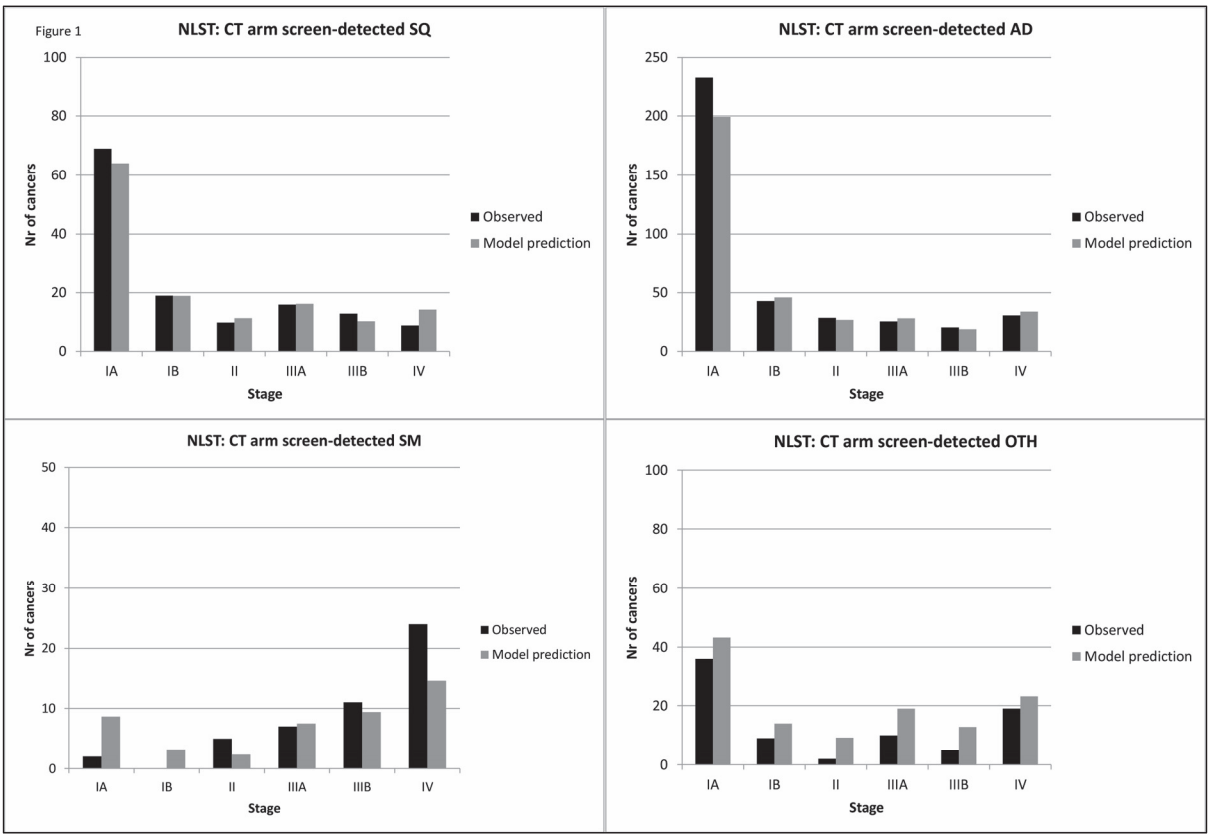
Abbreviations: AD, adenocarcinoma; SQ, squamous cell carcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

**The MPST estimates should be interpreted as follows: the time for an adenocarcinoma cancer to progress from preclinical stage IA to preclinical stage II (or be clinically detected in stage IB) in a male is on average 2.46 (1.82+0.64) years, of which 1.82 years are spent in the preclinical state of stage IA and 0.64 years are spent in the preclinical state of stage IB.*

***If discovered clinically in stage IV*.

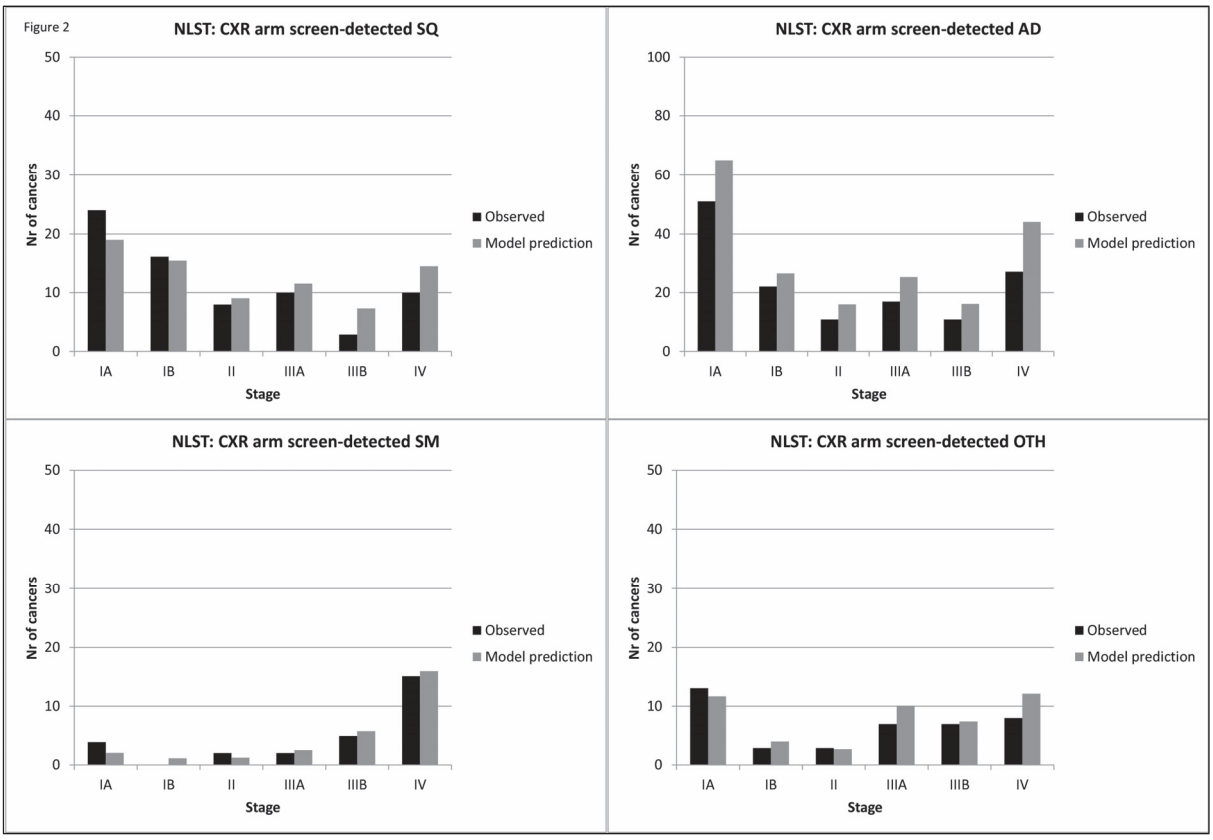
Figures 1–3 show the model fit to the screen-detected lung cancers in NLST and PLCO. The results for the number of otherwise detected cancers for both trials, the PLCO control arm and SEER, are shown in the supplementary material of this Chapter, in Figures S2-S7. The ranges of plausible values for the sensitivities and MPST estimates can be found the supplementary material of this Chapter, in the section: “Range of plausible values”.

Figure 1: NLST CT arm screen-detected lung cancers by histology and stage



Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Figure 2: NLST CXR arm screen-detected lung cancers by histology and stage

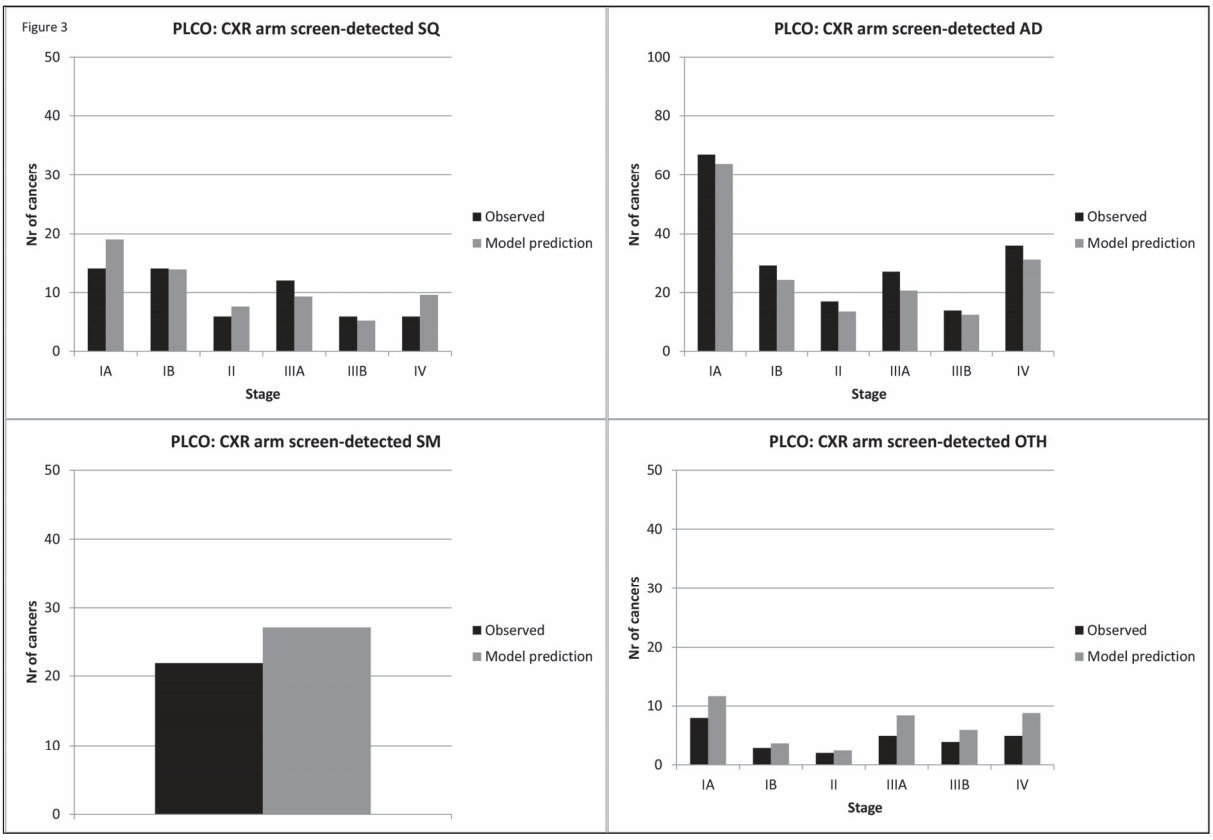


Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Discussion

The NLST and PLCO indicated that lung cancer mortality may be reduced by screening with CT, in contrast to screening with CXR, which may be due to a difference in sensitivity between CT and CXR. However, differences in sensitivity by stage and histology cannot be easily derived from the observed data. Our model enables us to quantify the differences in CT and CXR sensitivity by histology and stage. The main differences between CT and CXR sensitivity are found for the early stages, in particular stage IA. This difference may partially explain the difference in lung cancer mortality reduction between the CT and CXR arms of NLST, as even between stages IA and IB substantial differences in survival exist: persons diagnosed in stage IA have a median survival of 59 months compared with a median survival of 42 months for persons diagnosed in stage IB as indicated in Figure 7A of Groome and colleagues.[29]

Figure 3: PLCO CXR arm screen-detected lung cancers by histology and stage



Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

The differences between histologies indicate that the effects of screening may depend on the relative incidence present in the screened population. The low CT sensitivity for early-stage small cell carcinoma combined with its short MPST suggests that the relative benefits of screening may be lower in populations with a large proportion of small cell carcinoma.

Our study gives some indication on the effects of varying screening intervals. Stage IA cancers could be more difficult to detect with biennial or triennial strategies, as the MPST estimates for stage IA range from 1.25 to 2.16 years for men and 1.36 to 2.44 years for women, depending on the histology. As the survival of stage IA is considerably higher compared with other stages, this suggests that annual screening could lead to higher lung cancer mortality reductions compared with biennial or triennial screening.[29] This was also suggested in a recent analysis by the CISNET-Lung models, including MISCAN-Lung.[4] This study is the first to present sensitivity and MPST estimates for lung cancer by stage,

histology, and gender. However, the detectability of lung cancer has been investigated earlier. Pinsky estimated an MPST of 1.8 years for subjects clinically presenting with early-stage disease (stages I–II) and an MPST of 0.85 and 1.05 years in the early and late preclinical states respectively, for subjects clinically presenting with late-stage disease (stages III–IV).[6] CXR sensitivity estimates were 73% for early-stage cases and 93% for late-stage cases. Wu estimated an MPST of 2.24 years and a mean sensitivity of 89.4% for CXR, using Mayo Lung Project data. Chien estimated a median sojourn time of 2.06 years and a median CT sensitivity of 97%, based on a systematic literature review.[7] Patz estimated an MPST of 3.6 years for non–small cell lung cancers (excluding BACs) with mean CT and CXR sensitivities of 83% and 33% and an MPST of 32.1 years for BACs, with mean CT and CXR sensitivities of 38% and 4%, based on NLST.[9] The MPSTs for BACs are noticeably longer and the sensitivity estimates lower compared with other non–small cell lung cancers. However, these estimates are solely based on NLST, where both arms received screening. The lack of detailed data from a nonscreened population prevents an accurate assessment of the lead-time achieved by screening and the potential for overdiagnosis. As noted previously, most BACs found in NLST were screen-detected, found at an early stage and treated. This altered their natural history and may have influenced the BAC estimates. The same holds for non-BAC, non–small cell lung cancers found in NLST, but to a lesser extent, as 46.44% of them were screen-detected compared with 73.47% of BACs.[9] Furthermore, some adenocarcinomas found in NLST may have originated from BACs which progressed to adenocarcinoma, which is not taken into account in their investigation. Finally, due to the statistical dependence between the MPST and the sensitivity of the screening modality, a model estimating a low sensitivity and a long MPST may provide a similar fit compared with a model estimating a high sensitivity and a short MPST.[30] With higher sensitivity and shorter MPST estimates for BAC, a similar fit to NLST could have been obtained.

We estimate longer MPSTs for lung cancer compared with previous research, suggesting a greater window of opportunity for lung cancer screening. Noticeably, in contrast to previous studies, our model incorporates smoking history in the carcinogenesis process. A limitation of our study is the uncertainty in the parameter estimates caused by the statistical dependence between the MPST and sensitivity of the screening modalities.[30] However, we believe the joint fit of the natural history of lung cancer and the sensitivities of two different

screening modalities on multiple trials helps limit this uncertainty. Moreover, the incorporation of data from historical registries and the detailed individual-level data of the control arm of the PLCO in the calibration provides information on the natural history of lung cancer in the absence of screening, which is essential to accurately assess the lead-time achieved by screening and potential for overdiagnosis. Another limitation is that the estimated screening sensitivities and MPSTs are largely based on NLST and PLCO. However, these are the largest studies on lung cancer screening with CT and CXR, which makes them invaluable in investigating the natural history and detectability of lung cancer. Although differences in smoking prevalence exists between the NLST and the PLCO, the TSCE accounts for these differences in estimating a person's risk of lung cancer, through the use of individual-level data on smoking history, which allows us to adjust for differences in smoking prevalence between the two trials.

The model presented here has been used to analyze the harms and benefits of CT lung cancer screening for a 1950 birth cohort in the United States and will be used to investigate other populations.[4] Another topic for future research will be to validate the model on other lung cancer screening trials such as the Dutch–Belgian Lung Cancer Screening Trial (Nederlands-Leuvens Longkanker Screenings Onderzoek; the NELSON trial), which will provide information on the generalizability of our model.[31] In conclusion, this study provides detailed information on the detectability of lung cancer by histology, stage, and gender, in contrast to previous studies which had less detailed analyses, did not incorporate smoking in their models' carcinogenesis process, and used less extensive data sources.[6-9] This information will aid in the development of public health policies.

Acknowledgments

The authors thank their colleagues from the CISNET Lung working group (in particular R. Meza) and N. Horeweg (Department of Public Health, Erasmus MC) for providing useful comments. Furthermore, they also thank the NLST and PLCO investigators (in particular M.C. Tammemägi and W.C. Black) for useful comments and the NCI for access to NCI's data collected by the NLST and the PLCO trial. Finally, they also thank the NLST and PLCO study participants for their contributions to these studies.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: A Cancer Journal for Clinicians* 2011; **61**(2): 69-90.
2. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
3. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
4. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
5. Fracheboud J, Groenewoud JH, Boer R, et al. Seventy-five years is an appropriate upper age limit for population-based mammography screening. *International Journal of Cancer* 2006; **118**(8): 2020-5.
6. Pinsky PF. An Early- and Late-Stage Convolution Model for Disease Natural History. *Biometrics* 2004; **60**(1): 191-8.
7. Chien C-R, Chen TH-H. Mean sojourn time and effectiveness of mortality reduction for lung cancer screening with computed tomography. *International Journal of Cancer* 2008; **122**(11): 2594-9.
8. Wu D, Erwin D, Rosner GL. Sojourn time and lead-time projection in lung cancer screening. *Lung Cancer* 2011; **72**(3): 322-6.
9. Patz EF, Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Internal Medicine* 2014; **174**(2): 269-74.
10. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and Study Design. *Radiology* 2011; **258**(1): 243-53.
11. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Controlled Clinical Trials* 2000; **21**(6 Suppl): 273S-309S.
12. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER Stat Database: Incidence – SEER-17 Regs Research Data+Hurricane Katrina Impacted Louisiana Cases, Nov 2010 Sub (1973–2008 varying) – Linked To County Attributes – Total U.S., 1969–2009 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2011 (updated 10/28/2011), based on the November 2010 submission. Accessed Feb 24, 2012.
13. Schultz FW, Boer R, de Koning HJ. Chapter 7: Description of MISCAN-Lung, the Erasmus MC Lung Cancer Microsimulation Model for Evaluating Cancer Control Interventions. *Risk Analysis* 2012; **32**: S85-S98.
14. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
15. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Anaysisl* 2012; **32 Suppl 1**: S51-68.
16. Barton RR, Ivey JS. Nelder-Mead Simplex Modifications for Simulation Optimization. *Management Science* 1996; **42**(7): 954-73.
17. McMahon PM, Meza R, Plevritis SK, et al. Comparing Benefits from Many Possible Computed Tomography Lung Cancer Screening Programs: Extrapolating from the National Lung Screening Trial Using Comparative Modeling. *PLOS ONE* 2014; **9**(6): e99978.
18. Heidenreich WF, Luebeck EG, Moolgavkar SH. Some Properties of the Hazard Function of the Two-Mutation Clonal Expansion Model. *Risk Analysis* 1997; **17**(3): 391-9.
19. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.
20. Eheman C, Henley SJ, Ballard-Barbash R, et al. Annual Report to the Nation on the status of cancer, 1975-2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer* 2012; **118**(9): 2338-66.

21. SEER Stat Fact Sheets: Lung and Bronchus Cancer. [cited Accessed February 22nd 2013. Available from: http://seer.cancer.gov/statfacts/html/lungb.html.
22. Fritz A, Percy C, Jack A, et al. International classification of diseases for oncology. 3rd ed. Geneva: World Health Organization; 2000.
23. Eleazar JA, Borczuk AC. Chapter 23: molecular pathology of large cell carcinoma and its precursors. In: Zander DS, editor. Molecular pathology of lung diseases. Springer; 2010. p. 279–92.
24. Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *Journal of Thoracic Oncology* 2011; **6**(2): 244-85.
25. Pinsky PF, Church TR, Izmirlian G, Kramer BS. The National Lung Screening Trial: Results stratified by demographics, smoking history, and lung cancer histology. *Cancer* 2013; **119**(22): 3976-83.
26. Greene FL, Page DL, Fleming ID, et al. eds, for the American joint committee on cancer. AJCC Cancer Staging Manual. 6th ed. New York, NY: Springer-Verlag; 2002.
27. Aberle DR, DeMello S, Berg CD, et al. Results of the Two Incidence Screenings in the National Lung Screening Trial. *New England Journal of Medicine* 2013; **369**(10): 920-31.
28. Wever EM, Draisma G, Heijnsdijk EAM, et al. Prostate-Specific Antigen Screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer–Rotterdam. *Journal of the National Cancer Institute* 2010; **102**(5): 352-5.
29. Groome PA, Bolejack V, Crowley JJ, et al. The IASLC Lung Cancer Staging Project: Validation of the Proposals for Revision of the T, N, and M Descriptors and Consequent Stage-groupings in the Forthcoming (Seventh) Edition of the TNM Classification of Malignant Tumours. *Journal of Thoracic Oncology* 2007; **2**(8): 694-705.
30. van Ballegooijen M, Rutter CM, Knudsen AB, et al. Clarifying Differences in Natural History between Models of Screening. *Medical Decision Making* 2011; **31**(4): 540-9.
31. Horeweg N, van der Aalst CM, Thunnissen E, et al. Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *American Journal of Respiratory and Critical Care Medicine* 2013; **187**(8): 848-54.

Chapter 2

Supplementary material

Table S1: Classification of histologies by ICD-O-3 codes

MISCAN-Lung Histology	ICD-O-3 codes*[1]
AD	8012, 8013, 8014, 8140, 8141, 8143, 8147, 8200, 8201, 8230, 8231, 8250, 8251, 8252, 8253, 8254, 8255, 8260, 8310, 8323, 8480, 8481, 8490, 8550, 8570, 8571, 8572, 8573, 8574, 8575, 8576
SQ	8052, 8070, 8071, 8072, 8073, 8074, 8075, 8076, 8078, 8083, 8084
SM	8002, 8041, 8042, 8043, 8044, 8045
OTH	8000, 8001, 8003, 8004, 8005, 8010, 8011, 8015, 8020, 8021, 8022, 8030, 8031, 8032, 8033, 8034, 8035, 8046, 8050, 8051, 8120, 8121, 8122, 8123, 8124, 8240, 8241, 8242, 8243, 8244, 8245, 8246, 8249, 8320, 8430, 8510, 8551, 8560, 8562, 8800, 8801, 8802, 8803, 8804, 8805, 8806, 8810, 8811, 8813, 8814, 8815, 8830, 8890, 8891, 8894, 8895, 8896, 8900, 8901, 8902, 8910, 8912, 8972, 8973, 8980, 8981, 8982, 8990, 8991

Abbreviations: AD, adenocarcinoma; SQ, squamous cell carcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.
**Only invasive carcinoma were taken into account.*

Table S2: Transition probabilities by histology*

From	To	AD	SQ	SM	OTH
Preclinical IA	Preclinical IB	0.85	0.87	0.97	0.92
	Clinical detection IA	0.15	0.13	0.03	0.08
Preclinical IB	Preclinical II	0.88	0.85	0.97	0.94
	Clinical detection IB	0.12	0.15	0.03	0.06
Preclinical II	Preclinical IIIA	0.93	0.87	0.97	0.95
	Clinical detection II	0.07	0.13	0.03	0.05
Preclinical IIIA	Preclinical IIIB	0.87	0.81	0.89	0.87
	Clinical detection IIIA	0.13	0.19	0.11	0.13
Preclinical IIIB	Preclinical IV	0.76	0.65	0.80	0.80
	Clinical detection IIIB	0.24	0.35	0.20	0.20
Preclinical IV	Clinical detection IV	1.00	1.00	1.00	1.00

Abbreviations: AD, adenocarcinoma; SQ, squamous cell carcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.
**Parameters were estimated by model calibration. Transition probabilities are applicable when screening does not occur.*

Table S3: Additional duration parameter estimates

Parameter	Estimate (approximate lower and upper bounds*)
Weibull shape parameter (similar for all histologies and stages)	1.4411 (1.3131-1.6091)
Correlation between the duration of preclinical states in an individual (similar for all histologies and stages)	0.6403 (0.3515-0.7489)
SQ duration parameter	1.0974 (1.0264-1.2278)
AD duration parameter	0.9251 (0.8993-1.0360)
SM duration parameter	0.6339 (0.5933-0.8525)
Female SQ duration parameter	0.9987 (0.8104-1.1623)
Female AD duration parameter	1.3406 (1.2927-1.6582)
Female SM duration parameter	1.0884 (0.9234-1.4618)
Female OTH duration parameter	1.1753 (1.0347-1.3711)

Abbreviations: AD, adenocarcinoma; SQ, squamous cell carcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

**Profile likelihood confidence intervals were obtained by varying the value of one parameter at a time, holding all others constant at their maximum likelihood estimates. The deviance was defined as -2*(loglikelihood of the model – loglikelihood of the saturated model). Deviances were computed for a number of different values of each parameter. Quadratic functions were estimated based on these deviances. These functions were used to determine which values of the scaling parameters result in a deviance which is 3.84 (the critical value corresponding to the 95th percentile of a chi-square distribution with one degree of freedom) points higher compared to the minimum obtained through maximum likelihood estimation. These values of the parameters correspond to approximate lower and upper bounds for the parameters, as shown in Table S2.*

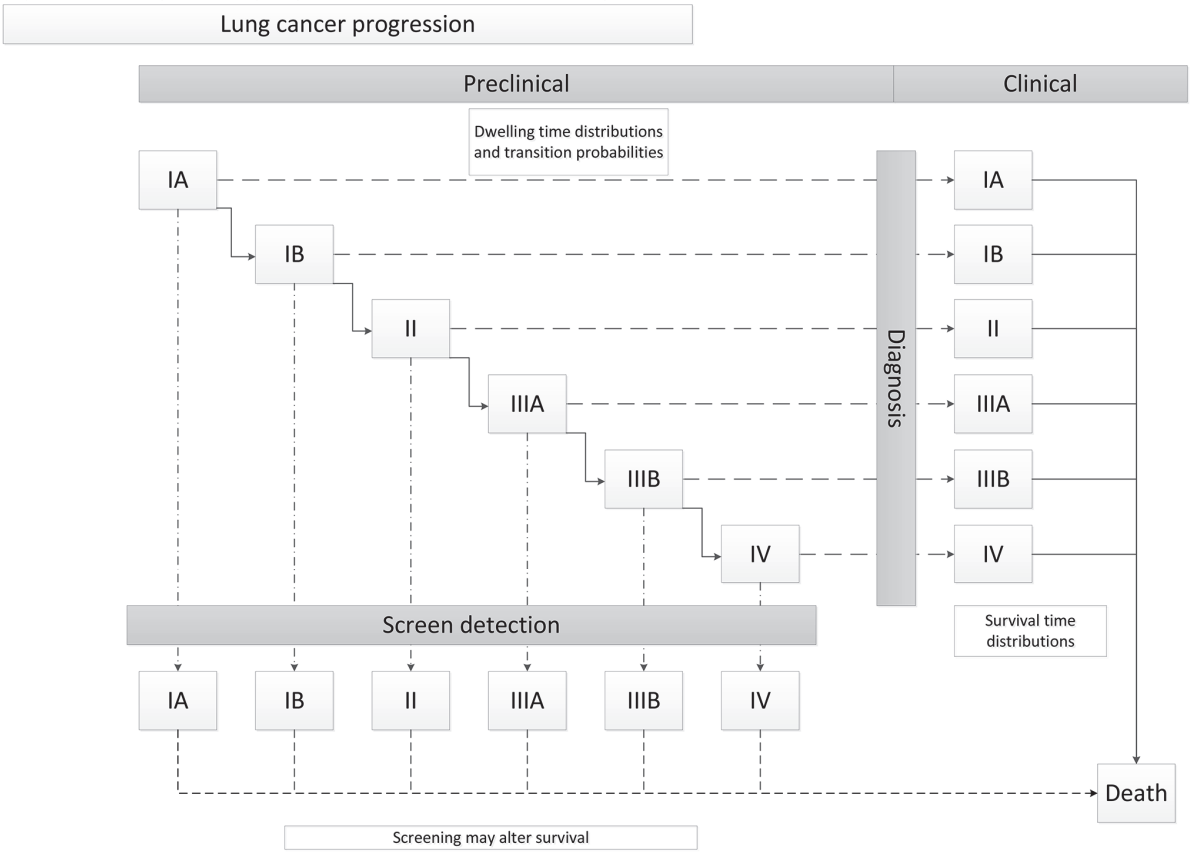
Table S4: Sensitivity estimates for round three in NLST by screening modality

CXR	AD	SQ	SM	OTH
IA	16.96%	9.75%	2.52%	6.29%
IB	27.21%	28.98%	4.27%	7.60%
II	27.34%	30.10%	6.66%	7.60%
IIIA	48.20%	46.40%	14.78%	29.86%
IIIB	49.39%	48.06%	53.28%	34.49%
IV	96.32%	76.68%	97.32%	37.03%

CT	AD	SQ	SM	OTH
IA	70.31%	44.84%	14.93%	32.23%
IB	76.42%	52.69%	17.20%	37.36%
II	76.70%	53.89%	18.59%	37.39%
IIIA	85.12%	80.64%	56.35%	73.44%
IIIB	88.02%	87.48%	92.42%	79.60%
IV	99.38%	98.69%	99.64%	97.57%

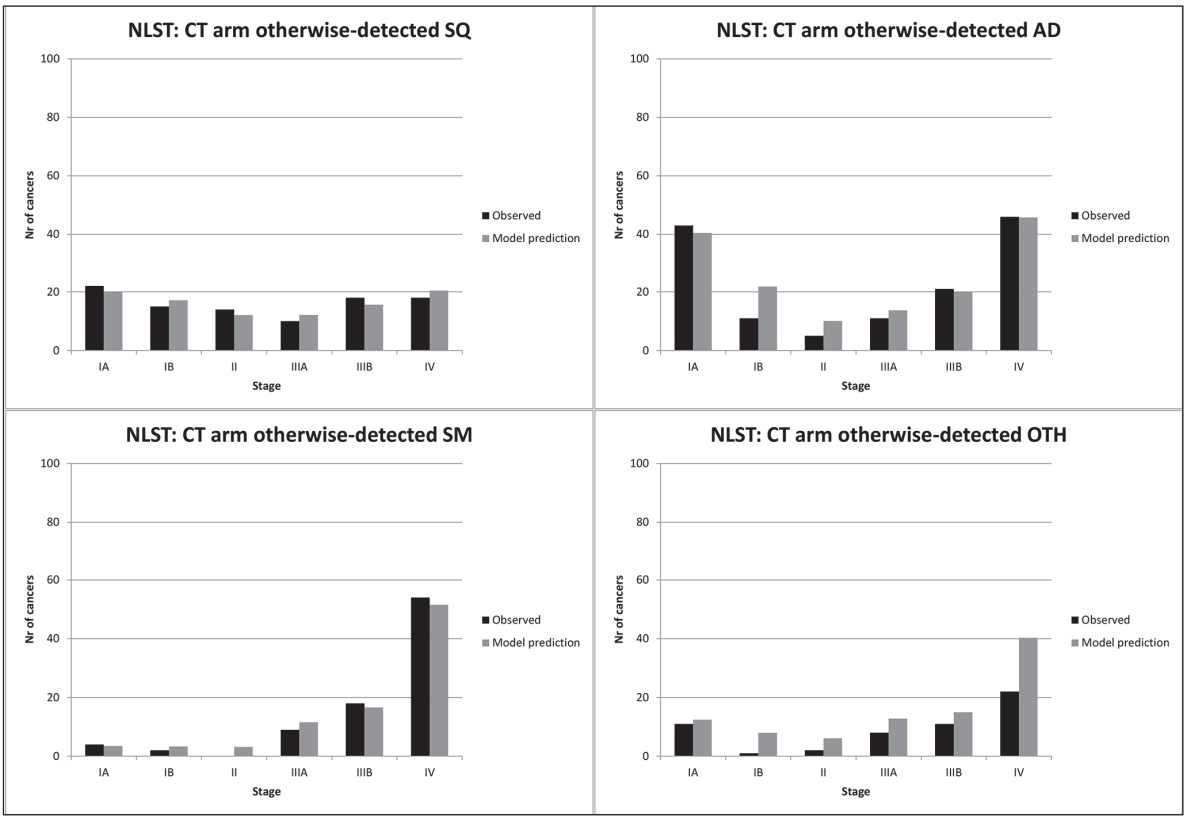
Abbreviations: AD, adenocarcinoma; SQ, squamous cell carcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Figure S1: Lung cancer progression in the MISCAN-Lung model



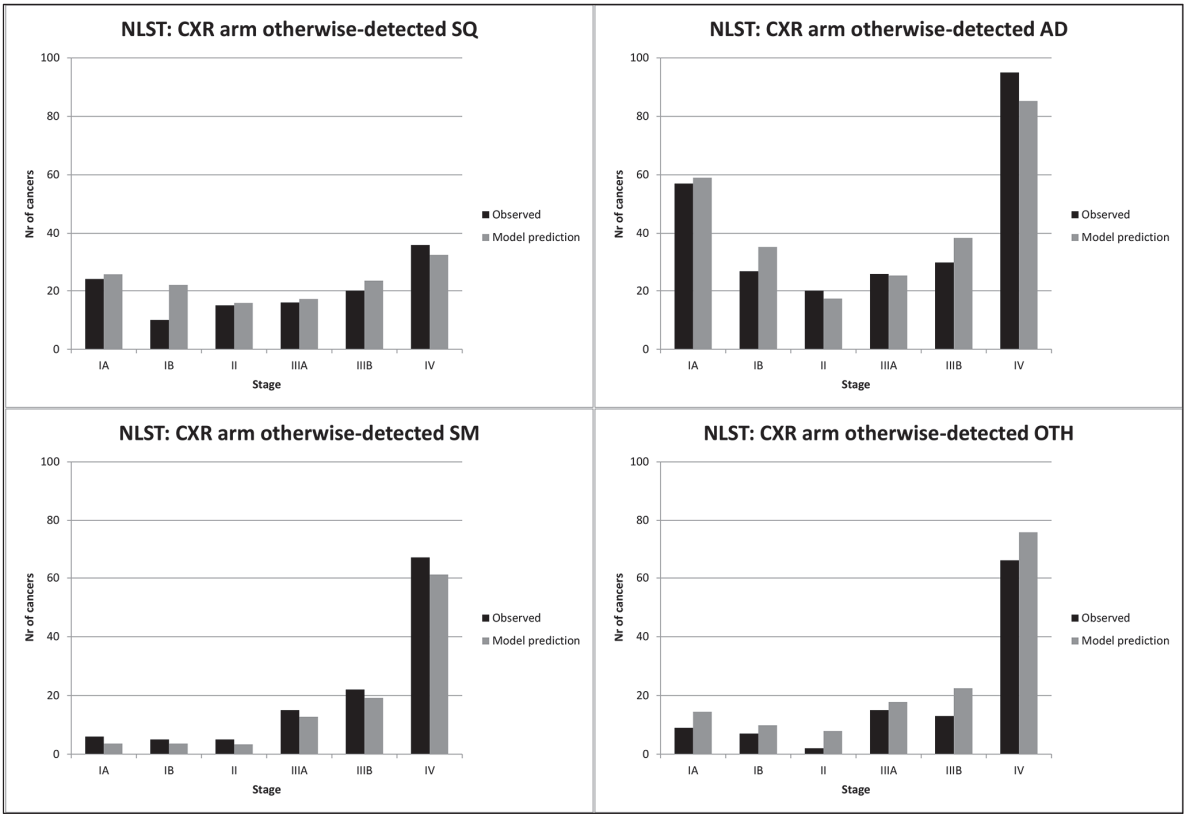
Once lung cancer has developed, it will progress from less advanced to more advanced preclinical stages until it is clinically detected. This process is similar for all histologies, however, the average time spent in the current state differs by histology, preclinical cancer stage and gender. The probability that a cancer progresses to a more advanced preclinical stage or is diagnosed clinically (e.g. diagnosed due to symptoms) is detailed by histology and stage in Table S1. Screening may detect cancers in each of the preclinical screen-detectable states, depending on the sensitivity of the screening test for the specific histology and preclinical detectable state. Upon detection of lung cancer by screening, a person's life history may be altered. Detection by screening may cure a patient, allowing him to resume his normal (lung cancer free) life history. The probability of cure differs by the stage of detection and between computed tomography and chest radiography for stages IA, IB and II. After clinical detection or screen-detection (without cure) the patient's duration of survival follows a histology and stage specific survival function, which is piecewise uniformly distributed. A person may also die from causes other than lung cancer. MISCAN-Lung incorporates the Smoking History Generator (SHG) application from the National Cancer Institute (NCI), which uses data on smoking habits in the U.S. population to provide probabilities for death from other causes depending on gender, smoking history and year of birth.[2]

Figure S2: NLST CT arm otherwise-detected lung cancers by histology and stage



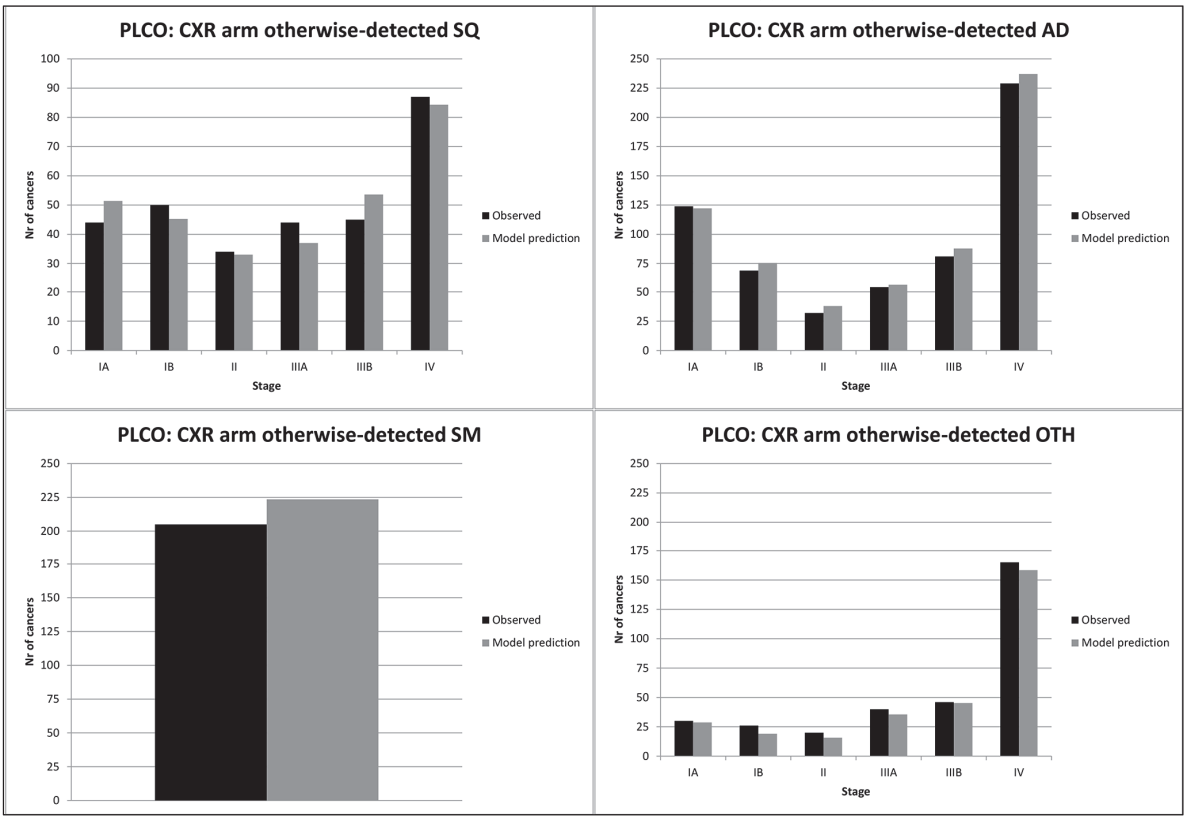
Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Figure S3: NLST CXR arm otherwise-detected lung cancers by histology and stage



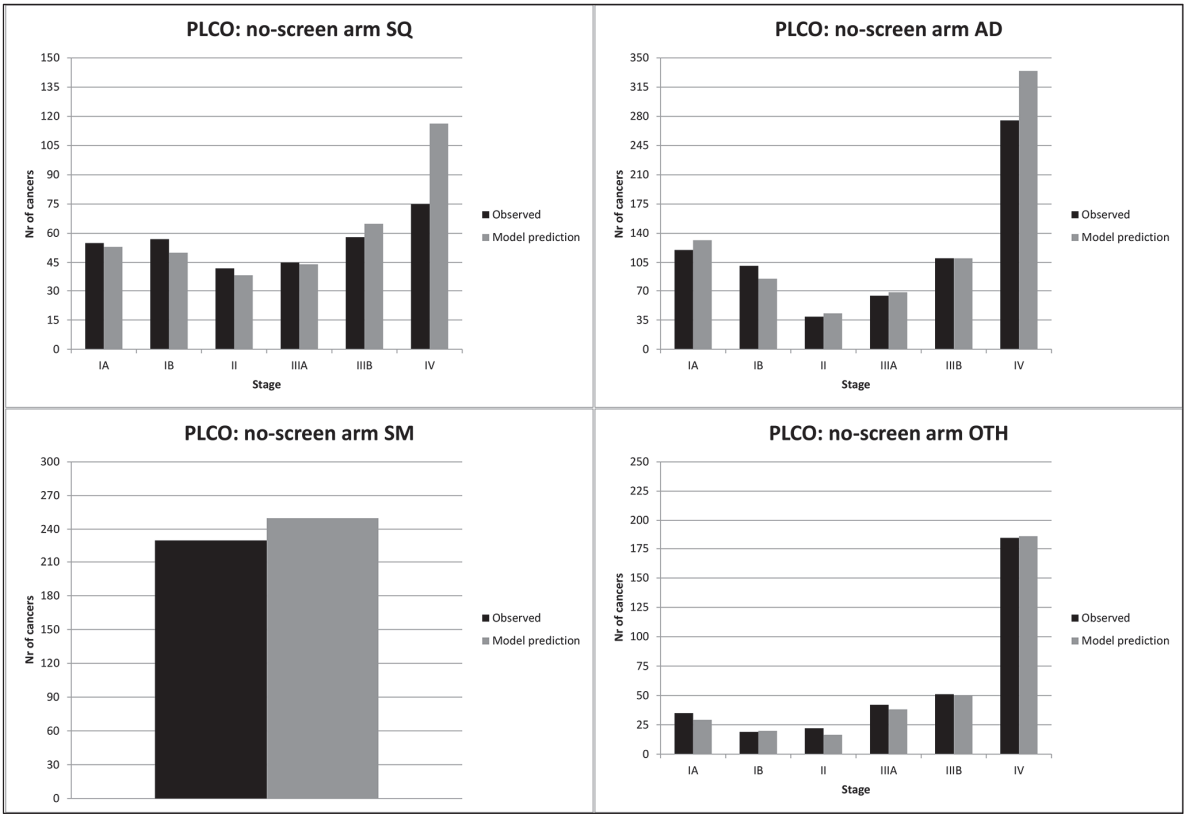
Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Figure S4: PLCO CXR arm otherwise-detected lung cancers by histology and stage



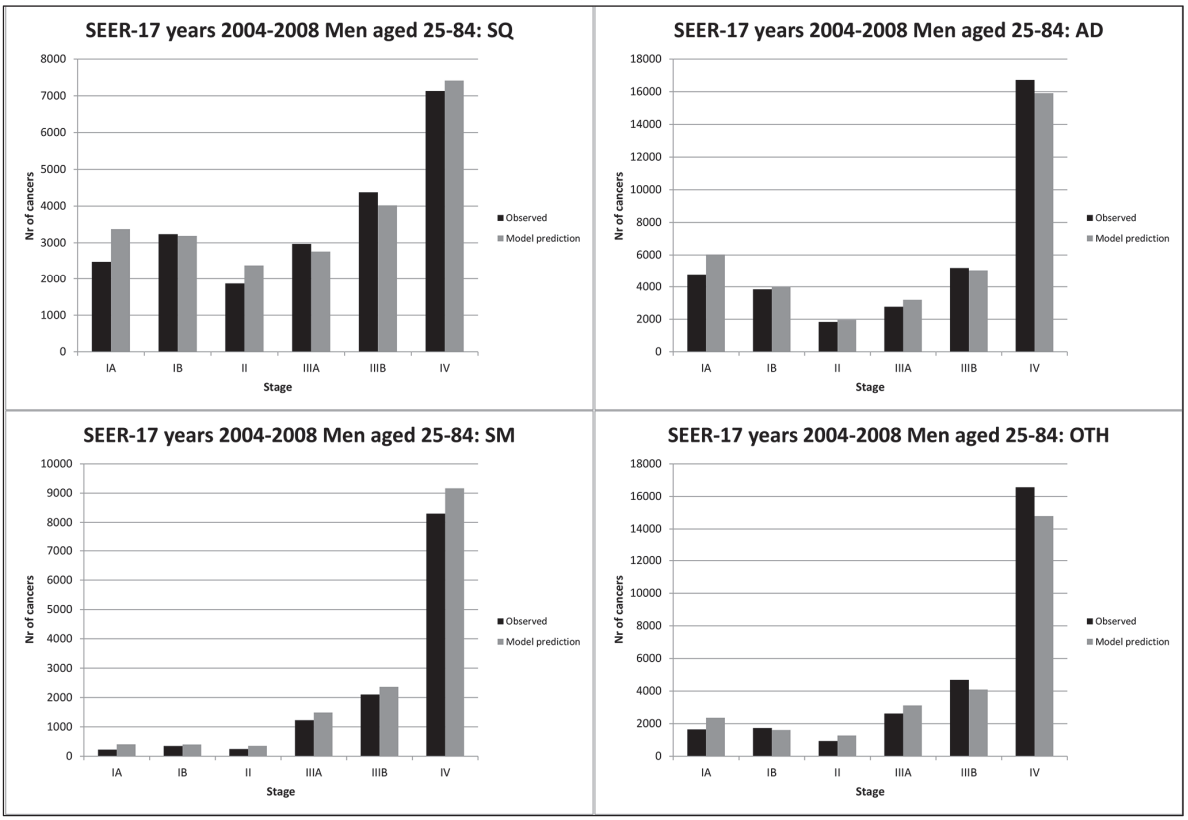
Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Figure S5: PLCO no-screen arm lung cancers by histology and stage



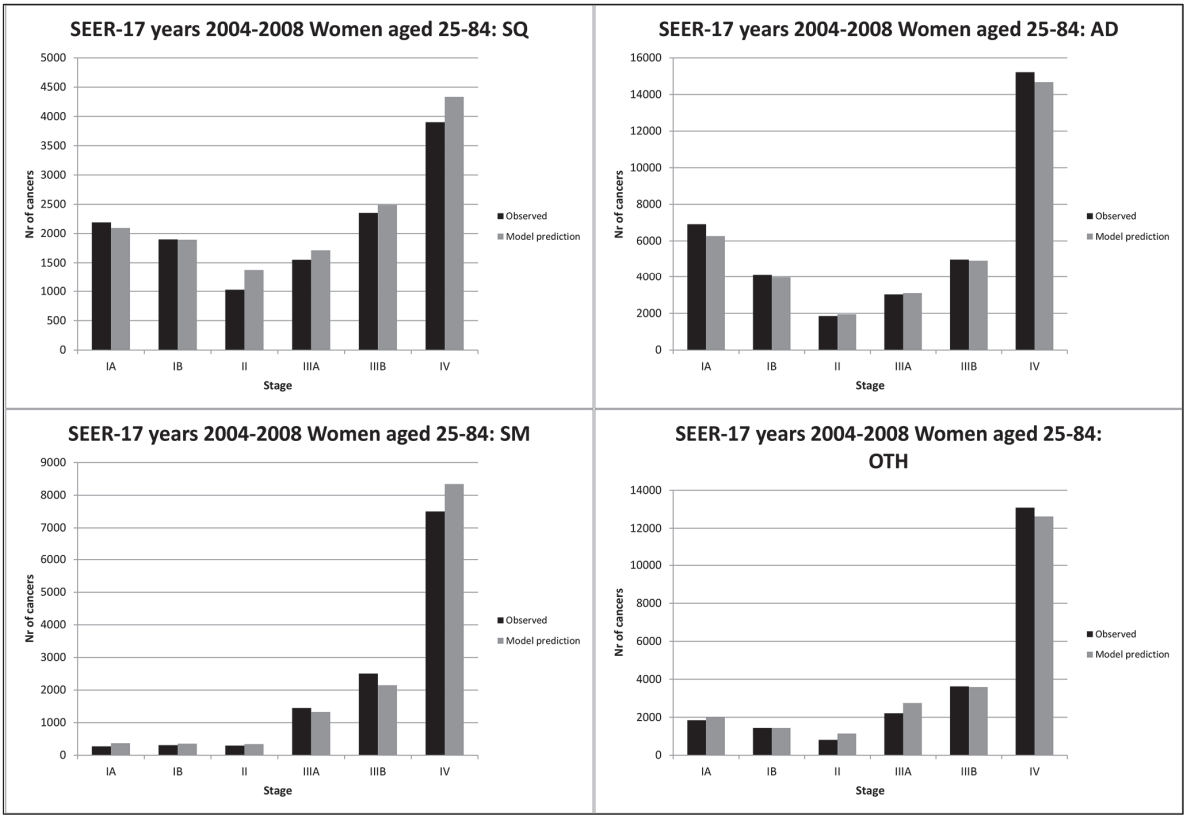
Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Figure S6: SEER-17 lung cancers in men aged 25-84 by histology and stage (years 2004-2008)



Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Figure S7: SEER-17 lung cancers in women aged 25-84 by histology and stage (years 2004-2008)



Abbreviations: SQ, squamous cell carcinoma; AD, adenocarcinoma; SM, small cell carcinoma; OTH, other non–small cell carcinoma.

Range of plausible values

Ranges of plausible values for the sensitivities of computed tomography (CT) and chest radiography (CXR) and the mean preclinical sojourn times (MPST) of the various histological types of lung cancer were calculated similar to Wever et al.[3]

To accomplish this, scaling parameters were added to each of the preclinical duration and sensitivity formulas. For example, the mean duration of stage IA AD for males:

$$Duration\_AD\_IA\_M = DurationAD*DurationIA$$

was rewritten as:

$$Duration\_AD\_IA\_M = Duration\_scale\_parameter*DurationAD*DurationIA$$

For the sensitivities a similar approach was taken, though an additive approach was taken for the scaling parameters to allow easier interpretation of the sensitivity adjustments. For example, the sensitivity of CT for stage IB AD in round two of the NLST was originally specified as:

$$\frac{\exp(SensADIB\_CT)}{1+\exp(SensADIB\_CT)}$$

Where:

$$SensADIB\_CT = Sensitivity\_AD\_IB\_CT + Sensitivity\_AD\_IA\_CT$$

And was rewritten as:

$$\max(0, \min(\frac{\exp(SensADIB\_CT)}{1+\exp(SensADIB\_CT)} + CTscalingparameter, 100))$$

A similar adjustment was made for the chest radiography sensitivity formulas, using a chest radiography specific scaling parameter. Note that sensitivities were truncated at 0% and 100%. For example, the CT sensitivity for stage IV AD was estimated to be 98.88%, so if the CT scaling parameter was set at +5%, the resulting stage IV sensitivity would be 100%.

Deviances were computed for different values of the scaling parameters (varying the value of one scaling parameter at the time), for two different random sequences (“seeds”) while holding the other model parameters constant at their maximum likelihood estimates. The deviance was defined as -2*(loglikelihood of

the model – loglikelihood of the saturated model). Figures S8, S9 and S10 show that the resulting deviances of the adjusted MPST, CXR and CT sensitivity estimates can be described as quadratic functions of the adjusted MPST and CXR and CT scaling parameters.

Quadratic functions were estimated based on these deviances, shown in Figures S8, S9 and S10. These functions were used to determine which values of the scaling parameters result in a deviance which is 3.84 (the critical value corresponding to the 95th percentile of a chi-square distribution with one degree of freedom) points higher compared to the minimum obtained through maximum likelihood estimation. These values of the scaling parameters correspond to approximate lower and upper bounds for the scaling parameters, as shown in Table S5. The estimated quadratic functions are shown below Table S5.

Table S5: Approximate bounds for the scaling parameters

Parameter	Parameter value of scaling value equivalent to Maximum likelihood estimates	Seed 1 lower bound	Seed 1 upper bound	Seed 2 lower bound	Seed 2 upper bound
Duration scale parameter	1	0.9805 (1.95% shorter durations)	1.1060 (10.60% longer durations)	0.9788 (2.12% shorter durations)	1.0959 (9.59% longer durations)
CXR sensitivity scaling parameter	0	-0.0046 (0.46% lower sensitivities)	0.0179 (1.79% higher sensitivities)	-0.0048 (0.48% lower sensitivities)	0.0174 (1.74% higher sensitivities)
CT sensitivity scaling parameter	0	-0.0232 (2.32% lower sensitivities)	0.0379 (3.79% higher sensitivities)	-0.0214 (2.14% lower sensitivities)	0.0406 (4.06% higher sensitivities)

Quadratic functions for duration scale parameters:

$$Quadratic\ Function\ Seed\ 1 = 11258 - 3881 * Sojournscale + 1860 * Sojournscale^2$$
$$Quadratic\ Function\ Seed\ 2 = 11278 - 3915 * Sojournscale + 1887 * Sojournscale^2$$

Quadratic functions for CXR sensitivity scaling parameters:

$$Quadratic\ Function\ Seed\ 1 = 9235.1 - 612.8 * CXRscale + 46238.2 * CXRscale^2$$
$$Quadratic\ Function\ Seed\ 2 = 9247.8 - 581.9 * CXRscale + 46204.7 * CXRscale^2$$

Quadratic functions for CT sensitivity scaling parameters:

$$Quadratic\ Function\ Seed\ 1 = 9235.31 - 64.34 * CTscale + 4375.81 * CTscale^2$$
$$Quadratic\ Function\ Seed\ 2 = 9247.97 - 85.35 * CTscale + 4424.04 * CTscale^2$$

Figure S8: Plausible range of values for the mean preclinical sojourn times (with quadratic function estimates)

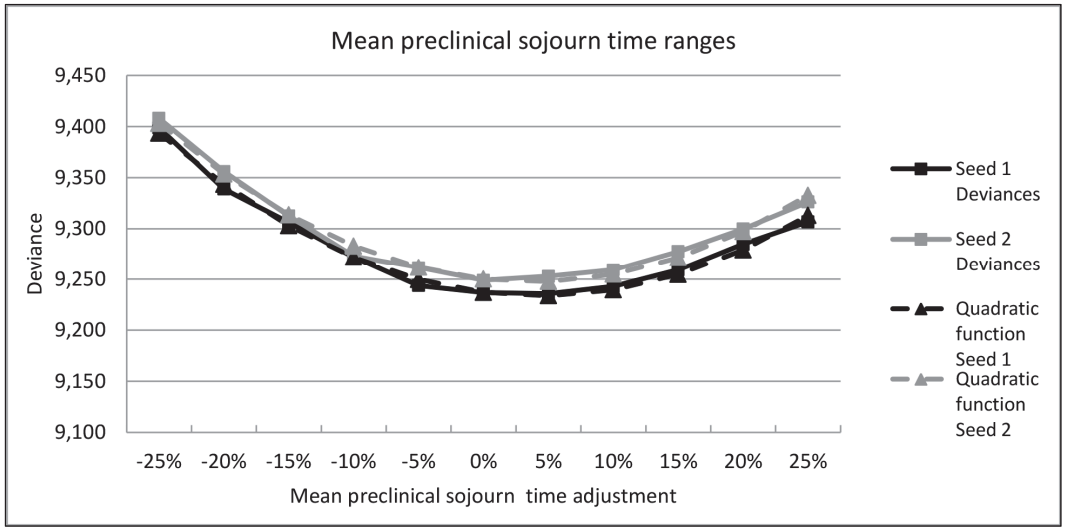


Figure S9: Plausible range of values for the chest radiography sensitivities (with quadratic function estimates)

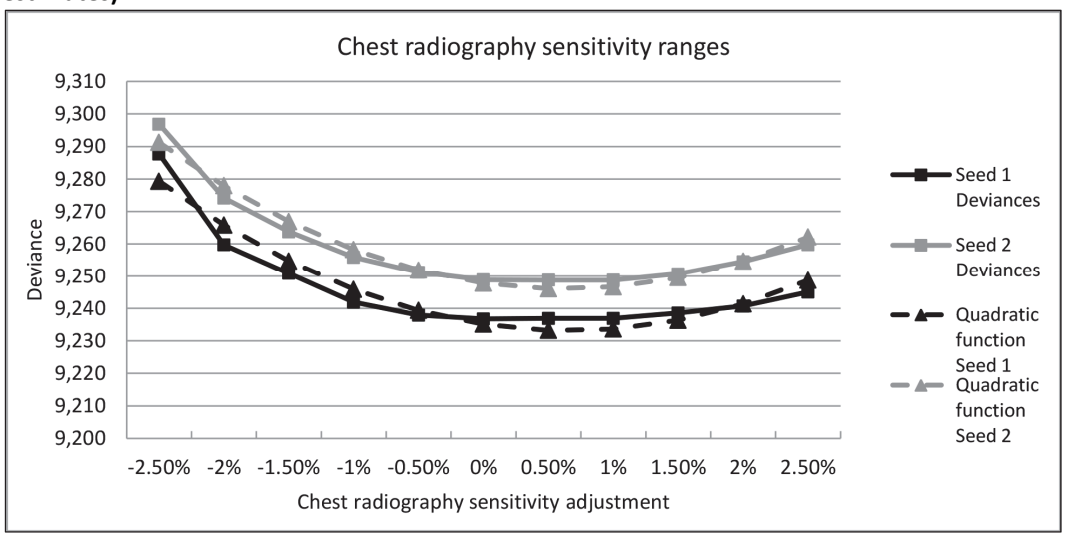
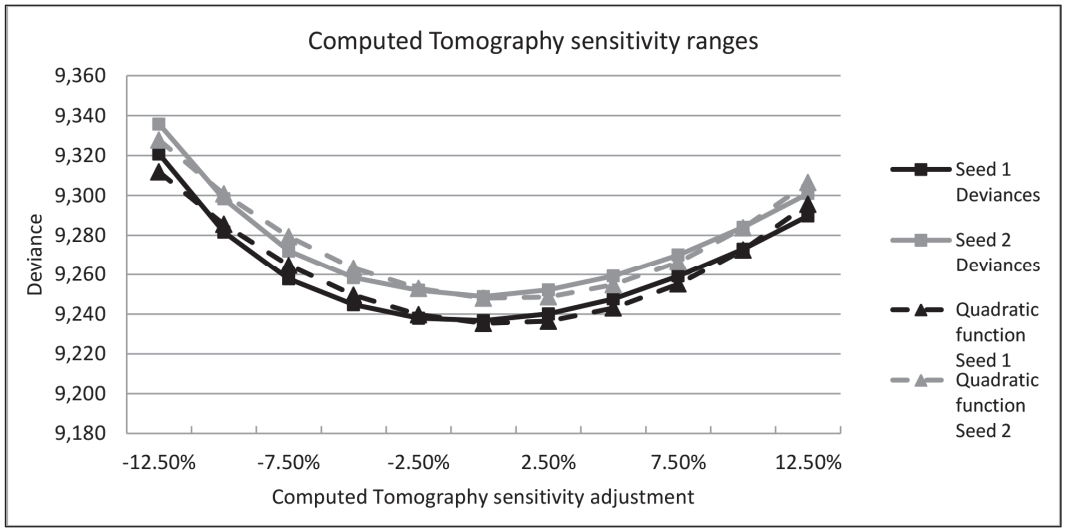


Figure S10: Plausible range of values for the computed tomography sensitivities (with quadratic function estimates)



References

1. ICD-O-3 SEER Site/Histology Validation List - 12/04/2009 - updated from 2/9/2001. Available from URL: http://seer.cancer.gov/icd-o-3/ .

2. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Analysis* 2012; **32 Suppl 1**: S51-68.

3. Wever EM, Draisma G, Heijnsdijk EAM, et al. Prostate-Specific Antigen Screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer–Rotterdam. *Journal of the National Cancer Institute* 2010; **102**(5): 352-5.

Chapter 3

Overdiagnosis in lung cancer screening: why modeling is essential

Kevin ten Haaf, Harry J. de Koning

Published as:

ten Haaf K, de Koning HJ.

Journal of Epidemiology and Community Health 2015; **69**(11): 1035-39.

Introduction

Screening for diseases is an important part of healthcare, as the detection of disease at an early stage may improve the chance of successful treatment of the disease. Although high-quality screening programs may provide substantial benefits, one of the major harms associated with screening is overdiagnosis. Overdiagnosis refers to the event that a disease is diagnosed that would not have been clinically detected; for example, screening may detect a slowly progressing tumor that would not have caused symptoms during the patient's lifetime. Overdiagnosis may lead to serious consequences, such as unnecessary treatments. Furthermore, overdiagnosis will lead to biased survival outcomes for screen-detected cases.[1]

Notable harms of CT lung cancer screening include the high proportion of false-positive results and radiation exposure.[2] The decision of the United States Preventive Services Task Force (USPSTF) to recommend lung cancer screening with CT for persons up to age 80 has also raised concerns on the magnitude of overdiagnosis.[2] The Medicare Evidence Development and Coverage Advisory Committee commented that the USPSTF's decision to extend the upper age of screening from 74 to 80 years '*was based upon modeling only, with no empirical data*'.[3]

The opinion that modeling cannot provide additional information beyond that of a clinical trial is a common misconception. Models are often criticized for depending on assumptions with regard to the processes of carcinogenesis and progression of cancer.[4] However, the same could be said of statistical tests used to evaluate the results of clinical trials, many of which are also based on implicit assumptions. Similarly to other statistical methods, if the underlying assumptions are properly stated, models can be valuable tools in bridging the gaps between the evidence provided by clinical trials and the evidence needed for the development of clinical guidelines.[5] In fact, modeling is essential to derive detailed estimates of the benefits and harms of screening beyond the time period of the clinical trial, especially for overdiagnosis.

Analyses using an excess incidence approach by Patz suggest that over 18% of all lung cancers detected by CT in the National Lung Screening Trial (NLST) were overdiagnosed.[6] However, five microsimulation models, which were calibrated to various data sources including the NLST, suggest that less than 10% of lung cancers detected by screening would be overdiagnosed in a 1950 U.S. birth cohort adhering to the annual screening policy recommended by the USPSTF.[7] Previous reports have discussed different methods to derive overdiagnosis.[8,9] In this report, we will address the difficulties of deriving estimates on overdiagnosis from clinical trials alone. Furthermore, we will detail how models can aid in extrapolating information from clinical trials to screening programs.

Potential biases in estimating overdiagnosis in clinical trials

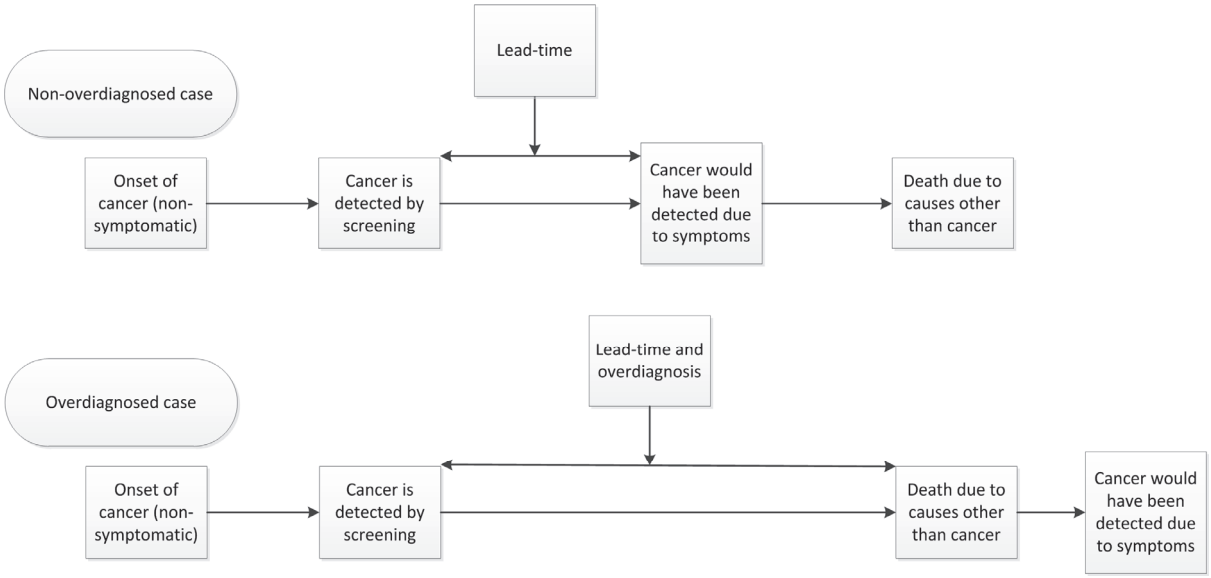
The NLST randomized 53,454 current smokers and former smokers (quit less than 15 years) between the ages of 55–74 with a minimum smoking exposure of 30 pack-years to receive either three CT or chest radiography (CXR) screens.[10] The NLST demonstrated a relative reduction in lung cancer mortality of 20% for CT compared to CXR (16% after extending the cutoff date for mortality analyses).[10,11] However, the magnitude of overdiagnosis in the NLST cannot be easily ascertained from the results of the trial alone. Assuming the excess number of lung cancers in the CT arm compared to the CXR arm are due to overdiagnosis (excess incidence approach) to derive the magnitude of overdiagnosis in the NLST, as suggested by Patz, is biased for two reasons.[6]

First, both arms of the NLST were screened, either with CT or CXR. While the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) indicated that CXR screening does not reduce lung cancer mortality, the possibility of overdiagnosis in the CXR arm of the NLST is still present.[12] Therefore, the CXR arm of the NLST does not provide an unbiased baseline incidence of lung cancer in the absence of screening.[8] Thus, ascertaining the magnitude of overdiagnosis in the CT arm of the NLST through an excess incidence analysis may underestimate the level of overdiagnosis.

Second, the median follow-up duration of NLST participants was limited to approximately 6.5 years.[10] A follow-up duration of sufficient length is essential to account for the effects of

lead-time.[9] Lead-time is the time interval between the detection of the cancer by screening and its clinical presentation (the time of diagnosis in the absence of screening), shown in Figure 1.

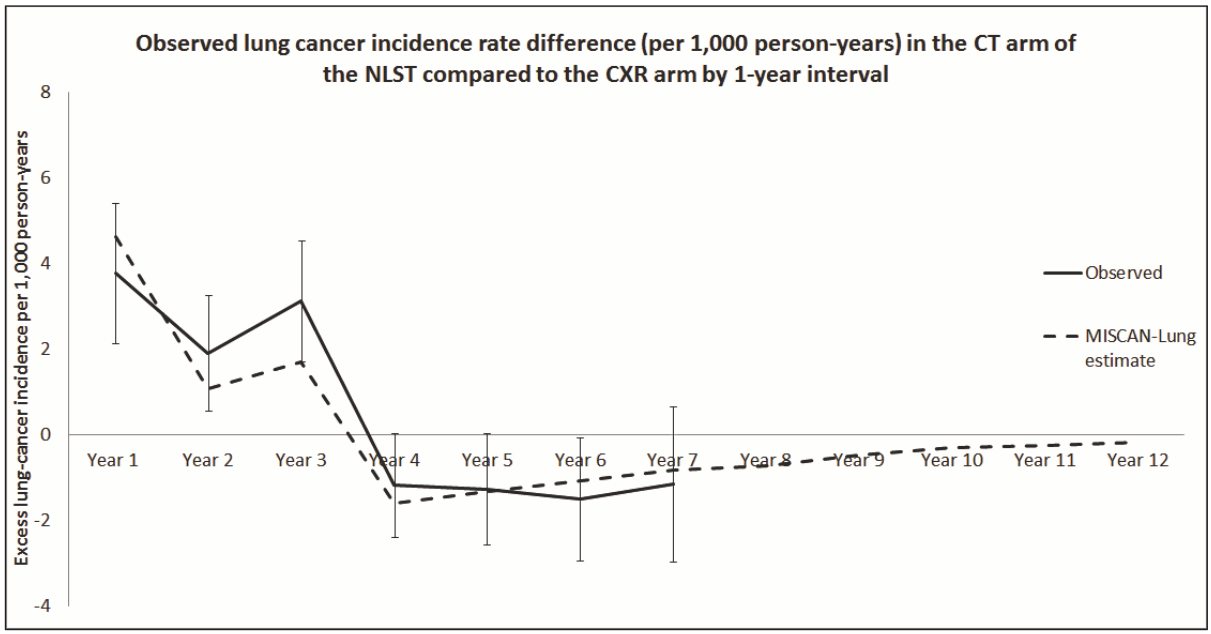
Figure 1: Lead-time and overdiagnosis



The two halves of the figure depict two scenarios: in the scenario depicted in the upper half of the figure, the person has developed a cancer that would have become symptomatic before the person would have died due to causes other than cancer. However, due to screening, this cancer is detected earlier. In the scenario depicted in the lower half of the figure, the person has developed a cancer that would not have become symptomatic before the person would have died due to causes other than cancer. However, due to screening, this cancer is detected, while it would not have been diagnosed without screening. The lead-time in both scenarios represents the time between the detection of the cancer by screening and the moment that the cancer would have become symptomatic.

A higher proportion of cancers were detected at an early stage in the CT arm compared to the CXR arm, due to the higher sensitivity of CT.[10,13] This suggests that the lead-time of CT is longer than that of CXR. Figure 2 (derived from Black et al, table S2–2) shows the effects of the difference in lead-time between CT and CXR in the NLST.[14]

Figure 2: Observed and estimated lung cancer incidence rate difference (per 1,000 person-years) in the CT arm of the National Lung Screening Trial (NLST) compared to the chest radiography (CXR) arm by 1-year interval



Observed data derived from table S2–2 of Black et al.[14] Error bars denote 95% CIs for the incidence rate difference. The Microsimulation SCreening ANalysis (MISCAN) Lung model was used to estimate the excess incidence of the CT arm compared to the CXR arm of the NLST and extrapolate beyond the follow-up duration of the trial.

In the years during which screening occurs (years 1–3), the incidence of lung cancer in the CT arm is higher compared to the CXR arm; due to the longer lead-time of CT. Non-overdiagnosed lung cancers would have become symptomatic at a later date, but screening advances their detection to an earlier moment in time. Thus, lung cancers of which the moment of detection was advanced will not be detected at their original moment of clinical presentation. As the moment of detection was advanced for more lung cancers in the CT arm, the relative number of lung cancers detected in the CXR arm is higher in the years after screening has ended (years 4–7), as shown in Figure 2. As the base incidence of lung cancer in the absence of screening is similar for both arms, the difference in lung cancer incidence between the arms, due to a difference in lead-time, will dissipate over time. The remaining excess of lung cancers will then represent the number of overdiagnosed lung cancers in the CT arm compared to the CXR arm.

Patz suggests that the difference in the absolute number of lung cancers converges in the last years of the trial.[6] However, this may be due to the low number of person-years observed in the last years of the trial, as noted in Table S2–2 of Black et al.[14] Black et al corrects for the lower number of person-years by investigating the difference in lung cancer incidence by a 0.5-year interval and also suggests a convergence Table S2–2 and Figure S2–1 in their report).[6,14] However, when one observes the 95% confidence intervals (CIs) of the incidence rate difference between the CT and CXR arms of the NLST by a 1-year interval instead, shown in Figure 2, the incidence rates are not suggested to have converged yet. Therefore, the follow-up duration of the NLST does not seem of sufficient length to account for the difference in lead-time between CT and CXR.

Limitations of extrapolating overdiagnosis estimates from clinical trials to screening programs

The design and limited follow-up duration complicate deriving the magnitude of overdiagnosis in the NLST. However, even if this information could be easily derived, it would provide little information on the magnitude of overdiagnosis in a screening program on the population level, due to a number of limitations.

The first limitation is the trial's fixed design. If the NLST had considered a different number of screening rounds or different intervals between screening rounds, the number of lung cancers detected by screening would have been different. Consequently, the proportion of overdiagnosed cases could have been different as well, which poses difficulties in extrapolating the results of the NLST to other designs (e.g., different number of screening rounds and/or different intervals between screenings).

The second limitation is the investigated population. Recent studies indicate that the risk of lung cancer (death) varies, even across participants of the NLST.[15,16] In addition, mortality for causes other than lung cancer varies greatly by smoking behavior and age.[17] Compared to the portion of the general U.S. population that meets the entry criteria of the NLST, the participants of the NLST were younger (26.6% of the NLST participants were older than 65 compared to 35.5% of the eligible U.S. population) and less likely to be current smokers

(48.2% of NLST participants compared to 57.1% of the eligible U.S. population).[18] Consequently, the average person eligible for a U.S. screening program utilising the same entry criteria as the NLST may be younger and more likely to be a current smoker compared to the average participant of the NLST. As a result, the average risk of lung cancer (death) and mortality for causes other than lung cancer may differ between the two groups.[15-17] Therefore, the magnitude of overdiagnosis in a population-based screening program could differ from that in the NLST.

Thus, even if information on the magnitude of overdiagnosis in the NLST can be ascertained, this information cannot be easily applied to different designs and populations. While the probability of overdiagnosis is constantly present, the rate of overdiagnosis is subject to different aspects of the screening program and the screened population, such as screening frequency and age. However, microsimulation models can aid in extrapolating the information obtained from the NLST to different designs and populations.

Estimating overdiagnosis through microsimulation modeling

Microsimulation models can simulate a person's entire life-history in the absence and presence of screening, which allows one to determine which cancers are overdiagnosed within the simulation. Each individual's probability of developing cancer due to biological processes and/or exposure to carcinogens is modeled, as well as the individual's probability of dying from the disease or other causes. However, to estimate the magnitude of overdiagnosis in lung cancer screening for different designs and populations, a model must meet a number of requirements.

First, the model must be able to provide an estimate of the baseline incidence of the disease for the population of interest in the absence of screening.[8] In modeling lung cancer, a smoking dose–response module, such as the Two-Stage Clonal Expansion model, is often used to determine the baseline incidence.[19] Smoking dose–response modules use age, smoking history and other risk-factors to estimate a person's risk of lung carcinogenesis.[19,20] When a model incorporates a smoking dose–response module calibrated to a wide range of risk profiles, including never-smokers, the model can be applied to any population.[19,20] The

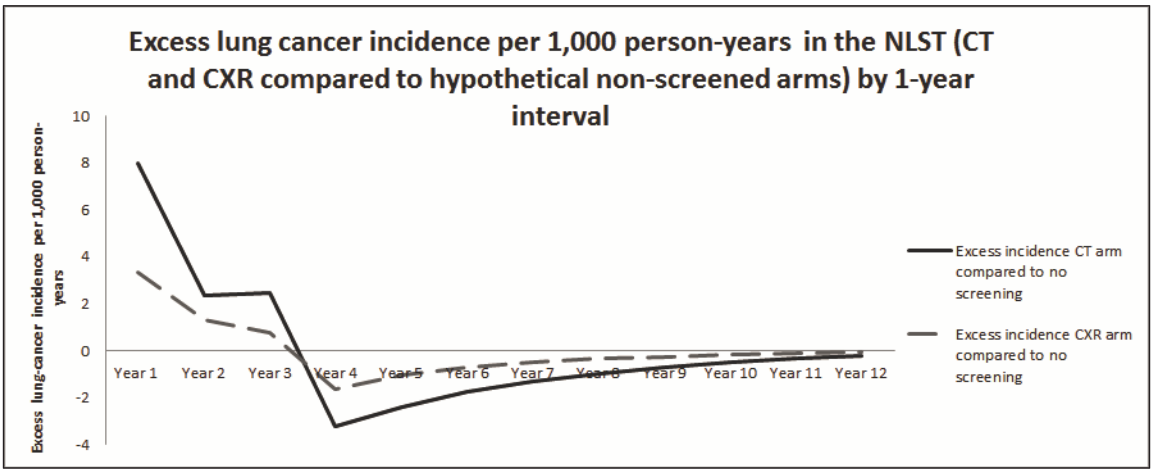
MIcrosimulation Screening ANalysis (MISCAN) Lung model demonstrates this by reproducing the incidence of lung cancer in populations with different risk profiles, such as the NLST and PLCO.[13,21] The assumptions, calibration process and sensitivity analyses of the MISCAN-Lung model used in this investigation were detailed previously.[13,21]

Furthermore, the model must explicitly consider the natural history (preclinical progression) and screen-detectability of lung cancer. This is essential to estimate the lead-time achieved by the screening test and the potential for overdiagnosis.[8] The model should take the statistical dependence between the natural history of the disease and the sensitivity of the screening test into account, which may cause uncertainty in the parameter estimates: a model with a low estimate for the sensitivity and a long preclinical duration may provide a similar fit compared to a model with a high estimate for the sensitivity and a short preclinical duration.[22] Microsimulation models can derive this information through synthesizing data from the NLST with data from clinical trials with non-screened control arms, such as the PLCO, which provides essential information on the natural history of lung cancer in the absence of screening.[12,13,21] By calibrating to the number of cancers detected by screening round and the number of interval cancers per year, models can estimate the natural history and screen-detectability of lung cancer.[13] This information allows the extrapolation of the findings of clinical trials to screening policies with different designs, for example, variations in number of screens and intervals between screens. In a previous investigation, we derived estimates for the natural history and screen-detectability of lung cancer by histological type, stage and gender.[13] Overall, our estimates suggested a greater window of opportunity for lung cancer screening compared to previous research.[13]

Figure 2 indicates that the incidence rate difference of the CT arm compared to the CXR arm estimated by MISCAN-Lung lies between the 95% CIs for the entirety of the observed follow-up duration of the trial. However, in contrast to other reports, the estimates of MISCAN-Lung suggest that the incidence rate difference between the two arms does not converge at the end of the observed follow-up duration of the trial (year 7).[6,14] Instead, the incidence rate difference between the two arms is suggested to converge in year 12, 9 years after screening has ended and 5 years after the observed follow-up duration of the trial. Through

modeling, one can investigate the excess incidence in both arms of the NLST compared to hypothetical non-screened arms, as shown in Figure 3.

Figure 3: Estimated lung cancer incidence rate difference (per 1,000 person-years) of the CT and chest radiography (CXR) arms of the National Lung Screening Trial (NLST) compared to hypothetical non-screened arms by 1-year interval



The MIcrosimulation SCreening ANalysis (MISCAN) Lung model was used to estimate the excess incidence of the CT and CXR arms of the NLST compared to hypothetical arms in which screening did not occur.

As expected, in the years in which screening occurs (years 1–3) the excess incidence is higher for both arms, compared to their hypothetical non-screened counterparts. The estimated excess incidence of the CT arm is higher than the CXR arm, indicating that CT is more sensitive than CXR: MISCAN-Lung estimates the excess incidence in the CT arm compared to the CXR arm, estimated as the excess number of lung cancers in the CT arm compared to the CXR arm divided by the number of screen-detected cancers in the CT arm, at year 8 to be 12.5% compared to the reported 18.5% (95% confidence interval: 5.4-30.6%).[6]

The estimated difference in yearly lung cancer incidence between the CXR arm and the hypothetical non-screened arm dissipates at approximately year 10, 7 years after screening has ended. For the CT arm, this occurs at approximately year 12, 9 years after screening has ended, which suggests a longer lead-time for CT compared to CXR. This information allows the model to estimate the amount of overdiagnosis, defined as the number of cancers that would not have been detected if screening had not occurred, divided by the number of

cancers detected by screening that occurred in both arms of the NLST. MISCAN-Lung estimates that 6.75% of all screen-detected cases in the CXR arm are overdiagnosed compared to 8.62% of all screen-detected cases in the CT arm. These percentages are relatively low, as approximately 75% of the participants of the NLST were aged younger than 65.[10] Furthermore, the participants of the NLST only received three screens, and over 50% of the detected cancers in the CT-arm were detected in persons aged under 65.[23] Therefore, the potential for overdiagnosis was relatively low in the NLST. The final requirement to extrapolate the results of clinical trials to different populations is that the model must be able to utilize data specific to the investigated population. Such information includes smoking behavior and mortality for causes other than lung cancer by smoking behavior, gender and age. For example, information on these aspects for a 1950 U.S. birth cohort has been used in modeling analyses to inform the USPSTF on its recommendation for lung cancer screening.[7]

Discussion

Clinical trials such as the NLST are essential to provide information on the efficacy of lung cancer screening.[10] However, information on overdiagnosis is difficult to ascertain from clinical trials alone, as the follow-up duration may be insufficient and, in trials without an unscreened control group, an unbiased baseline incidence of the disease in the absence of screening may not be available. Furthermore, an estimate for overdiagnosis based on a clinical trial only provides information with regard to the design and population investigated in that trial. Furthermore, the risk of lung cancer and (smoking-related) mortality from causes other than lung cancer can vary substantially across individuals, which must be taken into account when one considers implementing a lung cancer screening program.[15-17] This report shows that, while important, the results of the NLST by itself provide limited information on the magnitude of overdiagnosis in future lung cancer screening programs. Microsimulation models can provide estimates on overdiagnosis for lung cancer screening programs with designs and populations different from those considered in the NLST, which is essential for policy makers.[7,13,21 24-26] Furthermore, modeling can provide insights in situations where information from trials or observational studies alone is insufficient. For example, in colorectal screening, a colonoscopy can remove precancerous lesions

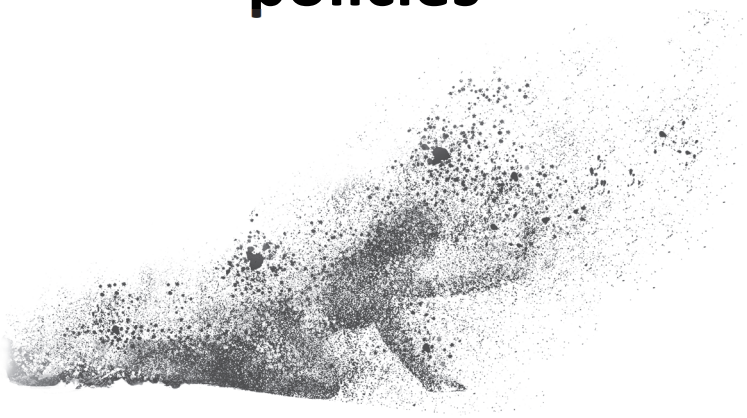
(adenomas), preventing the occurrence of colorectal cancer altogether. As a result, if one compares a screened group with a non-screened control group, the control group may have a higher incidence compared to the screened group. While overdiagnosis will still occur in the screened group, the magnitude cannot be ascertained with an excess incidence approach. However, sophisticated microsimulation models require detailed data, which can often only be obtained from clinical trials such as the NLST and PLCO, to provide accurate estimates.[13,21] Like any analysis, the assumptions and validity of the model should be clearly detailed.[4] Furthermore, one should take into account that models have other limitations, such as uncertainty in incidence trends and drifts in screening efficacy.[27] As lung cancer screening is implemented across the U.S., more data will become available on the effects of lung cancer screening in the general population, such as the effect of nodule cut points. Besides influencing the sensitivity of CT screening, it may also affect overdiagnosis. This information can be used to further improve the estimates on the magnitude of overdiagnosis in lung cancer screening.

References

1. Ruano-Ravina A, Heleno B, Fernández-Villar A. Lung cancer screening with low-dose CT (LDCT), or when a public health intervention is beyond the patient's benefit. *Journal of Epidemiology and Community Health* 2015; **69**(2): 99.
2. Moyer VA, on behalf of the USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 2014; **160**(5): 330-8.
3. Centers for Medicare & Medicaid Services. Proposed Decision Memo for Screening for Lung Cancer with Low Dose Computed Tomography (LDCT) (CAG-00439N). 2014. http://www.cms.gov/medicare-coverage-database/details/nca-proposed-decision-memo.aspx?NCAId=274 (accessed 20 Nov 2014).
4. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *British Medical Journal* 2015; **350**.
5. Habbema JF, Wilt TJ, Etzioni R, et al. Models in the development of clinical practice guidelines. *Annals of Internal Medicine* 2014; **161**(11): 812-8.
6. Patz EF, Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Internal Medicine* 2014; **174**(2): 269-74.
7. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
8. Etzioni R, Gulati R, Mallinger L, Mandelblatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Annals of Internal Medicine* 2013; **158**(11): 831-8.
9. de Gelder R, Heijnsdijk EAM, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting Overdiagnosis Estimates in Population-based Mammography Screening. *Epidemiologic Reviews* 2011; **33**(1): 111-21.
10. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
11. Pinsky PF, Church TR, Izmirlian G, Kramer BS. The National Lung Screening Trial: Results stratified by demographics, smoking history, and lung cancer histology. *Cancer* 2013; **119**(22): 3976-83.
12. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
13. ten Haaf K, van Rosmalen J, de Koning HJ. Lung Cancer Detectability by Test, Histology, Stage, and Gender: Estimates from the NLST and the PLCO Trials. *Cancer Epidemiology Biomarkers & Prevention* 2015; **24**(1): 154-61.
14. Black WC, Gareen IF, Soneji SS, et al. Cost-Effectiveness of CT Screening in the National Lung Screening Trial. *New England Journal of Medicine* 2014; **371**(19): 1793-802.
15. Kovalchik SA, Tammemägi M, Berg CD, et al. Targeting of Low-Dose CT Screening According to the Risk of Lung-Cancer Death. *New England Journal of Medicine* 2013; **369**(3): 245-54.
16. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.
17. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.
18. Aberle DR, Adams AM, Berg CD, et al. Baseline Characteristics of Participants in the Randomized National Lung Screening Trial. *Journal of the National Cancer Institute* 2010; **102**(23): 1771-9.
19. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.

20. Hazelton WD, Luebeck EG, Heidenreich WF, Moolgavkar SH. Analysis of a Historical Cohort of Chinese Tin Miners with Arsenic, Radon, Cigarette Smoke, and Pipe Smoke Exposures Using the Biologically Based Two-Stage Clonal Expansion Model. *Radiation Research* 2001; **156**(1): 78-94.
21. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
22. van Ballegooijen M, Rutter CM, Knudsen AB, et al. Clarifying Differences in Natural History between Models of Screening. *Medical Decision Making* 2011; **31**(4): 540-9.
23. Pinsky PF, Gierada DS, Hocking W, Patz EF, Jr, Kramer BS. National lung screening trial findings by age: Medicare-eligible versus under-65 population. *Annals of Internal Medicine* 2014; **161**(9): 627-33.
24. McMahon PM, Meza R, Plevritis SK, et al. Comparing Benefits from Many Possible Computed Tomography Lung Cancer Screening Programs: Extrapolating from the National Lung Screening Trial Using Comparative Modeling. *PLOS ONE* 2014; **9**(6): e99978.
25. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Analysis* 2012; **32 Suppl 1**: S51-68.
26. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort–Specific Smoking Histories, 1965–2009. *American Journal of Preventive Medicine* 2014; **46**(2): e31-e7.
27. Kramer BS, Elmore JG. Projecting the Benefits and Harms of Mammography Using Statistical Models: Proof or Proofiness? *Journal of the National Cancer Institute* 2015; **107**(7): djv145-djv.

Part 2: The long-term benefits and harms of lung cancer screening policies

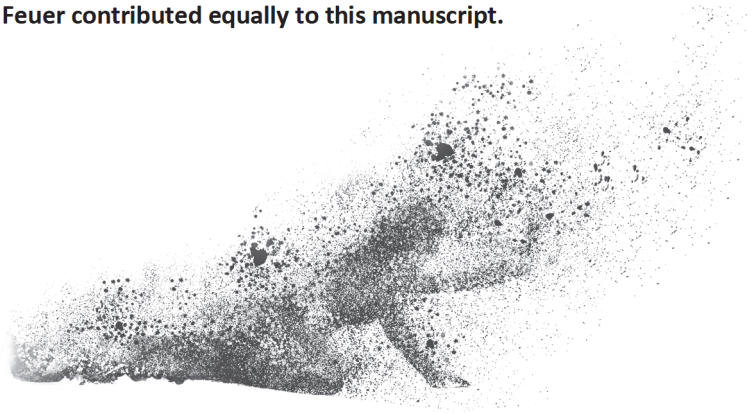


Chapter 4

Comparing benefits from many possible computed tomography lung cancer screening programs: extrapolating from the National Lung Screening Trial using comparative modeling

Pamela M. McMahon, Rafael Meza, Sylvia K. Plevritis, William C. Black, C. Martin Tammemägi, Ayca Erdogan, Kevin ten Haaf, William Hazelton, Theodore R. Holford, Jihyoun Jeon, Lauren Clarke, Chung Yin Kong, Sung Eun Choi, Vidit N. Munshi, Summer S. Han, Joost van Rosmalen, Paul F. Pinsky, Suresh Moolgavkar, Harry J. de Koning*, Eric J. Feuer*

*** Drs. de Koning and Feuer contributed equally to this manuscript.**



Published as:

McMahon PM, Meza R, Plevritis SK, et al.

PLOS ONE 2014; **9**(6): e99978.

Abstract

Background: The National Lung Screening Trial (NLST) demonstrated that in current and former smokers aged 55 to 74 years, with at least 30 pack-years of cigarette smoking history and who had quit smoking no more than 15 years ago, 3 annual computed tomography (CT) screens reduced lung cancer-specific mortality by 20% relative to 3 annual chest X-ray screens. We compared the benefits achievable with 576 lung cancer screening programs that varied CT screen number and frequency, ages of screening, and eligibility based on smoking.

Methods and Findings: We used five independent microsimulation models with lung cancer natural history parameters previously calibrated to the NLST to simulate life histories of the U.S. cohort born in 1950 under all 576 programs. 'Efficient' (within model) programs prevented the greatest number of lung cancer deaths, compared to no screening, for a given number of CT screens. Among 120 'consensus efficient' (identified as efficient across models) programs, the average starting age was 55 years, the stopping age was 80 or 85 years, the average minimum pack-years was 27, and the maximum years since quitting was 20. Among consensus efficient programs, 11% to 40% of the cohort was screened, and 153 to 846 lung cancer deaths were averted per 100,000 people. In all models, annual screening based on age and smoking eligibility in NLST was not efficient; continuing screening to age 80 or 85 years was more efficient.

Conclusions: Consensus results from five models identified a set of efficient screening programs that include annual CT lung cancer screening using criteria like NLST eligibility but extended to older ages. Guidelines for screening should also consider harms of screening and individual patient characteristics.

Introduction

In the National Lung Screening Trial (NLST), participants aged 55–74 years randomized to three annual CT examinations experienced a 20% reduction in lung cancer mortality at 6.5 years of follow up (16% at 7.5 years), compared to participants randomized to receive three annual chest radiographs.[1,2] The NLST was designed to determine the efficacy of CT screening, but the eligibility criteria and the number of screens offered were not meant to represent a population screening strategy. Multiple clinical guidelines, however, recommend lung cancer screening for individuals meeting the NLST eligibility criteria.[3,4] Other guidelines expanded recommendations for screening to individuals who would have been ineligible for the NLST.[5-7]

The NLST provided no direct evidence of further reductions in lung cancer mortality from additional screens, or of potential benefits of screening individuals with lighter smoking histories (fewer than 30 pack-years of cigarette smoking or former smokers who had quit more than 15 years prior) or individuals younger than 55 or older than 74 years at the beginning of screening. We extrapolated the findings of the NLST and compared various screening programs if adopted in the U.S. population. Five modeling groups used independent approaches to combine multiple sources of data to simulate the underlying natural history of lung cancer and to estimate the benefit of alternative screening programs.

In a single cohort of people born in 1950, each model estimated the benefits from 576 screening programs that varied eligibility criteria and frequency of screens, and two reference scenarios. We sought to rank programs according to a measure of efficiency, to reduce the number of programs that would require closer evaluation. The 1950 birth cohort was selected because they reach age 63 (about mid-range of participants in the NLST) in 2013. When independent models reach consensus on the characteristics of efficient screening programs, as reported here, the results can better inform screening guidelines. As in prior comparative modeling studies of important public health questions independent modeling groups collaborated, sharing inputs and standardizing analyses to remove uncertainty due to incongruent modeled populations, endpoints and metrics.[8,9]

Methods

Models

The microsimulation models used were developed independently by investigators at five institutions funded by the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET, www.cisnet.cancer.gov) consortium through a peer-reviewed, cooperative award (2010–2015) from the National Institutes of Health: Erasmus MC in the Netherlands (Model E), Fred Hutchinson Cancer Research Center (Model F),Massachusetts General Hospital (MGH) (Model M), Stanford University (Model S) and the University of Michigan (Model U). Additional investigators (referred to in the Acknowledgments section of this Chapter) collaborated to develop common inputs and standardize analyses. The analyses and results described in this report were part of a project to inform recommendations for lung cancer screening issued by the U.S. Preventive Services Task Force.[10]

Each of the five models simulated the underlying natural history of lung cancer, including dose-response modules that relate an individual's detailed, dynamic cigarette smoking history to lung cancer risk (by histology and sex), and estimated (as an output) the effect of early detection with CT screening on lung cancer survival (Table 1 and the supplementary material of this Chapter, in the section: "Supplementary Model Descriptions"). Algorithms for following up a positive screening test (defined in our analysis as suspicious for lung cancer) were simulated with varying detail (Table 1). Prior to this analysis, all models were populated with de-identified trial participant histories and adjusted to match the trial design (e.g., numbers of screens and screening modality). All models were calibrated to reproduce multiple endpoints consistent with NLST and the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial.[11,12] Because the models simulate the natural history of disease, they can predict outcomes in years after the last year of observed follow up and in what-if scenarios with hypothetical screening programs and participants.

Table 1: Comparison of features across five independent models

	Erasmus MC	Fred Hutchinson Cancer Research Center	University of Michigan	Massachusetts General Hospital	Stanford University
Model Features	Model E	Model F	Model U	Model M	Model S
Central dose-response model	Two-stage clonal expansion (TSCE)[13]	Longitudinal multistage observation	Multistage clonal expansion	Probabilistic[14]	TSCE with modifications[13]
Diagnostic follow-up algorithm	Implicit. Stochastic chance (separately for patients with lung cancer diagnoses versus false-positives) of receipt of a set number of follow-up exams, based on the observed frequency of exams per positive exam in the NLST CT arm.	Implicit (see model E).	Implicit (see model E).	Explicit. Detailed algorithms based on size thresholds and risk factors. Simulated less-aggressive algorithms than the Fleischner guidelines to approximate the observed frequency in the NLST, which did not specify an algorithm.[15]	Explicit (see Model M).
Screening effectiveness mechanism	Cure model. Screen-detected cases experience a reduced risk of dying from lung cancer (compared to the stage-specific survival had the same tumor been diagnosed clinically). The improved prognosis is represented as a cure fraction (specific to stage, estimated via calibration to screening trial results). If curative treatment fails, the patient survives as long as if the tumor had been diagnosed clinically, corrected for lead-time.	Combination of cure model and stage shift. Model F assumes that screen-detected cancers were treated according to clinical practice guidelines with estimated cure rates that depend on both tumor stage and histology.	Stage shift model, with adjustments for age. Time to death from lung cancer detection is based on survival models that define cure by histology, stage, gender, and age at diagnosis with better outcomes associated with younger age at detection. Screening can lead to improved survival due to detection at earlier stages.	Cure model with possibility of recurrence. Patients with early-stage non-small cell lung cancer undergo resection (lobectomy, consistent with consensus practice guidelines) which removes the primary cancer. For patients with neither undetected distant (lethal) metastases nor undetected primary lung cancers in another lobe of the lung, resection is curative for lung cancer.	Cure model. The probability of lethal metastases is estimated as a function of tumor size, histology and sex. With screening, patients are more likely to be detected at early stages and before the onset of lethal metastases, and cured following standard of care; patients are not cured if detected in early stages but after the onset of lethal metastases or in advanced stages.
Operative mortality and operative candidacy	Neither varied with age.	Neither varied with age.	Neither varied with age.	Neither varied with age in comparative analysis. In second analysis, simulated decreased rates of operative candidacy for older persons, and excluded from screening anyone who was not an operative candidate. Operative mortality (applied to operative candidates with early stage cancer) was constant.	Neither varied with age

The supplementary material of this Chapter (section: "Supplementary Model Descriptions") provides additional details, including data used to develop and verify the models.

Common Model Inputs

Publicly available data were used for this analysis. All models simulated U.S. men and women (all races) born in 1950. Detailed smoking histories (including non-smokers) and non-lung cancer mortality risks were created as described below and in the supplementary material of this Chapter (in the section: "Smoking Histories and Non-Lung Cancer Mortality Risks of Simulated Cohort (born in 1950)"), and used by all models as common inputs. Smoking histories and quit rates that were previously estimated through 2000 were updated to calendar year 2009 for this analysis and years past 2009 were projected; similarly, tables of non-lung-cancer mortality rates specific to smoking history (i.e., categories of current smokers had increased risks relative to never smokers, with former smoker mortality interpolated as a function of years since quitting)) were updated to 2009 and projected past 2009.[16-18] The proportion of the 1950 cohort that had accumulated the specified number of pack-years by a given age is shown in Figure S4 in the supplementary material of this Chapter. In the NLST and the PLCO trial, individuals had substantially lower non-lung cancer mortality than the general population even after adjusting for their smoking status. Our use of U.S. population other-cause mortality rates rather than the lower rates observed in the NLST or PLCO was based on an assumption that the ''healthy volunteer'' effect in the trials would not persist if screening for lung cancer disseminated widely.

Standardized analyses

Each model was used to simulate men and women who were born in 1950 from age 45 (calendar year 1995) to death or age 90, under 576 programs and 2 reference scenarios (a no screening scenario and a scenario with a maximum of 3 screens; Table 2). Screening programs varied according to five criteria: age to start screening (45, 50, 55, 60); age to stop screening (75, 80, 85); screen frequency (every 1, 2, or 3 years); minimum number of pack-years of cigarette exposure (10, 20, 30, 40); and (for former smokers) maximum years since quitting (10, 15, 20, 25). We refer to programs using shorthand for Periodicity (A, annual, B, biennial, or T, triennial), Start Age - Stop Age - Minimum Pack-Years - Maximum Years Since Quit. For example A-55-75-30-15 represents starting screening at age 55 years and ending screening at age 75, for individuals with a minimum smoking history of 30 pack-years, and a maximum years since quitting of 15 years. This program, which we refer to as 'NLST

eligibility' is similar to the NLST design except that screening was not limited to 3 screenings (a maximum of 21 screens are possible from ages 55 to 75). As individuals age, their accumulated pack-years or years since quitting may change. In this analysis, the models assessed eligibility annually; to be screened at a specific age within the qualifying age range, an individual also had to meet both the pack-years and the years-since-quitting criteria. Thus lighter smokers may not begin screening at the start age and former smokers may cease screening prior to the stop age. All simulations were performed assuming idealized, perfect screening adherence for eligible individuals and smoking cessation was assumed to be unaffected by screening results. For the biennial and triennial programs, the frequency of screening exams was changed while retaining each model's natural history parameters, which simulate the underlying progression of disease. Model M generated a second set of results that added operative candidacy (i.e. healthy enough for curative surgery) as an eligibility criteria for screening and reduced rates of operative candidacy in older patients (described in the supplementary material of this Chapter, in the section: "Supplementary Model Descriptions").[19]

Table 2: Screening programs evaluated

Program characteristic	Values	# of Combinations
Frequency of screening	Annual, every 2 years, every 3 years	3
Age to begin screening	45, 50, 55, 60	4
Age to end screening	75, 80, 85	3
Minimum PY for screening	10, 20, 30, 40	4
Maximum YSQ for screening	10, 15, 20, 25	4
Total (including 2 reference programs)		578

Abbreviations: PY, pack-years; YSQ, years since quitting.

Reference programs: no screening and an approximation of the National Lung Screening Trial design (at age 62, 3 annual screens for smokers with ≥30 PY, and ≤15 YSQ).

All screening programs simulated U.S. cohorts born in 1950. For individuals meeting the pack-year and (for former smokers) years since quitting cutoffs, the first screen occurs at the beginning age and last screen occurs at the ending age. Programs are labeled as follows: Frequency (Annual, Biennial, Triennial) Age-Start-Age Stop-minimum-PY-maximum YSQ. As an example, B-55-85-20-15 corresponds to biennial screening starting at age 55, ending at age 85, subject to a minimum pack-year history of 20 and a maximum years since quitting (for former smokers) of 15.

Outcome Metrics

For each program, each model generated counts of screening exams and lung cancer deaths avoided relative to no screening, separately for males and females. All events are 'per person in the population' rather than 'per person screened' because programs defining eligibility based on smoking history may screen *similar* proportions of the population but screen *dissimilar* people, even for identical starting and stopping ages. Counts of screening exams excluded follow-up and incidental CT exams. Counts of deaths avoided per screening scenario were expressed as the proportion of the (within-model) maximum possible deaths avoided from any of the screening programs evaluated.

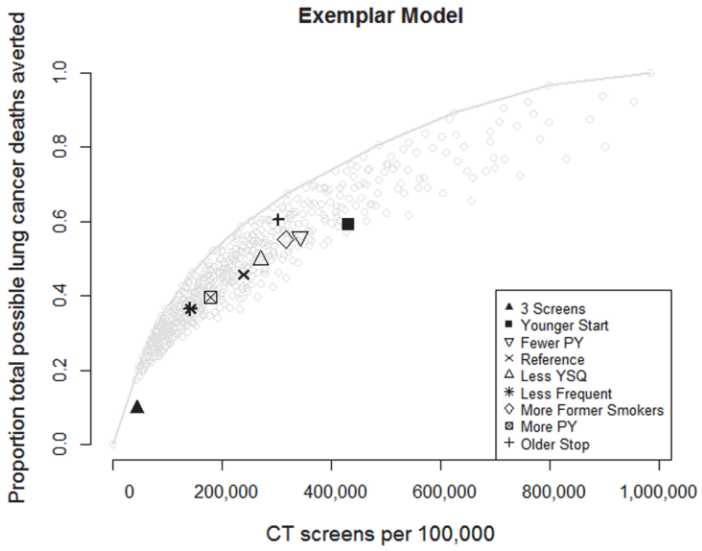
In this analysis, we sought to formally represent the trade-offs between maximizing the benefits (here, lung cancer deaths avoided) accruing to a specific screening program while simultaneously minimizing the harms (here, the numbers of screening exams required to avoid the lung cancer deaths). One way to compare alternative programs that represent different trade-offs is to generate an 'efficiency frontier'. Each model generated efficiency frontiers for each sex that connected the screening programs that prevented the most deaths for each possible value of the number of CT screens. (Note that our definition of efficiency is not equivalent to identifying the lowest ratio of screens per death avoided. As screening intensity increases, the number of screens per death avoided will increase, but among programs with similar numbers of screens, some [the most efficient] will prevent more deaths.) For each model's results, we generated a rank score (decile of distance from the model's frontier) for each program not on the frontier (described in the supplementary material of this Chapter, in the section: "Supplementary Analysis Methods").[20] Programs on or closest to the frontier (first three deciles) as predicted by at least 3 models were identified for males and females separately. Programs that were in both male and female lists were defined as consensus programs. For each consensus program, we combined counts per 100,000 persons from males and females and calculated the mean predicted counts of lung cancer cases, lung cancer deaths, life-years, and screening CT exams performed. We calculated the percent of the cohort receiving at least one screening exam and the number of persons ever screened per lung cancer death avoided (number needed to screen, NNS). A secondary set of consensus programs for which the benefit (i.e., the y-

axis) was measured as life-years saved (with the x-axis remaining counts of CT screens) was also identified, using the identical steps as above.

Results

Using eligibility criteria like those in NLST, neither 3 annual screens (A62-64-30-15) nor 21 annual screens (A55-75-30-15) appears on the frontier for any model (Figure 1 and Figure S7 in File S1). There was variability among the models with respect to the effects of the smoking criteria on distance from the frontier, but consensus was clear regarding age: compared with A55-75-30-15, all models placed A55-85-30-15 closer to (or on) the frontier, indicating that continuing screening to older ages was more efficient than stopping at age 75. Conversely, initiating screening at younger ages (A45-75-30-15) was farther from the frontier (less efficient). Less-frequent (B55-75-30-15) screens provided fewer benefits, as did increasing the pack-year minimum (A55-75-40-15). The most intensive annual program (A45-85-10-25) was the upper right of the frontier for all models.

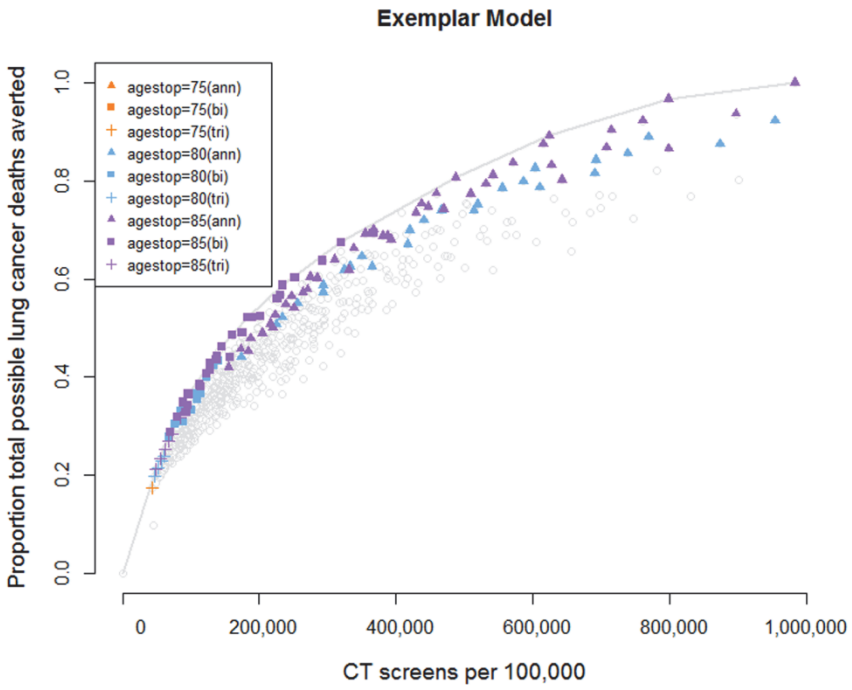
Figure 1: Systematic variation of reference screening program A-55-75-30-15



Vertical axis normalized so that 1.0 represents within-model prediction of lung cancer deaths avoided with most intensive screening program (A-45-85-10-25); values not directly interpretable as a hazard ratio. Compared to annual screening of individuals aged 55 to 74 with at least 30 pack-years of cigarette smoking and who quit with in the last 15 years (reference, x) a program of continuing annual screening to eligible individuals up to age 85 (+) was closer to the efficiency frontier. Results from one model shown; see Figure S7 in the supplementary material of this Chapter for results from all five models.

We identified 120 consensus programs. Of these, 119 had a stopping age of 80 or 85 (described in Table S2 and Figure S8 in the supplementary material of this Chapter).

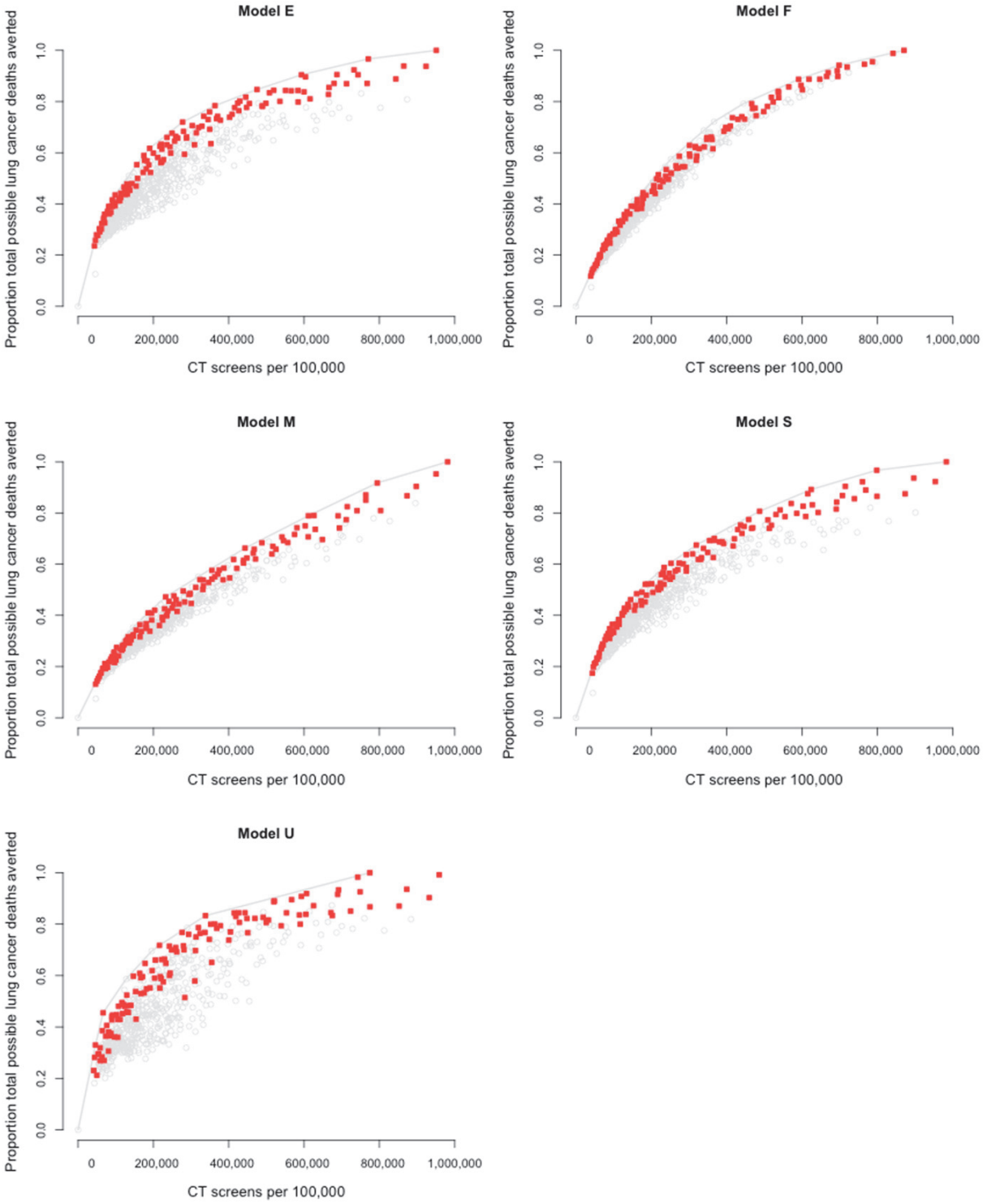
Figure 2: Exemplar model showing consensus programs



Vertical axis normalized as in Figure 1. Consensus programs were the 120 (out of 576 evaluated, see Table 2) that five models ranked as most efficient. Only a single consensus strategy (the single orange +) had a stop age of 75. The remaining consensus strategies continued screening of individuals meeting the smoking eligibility criteria to ages 80 (aqua) or 85 (purple). Annual screening (triangles) provided greater benefits (i.e., averted more lung cancer deaths) than triennial (+) or biennial (squares). Results from one model shown; see Figure S8 in the supplementary material of this Chapter for results from all five models.

Across the 120 consensus programs, the average start age (54.8 y) and the average minimum pack-years (27.1) were close to the NLST criteria but the average maximum years since quit was higher (19.9 years). For all models (Figure 3), the 120 consensus programs are close to the model's own frontier. Results from a selected subset of 41 (every third, sorted by percent ever screened) consensus programs are provided in Table 3 (mean and SD of results from the five models). Between 11% and 40% of the cohort was screened, requiring between 43,000 to over 920,000 CT screens per 100,000 persons (Table 3). The models predicted an average of 3,719 lung cancer deaths per 100,000 in the no screening scenario (SD 820.43; described in Figure S6 in the supplementary material of this Chapter). Per 100,000 persons, the 41 consensus programs would avoid between 153 and 846 lung cancer deaths and save between 1,883 and 9,851 years of life, relative to no screening, and the mean predicted NNS varied from 34.5 to 94.2. Based on results from one model (M), reducing the proportions of older individuals screened (due to ineligibility for surgical resection) resulted in fewer CT screens and fewer lung cancer deaths avoided (13.3% and 14.8%, respectively, across the consensus programs), but programs that extended screening to ages 80 and 85 remained on the efficiency frontier (described in Figure S9 in the supplementary material of this Chapter). When the benefit of screening was measured as life-years saved rather than lung cancer deaths avoided, the second set of consensus efficient programs had younger average start and stop ages (49.5 years and 80.9 years, respectively) but similar average minimum pack-years and maximum years since quit (described in Table S3 in the supplementary material of this Chapter).

Figure 3: Normalized plots from all models showing consensus programs



Shown are efficiency frontiers for all 5 models, with the 120 consensus programs marked. All vertical axes are normalized to within-model predictions, as in Figures 1 and 2.

Table 3: Mean (SD) predicted benefits from 5 models for 41 selected (of 120) consensus programs (both sexes combined)

Program characteristics: Freq-Start-Stop-PY-YSQ	% cohort ever screened^ (mean)	% cohort ever screened^ (SD)	Number of CT screens (mean)	Number of CT screens (SD)	Lung cancer deaths avoided** (mean)	Lung cancer deaths avoided** (SD)	NNS (mean)	NNS (SD)	Life-years saved** (mean)	Life-years saved** (SD)
T-60-75-40-10	11.1	1.0	42,893	2,757	153	72	94	64	1,896	1,093
T-60-80-40-10	11.2	1.0	45,685	3,223	173	78	85	60	1,883	1,201
B-60-85-40-10	11.3	1.1	69,662	4,466	256	115	59	44	2,771	1,639
T-60-85-40-15	12.0	1.2	55,316	3,573	201	93	77	52	2,085	1,426
T-60-80-40-20	12.6	1.0	56,712	3,502	197	88	81	52	2,138	1,344
B-60-85-40-20	12.7	1.0	88,781	4,802	288	138	57	37	2,943	1,957
T-60-80-40-25	12.9	0.9	60,570	3,483	202	92	80	47	2,299	1,352
T-60-85-40-25	13.0	0.9	66,333	3,578	225	106	73	44	2,344	1,559
A-60-85-40-25	13.0	0.9	185,451	8,027	449	219	38	25	4,394	2,859
A-55-85-40-15	13.7	0.8	200,575	10,864	445	223	41	29	4,740	2,844
T-55-85-40-25	13.9	0.9	83,043	4,633	252	120	70	44	2,767	1,702
A-55-85-40-20	14.0	0.9	220,505	10,542	485	237	38	26	4,958	3,029
B-50-80-40-25	14.5	0.6	137,944	6,221	358	167	51	32	4,012	2,216
B-50-85-40-25	14.6	0.7	143,621	6,835	376	178	49	30	4,090	2,377
A-50-85-40-25	14.6	0.7	281,218	11,061	542	261	35	22	5,955	3,161
A-60-85-30-10	15.6	1.0	180,599	7,772	412	200	50	34	4,212	2,603
A-60-85-30-15	16.9	1.1	213,400	8,568	457	232	49	32	4,666	2,964
B-60-85-30-20	17.9	1.2	127,046	4,888	358	166	64	41	3,591	2,304
A-60-85-20-10	18.3	1.0	214,153	7,742	452	218	53	35	4,613	2,839
A-55-80-30-15	19.3	1.0	286,813	11,098	521	268	49	31	5,603	3,278
A-55-85-30-20	20.2	0.8	331,990	11,705	593	305	44	27	6,237	3,642
A-55-85-30-25	20.4	0.9	361,001	11,107	628	323	42	25	6,469	3,822
A-50-85-30-15	21.2	0.7	382,439	15,625	608	316	45	27	6,998	3,596
A-50-85-30-20	21.4	0.8	419,782	15,070	653	336	42	25	7,244	3,781
A-45-85-30-25	22.0	0.7	520,793	18,498	707	362	39	22	7,775	3,959
B-60-85-20-20	23.2	1.0	158,397	4,474	399	185	73	44	4,070	2,508
A-60-85-20-25	24.8	1.0	348,894	6,919	624	314	51	30	6,120	3,857
A-55-80-20-20	26.6	0.9	410,565	10,425	631	342	55	32	6,928	3,892

B-55-85-20-25	27.4	1.1	247,058	6,305	501	256	69	39	5,256	3,153
A-50-85-20-15	27.9	0.9	496,010	15,834	685	378	53	30	7,688	4,118
A-60-85-10-20	28.0	2.0	370,825	19,139	605	296	59	34	6,108	3,671
A-50-85-20-20	28.7	1.0	557,513	15,580	737	411	50	28	8,028	4,450
A-50-85-20-25	29.0	0.9	610,443	14,822	787	427	47	25	8,746	4,512
A-45-80-20-25	29.9	1.1	721,956	19,536	780	453	49	25	9,206	4,531
A-55-85-10-15	29.9	2.3	448,193	26,722	651	332	59	34	6,876	3,909
A-60-85-10-25	31.1	2.1	427,669	21,334	660	322	59	32	6,474	3,951
A-50-80-10-15	34.6	2.3	583,756	35,681	700	388	63	34	8,036	4,143
A-55-85-10-25	36.0	2.0	590,101	31,172	768	397	59	31	8,109	4,454
A-50-85-10-20	37.5	2.0	685,484	39,445	795	422	59	31	8,772	4,509
A-50-85-10-25	38.9	1.9	767,313	40,320	851	443	57	28	9,151	4,735
A-45-80-10-25	40.3	1.9	920,505	45,739	846	479	60	29	9,851	4,737
Average CV	*0.06*		*0.04*		*0.50*		*0.61*		*0.58*	

Abbreviations: Frequency, A = annual, B = biennial (every 2 years), T = triennial (every 3 years); Start Age, Stop Age, PY = minimum pack-years, YSQ = maximum years since quit.

NNS, Number (people) needed to screen (ever) to prevent one lung cancer death.

Percentage of cohort screened, numbers of CT screens, lung cancer deaths avoided, and life-years saved are all normalized to cumulative counts per 100,000 people in the cohort at age 45 (including non-smokers and persons not screened), followed to age 90. See the Table S2 in the supplementary material of this Chapter for a complete list of 120 consensus programs identified from the 576 programs evaluated.

^Percent of cohort that received at least one screen; eligible individuals varied across programs.

*** Numbers of lung cancer deaths avoided and life-years saved were first calculated per model, comparing each model to its own results for lung cancer deaths in the no-screening arm. Shown are averages across models. The average (across models) number of lung cancer deaths in the no screening scenario was 3,719 (SD 820).*

Average Coefficient of Variation (CV) calculated as the average of (SD/mean) for each program in the table. Lower values indicate less dispersion of estimates from the models for that endpoint, across the selected consensus programs.

Discussion

Five independent models ranked 576 lung cancer screening programs by weighing one metric of their potential benefits (lung cancer deaths avoided) against one measure of harms or resource use (counts of CT screening exams) in the U.S. cohort born in 1950. The models had been previously calibrated to multiple endpoints in NLST, but heterogeneity in the underlying model structures and assumptions yielded heterogeneous predictions for absolute numbers of lung cancer deaths avoided when extrapolating beyond the trial data.[12] A key finding of our analysis was that despite differences in absolute benefits across the models, the ranking of programs was consistent; while accounting for the heterogeneity in model predictions, we were able to identify a set of consensus efficient programs.

Annual screening with eligibility based on NLST criteria (beginning at age 55, continuing to age 75 for current and former smokers with a minimum of 30 pack-years and less than 15 years since quitting) was not among the programs on the efficient frontier of any of the five models. Results from all models showed that programs that extended the screening age beyond 75 prevented more lung cancer deaths for relatively few additional screens. Note that in our modeling, the stopping age for a program was the last screen for any individuals who still met the smoking cutoffs, and not the last year to be invited to begin a screening program. In the NLST which had an upper eligibility age of 74 years, individuals were as old as (77 or, rarely, 78) at the third screen. Our finding that programs that screened eligible individuals past age 75 years were efficient was unchanged when more older patients were ineligible for screening due to comorbidities that categorized them as non-operative candidates (based on results from one model) or when life-years saved was substituted for the measure of benefit.

While in other cancers (e.g. breast and colorectal) screening is not generally recommended beyond age 75 and not generally recommended every year, in lung cancer annual screening to older ages can be beneficial because: (1) the age-specific incidence curve for lung cancer is quite steep, and (2) the high lethality of the disease makes early detection worthwhile, even among individuals with a somewhat modest life expectancy. It is also important to note that had we defined life-years saved (instead of lung cancer deaths avoided) as the

measure of benefit, one could logically predict that strategies with younger stopping ages would be more likely to emerge as 'consensus efficient'. Our predicted NNS for A-55-80-30-15 varied across models, ranging from 19.8 (Model F) to 100.5 (Model M), but all were below published estimates of NNS for only 3 screens (256) and closer to published NNS for mammography (95) or FOBT (roughly 130) for healthy 50 year-olds.[21,22] For consensus programs with screening until age 80, between 11% (for the least frequent programs with strictest eligibility, e.g., T-60-75-40-10) and 40% (for the annual programs with more inclusive eligibility, e.g., A-45-80-10-25) of the cohort born in 1950 would be screened at least once after age 45. Although not directly comparable to earlier estimates that 6% (8.7 million people) of U.S. adults over 40 would meet the NLST eligibility cutoffs for lung cancer screening each year, our estimate of 11% of individuals seems reasonable.[23,24]

We identified a set of consensus efficient programs rather than a single optimal strategy, because the efficiency frontiers did not identify a consensus inflexion point at which additional screens provided diminishing benefits. The least intensive programs at the lower left of the frontiers (Figure 2) may be less attractive, however, since annual screening consistently prevented more lung cancer deaths than did triennial or biennial programs. The most-intensive screening programs, on the other hand, will lead to more accumulated harms (radiation exposure from additional imaging examinations, overdiagnosis, invasive biopsies) and costs.

Screening programs cannot be evaluated in isolation from the follow-up algorithm. In the NLST, an average of 24% of individuals in a given round of screening (CT arm) had results requiring some follow-up, but the trial did not specify a follow-up regimen, leaving open the question of the optimal regimen for individuals with positive screens, most of whom are healthy.[4,25] In models (E, F, U) that used implicit follow-up algorithms based on the experience of participants in the NLST, extrapolating the rate of follow-up to less frequent screening programs was dependent on the assumption that the rates of follow up exams and early detection of lung cancers (defined in the NLST and models E, F, and U as 'screen-detected' even if first seen on a follow-up exam) would not change. In the models (M, S) that explicitly modeled follow-up programs based on size, follow-up exams could change the

timing of detection of a lung cancer, but the assumptions used here for frequency of follow-up imaging may not be representative of eventual practice patterns.

Several limitations of our analysis are important to note. The models do not simulate non-lung cancer incidental findings (e.g., coronary artery calcification, abdominal aortic aneurysm, or other malignancies), so our results do not include potential benefits (or harms) due to their detection and treatment. There are few data to predict adherence patterns for lung cancer screening, and many possibilities to model.[23,26] We conducted an idealized analysis with the goal of informing guidelines and did not consider that individuals will self-select for participation in screening based on their comorbidities, specific smoking history, or family history, as observed in screening trials.[27,28] It will be important to monitor how lung cancer screening is implemented in community settings (including recruitment, participation, positive screen evaluations, diagnosis, referral for treatment), and modeling can suggest the most important leverage points to optimize the process.

Definitive evidence on the relationship between smoking cessation and NLST screening results was not available in time for our analyses. Based on limited data with non-standardized definitions of 'quit' and the PLCO Trial, which found no correlation between CXR screening result and smoking behavior, we assumed screening did not affect background smoking patterns.[29-33] Efficient screening programs might differ in populations with different smoking patterns or other-cause mortality risks than the cohort we simulated. To simplify the comparison of hundreds of programs, we performed our analyses in a single birth cohort and did not estimate total lung cancer deaths avoided in the U.S.[34] Our requirement that individuals meet all eligibility criteria (including years since quitting) was transparent and is a step towards risk-based screening criteria (our models account for decreasing risks of death from lung cancer and other causes after quitting), but may not reflect guidelines, which typically define eligibility to begin screening. Future analyses to examine programs that define eligibility based on risk models will require that the models and population input files include additional characteristics (e.g., BMI, education) that go beyond age and smoking exposure.[35-39] We did not incorporate increases in operative mortality rates by age, or special clinical considerations individual to a particular patient.

Although the rankings of programs were consistent across models, uncertainty in absolute numbers of lung cancer deaths avoided (and life-years saved) remained, due to variation in the underlying assumptions regarding unobserved disease processes.[40] Underlying the differences across models in predicted absolute benefits is a variation in the predicted future number of lung cancer cases in the absence of screening (shown in Figure S5 in the supplementary material of this Chapter). Essentially, our consortium of 5 models served as a sensitivity analysis on model structure and demonstrated that even when model heterogeneity was specifically taken into account, the models identified similar efficient programs (i.e., the consensus set). Our results highlight trade-offs between preventing greater numbers of lung cancer deaths and the additional screening exams required. Guidelines for screening also consider trade-offs in gains in life expectancy and important harms, including invasive biopsies for benign disease, overdiagnosis, and lung cancers related to radiation from diagnostic imaging examinations.[10] Difficulties with estimating population effects of screening include the potential for concurrent smoking cessation programs to augment the benefits from screening, and the heterogeneity of the radiation dose attributable to a given CT exam, which could vary as much as 10-fold depending on the size of the patient, the generation of scanner, and the protocol in use at the clinical setting.[41] All smokers, whether undergoing screening or not, should receive cessation assistance and be encouraged to quit.[42]

Acknowledgments

In addition to the authors, a team of investigators from multiple institutions (NLST and PLCO investigators, MGH, Stanford, Yale, Cornerstone Systems Northwest, and Information Management Services contributed to the analysis.

References

1. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
2. Pinsky PF (2013) Subset and Histological Analysis of Screening Efficacy in NLST. National Cancer Advisory Board. Bethesda, MD: Department of Health and Human Services.
3. Wender R, Fontham ETH, Barrera E, et al. American Cancer Society lung cancer screening guidelines. *CA: A Cancer Journal for Clinicians* 2013; **63**(2): 106-17.
4. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of ct screening for lung cancer: A systematic review. *Journal of the American Medical Association* 2012; **307**(22): 2418-29.
5. Jacobson FL, Austin JHM, 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. *The Journal of Thoracic and Cardiovascular Surgery* 2012; **144**(1): 25-32.
6. Jaklitsch MT, Jacobson FL, Austin JHM, et al. The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. *The Journal of Thoracic and Cardiovascular Surgery* 2012; **144**(1): 33-8.
7. Wood DE, Eapen GA, Ettinger DS, et al. Lung Cancer Screening. *Journal of the National Comprehensive Cancer Network* 2012; **10**(2): 240-65.
8. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2009; **151**(10): 738-47.
9. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of Reduced Tobacco Smoking on Lung Cancer Mortality in the United States During 1975–2000. *Journal of the National Cancer Institute* 2012; **104**(7): 541-8.
10. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
11. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
12. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
13. Moolgavkar SH, Dewanji A, Venzon DJ. A Stochastic Two-Stage Model for Cancer Risk Assessment. I. The Hazard Function and the Probability of Tumor. *Risk Analysis* 1988; **8**(3): 383-92.
14. McMahon PM, Kong CY, Johnson BE, et al. Estimating Long-term Effectiveness of Lung Cancer Screening in the Mayo CT Screening Study. *Radiology* 2008; **248**(1): 278-87.
15. MacMahon H, Austin JHM, 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.
16. Anderson CM, Burns DM, Dodd KW, Feuer EJ. Chapter 2: Birth-Cohort-Specific Estimates of Smoking Behaviors for the U.S. Population. *Risk Analysis* 2012; **32**: S14-S24.
17. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort–Specific Smoking Histories, 1965–2009. *American Journal of Preventive Medicine* 2014; **46**(2): e31-e7.
18. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.
19. Mery CM, Pappas AN, Bueno R, et al. Similar Long-term Survival of Elderly Patients With Non-small Cell Lung Cancer Treated With Lobectomy or Wedge Resection Within the Surveillance, Epidemiology, and End Results Database. *Chest* 2005; **128**(1): 237-45.
20. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *European Journal of Operational Research* 1978; **2**(6): 429-44.
21. Bach PB, Gould MK. When the average applies to no one: Personalized decision making about potential benefits of lung cancer screening. *Annals of Internal Medicine* 2012; **157**(8): 571-3.

22. Walter LC, Covinsky KE. Cancer screening in elderly patients: A framework for individualized decision making. *Journal of the American Medical Association* 2001; **285**(21): 2750-6.
23. Doria-Rose VP, White MC, Klabunde CN, et al. Use of Lung Cancer Screening Tests in the United States: Results from the 2010 National Health Interview Survey. *Cancer Epidemiology Biomarkers & Prevention* 2012; **21**(7): 1049.
24. Pinsky PF, Berg CD. Applying the National Lung Screening Trial eligibility criteria to the U.S. population: what percent of the population and of incident lung cancers would be covered? *Journal of Medical Screening* 2012; **19**(3): 154-6.
25. Duke SL, Eisen T. Finding needles in a haystack: annual low-dose computed tomography screening reduces lung cancer mortality in a high-risk group. *Expert Review of Anticancer Therapy* 2011; **11**(12): 1833-6.
26. Jonnalagadda S, Bergamo C, Lin JJ, et al. Beliefs and attitudes about lung cancer screening among smokers. *Lung Cancer* 2012; **77**(3): 526-31.
27. Pinsky PF, Miller A, Kramer BS, et al. Evidence of a Healthy Volunteer Effect in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *American Journal of Epidemiology* 2007; **165**(8): 874-81.
28. Dominioni L, Rotolo N, Poli A, et al. Self-Selection Effects in Smokers Attending Lung Cancer Screening: A 9.5-Year Population-Based Cohort Study in Varese, Italy. *Journal of Thoracic Oncology* 2010; **5**(4): 428-35.
29. Ostroff JS, Buckshee N, Mancuso CA, Yankelevitz DF, Henschke CI. Smoking Cessation Following CT Screening for Early Detection of Lung Cancer. *Preventive Medicine* 2001; **33**(6): 613-21.
30. Park ER, Gareen IF, Jain A, et al. Examining whether lung screening changes risk perceptions: National Lung Screening Trial participants at 1-year follow-up. *Cancer* 2013; **119**(7): 1306-13.
31. van der Aalst CM, de Koning HJ, van den Bergh KAM, Willemsen MC, van Klaveren RJ. The effectiveness of a computer-tailored smoking cessation intervention for participants in lung cancer screening: A randomized controlled trial. *Lung Cancer* 2012; **76**(2): 204-10.
32. van der Aalst CM, van den Bergh KAM, Willemsen MC, de Koning HJ, van Klaveren RJ. Lung cancer screening and smoking abstinence: 2 year follow-up data from the Dutch–Belgian randomized controlled lung cancer screening trial. *Thorax* 2010; **65**(7): 600.
33. Barry SA, Tammemägi MC, Penek S, et al. Predictors of Adverse Smoking Outcomes in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Journal of the National Cancer Institute* 2012; **104**(21): 1647-59.
34. Ma J, Ward EM, Smith R, Jemal A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer* 2013; **119**(7): 1381-5.
35. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.
36. 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. *Annals of Internal Medicine* 2012; **157**(4): 242-50.
37. Tammemägi CM, Pinsky PF, Caporaso NE, et al. Lung Cancer Risk Prediction: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Models and Validation. *Journal of the National Cancer Institute* 2011; **103**(13): 1058-68.
38. Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *Journal of the National Cancer Institute* 2003; **95**(6): 470-8.
39. Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *British Journal of Cancer* 2007; **98**(2): 270-6.
40. McMahon PM, Hazelton WD, Kimmel M, Clarke LD. Chapter 13: CISNET lung models: comparison of model assumptions and model structures. *Risk Analysis* 2012; **32 Suppl 1**: S166-78.
41. 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. *Annals of Internal Medicine* 2013; **159**(6): 411-20.

42. Pirie K, Peto R, Reeves GK, Green J, Beral V. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. *The Lancet* 2013; **381**(9861): 133-41.

Chapter 4

Supplementary material

Supplementary Model Descriptions

See Table S1. Earlier versions of three of the models were described in a monograph, both individually and in comparison.[1-4] Additional details of some models are also available at www.cisnet.cancer.gov/profiles.

Calibration to NLST and PLCO trials.

The details of the methods and results were provided in Chapter 1 of this thesis.

Benign pulmonary nodules

Models M and S simulate the presence of benign pulmonary nodules. In both models, if benign nodules are detected on imaging exams, the patient's management and outcome may change.

Model M assumes that benign nodules arise according to age, and may resolve spontaneously also as a function of age. Individuals may have up to three benign nodules (and three lung cancers, for a maximum total of six nodules at one time). The sizes of benign nodules are drawn from a lognormal distribution (mean = 0.9, variance 0.36), and the locations of these nodules are based on Mayo Clinic data.[5] Recent results from the NLST support the use of benign pulmonary nodule (which cause false-positives) prevalences from areas in the 'histoplasmosis belt' for the rest of the US.[6]

Model S assumes that benign nodules arise as a function of age.

Operative mortality and operative candidacy

For all models in the base case, operative mortality was assumed constant by age. Model M assumed a 1% operative rate. Other models do not explicitly model surgical mortality. In the secondary analysis in which Model M varied the probability of operative candidacy by age, we used the following values, approximated from a figure in Mery, et al.[7] from an analysis of stage I/II NSCLC patients in SEER, 1992-1997 (n = 14,555): For ages younger than 65 years, 92% of stage I/II lung cancer patients were assumed to be operative candidates. For ages 65-74 years and over age 75 years, the percentages were 86% and 70%, respectively.

Follow-up

Across individuals in NLST with a positive CT-screen, the mean number of follow-up CTs was roughly 1, expressed as distributions around the mean numbers of follow-up CTs given gender and screening round (and other covariates that we are not using as of now). The mean number was not stratified by cases vs. false-positives, since the number of cases was negligible vs. the number of false-positives. Most of the follow-up CTs occurred within a year of the screen. For the biennial and annual extrapolation simulations, Model U scans a fraction of individuals (~ 26-27% - false-positive rate in NLST) exactly a year after the screen.

Models M and S simulate an explicit follow-up algorithm based on the size of the nodule (in Model M, based on the nodule with the largest diameter, if more than one nodule exists). For example, for nodules between 4 and 6 mm, Model M would simulate high-resolution CT exams at 12 and 24 months, and nodules between 6 and 8mm would undergo follow-up exams at 9 and 24 months. In the analyses, the screening exams that would have occurred during that 24 month period would be missed.

Table S1. Additional detail on models

	Erasmus	**FHCRC**	**Univ. of Michigan**	**MGH**	**Stanford**
	Model E	Model F	Model U	Model M	Model S
Data sources used for model development, calibration or validation	NLST and PLCO[8]; SEER 2000-2008 incidence by age, stage, histology; and NHS, HPFS.[9]	NLST and PLCO[8]; model developed using PLuSS CT and CARET.[10]	NLST and PLCO[8]; NHS/HPFS LC incidence;[9] SEER LC survival by sex, age, histology and stage; NLST and PLCO[8]; NHS/HPFS LC incidence[9]; SEER 2000-2008 LC survival by sex, age, histology and stage.	NLST and PLCO[8]; SEER 1990-2000 incidence by age, stage, histology; survival by stage; Mayo CT; LSS.[3,5]	NLST and PLCO[8]; NHS/HPFS- LC incidence[9], SEER 1988-2003 survival by histology.
Representative prior uses of the model	Evaluation of tobacco control programs[11] and screening programs in cancers other than lung.[12-15]			Evaluation of tobacco control programs[11] and screening and treatment interventions.[5,16,17]	Comparing lung and breast cancer screening.[18]
Model simulates metastasis explicitly	No	No	No	Yes	Yes

Supplementary Analysis Methods

From each model, separately for males and females, a .csv file containing counts of CT exams and counts of lung cancer deaths avoided (relative to no screening) was generated for each of the 576 scenarios (a total of 578 scenarios, including a no screening reference scenario and a reference scenario with 3 screens). Each model simulated at least 1 million individuals per scenario. Cumulative counts of events from age 45 to age 90 were normalized per 100,000 individuals in the cohort at age 45. For each model, we plotted CT screening exams (x-axis) and lung cancer deaths avoided, relative to a no screening reference strategy (y-axis). The y-axis

was normalized to display the proportion of the maximum possible lung cancer deaths avoided, as predicted by that model, from the most-intensive strategy modeled (A-45-85-10-25).

In R (version 2.15.2, RStudio version 0.97.248), the convex hull function was used to identify the efficiency frontier. Secondly, the nonparametric *nonparaeff* package in R, which uses data envelopment analysis (DEA) method to measure efficiency (productivity) was used to measure an 'efficiency score', a measure of distance from the efficiency frontier.[19,20] Data Envelopment Analysis (DEA) measures the relative efficiencies of strategies, units, or organizations with multiple inputs and multiple outputs. We used the output-oriented DEA model, which maximizes the output (deaths avoided) for each value of the input (counts of CT screens) held constant to generate efficiency scores. An efficiency score of 1 identifies scenarios on the frontier, and increasing scores (>1) indicate increasing distance from the frontier. We grouped the scores into deciles and defined each decile as an efficiency rank. Specifically, programs with ranks 1, 2, and 3 were on or near the frontier and were termed 'optimal' programs. We used the program's rank from each model to compare results across models.

We identified strategies that were in at least 3 models' acceptable region, which yields a list of consensus strategies in males and (separately) females, and identified the intersection of the male and female lists of consensus strategies to yield a list of consensus strategies applicable to both sexes. Outcomes shown are averages of five model's estimates, for males and females combined (weighted by population distribution at age 45: female contribution = 0.5188053 and male contribution = 0.4811947).

Smoking Histories and Non-Lung Cancer Mortality Risks of Simulated Cohort (born in 1950)

The 1950 birth cohort represents the U.S. population aged 45 in 1995, with respect to smoking patterns (including non-smokers) and other-cause (non-lung cancer) mortality. In the absence of screening, mortality rates in the cohort would correspond to U.S. rates (available from the National Center for Health Statistics) from 1995 to 2012, or ages 45 to 62. Estimates of smoking history parameters were derived from 33 National Health Interview Surveys (NHIS) conducted from 1965-2009 and temporal trends were used to extrapolate smoking patterns to future years, as described below.

Thirteen of the NHIS surveys (carried out from 1970-2001) included more detailed questions on age at initiation for subjects with a history of smoking and age at cessation for former smokers. This additional detail provided an approach for retrospectively constructing smoking histories for those surveyed, but because of the cross-sectional nature of these data, important details on the experience of those not surveyed were not available. Especially important were mortality differences caused by cigarette smoking, which would be expected to bias (downward) the estimated proportion of smokers in a generation as they are sampled at older ages. Correction for this bias is essential and in this analysis we have extended the approach used by

Anderson, et al. for NHIS surveys up to 2000 to include nine additional surveys up to 2009.[21-24] The approach essentially followed that used by Anderson et al., yielding estimates of (a) yearly age-specific initiation and cessation probabilities, (b) prevalence of current, former and never smokers, and (c) quintile of dose (mean cigarettes smoked per day).

Temporal trends in smoking were analyzed using age (*a*), period (*p*) and birth cohort (*c*) factors in the model. The well-known identifiability problem affects interpretation of the separate effects of each temporal component in models, but estimated outcomes resulting from model fitting are estimable functions of the model parameters and thus not affected by the problem.[25]

Ever smoker prevalence: Data from all surveys were partitioned into five year cohorts and the proportion who had ever smoked by single year of age provided estimates of prevalence among survivors. A nonlinear model for the proportion of ever smokers as a function of age gave a good description of the data, where the prevalence of ever smokers at age a for cohort *c*, $P_{\mathrm{E}}(a;c)$ is

$$P_{\mathrm{E}}(a;c) = \left(\beta_0(c) - \exp\left\{\beta_1(c)\left[a-30\right]\right\}\right)\Big/100$$

where $\beta_0(c)$ and $\beta_1(c)$ are unknown parameters that were estimated for each cohort from the NHIS data. Linear interpolation for these five year cohorts yielded estimates for single year cohorts. These estimates were used for prevalence for individuals 30 or older. For the entire age span we also included initiation probability estimates, described below.

Initiation probability: Surveys reporting age of initiation were used to estimate the conditional probability that a never smoker of a given age began to smoke during that year. In a manner similar to that used for cessation probabilities, a constrained additive spline for age, period and cohort was fitted to the logit of the conditional probabilities of smoking initiation. Multiplying subsequent conditional probabilities provide an estimate of the cumulative probability of smoking initiation, which is essentially the ever smoker prevalence. However, for some cohorts there is a considerable lag between age of initiation and age at survey, providing time for substantial effects of differential smoking related mortality to take hold. Alignment between this estimate and the cross sectional estimate described above was accomplished by finding the appropriate multiplicative constant that aligned the curves at the age for the cohort at the first survey, i.e., 1965. The result was an estimate of the prevalence of ever smokers for all ages for each cohort, $P_{\mathrm{E}}(a;c)$.

Cessation probability: Using data from surveys that reported age at cessation, birth cohort was calculated from interview age and calendar year (*c=p-a)*. For each age, the number of smokers who quit divided by the number who continued to smoke provided estimates of the conditional probability of cessation given the individual was a smoker. Constrained natural splines were used as additive effects for age, period and cohort in a linear logistic model for the yearly probability of cessation, providing estimates for the 1940, 1950 (used

for analyses in this report) and 1960 birth cohorts. The period effects for years 2010 and later were held constant at the value for 2009. These fitted conditional probabilities were then used to estimate the cumulative proportion of smokers in cohort *c* who had not ceased smoking by age *a*, $F_{\mathrm{Q}}\left(a;c\right)$.

Never smoker prevalence: The prevalence of never smokers is the complement of the prevalence of ever smokers, $P_{\mathrm{N}}\left(a;c\right)=1-P_{\mathrm{E}}\left(a;c\right)$.

Current smoker prevalence: Current smokers represent ever smokers who have not quit, which we estimate from the cumulative probability of cessation, i.e., $P_{\mathrm{C}}\left(a;c\right)=P_{\mathrm{C}}\left(a;c\right)F_{\mathrm{Q}}\left(a;c\right)$.

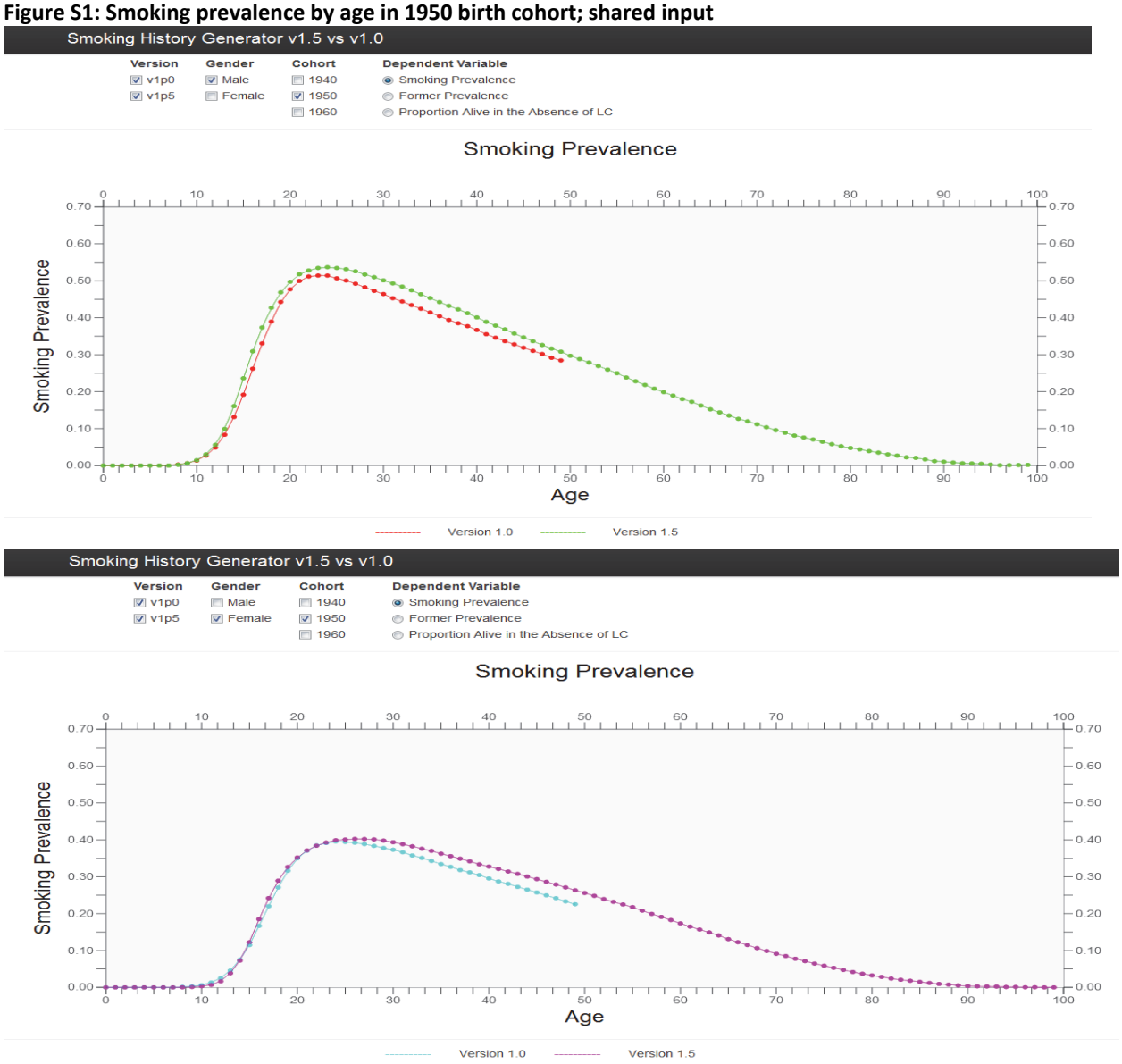
Former smoker prevalence: Former smokers are those that remain after removing never and current smokers, or equivalently, the ever smokers who have not quit,

$$P_{\mathrm{F}}\left(a;c\right)=1-P_{\mathrm{N}}\left(a;c\right)-P_{\mathrm{C}}\left(a;c\right)=P_{\mathrm{E}}\left(a;c\right)\left[1-F_{\mathrm{Q}}\left(a;c\right)\right]$$

Quintile of mean cigarette per day: Quintiles of reported cigarettes per day were calculated by ten year age (30-39, …, 80-89) and cohort (1935-1944, …, 1955-1964) groups. A linear regression model was fitted to these values for each quintile by cohort and gender, and the fitted line was used to derive estimates of the mean for each quintile by single year of age for the cohorts used in this report. Dose was ramped up from 0 at age at initiation until age 30, after which the person was simulated as smoking the mean observed dose for that quintile until cessation, if it occurred. Mean dose per quintile decreased with age after age 35, as observed in NHIS data.

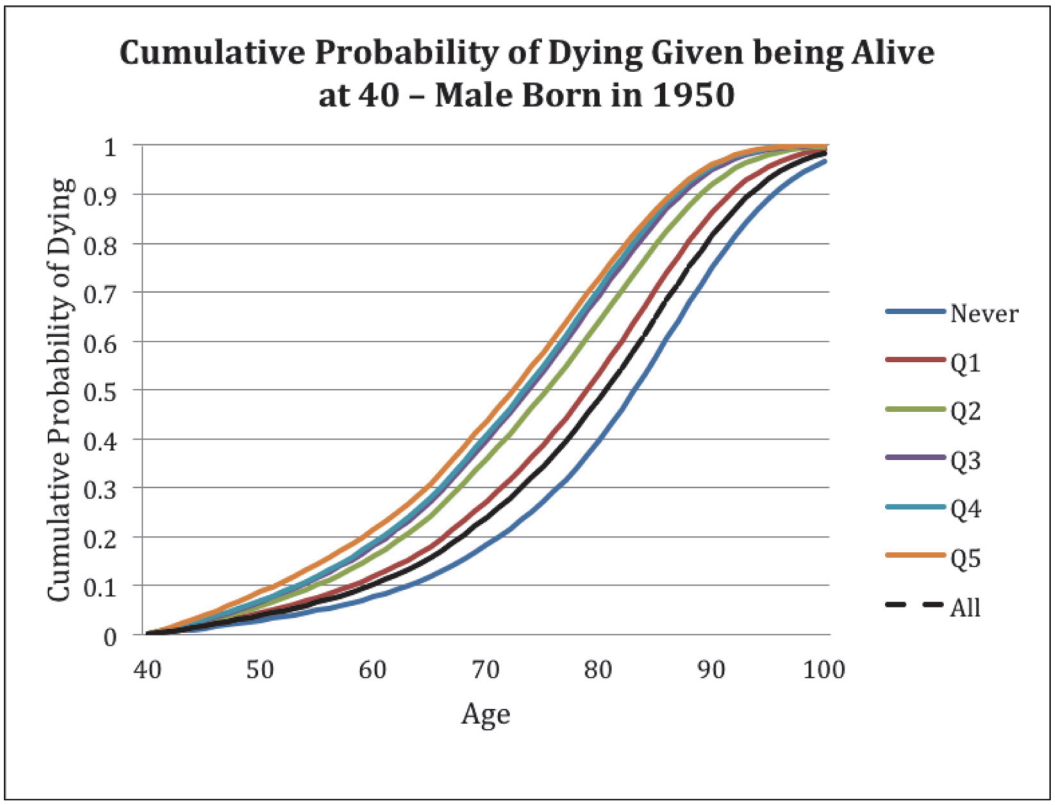
Other-cause mortality: We repeated prior methods in Rosenberg, et al. and developed quintile-based tables of non-lung cancer mortality rates (specific to the birth cohort, stratified by sex) for current and never smokers, with former smoker mortality calculated as a function of years since quit. Mortality rates were fixed at values for age 85 until age 99.[26]

Figure S1: Smoking prevalence by age in 1950 birth cohort; shared input



Summary of shared input data (used by all 5 models) on smoking patterns for the U.S. cohort born in 1950. Prevalence shown is estimated in the absence of lung cancer mortality. Version 1.0 of the Smoking History Generator (SHG) refers to published data through 2000 (Anderson, et al.), and version 1.5 supplies the 1950 birth cohort used for this analysis with data through 2009 and projections past 2009.

Figure S2: Non-lung cancer mortality by age, sex, and smoking; shared input



These curves show the other-cause (non-lung cancer) mortality for never smokers and for current smokers by smoking quintile (Q, of cigarettes per day) for the male birth cohort of 1950, out to age 99. Former smokers are intermediate to current and never smokers. There is a similar plot for females. These were shared inputs used by all the models. Note that the rates of non-lung cancer mortality represent the U.S. population, not trial (NLST or PLCO) participants.

Figure S3: Output from one model showing smoking prevalence by age (calendar year), in a no screening scenario. Proportions of current/former/never smokers are in the presence of lung cancer mortality as well as all-cause mortality

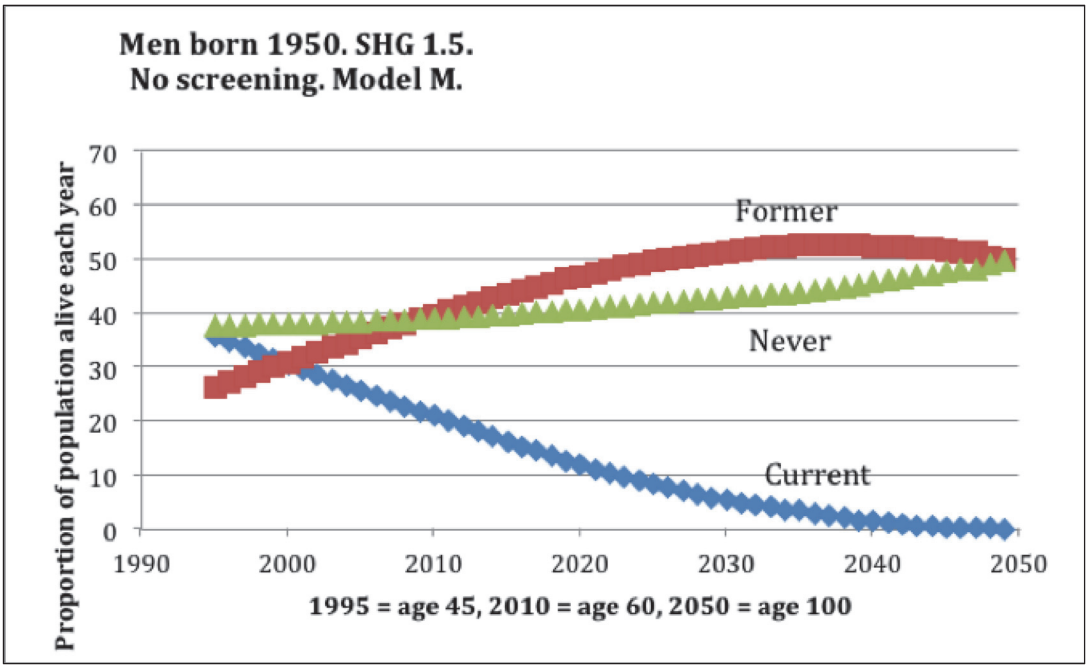


Figure S4: Output from one model showing smoking prevalence by category of pack-year and age. The proportion of the cohort by age that has accumulated the specified number of pack-years in the presence of lung cancer mortality and other-cause mortality

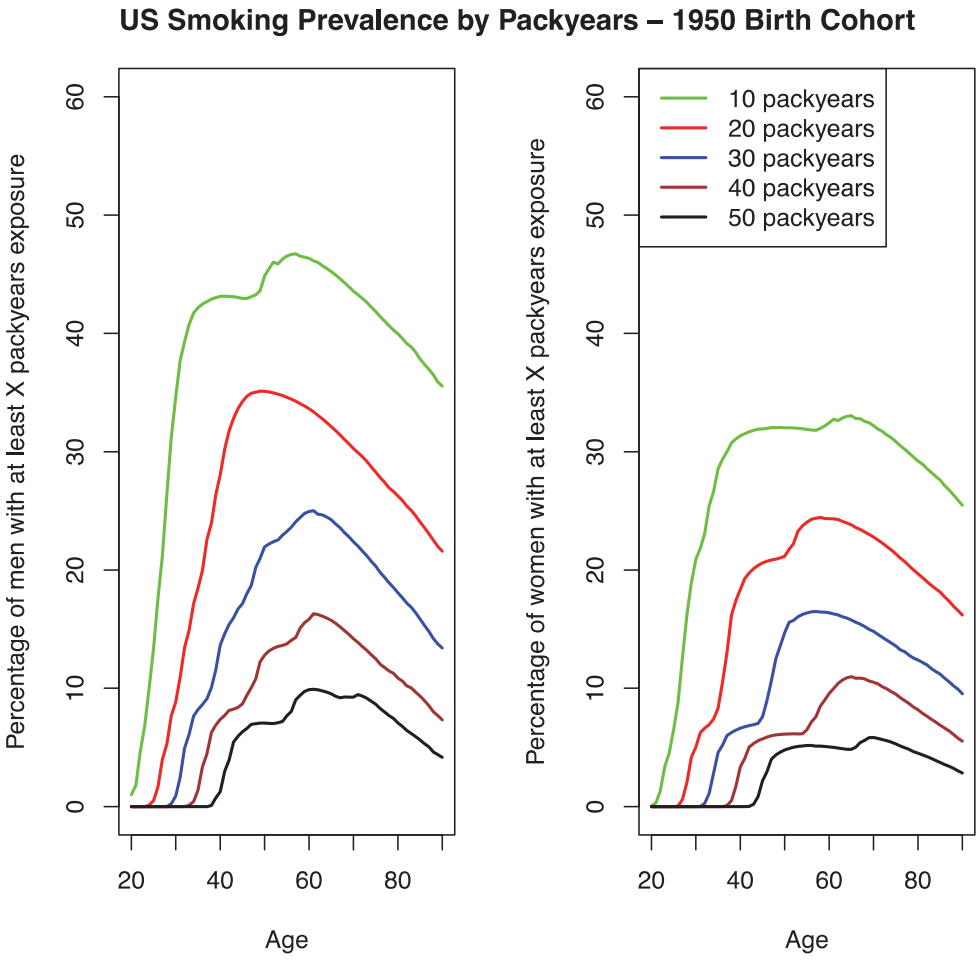
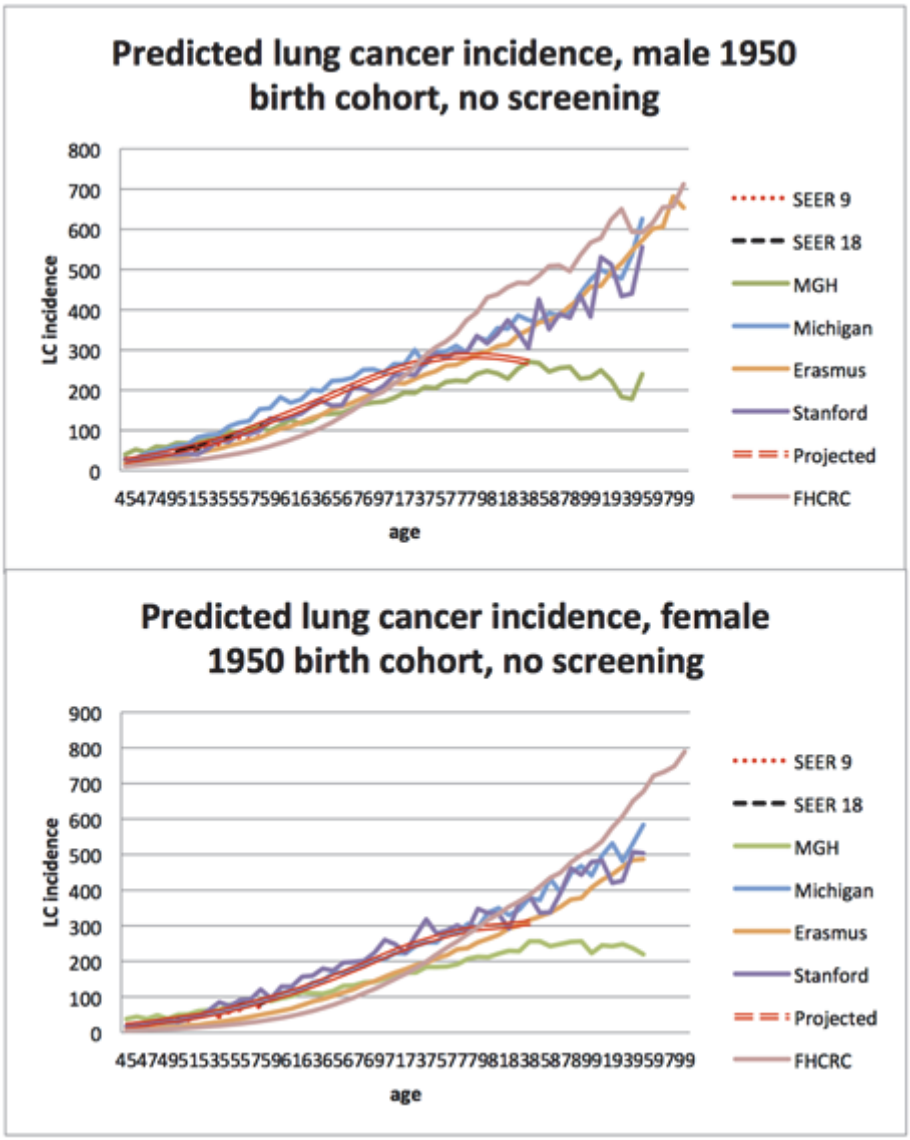
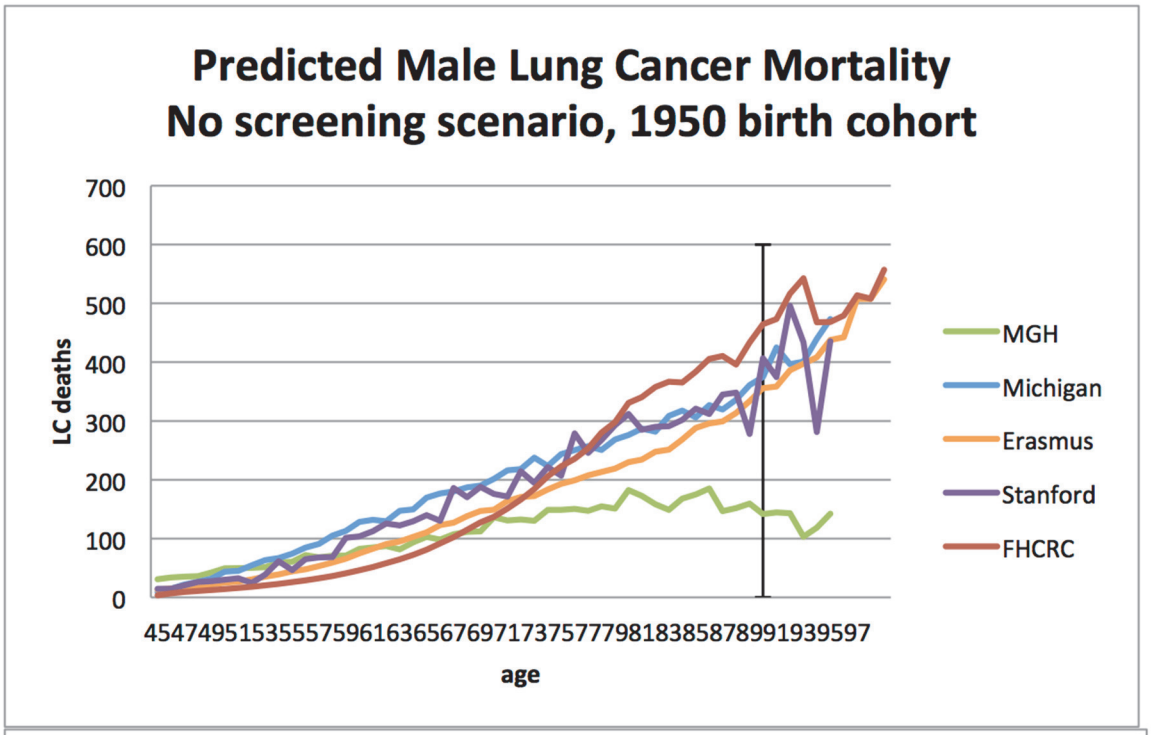


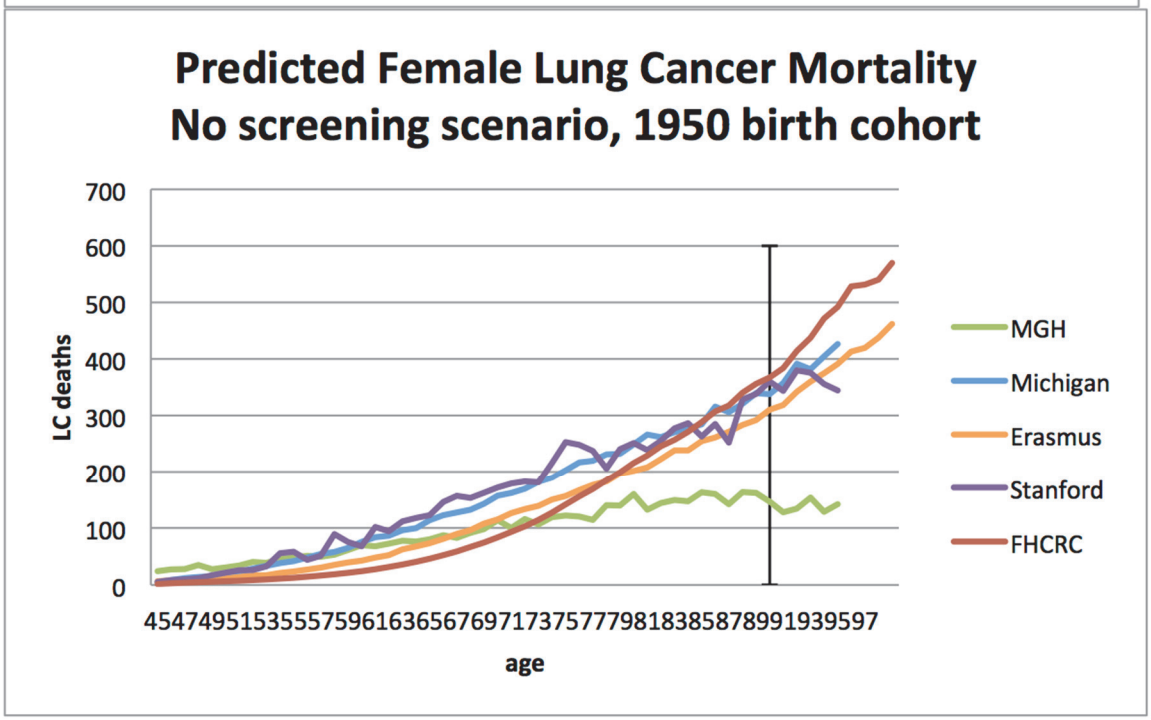
Figure S5: Models (no screening) vs. SEER9 and SEER18 incidence, and projections



For predictions past observed SEER data (over age 60) there are no observed data, but we used an age-period-cohort model to project past observed years ('Projected' red double line in plots below), which shows that the models are most divergent after age 85, when SEER data become most sparse. We cannot strictly compare incidence to that in prior birth cohorts since smoking patterns are dissimilar, and incidence varies by cohort.

Figure S6: Mortality in the no-screening scenario, as predicted by the models





The vertical line at age 90 indicates age at which all event counts (screens, deaths and deaths averted, and life-years gained) were truncated for the analyses reported here. Although the models ranked programs similarly, there was variability in the total numbers of predicted lung cancer cases, deaths, and therefore lung cancer deaths prevented. The differences in rates in the no screening scenario in large part explains the predicted differences between models. The four models (E, F, S, and U) which use two-stage or multi-stage clonal expansion models have more similarly shaped curves than the fifth model (M), which does not use a clonal expansion component (see Table S1).

Table S2: Complete list of 120 consensus efficient scenarios, selected as described in the methods. Measure of benefit was lung cancer deaths avoided and measure of harms was number of CT screenings. Table 3 in the main text provides average results across the models for selected scenarios

A45-80-10-20	A50-85-10-25	A55-85-10-25	A60-85-20-15	B55-80-40-10	B60-85-20-20
A45-80-10-25	A50-85-20-15	A55-85-20-15	A60-85-20-20	B55-80-40-15	B60-85-20-25
A45-80-20-20	A50-85-20-20	A55-85-20-20	A60-85-20-25	B55-80-40-20	B60-85-30-10
A45-80-20-25	A50-85-20-25	A55-85-20-25	A60-85-30-10	B55-80-40-25	B60-85-30-15
A45-80-30-25	A50-85-30-15	A55-85-30-20	A60-85-30-15	B55-85-20-20	B60-85-30-20
A45-85-10-15	A50-85-30-20	A55-85-30-25	A60-85-30-20	B55-85-20-25	B60-85-40-10
A45-85-10-20	A50-85-30-25	A55-85-40-15	A60-85-30-25	B55-85-30-20	B60-85-40-15
A45-85-10-25	A50-85-40-15	A55-85-40-20	A60-85-40-15	B55-85-30-25	B60-85-40-20
A45-85-20-15	A50-85-40-20	A55-85-40-25	A60-85-40-20	B55-85-40-10	B60-85-40-25
A45-85-20-20	A50-85-40-25	A60-80-10-20	A60-85-40-25	B55-85-40-25	T55-85-40-15
A45-85-20-25	A55-80-10-20	A60-80-10-25	B45-85-40-20	B60-80-30-15	T55-85-40-25
A45-85-30-25	A55-80-10-25	A60-80-20-20	B50-80-40-20	B60-80-30-20	T60-75-40-10
A50-80-10-15	A55-80-20-20	A60-80-20-25	B50-80-40-25	B60-80-40-10	T60-80-40-10
A50-80-10-20	A55-80-20-25	A60-80-30-20	B50-85-10-20	B60-80-40-15	T60-80-40-15
A50-80-10-25	A55-80-30-15	A60-80-30-25	B50-85-20-20	B60-80-40-20	T60-80-40-20
A50-80-20-20	A55-80-30-20	A60-80-40-25	B50-85-20-25	B60-80-40-25	T60-80-40-25
A50-80-20-25	A55-80-30-25	A60-85-10-15	B50-85-30-25	B60-85-10-20	T60-85-40-10
A50-80-30-25	A55-80-40-25	A60-85-10-20	B50-85-40-15	B60-85-10-25	T60-85-40-15
A50-85-10-15	A55-85-10-15	A60-85-10-25	B50-85-40-20	B60-85-20-10	T60-85-40-20
A50-85-10-20	A55-85-10-20	A60-85-20-10	B50-85-40-25	B60-85-20-15	T60-85-40-25

Table S3: Using life-years gained (LYG) as the benefit (versus CT screens performed on the x-axis) yielded a different set of consensus efficient programs. The starting and stopping ages are lower when LYG is maximized, as one might predict due to longer life expectancy among younger individuals. Note that radiation-related risks are not considered in these results.

Result	Benefit = LC Deaths Avoided (as in Table S2)	Benefit = LYG
Number of consensus efficient programs (DEA, 3rd decile)	120	152
Average age start	54.8 years	49.5 years
Average age stop	83.2 years	80.9 years
Average minimum pack-years	27.1	27.3
Average years since quit	19.9	17.4
Average frequency	1.5 years	1.7 years

Figure S7: Systematic variation of reference screening program A-55-75-30-15 across all models

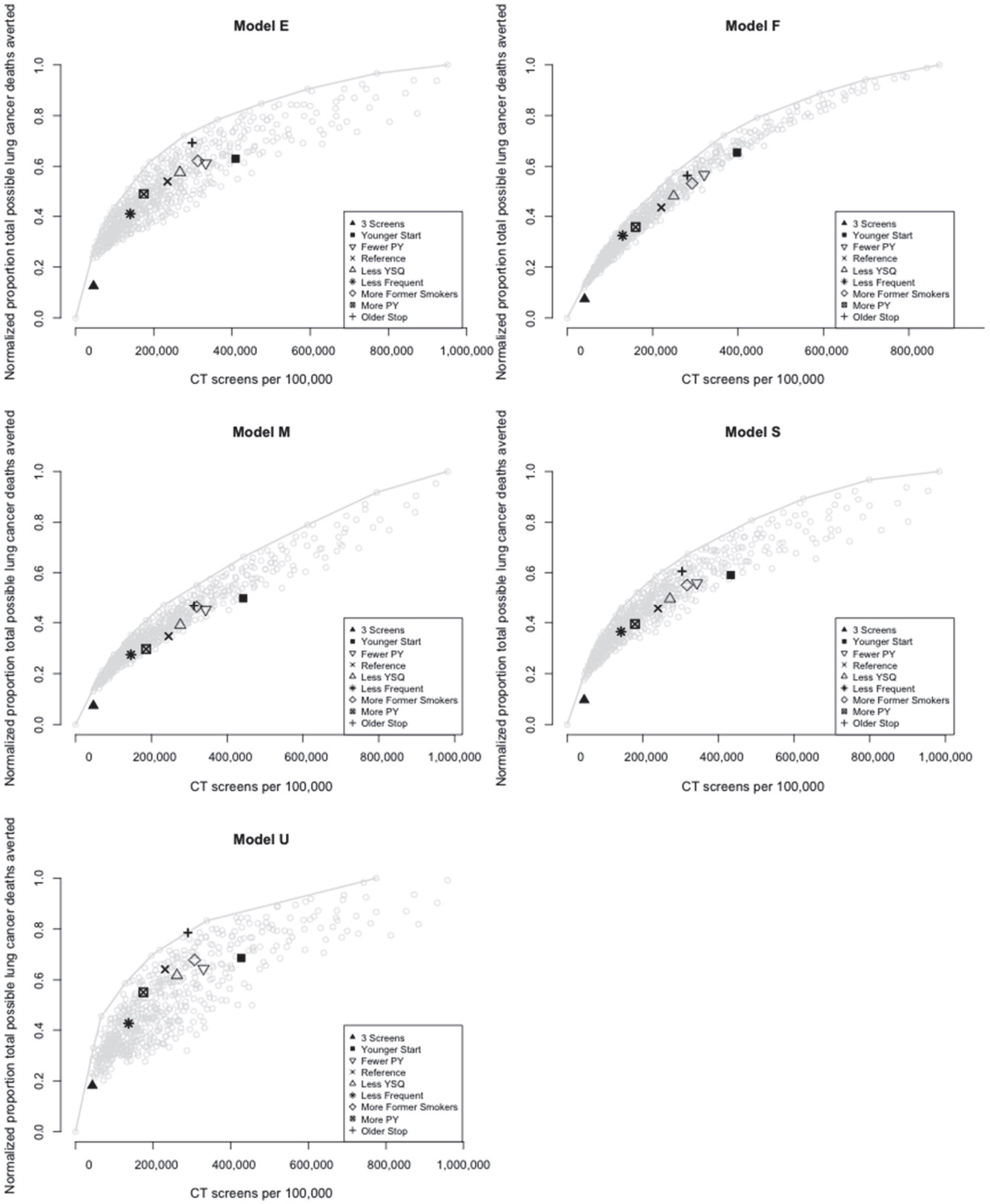


Figure S8: Consensus programs across all models

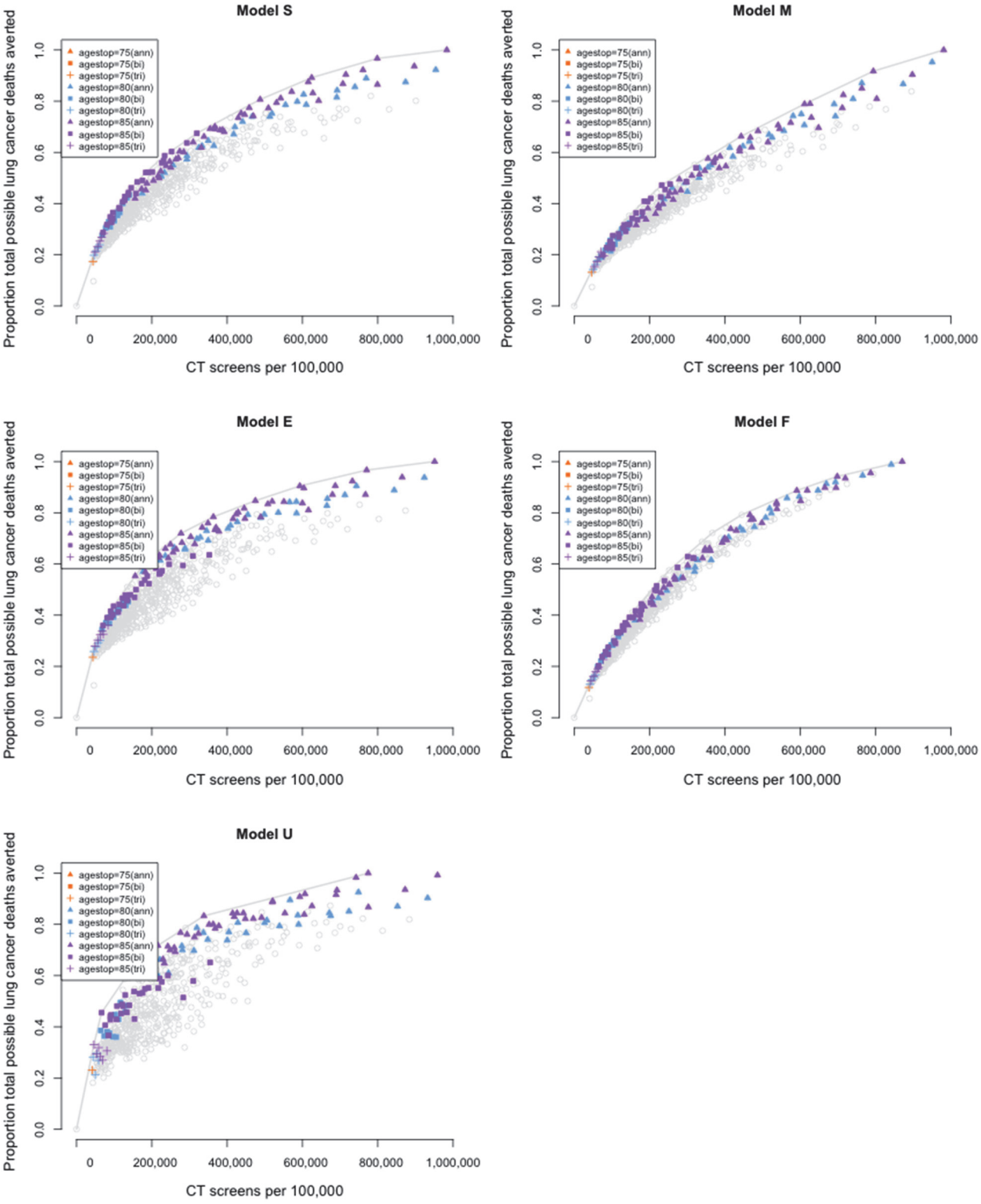
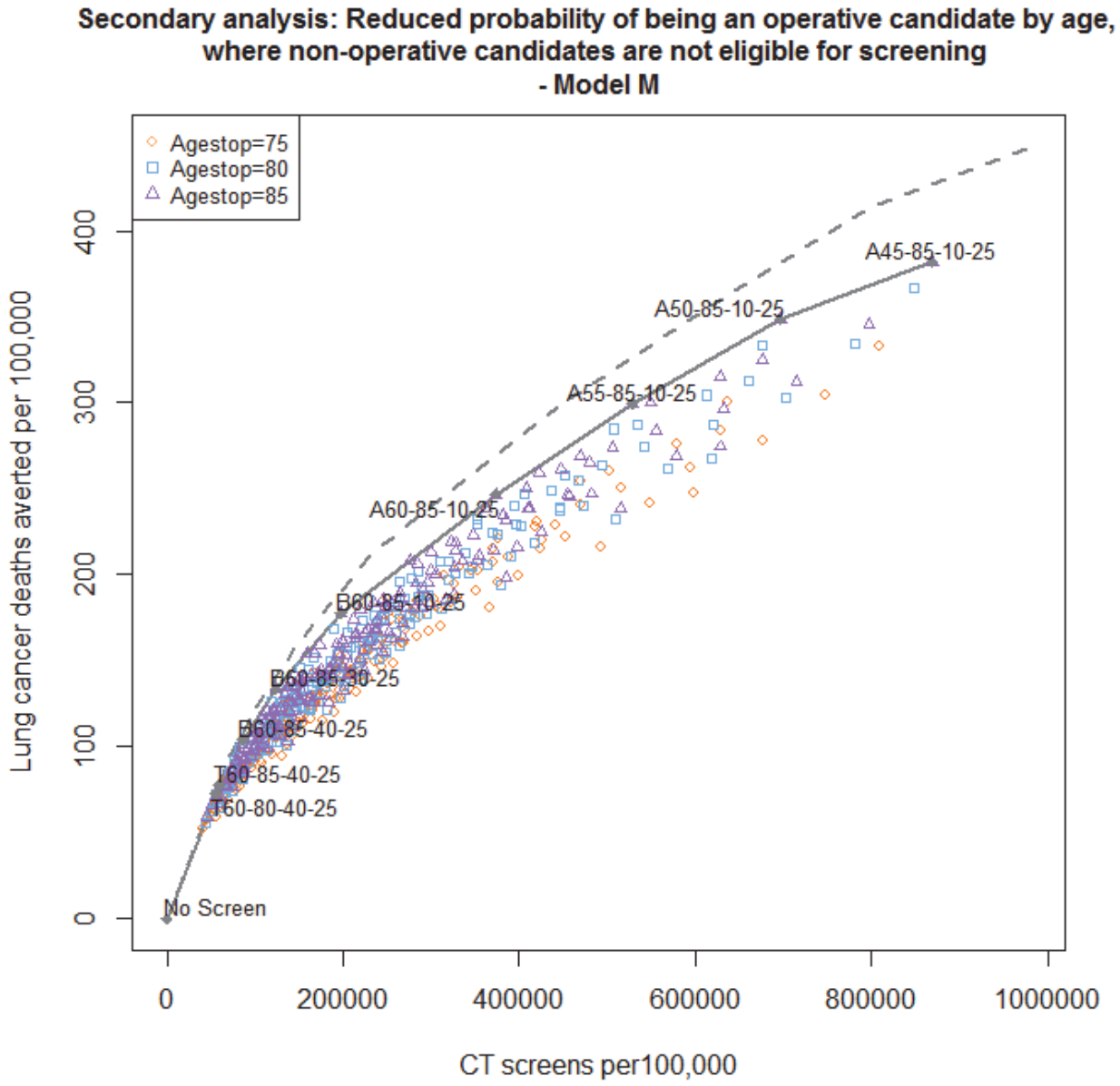


Figure S9: Effect of reducing the operative candidate probability by age



The dashed line denotes the efficiency frontier in the main analysis.

References

1. Hazelton WD, Jeon J, Meza R, Moolgavkar SH. Chapter 8: The FHCRC Lung Cancer Model. *Risk Analysis* 2012; **32**: S99-S116.
2. Schultz FW, Boer R, de Koning HJ. Chapter 7: Description of MISCAN-Lung, the Erasmus MC Lung Cancer Microsimulation Model for Evaluating Cancer Control Interventions. *Risk Analysis* 2012; **32**: S85-S98.
3. McMahon PM, Kong CY, Johnson BE, et al. Chapter 9: The MGH-HMS lung cancer policy model: tobacco control versus screening. *Risk Analysis* 2012; **32 Suppl 1**: S117-24.
4. McMahon PM, Hazelton WD, Kimmel M, Clarke LD. Chapter 13: CISNET lung models: comparison of model assumptions and model structures. *Risk Analysis* 2012; **32 Suppl 1**: S166-78.
5. McMahon PM, Kong CY, Johnson BE, et al. Estimating long-term effectiveness of lung cancer screening in the Mayo CT screening study. *Radiology* 2008; **248**(1): 278-87.
6. Pinsky PF, Gierada DS, Nath PH, Kazerooni E, Amorosa J. National lung screening trial: variability in nodule detection rates in chest CT studies. *Radiology* 2013; **268**(3): 865-73.
7. Mery CM, Pappas AN, Bueno R, et al. Similar long-term survival of elderly patients with non-small cell lung cancer treated with lobectomy or wedge resection within the surveillance, epidemiology, and end results database. *Chest* 2005; **128**(1): 237-45.
8. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
9. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.
10. Hazelton WD, Goodman G, Rom WN, et al. Longitudinal multistage model for lung cancer incidence, mortality, and CT detected indolent and aggressive cancers. *Mathematical Biosciences* 2012; **240**(1): 20-34.
11. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of Reduced Tobacco Smoking on Lung Cancer Mortality in the United States During 1975–2000. *Journal of the National Cancer Institute* 2012; **104**(7): 541-8.
12. Draisma G, Etzioni R, Tsodikov A, et al. Lead-Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. *Journal of the National Cancer Institute* 2009; **101**(6): 374-83.
13. Heijnsdijk EAM, Wever EM, Auvinen A, et al. Quality-of-Life Effects of Prostate-Specific Antigen Screening. *New England Journal of Medicine* 2012; **367**(7): 595-605.
14. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2009; **151**(10): 738-47.
15. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: A decision analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 2008; **149**(9): 659-69.
16. Tramontano AC, Cipriano LE, Kong CY, et al. Microsimulation model predicts survival benefit of radiofrequency ablation and stereotactic body radiotherapy versus radiotherapy for treating inoperable stage I non-small cell lung cancer. *American Journal of Roentgenology* 2013; **200**(5): 1020-7.
17. McMahon PM, Kong CY, Bouzan C, et al. Cost-Effectiveness of Computed Tomography Screening for Lung Cancer in the United States. *Journal of Thoracic Oncology* 2011; **6**(11): 1841-8.
18. Lin RS, Plevritis SK. Comparing the benefits of screening for breast cancer and lung cancer using a novel natural history model. *Cancer Causes & Control* 2012; **23**(1): 175-85.
19. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *European Journal of Operational Research* 1978; **2**(6): 429-44.
20. Oh D, Suh D. nonpareff: Nonparametric Methods for Measuring Efficiency and Productivity. Version 0.5-8, 2013.
21. Harris JE. Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80. *Journal of the National Cancer Institute* 1983; **71**(3): 473-9.

22. Burns D, Lee L, Shen L, et al. Chapter 2. Cigarette Smoking Behavior in the United States. Washington, DC: National Institutes of Health; 1997. Report No.: NIH Publication No.97-4213.
23. Anderson CM, Burns DM, Dodd KW, Feuer EJ. Chapter 2: Birth-Cohort-Specific Estimates of Smoking Behaviors for the U.S. Population. *Risk Analysis* 2012; **32**: S14-S24.
24. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort–Specific Smoking Histories, 1965–2009. *American Journal of Preventive Medicine* 2014; **46**(2): e31-e7.
25. Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics* 1983; **39**(2): 311-24.
26. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.

Chapter 5

Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force

Harry J. de Koning, Rafael Meza, Sylvia K. Plevritis, Kevin ten Haaf, Vidit N. Munshi, Jihyoun Jeon, Saadet Ayca Erdogan, Chung Yin Kong, Summer S. Han, Joost van Rosmalen, Sung Eun Choi, Paul F. Pinsky, Amy Berrington de Gonzalez, Christine D. Berg, William C. Black, Martin C. Tammemägi, PhD, William D. Hazelton, Eric J. Feuer*, Pamela M. McMahon*

*** Drs. Feuer and McMahon contributed equally to this manuscript.**

Published as:

de Koning HJ, Meza R, Plevritis SK, et al.

Annals of Internal Medicine 2014; **160**(5): 311-20

Abstract

Background: The optimum screening policy for lung cancer is unknown.

Objective: To identify efficient computed tomography (CT) screening scenarios in which relatively more lung cancer deaths are averted for fewer CT screening examinations.

Design: Comparative modeling study using 5 independent models.

Data Sources: The National Lung Screening Trial; the Prostate, Lung, Colorectal, and Ovarian Cancer Screening trial; the Surveillance, Epidemiology, and End Results program; and the U.S. Smoking History Generator.

Target Population: U.S. cohort born in 1950.

Time Horizon: Cohort followed from ages 45 to 90 years.

Perspective: Societal.

Intervention: 576 scenarios with varying eligibility criteria (age, pack-years of smoking, years since quitting) and screening intervals.

Outcome Measures: Benefits included lung cancer deaths averted or life-years gained. Harms included CT examinations, false-positive results (including those obtained from biopsy/surgery), overdiagnosed cases, and radiation-related deaths.

Results of Best-Case Scenario: The most advantageous strategy was annual screening from ages 55 through 80 years for ever-smokers with a smoking history of at least 30 pack-years and ex-smokers with less than 15 years since quitting. It would lead to 50% (model ranges, 45% to 54%) of cases of cancer being detected at an early stage (stage I/II), 575 screening examinations per lung cancer death averted, a 14% (range, 8.2% to 23.5%) reduction in lung cancer mortality, 497 lung cancer deaths averted, and 5,250 life-years gained per the 100,000 member cohort. Harms would include 67,550 false-positive test results, 910 biopsies or surgeries for benign lesions, and 190 overdiagnosed cases of cancer (3.7% of all cases of lung cancer [model ranges, 1.4% to 8.3%]).

Results of Sensitivity Analysis: The number of cancer deaths averted for the scenario varied across models between 177 and 862; the number of overdiagnosed cases of cancer varied between 72 and 426.

Limitations: Scenarios assumed 100% screening adherence. Data derived from trials with short duration were extrapolated to lifetime follow-up.

Conclusion: Annual CT screening for lung cancer has a favorable benefit–harm ratio for individuals aged 55 through 80 years with 30 or more pack-years' exposure to smoking.

Primary Funding Source: National Cancer Institute

Introduction

The burden of lung cancer in the world remains extremely high: The International Agency for Research on Cancer estimated 1.6 million new diagnoses in 2008 (12.7% of total cases of cancer) and 1.4 million death (18.2% of total cancer mortality).[1] In the United States and Canada, incidence (per 100,000) is 48.5 for men and 35.8 for women; mortality (per 100,000) is 37.9 and 24.2, respectively; and cumulative risk (to age 74 years) of dying of lung cancer is 3% in women and 4.6% in men. In the United States, 228,000 new cases of lung cancer and about 160,000 deaths are estimated for 2013.[2] Despite substantial reductions in smoking prevalence in the United States, which translated into an approximately 32% reduction in lung cancer mortality between 1975 and 2000 at the population level, lung cancer remains the leading cause of cancer death.[3]

Recently, the National Lung Screening Trial (NLST) demonstrated that in a volunteer population of current and former smokers who were aged 55 to 74 years at entry, had at least 30 pack-years of cigarette smoking history, and had quit no more than 15 years previously (for former smokers), 3 annual computed tomography (CT) screening examinations reduced lung cancer–specific mortality by 20% relative to 3 annual chest radiography screening examinations at a median follow-up of 6.5 years.[4] This trial did not directly address the effects of additional rounds of screening, long-term benefits or harms, or multiple alternative screening policies with different screening intervals and different eligibility criteria. Moreover, long-term outcomes must be quantified to understand the trade-offs between benefits and potential harms involved with alternative screening strategies.[5] In this study, we estimate future harms and benefits of lung cancer screening and identify a set of possible efficient lung cancer screening policies by using 5 separately developed microsimulation models calibrated to the two largest randomized, controlled trials on lung cancer screening. This work was initiated by the U.S. Preventive Services Task Force (USPSTF) to inform its recommendations on lung cancer screening.

Methods

Calibration of 5 Models to de-identified Lung Cancer Screening Data From the NLST and Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial

We used 5 models calibrated to individual-level, de-identified data from the NLST and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening trial (PLCO).[6,7] The NLST enrolled 53,452 persons at high risk for lung cancer at 33 U.S. centers from August 2002 through April 2004. Participants were randomly assigned to undergo 3 annual screening examinations with low-dose CT (26,722 participants) or single-view posterior–anterior chest radiography (26,730 participants). The PLCO trial randomly assigned 154,901 participants aged 55 through 74 years at entry, 77,445 of whom were assigned to annual chest radiography and 77,456 to usual care between November 1993 and July 2001. There was no eligibility requirement concerning smoking. Although the PLCO trial compared chest radiography with no screening, it provided information on the natural history of lung cancer.

Groups of investigators at the following 5 institutions independently developed the models: Erasmus Medical Center in Rotterdam, the Netherlands (model E); Fred Hutchinson Cancer Research Center in Seattle, Washington (model F); the Massachusetts General Hospital in Boston, Massachusetts (model M); Stanford University in Stanford, California (model S); and the University of Michigan in Ann Arbor, Michigan (model U). Each model estimates screening effectiveness on the basis of a different set of assumptions that are key in predicting the effects of earlier treatment, and each model uses different mathematical formalisms and model structures. In essence, all the models account for the individual's age-specific smoking-related risk for lung cancer, date and stage of lung cancer diagnosis, the corresponding lung cancer mortality, and the individual's life expectancy in the presence and absence of screening (described in Figure S1 and Table S1 in the supplementary material of this Chapter).

For correct extrapolation of different possible screening scenarios, one must obtain the best estimates on several key parameters, including the duration of the screening-detectable preclinical period (by test, age, histologic characteristics, and sex), sensitivity (by test, age, and sex), and improvement in prognosis by earlier detection and treatment. All models were

first set to mimic the design of both trials (for example, setting the numbers of screening examinations and screening method, ages at screening, smoking history and sex of enrollees, and screening intervals). The models are validly calibrated when the key parameters—which may differ by model—can be estimated or adjusted to replicate the trial data closely. After calibration, the models reproduce the observed cumulative incidence of lung cancer (by stage, histologic characteristics, sex, age, type of detection, and round of screening) and lung cancer mortality in both groups of the trials. Close calibration to the 19% (95% CI, 7% to 25%) lung cancer mortality difference between groups of the NLST at 6 years of follow-up was prioritized, shown in Figure S2 in the supplementary material of this Chapter. Key similarities and differences among the 5 models in calibration targets are described in Table S1 of the supplementary material of this Chapter.

Choosing Screening Programs and Expressing Harms and Benefits

The modeling groups standardized input data on smoking histories and non–lung cancer mortality to simulate life histories of the U.S. cohort born in 1950 by using an updated version of the National Cancer Institute's Smoking History Generator.[8-11] All models included other-cause mortality to differ by sex, age, smoking status, and smoking intensity. A set of 576 programs that varied frequency of CT screening for lung cancer (1-, 2-, or 3-year intervals), ages of starting (45, 50, 55, or 60 years) and stopping (75, 80, or 85 years) screening (assuming that a last screening examination is included at this age), and eligibility based on smoking history (10, 20, 30, or 40 pack-years; having quit smoking 10, 15, 20, or 25 years previously) was examined, with analyses run separately for men and women. In all scenarios, perfect screening adherence was assumed. Once a person's characteristics did not satisfy the eligibility criteria (such as passing the limit of years since smoking cessation), he or she would not be invited for future screenings.

Potential benefits are expressed as lung cancer deaths averted and life-years gained. Potential harms are expressed as the number of screening examinations plus follow-up imaging examinations, number of false-positive results (including findings on surgery and biopsy), number of overdiagnosed lung cancer cases, and number of radiation-related lung cancer deaths. Follow-up procedures were assumed to be consistent with the observed rate of examinations per positive screening examination in the NLST; 2 models used explicit

follow-up algorithms based on nodule size thresholds. False-positive results were estimated as a direct proportion to the number of CT screening examinations, as based on the average in 3 rounds of the NLST; we assumed that a false-positive result in a given round did not influence the probability of a false-positive result in subsequent rounds. Overdiagnosed cases are the additional number of lung cancer cases detected in the screening scenarios compared with the estimated number of cases diagnosed in the absence of screening.[12] All models simulate the underlying natural history of lung cancer (separately by histologic type) in individuals and include dose–response modules that relate a detailed cigarette smoking history over time to lung cancer risk. Each comparison is based on an identical underlying simulated cohort of individuals with the same smoking histories, sex composition, and potential times of other-cause death. Another scenario that reflects overdiagnosis is a person who has lung cancer that is expected to be clinically detected after death from other causes but whose cancer in the screening scenario is detected before death from other causes. For all measures of benefits and harms, expressed per 100,000 persons, a cohort of persons born in 1950 was followed from ages 45 to 90 years.

We identified “efficient” scenarios as those that prevented the most lung cancer deaths for the same number of CT screening examinations (not including follow-up scans). Model results were compared by using the data envelopment analysis method, which is an engineering-based approach for selecting efficient scenarios from among a collection of alternatives.[13] In simple terms, it finds programs that are near the efficient frontier, with consideration given to whether one is prioritizing maximizing benefits (that is, deaths averted [y-axis]) or minimizing harms (that is, CT screening examinations [x-axis]). For each model’s results, we generated a rank score (decile of distance from the model’s frontier) for each scenario not on the frontier. We identified scenarios on (score 0) or closest to (first 3 deciles) the frontier of at least 3 models. Two models (F and M) were used to estimate radiation-related lung cancer cases (described in the supplementary material of this Chapter, in the section: “Estimating Radiation-Related Lung Cancer Cases”). All results were averaged across all 5 models. Finally, an advantageous scenario was selected that was efficient and led to a substantial reduction in lung cancer mortality and life-years gained at reasonable harms. The number of screening examinations needed was similar to (a

continuous) NLST scenario and to breast and colorectal cancer screening guidelines, given manpower and resources.

Results

Benefits and Harms of Efficient Scenarios

Of the 576 possible programs, 120 were on or close to the efficient frontier, where no alternative that provides more lung cancer deaths averted for fewer CT screening examinations exists. Table 1 shows the benefits of 26 top-ranked triennial, biennial, and annual scenarios, as well as benefits of a 27th program that was most similar to the NLST strategy but was not among the consensus efficient programs. None of the top-ranked scenarios had a starting age of 45 years. For the top-ranked triennial and biennial efficient programs, the starting age was 60 years and the minimum number of pack-years was 40 (with 1 exception). Triennial screening programs led to limited reductions in lung cancer mortality: from 4.6% to 6% in this cohort (range, 1.7% to 9.5% across models). Biennial programs led to 6.5% to 9.6% reductions in lung cancer mortality (range, 2.3% to 14.8% across models). When we compared the least intensive program (60-80-40-10, with values arranged per the following order: start age–stop age–minimum pack-years–maximum years since quitting smoking) of triennial to biennial screening, the additional percentage of lung cancer deaths averted was about 40%, at the expense of about 50% additional screening examinations (described in Figure S3 in the supplementary material of this Chapter). Annual screening scenarios provided substantially more benefit, leading to 11.0% to 21.2% reductions in lung cancer mortality (range, 4.3% to 39.1% across models). In these scenarios, 48.1% to 56.9% of lung cancer cases were detected at stage I/II, compared with 37.4% without screening. The scenario most similar to the NLST criteria (A [annual]-55-75-30-15) led to fewer lung cancer deaths averted but more screenings compared with the next-most-intensive program (A-60 80-30-25).

Table 2 summarizes the most important harms associated with the scenarios. The number of follow-up imaging procedures and false-positive results increased proportionally to the number of CT screening examinations needed in each scenario, leading to 1.0 to 4.9 false-positive results per person screened. Decreasing the minimum pack-years eligibility criteria

from 30 to 20 pack-years and to 10 pack-years in annual scenarios provided a relatively small increase in lung cancer deaths averted versus the large number of additional CT scans. Although these are still efficient scenarios, they require substantially more CT screening examinations (both overall and per person) and follow-up procedures, and false-positive results increase proportionally. Overdiagnosis ranged from 1.5% to 6.6% of all lung cancer cases, or 8.7% to 13.5% of screening-detected lung cancer cases.

Table 1: Benefits of 26 Selected Efficient Screening Programs and a Screening Program Most Similar to Eligibility Criteria for the National Lung Screening Trial*

Frequency–Start Age (y)–Stop Age (y)–Pack-Years–Years Since Quitting	Cohort Eligible, %	CT Screening Examinations, n	Screening-Detected Cases, n	Total Cases Detected at an Early Stage, %†	Reduction in Lung Cancer Mortality, %	Lung Cancer Deaths Averted, n‡	Life-Years Gained	Life-Years Gained per Lung Cancer Death Averted	Screening Examinations per Life-Year Gained, n	Screening Examinations per Lung Cancer Death Averted, n
T-60-80-40-10	11.2	45,685	787	42.0	4.6	172	1,823	10.6	25	265
T-60-85-40-10	11.3	48,317	943	42.6	5.1	190	1,894	10.0	26	254
T-60-85-40-15	12.0	55,316	1,043	43.3	5.4	201	2,000	10.0	28	275
T-60-85-40-25	13.0	66,333	1,139	44.1	6.0	225	2,252	10.0	29	294
B-60-80-40-10	11.2	67,167	1,072	44.0	6.5	241	2,526	10.5	27	278
B-60-85-40-10	11.3	69,662	1,181	44.3	6.9	256	2,665	10.4	26	272
B-60-85-40-15	12.0	79,757	1,279	45.3	7.4	275	2,882	10.5	28	290
B-60-80-40-25	13.0	90,337	1,279	45.5	7.7	286	3,017	10.6	30	315
B-60-85-40-25	13.0	95,914	1,536	46.3	8.4	312	3,045	9.8	32	307
B-60-85-30-20	17.9	127,046	1,744	47.5	9.6	358	3,451	9.6	37	354
A-60-80-40-25§	13.0	171,924	1,664	48.1	11.0	410	4,211	10.3	41	419
A-60-85-40-25	13.0	185,451	1,911	49.4	12.1	449	4,203	9.4	44	413
A-55-85-40-20	14.0	220,505	1,967	50.0	13.0	485	4,811	9.9	46	454
A-55-80-40-25§	13.9	221,606	1,782	49.2	12.3	458	4,777	10.4	46	483
A-60-80-30-25§	18.8	253,095	1,983	50.4	13.3	495	4,940	10.0	51	511
A-55-75-30-15 ‖	19.2	265,049	1,646	48.4	12.3	459	5,375	11.7	49	577
A-60-85-30-25	18.8	271,152	2,263	52.1	14.7	547	5,322	9.7	51	495
A-50-85-40-25	14.6	281,218	2,159	51.4	14.6	542	5,908	10.9	48	518
A-55-80-30-15§	19.3	286,813	1,971	50.5	14.0	521	5,517	10.6	52	550
A-60-80-20-25§	24.8	372,024	2,419	51.9	15.4	573	5,707	10.0	57	570
A-55-80-30-25§	20.4	342,880	2,288	52.1	15.8	588	6,321	10.8	54	583
A-60-85-20-25	24.8	348,894	2,779	53.7	16.8	624	5,934	9.5	59	559
A-55-80-20-25§	27.4	455,381	2,543	53.9	17.9	664	7,092	10.7	64	685
A-55-85-20-25	27.4	477,334	2,955	55.6	19.1	712	7,490	10.5	64	670
A-55-80-10-25§	36.0	561,744	2,803	55.2	19.4	721	7,693	10.7	73	777
A-50-80-20-25	29.0	588,516	2,732	55.2	20.0	743	8,530	11.5	69	792
A-50-85-20-25	29.0	610,443	3,153	56.9	21.2	787	8,948	11.4	68	775

Abbreviations: A = annual; B = biennial; CT = computed tomography; T = triennial.

** Numbers are per a 100,000-person cohort followed from ages 45 to 90 years and are based on averaged estimates across the 5 models. The screening programs are labeled as follows: Frequency–age start–age stop–minimum pack-years–maximum years since quitting. Note that these mortality reductions are different from the observed point estimate in the National Lung Screening Trial because in our cohort analysis only eligible persons are screened (dilution) and it is a lifetime reduction compared with 6-years' follow-up in the trial.*

† Average percentage of cases detected at an early (I/II) stage in the no-screening scenario was 37.4%. Incident number of cases in the no-screening scenario was 5,119 per 100,000 persons.

‡ Average number of lung cancer deaths in the no-screening scenario was 3,719 (per 100,000-person cohort).

§ Seven programs are the consensus-efficient, annual programs (minimum start age 55) with a stop age of 80 years and screening counts between 200,000 and 600,000 plus an eighth program (A-60-80-40-25) with just under 200,000 screening examinations included as a reference program. Note that in this table the columns that include lung cancer deaths averted do not include radiation-related lung cancer deaths (these are presented in Table 2).

|| Denotes eligibility most similar to that in the National Lung Screening Trial.

Table 2: Harms of 26 Selected Efficient Screening Programs and a Screening Program Most Similar to Eligibility Criteria for the National Lung Screening Trial*

Frequency–Start Age (y)–Stop Age (y)–Pack-Years–Years Since Quitting	CT Screening Examinations, n	Total CT Examinations Including Screening, n	Average Screening Examinations per Person Screened, n	Average False-Positive Results per Person Screened, n	Overdiagnosed Cases, n	Overdiagnosis, % of all cases†	Overdiagnosis, % of screening-detected cases	Radiation-Related Lung Cancer Deaths, n‡
T-60-80-40-10	45,685	55,696	4.1	1.0	79	1.5	10.1	9
T-60-85-40-10	48,317	58,677	4.3	1.0	98	1.9	10.5	10
T-60-85-40-15	55,316	66,677	4.6	1.1	119	2.3	11.6	10
T-60-85-40-25	66,333	79,267	5.1	1.2	147	2.8	13.1	11
B-60-80-40-10	67,167	80,068	6.0	1.4	116	2.2	10.9	11
B-60-85-40-10	69,662	82,874	6.2	1.4	129	2.5	11.0	11
B-60-85-40-15	79,757	94,383	6.7	1.6	156	3.0	12.4	12
B-60-80-40-25	90,337	106,512	7.0	1.6	151	2.9	12.0	13
B-60-85-40-25	95,914	112,810	7.4	1.7	184	3.5	12.2	13
B-60-85-30-20	127,046	148,518	7.1	1.7	197	3.8	11.5	16
A-60-80-40-25§	171,924	199,035	13.3	3.1	183	3.5	11.2	17
A-60-85-40-25	185,451	214,351	14.3	3.3	241	4.6	12.9	17
A-55-85-40-20	220,505	254,083	15.8	3.7	224	4.3	11.6	19
A-55-80-40-25§	221,606	255,398	15.9	3.7	194	3.7	11.1	20
A-60-80-30-25§	253,095	291,667	13.5	3.1	231	4.4	11.9	21
A-55-75-30-15‖	265,049	305,181	13.8	3.2	141	2.7	8.7	24
A-60-85-30-25	271,152	312,130	14.4	3.4	296	5.6	13.5	20
A-50-85-40-25	281,218	323,024	19.3	4.5	243	4.6	11.5	22
A-55-80-30-15§	286,813	329,809	14.9	3.5	190	3.7	9.9	24
A-60-80-20-25§	372,024	376,098	13.2	3.1	232	4.4	9.8	25
A-55-80-30-25§	342,880	393,611	16.9	3.9	224	4.3	10.0	25
A-60-85-20-25	348,894	400,898	14.1	3.3	328	6.2	12.2	23
A-55-80-20-25§	455,381	521,943	16.6	3.9	258	4.9	10.4	31
A-55-85-20-25	477,334	546,838	17.4	4.1	348	6.6	12.2	30
A-55-80-10-25§	561,744	643,001	15.6	3.6	259	4.9	9.5	35
A-50-80-20-25	588,516	673,103	20.3	4.7	256	4.9	9.6	38
A-50-85-20-25	610,443	697,962	21.1	4.9	344	6.5	11.3	37

Abbreviations: A = annual; B = biennial; CT = computed tomography; T = triennial.

** Numbers are per a 100 000-person cohort followed from ages 45 to 90 years and are based on averaged estimates across the 5 models. The screening programs are labeled as follows: Frequency–age start–age stop–minimum pack-years–maximum years since quitting. Overdiagnosed cases are slightly overestimated because all counts are up to age 90 years (some cases detected early will appear clinically after age 90 years).*

† Incident number of cases in the no-screening scenario was 5,119 per 100,000 persons; the number of lung cancer deaths was 3,719 per 100,000 persons.

‡ Average of two models.

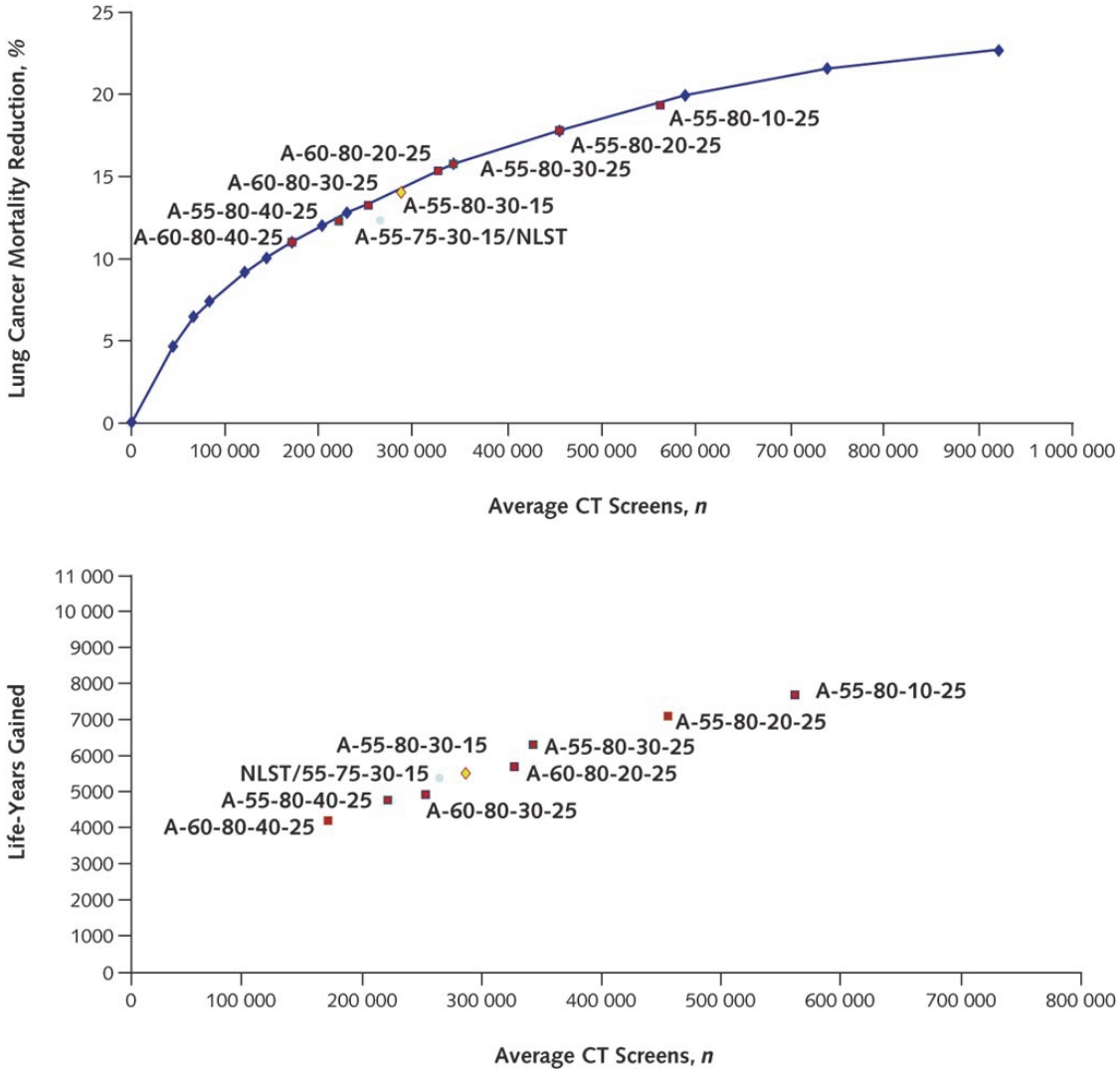
§ Seven programs are the consensus-efficient, annual programs (minimum start age 55) with a stop age of 80 years and screening counts between 200,000 and 600,000 plus an eighth program (A-60-80-40-25) with just under 200,000 screening examinations included as a reference program. Note that in this table the columns that include lung cancer deaths averted do not include radiation-related lung cancer deaths (these are presented in Table 2).

|| Denotes eligibility most similar to that in the National Lung Screening Trial.

Lung Cancer Deaths Averted and Life-Years Gained

Clinical concerns about the potential for increased operative mortality in older individuals with a history of heavy smoking, as well as increased comorbidity and reduced eligibility for surgery with curative intent at higher age limits, led us to focus on scenarios with stopping ages of 80 years or younger. Figure 1 shows the effect of expanding the smoking eligibility in the age range of 55 through 80 years, beyond the criteria similar to those used in NLST: for example, to 25 years since quitting (A-55-80-30-25) or 20 or fewer pack-years (A-55-80-20-25 or A-55-80-10-25). More lung cancer deaths may be averted with more CT screening examinations, but there are diminishing returns (although not at a single distinct point). The A-60-80-20-25 scenario, which extends eligibility to individuals with fewer pack-years but starting at a later age, was still efficient with respect to number of screening examinations and lung cancer deaths averted, but it provided relatively fewer life-years gained than did A-55-80-30-15. For the 3 consecutive scenarios of A-55-80-30-15, A-60- 80-20-25, and A-55-80 30-25, the number of screening examinations per lung cancer death averted increased progressively (550, 570, and 583 examinations), whereas the number of screening examinations per life-year gained was the highest (that is, the worst) for A-60-80-20-25 (52, 57, and 54 examinations). The A-60-80-20-25 scenario also resulted in the highest number and percentage of overdiagnosed cases.

Figure 1: Estimated lung cancer mortality reduction (as percentage of total lung cancer mortality in cohort) and life-years gained (averages of 5 models) from annual CT screening, for programs with minimum eligibility age of 55 years and maximum of 80 years at different smoking eligibility cutoffs and NLST scenario (A-55-75-30-15)



Abbreviations: A = annual; CT = computed tomography; NLST = National Lung Screening Trial.

The average number of CT screening examinations (5 models) is shown on x-axis. The graph plots the average number of CT screening examinations against the percentage reduction of lung cancer mortality (top) or life-years gained (bottom) for each screening scenario (versus no screening) that was estimated for 100,000 individuals of the 1950 cohort followed from ages 45 to 90 years. Programs are labeled as follows: frequency–start age–stop age–minimum pack-years–maximum years since quitting smoking. The reductions in lung cancer mortality differ from the point estimate of the reduction reported at the 6.5-year follow-up in the NLST because only eligible persons are screened in this cohort analysis (dilution) and lifetime reduction in lung cancer mortality is modeled. The top panel shows the efficiency frontier for all models combined. When the slope in the efficiency frontier plot levels off, the additional reductions in mortality per unit increase in use of CT screening examinations are small relative to the previous strategies.

Advantageous Scenario

Of the efficient scenarios, annual screening in the age range of 55 through 80 years had substantial benefits while maintaining a moderate level of harms. We judged a strategy that was similar to the NLST criteria—starting screening at age 55 years, but ending through age 80 years for ever-smokers with a smoking history of at least 30 pack-years, and no more than 15 years since quitting for former smokers (A-55-80-30-15)—as the advantageous scenario with the optimum balance of benefits and harms. Table 3 summarizes the modeled data about harms and benefits associated with that scenario expressed per 100,000 45-year-old persons born in 1950 and followed through age 90 years.[14-19] The upper- and lower-bound estimates presented in the table are ranges found across the 5 different models and not confidence intervals (CIs). The table illustrates that 19,300 of 100,000 individuals would be eligible for screening at some point in their lifetime. Without screening, lung cancer will be diagnosed in 5,119 and 3,719 will die of the disease. Assuming 100% adherence to screening, 50.5% of cases of lung cancer will be detected at an early (I/II) stage. There will be 497 fewer lung cancer deaths, and these persons will on average gain 10.6 life-years per death averted. They will also be prevented from experiencing advanced disease and its treatment. On the negative side, 67,550 false-positive results would be expected (19,300 individuals * 3.5 average false-positive results per individual), leading to 910 surgeries or biopsies for benign disease. There would be 1,970 persons with a diagnosis of lung cancer made earlier than would have occurred if they had not been screened, and about 10% of these cancer cases would otherwise never have been diagnosed during their lifetime (190 cases).

Table 3: Number of Individuals Having Benefits and Harms of Annual CT Screening From Ages 55 Through 80 years*

	Average of Five Models	Lower-Bound Estimate	Upper-Bound Estimate
Benefits			
Persons no longer dying of lung cancer†	497	177	862
Life-years gained	5,250	2,020	10,153
Persons no longer needing treatment for advanced lung cancer	550	200	950
Life-years with advanced disease prevented‡	550	200	950
Not estimated here			
Persons receiving less intensive or mutilating primary treatment			
Possible additional effect of quitting smoking when offered together			
Harms			
Times persons undergo CT screening examination, n	287,000	272,000	301,000
False-positive test results experienced, n	67,550	61,250	70,700
Times persons undergo CT follow-up (regular dose), n	43,000	23,175	50,100
Persons receiving the diagnosis of lung cancer earlier, n	1,970	1,370	2,845
Persons undergoing surgery/biopsy for lesions that ultimately seem benign, n	910	825	955
Persons diagnosed with lung cancer who would otherwise never have had the diagnosis (overdiagnosed cases), n	190	72	426
Not estimated here			
Persons possibly falsely reassured by a negative test result (postponing future visits when noticing symptoms or signs)			
Persons possibly increasing smoking after a negative test result			

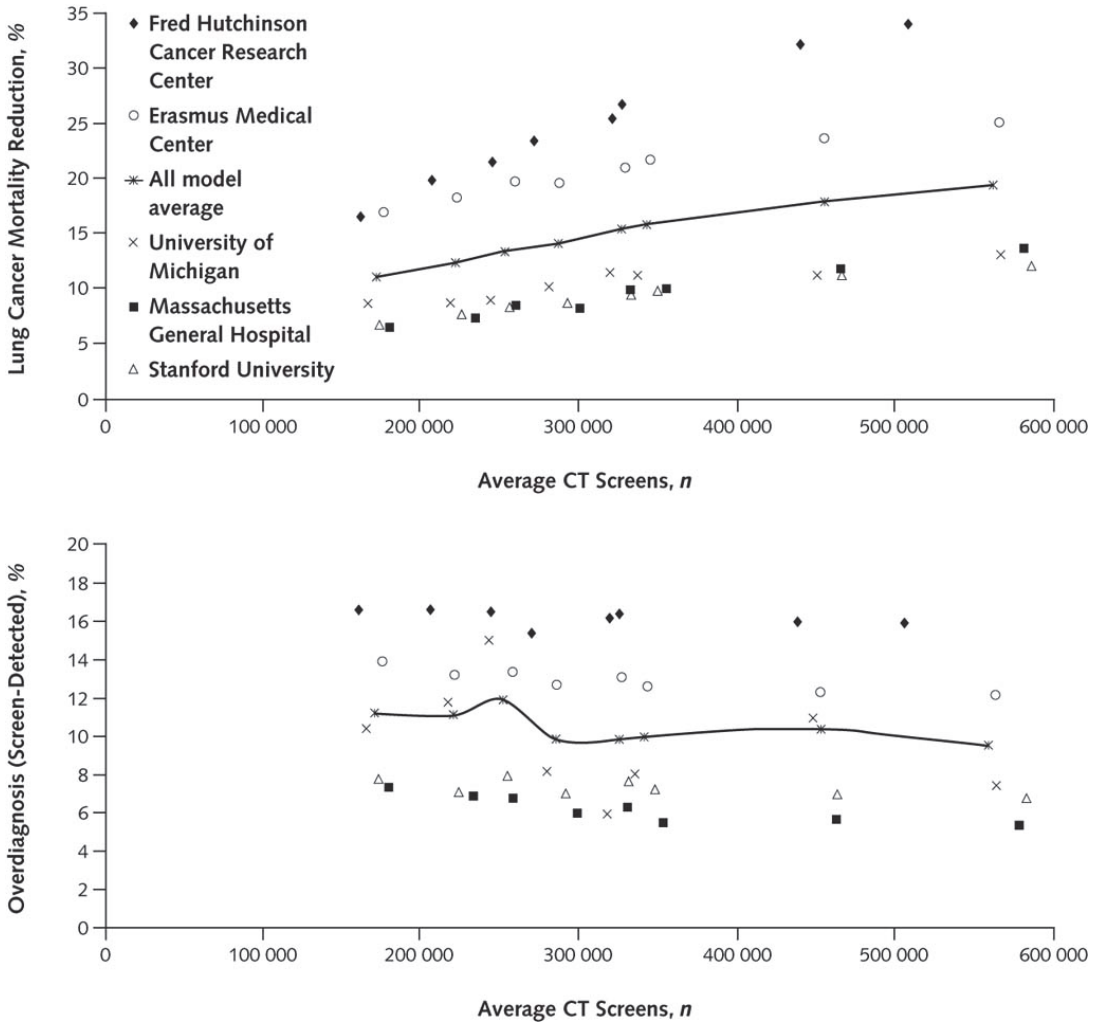
Abbreviations: CT = computed tomography.

** Current and former smokers have a minimum smoking history of 30 pack-years, and former smokers quit in the past 15 years. Numbers for 100,000 individuals from the 1950 U.S. cohort followed from ages 45 to 90 years; 19,300 persons eligible, assuming 100% compliance. A total of 5,119 lung cancer cases was diagnosed without screening, and 5,307 were diagnosed with screening; 37% were screening-detected (average of 5 models). A total of 3,719 lung cancer deaths occurred without screening. † Incorporates 24 radiation-related deaths. ‡ The number of life-years with advanced disease prevented was derived from Goldstraw et al.[14] The number of biopsies performed for ultimately benign lesions was based on the National Lung Screening Trial: There were 164 surgeries and 66 needle biopsies for benign nodules out of 17,053 false-positive test results or 75,126 CT screening examinations, making up 1.3% of false-positive test results. Differences in the range of results reflect differences in modeling approaches but should not be seen as formal 95% CIs. We did not consider that the earlier knowledge of the diagnosis of cancer has been shown to negatively affect quality of life, including adverse effects of treatment, anxiety regarding assessment, and longer hospitalizations; the possible risks of false reassurance (a false-negative screening test may lead to postponing access to care); or the possibility of a behavioral change (that is, relapsing to smoking) after the screening examination.[15-19]*

Increasing the number of pack-years to 40 but extending the years since having quit smoking from 15 to 25 in the age range of 55 to 80 years decreased the percentage eligible from 19.3% to 13.9%; however, these individuals would have an average of 1 more screening examination during their lifetime, thereby increasing harms (Table 2). The same scenario for persons with 20 pack-years (instead of 30) or more increased the number of lung cancer deaths averted by 12.9%, but this means 33% more CT screening examinations in the population, with a proportional increase in false-positive test results (Tables 1 and 2). Extending eligibility to 10 pack-years or more resulted in 8.6% more deaths averted for 23% more CT scans.

Figure 2 (top) shows the reductions in lung cancer mortality for the 8 (labeled) scenarios for each model group and the average of the 5 models. The reference scenario (A-60-80-40-25) led to a 6.5% to 17.0% decrease in lung cancer mortality, and the A-55-80-30-15 scenario resulted in an 8.6% to 23.5% reduction. The most intensive scenario (A-55-80-10-25) led to an estimated 12% to 34% decrease in lung cancer mortality. The percentage of overdiagnosed screening-detected cases varied between 5% and 17% almost uniformly across inclusion criteria.

Figure 2: Estimated percentage of reduction in lung cancer mortality and overdiagnosed cases (of screening-detected cases) for the highlighted scenarios in Tables 1 and 2 (average number of CT screening examinations is shown on the x-axis) for all individual models and the average of the 5 models



Abbreviations: CT = computed tomography.

Presentation of a 100,000-person 1950 cohort followed from ages 45 to 90 years.

Discussion

Our models show that annual lung cancer screening of individuals aged 55 to 80 years with a smoking history of at least 30 pack-years offers substantial benefits. There would be a 14% overall lung cancer mortality reduction and a 25% reduction in those eligible for screening, with relatively limited harms. Extending eligibility to individuals with fewer pack-years, although still efficient, leads to additional benefits along with relatively more harms. The models provide valuable tools to project trial results to different screening scenarios over the course of a lifetime and show the strategies that provide the greatest benefits for a

specified level of resources. The advantageous scenario for lung cancer screening compares favorably with the USPSTF guidelines for breast and colorectal cancer screening. Applying current USPSTF breast cancer screening recommendations to a similar 1950 U.S. cohort translates to about 1.1 million screening examinations (per 100,000 women) and 700 breast cancer deaths averted through use of the Erasmus model.[20] Applying current colorectal screening guidelines translates to about 227,000 screening colonoscopies and 1,910 colorectal cancer deaths averted through use of the Erasmus model.[21]

If we examine eligibility of the advantageous scenario by age in 2013 for the 1950 birth cohort, 17% of the 55- to 64-year-old age-group, 12.5% of the 65- to 74-year-old age-group, and 7% of the 75- to 80-year-old age-group would be eligible for lung cancer screening. Applying these percentages to the current U.S. population means that about 10.5 million persons in the United States would be eligible for screening and that more than 18,000 lung cancer deaths per year might be avoided. That estimate is more optimistic than the recently reported estimates under the NLST criteria of about 8.6 million persons eligible for screening and about 12,000 averted lung cancer deaths.[22]

Simulating 3 screening examinations as was done in NLST would, in our cohort-based approach, have led to a lifetime 3.7% reduction in lung cancer mortality (not shown). This is notably different from the observed mortality reduction point estimate reported at 6 years' follow-up in the NLST because only eligible persons are screened in our cohort analysis (dilution) and we modeled a lifetime reduction rather than short-term follow-up. Overdiagnosis is a general concern with screening. There are few estimates of the magnitude of overdiagnosis with CT lung cancer screening.[23] We estimated overdiagnosis with CT screening to be less than 17% of screening-detected cases (upper range in Figure 2). Although most published reports describing overdiagnosis in breast cancer screening apply to populations (that is, multiple cohorts), our average of 10% of overdiagnosed screening-detected lung cancer cases in the advantageous scenario is equal to that for breast cancer screening every 2 years in women aged 50 to 74 years in relatively low-referral programs and far less than that for breast cancer screening in high-referral countries, such as the United States.[24,25] Two groups explicitly modeled radiation risk and found the number of radiation-related lung cancer deaths to be very small, in line with earlier reports.[26]

Several limitations that affect generalizability and certainty of findings are worth noting. First, models assumed 100% adherence to screening. Second, models extrapolated benefits and harms derived from trials with short-term duration to lifetime follow-up in the U.S. population. Although the models were calibrated and are consistent with the NLST and PLCO trial, extrapolations beyond those trials' time horizons, screening intervals, and eligibility criteria introduce uncertainty. Third, 5 models, with different structures and assumptions, showed some variability in their absolute predictions of benefits and harms (Table 3), although the ranking of strategies was consistent across models (shown in Figure S4 of the supplementary material of this Chapter). Moreover, there is variance in the absolute level of lung cancer deaths averted between the models, ranging from 177 to 862, and variance in the overdiagnosis estimates, ranging from 72 to 426, for the A-55-80-30-15 scenario. Fourth, although extrapolations to the age span of 75 to 80 years seem reasonable (for example, the oldest participants in the NLST were screened until age 78 years), there are still limited observational data on screening in older individuals. We did note, however, that sicker individuals who were deemed less favorable candidates for possible surgical cure did not affect ranking of the strategies (data not shown).

Benefits were extrapolated from 1 large-scale trial in the United States with positive results, whereas 2 small fair-quality European trials have published negative interim results.[27-29] However, these are not large enough to have statistical power to show a clinically plausible effect, in contrast to the NELSON (Nederlands-Leuven Screening Onderzoek) trial, which enrolled 15,822 individuals aged 50 to 75 years and compared CT screening with no screening.[30] Preliminary analyses showed that the percentage of lung cancer detected early is more favorable than in the NLST.[31-33] Mortality results are still pending. The NELSON trial has primarily used volume-doubling times and volume measurements of lung nodules to define its referral strategy, thereby substantially reducing the number of positive and false-positive results: About 60% of referrals were for false-positive results, and the percentage of referrals was about 2%.[32] It may therefore be feasible to reduce one of the important harms of lung cancer screening via changes in follow-up guidelines. The criteria we simulated in our scenarios may not be ideal in clinical practice. Number of pack-years is a known moderate surrogate measure of risk; its use for inviting persons to participate in a program may lead to "screening desirable" answers.[34] Use of a risk prediction model in the PLCO trial,

as compared with the NLST criteria, would have led to 41% fewer lung cancer cases being missed.[35] In the coming years, it may be possible to improve eligibility criteria for screening and adapt our models to incorporate broader eligibility criteria based on more complex measures of risk.[36] It will also be important to investigate possible important differences between men and women. In general, studies demonstrate that women receive diagnoses at an earlier age and at a more favorable cancer stage and more frequently are identified as having adenocarcinomas compared with men.[14,37-41] Recently, subgroup analyses of NLST showed statistically significant reductions in lung cancer mortality in persons diagnosed with adenocarcinoma (relative risk, 0.75 [95% CI, 0.60 to 0.94]) and not for other histologic types. These results also showed borderline-significant interaction with sex (relative risk, 0.73 for women vs. 0.92 for men; $p = 0.08$).[42] Inviting asymptomatic individuals for screening and implementing a large-scale screening program should be considered only when the benefits clearly outweigh the harms. Our analysis provides a detailed account of the balance between harms and benefits of annual lung cancer screening to inform individuals, clinicians, and policymakers. However, our predictions have some uncertainty and are contingent on high-quality screening, 100% adherence to screening, and closely coordinated follow-up and treatment protocols. Future providers and possible recipients of lung cancer screening should be fully aware of this and opt for screening only after having been informed about these harms and benefits.

Acknowledgments:

The authors thank Melecia Miller, MPH (formerly of Massachusetts General Hospital); Suresh Moolgavkar (Fred Hutchinson Cancer Research Center); and Arry de Bruijn (Erasmus Medical Center).

References

1. Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer* 2010; **127**(12): 2893-917.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians* 2013; **63**(1): 11-30.
3. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of Reduced Tobacco Smoking on Lung Cancer Mortality in the United States During 1975–2000. *Journal of the National Cancer Institute* 2012; **104**(7): 541-8.
4. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
5. Heijnsdijk EAM, Wever EM, Auvinen A, et al. Quality-of-Life Effects of Prostate-Specific Antigen Screening. *New England Journal of Medicine* 2012; **367**(7): 595-605.
6. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and Study Design. *Radiology* 2011; **258**(1): 243-53.
7. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
8. Anderson CM, Burns DM, Dodd KW, Feuer EJ. Chapter 2: Birth-Cohort-Specific Estimates of Smoking Behaviors for the U.S. Population. *Risk Analysis* 2012; **32**: S14-S24.
9. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.
10. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort–Specific Smoking Histories, 1965–2009. *American Journal of Preventive Medicine* 2014; **46**(2): e31-e7.
11. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Analysis* 2012; **32 Suppl 1**: S51-68.
12. Draisma G, Etzioni R, Tsodikov A, et al. Lead-Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. *Journal of the National Cancer Institute* 2009; **101**(6): 374-83.
13. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *European Journal of Operational Research* 1978; **2**(6): 429-44.
14. 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 Tumors. *Journal of Thoracic Oncology* 2007; **2**(8): 706-14.
15. Korfage IJ, de Koning HJ, Roobol M, Schröder FH, Essink-Bot M-L. Prostate cancer diagnosis: The impact on patients' mental health. *European Journal of Cancer* 2006; **42**(2): 165-70.
16. de Gelder R, van As E, Tilanus-Linthorst MMA, et al. Breast cancer screening: Evidence for false reassurance? *International Journal of Cancer* 2008; **123**(3): 680-6.
17. van der Aalst CM, van den Bergh KAM, Willemsen MC, de Koning HJ, van Klaveren RJ. Lung cancer screening and smoking abstinence: 2 year follow-up data from the Dutch–Belgian randomized controlled lung cancer screening trial. *Thorax* 2010; **65**(7): 600.
18. van der Aalst CM, van Klaveren RJ, de Koning HJ. Does participation to screening unintentionally influence lifestyle behaviour and thus lifestyle-related morbidity? *Best Practice & Research Clinical Gastroenterology* 2010; **24**(4): 465-78.
19. van der Aalst CM, van Klaveren RJ, van den Bergh KAM, Willemsen MC, de Koning HJ. The impact of a lung cancer computed tomography screening result on smoking abstinence. *European Respiratory Journal* 2011; **37**(6): 1466.
20. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2009; **151**(10): 738-47.

21. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: A decision analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 2008; **149**(9): 659-69.
22. Ma J, Ward EM, Smith R, Jemal A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer* 2013; **119**(7): 1381-5.
23. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of ct screening for lung cancer: A systematic review. *Journal of the American Medical Association* 2012; **307**(22): 2418-29.
24. de Gelder R, Heijnsdijk EAM, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting Overdiagnosis Estimates in Population-based Mammography Screening. *Epidemiologic Reviews* 2011; **33**(1): 111-21.
25. van Ravesteyn NT, Miglioretti DL, Stout NK, et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: A comparative modeling study of risk. *Annals of Internal Medicine* 2012; **156**(9): 609-17.
26. Berrington de González A, Kim KP, Berg CD. Low-dose lung computed tomography screening before age 55: estimates of the mortality reduction required to outweigh the radiation-induced cancer risk. *Journal of Medical Screening* 2008; **15**(3): 153-8.
27. 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. *Annals of Internal Medicine* 2013; **159**(6): 411-20.
28. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomized Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; **67**(4): 296-301.
29. 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. *European Journal of Cancer Prevention* 2012; **21**(3): 308-15.
30. 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 randomized lung cancer multi-slice CT screening trial (NELSON). *International Journal of Cancer* 2007; **120**(4): 868-74.
31. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *European Respiratory Journal* 2013; **42**(6): 1659.
32. Horeweg N, van der Aalst CM, Thunnissen E, et al. Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *American Journal of Respiratory and Critical Care Medicine* 2013; **187**(8): 848-54.
33. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of Lung Nodules Detected by Volume CT Scanning. *New England Journal of Medicine* 2009; **361**(23): 2221-9.
34. Peto J. That the effects of smoking should be measured in pack-years: misconceptions 4. 2012; **107**(3): 406-7.
35. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.
36. Kovalchik SA, Tammemägi M, Berg CD, et al. Targeting of Low-Dose CT Screening According to the Risk of Lung-Cancer Death. *New England Journal of Medicine* 2013; **369**(3): 245-54.
37. 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. *Annals of Oncology* 2002; **13**(7): 1087-93.
38. Sagerup CMT, Småstuen M, Johannesen TB, Helland Å, Brustugun OT. Sex-specific trends in lung cancer incidence and survival: a population study of 40 118 cases. *Thorax* 2011; **66**(4): 301.
39. Caldarella A, Crocetti E, Comin CE, Janni A, Pegna AL, Paci E. Gender differences in non-small cell lung cancer: A population-based study. *European Journal of Surgical Oncology* 2007; **33**(6): 763-8.
40. Schwartz KL, Crossley-May H, Vigneau FD, Brown K, Banerjee M. Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes & Control* 2003; **14**(8): 761-6.

41. Ringer G, Smith JM, Engel AM, Hendy MP, Lang J. Influence of sex on lung cancer histology, stage, and survival in a midwestern United States tumor registry. *Clinical Lung Cancer* 2005; **7**(3): 180-2.
42. Pinsky PF, Church TR, Izmirlian G, Kramer BS. The National Lung Screening Trial: Results stratified by demographics, smoking history, and lung cancer histology. *Cancer* 2013; **119**(22): 3976-83.

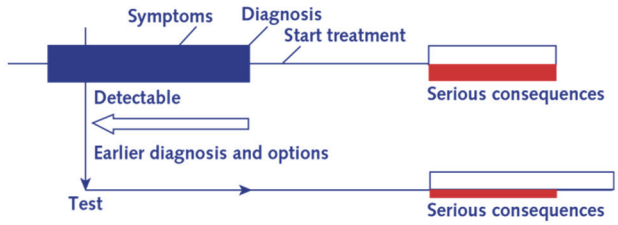
Chapter 5

Supplementary material

Estimating Radiation-Related Lung Cancer Cases

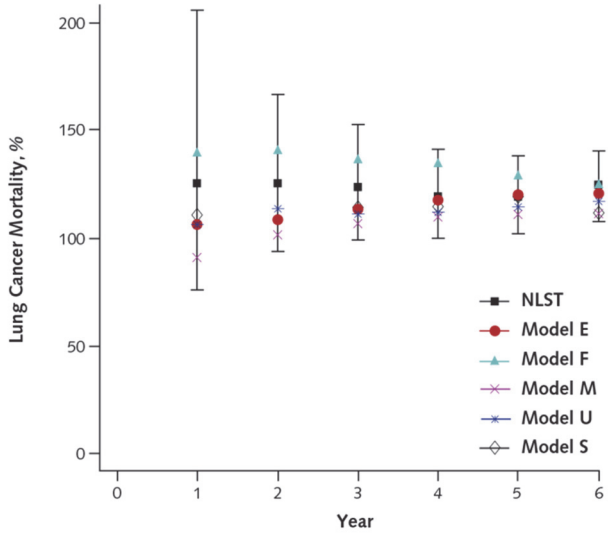
The Fred Hutchinson Cancer Research Center group included a radiation dose–response component in its lung cancer model to estimate lung cancer risk associated with each simulated individual's history of CT screening and CT follow-up examinations. Each scheduled CT screening examination was modeled as contributing a radiation exposure of 1 millisievert (mSv), and each was assumed to have a 25% chance of leading to a follow-up CT examination that included a 4-mSv exposure. The biological effects of these radiation exposures were modeled as occurring within a 1-year window through an increase in the premalignant clonal expansion rate and the malignant transformation rate. The nonlinear dose–response relationship and parameters affecting premalignant clonal expansion and malignant transformation were based on an earlier calibration to radiation risk for lung cancer incidence using the two-stage clonal expansion model. Data for the calibration included 191,042 individuals with complete records in the Canadian National Dose Registry cohort for low-dose exposure to x-ray, gamma, and other types of ionizing radiation exposure between 1951 and 1988, with follow-up for lung cancer between 1969 and 1988.[1] This two-stage clonal expansion model of radiation-related lung cancer risk was also used to estimate risk among Japanese atomic bomb survivors in the Life Span Study cohort, finding good agreement with observed lung cancer cases, and was found to have reasonable estimates of lung cancer risk compared with other studies.[1,2] Radiation risk has generally been extrapolated from individuals with atomic bomb exposure in Japan. Although there is no evidence that a single large exposure to radiation equates proportionally to multiple miniscule radiation exposures as in CT lung cancer screening, this method has been adopted because it seems to be the safest assumption to make. Massachusetts General Hospital used a radiation risk model with the parameters from the Biological Effects of Ionizing Radiation (BEIR) VII report and the absorbed doses to the lung to estimate the number of excess lung cancer cases and deaths. Estimated from the historical data, organ-absorbed (lung) doses for screening CT examinations were 3.8 milligray (mGy) (men) and 3.9 mGy (women).[3] For follow-up CT examinations, the doses are 15.4 mGy (men) and 15.0 mGy (women). The BEIR VII report recommended combining excess additive risk and excess relative risk estimates, but the former were not available by histologic cell type or by smoking status.

Figure S1: Diagram of how earlier detection (followed by treatment) may have an effect on reducing serious consequences of the disease and/or increasing life expectancy



All models account for the individual's age-specific smoking-related risk for lung cancer, the date and stage of lung cancer diagnosis, the corresponding lung cancer mortality, and the individual's life expectancy in the presence and absence of screening. By replicating trial detection, models estimate key parameters of the screening-detectable period and/or sensitivity and can subsequently estimate cancer detected in the screening scenarios. In essence, when a model incorporates the exact demographic characteristics of participants and the design of a trial, it should be able to reproduce cumulative incidence of lung cancer (by stage, histologic features, sex, age, type of detection, and round) and lung cancer mortality in both groups as closely as possible. The best fit is often defined as the lowest deviance between observed and model-expected numbers.

Figure S2: Percentage and 95% CI of lung cancer mortality in chest radiography group compared with computed tomography group in the NLST, by follow-up duration and comparison with 5 model group results



Abbreviations: E = Erasmus Medical Center; F = Fred Hutchinson Cancer Research Center; M = Massachusetts General Hospital; NLST = National Lung Screening Trial; S = Stanford University; U = University of Michigan.
As stated in the Methods section of Chapter 5, close calibration to difference in lung cancer mortality between groups of the NLST at 6 years' follow-up was prioritized, but not the slope before year 6. This was done on purpose because mortality differences in the first years of trials are subject to chance and small numbers.

Table S1: Key Similarities and Differences Between the Models in Estimating Effects on Life Expectancy With an Effective Lung Cancer Screening Test*

Variable	Erasmus Model	FHCRC Model	MGH Model	Stanford Model	University of Michigan Model
Simulation runs	10 million per scenario	100,000 per scenario (likelihood-based approach)	500,000	100,000	2 million per scenario
Risk mechanism	Two-stage clonal expansion model	Longitudinal multistage observation model	Probabilistic	Two-stage clonal expansion model	Multistage clonal expansion model
Incidence per 100,000 (no screening), *n*†	4,683	5,275	4,152	5,857	5,628
Mean age at diagnosis (no screening), y	74.01	76.32	69.28	71.90	71.14
Early (I/II) clinical stage (no screening), %	28	37	47	36	40
Adenocarcinoma (no screening), %	45.57	50.12	54.43	48.41	48.29
Clinical survival	SEER 2004-2008	NLST/PLCO	SEER 2004-2008	SEER 1988-2003	SEER 2004-2008
5-year survival rate (no screening), %	19	20	30	22	24
Lung cancer deaths in no-screening scenario, n†	3,513	3,773	2,449	4,531	4,331
Sensitivity/screening detectability	By stage/histologic features/sex	By number of cells/ histologic features/sex	By size and location in lung (correlates with histologic features)	By size/histologic features	By number of cells/histologic features/sex
Follow-up procedures	Based on NLST (1.5 examinations per positive screening result)	Based on NLST	Algorithms based on size thresholds and risk factors, adjusted to NLST rates	Algorithms based on size thresholds, calibrated to NLST rates for annual screening	Based on NLST
Stages	IA, IB, II, IIIA, IIIB, IV	IA1, IA2, IB, II, IIIA, IIIB, IV	IA1, IA2, IB, II, IIIA, IIIB, IV	Early (I-II), late (III-IV)	IA1, IA2, IB, II, IIIA, IIIB, IV
Early stage (in A-55-80-30-15‡ screening scenario), %	48	45	54	53	53

General mechanism of effect	Cure model	Cure and stage shift	Not stage-shift model	Cure model	Stage shift
Effect of earlier detection	Screening-detected cases (which are treated earlier) are associated with a reduced risk for dying of lung cancer compared with stage-specific survival had the same tumor been clinically diagnosed later. The improved prognosis is represented as a cure fraction specific to stage at detection, but if curative treatment fails, patient survival will equal the survival in the case that the tumor had been clinically diagnosed.	Estimates cure rates, which depend on sex and tumor stage, size, and histologic features	Assumes that most patients with early-stage non–small-cell lung cancer would undergo resection; therefore (for patients without undetected distant metastases or additional primary lung cancers in another lobe), this resection is curative	Estimates the probability of fatal metastases as a function of tumor size, histologic features, and sex. Patients with advanced-stage lung cancer are, by definition, identified after the onset of fatal metastases, but some early-stage patients are identified before this occurs. With screening, patients are more likely to be identified at an early stage and cured of their disease following standard of care.	Time to death from lung cancer is based on survival models that define cure by histologic features, stage, sex, and age at diagnosis. Mortality reduction due to screening is due to the earlier stage and younger age at detection.
Estimated relative risk for death from lung cancer 6 years after randomization (observed, 0.81; updated results, 0.84 [95% CI, 0.75–0.95])§	0.83	0.81	0.89	0.90	0.85
Life-years gained per lung cancer death averted (A-55-80-30-15)‡	10.6	11.8	11.4	8.7	10.4

Abbreviations: FHCRC = Fred Hutchinson Cancer Research Center; MGH = Massachusetts General Hospital; NLST = National Lung Screening Trial; PLCO = Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SEER = Surveillance, Epidemiology, and End Results.

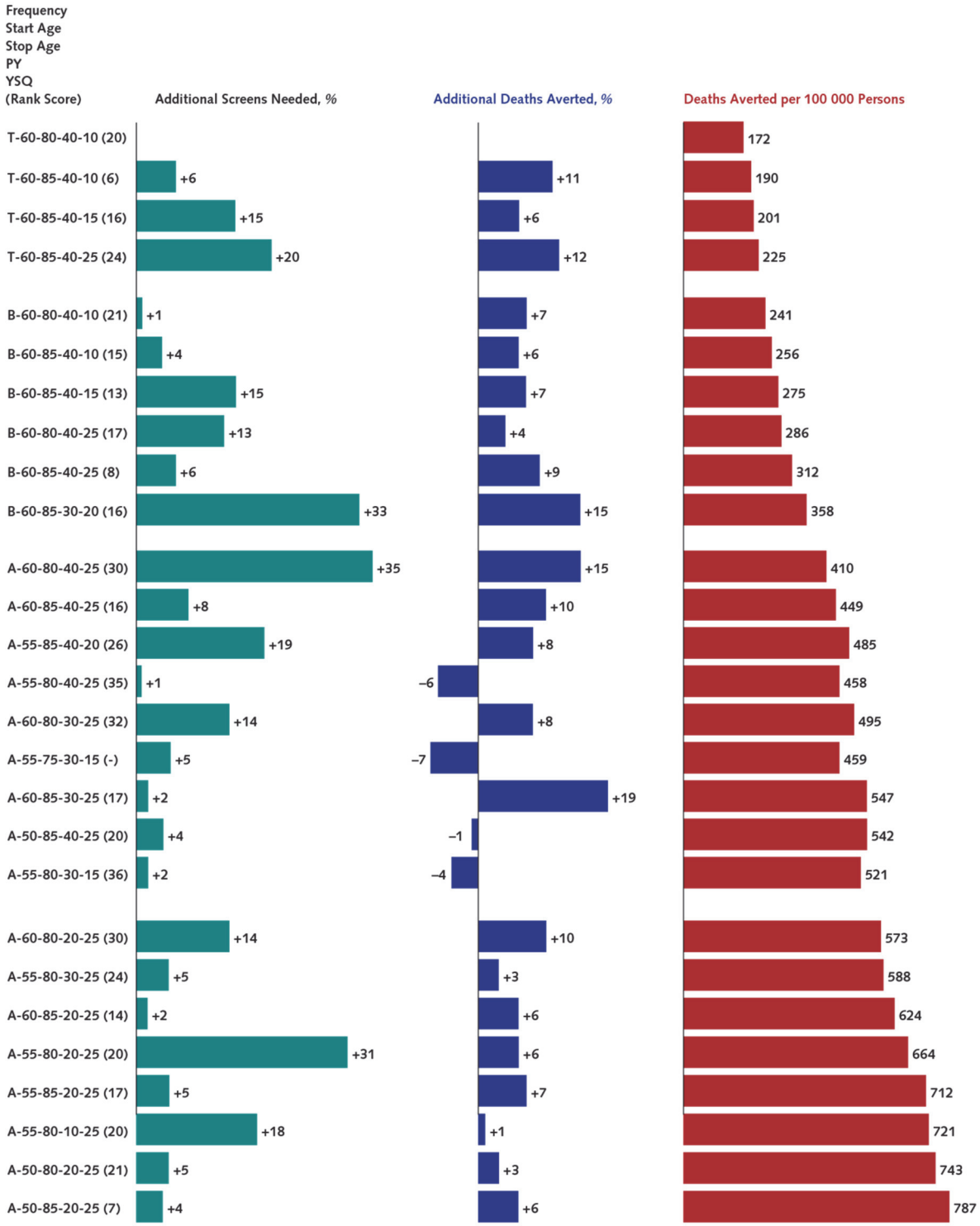
** If applicable, numbers per 100 000-person 1950 cohort followed from ages 45 to 90 years. Key parameters relate to the diagram in Figure 1.*

† There is no direct comparison with observed data for this specific 1950 cohort.

‡ Numbers are arranged as follows: Frequency (A = annual)–age start–age stop–minimum pack-years–maximum years since quitting.

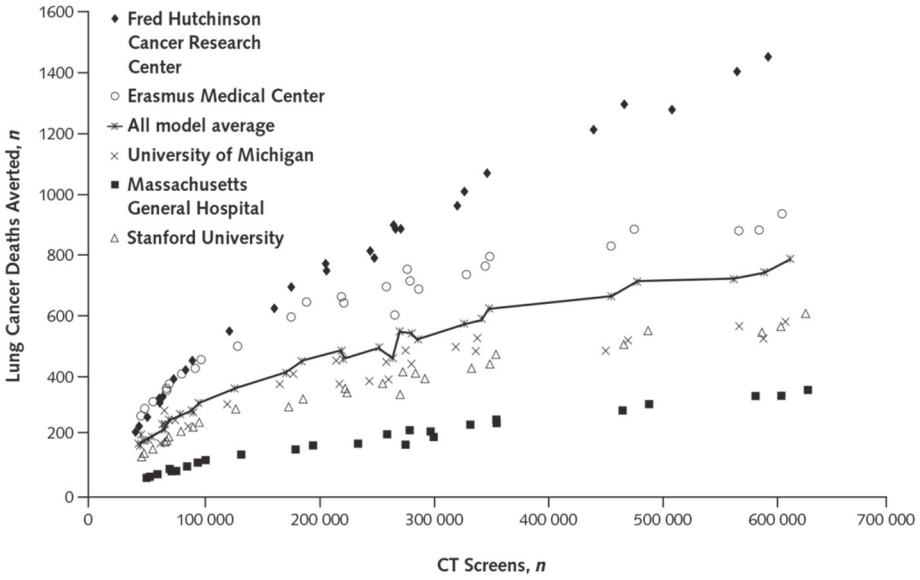
§ Data obtained from Pinsky (personal communication). Model estimates of relative risks were obtained from Meza et al (Chapter 1 of this thesis).[4]

Figure S3: Twenty-seven screening scenarios in order of increasing number of CT examinations needed, with the relative increase in screening examinations and lung cancer deaths averted (compared with the prior scenario), and the average number of lung cancer deaths averted in each scenario, for a 100,000-person 1950 cohort followed from ages 45 to 90 years



Abbreviations; A = annual; B = biennial; CT = computed tomography; PY = minimum pack-years; T = triennial; YSQ = maximum years since quitting.
Number of CT scans is given in Table 1 in Chapter 5. The bars show the absolute number of lung cancer deaths averted in 27 screening scenarios and the percentage increases in both screening examinations and deaths averted when inclusion criteria are relaxed. The first two triennial scenarios show the effect of stopping through age 80 or 85 years: about 6% more screenings when stopping through age 85 years, leading to 11% more lung cancer deaths averted (compared with stopping through age 80 years). Extending the maximum (quit) time from 10 years to 15 years leads to a 6% increase in deaths averted (at the expense of 15% additional screening examinations), and extending it to 25 years yields an additional 12% in lung cancer deaths averted (at the expense of 20% additional screening examinations). Decreasing the minimum patient-year eligibility criteria from 30 to 20 and to 10 patient-years in annual scenarios shows relatively large increases in additional CT scans needed compared with additional lung cancer deaths averted. The scenario's rank score among 576 possible scenarios (that is, the average distance to the efficient frontier for the 5 models) is shown in parentheses.

Figure S4: Absolute number of lung cancer deaths averted for the scenarios in Table 1, for all model groups separately and the average of 5 models



Abbreviations: CT = computed tomography.
Presentation of 100,000 individuals of the 1950 cohort followed from ages 45 to 90 years. The x-axis shows the number of CT screening examinations. Ranking of strategies is similar across models. There is no direct comparison with observed data for this specific 1950 cohort.

References

1. Hazelton WD, Moolgavkar SH, Curtis SB, Zielinski JM, Ashmore JP, Krewski D. Biologically based analysis of lung cancer incidence in a large Canadian occupational cohort with low-dose ionizing radiation exposure, and comparison with Japanese atomic bomb survivors. *Journal of Toxicology and Environmental Health Part A* 2006; **69**(11): 1013-38.

2. Cardis E, Gilbert ES, Carpenter L, et al. Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries. *Radiation Research* 1995; **142**(2): 117-32.

3. Berrington de González A, Kim KP, Berg CD. Low-dose lung computed tomography screening before age 55: estimates of the mortality reduction required to outweigh the radiation-induced cancer risk. *Journal of Medical Screening* 2008; **15**(3): 153-8.

4. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.

Chapter 6

The impact of overdiagnosis on the selection of efficient lung cancer screening strategies

Summer S. Han, Kevin ten Haaf, William D. Hazelton, Vidit N. Munshi, Jihyoun Jeon, Saadet A. Erdogan, Colden Johanson, Pamela M. McMahon, Rafael Meza, Chung Yin Kong, Eric J. Feuer, Harry J. de Koning, Sylvia K. Plevritis

Published as:

Han SS, ten Haaf K, Hazelton WD, et al.

International Journal of Cancer 2017; **140**(11): 2436-43.

Abstract

The U.S. Preventive Services Task Force (USPSTF) recently updated their national lung screening guidelines and recommended low-dose computed tomography (LDCT) for lung cancer (LC) screening through age 80. However, the risk of overdiagnosis among older populations is a concern. Using four comparative models from the Cancer Intervention and Surveillance Modeling Network, we evaluate the overdiagnosis of the screening program recommended by USPSTF in a U.S. 1950 birth cohort. We estimate the number of LC deaths averted by screening (D) per overdiagnosed case (O), yielding the ratio D/O, to quantify the trade-off between the harms and benefits of LDCT. We analyze 576 hypothetical screening strategies that vary by age, smoking, and screening frequency and evaluate efficient screening strategies that maximize the D/O ratio and other metrics including D and life-years gained (LYG) per overdiagnosed case. The estimated D/O ratio for the USPSTF screening program is 2.85 (model range: 1.5–4.5) in the 1950 birth cohort, implying LDCT can prevent ~3 LC deaths per overdiagnosed case. This D/O ratio increases by 22% when the program stops screening at age 75 instead of 80. Efficiency frontier analysis shows that while the most efficient screening strategies that maximize the mortality reduction (D) irrespective of overdiagnosis screen through age 80, screening strategies that stop at age 75 versus 80 produce greater efficiency in increasing life-years gained per overdiagnosed case. Given the risk of overdiagnosis with LC screening, the stopping age of screening merits further consideration when balancing benefits and harms.

Introduction

The National Lung Screening Trial (NLST) recently demonstrated that low-dose computed tomography (LDCT) screening is effective in reducing lung cancer (LC) mortality.[1] However, overdiagnosis of LC in LDCT screening is a significant concern.[2-4] Overdiagnosis is defined as a screen-detected disease that in the absence of screening, would not have become clinically evident within one's lifetime.[5] Overdiagnosis can lead to unnecessary treatment and costs and negatively impact well-being and life expectancy.[5]

The Cancer Intervention and Surveillance Modeling Network (CISNET) is a National Cancer Institute (NCI) sponsored consortium that uses a comparative statistical modeling approach to estimate the population-level impact of cancer screening. In prior work, the CISNET lung cancer screening models were used to evaluate the comparative effectiveness of 576 screening strategies that varied by smoking history, age, and screening frequency.[6-8] These analyses were used by the U.S. Preventive Services Task Force (USPSTF) as secondary evidence to support the recent recommendation to annually screen persons aged 55 to 80 with the same smoking criteria as the NLST.[9] One notable aspect of the USPSTF recommendation is the increased stopping age to 80 from 74 compared to the NLST. Although several harms (including overdiagnosis) were considered by the USPSTF, harms were not explicitly incorporated when ranking the efficient scenarios provided by CISNET; instead these scenarios were selected by maximizing the LC mortality reduction, i.e. the number of LC deaths (D) prevented due to screening over the number of CT screening examinations.[7]

While the lung cancer screening guidelines by the USPSTF recommended targeting the age-group of 55 to 80, there is still considerable debate over the potential benefits and harms of screening in the older population. The Centers for Medicare & Medicaid Services (CMS) issued a national coverage determination for Medicare coverage of CT screening for individuals aged 55 to 77 (www.cms.gov), whereas the USPSTF recommends screening up to age 80. Other guidelines such as those proposed by the American Cancer Society (ACS) recommended CT screening for individuals ages 55 to 74.[10] Hence, the stopping age of screening varies widely across different guidelines (ages 74, 77, 80) while the starting age

(age 55) and smoking criteria (30 pack-years and 15 years since cessation) are consistent across recommendations. Not surprisingly, the age of LC patients also has shown to be associated with increased risk of postoperative complications: patients aged 50 to 69 have a threefold higher risk for life threatening complications compared to patients aged <50, while the risk is nine-fold higher for patients aged >70.[11] Given the importance of the effect of screening age on the potential harms and the divergence on the recommended stopping for lung screening, it is essential to evaluate the optimal stopping age of lung cancer screening by more directly accounting for screening associated harms.

In this study, CISNET re-examines efficient screening strategies for lung cancer using a range of metrics that incorporate overdiagnosis. One useful metric for assessing the impact of overdiagnosis is the ratio between LC deaths prevented due to screening (D) and overdiagnosed cases (O), represented by D/O. This metric has been previously used to quantify the trade-off between the harms and benefits of screening.[5] Other measures are also considered such as life-years gained (LYG) due to screening and the LYG per overdiagnosed case (LYG/O). We use four independent CISNET lung models to estimate LC overdiagnosis for 576 alternative CT screening scenarios that vary by smoking, age, and screening frequency in the general U.S. population. Included is a direct comparison between the USPSTF recommended scenario and the NLST-like scenario (i.e. ACS-like scenario), which only differ in the screening stopping age (80 vs. 75). We evaluate screening strategies that optimize a range of metrics integrating overdiagnosis, comparing their outcomes to those based on LC mortality reduction (D) alone. These findings can provide insights into the impact of incorporating overdiagnosis on the selection of efficient lung screening programs, providing a more balanced consideration of screening benefits and harms.

Methods

CISNET models

Four CISNET LC screening models were independently developed based on different sets of assumptions and mathematical model structures at the following institutions: Erasmus Medical Center; Fred Hutchinson Cancer Research Center; Massachusetts General Hospital, and Stanford University. The common model components are: age-specific LC risk in the

absence of screening, natural history model for tumor growth and progression, screening component for predicting detection age and stage of LC in the presence of screening, diagnostic workup component for following up lung nodules; and corresponding LC mortality and death from other causes (described in Table S1 in the supplementary material of this Chapter).[6-8] Each model was calibrated and validated using data from NLST and PLCO (Prostate, Lung, Colorectal, and Ovarian Cancer Screening) to obtain estimates on screening-related parameters such as tumor size thresholds for diagnostic follow-up (described in Chapter 1 of this thesis).[6,12] Each model reproduced the observed incidence and mortality of LC (stratified by cancer stage at diagnosis, histology, sex, and detection mode) in both arms of these trials.[6]

Target population and screening scenarios

The models were used to simulate life histories of a U.S. cohort born in 1950, whose smoking histories and other-cause mortalities were generated using the Smoking History Generator.[13] We chose the 1950 birth cohort because it was considered in the USPSTF report.[9] We evaluated a total of 576 screening scenarios, varying the frequency of screening (annual, biennial, or triennial), starting age (45, 50, 55, or 60), stopping age (75, 80, or 85), minimum pack-years of smoking (25, 30, 35, or 40) and maximum years since quitting smoking (5, 10, 15, or 20) as considered in our previous reports.[7,8] Each model was run for each of the 576 screening strategies, assuming perfect screening compliance. For each strategy, each model produced several population-level outcomes, including number of LC deaths, number of LDCT screening examinations, number of prevented LC deaths (D) and life-years gained (LYG) due to screenings compared to a no-screening scenario. All counts were normalized per 100,000 persons in the cohort, who were followed up from age 45 to 90. False-positives, radiation-related harms, and follow-up examinations were also previously quantified by some of the models.[7]

Quantification of overdiagnosis

A patient is defined as overdiagnosed if their LC is detected in the screening scenario, but the tumor would not have been clinically detected before death from other causes in the no-screening scenario. We took the probability that a lung cancer detected by screening was

overdiagnosed as a measure for overdiagnosis. Using each simulation model, the risk of overdiagnosis was calculated as the number of overdiagnosed cases divided by the number of screen-detected cases, i.e., the proportion of screen-detected cases that are overdiagnosed. We compared overdiagnosis risk of 576 strategies stratified by screening starting/stopping age, smoking (pack-years and year-since-quit), gender or histology. To compare the estimated median overdiagnosis risk by groups, we applied the non-parametric Kruskal-Wallis (K-W) test. D/O was calculated by dividing the number of LC deaths prevented due to screening by the number of overdiagnosed cases.

Selection of consensus scenarios

For each model, we selected a series of scenarios that maximize the D/O ratio over the number of screening examinations by identifying the convex hull on a scatter plot between the D/O ratio (y-axis) and the number of CT screens (x-axis).[7] An efficient frontier is defined as this convex hull, that is, a curve that connects a set of scenarios that maximize the y-values over the x-values. A scenario was labeled as an "efficient scenario" if it is among the top 25% closest scenarios to the efficiency frontier. A set of consensus scenarios was identified by choosing scenarios that are defined as an "efficient scenario" by at least three out of the four models. We selected consensus scenarios by maximizing the D/O ratio and, separately D, as a function of the number of screening examinations, by sex. In selecting efficient and consensus scenarios for both metrics, we focused on annual screening scenarios with starting age ≥55 and stopping age ≤80, since they are considered to be the most feasible for implementation. When analyzing the findings, we focused on scenarios that are near the NLST and USPSTF scenarios, associated with the number of CT screens ranging between 250,000 and 350,000 (per 100,000 persons in the cohort) for males and 160,000 and 260,000 for females.

In a sensitivity analysis, we considered several alternative metrics that incorporate overdiagnosis to examine how the selection of efficient and consensus scenarios is affected by using different metrics. First, we considered LYG instead of D. While LYG takes into account different life expectancy among LC cases when measuring the benefit of screening, it does not explicitly incorporate overdiagnosis. Therefore, we also considered LYG/O as a

metric for selecting efficient screening strategies. Another alternative metric that incorporates overdiagnosis is defined as D-O, which is the net prevented LC deaths subtracted by the number of overdiagnosed cases; this metric measures the benefit of screening (D) penalized by overdiagnosis (O). The last metric that was considered is defined as D/(O/S), the number of LC deaths prevented per overdiagnosis risk, where S is the number of screen-detected cases.

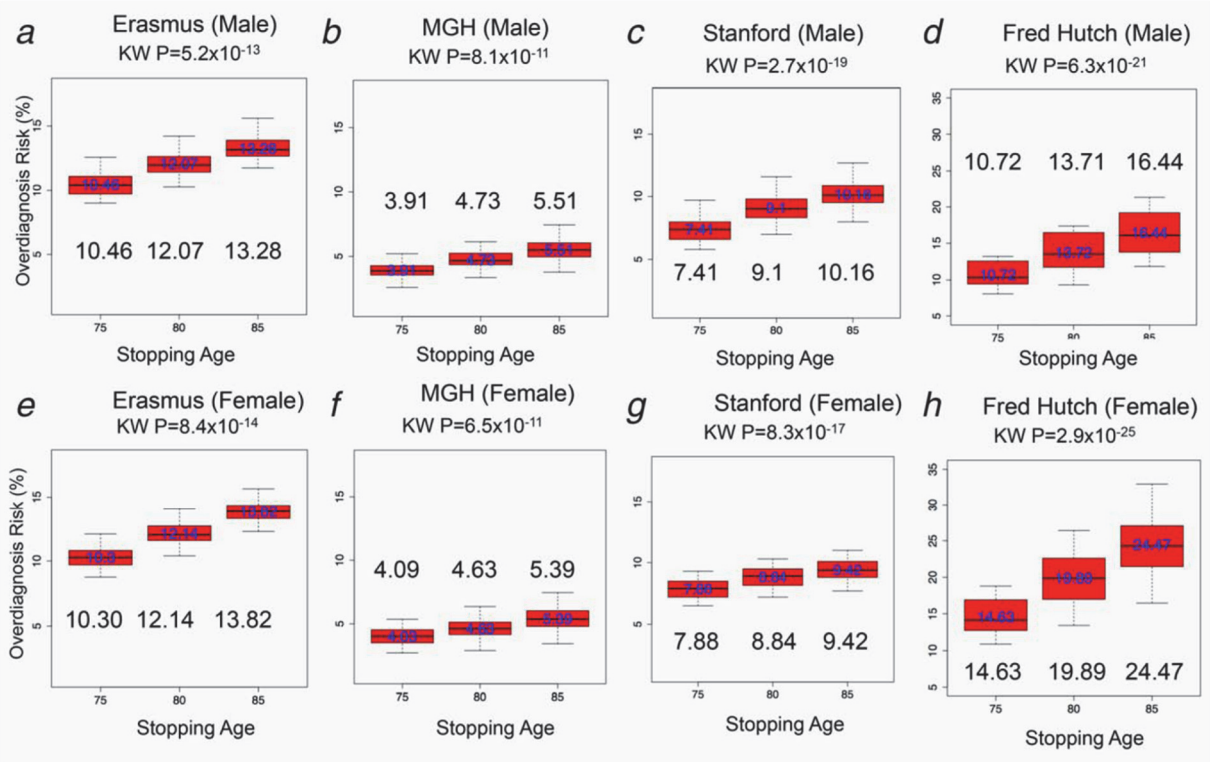
Results

The calibration results of the four models using the NLST data are shown in Figure S1 in the supplementary material of this Chapter. This figure shows that the four models reliably reproduce the outcomes of the NLST data, in which the model-based estimates for excess LC incidence rate in the CT arm compared to the chest X-ray arm are consistent with the reported value of 18.5% (95% confidence interval 5.5–30.6%) based on the NLST data.[2]

Overdiagnosis risk across 576 LC screening strategies

The analysis of 576 screening scenarios indicates that the overdiagnosis risk is highly influenced by screening stopping age. Figures 1a to 1d display the results for males, in which the screening programs with stopping age 85 have higher overdiagnosis rates (model median range: 5.51–16.44%) than the programs with lower stopping ages (model median range: 3.9–10.7% for stopping age 75 and 4.7–13.7% for stopping age 80). These patterns are similar for females and across the four models with all p-values of the eight K-W tests $<10^{-10}$. The comparisons of overdiagnosis risks by screening frequency, starting age, and pack-years of smoking are presented in Figure S2 in the supplementary material of this Chapter. In these analyses, we find overdiagnosis is higher for more frequent screening, older starting age, and higher smoking pack-years. Among histologic subtypes, BAC (bronchioloalveolar carcinoma) has the highest overdiagnosis risk.

Figure 1: Overdiagnosis risk (%) of 576 scenarios by stopping age of screening programs for each model and gender



Overdiagnosis risk is calculated as the number of overdiagnosed cases divided by the number of screen-detected cases. The number for each box represents a median of overdiagnosis risk of screening programs with given stopping age. "KW" denotes Kruskal-Wallis (K-W) test.

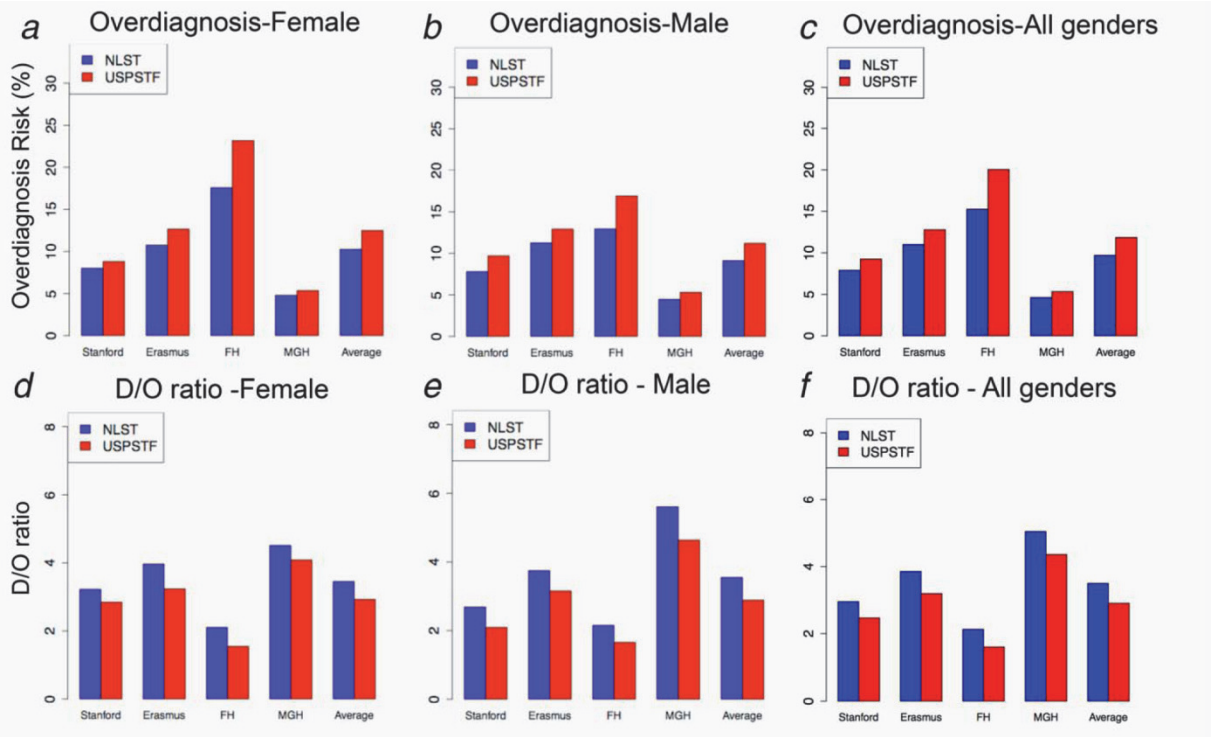
Comparisons of the USPSTF and NLST-like scenarios

Comparisons of the USPSTF and the NLST-like scenarios (Figures 2a–2c) show that overdiagnosis risk is higher for the USPSTF scenario (mean: 11.9%; model range: 5.5–23.2%) than the NLST-like scenario (mean: 9.7%; model range: 4.4–17.6%) by 21.7% (model range for percentage increase: 10–31.8%) due to the extended stopping age. This pattern is consistently observed across the models for each gender. The analysis of the D/O ratio (shown in Figure S6 in the supplementary material of this Chapter) shows that the USPSTF scenario prevents 2.85 LC deaths per overdiagnosed case (mean D/O ratio 2.85; model range: 1.5–4.5). Notably, the D/O ratio increased by 22% when the program stops screening at age 75 instead of 80 as shown in the NLST-like scenario (mean: 3.49, model range: 2.10–5.61). The range of D/O ratios of all 576 scenarios are shown in Table S3 in the supplementary material of this Chapter, by model and gender.

Consensus scenarios

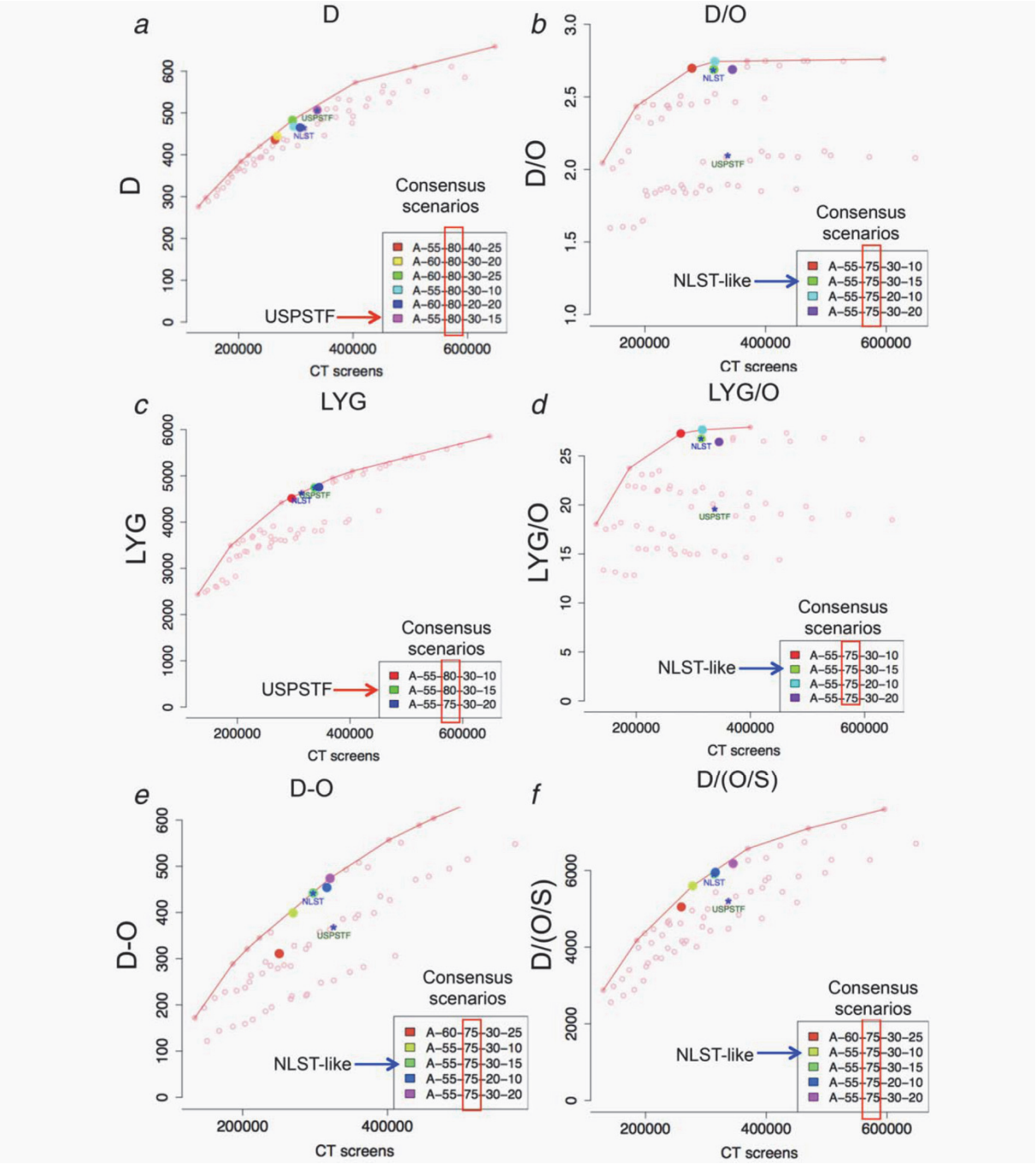
The scenarios that maximize D over the numbers of CT screens among males are shown in Figure 3(a), plotted for a representative model, with all other model results for both genders shown in Figures S5 and S6 in the supplementary material of this Chapter. It is notable that all the consensus scenarios have a stopping age 80 (instead of 75), which reflects the greater efficiency of programs that screen through older age 80, as opposed to 75, for reducing the number of LC deaths. Notably, the consensus scenarios include the USPSTF scenario (A-55–80-30–15). Other measures of benefits of LDCT screening for the consensus scenarios such as life-years and mortality reduction rate are shown in Table 1. The consensus scenarios that maximize D/O over the number of CT screens are shown in Figure 3(b). Interestingly, all the consensus scenarios maximizing D/O have stopping age 75, instead of 80, which suggests that programs that stop screening earlier are more efficient in reducing the number of LC deaths per overdiagnosed case. Notably, the NLST-like scenario (A-55–75-30–15) is included among the consensus scenarios selected by four models for each gender. The gray colored scenarios in Table 1 are consensus scenarios that overlap across genders within each metric for D and D/O. Overall, higher consensus was observed using the metric D/O across genders, where 80% of all the consensus scenarios (four of five) are selected in both genders while for metric D, around 44% of all consensus scenarios (four of nine) appear in both genders.

Figure 2: Comparisons of overdiagnosis risks and D/O ratios (the number of LC deaths prevented per overdiagnosed case) of the USPSTF and the NLST-like scenarios, by gender and both genders combined



Overdiagnosis risk is calculated as the number of overdiagnosed cases divided by the number of screen-detected cases.

Figure 3: Consensus screening scenarios chosen for males by maximizing: (i) D, the number of LC deaths prevented (a); (ii) D/O, the number of LC deaths prevented per overdiagnosed case (b); (iii) life-years gained (LYG) (c); (iv) life-year gained per overdiagnosed case (LYG/O) (d); (v) D-O, net LC deaths prevented subtracting overdiagnosed cases (e); (vi) and D/(O/S), the number of LC deaths prevented per overdiagnosis risk (f), where S is the number of screen-detected cases



In each Figure, we show the outcomes under several screening strategies that vary by age and smoking eligibility criteria. Each dot represents a specific screening strategy, with selected scenarios highlighted in color. Here, the x-axis is the number of CT screens that need to be performed under each strategy. (a) The number of LC deaths avoided versus no-screening (D, y-axis) under the given strategy. (b–f) show alternative outcome metrics: LC deaths avoided per overdiagnosed case (D/O, b), life-years gained (LYG, c), life-years gained per overdiagnosed case (LYG/O, d), the net prevented LC deaths subtracted by the number of overdiagnosed cases (D-O, e), and the number of LC deaths prevented per overdiagnosis risk (D/(O/S), f). Within each metric, a consensus scenario was identified by choosing a scenario that is defined as an "efficient scenario" (i.e. top 25% closest scenarios to the efficient frontier) by at least three out of the four models under each metric. The consensus scenarios are listed in the legend box and highlighted for a representative model. For each panel, the NLST-like and the USPSTF scenarios are plotted for reference purposes, regardless of whether or not they are included in the consensus list. The results for females are shown for D and D/O in Figures S5 and S6 in the supplementary material of this Chapter. Each legend box shows the scenarios selected by consensus across the four models and annotated as Frequency–Start Age (y)–Stop Age (y)–Pack-Years–Years Since Quitting.

Table 1: Consensus scenarios chosen by maximizing the number of prevented LC deaths (D); and the number of prevented LC deaths per overdiagnosed case (D/O)

Metric	Efficient scenario Frequency - Start age (y) – Stop age (y) - pack-years – years since quitting	No. CT scans	Overdiagnosis (%)	No. Overdiagnosed cases (O)	No. Prevented LC deaths (D)	Mortality reduction (%)	Life-years saved	D/O
Female	A-55-80-40-25	166,177	13.82	241	453	14.60	5,983	2.56
	A-60-80-20-10	186,932	12.29	224	456	14.70	5,889	2.74
	A-60-80-30-20	189,433	13.55	264	480	15.70	6,212	2.55
D	A-60-80-30-25	208,614	13.69	290	510	16.90	6,593	2.47
	A-60-80-20-15	227,049	12.81	272	527	17.60	6,833	2.65
	A-55-80-30-15	232,461	12.50	239	528	17.70	7,342	2.94
	A-60-80-10-15	261,556	12.77	281	551	18.30	7,209	2.69
	A-55-75-30-10	186,549	9.98	145	422	13.60	6,194	3.56
	A-55-75-30-15	214,158	10.28	166	466	15.30	6,831	3.45
D/O	A-55-75-30-20	235,702	10.39	180	495	16.50	7,340	3.48
	A-55-75-30-25	250,305	10.49	189	512	17.20	7,616	3.46
	A-55-75-20-10	253,105	9.66	169	504	16.90	7,350	3.63
Male	A-55-80-40-25	260,832	11.78	256	526	14.20	7,700	2.70
	A-60-80-30-20	261,778	11.88	277	528	14.20	7,122	2.55
D	A-55-80-30-10	286,878	11.20	250	547	14.70	7,936	2.87
	A-60-80-30-25	287,521	11.89	294	563	15.20	7,616	2.58
	A-60-80-20-20	307,380	11.77	288	561	15.30	7,779	2.57
	A-55-80-30-15	326,549	11.21	270	597	16.30	8,698	2.90
	A-55-75-30-10	267,730	9.13	180	498	13.20	7,513	3.55
D/O	A-55-75-30-15	301,853	9.12	192	536	14.40	8,170	3.54
	A-55-75-20-10	312,252	9.09	186	532	14.30	8,164	3.57
	A-55-75-30-20	330,807	9.02	200	566	15.40	8,679	3.60

A consensus scenario was identified by choosing a scenario that is defined as an "efficient scenario" (i.e. top 25% closest scenarios to the efficient frontier) by at least three out of the four models. For the numbers in each cell below, model average values were used. The USPSTF scenario and NLST-like scenarios are highlighted in red and blue, respectively. Grayed scenarios are the ones that overlap between genders within each metric.

Sensitivity analysis to outcomes metric

The selection of consensus scenarios using LYG instead of D is shown in Figure 3(c), which is similar to the selection under D, in the sense that both sets of consensus scenarios include the USPSTF scenario. However, a tendency was observed that using LYG (vs. D) penalizes screening through older ages; while all consensus scenarios chosen under D screen through 80, the consensus scenarios using LYG includes a scenario that stops screening at 75 (A-55 75-30–20). When overdiagnosis is taken into account by using LYG/O, however, the selection of consensus scenarios was remarkably similar to those using D/O (see Figure 3d). Further sensitivity analyses using alternative outcomes metrics, namely D-O and D/(O/S), show that the consensus scenarios that incorporate overdiagnosis are consistent with the ones selected by maximizing D/O (Figures 3e and 3f). Most of the consensus scenarios have stopping age 75 (except for one scenario) and the NLST-like scenario is included among the consensus scenarios. Using the metric D-O (Figure. 3e), four out of the five consensus scenarios are shown to be selected as consensus scenarios using the metric D/O (Table 1). Using the outcomes metric D/(O/S) (Figure 3f), four out of the five consensus scenarios also appear in the list selected using the metric D/O (Table 1).

Discussion

We presented a comparative model-based analysis of overdiagnosis in lung cancer screening by quantifying the trade-off between harms and benefits of LDCT screening. Our analysis shows that the lifetime screening program recommended by the USPSTF can prevent approximately three LC deaths by per overdiagnosed case (mean D/O ratio 2.85). The D/O ratio increases by 22% when the program stops screening at age 75 instead of 80, as shown in the NLST-like scenario (mean D/O ratio 3.49). Given that the USPSTF scenario prevents more LC deaths than the NLST-like scenario (i.e. a larger value of D in the USPSTF scenario), the lower D/O of the USPSTF scenario implies that the number of overdiagnosed cases (O) increases more quickly than the number of LC deaths prevented (D) as screening is extended to older ages. Overall, overdiagnosis was significantly associated with increased screening stopping age in our analysis of 576 hypothetical screening strategies (p <10^{-10}).The efficiency frontier analysis shows that the most efficient screening strategies that maximize the outcomes metrics incorporating overdiagnosis, namely D/O, LYS/O, D-O, and D/(O/S), are

consistently the strategies that stop screening at age 75 (which include the NLST-like scenario) compared to programs that screen through age 80. On the other hand, efficient programs chosen based on maximizing the number of LC deaths prevented (D) irrespective of overdiagnosis are the ones that screen through 80, which includes the USPSTF recommendation.

While previous model-based studies considered mortality reduction (D) when identifying efficient screening strategies, including our earlier work, we examined the impact of incorporating overdiagnosis on the selection of efficient scenarios by investigating various metrics that integrate overdiagnosis.[7,8,14-16] Undoubtedly there are other useful metrics to more explicitly quantify the harms associated with overdiagnosis such as quality-adjusted life-years (QALY); such metrics were not used in the current study because we intended our analysis to be directly comparable to the recent CISNET analyses performed for the USPSTF, which used D, not QALY, in ranking the efficient scenarios.[7,9]

A noteworthy aspect in the analysis of the D/O ratios for 576 screening strategies is that in most scenarios (99%), the D/O values are larger than one (i.e. D/O>1) across the four models. This finding implies that the number of LC deaths prevented by screening is greater than the number of overdiagnosed cases over a wide range of screening scenarios. In comparison with screening programs for other cancers, such as prostate cancer and breast cancer, which have been estimated to have D/O values less than one (0.2 for prostate cancer and 0.3 for breast), our results suggest that the negative impact of screening could be lower for LC compared to screening for other cancers.[17,18] However, the morbidity and mortality associated with overdiagnosis of LC may be higher than those of other cancers, hence a direct comparison of overdiagnosis-related outcomes across different cancers is not warranted.

Our model-based approach for analyzing overdiagnosis in lung cancer screening has several advantages compared to a trial data based approach. Recently an upper bound of the overdiagnosis risk of 18.5% for LC was estimated based on excess incidence using the NLST data.[2] This estimate is an upper bound because it quantifies the excess incidence observed in the CT arm compared to the chest X-ray arm after a short follow-up period (8 years from

trial entry), and likely includes screen-detected cases that would have been clinically detected. Longer follow-up would be needed to observe "catch-up" cancers in the control arm to allow a more accurate estimate of overdiagnosis based on excess incidence.[19,20] A trial-derived estimate of the overdiagnosis risk would be of limited generalizability, even if based on a sufficient follow-up period, because the estimate would be associated with the specific screening strategy of the trial. For example, most participants in NLST were screened annually for 3 years and aged between 55 and 74 who accumulated at least 30 pack-years of smoking and were current smokers or quit <15 years at the time of enrollment. However, different screening strategies (e.g. variations in smoking, stopping age or screening frequency) would likely yield different overdiagnosis risks. Given of all these challenges, a model-based approach is valuable for estimating overdiagnosis over a lifetime period by providing insights into how different screening strategies affect overdiagnosis.

While our overdiagnosis analyses across numerous screening scenarios could not be performed without modeling, modeling has limitations. Firstly, we found that the absolute values of the overdiagnosis risks vary across the four models. This variation is due in part to the fact that the four models were developed independently based on different assumptions, datasets, and mathematical formulations. Given these differences, the model variation captures a range of uncertainty associated with model building that could not be captured by one model alone. However, despite this model variation, relative magnitudes of overdiagnosis risks and related statistics such as D/O were notably consistent across the models. For example, overdiagnosis was higher in the USPSTF scenario than in the NLST-like scenario in all four models, as was D/O in the NLST-like scenario compared to the USPSTF. Second, while we accounted for smoking-related effects on other-cause mortality, we assumed that screening was performed on any individual who met the smoking and age criteria without explicit consideration of existing comorbidities. A screen-detected cancer patient with significant comorbidities is more likely to be overdiagnosed than one without significant comorbidities, because the patient with comorbidities has a higher competing risk of death. As additional data becomes available to associate other-cause mortality with comorbidities in screening eligibility, the risk of overdiagnosis will likely to decrease. Third, our study assumed perfect screening compliance but if screening compliance reduces at the older ages, then overdiagnosis risks will decrease. Fourth, our analysis is based on

calibrations to the practice patterns of NLST; should practice patterns change, particularly for the management of small indeterminate nodules on CT, our D/O estimates would need to be modified. Finally, our study evaluated hypothetical screening strategies that vary by age, smoking, and screening frequency as considered in the USPSTF guidelines, but we did not vary other factors such as nodule size or screening results that may also have impacts on LC mortality or overdiagnosis.

Our future research includes the evaluation of efficient diagnostic work-up strategies by varying several factors for follow-up, such as nodule size, features, follow-up interval, use of biomarkers, and prior screening results to examine how these factors affect the benefits and harms of CT in the population setting. In summary, our model-based analysis shows that incorporating overdiagnosis affects the selection of efficient screening strategies. Consistent results across four independent models indicate that our findings are robust. We conclude that while screening through age 80 is efficient in reducing LC mortality irrespective of overdiagnosis, stopping screening at a younger age of 75 provides a greater efficiency in reducing LC deaths and increasing life-years gained per overdiagnosed case, which merits further consideration when balancing the benefits and harms of screening.

References

1. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
2. Patz EF, Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Internal Medicine* 2014; **174**(2): 269-74.
3. Dammas S, Patz EF, Goodman PC. Identification of small lung nodules at autopsy: implications for lung cancer screening and overdiagnosis bias. *Lung Cancer* 2001; **33**(1): 11-6.
4. Black WC. Overdiagnosis: An underrecognized cause of confusion and harm in cancer screening. *Journal of the National Cancer Institute* 2000; **92**(16): 1280-2.
5. Welch HG, Black WC. Overdiagnosis in Cancer. *Journal of the National Cancer Institute* 2010; **102**(9): 605-13.
6. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
7. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
8. McMahon PM, Meza R, Plevritis SK, et al. Comparing Benefits from Many Possible Computed Tomography Lung Cancer Screening Programs: Extrapolating from the National Lung Screening Trial Using Comparative Modeling. *PLOS ONE* 2014; **9**(6): e99978.
9. Moyer VA, on behalf of the USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 2014; **160**(5): 330-8.
10. Wender R, Fontham ETH, Barrera E, et al. American Cancer Society lung cancer screening guidelines. *CA: A Cancer Journal for Clinicians* 2013; **63**(2): 106-17.
11. Yano T, Yokoyama H, Fukuyama Y, Takai E, Mizutani K, Ichinose Y. The current status of postoperative complications and risk factors after a pulmonary resection for primary lung cancer. A multivariate analysis. *European Journal of Cardio-Thoracic Surgery* 1997; **11**(3): 445-9.
12. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
13. Holford TR, Clark L. Chapter 4: Development of the Counterfactual Smoking Histories Used to Assess the Effects of Tobacco Control. *Risk Analysis* 2012; **32**: S39-S50.
14. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2009; **151**(10): 738-47.
15. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: A decision analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 2008; **149**(9): 659-69.
16. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. *Journal of the National Cancer Institute* 2002; **94**(3): 193-204.
17. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen–based prostate cancer screening strategies: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2013; **158**(3): 145-53.
18. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. 2012; *The Lancet*; **380**(9855): 1778-86.
19. Marcus PM, Bergstralh EJ, Zweig MH, Harris A, Offord KP, Fontana RS. Extended Lung Cancer Incidence Follow-up in the Mayo Lung Project and Overdiagnosis. *Journal of the National Cancer Institute* 2006; **98**(11): 748-56.

20. Etzioni R, Xia J, Hubbard R, Weiss NS, Gulati R. A Reality Check for Overdiagnosis Estimates Associated With Breast Cancer Screening. *Journal of the National Cancer Institute* 2014; **106**(12): dju315-dju.

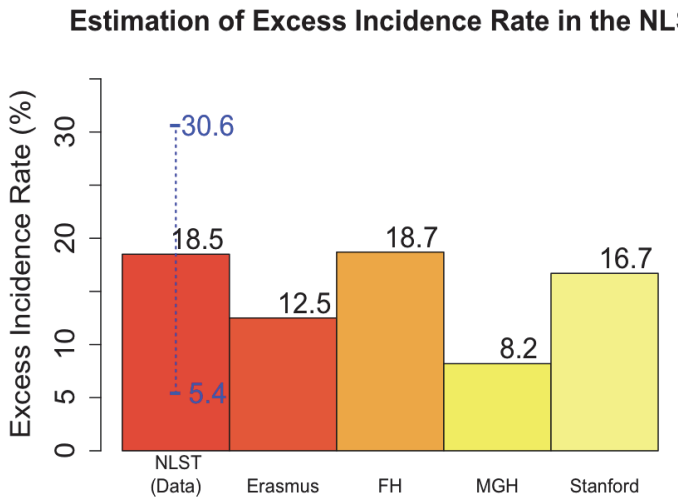
Chapter 6

Supplementary material

Overdiagnosis by screening frequency, starting age, and smoking pack-years

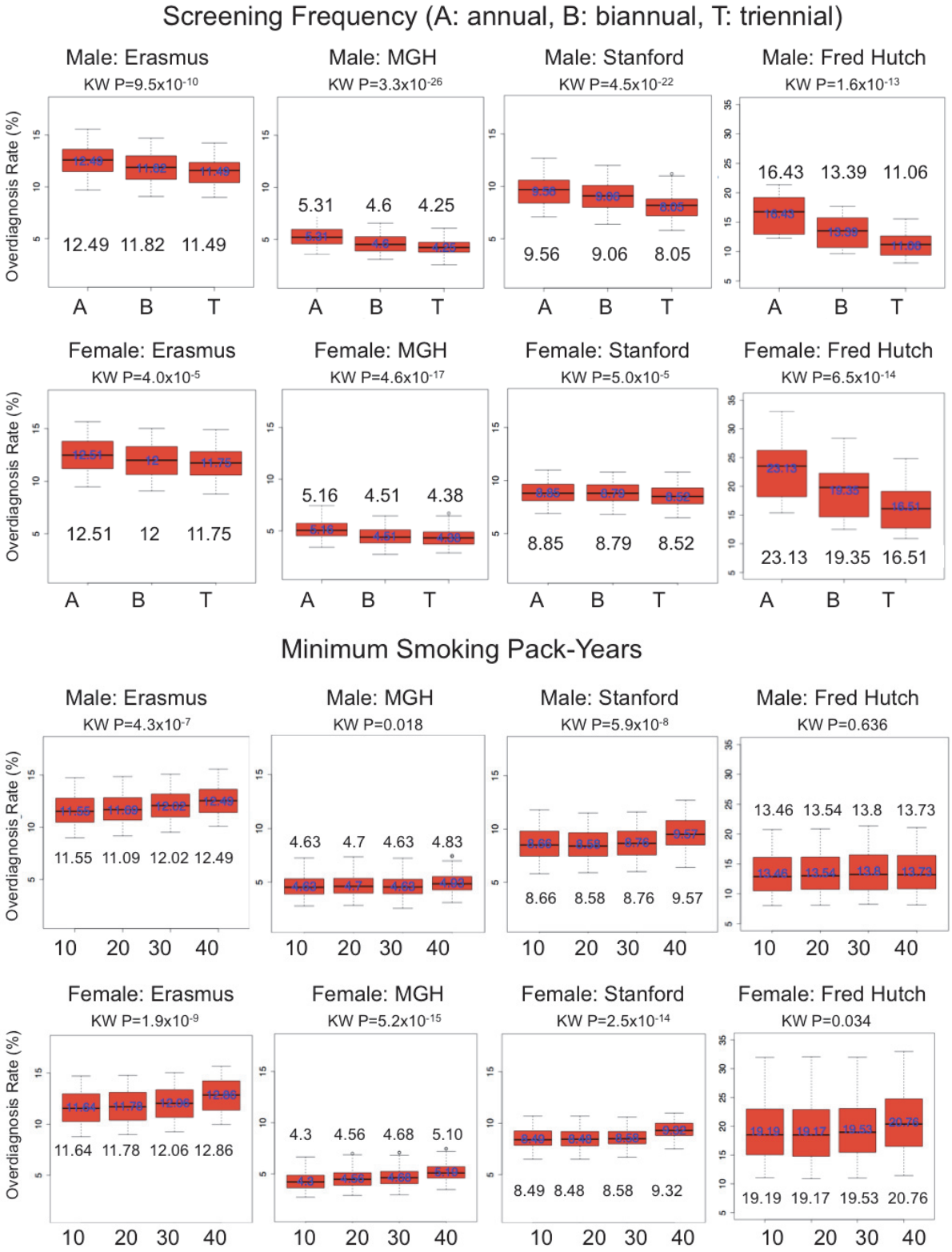
Overdiagnosis by screening frequency, starting age, and smoking pack-yearsOverdiagnosis is influenced by screening frequency, starting age and pack-years of smoking (**Figure S2**). Annual screening programs have higher overdiagnosis rates compared to triennial or biennial programs with K-W test p-value < 10^{-5}. Programs focusing on heavy smokers who smoked a minimum of 40 pack-years also have increased overdiagnosis rates compared to programs with low pack-years. Overall, screening programs with a starting age of 60 show higher overdiagnosis rates than programs with younger starting ages. Analyses by histological subtype show that BAC (bronchioloalveolar carcinoma) has the highest overdiagnosis rate (model range 10.2-42.6%) while the lowest overdiagnosis rate is observed in small cell lung cancer (model range 0-5.7%) (**Figure S3**). Despite the recent reclassification of in situ histologies, we used BAC as one of the histologic subtypes in analyzing overdiagnosis rates because it was used in the NLST the NLST-based overdiagnosis report.[1] The difference in overdiagnosis rates by gender is not statistically significant with directions being inconsistent across the models (**Figure S4**).

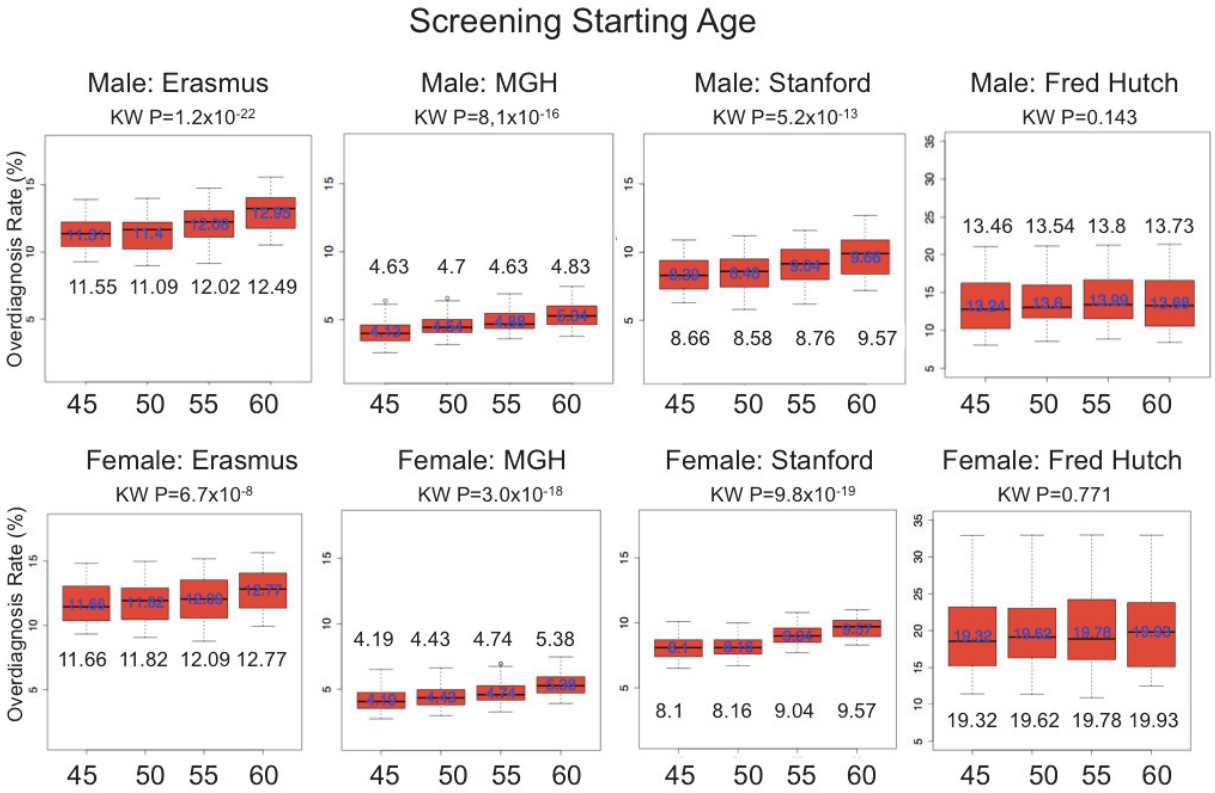
Figure S1: Comparing model-based estimation vs. observed excess incidence of lung cancer in the CT arm compared to the chest x-ray arm in the NLST



The first bar is the observed excess incidence of lung cancer using the NLST data at year 8 and the blue dotted lines provide a 95% confidence interval. The other four bars are model-based estimates of the excess incidence of lung cancer in the CT arm compared to the chest radiography arm at year 8 from trial entry. All model estimates are within the 95% confidence interval.

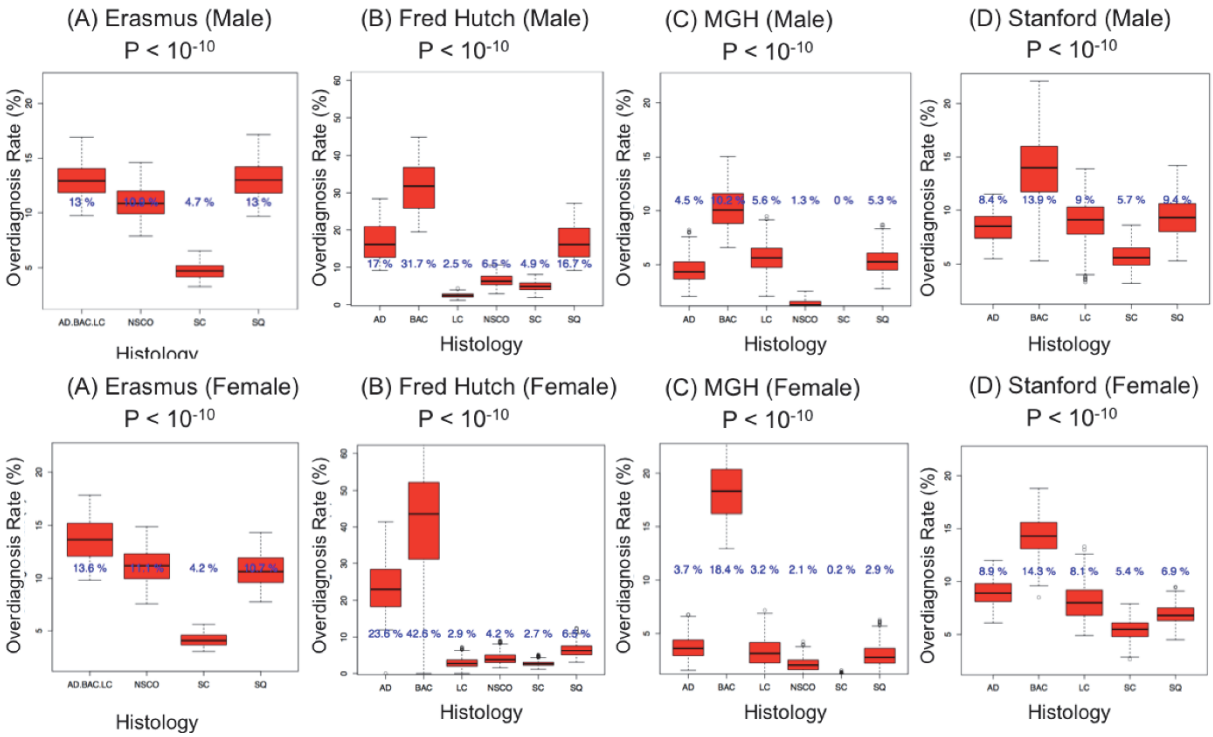
Figure S2: Overdiagnosis rates (%) of 576 scenarios by screening frequency (rows 1-2), smoking pack-years (rows 3-4) and screening starting age (rows 5-6)





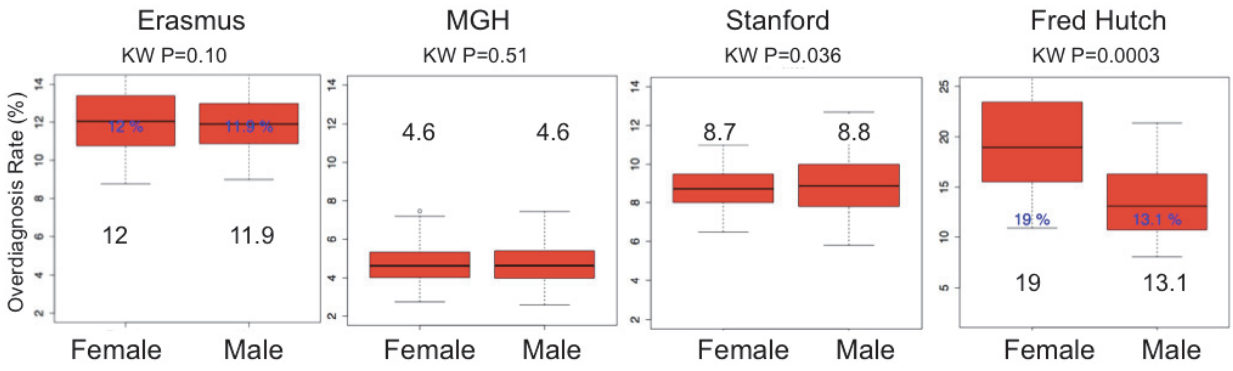
Overdiagnosis is defined as the number of overdiagnosed cases divided by the number of screen-detected cases. Model average was used. P-values are from K-W test comparing medians of groups.

Figure S3: Overdiagnosis of 576 scenarios by histological subtype by model



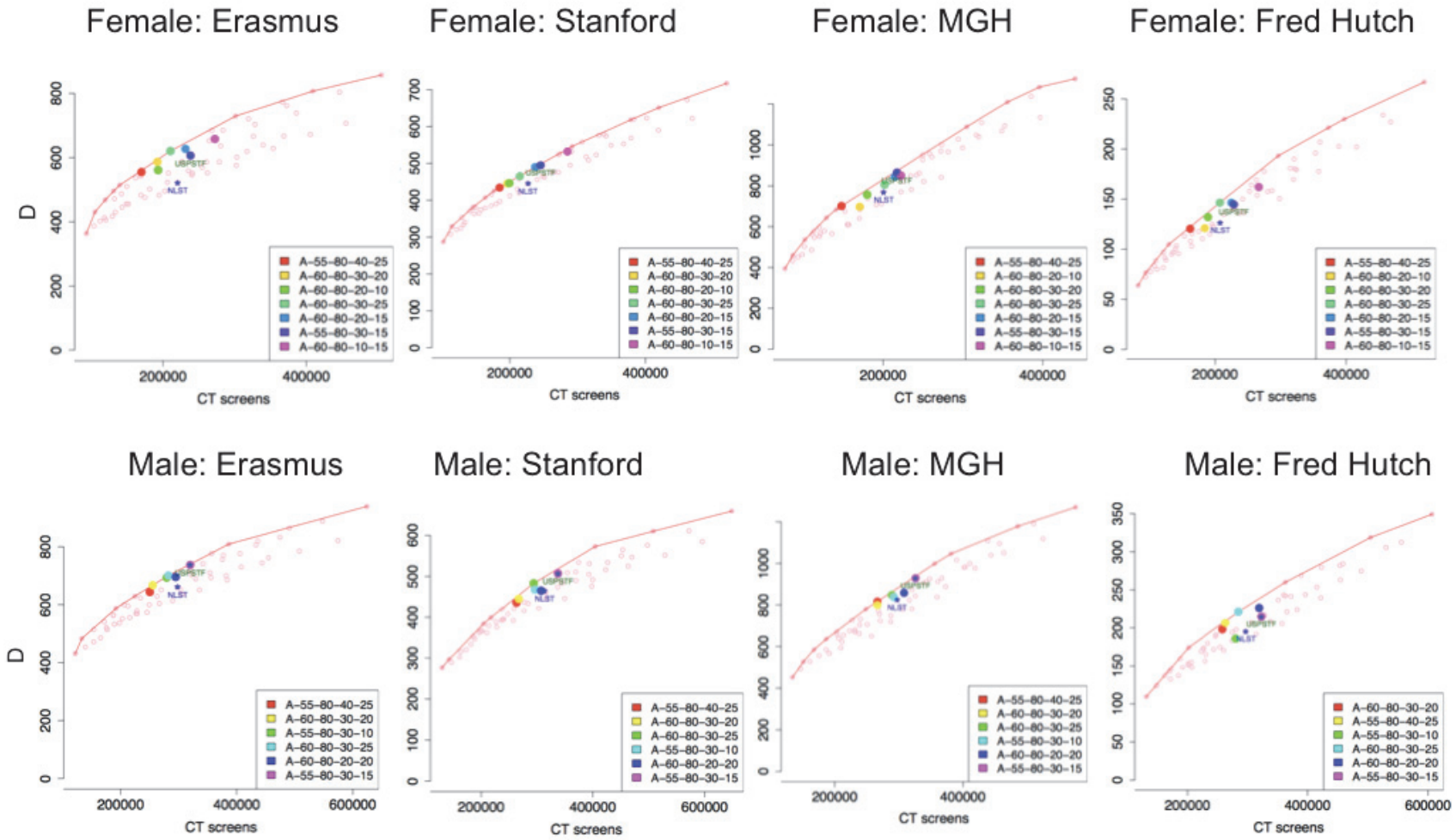
Abbreviations: AD: adenocarcinoma, BAC: bronchioloalveolar carcinoma, LC: large cell lung cancer, SC: small cell lung cancer, SQ: squamous lung cancer, NSCO: non-small cell lung cancer other than AD, BAC, LC and SQ. Overdiagnosis is defined as the number of overdiagnosed cases divided by the number of screen-detected cases. P-values are from K-W test comparing medians of histologic subtype groups.

Figure S4: Overdiagnosis of 576 scenarios by gender by model



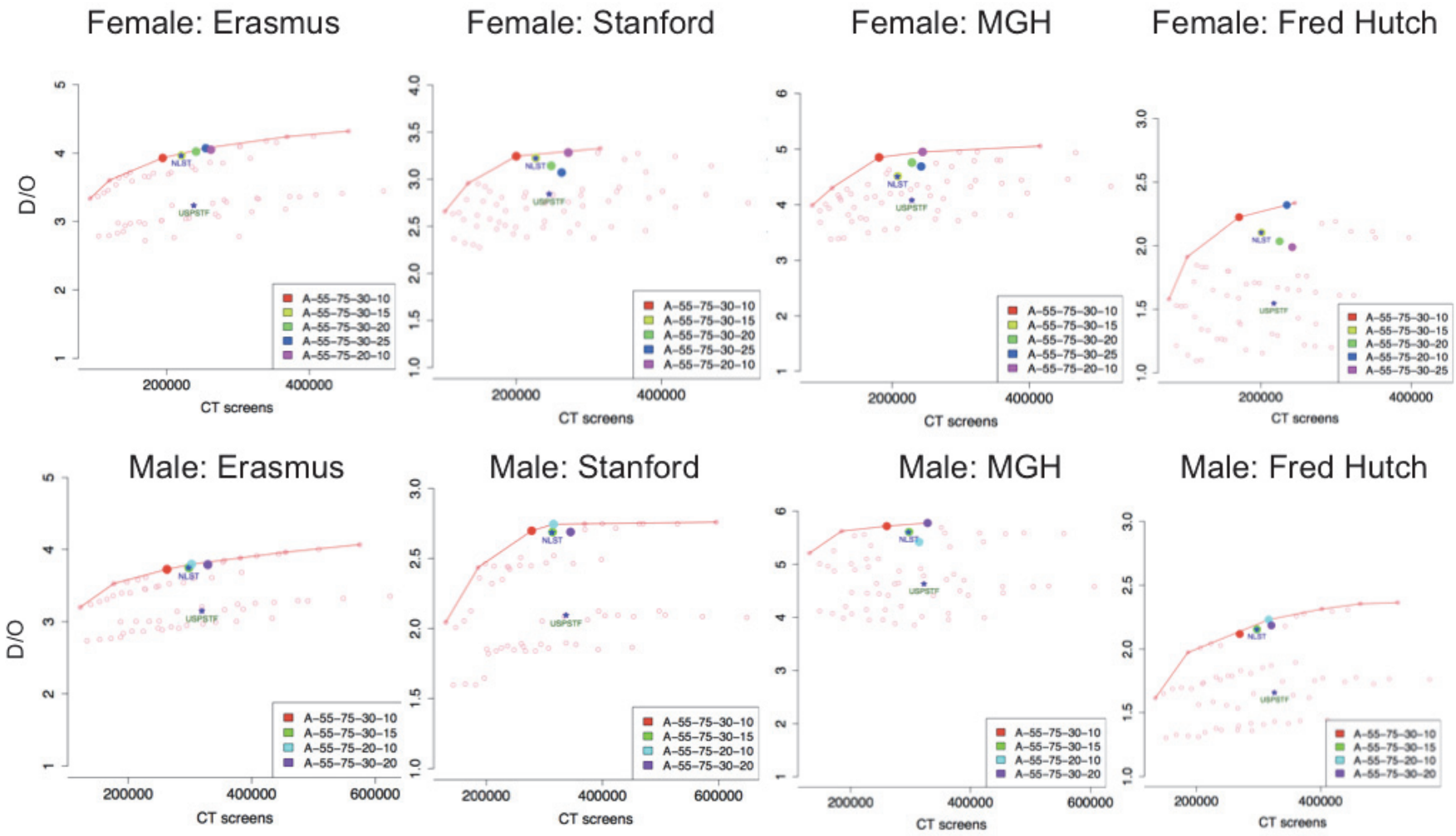
Overdiagnosis is defined as the number of overdiagnosed cases divided by the number of screen-detected cases. P-values are from K-W test comparing medians of two gender groups.

Figure S5: **The consensus scenarios chosen by maximizing the number of prevented LC deaths (D). A consensus scenario was identified by choosing a scenario that is defined as an "efficient scenario" (i.e. top 25% closest scenarios to the efficient frontier) by at least three out of the four models. The consensus scenarios are plotted against each model. The NLST-like scenario is not included in the list of consensus scenarios, but is plotted for reference**



Each legend box shows the selected scenarios by consensus of the four models (Frequency–Start Age (y)–Stop Age (y)–Pack-Years–Years Since Quitting).

Figure S6: The consensus scenarios chosen by maximizing D/O, number of prevented LC deaths per overdiagnosed case. A consensus scenario was identified by choosing a scenario that is defined as an "efficient scenario" (i.e. top 25% closest scenarios to the efficient frontier) by at least three out of the four models. The scenarios are plotted against each model. The USPSTF scenario is not included in the list of consensus scenarios, but is plotted for reference



Each legend box shows the selected scenarios by consensus of the four models (Frequency–Start Age (y)–Stop Age (y)–Pack- Years–Years Since Quitting).

Table S1: Model Comparisons

		Erasmus	Fred Hutch	MGH	Stanford
Lung cancer Incidence	Main reference(s) for mathematical model description	2,3	4	5-8	9
	Incidence model (Central smoking dose-response model)	Two-stage clonal expansion model (TSCE)[10]	Longitudinal multistage observation by histology[4]	Probabilistic[5-8]	Two-stage clonal expansion model (TSCE)[10]
	Data sources for estimating incidence model	NHS/HPFS, NLST, PLCO, and SEER	NLST and PLCO; PLuSS CT and CARET	SEER, NLST, PLCO	NHS/HPFS
	Smoking dose-response model histology specific	No	Yes	Yes	No
	Histological types	Adenocarcinoma/large cell/BAC, squamous, small cell, and other	Adenocarcinoma, large cell, squamous, BAC, other non-small cell, small cell	Adenocarcinoma, BAC, large cell, squamous, small cell, and other	Adenocarcinoma, large cell, squamous, small cell
Tumor progression	Tumor and stage progression model	Markov state-transition by histology	Based on tumor size and presence of metastasis	Based on tumor volume and metastatic burden	Tumor growth is modeled as exponential and stage is modeled based on tumor volume and metastatic burden
	Lung cancer stages	Ia, Ib, II, IIIa, IIIb, IV	Ia1,Ia2, Ib, II, IIIa, IIIb, IV	Ia1, Ia2, Ib, II, IIIa, IIIb, IV	Early (I-II), Advanced (III-IV)
	Data sources for estimating lung cancer survival	SEER-17 2004-2008 survival	NLST and PLCO	SEER-17 1973-2008 survival	SEER-17 1988-2003 survival
	Model estimation/ Calibration method	Nelder-Mead optimization of likelihood-based deviance criterion	Maximum likelihood approach	Simulated annealing based on weighted-sum total deviance	Nelder-Mead simplex for Natural History Model calibration to SEER, and multi-dimensional grid search for calibration to trials
	Data sources used for calibration	NLST; PLCO; SEER-17 2004-2008 incidence by age, stage, histology; NHS, HPFS	NLST; PLCO; model originally developed using PLuSS CT and CARET	NLST; SEER 1990-2000 incidence by age, stage, histology; survival by stage; Mayo CT; LSS	NLST; PLCO; SEER 1988-2003 survival by histology and sex

Screening	Screening sensitivity model	By stage and histology	By size (number of cells), histology and gender	By size (mm) and location in the lung (central/peripheral)	By size (mm) and histology
	Screening effectiveness mechanism	Cure model	Combination cure model and stage shift	Not a stage shift model	Not a stage shift model
	Positive Nodule Follow-up algorithm	Implicit	Implicit based on NLST follow-up rates	Explicit. Based on size at diagnosis and smoking history. Lung cancers diagnosed on follow-up are categorized as 'non-screen detected'	Explicit. Based on modified version of Fleischner society guidelines, which aim to more closely capture practices conducted in the NLST.
Other-cause mortality		U.S. rates (NCI Smoking History Generator)[11,12]	U.S. rates (NCI Smoking History Generator)[11,12]	U.S. rates (NCI Smoking History Generator)[11,12]	U.S. rates (NCI Smoking History Generator)[11,12]

Table S2: Comparisons of overdiagnosis rates (%) and D/O values of the USPSTF and the NLST-like scenarios by models

		USPSTF								
	Subtype	Stanford		MGH		Fred Hutch		Erasmus		Model
		F	M	F	M	F	M	F	M	Average
	AD	8.3	9.5	5.8	5.2	29.9	22.1	14.4*	13.9*	-
Overdiagnosis	SQ	7.0	10.4	4.0	7.1	7.2	20.7	11.4	14.4	-
	LC	9.1	10.1	0.0	4.6	3.1	2.7			-
	SC	6.4	5.0	0.0	0.0	2.6	5.9	4.3	5.2	-
	BAC	15.8	14.3	16.4	15.9	49.0	35.8			-
	NSCO					4.3	7.6	12.1	12.3	-
	Overall	8.8	9.7	5.3	5.4	23.2	16.9	12.7	12.9	11.9
# of overdiagnosed cases (O)		174	242	36	45	559	560	187	233	254.44
# of prevented deaths (D)		495	507	137	205	865	929	611	707	556.99
Mortality reduction		10.59	9.83	5.29	7.9	36.07	27.7	18.1	19.3	16.85
D/O		2.84	2.10	3.85	4.52	1.55	1.66	3.27	3.04	2.85
		NLST-derived								
		Stanford		MGH		FH		Erasmus		Model
		F	M	F	M	F	M	F	M	Average
	AD	7.3	7.9	4.5	4.4	22.8	16.2	12.2*	12.1*	-
Overdiagnosis	SQ	7.0	8.4	2.9	5.6	5.4	14.4	9.7	12.5	-
	LC	7.8	7.3	0.0	3.2	2.0	2.3			-
	SC	5.1	5.1	0.0	0.0	2.1	4.9	3.6	4.5	-
	BAC	14.2	11.0	15.4	12.4	32.9	27.6			-
	NSCO					3.5	6.5	10.3	10.8	-
	Overall	8.0	7.8	4.4	4.4	17.6	12.9	10.8	11.3	9.7
# of overdiagnosed cases(O)		138	173	26	32	366	383	132	176	178.14
# of prevented deaths (D)		445	465	118	178	770	825	527	634	495.17
Mortality reduction		9.52	9.02	4.57	6.84	30.89	23.86	15.6	17.3	14.70
D/O		3.22	2.69	4.59	5.61	2.10	2.15	4.01	3.61	3.50
Differences of overdiagnosis rates (USPSTF – NLST)		0.8	1.9	0.9	1.0	5.6	4.0	1.9	1.6	2.21
Differences of D/O (NLST-USPSTF)		0.38	0.59	0.74	1.09	0.55	0.49	0.74	0.57	0.64
Differences of D/O (NLST-USPSTF) Per 1,000 overdiagnsed cases		380	593	741	1095	553	491	736	570	645

Table S3: The range of D/O ratios of 576 scenarios by model and gender

	Male	**Female**
Erasmus	2.43-4.55	2.38-4.66
Fred Hutch	1.09-3.21	0.86-3.33
MGH	3.26-7.01	2.84-6.49
Stanford	1.44-3.05	2.10-3.75

References

1. Patz EF, Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Internal Medicine* 2014; **174**(2): 269-74.
2. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
3. ten Haaf K, van Rosmalen J, de Koning HJ. Lung Cancer Detectability by Test, Histology, Stage, and Gender: Estimates from the NLST and the PLCO Trials. *Cancer Epidemiology Biomarkers & Prevention* 2015; **24**(1): 154-61.
4. Hazelton WD, Goodman G, Rom WN, et al. Longitudinal multistage model for lung cancer incidence, mortality, and CT detected indolent and aggressive cancers. *Mathematical Biosciences* 2012; **240**(1): 20-34.
5. McMahon PM, Kong CY, Johnson BE, et al. Estimating long-term effectiveness of lung cancer screening in the Mayo CT screening study. *Radiology* 2008; **248**(1): 278-87.
6. McMahon PM, Kong CY, Weinstein MC, et al. Adopting helical CT screening for lung cancer: potential health consequences during a 15-year period. *Cancer* 2008; **113**(12): 3440-9.
7. McMahon PM, Kong CY, Johnson BE, et al. The MGH-HMS Lung Cancer Policy Model: Tobacco Control versus Screening. *Risk Analysis* 2012; **32**(Suppl 1): S117-S24.
8. Kong CY, McMahon PM, Gazelle GS. Calibration of disease simulation model using an engineering approach. *Value Health* 2009; **12**(4): 521-9.
9. Lin RS, Plevritis SK. Comparing the benefits of screening for breast cancer and lung cancer using a novel natural history model. *Cancer Causes & Control* 2012; **23**(1): 175-85.
10. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.
11. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
12. Holford TR, Clark L. Chapter 4: Development of the Counterfactual Smoking Histories Used to Assess the Effects of Tobacco Control. *Risk Analysis* 2012; **32**: S39-S50.

Chapter 7

Lung cancer screening: latest developments and unanswered questions

Carlijn M. van der Aalst, Kevin ten Haaf, Harry J. de Koning

Published as:

van der Aalst CM, ten Haaf K, de Koning HJ.

The Lancet Respiratory Medicine 2016; **4**(9): 749-61.

Abstract

The U.S. National Lung Screening Trial showed that individuals randomly assigned to screening with low-dose CT scans had 20% lower lung cancer mortality than those screened with conventional chest radiography. On the basis of a review of the literature and a modeling study, the U.S. Preventive Services Task Force recommends annual screening for lung cancer for individuals aged 55–80 years who have a 30 pack-year smoking history and either currently smoke or quit smoking within the past 15 years. However, the balance between benefits and harms of lung cancer screening is still greatly debated. The large number of false-positive results and the potential for overdiagnosis are causes for concern. Some investigators suggest the ratio between benefits and harms could be improved through various means. Nevertheless, many questions remain with regard to the implementation of lung cancer screening. This paper highlights the latest developments in CT lung cancer screening and provides an overview of the main unanswered questions.

Introduction

Lung cancer is the most important tobacco-related health problem worldwide, accounting for an estimated 1.3 million deaths each year, representing 28% of all deaths from cancer.[1] Only 15% of patients with clinically detected lung cancer are still alive after 5 years, mainly because of the advanced stage at diagnosis.[2] Smoking cessation is considered the most effective method of prevention for current smokers, as 90% of lung cancers can be attributed to smoking.[3] Although smoking is becoming less prevalent in many developed countries, further reduction of smoking-related lung cancer mortality will take decades. Moreover, smoking prevalence rates are still high in developing countries, and currently half of all lung cancers are diagnosed in former smokers.[4] Therefore, action is needed at both primary and secondary prevention levels.

Lung cancer screening aims to reduce lung cancer related mortality with relatively limited harm through early detection and treatment. Years can pass between the onset of lung cancer and when the tumor presents itself through obvious signs and symptoms, by which time it is often in an advanced state.[5] Previous research sought to determine whether chest radiography screening—often combined with sputum cytology—could reduce lung cancer mortality, but no statistically significant reduction was found.[6] However, the introduction of low-dose CT (LDCT) renewed interest in lung cancer screening. Single-arm studies showed a stage shift, because 48–85% of screen-detected lung cancers were detected at stage I—with more favorable treatment options—compared with 30–35% of clinically detected lung cancers.[7] Therefore, several randomized controlled trials were started to investigate the cost-effectiveness and efficacy of lung cancer screening with LDCT (Tables 1, 2).

In November 2011, the National Lung Screening Trial (NLST) in the U.S. showed a 20.0% reduction in lung cancer mortality for LDCT compared with chest radiography screening, in addition to a 6.7% all-cause mortality reduction.[8] Several medical associations issued guidelines which mostly adopted the NLST criteria for the inclusion of high-risk participants for LDCT screening. The United States Preventive Services Task Force (USPSTF) requested an independent review and an investigation of the long-term harms and benefits of different screening policies.[9] On the basis of this review and modeled screening policies, in December

2013, the USPSTF recommended LDCT screening for lung cancer; annual screening for men and women aged 55–80 years with a smoking history of at least 30 pack-years, who currently smoke or quit smoking within the past 15 years, provides a beneficial trade-off between benefits and harms.[10,11] This recommendation implies that over 10 million individuals in the U.S. would be eligible for lung cancer screening.[8] In early 2015, the recommendation to screen for lung cancer in high-risk individuals was also adopted in the Final Coverage Decision of the Centers for Medicare and Medicaid Services under the Affordable Care Act.[12,13] Although CT lung cancer screening is being implemented in the U.S., the debate about the balance between the benefits and harms is still ongoing. This paper aims to give an overview of the latest developments in lung cancer screening and to discuss unanswered questions.

Search strategy and selection criteria

We searched PubMed, Embase, MEDLINE, and Cochrane Library for relevant studies published in English from June 29, 2011 (the date of the online publication of the NLST mortality reduction) to April 1, 2016. References of included articles were also checked for further relevant articles. The initial search included the following Medical Subject Headings terms or words in the title or abstract: “lung cancer”, “cancer screening”, “screening”, and “early detection”. All titles were checked for relevance, but we focused primarily on studies published in 2015 or 2016 to provide an overview of the latest developments. We included only original articles or reviews with access to the full text as potentially relevant. Editorials, personal views, columns, comments, and letters to the editor were not included. The included articles had to be relevant to lung cancer screening in an adult population at high risk for developing lung cancer. Studies of biomarkers, genetics, and work-up after diagnosis or treatment were excluded. We selected the most relevant and recent articles for this Review that described the latest developments and unanswered questions (68 of 496 identified papers were ultimately included).

Table 1: Large-scale randomized controlled lung cancer screening trials

	Sample size	Recruitment	Follow-up (years)	Comparison (nodule measurement)	Smoking history	Smoking cessation	Age-group (years)	Screening interval
NLST (U.S., 2002)[14]	53,454	Volunteers	6.4*	CT vs chest radiography (D)	≥30 pack-years	<15 years	55-74	Three annual screenings
NELSON (Netherlands/ Belgium, 2004)[15]	15,822	Population-registry	7.4*	CT vs usual care (V)	≥15 cigarettes per day for ≥25 years or ≥10 cigarettes per day for ≥30 years	≤10 years	50-75	Four screenings with different screening intervals: 1 year, 2 years, and 2.5 years
DLCST (Denmark, 2004)[16]	4,104	Volunteers	9.8*	CT vs usual care (V)	≥20 pack-years	≤10 years	50-70	Five annual screenings
MILD (Italy, 2000 and 2005)[17]	1,035 (pilot) and 4,099 (main study)	Volunteers	4.4*	CT vs usual care (V)	≥20 pack-years	≤10 years	≥49	Five annual vs three biennial screenings
UKLS (U.K., 2011-2012)[18]	4,055	Population-registry	-	CT vs usual care (V)	Predicted risk of ≥5% of lung cancer diagnosis within 5 years, and 5-year risk for lung cancer	≤10 years	50-75	One screening
LUSI (Germany, 2007)[19]	4,052	Population	3-6.5	CT vs usual care (D)	≥15 cigarettes per day for ≥25 years or ≥10 cigarettes per day for ≥30 years	≤10 years	50-69	Four annual screenings
ITALUNG (Italy, 2003)[20]	3,206	General practitioners	6	CT vs usual care (D)	≥20 pack-years	≤10 years	55-69	Four annual screenings
DANTE (Italy, 2005)[21]	2,450	Volunteers	8.4	CT vs clinical review (D)	≥20 pack-years	≤10 years	60-74	Four annual screenings

Abbreviations: D=diameter. V=volume.

**Median.*

Dates provided are trial start dates.

Table 2: Overview of latest data from randomized controlled CT lung cancer screening trials

	High referral protocol	Low referral protocol with published mortality analyses			Low referral protocol with no published mortality analyses			
	NLST (n=53,454)[22]	DLCST (n=4,104)[23,24]	DANTE (n=2,450)[25]	MILD annual and biennial (n=4,099)[11,17]	NELSON (n=15,822)[26,27]	LUSI (n=4,052)[28]	ITALUNG (n=3,206)[29]	UKLS (n=4,055)[30]
Age at T0, mean (SD) or median (IQR)	61 (5)	58 (5)	64 (IQR 5)	Annual 57*; Biennial: 58*	59 (IQR 6)	58 (IQR 5)	61 (4)	67 (4)
Current smokers at T0 (%)	48%	75%	57%	Annual: 69%; Biennial: 68%	55%	61%	65%	39%
Pack-years at T0, median (IQR)	48 (27)	36 (13)	47 (30)	Annual: 39*; Biennial: 39*	42 (19)	36 (18)	43 (18)	-
Gender (s)								
Male	59%	56%	100%	Annual 68%; Biennial: 69%	84%	66%	64%	75%
Female	41%	44%	0%	Annual: 32%; Biennial: 31%	16%	34%	36%	25%
Follow-up years (mean/median or total years)	../6.4	9.5/9.8	../8.4	../4.4	../7.4	3-6.5	6	..
Person-years of follow-up (screen arm)	144,103	19,439	10,875	Annual: 5,557; Biennial: 5,5517	..	..	..	..
Lung cancer detection rate (%)†								
T0	1.0%	0.8%	2.3%	0.8%	0.9%	1.1%	1.5%	1.7%
T1	0.7%	0.6%	..	0.5%	0.8%	0.6%	0.4%	..
T2	0.9%	0.7%	..	..	1.1%	0.5%	0.7%	..
T3	..	0.6%	..	..	0.8%	0.4%	0.5%	..
T4	..	0.9%	..	..	..	0.4%	..	..
CT screen-detected cases of lung cancer	649	69‡	66	Annual: 29; Biennial: 20	255	58	41	42
Total lung cancers CT arm	1060	100	104	Annual: 34; Biennial: 25	307	62	43	42
Histology of screen-detected lung cancers; n/N (%) §								
Adenocarcinoma	258/649 (40%)	48/69 (70%)	44/104 (42%)	Annual: 15/29 (52%); Biennial: 17/20 (85%)	130/255 (51%)	45/62 (73%)	27/41 (66%)	25/42 (60%)
Squamous-cell carcinoma	136/649 (21%)	7/69 (10%)	25/104 (24%)	Annual: 10/29 (34%); Biennial: 1/20 (5%)	44/255 (17%)	10/62 (16%)	6/41 (15%)	12/42 (29%)
Small-cell carcinoma	49/649 (8%)	3/69 (4%)	9/104 (9%)	..	11/255 (4%)	3/62 (5%)	3/41 (7%)	3/42 (7%)
Others	203 /649 (31%)	11/69 (16%)	26/104 (25%)	Annual: 4/29 (14%); Biennial: 2/20 (10%)	70/255 (27%)	4/62 (6%)	5/41 (12%)	2/42 (5%)
Missing	3/649 (0.5%)	..	..	..	..	..	..	..
Stage of screen-detected lung cancers; n/N (%) ¶								
IA	329/649 (51%)	37/69 (54%)	31/104 (30%)	Annual: 17/29 (59%); Biennial: 11/20 (55%)	159/255 (62%)	32/58 (56%)	21/41 (51%)	26/42 (62%)
IB	71/649 (11%)	10/69 (14%)	16/104 (15%)	Annual: 1/29 (3%); Biennial 3/20 (15%)	17/255 (7%)	10/58 (17%)	4/41 (10%)	2/42 (5%)
IIA	26/649 (4%)	..	7/104 (7%)	Annual: 1/29 (3%); Biennial: ..	14/255 (5%)	3/58 (5%)	1/41 (2%)	7/42 (17%)
IIB	20/649 (3%)	..	..	Annual: 1/29 (3%); Biennial 1/20	7/255 (3%)	3/58 (5%)	3/41 (7%)	..

				(5%)				
IIIA	59/649 (9%)	10/69 (14%)	9/104 (9%)	Annual: 4/29 (14%); Biennial: 1/20 (5%)	33/255 (13%)	6/58 (10%)	2/41 (5%)	3/42 (7%)
IIIB	49/649 (8%)	3/69 (4%)	8/104 (8%)	Annual: ..; Biennial: 1/20 (5%)	9/255 (4%)	1/58 (2%)	3/41 (7%)	1/42 (2%)
IV	81/649 (12%)	6/69 (9%)	26/104 (25%)	Annual: 5/29 (17%); Biennial: 3/20 (15%)	16/255 (6%)	3/58 (5%)	4/41 (10%)	3/42 (7%)
Missing	14/649 (2%)	3/69 (4%)	7/104 (7%)	..	..	..	3/41 (7%)	..
Interval lung cancers	44	32	38	Annual: 5; Biennial: 5	52	4	2	..
Positive screening tests (%)								
Overall	24.2%	3.8%	37.3%	..	2.0%	8.8%	19.6%	5.7%
At least one positive test	39.1%	..	..	..	6.0%||	..	52.7%	..
T0	27.3%	8.7%	..	Annual: 14.0%; Biennial: 15.0%	2.6%	22.2%	30.3%	5.7%
T1	27.9%	2.3%	..	..	1.8%	4.7%	17.3%	..
T2	16.8%	2.7%	..	..	2.4%	4.0%	16.1%	..
T3	..	2.2%	..	..	2.0%	5.7%	13.7%	..
T4	..	2.8%	..	..	..	5.7%	..	..
False-positives (%)								
Of all scans	17,497/75,126 (23.3%)	302/9,800 (3.1%)	..	..	355/29,735 (1.2%)	747/9,121 (8.2%)	1,003/5,333 (18.8%)	72/1,994 (3.6%)
Of all positive scans	17,497/18,146 (96.4%)	302/371 (81.4%)	..	..	355/598 (59.4%)	747/805 (92.8%)	1,003/1,044 (96.1%)	72/114 (63.2%)
Per participant basis	..	..	22.9%	..	273/7,582 (3.6%)||	..	..	..
Contamination (%)	4.3%	2.8%	..	..	3.1%	..	..	..
Incidental findings (%)	7.5%	6.8%	..	..	6.7%	..	..	..
Lung cancer mortality, RR (95% CI)	0.80 (0.73-0.93)	1.03 (0.66-1.60)	0.99 (0.69-1.43)	Annual: 1.98 (1.57-2.50); Biennial: 1.99 (0.80-4.96); Annual+Biennial: 1.64 (0.73-4.01)**	..	..	..	..
All-cause mortality, RR (95% CI)	0.93 (0.86-0.99)	1.02 (0.82-1.27)	0.95 (0.77-1.17)	Annual: 1.80 (1.03-3.13); Annual+Biennial: 1.40 (0.82-2.38)**	..	..	..	..

Abbreviations: RR=relative risk. T0=baseline. T1=first incidence screening. T2=second incidence screening. T3=third incidence screening.

**IQR not applicable. **Adjusted for age and smoking. †Mean lung cancer detection rate over T1– T4 was 2.9% for DANTE. ‡Contamination rate within first 24 months after baseline screening (detailed overview is about 69 screen-detected lung cancers, although latest data show 68 screen-detected lung cancers and 32 interval cancers, but without further details). §Data about both screen-detected as well as interval cancers in the screen arm for DANTE and LUSI. ¶Data about both screen-detected as well as interval cancers in the screen arm for DANTE. ||Data from round 1, round 2, and round 3.*

Benefits of lung cancer screening

Lung cancer mortality

Tables 1 and 2 summarize the basic characteristics and main outcomes of randomized controlled lung cancer screening trials from the U.S. and Europe since 2000. In 2011, the NLST reported a 20% lung cancer-related mortality reduction (relative risk [RR] 0.80, 95% confidence interval [CI] 0.73–0.93) for CT screening compared with chest radiography screening.[8] Modeling studies that used NLST data suggest that more than 12,000 lung cancer deaths per year in the U.S. are preventable.[8,10,31] The NLST compared CT screening with chest radiography screening, which probably had a small effect on mortality reduction, but the effects of CT screening compared with an unscreened population were estimated in the comparative modeling analyses for the USPSTF.[5]

Some European trials have reported conflicting results, although these trials had ratings of fair (DLCST and DANTE) and poor quality (MILD).[11,17,23,25] Comparing CT screening with no screening, the DLCST reported an RR for lung cancer mortality of 1.03 (95% CI 0.66–1.60), DANTE reported an RR of 0.99 (95% CI 0.69–1.43), and MILD reported an RR of 1.99 (95% CI 0.80–4.96).[17,23,25] Importantly, these trials had insufficient statistical power to show a reduction in lung cancer mortality in terms of sample sizes; these three studies combined have a smaller number of person-years at follow-up than did the NLST or NELSON.[22,27] Recent post-hoc analyses of the DLCST showed more favorable, but statistically insignificant, mortality outcomes in high-risk participants in the screening group compared with the control group.[23] More evidence of the effectiveness of CT lung cancer screening (in subgroups) might be obtained from pooled analyses of the European trials (n=38,793; NELSON, DLCST, ITALUNG, MILD, DANTE, UKLS). However, such analyses require agreement on aspects of quality assurance, such as proper randomization, appropriate follow-up in both study arms, and proper and masked determination of outcome measures. Until these aspects are thoroughly evaluated for each trial, no decision can be made on which trials should be pooled. Assessment of the screening performance of each individual trial would be possible through comparative data analyses and modeling of the trials, and could be related to the potentially associated outcome (e.g., mortality reduction). Such an analysis has already been performed with the NLST, but could still be useful for subgroups (several

groups related from the high-risk group to the screened population).[10] Post-hoc analyses of the NLST showed weak evidence for differences in lung cancer mortality reduction by sex.[32] In addition, a smaller lung cancer mortality reduction was found after an extended follow-up, but this effect could have been because of a more selective follow-up.[32]

Therefore, the awaited results of NELSON (trial number: ISRCTN63545820), which has sufficient statistical power to detect a lung cancer mortality reduction of at least 25%, are needed to provide a clear picture. Although NELSON has a substantially smaller sample size than did NLST, it encompasses a longer follow-up, no scheduled screening in the control arm, four screening rounds with different intervals, and a screening protocol that was substantially the same between rounds. NELSON has also been shown to be generalizable to the high-risk group in the general Dutch population, whereas a healthy-volunteer effect could be present in NLST.[33,34] Moreover, higher proportions of stage I lung cancers and adenocarcinomas were found in European trials than in the NLST (Table 2). A large part of the mortality reduction found in the NLST might be caused by the adenocarcinomas, and detection time and CT sensitivity might be increased for early stages of adenocarcinomas.[5,32]

Furthermore, lung cancer detection rates were roughly comparable between all trials to date. The UKLS, which used a risk prediction model, reported a slightly higher detection rate (1.7%) than that which was reported by the NLST.[30] The DANTE trial reported a 2.3% lung cancer detection rate at baseline, which was high compared with the other trials, but could possibly be explained by a potential preselection of male volunteers.[25] Thorough investigation of follow-up data is warranted before implementation of LDCT screening to allow detailed exploration of the effect of these differences on lung cancer mortality, as the implication is that lung cancer screening in Europe could have different effects compared with the U.S. The policy recommended by the USPSTF would lead to 50.5% (range 48.1–56.9%) of lung cancers being detected at an earlier stage (stage I–II) compared with 37.4% of early-stage cancers in a situation without screening. 575 screenings would have to be performed per lung cancer death averted, yielding a 14.0% (8.2–23.5) reduction in lung cancer-specific mortality and 5,250 life-years gained per 100,000 people.[10] Although all trials were underpowered to show a statistically significant effect on overall mortality reduction, the modeling analyses suggest an average of 10 life-years gained for each lung cancer death

prevented (corrected for increased other-cause mortality in smokers and former smokers).[10] An estimated 8.4 million individuals met the eligibility criteria for lung cancer screening as proposed by the USPSTF in 2013.[31] However, the potential screening-eligible population was older, had a higher proportion of current smokers, and had more comorbidities than did the NLST population, suggesting that the effects of lung cancer screening in the general population should be carefully examined.[35]

Cost-effectiveness

Recent systematic reviews indicated that cost-effectiveness estimates for LDCT screening for lung cancer range from U.S. $18,452 to $66,480 per life-year gained and $27,756 to $243,077 per quality-adjusted life-year (QALY) gained.[36,37] The NLST estimated its cost-effectiveness (CT vs chest radiography) at U.S. $52,000 per life-year gained and U.S. $81,000 per QALY gained.[38] A recent modeling study suggests a cost of CAD $52,000 per QALY gained, which could even be improved by, for example, incorporation of a smoking cessation program.[39] In addition, the UKLS estimated a cost-effectiveness ratio of £8,466 per QALY gained (95% CI 5,542–12,569) based on the baseline screening round.[30] Most studies, perhaps unexpectedly, reported cost-effectiveness estimates below the threshold of U.S. $100,000 per QALY gained.

However, the threshold value for cost-effectiveness of a cancer screening program varies across countries, with some program not even defining a threshold. This variation causes uncertainty about the acceptability of lung cancer screening in European countries, because the U.S. threshold of $100,000 is not likely to be considered acceptable. According to the WHO, an intervention that costs less than three times the national annual per capita gross domestic product is considered cost-effective. However, cost-effectiveness depends on the lung cancer risk of those being screened. Therefore, potential harms—such as the high positive and false-positive rate found in the NLST—might not be acceptable in European countries, both in terms of emotional burden and the high demand on health resources and costs. This problem suggests that the selection of people eligible for lung cancer screening can or should be based on an individual's risk, for example, through the application of risk-

prediction models, thus improving the cost-effectiveness ratio.[40] In this way, implementation of cost-effective programs seems to be feasible.

Risks of CT lung cancer screening

Radiation risk

Radiation-induced carcinogenesis is one of the potential harms of CT lung cancer screening. An individual who attends annual screening could receive as many as 25 CT examinations. Some of these individuals will also be exposed to additional radiation due to work-up procedures. However, in the NLST, screened individuals received an estimated 8 millisievert (mSv) in a 3-year period, which corresponds to potentially causing one radiation-induced cancer death per 2,500 individuals screened in a 10–20-year period.[41] Moreover, the radiation dose could be reduced to an exposure comparable to chest radiography.[42] Further improvements in CT imaging could aid further reductions in radiation doses.

False-positives

One or more benign or malignant lung nodules will be found in about half of individuals screened for lung cancer, although only a few nodules actually represent lung cancer after (invasive) follow-up. False-positive screening test results are a major concern because of the potential for subsequent unnecessary diagnostic work-up procedures, such as a biopsy sampling, bronchoscopy, or lung surgery; these procedures have accompanying risks of complications, negative effects on health-related quality of life, and increased health-care costs. In the NLST, 39.1% of the participants received at least one positive test result across the three annual CT screens (Table 2). Of those with a positive test, 96.4% were found to be false-positive, of which 72.1% required referral to the pulmonologist for further (invasive) work-up. Overall, 23.3% of all CT screens in the NLST were false-positive and only 3.6% of the CT screens led to a diagnosis of lung cancer, with 2.7% of the participants with a false-positive screening test result facing complications after undergoing work-up.[8,14,43] In NELSON, 59.4% of those with an initial positive test in four screening rounds were found to be false-positive, leading to an overall low false-positive rate of 1.2% (Table 2).[26,27] The main reason for the lower false-positive rate in NELSON compared with the NLST was the difference in nodule management—i.e., volume measurement and the use of an

indeterminate screening test result (9.3% [2,503 of 29,733] of all regular round scans). In particular, NELSON had a higher threshold for a positive screening test result than did the NLST. Participants of the NELSON trial with an indeterminate screening test result received a letter, which stated the presence of a "very small abnormality" in the lung and that there was no immediate need for further investigations, suggesting a follow-up CT scan of the lungs after 3–4 months to monitor for any changes.[44] This message is assumed to have less impact than an immediate referral, as is the case with a positive screening result. However, no method is known to reduce the false-positive rate without loss of sensitivity for detection of lung cancer. Therefore, a better distinction between benign and malignant nodules is required. In 2014, the quality assurance tool Lung-RADS (American College of Radiology; Reston, VA, USA) was designed to standardize CT lung cancer screening with a set of recommendations for interpretation, reporting, and management of lung nodules, and to facilitate outcome monitoring for screen-detected pulmonary nodules based on the highest expected malignancy of the nodules. The main difference with the NLST protocol was the increased threshold (from 4 mm to 6 mm diameter) for a positive baseline screening test and the use of growth for pre-existing nodules.[45] Retrospective analyses of NLST data showed that the use of Lung-RADS criteria could reduce the false-positive rate (baseline screening: 26.6% [NLST] vs 12.8% [Lung-RADS]; incidence screening rounds: 21.8% vs 5.3%) at the cost of sensitivity (baseline screening: 93.5% vs 84.9%; incidence screening rounds: 93.8 vs 78.6).[45] In 2015, the British Thoracic Society also issued renewed guidelines for the management of pulmonary nodules that incorporates both nodule size and volume or volume-doubling time.[46]

Overdiagnosis and overtreatment

An overdiagnosed lung cancer refers to a screen-detected lung cancer that would have never been clinically detected if screening had not occurred. Consequences of overdiagnosis are unnecessary diagnostic work-up procedures and even cancer treatment, with all its physical consequences, negative psychological effects (e.g., anxiety, distress), and health-care costs. Both trials and modeling studies are needed to determine the degree of overdiagnosis in lung cancer screening.[47] Use of trial data alone is flawed owing to the fixed design of trials (e.g., specific study population and screening protocol, short duration of

follow-up). Modeling can be used to extrapolate trial data to many different screening scenarios; however, the model must be able to incorporate the natural history and screen-detectability of lung cancer as well as birth cohort data on relevant information about risk factors (smoking behavior) and other-cause mortality (by age, sex, smoking history). Excess incidence analyses based on the NLST suggest that 18.5% of all lung cancers detected by CT screening were overdiagnosed.[48] Estimates from microsimulation modeling suggest that 8.6% of lung cancers detected by CT were over diagnosed in the NLST.[47] Furthermore, an estimated 9.9% of all lung cancers detected by screening are overdiagnosed in the screening policy as recommended by the USPSTF, which is relatively low compared with other screening programmes.[10,40]

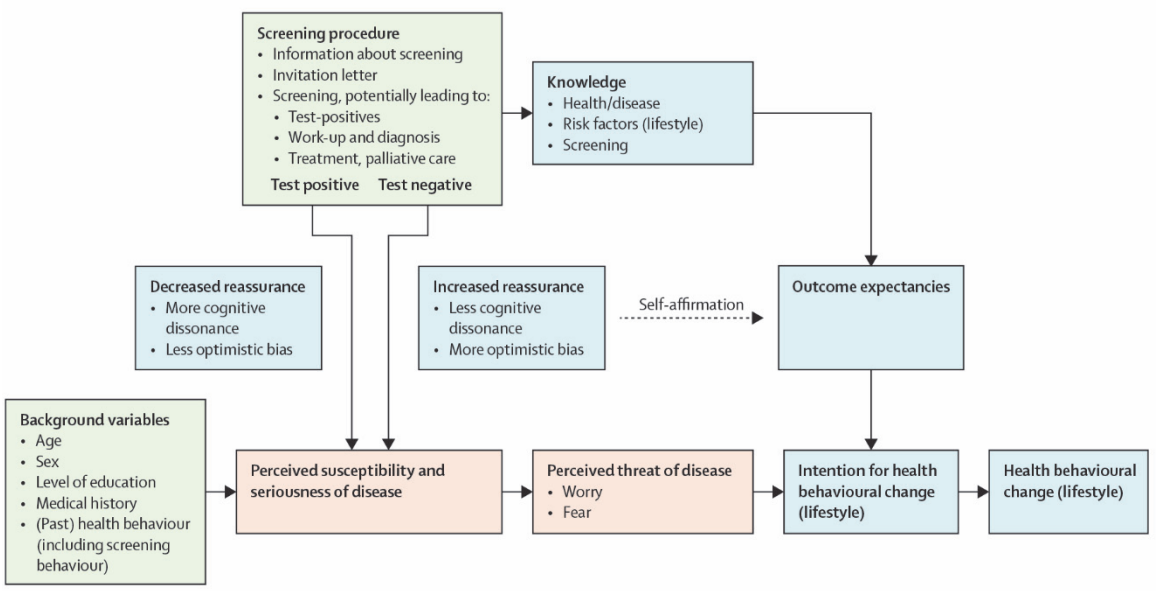
Psychological consequences

Lung cancer screening affects health-related quality of life, but estimates of the magnitude of these effects are scarce because of low lung cancer survival. False-positives and overdiagnosed cases of lung cancer could have negative effects on quality of life.[49] Some trials (NLST, NELSON, DLCST) reported no or only temporary short-term effects on health-related quality of life, without long-term consequences.[50-52] However, parts of the screening process (invitation, screening test, awaiting the screening test result, receiving the screening result, work-up) will definitely affect people in terms of anxiety and lung cancer-specific distress.

In a hypothetical pathway of the psychological impact of screening (Figure 1), perceived susceptibility to the risk of lung cancer and subsequent anxiety and distress—which are predictors of health behavior in current models (e.g., the Health Belief Model)—increase because of increased attention from invitation to screening until the moment of screening (Figure 2).[53] A negative screening test result might then lead to a feeling of reassurance, whether realistic or not, that could disproportionally decrease people's perceived susceptibility to the perceived threat. The feeling of reassurance might cause people to unrealistically underestimate their susceptibility and the influence of their lifestyle on developing the disease. People might compensate for their screening-induced reassurance by continuing current risk behavior or by reducing preventive behavior (risk compensation

or homoeostasis). A positive screening test result could enhance the elevated perceived susceptibility. However, evidence about the working mechanism is needed. An option to reduce potential negative psychological consequences is to provide participants with sufficiently useful information (information about risk factors, screening procedures, the screening result) in the right form at the right time (e.g., personal vs written information before screening, before the screening test result, before follow-up procedures, and after diagnosis). Nonetheless, more than 300 individuals are estimated to need screening to avert one lung cancer death, meaning many screened individuals will potentially be exposed to unfavorable psychological effects.[8] Future research should investigate whether the effect of screening on health-related quality of life in the general population is comparable to that in trial participants.

Figure 1: Hypothetical pathway of the psychological effect of screening on health behavioral change



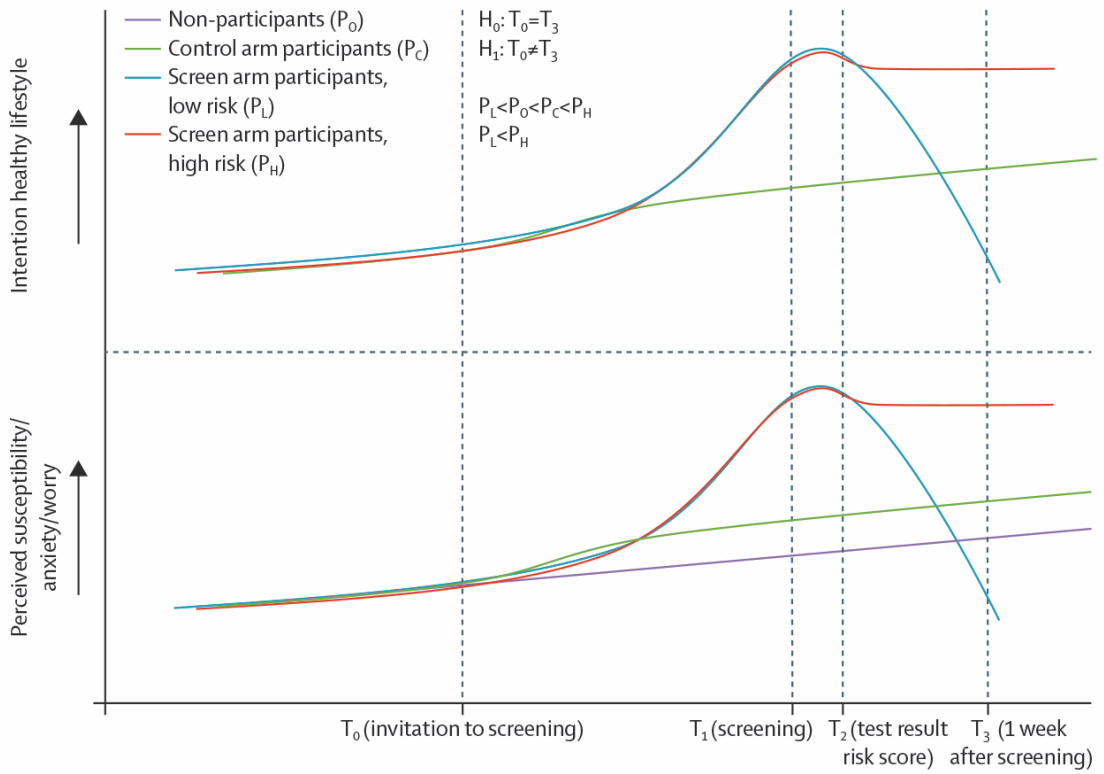
The screening process is hypothesized to have psychological consequences, with different levels of reassurance depending on positive or negative test results. These varying levels can lead to unrealistic estimations of susceptibility to the disease and the influence of lifestyle on development of the disease, which affects (the intention towards) health behavioral change.

The optimal screening policy

Screening population

Who should be selected for screening to further optimize the harm–benefit ratio? In the U.S., an estimated 26% of patients with lung cancer and 8.6 million of the general population would have met the NLST eligibility criteria.[54] Notably, data on non-white populations are scarce. Furthermore, although CT screening trials have not examined never-smokers, model analyses show that never-smokers at increased risk of development of lung cancer require a relative risk of at least 15–20 compared with never-smokers at average risk of development of lung cancer to achieve the harm–benefit ratio of the USPSTF eligibility criteria. However, analyses based on the Prostate Lung Colorectal Ovarian (PLCO) cancer screening trial suggest that most never-smokers would not reach the level of risk required to benefit from screening.[55,56] Risk-based analyses based on the NLST showed that 60% of the participants accounted for 88% of the averted lung cancer deaths and for 64% of people with false-positive screening test results. These results highlight the importance of risk-based selection to enhance the efficacy and cost-effectiveness of lung cancer screening, which other research supports.[57,58]

Figure 2: Hypothetical pathway of teachable moments for smoking cessation during the screening process



The null hypothesis (H0: T0 = T3) implies that screening will not lead to a significant increase in perceived susceptibility, anxiety, and worry. This situation would lead to no increase or even a decrease in motivation to adopt a healthy lifestyle (smoking cessation). The alternative hypothesis implies that screening will lead to an increase in perceived susceptibility, anxiety, and worry from T0 until T2 (teachable moments), resulting in increased motivation to adopt a healthy lifestyle.

The PLCOm2012 model included a number of risk factors to select individuals at high risk for development of lung cancer.[57] Compared with the NLST criteria (age, pack-years, and smoking quit years), the criteria derived from PLCOm2012 (age, race, education, body-mass index, chronic obstructive pulmonary disease [COPD], personal history of cancer, family history of lung cancer, smoking status, smoking intensity, smoking duration, and smoking quit time) had a significantly improved sensitivity (83.0% vs 71.1%, p <0.001) and positive predictive value (4.0% vs 3.4%, p = 0.01) for selection of individuals for lung cancer screening, without loss of specificity (62.9% vs 62.7%, p = 0.54).[57] Moreover, the UKLS used the Liverpool Lung Project (version 2) model to stratify a very high-risk population (5-year lung cancer risk >5%), leading to the highest lung cancer detection rate (1.7%) in a population-based CT lung cancer screening trial.[30] Based on Continuous Observation of Smoking Subject (COSMOS) data, incorporation of baseline nodule characteristics and the

presence of emphysema in the Bach model were found to help to predict lung cancer in subsequent screening rounds. Thus, risk-prediction models can assist in lung cancer screening by optimizing the selection of individuals who could benefit the most (i.e., those with the highest probability of developing lung cancer), which should lead to fewer people having to be screened to avert one lung cancer death.

Screening interval

The screening interval has a direct impact on the balance between the benefits and harms. Annual screening has been recommended, as the mortality reduction found in the NLST was based on annual screening.[22] Recent NELSON data showed a statistically significant increase in advanced stage lung cancers (stage IIIB–IV) and more interval cancers in round four compared with the first three rounds, which suggests that an interval of 2.5 years might be too long.[27] Researchers from MILD have suggested that biennial screening could be as effective as annual screening, which was based on comparable lung cancer stage distributions between the annual and biennial screening arms.[59] Modeling analyses suggest that a uniform 2-year interval could lead to difficulties in detection of lung cancer in stage IA.[5] However, retrospective analyses of the NLST suggest that a 1-year interval might not be necessary for participants with a negative baseline screen or subsequent negative screens.[22] The group with negative baseline screen results had a lower lung cancer incidence and mortality compared with all screened participants: the negative baseline results group had lung cancer detection rates of 0.34% (62 of 18,121 participants) compared with 1.02% (267 of 26,231 participants) of all screened participants. These rates were even lower in individuals with subsequent negative screens. If participants with a negative baseline screen had not received the first incidence screening, lung cancer mortality would have hypothetically risen from 186 per 100,000 person-years to 212 per 100,000 person-years. However, this change in first incidence screenings could reduce the number of CT screens for the majority of screening participants by about 73% (19,066 of 26,231 participants), which could reduce negative aspects such as false-positive screening test results and health-care costs.[22] NELSON also found that the probability of a lung cancer diagnosis in the following 2 years after the baseline screen was lowest for participants with a negative baseline screening test result (0.4% [95% CI 0.3%–0.6%]). This probability increased for a

larger volume, diameter, or both, of the largest lung nodule up to a lung cancer probability of 25.7% (95% CI 20.6%–31.6%) for a volume of 1,000 mm^3 or greater, or 31.6% (95% CI 15.2%–54.2%) for a diameter of 30 mm or greater.[60] Although annual screening is more likely to be effective, these results imply that information obtained from the baseline screening can be used to optimize screening intervals for different risk groups.

Nodule management protocol

Clinicians increasingly face challenges with how to approach lung nodules detected as a result of the introduction of LDCT screening. The nodule management strategies used in the different clinical trials differ notably. In the two largest CT lung cancer screening trials (the NLST and NELSON), researchers used different measurement techniques (diameter vs volume and growth) and cutoff points for a positive screening test result.[14,61] NELSON also introduced a third screening test result—the indeterminate screening test result—to reduce the number of false-positive screening test results. After an indeterminate screening test result, participants received a repeat scan after a period ranging from 6 weeks to 3 months to examine changes in nodule volume (i.e., growth). The results suggest that this approach is successful in reducing false-positive screening test results, along with the accompanying reduction in follow-up procedures and psychological impact.[61] Sensitivity and negative predictive value appear to be comparable between volume-based protocols (90.9% [95% CI 81.2%–96.1%]) and diameter-based protocols (90.0% [89.3%–90.7%]), although a higher specificity (94.9% [94.4%–95.4%] vs 90.0% (89.3–90.7]) and positive predictive value (14.4% [11.3%–18.1%] vs 7.9% [6.2%–10.1%]) were found with use of a volume-based protocol. A risk prediction model for lung nodules showed great promise for discrimination between benign and malignant lung nodules, with an area under the curve above 0.90, which suggests a non-linear relationship between nodule size and cancer probability.[60,62] Furthermore, nodules located in the upper lobes had increased probability of being malignant.[62] Analyses from the NLST showed that an increased nodule size threshold (from 5.0 mm to 6.0 mm, 7.0 mm, 8.0 mm, and 9.0 mm) could reduce the number of positive screens (to 10.5%, 7.2%, 5.3%, and 4.2%, respectively) and subsequent work-up (by 33.8%, 54.7%, 66.6%, and 73.8%, respectively).[63] NELSON presented an approach based on nodule size and growth that might reduce the number of false-positive screening test results.[26]

People with nodules smaller than 100 mm³ or 5 mm diameter had comparable lung cancer probability to those without lung nodules, which indicates that these people should not be referred. Individuals with a lung nodule of 100–300 mm³ or 5–10 mm diameter are eligible for growth measurement (volume doubling time) assessment. A volume doubling time of less than 600 days requires further work-up. Nodules greater than 300 mm³ or 10 mm diameter require immediate work-up due to the high probability of lung cancer (>16.9% and >15.2%, respectively).[64] Little evidence suggests that biomarkers can substantially improve the identification of those eligible for lung cancer screening or aid in discrimination of benign nodules from malignant nodules. However, future studies might identify biomarkers with these abilities.

Smoking cessation intervention

Screening is not an alternative to smoking cessation, which is an important message to give to current smokers eligible for screening. Lung cancer screening has been argued to be a teachable moment for smoking cessation services: screened participants have quit smoking more often than have the general smoking population, therefore screening could be used to motivate individuals to quit.[65] Combination of lung cancer screening with a smoking cessation program could further reduce lung cancer mortality. In the NLST, lung cancer specific and all-cause mortality was increased for current smokers (hazard ratio [HR] range 2.14–2.29) compared with former smokers (HR range 1.79–1.85).[66] 7 years of smoking abstinence in smokers who smoked at least 30 pack-years led to a reduction in lung cancer mortality of 20%, which is comparable with three annual CT lung cancer screenings. An optimum was reached after CT lung cancer screening for former smokers who had not smoked for at least 15 years (HR 0.62, 95% CI 0.51–0.76).[66] Moreover, smoking cessation not only affects lung cancer risk but a wide spectrum of tobacco-related health problems, such as cardiovascular diseases and COPD, with similar results found in a modeling study.[3,67] For this high-risk population, insufficient evidence exists regarding the best type of intervention (behavioral, pharmacological, or both), frequency (once vs multiple contacts), modality (personal vs written communication), or content of the communication (e.g., generic, targeted, personalized, tailored, or inter personal).[68-73] Research is needed to explore how to help specific groups of smokers to quit smoking, based on current evidence

of effective smoking cessation services. Counseling is known to be more effective than self-help interventions, and the more intensive the counseling is in terms of duration and frequency, the more effective it is.[74] Moreover, a combination of counseling and medication (nicotine replacement therapy, behavioral medication, or both) is more effective than is a single intervention.[74] One intervention adopted by several medical associations related to lung cancer care is the use of the 5 As (ask, advise, assess, assist, and arrange) for motivated smokers and 5 Rs (relevance, risk, rewards, roadblocks, and repeat) for unmotivated smokers.[75] However, evaluation in a case-control study with 3,336 NLST participants indicates the need for more intensive treatment.[76]

Implementation

Multiple logistical questions about the implementation of lung cancer screening need to be addressed. Are there enough radiologists and other trained individuals to service a screening program? Are there correct quality assurance systems in place to undertake volumetric analysis? Are there correct CT screen-detected nodule clinical work-up practices in place? What is the most appropriate surgical approach? Lung cancer screening should ideally be implemented for high-risk groups only, as indicated by research showing the effectiveness of risk-based selection. The concept of stratified prevention has been widely promoted as part of the discourse on stratified medicine. However, how to scale up risk stratification-based prevention programs has had little—or no—consideration. Largescale implementation requires an evidence-based, organized, and protocolized approach consisting of risk selection and recruitment strategies (including engaging the hard-to-reach high-risk population); training in image acquisition, image quality, and radiological image interpretation of the CT scans; nodule (volumetric) management; multidisciplinary work-up procedures; multidisciplinary oncology treatment of (early-stage screen-detected) lung cancers; work-up of screen-detected treated lung cancers; and quality assessment and defining performance indicators to ensure that screening achieves the greatest benefit with the least harm.[77] Most guidelines have been based on NLST or USPSTF recommendations.

Furthermore, scaling up requires adequate resources in terms of—for example—radiologists, pulmonologists, thoracic surgeons, and screening sites, to ensure that a

national program could be feasible. Decision modeling can help to investigate whether sufficient capacity is available for implementation of national screening program, as done for the implementation of the Dutch colon cancer screening programme.[78] Although volume-based protocols have been shown to be superior to diameter-based protocols, a nodule management protocol based on volume and growth is not yet in common practical guidelines.[60] Evidence-based guidelines should be formulated to provide guidance for current practitioners, who have different levels of expertise in handling lung nodules found by CT screening. Radiologists also need to be informed about the advantages of using volume-based nodule management protocols. The incorporation of software to allow CT scanners to implement volume-based protocols could take some time; hence, a transitional period of using both diameter-based and volume-based protocols might be necessary.

According to the European Code against Cancer, screening with adequate quality is only recommended for cancers for which a life-saving effect has been demonstrated to substantially outweigh the potential harm of examining very large numbers of individuals who might otherwise never have felt the consequences of these cancers. EU citizens are recommended to participate in cancer screening each time an invitation is received, after having read the information materials provided and carefully considered the potential benefits and harms of screening (i.e., informed decision making). The alternative would be a quality assessment system for health check-ups, in which ineligible people should not be advised to be screened. However, the difficulty of this approach lies in the validity of the risk assessment—which is often based on self-reported data—as well as the messages given in screening invitations to the population and the attitude of low-risk, and thus ineligible, people. We could learn from the process of treating early-stage lung cancer, in which information is used to distinguish patients who might benefit from specific treatments and those who will not benefit from specific treatments. Some unanswered questions remain: how to communicate to individuals that they are not eligible for lung cancer screening although their acquaintances might be; what to do when low-risk individuals are clinically diagnosed with lung cancer while being excluded from screening; and what to do when those who were falsely reassured present with signs and symptoms. These problems seem much more difficult to address in a stratified system, in which selection would be primarily based on initial answers, compared with the running of large-scale population-based

screening, in which everyone within a certain age range is invited. Communication of the concepts of low risk, harms, overdiagnosis, comorbidity, and limited added value seem more important than ever.

One concern is that screening in the general population will insufficiently reach certain groups who might benefit the most, such as those with a lower socioeconomic status, minorities, and ethnic groups. For example, McRonald and colleagues found that although current smokers and individuals with lower socioeconomic status were more likely to be at risk for lung cancer than non-smokers and individuals with higher socioeconomic status, they were also less likely to take part in the first stage of recruitment of the UKLS trial.[79] Further research suggests that the main barriers for participation were related to comorbidities and emotional barriers (the latter in particular for current smokers).[80] Data from NELSON showed that the study population is comparable to the Dutch general high-risk population.[33] Therefore, recruitment methods that use population registries to ensure that all inhabitants are approached could also be effective in reaching high-risk groups. The UKLS showed that the use of a risk calculator could be a feasible approach for identification of high-risk groups.[30] The question is whether these high-risk groups include the hard-to-reach groups. However, efforts should be made to measure actual screening uptake in both groups. Strategies on how to reach these groups can improve their screening uptake. For example, information brochures in different languages and recruitment via population registries or general practitioners might be useful. Lung cancer screening trials were conducted in specialized medical centers with sufficient expertise and resources. Whether nationally implemented programs can provide similar levels of quality as achieved in these trials remains unclear.

Incidental findings and detection of other diseases

During LDCT screening, abnormalities other than lung cancer nodules can be found, which can be clinically relevant. These incidental findings might cause effects— both physical (consequences of follow-up procedures) and psychological (anxiety, distress) harms—similar to those resulting from positive screening test results.[81] The NLST reported that incidental findings were found in 7.5% of the screen participants.[8] In a retrospective analysis of a

subset of CT scans from one out of four screening sites, NELSON researchers found incidental findings in 129 (6.7%) of 1,929 screen participants.[82] However, only 21 (1%) of the screen participants in NELSON had incidental findings with clinical implications, because these findings required additional work-up procedures or treatment.[82] In DLCST, incidental findings were found in 140 (7%) of 2,052 screened participants, of which 22 (16%) underwent invasive workup procedures.[83] However, some argue that despite the occurrence of incidental findings, lung cancer screening could be useful in the early detection of other smoking-related diseases, such as cardiovascular diseases and COPD.[84-86] Further research is required on the consequences of incidental findings and the benefits and harms of incorporating screening for other diseases.

Discussion

In many countries, the population at risk for lung cancer will remain high for decades, but CT lung cancer screening could help to reduce the burden of lung cancer mortality. The benefits, harms, and cost-effectiveness of CT lung cancer screening have been assessed—on the basis of one large-scale randomized controlled trial with high referral and false-positive rates—and screening recommendations have been made, partly on the basis of modeling studies. Risk prediction models have the potential to optimize selection of individuals who could benefit most from lung cancer screening. However, the most crucial unanswered question is how to substantially reduce the referral and false-positive rates due to the detection of benign lung nodules without affecting the reductions in lung cancer mortality. Although smaller European trials are insufficiently powered to provide definite answers, the NELSON trial—possibly combined with the pooling of high-quality data from other trials— is likely to provide answers within the next few years. In the meantime, pilot studies with different recruitment strategies will provide information on the feasibility of population-based organized screening programs. Limited evidence is available on how to implement large- scale individualized screening based on risk stratification. Other essential elements to consider for implementation will be quality assurance with regard to providing accurate information to participants about the potential risks, harms, and benefits across all phases of the screening process. Possible unintended side-effects—such as unintentional screening in low-risk groups, false reassurance, and the impact on risk perception and quality of life—

should be carefully monitored. Nonetheless, CT lung cancer screening is likely to become an important element in addressing the burden of lung cancer on global health.

Acknowledgments

We thank a medical writer from Metamorfose Vertalingen who checked the manuscript for spelling and grammar.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* 2015; **136**(5): E359-E86.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians* 2015; **65**(1): 5-29.
3. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Published January 17, 2014. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, Ga.
4. American Cancer Society. Cancer facts and figures 2010. Atlanta: American Cancer Society, 2010.
5. ten Haaf K, van Rosmalen J, de Koning HJ. Lung Cancer Detectability by Test, Histology, Stage, and Gender: Estimates from the NLST and the PLCO Trials. *Cancer Epidemiology Biomarkers & Prevention* 2015; **24**(1): 154-61.
6. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
7. The International Early Lung Cancer Action Program I. Survival of Patients with Stage I Lung Cancer Detected on CT Screening. *New England Journal of Medicine* 2006; **355**(17): 1763-71.
8. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
9. Smith RA, Manassaram-Baptiste D, Brooks D, et al. Cancer screening in the United States, 2015: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA: A Cancer Journal for Clinicians* 2015; **65**(1): 30-54.
10. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
11. 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. *Annals of Internal Medicine* 2013; **159**(6): 411-20.
12. Parker MS, Groves RC, Fowler AA, III, et al. Lung Cancer Screening With Low-dose Computed Tomography: An Analysis of the MEDCAC Decision. *Journal of Thoracic Imaging* 2015; **30**(1): 15-23.
13. Pinsky PF, Gierada DS, Hocking W, Patz EF, Jr, Kramer BS. National lung screening trial findings by age: Medicare-eligible versus under-65 population. *Annals of Internal Medicine* 2014; **161**(9): 627-33.
14. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and Study Design. *Radiology* 2011; **258**(1): 243-53.
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 randomized lung cancer multi-slice CT screening trial (NELSON). *International Journal of Cancer* 2007; **120**(4): 868-74.
16. Pedersen JH, Ashraf H, Dirksen A, et al. The Danish Randomized Lung Cancer CT Screening Trial—Overall Design and Results of the Prevalence Round. *Journal of Thoracic Oncology* 2009; **4**(5): 608-14.
17. 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. *European Journal of Cancer Prevention* 2012; **21**(3): 308-15.
18. Field JK, Devaraj A, Baldwin DR, et al. 66 UK Lung Cancer Screening trial (UKLS): Prevalence data at baseline. *Lung Cancer*; **83**: S24-S5.
19. Becker N, Delome S, Kauczor HU. LUSI: the German component of the European trial on the efficacy of multislice-CT for the early detection of lung cancer. *Onkologie* 2008; 31 (suppl 1): PO320 (abstr).
20. 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.

21. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT. *Lung Cancer* 2008; **59**(3): 355-63.
22. Patz Jr EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomized, multicentre, diagnostic screening trial. *The Lancet Oncology* 2016; **17**(5): 590-9.
23. Wille MM, Dirksen A, Ashraf H, et al. Results of the Randomized Danish Lung Cancer Screening Trial with Focus on High-Risk Profiling. *American Journal of Respiratory and Critical Care Medicine* 2016; **193**(5): 542-51.
24. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomized Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. *Thorax* 2012; **67**(4): 296.
25. Infante M, Cavuto S, Lutman FR, et al. Long-Term Follow-up Results of the DANTE Trial, a Randomized Study of Lung Cancer Screening with Spiral Computed Tomography. *American Journal of Respiratory and Critical Care Medicine* 2015; **191**(10): 1166-75.
26. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *European Respiratory Journal* 2013; **42**(6): 1659.
27. Yousaf-Khan U, van der Aalst C, de Jong PA, et al. Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval. Thorax 2017;72(1):48-56.
28. Becker N, Motsch E, Gross ML, et al. Randomized Study on Early Detection of Lung Cancer with MSCT in Germany: Results of the First 3 Years of Follow-up After Randomization. *Journal of Thoracic Oncology* 2015; **10**(6): 890-6.
29. Pegna AL, Picozzi G, Falaschi F, et al. Four-Year Results of Low-Dose CT Screening and Nodule Management in the ITALUNG Trial. *Journal of Thoracic Oncology* 2013; **8**(7): 866-75.
30. Field JK, Duffy SW, Baldwin DR, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax* 2016; **71**(2): 161-70.
31. Ma J, Ward EM, Smith R, Jemal A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer* 2013; **119**(7): 1381-5.
32. Pinsky PF, Church TR, Izmirlian G, Kramer BS. The National Lung Screening Trial: Results stratified by demographics, smoking history, and lung cancer histology. *Cancer* 2013; **119**(22): 3976-83.
33. van der Aalst CM, van Iersel CA, van Klaveren RJ, et al. Generalisability of the results of the Dutch–Belgian randomized controlled lung cancer CT screening trial (NELSON): Does self-selection play a role? *Lung Cancer* 2012; **77**(1): 51-7.
34. Moyer VA, on behalf of the USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 2014; **160**(5): 330-8.
35. Howard DH, Richards TB, Bach PB, Kegler MC, Berg CJ. Comorbidities, smoking status, and life expectancy among individuals eligible for lung cancer screening. *Cancer* 2015; **121**(24): 4341-7.
36. Raymakers AJN, Mayo J, Lam S, FitzGerald JM, Whitehurst DGT, Lynd LD. Cost-Effectiveness Analyses of Lung Cancer Screening Strategies Using Low-Dose Computed Tomography: a Systematic Review. *Applied Health Economics and Health Policy* 2016; **14**(4): 409-18.
37. Puggina A, Broumas A, Ricciardi W, Boccia S. Cost-effectiveness of screening for lung cancer with low-dose computed tomography: a systematic literature review. *European Journal of Public Health* 2016; **26**(1): 168-75.
38. Black WC, Gareen IF, Soneji SS, et al. Cost-Effectiveness of CT Screening in the National Lung Screening Trial. *New England Journal of Medicine* 2014; **371**(19): 1793-802.
39. Goffin JR, Flanagan WM, Miller AB, et al. Cost-effectiveness of lung cancer screening in canada. *JAMA Oncology* 2015; **1**(6): 807-13.

40. Tammemägi MC. Application of Risk Prediction Models to Lung Cancer Screening: A Review. *Journal of Thoracic Imaging* 2015; **30**(2): 88-100.
41. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of ct screening for lung cancer: A systematic review. *Journal of the American Medical Association* 2012; **307**(22): 2418-29.
42. Huber A, Landau J, Ebner L, et al. Performance of ultralow-dose CT with iterative reconstruction in lung cancer screening: limiting radiation exposure to the equivalent of conventional chest X-ray imaging. *European Radiology* 2016; **26**(10): 3643-52.
43. The National Lung Screening Trial Research T. Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. *New England Journal of Medicine* 2013; **368**(21): 1980-91.
44. van der Aalst CM, van Klaveren RJ, van den Bergh KAM, Willemsen MC, de Koning HJ. The impact of a lung cancer computed tomography screening result on smoking abstinence. *European Respiratory Journal* 2011; **37**(6): 1466.
45. Pinsky PF, Gierada DS, Black W, et al. Performance of Lung-RADS in the National Lung Screening Trial: A retrospective assessment. *Annals of Internal Medicine* 2015; **162**(7): 485-91.
46. Baldwin DR, Callister MEJ. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax* 2015; **70**(8): 794.
47. ten Haaf K, de Koning HJ. Overdiagnosis in lung cancer screening: why modeling is essential. *Journal of Epidemiology and Community Health* 2015; **69**(11): 1035-39.
48. Patz EF, Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Internal Medicine* 2014; **174**(2): 269-74.
49. Slatore CG, Sullivan DR, Pappas M, Humphrey LL. Patient-Centered Outcomes among Lung Cancer Screening Recipients with Computed Tomography: A Systematic Review. *Journal of Thoracic Oncology* 2014; **9**(7): 927-34.
50. Aggestrup LM, Hestbech MS, Siersma V, Pedersen JH, Brodersen J. Psychosocial consequences of allocation to lung cancer screening: a randomized controlled trial. *BMJ Open* 2012; **2**(2).
51. Rasmussen JF, Siersma V, Pedersen JH, Brodersen J. Psychosocial consequences in the Danish randomized controlled lung cancer screening trial (DLCST). *Lung Cancer* 2015; **87**(1): 65-72.
52. van den Bergh KAM, Essink-Bot ML, Borsboom GJJM, Scholten ET, van Klaveren RJ, de Koning HJ. Long-term effects of lung cancer computed tomography screening on health-related quality of life: the NELSON trial. *European Respiratory Journal* 2011; **38**(1): 154.
53. Glanz K, Rimer BK, Lewis FM. Health Behavior and Health Education. Theory, Research and Practice. San Francisco: Wiley & Sons, 2002.
54. Paul FP, Christine DB. Applying the National Lung Screening Trial eligibility criteria to the U.S. population: what percent of the population and of incident lung cancers would be covered? *Journal of Medical Screening* 2012; **19**(3): 154-6.
55. ten Haaf K, de Koning HJ. Should Never-Smokers at Increased Risk for Lung Cancer Be Screened? *Journal of Thoracic Oncology* 2015; **10**(9): 1285-91.
56. Tammemägi MC, Church TR, Hocking WG, et al. Evaluation of the Lung Cancer Risks at Which to Screen Ever- and Never-Smokers: Screening Rules Applied to the PLCO and NLST Cohorts. *PLOS Medicine* 2014; **11**(12): e1001764.
57. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.
58. Kovalchik SA, Tammemägi M, Berg CD, et al. Targeting of Low-Dose CT Screening According to the Risk of Lung-Cancer Death. *New England Journal of Medicine* 2013; **369**(3): 245-54.
59. Sverzellati N, Silva M, Calareso G, et al. Low-dose computed tomography for lung cancer screening: comparison of performance between annual and biennial screen. *European Radiology* 2016; **26**(11): 3821-9.
60. Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *The Lancet Oncology* 2014; **15**(12): 1332-41.

61. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of Lung Nodules Detected by Volume CT Scanning. *New England Journal of Medicine* 2009; **361**(23): 2221-9.
62. McWilliams A, Tammemägi MC, Mayo JR, et al. Probability of Cancer in Pulmonary Nodules Detected on First Screening CT. *New England Journal of Medicine* 2013; **369**(10): 910-9.
63. Yip R, Yankelevitz DF, Henschke CI. CT screening for lung cancer: definition of positive test result in the National Lung Screening Trial CT cohort compared with I-ELCAP. *Journal of Thoracic Oncology* 2013; 8: S326.
64. Gierada DS, Pinsky P, Nath H, Chiles C, Duan F, Aberle DR. Projected Outcomes Using Different Nodule Sizes to Define a Positive CT Lung Cancer Screening Examination. *Journal of the National Cancer Institute* 2014; **106**(11): dju284-dju.
65. Slatore CG, Baumann C, Pappas M, Humphrey LL. Smoking behaviors among patients receiving computed tomography for lung cancer screening. Systematic review in support of the U.S. Preventive Services Task Force. *Annals of the American Thoracic Society* 2014; **11**(4): 619-27.
66. Tanner NT, Kanodra NM, Gebregziabher M, et al. The Association between Smoking Abstinence and Mortality in the National Lung Screening Trial. *American Journal of Respiratory and Critical Care Medicine* 2016; **193**(5): 534-41.
67. Tramontano AC, Sheehan DF, McMahon PM, et al. Evaluating the impacts of screening and smoking cessation programmes on lung cancer in a high-burden region of the USA: a simulation modeling study. *BMJ Open* 2016; **6**(2).
68. Pua BB, Dou E, O'Connor K, Crawford CB. Integrating smoking cessation into lung cancer screening programs. *Clinical Imaging*; **40**(2): 302-6.
69. Marshall HM, Courtney DA, Passmore LH, et al. Brief Tailored Smoking Cessation Counseling in a Lung Cancer Screening Population is Feasible: A Pilot Randomized Controlled Trial. *Nicotine & Tobacco Research* 2016; **18**(7): 1665-9.
70. Bade M, Bähr V, Brandt U, et al. Effect of smoking cessation counseling within a randomized study on early detection of lung cancer in Germany. *Journal of Cancer Research and Clinical Oncology* 2016; **142**(5): 959-68.
71. Pozzi P, Munarini E, Bravi F, et al. A combined smoking cessation intervention within a lung cancer screening trial: a pilot observational study. *Tumori Journal* 2016; **101**(3): 306-11.
72. Ashraf H, Saghir Z, Dirksen A, et al. Smoking habits in the randomized Danish Lung Cancer Screening Trial with low-dose CT: final results after a 5-year screening programme. *Thorax* 2014; **69**(6): 574.
73. van der Aalst CM, de Koning HJ, van den Bergh KAM, Willemsen MC, van Klaveren RJ. The effectiveness of a computer-tailored smoking cessation intervention for participants in lung cancer screening: A randomized controlled trial. *Lung Cancer* 2012; **76**(2): 204-10.
74. Lemmens V, Oenema A, Knut IK, Brug J. Effectiveness of smoking cessation interventions among adults: a systematic review of reviews. *European Journal of Cancer Prevention* 2008; **17**(6): 535-44.
75. Fucito LM, Czabafy S, Hendricks PS, et al. Pairing smoking-cessation services with lung cancer screening: A clinical guideline from the Association for the Treatment of Tobacco Use and Dependence and the Society for Research on Nicotine and Tobacco. *Cancer* 2016; **122**(8): 1150-9.
76. Park ER, Gareen IF, Japuntich S, et al. Primary care provider-delivered smoking cessation interventions and smoking cessation among participants in the national lung screening trial. *JAMA Internal Medicine* 2015; **175**(9): 1509-16.
77. Armaroli P, Villain P, Suonio E, et al. European Code against Cancer, 4th Edition: Cancer screening. *Cancer Epidemiology* 2015; **39 Suppl 1**: S139-52.
78. van Hees F, Zauber AG, van Veldhuizen H, et al. The value of models in informing resource allocation in colorectal cancer screening: the case of the Netherlands. *Gut* 2015; **64**(12): 1985.
79. McRonald FE, Yadegarfar G, Baldwin DR, et al. The UK Lung Screen (UKLS): Demographic Profile of First 88,897 Approaches Provides Recommendations for Population Screening. *Cancer Prevention Research* 2014; **7**(3): 362.

80. Ali N, Lifford KJ, Carter B, et al. Barriers to uptake among high-risk individuals declining participation in lung cancer screening: a mixed methods analysis of the UK Lung Cancer Screening (UKLS) trial. *BMJ Open* 2015; **5**(7).
81. Priola AM, Priola SM, Giaj-Levra M, et al. Clinical Implications and Added Costs of Incidental Findings in an Early Detection Study of Lung Cancer by Using Low-Dose Spiral Computed Tomography. *Clinical Lung Cancer* 2013; **14**(2): 139-48.
82. van de Wiel JCM, Wang Y, Xu DM, et al. Neglectable benefit of searching for incidental findings in the Dutch--Belgian lung cancer screening trial (NELSON) using low-dose multidetector CT. *European Radiology* 2007; **17**(6): 1474-82.
83. Saghir Z, Dirksen A, Rasmussen JF, Heleno BM, Brodersen J, Pedersen JH. In lung cancer screening by CT incidental findings are frequent and often of clinical importance. American Thoracic Society 2012 International Conference; San Francisco, USA; May 18–May 23, 2012: abstr 5072.
84. Watts JR, Jr., Sonavane SK, Snell-Bergeon J, Nath H. Visual scoring of coronary artery calcification in lung cancer screening computed tomography: association with all-cause and cardiovascular mortality risk. *Coronary Artery Disease* 2015; **26**(2): 157-62.
85. Mets OM, Vliegenthart R, Gondrie MJ, et al. Lung Cancer Screening CT-Based Prediction of Cardiovascular Events. *JACC: Cardiovascular Imaging* 2013; **6**(8): 899-907.
86. Mets OM, Buckens CM, Zanen P, et al. Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. *Journal of the American Medical Association* 2011; **306**(16): 1775-81.

Part 3: Optimization through risk stratification

Chapter 8

Should never-smokers at increased risk for lung cancer be screened?

Kevin ten Haaf, Harry J. de Koning

Published as:

ten Haaf K, de Koning HJ.

Journal of Thoracic Oncology 2015; **10**(9): 1285-91.

Abstract

Introduction: Lung cancer in never-smokers ranks among the 10 most common causes of death due to cancer worldwide and in the United States. However, it is unknown whether never-smokers at elevated risk for developing lung cancer may benefit from lung cancer screening.

Methods: The MIcrosimulation SCreening ANalysis (MISCAN)-Lung microsimulation model as used to assess the effects of lung cancer screening for simulated cohorts of never-smokers at different levels of relative risk (RR) for lung cancer compared with never-smokers at average risk. The benefits and harms of screening were estimated for each cohort and compared with those of a cohort of ever-smokers eligible for lung cancer screening according to the United States Preventive Services Task Force (USPSTF) criteria.

Results: The relative lung cancer mortality reduction in never-smokers was higher than the USPSTF eligible cohort (37% compared with 32%). However, the number of life-years gained per lung cancer death averted was lower (10.4 compared with 11.9) and the proportion of overdiagnosed cancers was higher (9.6% compared with 8.4%) for never-smokers compared with the USPSTF eligible cohort, as never-smokers are diagnosed at a later age. The estimated number of screens per lung cancer death averted ranged from 6,162 for never-smokers at average risk to 151 for never-smokers with an RR of 35 compared with 353 for the USPSTF eligible cohort.

Conclusions: Never-smokers with RRs of 15 to 35 have similar to better trade-offs between benefits and harms compared with ever-smokers recommended for lung cancer screening by the USPSTF guidelines. For most never-smokers, lung cancer screening is not beneficial.

Introduction

Although smoking is considered a main risk factor for developing lung cancer, 10% to 25% of all lung cancers occur in never-smokers.[1,2] Lung cancer in never-smokers is a significant public health problem, as it ranks among the 10 most common causes of death due to cancer worldwide and in the United States.[2-4] The results of the National Lung Screening Trial (NLST) have indicated that lung cancer mortality can be reduced by screening ever-smokers with computed tomography (CT).[5] The United States Preventive Services Task Force (USPSTF) recently published the recommendation to implement annual lung cancer screening for ever-smokers aged 55 to 80 years who have smoked at least 30 pack-years and, if quit smoking, quit less than 15 years ago.[6] Other organizations have recommended screening using the NLST eligibility criteria or variations thereof.[7-9] To our knowledge, no organization currently recommends lung cancer screening for never-smokers. Some lung cancer screening studies have included never-smokers, but these studies used chest radiography or were single-arm studies.[10-12] A survey on attitudes toward lung cancer screening in the United States showed that a large proportion of never-smokers were willing to consider lung cancer screening, even though few believed that they were at risk for developing lung cancer.[13]

In addition to tobacco smoking, various risk factors for developing lung cancer have been identified for ever- and never-smokers, such as environmental tobacco smoke (e.g., “second hand smoking”), exposure to carcinogens (e.g., asbestos, radon gas, and ionizing radiation), and genetic susceptibility.[3,14-16] A number of risk models incorporate these and other risk factors to identify ever- and never-smokers at elevated levels of risk.[17-21] Recent studies have identified subpopulations within the NLST who were at a higher level of risk for developing lung cancer compared with the average population of the trial.[20,22,23] Screening was more effective for these subpopulations, which indicates that screening recommendations based on an individual’s risk could lead to more effective screening programs.[20,22,23] Therefore, some researchers argue that lung cancer screening may be recommended for never-smokers, provided that they have a high risk for developing lung cancer.[24] However, the long-term benefits and harms of implementing a lung cancer screening program for never-smokers are unknown. The USPSTF recommendations were in

part based on modeling analyses, which investigated the trade-offs between the long-term benefits and harms of different screening policies for ever-smokers.[25] This study aims to investigate the trade-offs between the benefits and harms of lung cancer screening for never-smokers at different levels of risk.

Methods

MISCAN-Lung

The MIcrosimulation SCreening ANalysis (MISCAN)-Lung model is used in this investigation. MISCAN-Lung has been calibrated to the NLST, the Prostate, Lung, Colorectal and Ovarian Cancer Screening trial (PLCO), and data from the Surveillance, Epidemiology and End Results (SEER) Program, from which it derived information on the preclinical duration of lung cancer and CT screening effectiveness.[26,27] Lung cancer incidence and mortality in never-smokers in the PLCO were among the calibration targets of the model.[26,27] MISCAN-Lung aided in informing the USPSTF on their recommendations or lung cancer screening.[25,28]

Histologic Types

There are indications that smoking behavior affects not only a person's risk of developing lung cancer, but also the histologic type that develops.[29,30] This suggests that the distribution of histological types of lung cancer in never-smokers may differ from ever-smokers. Subramanian and Govindan provided an overview of the distribution of histological types of lung cancer in never-smokers across different studies.[16] This overview was used to derive the distribution of histological types of lung cancer in never-smokers for this investigation, shown in Table 1.[16] To our knowledge, little information is available on differences in the distribution of histological types of lung cancer in never-smokers between sexes. Therefore, we assumed that the distribution of histological types of lung cancer in never-smokers did not differ by sex.

Table 1: Histological Types Considered in MISCAN-Lung

Histological Types Considered in MISCAN-Lung	Proportions Considered in Never-Smokers (Both Sexes)	Proportions Considered in Ever-Smokers (Men)	Proportions Considered in Ever-Smokers (Women)
Adenocarcinoma/ large cell carcinoma/ bronchioalveolar carcinoma	66.68%	41.01%	50.33%
Squamous cell carcinoma	13.68%	25.22%	15.78%
Small cell carcinoma	2.53%	13.75%	13.26%
Other non-small-cell carcinoma	17.12%	20.02%	20.63%

Abbreviations: MISCAN, Microsimulation Screening Analysis model.

Lung Cancer Survival

It has been suggested that never-smokers may have a better response to certain treatments compared with ever-smokers, such as treatment with epidermal growth factor receptor inhibitors, which could lead to differences in survival.[31,32] Some studies suggest that never-smokers have a better survival compared with ever-smokers, whereas other studies indicate that no significant differences in survival exist.[33-36] To our knowledge, no study provides detailed data on lung cancer survival for never-smokers by stage, histology, and sex.[33-36] Therefore, survival data from SEER were used, which provides detailed information on survival by stage, histology, and sex for ever- and never-smokers combined.[37]

Lung Carcinogenesis

MISCAN-Lung uses the two-stage clonal expansion model (TSCE) to estimate a person's risk of developing lung cancer as a function of age and smoking history.[26,27,38,39] The TSCE has been used to investigate the age-specific incidence of lung cancer in never-smokers previously.[14,39,40] To assess whether MISCAN-Lung is suitable for investigating the effectiveness of lung cancer screening for never-smokers, the estimated age-group–specific mortality rates of lung cancer in never-smokers were compared with those reported by Thun et al.[41]

Considered Levels of Relative Risk

If lung cancer screening is to be considered for never-smokers, eligible individuals will need to be identified, for example, through the application of risk models. To our knowledge, the

following lung cancer risk models consider never-smokers: the Spitz, PLCOm2011, PLCOm2014, and the Liverpool Lung Project (LLP) models.[17-20] The Spitz model incorporates environmental tobacco smoke exposure (odds ratio [OR], 1.80; 95% confidence interval [CI], 1.20–2.69)) and a family history of any cancer in two or more first-degree relatives (OR: 2.00; 95% CI, 1.39–2.90).[18] Spitz et al noted that the ORs of these variables closely approximated the relative risks (RRs).[18] Thus, the Spitz model considers RRs up to 3.6. Recently, this model was extended to incorporate micronuclei in binucleated cells (BN-MN) (OR: 16.72 per unit increase; 95% CI, 9.01–31.02) alongside environmental tobacco smoke exposure (OR: 1.12; 95% CI, 0.47–2.68) and a family history of cancer in two or more first-degree relatives (OR: 1.06; 95% CI, 0.47–2.43).21 The average difference in BN-MN between cases and controls in the model's development and validation data sets was 1.78 to 1.79 units. [21] Assuming the ORs of the model variables closely approximate the RRs and an increase of 1.80 units of BN-MN compared with a never-smoker at average risk is considered, the model considers RRs up to at least 35.73 for never-smokers. The PLCOm2011 model was the first model based on data from PLCO to provide risk estimates for never-smokers.[17] Recently, an updated version of this model (PLCOm2014) was published that incorporates five risk factors (excluding age and race) for never-smokers: education (OR: 0.92 per one of six levels change; 95% CI, 0.87–0.96), body mass index (BMI) (OR: 0.97 per one unit change; 95% CI, 0.95–0.99), chronic obstructive pulmonary disease (OR: 1.41; 95% CI, 1.15–1.73), a personal history of cancer (OR: 1.62; 95% CI, 1.22–2.16), and a family history of lung cancer (OR: 1.80; 95% CI, 1.48–2.18).20 The calculator provided by the authors was used to verify that the ORs closely approximate the RRs (available at http://www.brocku.ca/lung-cancer-risk-calculator). A BMI of 18 is the lowest for which the model is valid and assuming the base BMI and education levels are similar to those in the PLCOm2012 model (a BMI of 27 and "some college education"), implies that the PLCOm2014 model considers RRs up to 6.98 (disregarding age and race).[20,23] The LLP model incorporates four risk factors for never-smokers: a history of pneumonia (OR: 1.83; 95% CI, 1.26– 2.64), asbestos exposure (OR: 1.89; 95% CI, 1.35–2.62), a history of cancer (OR: 1.96; 95% CI, 1.22–3.14), and a family history of lung cancer (OR: 2.02 for age of onset <60; 95% CI, 1.18–3.45; and 1.18 for age of onset ≥60; 95% CI, 0.79–1.76).[19] The model was replicated in R software (version 3.0.1) and analyses indicate that the ORs of the risk factors closely approximate the RRs. Thus, the LLP model considers RRs up to 13.69. Therefore, cohorts of

never-smokers with the following levels of RR will be simulated: 1, 2, 5, 10, 15, 20, and 35. For cohorts with RRs higher than 1, the hazard function of the TSCE at each age was multiplied by the considered level of RR. In addition, a "USPSTF eligible" cohort is simulated composed of individuals who would be eligible for at least one screening at some point in their life according to the USPSTF recommendations. [25] The Smoking History Generator developed by the National Cancer Institute was used to generate the probability of death from causes other than lung cancer by smoking behavior (including never smoking).[42-44]

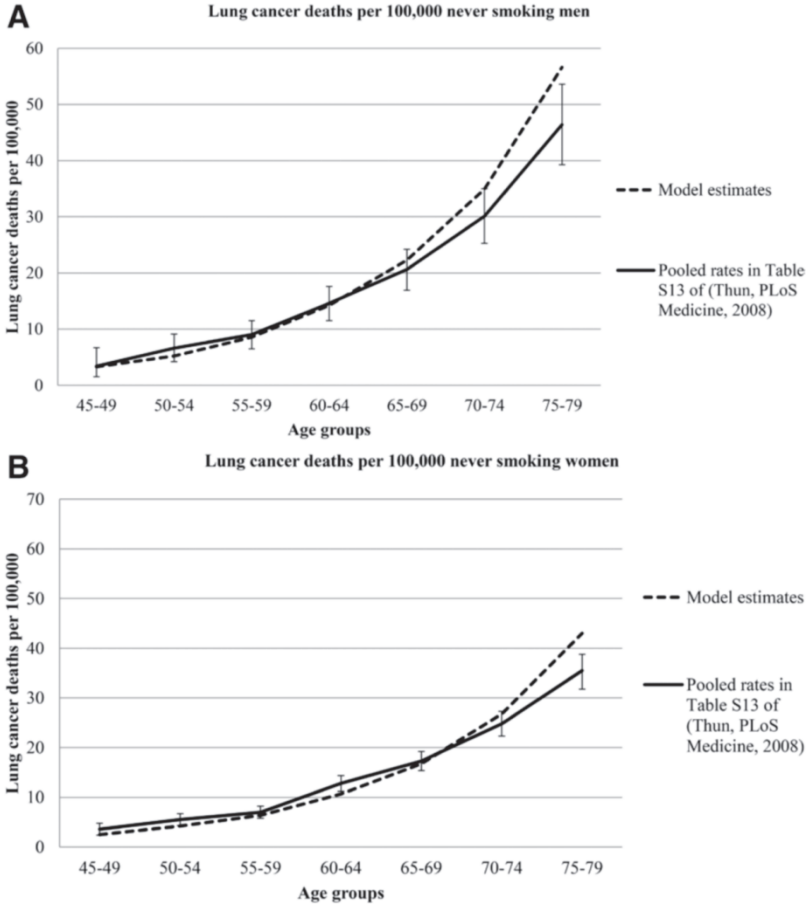
Considered Screening Programs, Benefits, and Harms

The investigated cohorts are assumed to be born in 1950 and followed from ages 45 to 90 years, similar to de Koning et al.[25] To allow comparison with the screening policy recommended by the USPSTF, never-smokers are assumed to be screened annually from ages 55 to 80 years with a perfect adherence to screening. The investigated benefits and harms of lung cancer screening include the relative reduction in lung cancer mortality, the number of life-years gained, and overdiagnosis.

Results

The age-group–specific lung cancer mortality rates for never-smoking men and women estimated by MISCAN-Lung were compared with those reported by Thun et al (Table S13 in their report), in Figures 1A and B, respectively.[41] Overall, the model reproduced the reported age-group–specific lung cancer mortality rates well for both sexes but somewhat overestimated the lung cancer mortality rate for ages 75 to 79 years.

Figure 1: Observed and estimated lung cancer death rates in never-smoking men (A) and women (B)



Adapted from Thun et al.[41] Error bars denote 95% confidence intervals for the incidence rate difference.

Table 2 shows the benefits of lung cancer screening for the investigated cohorts. The proportion of lung cancers detected at an early stage was higher for never-smokers compared with the USPSTF eligible cohort (65.8–65.9% of all cases compared with 59.4%). This may be a result of the higher proportion of adenocarcinomas in never-smokers (see Table 1), which have a longer preclinical sojourn time and are more likely to be detected at an early stage by CT screening compared with other histologies.[26] As a result of the larger proportion of lung cancers detected at an early stage, the relative reduction in lung cancer mortality was higher for never-smokers compared with the USPSTF eligible cohort: 37.0% to 37.3% compared with 32.7%.

Table 2: Benefits of screening

Scenario	Lung Cancers Detected at an Early Stage (Stage I-II) (%)	Lung Cancer Mortality Reduction (%)	Absolute Number of Lung Cancer Deaths Averted per 100,000	Life-Years Gained per 100,000	Life-Years Gained per Lung Cancer Death Averted	Screens per Life-Year Gained	Screens per Lung Cancer Death Averted
USPSTF (ever eligible only)	59.4	32.7	4,305	51,035	11.9	30	353
Never-smokers at average risk	65.8	37.1	354	3,669	10.4	594	6,162
Never-smokers at two times average risk	65.9	37	706	7,332	10.4	296	3,075
Never-smokers at five times average risk	65.8	37	1,764	18,359	10.4	117	1,216
Never-smokers at 10 times average risk	65.8	37.1	3,541	36,809	10.4	57	593
Never-smokers at 15 times average risk	65.8	37.1	5,322	55,247	10.4	37	387
Never-smokers at 20 times average risk	65.8	37.1	7,118	73,892	10.4	27	283
Never-smokers at 35 times average risk	65.9	37.3	12,509	129,786	10.4	15	151

Abbreviations: USPSTF, United States Preventive Services Task Force.

However, the number of lung cancer deaths averted (per 100,000) was lower for most cohorts of never-smokers, ranging from 354 deaths averted for never-smokers at average risk to 12,509 for never-smokers with an RR of 35 compared with 4,305 for the USPSTF eligible cohort. The same holds for the number of life-years gained (per 100,000), which ranged from 3,669 for never-smokers at average risk to 129,509 for never-smokers with an RR of 35 compared with 51,035 for the USPSTF eligible cohort. The number of lung cancer deaths averted and life-years gained for never-smokers with an RR of 15 were higher compared with the USPSTF cohort. However, because of the high number of screens, the number of screens per life-year gained and the number of screens per lung cancer death averted were still higher compared with the USPSTF cohort. However, screening never-smokers with an RR of 20 leads to a slightly lower number of screens per life-year gained and a much lower number of screens per lung cancer death averted compared with the USPSTF cohort. Table 3 shows the harms of lung cancer screening for the investigated cohorts.

Table 3: Harms of screening

Scenario	Screens per 100,000	CT Examinations per 100,000 (Includes Screenings)	Average Number of Screening Examinations per Person Screened	Percentage of Lung Cancers Overdiagnosed	Percentage of Screen-Detected Lung Cancers Overdiagnosed	Screens per 100,000	CT Examinations per 100,000 (Includes Screenings)
USPSTF (ever eligible only)	1,520,632	1,776,046	16	5	8.4	1,520,632	1,776,046
Never-smokers at average risk	2,179,173	2,544,605	22.5	5	9.5	2,179,173	2,544,605
Never-smokers at two times average risk	2,170,544	2,534,532	23	5	9.6	2,170,544	2,534,532
Never-smokers at five times average risk	2,144,601	2,504,249	22	5	9.6	2,144,601	2,504,249
Never-smokers at 10 times average risk	2,101,248	2,453,645	22	5	9.6	2,101,248	2,453,645
Never-smokers at 15 times average risk	2,057,745	2,402,865	22	5	9.6	2,057,745	2,402,865
Never-smokers at 20 times average risk	2,014,118	2,351,940	21	5	9.6	2,014,118	2,351,940
Never-smokers at 35 times average risk	1,882,721	2,198,562	20	5	9.6	1,882,721	2,198,562

Abbreviations: USPSTF, United States Preventive Services Task Force.

The number of screens (per 100,000) was much greater for the USPSTF eligible and never-smoker cohorts compared with the 1950 cohort examined in de Koning et al, as in the latter cohort only 19.3% of the cohort received at least one screen.[25] The number of screens was higher for the never-smoker cohorts compared with the USPSTF eligible cohort: approximately 1.8 to 2.1 million screens compared with 1.5 million screens. This is because of two reasons: first, never-smokers live longer compared with ever-smokers and will be able to attend more screenings during their lifetime.[43,45] Second, the USPSTF criteria

indicate that ever-smokers may not be eligible for screening at the earliest starting age, as some current, continuing smokers may not reach the minimum number of pack-years at age 55 but at a later age. In addition, eligible former smokers may not complete the full screening program because of becoming ineligible by reaching the maximum years since cessation. As a result, the average number of screening examinations per person screened was higher for the cohorts of never-smokers compared with the USPSTF eligible cohort (20–22 compared with 16). As the never-smoker cohorts receive a higher number of screening examinations and are diagnosed at a later age (when no screening occurs), the proportion of overdiagnosis was higher compared with the USPSTF eligible cohort (9.5–9.6% of all screen-detected cases compared with 8.4%).

Discussion

Suggestions to recommend lung cancer screening based on an individual's risk and the growing awareness of lung cancer in never-smokers have raised the question of whether never-smokers at high risk for lung cancer should be screened.[2-4,12,14,20,23,30,33,35,41,46] Our study is the first to provide indications whether never-smokers may benefit from lung cancer screening through quantifying the benefits and harms of screening never-smokers at different levels of risk. Screening never-smokers at average risk or an RR of 2 compared with average risk has unfavorable trade-offs between benefits and harms, requiring 3,000 to 6,000 screens to prevent one death. However, the trade-off for never-smokers with an RR of 5 is more favorable than breast cancer screening: 1,216 screens per death averted compared with 1,558 (Model E in Mandelblatt et al); however, the number of screens per life-year gained is less favorable: 117 compared with 91.[47] Never-smokers with an RR of 10 have more favorable trade-offs in deaths prevented and life-years gained per screen compared with breast cancer screening but less favorable compared with ever-smokers for whom the USPSTF recommends screening.[25,47] However, never-smokers with an RR of 15 to 35 have similar to more favorable trade-offs between the benefits and the harms compared with smokers for whom the USPSTF recommends screening.

Lung cancer screening for never smokers may lead to a higher relative reduction in lung cancer mortality compared with screening ever-smokers. However, although the cohorts of

never-smokers have a higher relative reduction in lung cancer mortality compared with the USPSTF eligible cohort, the increase in number of life-years gained is less than one would anticipate. For example, the number of lung cancer deaths averted for never-smokers with an RR of 15 was 23.62% higher compared with the USPSTF eligible cohort, whereas the number of life-years gained was only 8.25% higher (Table 2). This can be explained by the lower number of life-years gained per lung cancer death averted, which was 11.9 years for the USPSTF eligible cohort compared with 10.4 for the never-smoker cohorts. This may seem counterintuitive, as ever-smokers have a higher all-cause mortality compared with never-smokers but can be explained through the differences in the average age of lung cancer diagnosis and average age of death between these groups.[43,45] Table S1 in the supplementary material of this Chapter shows the average age of lung cancer diagnosis (given that the cancer is diagnosed after age 45) and the average age of death (given that the person is alive at age 45 years) for the investigated cohorts. Table S1 in the supplementary material of this Chapter indicates that persons in the USPSTF eligible cohort die younger compared with never-smokers because of the detrimental effects of smoking.[43,45] However, because of the carcinogenic effects of smoking, patients in the USPSTF eligible cohort developed lung cancer at a younger age compared with never-smokers. In addition, the high proportion of adenocarcinoma in never-smokers, which have a longer preclinical sojourn time compared with other histologies, may further contribute to the later age of diagnosis.[26]

Our investigation has some limitations. We assume that the preclinical duration of lung cancer in never-smokers is similar to that of ever-smokers, whereas there are indications that lung cancer biology may differ in never-smokers.[3] However, although the carcinogenesis process may differ in never-smokers, to our knowledge, there are no indications that differences in the preclinical duration of lung cancer exist between never- and ever-smokers.[3] Another limitation is that the investigated levels of RR are assumed to be constant over a person's life. Although the elevation in risk may be constant over a person's life for some risk factors, such as genetic susceptibility, this may not be the case for risk factors such as asbestos or radon exposure.[3,14,16] However, the benefits and harms of screening never-smokers at specific levels of RR are more easily interpreted by assuming that the RR is constant over a person's lifetime. Finally, although our research indicates at

what level of risk for developing lung cancer never-smokers could benefit from lung cancer screening, our findings are based on model-based extrapolations. Further research is needed to accurately identify never-smokers at high risk for developing lung cancer, which would allow us to further validate our findings. However, although a number of risk factors for developing lung cancer in never-smokers have been identified, the etiology of lung cancer in never-smokers is not well understood.[3,14-16,46] Although lung cancer risk models for never-smokers exist, the performance of the majority of these models is limited.[17-19,48] This is further demonstrated by comparing the risk of developing lung cancer for an average 67-year-old (the age between the USPSTF recommended screening ages of 55–80 years) never-smoker (by sex) for different time frames in MISCAN-Lung with those of the investigated risk models (shown in the supplementary material of this Chapter, in Table S2). Table S3 in the supplementary material of this Chapter indicates that the investigated lung cancer risk models for never-smokers generally predict higher lung cancer risks compared with MISCAN-Lung. However, the age-group–specific lung cancer mortality rates for never-smoking men and women estimated by MISCAN-Lung are comparable with those reported by Thun et al (Table S13 in their report).[41] This indicates that the majority of lung cancer risk models may overestimate the risk of lung cancer for never-smokers.

Finally, the proportion of never-smokers with RRs of 15 compared with never-smokers at average risk is uncertain, as information on many risk factors (and the joint distribution thereof) is scarcely available at the population level. The application of the PLCOm2014 model to the PLCO data set and the application of the LLP model for recruiting participants for the UK Lung Cancer Screening Trial (UKLS) may currently provide the best information on the expected levels of risk for never-smokers.[20,49] The PLCOm2014 model suggests that the maximum observed risk in 65,711 never-smokers in the PLCO was 1.47% over a 6-year period.[20] A white never-smoker attaining the highest level of RR (6.98, disregarding age and race) would not reach this level of risk until age 73 years (verified using the calculator provided by the authors, available at http://www.brocku.ca/lung-cancer-risk-calculator). The theoretical maximum possible 6-year risk of lung cancer for never-smokers in the PLCOm2014 model is 3.5%; however, the necessary combination of risk factors to achieve this level of risk is expected to be rare.[20] The LLP model was used to recruit participants for the UKLS trial, including never-smokers.[49] Only four (0.04% of 10,697) never-smokers had a

high LLP risk (a risk of 5% or higher over a 5-year period) and all were aged at least 73 years.[49] Analyses using the model (replicated in R software) suggest that both men and women can only achieve this absolute level of risk at this age at the highest level of RR (13.69). In conclusion, this study is the first to investigate the long-term benefits and harms of lung cancer screening for never-smokers. Screening never-smokers at high levels of elevated risk for developing lung cancer (RRs of 15 or higher compared with average risk) is indicated to have similar or better trade-offs between benefits and harms as the population for which the USPSTF recommends screening. However, most lung cancer risk models for never-smokers consider RRs of lower than 15 for never-smokers at elevated risk compared with never-smokers at average risk. In addition, the majority of lung cancer risk models for never-smokers may overestimate the average risk of never-smokers. Applications of lung cancer risk models to populations of never-smokers suggest that few never-smokers attain high levels of risk.[20,49] Thus, few never-smokers are expected to attain RRs of 15 compared with never-smokers at average risk. Therefore, for most never-smokers, lung cancer screening is not beneficial.

Acknowledgements

We thank M.C. Tammemägi for providing useful comments with regard to the PLCOm2011 and PLCOm2014 models. We thank our colleagues from the CISNET Lung working group (in particular J. Jeon) and F. van Hees (Department of Public Health, Erasmus MC) for providing useful comments.

References

1. Koo LC, Ho JHC. Worldwide Epidemiological Patterns of Lung Cancer in Nonsmokers. *International Journal of Epidemiology* 1990; **19**(Supplement_1): S14-S23.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. *CA: A Cancer Journal for Clinicians* 2005; **55**(2): 74-108.
3. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers-a different disease. *Nature Reviews Cancer* 2007; **7**(10): 778-90.
4. Thun MJ, Henley SJ, Burns D, Jemal A, Shanks TG, Calle EE. Lung Cancer Death Rates in Lifelong Nonsmokers. *Journal of the National Cancer Institute* 2006; **98**(10): 691-9.
5. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
6. Moyer VA, on behalf of the USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 2014; **160**(5): 330-8.
7. Wood DE, Eapen GA, Ettinger DS, et al. Lung Cancer Screening. *Journal of the National Comprehensive Cancer Network* 2012; **10**(2): 240-65.
8. Smith RA, Manassaram-Baptiste D, Brooks D, et al. Cancer screening in the United States, 2014: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA: A Cancer Journal for Clinicians* 2014; **64**(1): 30-51.
9. Detterbeck FC, Mazzone PJ, Naidich DP, Bach PB. Screening for Lung Cancer. *CHEST* 2013; **143**(5): e78S-e92S.
10. Sone S, Nakayama T, Honda T, et al. Long-term follow-up study of a population-based 1996–1998 mass screening programme for lung cancer using mobile low-dose spiral computed tomography. *Lung Cancer* 2007; **58**(3): 329-41.
11. Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. *The Lancet* 1998; **351**(9111): 1242-5.
12. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
13. Silvestri GA, Nietert PJ, Zoller J, Carter C, Bradford D. Attitudes towards screening for lung cancer among smokers and their non-smoking counterparts. *Thorax* 2007; **62**(2): 126.
14. McCarthy WJ, Meza R, Jeon J, Moolgavkar SH. Chapter 6: Lung Cancer in Never Smokers: Epidemiology and Risk Prediction Models. *Risk Analysis* 2012; **32**: S69-S84.
15. Couraud S, Zalcman G, Milleron B, Morin F, Souquet P-J. Lung cancer in never smokers – A review. *European Journal of Cancer* 2012; **48**(9): 1299-311.
16. Janakiraman S, Ramaswamy G. Lung Cancer in Never Smokers: A Review. *Journal of Clinical Oncology* 2007; **25**(5): 561-70.
17. Tammemägi CM, Pinsky PF, Caporaso NE, et al. Lung Cancer Risk Prediction: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Models and Validation. *Journal of the National Cancer Institute* 2011; **103**(13): 1058-68.
18. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *Journal of the National Cancer Institute* 2007; **99**(9): 715-26.
19. 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. *Annals of Internal Medicine* 2012; **157**(4): 242-50.
20. Tammemägi MC, Church TR, Hocking WG, et al. Evaluation of the Lung Cancer Risks at Which to Screen Ever- and Never-Smokers: Screening Rules Applied to the PLCO and NLST Cohorts. *PLOS Medicine* 2014; **11**(12): e1001764.

21. El-Zein RA, Lopez MS, Amelio AM, et al. The Cytokinesis-Blocked Micronucleus Assay as a Strong Predictor of Lung Cancer: Extension of a Lung Cancer Risk Prediction Model. *Cancer Epidemiology Biomarkers & Prevention* 2014; **23**(11): 2462.
22. Kovalchik SA, Tammemägi M, Berg CD, et al. Targeting of Low-Dose CT Screening According to the Risk of Lung-Cancer Death. *New England Journal of Medicine* 2013; **369**(3): 245-54.
23. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.
24. Baldwin DR, Duffy SW, Wald NJ, Page R, Hansell DM, Field JK. UK Lung Screen (UKLS) nodule management protocol: modeling of a single screen randomized controlled trial of low-dose CT screening for lung cancer. *Thorax* 2011; **66**(4): 308.
25. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
26. ten Haaf K, van Rosmalen J, de Koning HJ. Lung Cancer Detectability by Test, Histology, Stage, and Gender: Estimates from the NLST and the PLCO Trials. *Cancer Epidemiology Biomarkers & Prevention* 2015; **24**(1): 154-61.
27. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
28. McMahon PM, Meza R, Plevritis SK, et al. Comparing Benefits from Many Possible Computed Tomography Lung Cancer Screening Programs: Extrapolating from the National Lung Screening Trial Using Comparative Modeling. *PLOS ONE* 2014; **9**(6): e99978.
29. Kenfield SA, Wei EK, Stampfer MJ, Rosner BA, Colditz GA. Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco Control* 2008; **17**(3): 198.
30. Haiman CA, Stram DO, Wilkens LR, et al. Ethnic and Racial Differences in the Smoking-Related Risk of Lung Cancer. *New England Journal of Medicine* 2006; **354**(4): 333-42.
31. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, et al. Erlotinib in Previously Treated Non–Small-Cell Lung Cancer. *New England Journal of Medicine* 2005; **353**(2): 123-32.
32. Thatcher N, Chang A, Parikh P, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomized, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *The Lancet* 2005; **366**(9496): 1527-37.
33. Kawaguchi T, Matsumura A, Fukai S, et al. Japanese Ethnicity Compared with Caucasian Ethnicity and Never-Smoking Status Are Independent Favorable Prognostic Factors for Overall Survival in Non-small Cell Lung Cancer: A Collaborative Epidemiologic Study of the National Hospital Organization Study Group for Lung Cancer (NHSGLC) in Japan and a Southern California Regional Cancer Registry Databases. *Journal of Thoracic Oncology* 2010; **5**(7): 1001-10.
34. Meguid RA, Hooker CM, Harris J, et al. Long-term survival outcomes by smoking status in surgical and nonsurgical patients with non-small cell lung cancer: comparing never smokers and current smokers. *Chest* 2010; **138**(3): 500-9.
35. Subramanian J, Velcheti V, Gao F, Govindan R. Presentation and Stage-Specific Outcomes of Lifelong Never-smokers with Non-small Cell Lung Cancer (NSCLC). *Journal of Thoracic Oncology* 2007; **2**(9): 827-30.
36. Chee-Keong T, Fei G, Wan-Teck L, et al. Never-Smokers With Lung Cancer: Epidemiologic Evidence of a Distinct Disease Entity. *Journal of Clinical Oncology* 2006; **24**(15): 2245-51.
37. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence – SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2012 Sub (1973–2010 varying) – Linked To County Attributes – Total U.S., 1969–2011 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2013, based on the November 2012 submission. Accessed March 29, 2014.
38. Heidenreich WF, Luebeck EG, Moolgavkar SH. Some Properties of the Hazard Function of the Two-Mutation Clonal Expansion Model. *Risk Analysis* 1997; **17**(3): 391-9.

39. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.
40. Hazelton WD, Clements MS, Moolgavkar SH. Multistage Carcinogenesis and Lung Cancer Mortality in Three Cohorts. *Cancer Epidemiology Biomarkers & Prevention* 2005; **14**(5): 1171.
41. Thun MJ, Hannan LM, Adams-Campbell LL, et al. Lung Cancer Occurrence in Never-Smokers: An Analysis of 13 Cohorts and 22 Cancer Registry Studies. *PLOS Medicine* 2008; **5**(9): e185.
42. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Analysis* 2012; **32 Suppl 1**: S51-68.
43. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.
44. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort–Specific Smoking Histories, 1965–2009. *American Journal of Preventive Medicine* 2014; **46**(2): e31-e7.
45. Thun MJ, Carter BD, Feskanich D, et al. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine* 2013; **368**(4): 351-64.
46. Heather AW, Ellen TC, Scarlett LG, et al. Lung Cancer Incidence in Never Smokers. *Journal of Clinical Oncology* 2007; **25**(5): 472-8.
47. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2009; **151**(10): 738-47.
48. Brenner DR, Hung RJ, Tsao M-S, et al. Lung cancer risk in never-smokers: a population-based case-control study of epidemiologic risk factors. *BMC Cancer* 2010; **10**(1): 285.
49. McRonald FE, Yadegarfar G, Baldwin DR, et al. The UK Lung Screen (UKLS): Demographic Profile of First 88,897 Approaches Provides Recommendations for Population Screening. *Cancer Prevention Research* 2014; **7**(3): 362.

Chapter 8

Supplementary material

Table S1: Average age of lung cancer diagnosis and average age of death by cohort

Cohort	Average age of lung cancer diagnosis (given that cancer is not detected before age 45)	Average age of death (given that the person is alive at age 45)
Cohort ever eligible by USPSTF criteria	74.78	78.92
Never-smokers at average risk	81.06	82.25
Never-smokers at 2 times average risk	81.04	82.13
Never-smokers at 5 times average risk	81.06	81.74
Never-smokers at 10 times average risk	81.05	81.11
Never-smokers at 15 times average risk	81.00	80.47
Never-smokers at 20 times average risk	80.51	79.83
Never-smokers at 35 times average risk	78.13	78.06

Abbreviations: United States Preventive Services Task Force (USPSTF)

Table S2: Probability of a 67-year old never-smoker at average risk developing lung cancer in the MISCAN-Lung model

MISCAN-Lung estimates	Male	Female
1-year probability	0.03%	0.02%
5-year probability	0.17%	0.14%
6-year probability	0.22%	0.18%

Table S3: Probability of a 67-year old never-smoker for developing lung cancer for different risk-models

Examined model / person	Male	Relative risk compared to MISCAN-Lung estimate for never-smoker at regular risk	Female	Relative risk compared to MISCAN-Lung estimate for never-smoker at regular risk
Spitz model: Never-smoker with no risk-factors (1-year probability)	0.04%	1.33	0.04%	2.00
Extended Spitz model: Never-smoker with no risk-factors (1-year probability, assuming never-smokers on average have 1.80 units of micronuclei in binucleated cells (BN-MN))	0.04%	1.33	0.04%	2.00
LLP model: Never-smoker with no risk factors (5-year probability)	0.43%	2.53	0.32%	2.29
PLCOm2014 model: White never-smoker with some college education and a body mass index of 27 (6-year probability)	0.13%	0.59	0.13%	0.72

Chapter 9

Performance and cost-effectiveness of computed tomography lung cancer screening scenarios in a population-based setting: a microsimulation modeling analysis in Ontario, Canada

Kevin ten Haaf, Martin C. Tammemägi, Susan J. Bondy, Carlijn M. van der Aalst, Sumei Gu, S. Elizabeth McGregor, Garth Nicholas, Harry J. de Koning, Lawrence F. Paszat

Published as:

ten Haaf K, Tammemägi MC, Bondy SJ, et al.

PLOS Medicine 2017; **14**(2): e1002225

Abstract

Background

The National Lung Screening Trial (NLST) results indicate that computed tomography (CT) lung cancer screening for current and former smokers with three annual screens can be cost-effective in a trial setting. However, the cost-effectiveness in a population-based setting with >3 screening rounds is uncertain. Therefore, the objective of this study was to estimate the cost-effectiveness of lung cancer screening in a population-based setting in Ontario, Canada, and evaluate the effects of screening eligibility criteria.

Methods and Findings

This study used microsimulation modeling informed by various data sources, including the Ontario Health Insurance Plan (OHIP), Ontario Cancer Registry, smoking behavior surveys, and the NLST. Persons, born between 1940 and 1969, were examined from a third-party health care payer perspective across a lifetime horizon. Starting in 2015, 576 CT screening scenarios were examined, varying by age to start and end screening, smoking eligibility criteria, and screening interval. Among the examined outcome measures were lung cancer deaths averted, life-years gained, percentage ever screened, costs (in 2015 Canadian dollars), and overdiagnosis. The results of the base-case analysis indicated that annual screening was more cost-effective than biennial screening. Scenarios with eligibility criteria that required as few as 20 pack-years were dominated by scenarios that required higher numbers of accumulated pack-years. In general, scenarios that applied stringent smoking eligibility criteria (i.e., requiring higher levels of accumulated smoking exposure) were more cost-effective than scenarios with less stringent smoking eligibility criteria, with modest differences in life-years gained. Annual screening between ages 55-75 for persons who smoked 40 pack-years and who currently smoke or quit 10 years ago yielded an incremental cost-effectiveness ratio of $41,136 Canadian dollars ($33,825 in May 1, 2015, United States dollars) per life-year gained (compared to annual screening between ages 60-75 for persons who smoked 40 pack-years and who currently smoke or quit 10 years ago), which was considered optimal at a cost-effectiveness threshold of $50,000 Canadian dollars ($41,114 May 1, 2015, U.S. dollars). If 50% lower or higher attributable costs were assumed, the incremental cost-effectiveness ratio of this scenario was estimated to be $38,240 ($31,444 May 1, 2015, U.S. dollars) or $48,525 ($39,901 May 1, 2015, U.S. dollars), respectively. If

50% lower or higher costs for CT examinations were assumed, the incremental cost-effectiveness ratio of this scenario was estimated to be $28,630 ($23,542 May 1, 2015, U.S. dollars) or $73,507 ($60,443 May 1, 2015, U.S. dollars), respectively. This scenario would screen 9.56% (499,261 individuals) of the total population (ever- and never-smokers) at least once, which would require 4,788,523 CT examinations, and reduce lung cancer mortality in the total population by 9.05% (preventing 13,108 lung cancer deaths), while 12.53% of screen-detected cancers would be overdiagnosed (4,282 overdiagnosed cases). Sensitivity analyses indicated that the overall results were most sensitive to variations in CT examination costs. Quality of life was not incorporated in the analyses, and assumptions for follow-up procedures were based on data from the NLST, which may not be generalizable to a population-based setting.

Conclusions

Lung cancer screening with stringent smoking eligibility criteria can be cost-effective in a population-based setting.

Introduction

The National Lung Screening Trial (NLST) showed that screening with low-dose computed tomography (CT) can reduce lung cancer mortality.[1] Although the sensitivity of CT screening in the NLST was reported to be over 90% across the three screening rounds, the reported specificity ranged from 73.4% in the first round to 83.9% in the third round.[1] Overall, 23.3% of the CT screens in the NLST were false-positive, which often required additional follow-up CT examinations and, infrequently, invasive procedures (such as a biopsy, bronchoscopy, or thoracotomy) to determine the malignancy of one or more suspicious pulmonary nodules detected by CT screening.[1] Lung cancer screening with three annual screens, as performed in the NLST, was reported to be cost-effective by U.S. standards, yielding estimated cost-effectiveness ratios of U.S. $52,000 per life-year gained and U.S. $81,000 per quality-adjusted life-year gained.[2,3] However, although the cost-effectiveness of lung cancer screening in a population-based setting has been examined previously, it has not been examined extensively.[4-9]

To determine the cost-effectiveness of implementing cancer screening programs, microsimulation modeling is invaluable.[10,11] The United States Preventive Services Task Force (USPSTF) recommended lung cancer screening for current and former smokers who have quit within the past 15 years, aged 55 through 80 who smoked at least 30 pack-years.[12] This recommendation was in part based on a comparative modeling study using microsimulation models, as modeling allows one to extrapolate the results of randomized clinical trials and provide information on the long-term benefits and harms for screening programs with different designs and populations than those considered in clinical trials.[13] However, although the modeling study that informed the USPSTF provides an understanding of the trade-offs between the benefits and harms of different screening scenarios, it did not formally consider their cost-effectiveness.[13]

In Ontario, Canada, lung cancer is responsible for the largest proportion of cancer deaths (49.9 per 100,000 individuals) in the population of 13.8 million individuals, despite falling smoking rates.[14-17] The implementation of a lung cancer screening program, in addition to continued efforts in primary prevention of smoking, could reduce lung cancer mortality.

However, concerns have been raised about whether and how such a program can be implemented in a cost-effective manner.[18,19] Previous studies on the cost-effectiveness of population-based lung cancer screening have yielded inconclusive results, ranging from U.S. $18,452-$66,480 per life-year gained and U.S. $27,756-$243,077 per quality-adjusted life-year gained.[4-9] However, many of these studies reported the average cost-effectiveness ratios (ACER, the ratio of differences in costs to differences in health effects compared to no screening) of the investigated screening scenarios as the incremental cost-effectiveness ratios (ICER, the ratio of incremental costs to incremental health effects of a screening policy relative to its next best alternative), which can give misleading cost-effectiveness estimates.[20] Furthermore, these studies considered limited numbers of screening scenarios, providing little information on the effects of screening eligibility criteria, and may have had insufficient numbers of comparator scenarios to yield correct ICERs.[21] The aim of this study was to investigate the benefits (such as lung cancer mortality reduction and the number of life-years gained), harms (such as the number of false-positive results and occurrence of overdiagnosis), and cost-effectiveness of many different lung cancer screening scenarios for the population of Ontario, overcoming some of the limitations of previous studies.

Methods

Ethics Statement

This study was approved by the Research Ethics Board of Sunnybrook Health Sciences Centre on behalf of the Institute for Clinical Evaluative Sciences (ICES). Individual consent for access to de-identified data was not required.

MISCAN-Lung

The MIcrosimulation SCreening ANalysis (MISCAN) Lung model was used for this analysis. Other versions of the MISCAN model have been used to investigate the cost-effectiveness of screening programs for breast, colorectal, cervical, and prostate cancers.[22-25] The MISCAN-Lung model used in these analyses was previously calibrated to individual-level data from the NLST and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial, from which information on the preclinical duration of lung cancer and the effectiveness of CT screening were derived.[26,27] MISCAN-Lung was one of the models used to inform the USPSTF on their

recommendations for lung cancer screening.[13] The structure of the model and its underlying assumptions have been described previously in Chapters 1 and 2 of this thesis; the characteristics of the investigated population are described in the following section and in the supplementary material of this Chapter, in the section: “Smoking behavior and smoking related mortality”.[26,27] In brief, MISCAN-Lung simulates life histories for each individual in the considered population from birth until death in the presence and absence of screening. For each individual, a smoking history is generated based on data on the investigated population. A person's smoking history influences the probability of developing preclinical lung cancer as well as the probability of dying from other causes. The model considers four histological types of lung cancer: adenocarcinoma/large cell carcinoma/bronchioloalveolar carcinoma, squamous cell carcinoma, other non-small cell carcinoma, and small cell carcinoma. Once preclinical lung cancer has developed, it is assumed to progress through stages IA to IV. During each stage, the cancer may be detected due to symptoms, after which the person is assumed to undergo treatment with associated treatment costs. Lung cancer survival after clinical diagnosis is dependent on the histology and stage of the cancer and the person's gender. If the screening component is activated, preclinical lung cancers may be detected by screening (at the expense of screening-related costs), which may alter a person's life history: detection by screening allows treatment at an earlier stage, which may cure the individual, allowing him or her to resume a normal (lung cancer-free) life history. The probability that an individual is cured due to early detection differs by the stage at detection.

Screening may also result in serious harms, such as overdiagnosis (the detection of a disease that would never have been detected if screening had not occurred), which may lead to unnecessary (invasive) follow-up procedures, treatments, and anxiety. The effects of screening are derived through utilizing information on the preclinical duration of lung cancer, the screen-detectability of lung cancer, and relevant information on the examined population (such as smoking behavior and other-cause mortality corrected for smoking history) to model the life histories of individuals in the presence and absence of screening.[28] Through comparing the life histories in the presence of screening with the corresponding life histories in the absence of screening, MISCAN-Lung can estimate the effectiveness and costs of screening scenarios.

Simulated Population

Three different birth cohorts were investigated: 1940-1949 (ages 66-75 in 2015), 1950-1959 (ages 56-65 in 2015), and 1960-1969 (ages 46-55 in 2015). These cohorts represent approximately 5.2 million individuals in 2016 in the age range for which the USPSTF currently recommends lung cancer screening.[12,29] Birth tables for each cohort were derived from information from Statistics Canada and the Ontario Ministry of Finance.[29,30] Ontario-specific data on smoking behavior were used to model smoking initiation and cessation probabilities and the average number of cigarettes smoked per day (divided into five categories) by age and gender for each cohort.[15,31-34] Life tables by birth year and gender were extracted from the Canadian Human Mortality Database and adjusted for smoking behavior and lung cancer mortality, as shown graphically for persons born in 1955 in Figures S5 and S6 in the supplementary material of this Chapter.[34] Further information on the methods used to model smoking behavior and adjustment of the life tables for smoking behavior is provided in the supplementary material of this Chapter, in the section: "Smoking behavior and smoking related mortality". The age- and gender-specific lung cancer incidence, histology proportions, and stage proportions estimated by MISCAN-Lung were compared to observed data from the Ontario Cancer Registry from 2007-2009, in which screening did not occur, and are detailed in the supplementary material of this Chapter, in the section: "Lung cancer incidence in Ontario".[35]

Screening Scenarios

In total, 576 potential screening scenarios were evaluated. The evaluated scenarios considered different combinations of the following characteristics: age to start screening; age to stop screening; screening interval; and screening eligibility regarding cumulative smoking exposure, years since smoking cessation (for former smokers, defined as individuals who have quit smoking permanently), and whether or not former smokers were excluded from further screening after they reach a maximum number of years since cessation (Table 1).

Table 1: Characteristics of the lung cancer screening scenarios evaluated by the MISCAN-Lung model

Scenario characteristic	Considered values
Age to start screening	50, 55, 60
Age to stop screening	75, 80
Screening interval	Annual, Biennial
Minimum cumulative smoking criteria*	
Pack-years (NLST-like scenarios)	20 pack-years, 30 pack-years, 40 pack-years
Minimum number of years smoked and minimum number of cigarettes per day during years smoked (NELSON-like scenarios)	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day, 20 years of smoking at least 15 cigarettes per day **or** 25 years of smoking at least 10 cigarettes per day, 25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day, 30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day, 25 years of smoking at least 20 cigarettes per day **or** 30 years of smoking at least 15 cigarettes per day
Additional smoking related criteria	
Maximum number of years since smoking cessation to be eligible for first screening invitation	10, 15, 20
Exclusion from further screening after reaching the maximum number of years since smoking cessation	No, Yes

**Either pack-years or minimum number of years smoked and minimum number of cigarettes per day during years smoked are used in an evaluated screening scenario.*

Two types of cumulative smoking criteria, of which one is used at a time in an evaluated scenario, were distinguished: the first type was based on the cumulative number of pack-years, used in the NLST ("NLST-like"); the second type of cumulative smoking criteria was based on the criteria used in the Dutch-Belgian lung cancer screening trial (NELSON), which evaluated a person's smoking duration and average number of cigarettes per day separately ("NELSON-like").[1,36] Of the 576 screening scenarios, 216 screening scenarios considered "NLST-like" criteria (including the criteria currently recommended by the USPSTF), whereas 360 screening scenarios considered "NELSON-like" criteria. Perfect attendance to screening was assumed for the base-case investigation. Estimations regarding screen-related procedures such as false-positive results (defined as receiving a positive screening test result when lung cancer is not found after diagnostic workup), follow-up CT examinations, screen-related biopsies/bronchoscopies, and non-lung cancer related surgeries were derived from individual-level data from the CT-arm of the NLST and are described in the supplementary material of this Chapter, in the section: "Screening related follow-up procedures". Limited

information on morbidity and mortality has been reported from randomized controlled trials. In a subgroup analysis of NELSON participants, 1% of participants were found to have incidental findings that required additional work-up procedures or treatment.[37] However, no information on morbidity or mortality was reported. In the NLST, 0.06% of the positive screening tests in the low-dose CT group that did not result in a diagnosis of lung cancer were associated with a major complication after an invasive procedure.[1] Overall, six individuals with a positive screening test in the low-dose CT group that did not result in a diagnosis of lung cancer died within 60 days after an invasive diagnostic procedure (0.04%), but it was unknown whether these deaths were caused by complications of the diagnostic procedures.[1] Thus, given the low occurrence of invasive procedures along with a low frequency of major complications, the occurrence of morbidity or death related to screen-related follow-up procedures is expected to be minor.

Costs

The analyses were conducted from a third-party health care payer perspective. Fully allocated costs for lung cancer treatment were estimated by stage, age, and gender from the date of diagnosis until the date of death or last known date of follow-up (by person-month) through data from the Ontario Health Insurance Plan (OHIP), the Canadian Institute for Health Information (CIHI), the Ontario Drug Benefit Plan database, the Ontario Chronic Care database, the Ontario Home Care database, and the Ontario New Drug Funding Program. These datasets were linked using unique encoded identifiers and were analyzed at the ICES. Controls without a lung cancer diagnosis from the Registered Persons Database (a roster of all OHIP beneficiaries) were matched to 12,713 staged cases of lung cancer from the Ontario Cancer Registry (10 controls matched per case), based on age, sex, median household income, and census tract on the date of diagnosis of the case. Fully allocated costs were estimated similarly for controls. The fully allocated costs of controls were subtracted from the fully allocated costs for cases in order to obtain the attributable costs of lung cancer care by phase of care (initial, continuing, and terminal care).[38] By incorporating the fully allocated costs of lung cancer care by phase of care, it is taken into account that individuals whose lung cancer death is averted will on average incur higher costs over their remaining lifetime. In MISCAN-Lung, the attributable costs for stage I were assumed for the

modeled stages IA and IB, whereas the attributable costs of stage III were assumed for the modeled stages IIIA and IIIB.

Each person's eligibility for lung cancer screening was assumed to be free of misclassification error. Therefore, upon entering the eligible age range for the considered screening scenarios, each ever-smoking individual was assumed to receive an invitation for a lung cancer risk assessment. It was assumed that half of all ever-smokers would accept this invitation; half of the individuals who participated in the risk assessment were assumed to request a consultation with a primary care physician about their risk. Costs for screening-related events were determined using 2013 data from OHIP and CIHI. The costs for screening invitations and fixed costs related to the screening program, such as costs for the screening registry, program infrastructure, communications, and advertising, were derived from those incurred in the recent establishment of the colorectal screening program administered by Cancer Care Ontario. The costs for risk assessments were estimated assuming that screening program staff trained in health communication would administer the assessments. Fixed costs were counted up to the year in which the last individuals in the cohorts are eligible for screening (2045 for screening scenarios that end at age 75, 2050 for screening scenarios that end at age 80). All costs were expressed in Canadian dollars (using May 2013 levels as a base) and were adjusted to reflect the May 2015 prices for health care services using the Ontario Consumer Price Index derived from Statistics Canada.[39]

A lifetime time horizon for the costs and effects of screening was applied to each simulated person. Annual discount rates of 3% were applied to both costs and effects, using 2015 as the reference year.[40] The estimated attributable costs of lung cancer care by phase of care and the estimated costs related to the screening program are presented in Tables 2 and 3, respectively. To reflect the uncertainty in these cost estimates, sensitivity analyses were performed, which varied the costs by 50%, as described in a later section of this manuscript. Although cost-effectiveness thresholds have been proposed in the past, there is no official cost-effectiveness threshold employed in the Canadian health care system.[39] Therefore, a cost-effectiveness threshold of $50,000 Canadian dollars ($41,114 in May 1, 2015, U.S. dollars) per life-year gained was chosen, similar to previous Canadian cost-effectiveness studies.[41,42]

Table 2: Attributable costs (in Canadian dollars) estimates used in the MISCAN-Lung model

Attributable costs of the initial care phase of lung cancer					Men				Women
Age-group / Lung cancer stage	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	
Younger than 60	$30,272	$41,283	$49,676	$45,374	$27,172	$42,441	$44,576	$43,170	
60-69	$29,440	$37,959	$45,788	$38,073	$27,454	$37,773	$47,589	$39,411	
70-79	$31,603	$39,444	$41,832	$32,659	$27,178	$43,439	$42,779	$32,493	
Older than 80	$27,391	$24,669	$32,569	$29,886	$27,284	$32,394	$29,414	$29,035	
Attributable costs of the continuing care phase of lung cancer					**Men**				**Women**
Age-group / Lung cancer stage	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	
Younger than 60	$5,617	$8,451	$10,458	$18,683	$4,859	$6,777	$8,626	$13,716	
60-69	$3,371	$4,276	$7,270	$5,004	$5,559	$7,424	$9,214	$12,062	
70-79	$3,019	$1,938	$5,011	$5,616	$2,124	$6,447	$6,999	$5,656	
Older than 80	$4,785	$1,938	$1,689	$5,616	$2,022	$6,447	$7,475	$10,882	
Attributable costs of the terminal care phase of lung cancer (death due to causes other than lung cancer)					**Men**				**Women**
Age-group / Lung cancer stage	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	
Younger than 60	$17,174	$21,061	$17,775	$23,884	$13,741	$2,332	$4,689	$14,284	
60-69	$13,596	$10,586	$13,690	$17,548	$13,741	$20,660	$5,591	$6,869	
70-79	$15,887	$15,887	$9,858	$9,032	$15,875	$10,999	$12,917	$4,300	
Older than 80	$15,887	$15,887	$9,368	$8,243	$21,554	$28,188	$2,217	$6,030	
Attributable costs of the terminal care phase of lung cancer (lung cancer death)					**Men**				**Women**
Age-group / Lung cancer stage	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	
Younger than 60	$72,167	$70,323	$84,041	$98,611	$51,164	$71,024	$89,322	$94,906	
60-69	$73,085	$82,296	$84,828	$95,000	$51,164	$81,256	$79,563	$87,113	
70-79	$68,187	$91,114	$77,067	$97,320	$68,844	$73,176	$77,424	$89,056	
Older than 80	$55,413	$141,182	$69,807	$78,002	$68,844	$59,590	$83,353	$73,453	

Costs were estimated per phase of lung cancer care per person-year of treatment by gender, age, and stage of disease for Ontario, Canada. Costs were estimated from a third-party health care payer perspective by matching lung cancer patients in the Ontario Cancer Registry to persons registered in the OHIP, free of lung cancer, by age, sex, median household income, and census tract on the date of diagnosis of the case. Additionally, data from the CIHI, the Ontario Drug Benefit Plan database, the Ontario Chronic Care database, the Ontario Home Care database, and the Ontario New Drug Funding Program were used.

Table 3: Costs (in Canadian dollars) of screening-related events and fixed program costs used in the MISCAN-Lung model

Cost estimates related to screening invitations	
Description	**Unit costs**
Invitation to assess lung cancer risk	$5
Lung cancer risk assessment	$32
Visit with primary care physician with regards to lung cancer risk assessment	$67
Initial and repeat screening invitations	$3
Cost estimates for screening and follow-up related procedures	
Description	**Unit costs**
Screening CT examination	$430
Follow-up CT examination	$430
Visit with primary care physician with regards to the results of a follow-up chest CT	$41
Percutaneous cytologic analysis/bronchoscopy/biopsy	$1,355
Non-lung cancer surgery for potentially benign disease	$11,844
Fixed program cost estimates per year (per 100,000 individuals alive in 2015)*	
Description	**Yearly costs**
First year	$823,321
Second year up to the year in which the last individuals in the cohorts are eligible for screening (2045 for screening scenarios that end at age 75, 2050 for screening scenarios that end at age 80)	$411,660

Costs for screening-related events were estimated using 2013 data from the OHIP and the CIHI. The costs for program invitations and fixed costs related to the screening program, such as costs for the screening registry, program infrastructure, communications, and advertising, were derived from those incurred in the recent establishment of ColonCancerCheck, the colorectal screening program administered by Cancer Care Ontario. The costs for lung cancer risk assessments were estimated assuming that screening program staff trained in health communication would administer the assessments.

**The fixed costs per 100,000 individuals alive in 2015 consist of one-time, first year only startup costs for Information Technology infrastructure ($411,661 Canadian dollars), annual maintenance costs for Information Technology infrastructure ($61,749 Canadian dollars), annual costs for maintaining main screening centers ($144,081 Canadian dollars), annual costs for communications and advertising ($102,915 Canadian dollars), and annual costs for provincial program management and evaluation ($102,915 Canadian dollars).*

Benefits, Harms, and Cost-Effectiveness of Screening Scenarios

For each screening scenario, the number of lung cancer deaths prevented, life-years gained, proportion of individuals ever screened, number of CT examinations, screen-related biopsies/bronchoscopies, false-positive screens, non-lung cancer-related surgeries, overdiagnoses, and costs were compared with a situation in which screening does not occur, from 2015 onward. Screening scenarios that were more costly and less effective (i.e., fewer life-years gained) than other scenarios were ruled out as non-efficient by simple dominance. Scenarios that were more costly and less effective than a combination of other scenarios were also ruled out as non-efficient by extended dominance. The remaining screening scenarios constitute the frontier of efficient screening scenarios, i.e., the efficient frontier. For each efficient screening scenario, the ICER was determined, calculated as the incremental net costs per incremental life-year gained compared to the previous efficient screening scenario.

Sensitivity Analyses

A number of sensitivity analyses were performed to investigate which groups of cost estimates and attendance assumptions have the greatest influence on the cost-effectiveness estimates by varying the costs for CT examinations by 50% compared with the base-case analyses, varying the attributable costs of lung cancer care by phase of care by 50% compared with the base-case analyses, and imperfect attendance rates for screening: low attendance (33% overall compliance rate), average attendance (55% overall compliance rate), and high attendance (64% overall compliance rate). Sensitivity analyses were performed for all 576 scenarios to investigate the effects of variations in assumptions on the composition of the efficient frontier.

Results

Screening Scenarios on the Efficient Frontier

The net discounted costs and life-years gained (from 2015 onwards) for each scenario were used to determine the screening scenarios on the efficient frontier, i.e., the scenarios that provide the highest number of life-years gained for their costs, in the base-case analysis, as shown in Figures 1 and 2. The scenarios that are on the efficient frontier are described in

Table 4 and shown in Figure 2. A complete overview of the net discounted costs and life-years gained of all investigated screening scenarios is presented in Figures S13-S20 in the supplementary material of this Chapter. All outcomes were reported per 100,000 individuals alive in 2015. All scenarios on the efficient frontier consist of annual screening (Table 4), while biennial screening is dominated. Assuming a cost-effectiveness threshold of $50,000 Canadian dollars ($41,114 May 1, 2015, U.S. dollars) per life-year gained as acceptable for the Canadian health care system, Scenario #2 was considered optimal: current and former smokers (who quit 10 years ago) who smoked 40 pack-years would be screened annually between ages 55-75, yielding an ICER of $41,136 Canadian dollars ($33,825 May 1, 2015, U.S. dollars) per life-year gained. If 50% lower or higher attributable costs were assumed, the ICER of this scenario was estimated to be $38,240 ($31,444 May 1, 2015, U.S. dollars) or $48,525 ($39,901 May 1, 2015, U.S. dollars), respectively. If 50% lower or higher costs for CT examinations were assumed, the ICER of this scenario was estimated to be $28,630 ($23,542 May 1, 2015, U.S. dollars) or $73,507 ($60,443 May 1, 2015, U.S. dollars), respectively.

In addition, the benefits and harms of all scenarios on the efficient frontier were examined (Table 5). Scenario #2 would reduce lung cancer mortality in the overall population (which includes non-eligible individuals) by 9.05%, preventing 251 lung cancer deaths and gaining 2,531 life-years (undiscounted) over the lifetime of the program (i.e., on average, 10.08 life-years would be gained for each lung cancer death prevented). However, in Scenario #2, 9.56% of the overall population would receive at least one screen, requiring 91,692 CT screens and follow-up examinations. This scenario would lead to 14,729 false-positive screens and 163 surgeries for potentially benign disease (in persons in whom lung cancer is not detected) and 350 biopsies/bronchoscopies (in persons in whom lung cancer is not detected). Ultimately, 12.53% of all screen-detected cancers would be overdiagnosed, leading to 82 overdiagnosed cases.

Figure 1: The cost-effectiveness of all 576 investigated lung cancer screening scenarios in the base-case analysis

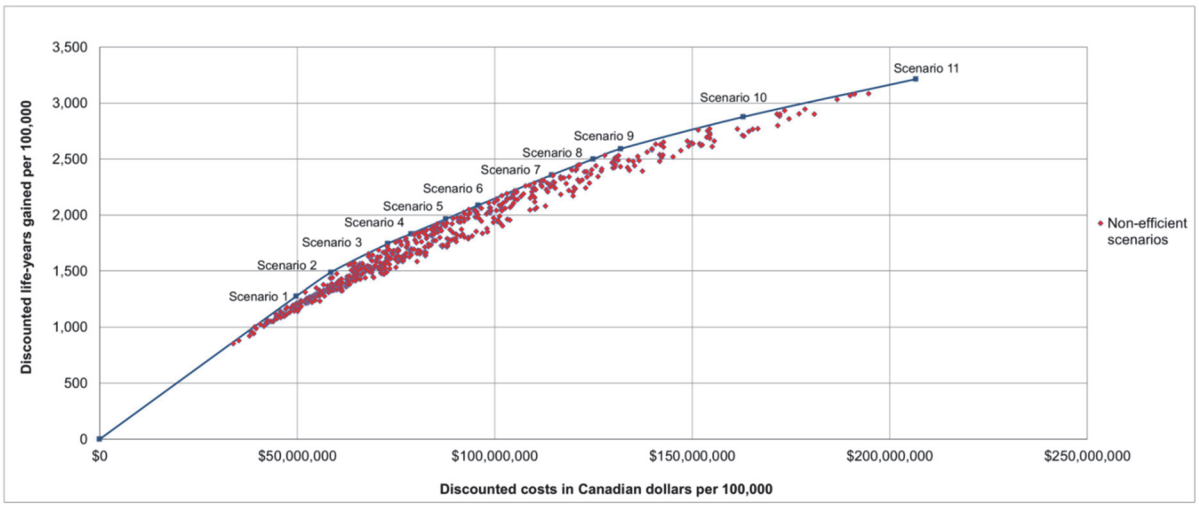
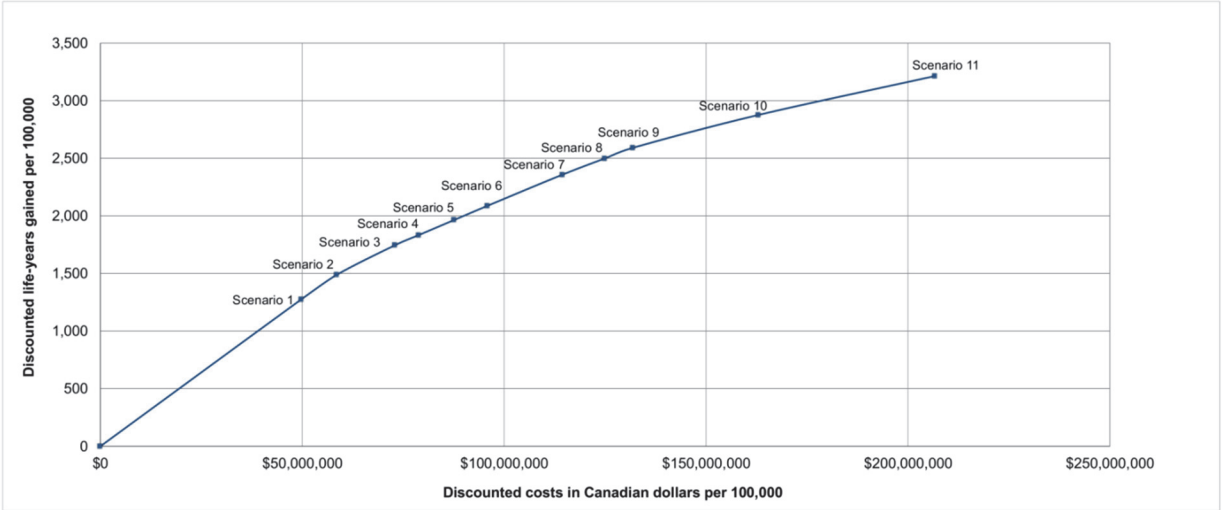


Figure 2: The incremental cost-effectiveness of the lung cancer screening scenarios on the efficient frontier



Results in both Figures are presented per 100,000 individuals alive in 2015 and are discounted by 3% annually. Scenarios on the efficient frontier are described in Table 4.

Table 4: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier

Scenario #	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs compared to no screening (in Canadian dollars) per 100,000	Discounted life-years gained per 100,000	Costs (in Canadian dollars) per life-year gained (discounted) / ACER compared to no screening	ICER compared to the previous efficient scenario
#1	60	75	Annual	10	40 pack-years (NLST-like)	Yes	$49,768,886	1,276	$39,006	-
#2	55	75	Annual	10	40 pack-years (NLST-like)	Yes	$58,549,938	1,489	$39,311	$41,136
#3	55	75	Annual	10	30 pack-years (NLST-like)	Yes	$72,978,421	1,746	$41,801	$56,262
#4	55	80	Annual	10	30 pack-years (NLST-like)	Yes	$78,858,485	1,834	$43,001	$66,802
#5	55	75	Annual	15	30 pack-years (NLST-like)	Yes	$87,658,495	1,965	$44,600	$66,885
#6	55	80	Annual	15	30 pack-years (NLST-like)	Yes	$95,859,980	2,088	$45,916	$67,065
#7	55	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$114,462,449	2,359	$48,530	$68,675
#8	55	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$124,978,314	2,500	$49,998	$74,557
#9	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$131,929,978	2,592	$50,901	$75,370
#10	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$162,994,771	2,877	$56,661	$109,083
#11	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$206,703,139	3,214	$64,304	$129,394

Results are per 100,000 individuals alive at the start of 2015. An annual discount rate of 3% annually was applied to costs and life-years gained.

Table 5: Overview of selected benefits and harms (per 100,000 individuals alive at the start of 2015) of the screening scenarios on the efficient frontier (effect estimates are not discounted)

Scenario #*	Percentage of the total population ever screened	CT screens and follow-up examinations (per 100,000)	Lung cancer mortality reduction in the total population (%)	Lung cancer deaths prevented† (per 100,000)	Life-years gained (per 100,000)	Average number of life-years gained per lung cancer death averted	Percentage of screen-detected cancers that is overdiagnosed	Number of overdiagnosed lung cancers (per 100,000)‡	False-positive screens (per 100,000)	Number of non-lung cancer surgeries due to screening§ (per 100,000)	Biopsies due to screening§ (per 100,000)
#1	8.74%	73,248	8.24%	229	2,170	9.48	13.06%	80	11,937	132	283
#2	9.56%	91,692	9.05%	251	2,531	10.08	12.53%	82	14,729	163	350
#3	13.03%	125,320	10.50%	292	2,993	10.25	12.31%	93	20,145	223	479
#4	13.04%	135,410	11.71%	325	3,159	9.72	14.43%	127	21,575	239	514
#5	15.41%	161,159	11.93%	331	3,388	10.24	12.34%	106	25,698	285	612
#6	15.42%	177,014	13.58%	377	3,624	9.61	14.68%	150	27,947	311	667
#7	16.06%	225,062	15.71%	436	4,129	9.47	14.89%	176	34,933	390	838
#8	16.06%	255,207	17.32%	481	4,422	9.19	15.48%	203	39,228	438	943
#9	16.19%	270,354	17.59%	489	4,577	9.36	15.36%	204	41,414	463	997
#10	19.99%	355,448	19.51%	542	5,142	9.49	15.15%	221	54,259	607	1,308
#11	26.13%	473,383	21.92%	609	5,774	9.48	15.14%	248	72,221	809	1,742

** Scenario details are provided in Table 4.*

[2] Number of lung cancer deaths per 100,000 without screening: 2,777.

[3] Number of lung cancer cases per 100,000 without screening: 3,522.

§ For persons in whom lung cancer was not detected by screening.

Based on the estimated number of individuals in the examined cohorts in 2016, Scenario #2 is estimated to screen 499,261 individuals at least once, require 4,788,523 CT examinations, and prevent 13,108 lung cancer deaths, while 4,282 cases of lung cancer would be overdiagnosed.[29,30] The average annual non-discounted costs compared to no screening would be approximately $1,400,000 Canadian dollars ($1,151,178 May 1, 2015, U.S. dollars) over the considered time period; however, the annual costs are higher in the first years compared to later years, due to diminishing numbers of individuals meeting the eligibility criteria. For example, the average non-discounted costs compared to no screening are approximately $5,000,000 Canadian dollars ($4,111,350 May 1, 2015, U.S. dollars) for 2015-2020 compared with approximately $1,600,000 Canadian dollars ($1,315,632 May 1, 2015, U.S. dollars) in 2030-2035.

Effects of Screening Scenario Characteristics on Cost-Effectiveness

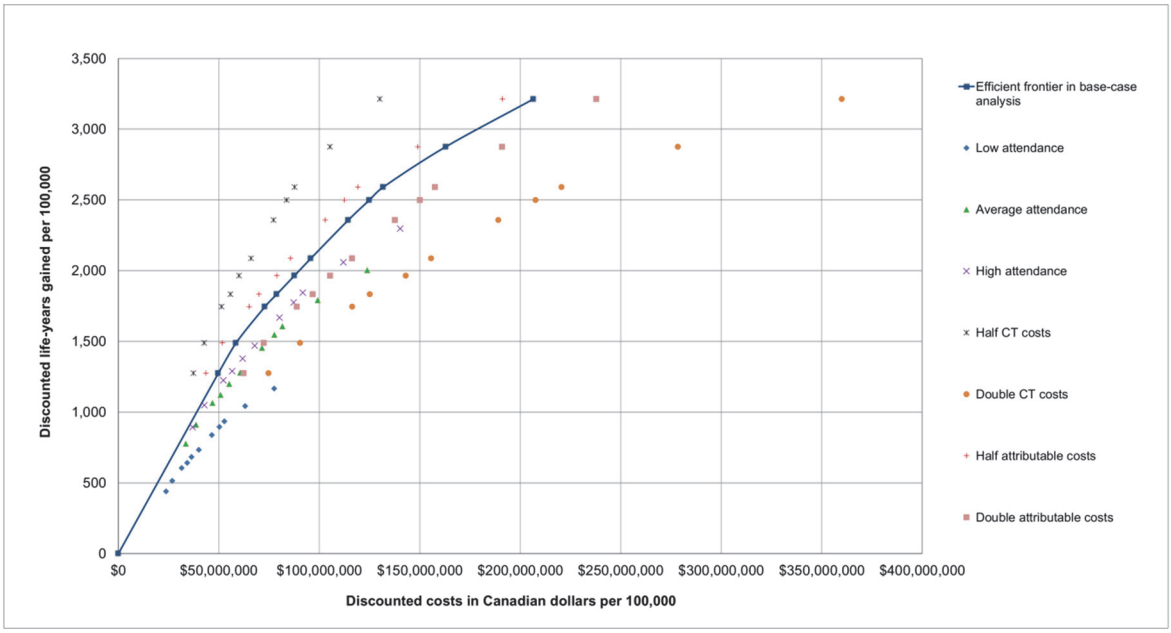
Scenarios with older starting ages have lower costs compared with scenarios that start at younger ages but also yield a smaller number of life-years gained (shown in Figure S13 in the supplementary material of this Chapter). Raising the age to stop screening from 75 to 80 increases the costs and the number of life-years gained, but differences are modest (shown in Figure S14 in the supplementary material of this Chapter). A comparison of scenarios by smoking eligibility criteria indicates that there is little difference between using NLST-like or NELSON-like smoking eligibility criteria (shown in Figure S15 in the supplementary material of this Chapter). Figures S16-S18 in the supplementary material of this Chapter demonstrates the importance of cumulative smoking criteria on cost-effectiveness. Each increase in the cumulative smoking requirement for enrollment substantially decreases the costs while modestly decreasing the number of life-years gained in both NLST-like and NELSON-like (shown in Figures S16 and S17 the supplementary material of this Chapter) screening scenarios. In other words, scenarios that apply stringent cumulative smoking eligibility criteria are closer to the efficient frontier than those that apply less restrictive cumulative smoking eligibility criteria. In general, scenarios that require only 20 pack-years are dominated by scenarios that apply more stringent (higher) pack-year criteria. Increasing the maximum number of years since smoking cessation (shown in Figure S18 in the supplementary material of this Chapter) and not excluding individuals from further screening once they reach the maximum number of years since cessation (shown in Figure

S19 in the supplementary material of this Chapter) both increase the costs and the numbers of life-years gained. However, the effects of these criteria are less pronounced than those related to cumulative smoking requirements. Figure S20 in the supplementary material of this Chapter shows the effects of annual screening compared with biennial screening. Although biennial screening scenarios have substantially lower costs compared with annual screening scenarios, the life-years gained are also substantially lower. Figure S20 demonstrates that annual screening scenarios dominate biennial screening scenarios.

Sensitivity Analyses

Altering assumptions about attendance rates, CT examination costs, and attributable costs impacted the scenarios on the efficient frontier in the base-case analysis to varying degrees. Figure 3 provides an overview of the scenarios on the efficient frontier in the base-case analysis along with the discounted life-years gained and costs for these scenarios in the sensitivity analyses.

Figure 3: The incremental cost-effectiveness of the lung cancer screening scenarios on the efficient frontier and their corresponding cost-effectiveness throughout different sensitivity analyses



Results are presented per 100,000 individuals alive in 2015 and are discounted by 3% annually. The relative ranking of the scenarios is consistent across sensitivity analyses (i.e., if a scenario is more costly and gains more life-years than another scenario in the base-case analysis, this is also the case in all sensitivity analyses).

Altering assumptions also impacted the composition of the efficient frontier, as shown in the supplementary material of this Chapter, in the section: “Impact of sensitivity analyses”. When the attendance rates were varied, lower attendance rates were observed to shift scenarios with less restrictive criteria, especially with regards to smoking behavior, on the efficient frontier (shown in Tables S7-S9 in the supplementary material of this Chapter). This may be due to the fixed costs of the screening scenarios, which are independent of the number of screened individuals; at lower levels of participation these costs have a greater influence on the cost-effectiveness than the costs of CT examinations.

When the attributable costs were varied, it was observed that halving the attributable costs had little effect on the scenarios on the efficient frontier (shown in Table S10 in the supplementary material of this Chapter). When the attributable costs were doubled, it was observed that scenarios with less restrictive criteria, especially with regards to smoking cessation, were shifted on the efficient frontier (shown in Table S11 in the supplementary material of this Chapter).

When the costs of CT examinations were varied, it was observed that halving the costs of CT examinations also shifted scenarios with less restrictive criteria, in particular with regards to smoking cessation, on the efficient frontier (shown in Table S12 in the supplementary material of this Chapter). Doubling the costs of CT examinations had the greatest effect of all sensitivity analyses; the scenarios with the most restrictive criteria with regards to age and smoking were shifted on the efficient frontier and the least costly scenarios on this efficient frontier favored biennial screening (shown in Table S13 in the supplementary material of this Chapter).

Scenario #2 was on the efficient frontier across all sensitivity analyses, with the exception of assuming the lowest attendance rates (shown in table S14 in the supplementary material of this Chapter). In contrast, although Scenario #5 closely resembles the eligibility criteria that were used in the NLST, it was not on the efficient frontier in any of the sensitivity analyses.

Discussion

This simulation study indicates that lung cancer screening can be cost-effective in a population-based setting when eligibility is restricted to high-risk groups. In contrast, utilizing loose eligibility criteria yields non-optimal and potentially cost-ineffective scenarios, as the cost-effectiveness of lung cancer screening is highly dependent on scenario characteristics, primarily the smoking eligibility criteria. Scenarios that utilize stringent smoking eligibility criteria are more cost-effective than scenarios that utilize less restrictive smoking eligibility criteria due to a focus on individuals at higher risk of developing lung cancer. This greatly reduces the number of screening examinations while still screening those at highest risk. Thus, the level of lung cancer risk at which an individual is eligible for lung cancer screening should be considered before implementing lung cancer screening policies. Future research should investigate the cost-effectiveness of lung cancer screening selection based on accurate lung cancer risk prediction models using suitable risk thresholds.[43-45]

The results of this study suggest that the greater reduction in lung cancer mortality and number of life-years gained by annual screening outweigh the costs of the additional number of CT examinations compared with biennial screening, which has previously been suggested to be equally or more cost-effective than annual screening.[18,46] However, previous studies indicated that lung cancer may be more difficult to detect in stage IA with biennial screening.[27] As survival in stage IA is considerably higher compared with other stages, the potential for mortality reduction and life-years gained is higher for annual screening compared to biennial screening.[47] This is supported by the modeling study that informed the USPSTF, which showed that annual screening provides substantial benefits over biennial screening at modest diminishing returns.[13] Previous studies that examined the cost-effectiveness of lung cancer screening only considered limited numbers of screening scenarios, which provided limited information on the effects of scenario characteristics.[4-8]

The results of this study suggest that scenario characteristics, especially smoking eligibility criteria and screening interval, influence the cost-effectiveness of a scenario and suggest that a large variety of scenarios should be considered. In addition, considering a wide variety

of screening scenarios provides sufficient comparator scenarios to yield appropriate ICERs.[21] Previous studies often reported the ACERs of the investigated screening scenarios as the ICERs, which can give misleading cost-effectiveness estimates.[20] This study provides both the ACERs and the ICERs of the scenarios on the efficient frontier, in contrast to previous studies that generally did not report an efficient frontier.[4-8] The robustness of the scenarios on the efficient frontier in this study is demonstrated by the sensitivity analyses of all 576 scenarios.

This study incorporates both allocated costs for all screening-related procedures and attributable costs for various stages of lung cancer care, which were derived from government data in a province with universal health care, which allows for more comprehensive and accurate cost estimates compared with other studies. Furthermore, this study incorporates detailed information on smoking behavior and smoking-related mortality in contrast to previous studies. Finally, although the majority of previous studies only reported the number of life-years gained, this study reports a variety of benefits (such as lung cancer mortality reduction and the number of life-years gained) and harms (such as the number of false-positive results and the occurrence of overdiagnosis).

This study has some limitations; for example, quality of life was not incorporated in the analyses. There may be some differences in quality of life between annual and biennial screening, as more frequent screening will increase the impact of screening and follow-up-related effects on quality of life. However, results from the NELSON trial indicate that although CT lung cancer screening has a minor impact on quality of life in the short term, the long-term effects are negligible.[48,49] In addition, utility estimates for lung cancer care are highly variable.[50] Another limitation is that assumptions for follow-up procedures were based on data from the NLST, which may not be generalizable to a population-based setting, as screening algorithms with reduced false-positive rates are being investigated.[1,51-53] By reducing the false-positive rates, the number of unnecessary follow-up CTs and invasive diagnostic procedures may be reduced as well, further improving the cost-effectiveness of lung cancer screening. Finally, although fully allocated costs for lung cancer care and observed costs for the administration of a cancer screening program were incorporated in the analyses, the government of Ontario only reimburses the physician costs of a CT

examination. However, capital investments would be required to acquire the CT scanners necessary to implement a lung cancer screening program, which could influence the costs per CT examination. Conversely, the increased CT capacity could potentially lead to discounts on the costs per CT examination.

This study used a cost-effectiveness threshold of $50,000 Canadian dollars ($41,114 May 1, 2015, U.S. dollars) per life-year gained, similar to previous Canadian cost-effectiveness studies.[41] However, the acceptable ratio between costs and effects differs between countries. For example, although a cost-effectiveness threshold of U.S. $100,000 per quality-adjusted life-year has been proposed for the U.S., the United Kingdom's National Institute for Health and Care Excellence uses a £20,000-£30,000 ($30,274-$45,411 May 1, 2015, U.S. dollars) threshold to determine cost-effectiveness.[3,54,55] Thus, the optimal screening scenario depends in part on the chosen cost-effectiveness threshold: if a cost-effectiveness threshold of $60,000 Canadian dollars ($49,336 May 1, 2015, U.S. dollars) per life-year gained was chosen, Scenario #3 (annual screening for persons aged 55-75 who smoked at least 30 pack-years and currently smoke or quit smoking less than 10 years ago) would have been considered the optimal scenario. However, the ICER of Scenario #2 (annual screening for persons aged 55-75 who smoked at least 40 pack-years and currently smoke or quit smoking less than 10 years ago) remained below the proposed cost-effectiveness threshold of $50,000 Canadian dollars per life-year gained in 5 out of 7 sensitivity analyses (71.4%) with a range of $28,630-$73,507 Canadian dollars per life-year gained. This suggests that both the dominance and cost-effectiveness of Scenario #2 are robust across various sensitivity analyses.

Although our results suggest that a uniform biennial screening interval is dominated by a uniform annual screening interval, recent studies suggest it may be possible to identify individuals for whom biennial screening intervals could be recommended. NLST participants with a negative prevalence screen had a substantially lower risk of developing lung cancer compared to individuals with a positive prevalence screen.[56] Results from the NELSON trial suggest that the 2-year probability of developing lung cancer after a CT screen varied substantially by nodule size and volume doubling time.[53] Future research should evaluate whether the interval between screens can be varied based on previous screening results and

what impact this has on cost-effectiveness. In addition, precision medicine could improve the treatment of selected individuals, and biomarkers might help to distinguish between indolent nodules and aggressive nodules requiring rapid diagnosis and treatment. The impacts of these future developments need to be assessed. In conclusion, this study indicates that lung cancer screening can be cost-effective in a population-based setting if stringent smoking eligibility criteria are applied. Annual screening scenarios are more cost-effective than biennial screening scenarios.

Acknowledgments

We thank the National Cancer Institute (NCI) for access to NCI's data collected by the National Lung Screening Trial (NLST). The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI. Finally, we thank the NLST study participants for their contributions to this study. The opinions, results, views, and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by the Institute for Clinical Evaluative Sciences (ICES), which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC), the MOHLTC, Cancer Care Ontario (CCO), and the Ontario Institute for Cancer Research is intended or should be inferred. This work incorporates data analysis provided by the Ontario Tobacco Research Unit, Toronto, Ontario. Parts of this manuscript are based on data and information compiled and provided by the Canadian Institute for Health Information (CIHI). However, the analyses, conclusions, opinions, and statements expressed herein are those of the authors and not necessarily those of CIHI. Parts of this manuscript are based on data and information provided by CCO. The opinions, results, views, and conclusions reported in this paper are those of the authors and do not necessarily reflect those of CCO. No endorsement by CCO is intended or should be inferred.

References

1. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
2. Black WC, Gareen IF, Soneji SS, et al. Cost-Effectiveness of CT Screening in the National Lung Screening Trial. *New England Journal of Medicine* 2014; **371**(19): 1793-802.
3. Neumann PJ, Cohen JT, Weinstein MC. Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold. *New England Journal of Medicine* 2014; **371**(9): 796-7.
4. McMahon PM, Kong CY, Bouzan C, et al. Cost-Effectiveness of Computed Tomography Screening for Lung Cancer in the United States. *Journal of Thoracic Oncology* 2011; **6**(11): 1841-8.
5. 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 Affairs* 2012; **31**(4): 770-9.
6. Villanti AC, Jiang Y, Abrams DB, Pyenson BS. A Cost-Utility Analysis of Lung Cancer Screening and the Additional Benefits of Incorporating Smoking Cessation Interventions. *PLOS ONE* 2013; **8**(8): e71379.
7. Mahadevia PJ, Fleisher LA, Frick KD, Eng J, Goodman SN, Powe NR. Lung cancer screening with helical computed tomography in older adult smokers: A decision and cost-effectiveness analysis. *Journal of the American Medical Association* 2003; **289**(3): 313-22.
8. Goffin JR, Flanagan WM, Miller AB, et al. Cost-effectiveness of lung cancer screening in canada. *JAMA Oncology* 2015; **1**(6): 807-13.
9. Raymakers AJN, Mayo J, Lam S, FitzGerald JM, Whitehurst DGT, Lynd LD. Cost-Effectiveness Analyses of Lung Cancer Screening Strategies Using Low-Dose Computed Tomography: a Systematic Review. *Applied Health Economics and Health Policy* 2016; **14**(4): 409-18.
10. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2009; **151**(10): 738-47.
11. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: A decision analysis for the u.s. preventive services task force. *Annals of Internal Medicine* 2008; **149**(9): 659-69.
12. Moyer VA, on behalf of the USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 2014; **160**(5): 330-8.
13. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
14. Corsi DJ, Boyle MH, Lear SA, Chow CK, Teo KK, Subramanian SV. Trends in smoking in Canada from 1950 to 2011: progression of the tobacco epidemic according to socioeconomic status and geography. *Cancer Causes & Control* 2014; **25**(1): 45-57.
15. Statistics Canada 2011. Canadian Community Health Survey (CCHS), Cycle 1.1, 2.1, 3.1, 4.1 and 2008, 2009, 2010 Annual Component surveys [Microdata]. Ottawa Ontario: Statistics Canada Health Statistics Division [producer and distributor]: Ottawa Ontario: Data Liberation Initiative [distributor].
16. Cancer Care Ontario. Cancer Fact: Lung cancer accounts for largest proportion of premature Ontario cancer deaths. Available from: http://www.cancercare.on.ca/cancerfacts.
17. Cancer Care Ontario. Cancer statistics 2012. Available from: http://www.cancercare.ca/ocs/csurv/stats/ontario/.
18. Field JK, Oudkerk M, Pedersen JH, Duffy SW. Prospects for population screening and diagnosis of lung cancer. *The Lancet*; **382**(9893): 732-41.
19. Sox HC. Better Evidence about Screening for Lung Cancer. *New England Journal of Medicine* 2011; **365**(5): 455-7.
20. O'Mahony JF, Normand C. HIQA's CEA of Breast Screening: Pragmatic Policy Recommendations are Welcome, but ACERs Reported as ICERs are Not. *Value in Health* 2015; **18**(8): 941-5.

21. O'Mahony JF, Naber SK, Normand C, Sharp L, O'Leary JJ, de Kok IMCM. Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. *Value in Health* 2015; **18**(8): 1138-51.
22. de Gelder R, Bulliard J-L, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. *European Journal of Cancer* 2009; **45**(1): 127-38.
23. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Boer R, Wilschut J, Habbema JDF. At what costs will screening with CT colonography be competitive? A cost-effectiveness approach. *International Journal of Cancer* 2009; **124**(5): 1161-8.
24. Heijnsdijk EAM, de Carvalho TM, Auvinen A, et al. Cost-effectiveness of Prostate Cancer Screening: A Simulation Study Based on ERSPC Data. *Journal of the National Cancer Institute* 2015; **107**(1): dju366-dju.
25. de Kok IMCM, van Rosmalen J, Dillner J, et al. Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost-effectiveness analysis based on a Dutch microsimulation model. *British Medical Journal* 2012; **344**.
26. Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014; **120**(11): 1713-24.
27. ten Haaf K, van Rosmalen J, de Koning HJ. Lung Cancer Detectability by Test, Histology, Stage, and Gender: Estimates from the NLST and the PLCO Trials. *Cancer Epidemiology Biomarkers & Prevention* 2015; **24**(1): 154-61.
28. ten Haaf K, de Koning HJ. Overdiagnosis in lung cancer screening: why modeling is essential. *Journal of Epidemiology and Community Health* 2015; **69**(11): 1035-39.
29. Ontario Ministry of Finance, Ontario Population Projections Update, 2011-2036. Spring 2012.
30. Statistics Canada, CANSIM, table 051-0001., http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=0510001&&pattern=&stByVal=1&p1=1&p2=1&tabMode=dataTable&csid=, Accessed on 22-01-2016.
31. Statistics Canada 2012. National Population Health Survey (NPHS), Cycle 1-9: Household Component-Longitudinal [Canada] [Microdata]. Ottawa Ontario: Statistics Canada Health Statistics Division [producer and distributor]: Ottawa Ontario: Data Liberation Initiative [distributor].
32. Statistics Canada. 1986. General Social Survey, Cycle 1, 1985 [Canada]: Health and Social Support [public use microdata file]. Ottawa, Ontario: Statistics Canada [producer and distributor].
33. Statistics Canada. Survey of Smoking Habits, 1971±1986 [Canada] [Public-use microdata file(s)]. Ottawa, Ontario: Special Surveys Division, Statistics Canada [producer and distributor]. Ottawa, Ontario: Data Liberation Initiative [distributor].
34. Canadian Human Mortality Database. Department of Demography, Université de Montréal (Canada). www.demo.umontreal.ca/chmd/ (data downloaded on 14-03-2014). [Internet].
35. https://www.cancercare.on.ca/siteinfo/contact/.
36. 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 randomized lung cancer multi-slice CT screening trial (NELSON). *International Journal of Cancer* 2007; **120**(4): 868-74.
37. van de Wiel JCM, Wang Y, Xu DM, et al. Neglectable benefit of searching for incidental findings in the Dutch--Belgian lung cancer screening trial (NELSON) using low-dose multidetector CT. *European Radiology* 2007; **17**(6): 1474-82.
38. Yabroff KR, Mariotto AB, Feuer E, Brown ML. Projections of the costs associated with colorectal cancer care in the United States, 2000–2020. *Health Economics* 2008; **17**(8): 947-59.
39. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Canadian Medical Association Journal* 1992; **146**(4): 473-81.
40. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. *Journal of the American Medical Association* 1996; **276**(15): 1253-8.

41. Telford JJ, Levy AR, Sambrook JC, Zou D, Enns RA. The cost-effectiveness of screening for colorectal cancer. *Canadian Medical Association Journal* 2010; **182**(12): 1307-13.
42. http://www.exchange-rates.org/Rate/CAD/USD/5-1-2015, accessed on 01-09-2016.
43. Kovalchik SA, Tammemägi M, Berg CD, et al. Targeting of Low-Dose CT Screening According to the Risk of Lung-Cancer Death. *New England Journal of Medicine* 2013; **369**(3): 245-54.
44. 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. *Annals of Internal Medicine* 2012; **157**(4): 242-50.
45. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.
46. Duffy SW, Field JK, Allgood PC, Seigneurin A. Translation of research results to simple estimates of the likely effect of a lung cancer screening programme in the United Kingdom. 2014; **110**(7): 1834-40.
47. Groome PA, Bolejack V, Crowley JJ, et al. The IASLC Lung Cancer Staging Project: Validation of the Proposals for Revision of the T, N, and M Descriptors and Consequent Stage-groupings in the Forthcoming (Seventh) Edition of the TNM Classification of Malignant Tumours. *Journal of Thoracic Oncology* 2007; **2**(8): 694-705.
48. van den Bergh KAM, Essink-Bot ML, Borsboom GJJM, Scholten ET, van Klaveren RJ, de Koning HJ. Long-term effects of lung cancer computed tomography screening on health-related quality of life: the NELSON trial. *European Respiratory Journal* 2011; **38**(1): 154.
49. van den Bergh KAM, Essink-Bot ML, Borsboom GJJM, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *British Journal of Cancer* 2009; **102**(1): 27-34.
50. Sturza J. A Review and Meta-Analysis of Utility Values for Lung Cancer. *Medical Decision Making* 2010; **30**(6): 685-93.
51. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of ct screening for lung cancer: A systematic review. *Journal of the American Medical Association* 2012; **307**(22): 2418-29.
52. Pinsky PF, Gierada DS, Black W, et al. Performance of Lung-RADS in the National Lung Screening Trial: A retrospective assessment. *Annals of Internal Medicine* 2015; **162**(7): 485-91.
53. Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *The Lancet Oncology* 2014; **15**(12): 1332-41.
54. http://www.exchange-rates.org/Rate/GBP/USD/5-1-2015, accessed on 01-09-2016.
55. National Institute for Health and Care Excellence. Developing NICE Guidelines: The Manual [Internet]. London: National Institute for Health and Care Excellence (NICE); 2015 Jul 22. Process and Methods Guides No. 20. http://www.ncbi.nlm.nih.gov/books/NBK310375/.
56. Patz Jr EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomized, multicentre, diagnostic screening trial. *The Lancet Oncology* 2016; **17**(5): 590-9.

Chapter 9

Supplementary material

Smoking behavior and smoking related mortality

The MISCAN-Lung model incorporates detailed information on smoking behavior in Ontario, such as smoking initiation and cessation probabilities and the average number of cigarettes smoked per day (divided into five categories) by cohort, age and gender. In addition, the effects of smoking on non-lung cancer mortality are incorporated. The following sections detail the methods and assumptions used to generate the smoking behaviors of individuals and the effects of smoking on non-lung cancer mortality in the MISCAN-Lung model.

Smoking initiation

Little data are available on Ontario-specific smoking initiation by age. Previous investigations in the U.S. indicate that the majority of smoking initiation occurs before the age of 30 for cohorts born after the early 1900's.[1,2] Anderson et al. have shown that age-specific smoking initiation probabilities (the probability that a never-smoker, a person that smoked less than 100 cigarettes during his or her lifetime, at the beginning of that year of age starts smoking) can be estimated by calibrating them to the ever-smoking prevalence at age 30 (when the majority of initiation has occurred).[2] Therefore, Ontario-specific data on the ever-smoking prevalence at age 30 for each investigated birth cohort (persons born between 1940-1949, 1950-1959 and 1960-1969) was obtained from the Canadian Community Health Surveys (CCHS), the National Population Health Surveys (NPHS), the Canada Health Survey, the General Social Surveys (years 1985 and 1991) and the Smoking Habits of Canadians Surveys.[3-8] For each cohort, age and gender specific smoking initiation probabilities were estimated in R package, using exponential functions, to match the observed ever-smoking prevalence at age 30.[9] The smoking initiation probabilities were corrected for all-cause mortality using Ontario specific all-cause mortality life tables by birth-year, age and gender, obtained from the Canadian Human Mortality Database (CHMD).[10] It was assumed persons in each cohort can initiate smoking from ages 8 to 29. Furthermore, it was assumed that the age-specific smoking initiation probabilities increase with age until age 17, after which the probability of smoking initiation decreases, as shown in previous investigations.[1] Table S1 shows the observed ever-smoking prevalence at age 30 compared to the estimated ever-smoking prevalence at age 30 for each cohort, by gender. Overall, the estimated prevalences at age 30 closely match the observed data.

Smoking cessation probabilities

Information on Ontario-specific current-, former- and never-smoker prevalences for each cohort, by gender, was obtained from the CCHS, the NPHS, Canada Health Survey, General Social Surveys (years 1985 and 1991) and the Smoking Habits of Canadians Surveys.[3-8] Smokers who attempt to quit have a high probability to relapse in their first two years since cessation, therefore, we defined former smokers as smokers who reported having quit for at least two years.[2] However, the available surveys did not inquire about the time since cessation until 2000 and surveys held after 2000 had limited sample sizes. Therefore, for each cohort, information on current-, former- and never-smoking prevalences in the year 2000 were used as calibration targets to estimate the age-specific smoking cessation probabilities (the probability that a current smoker at the beginning of that year of age ceases smoking permanently) for each cohort, by gender.

It was assumed that current smokers can cease smoking from age 9 to 100 (the maximum age in MISCAN-Lung) onwards. The probability of successful smoking cessation is assumed to increase with age until age 85, after which the probability of smoking cessation is assumed to be similar to that at age 85.[1,11,12] The smoking cessation probabilities were estimated in R package by age, gender and cohort, using logistic functions.[9] The smoking cessation probabilities were estimated simultaneously with the mortality probabilities by smoking behavior (detailed in the "Mortality by smoking behavior" section of this Supplement).

Table S2 shows the observed current-, former- and never-smoking prevalence in 2000 compared to the estimated prevalence for each cohort, by gender. Overall, the estimated probabilities reproduce the current-, former- and never-smoking prevalences in 2000.

Cigarettes Smoked per Day

Data on the average number of cigarettes smoked per day (CPD) were obtained from the CCHS, the NPHS, Canada Health Survey, General Social Surveys (years 1985 and 1991) and the Smoking Habits of Canadians Surveys.[3-8] Smokers were divided into five smoking-intensity quintiles, ranging from the lightest to heaviest smokers by the reported average number of CPD at each age, similar to Anderson et al.[2] Age-specific values for the average number of CPD per quintile for ages ≥30 were calculated for ages 30, 40, 50, 60 and 70, depending on the availability of data for each cohort. To account for underreporting of the number of cigarettes per day, due to digit preference and the social undesirability of smoking, the observed values were increased by 7.5%, as this provided the best fit to the observed overall mortality and lung cancer incidence in Ontario (see the following sections "Mortality by smoking behavior" and "Lung cancer incidence in Ontario)".[13,14] Linear interpolation was used to fit the average number of CPD by age and quintile between observed CPD values. CPD values were extrapolated by assuming the average CPD value in each quintile decreases by 1% yearly, to reflect the reduction in CPD smoked as individuals age beyond the age of 45-50.[15] Figures S1 and S2 illustrate the variation in average CPD across the five quintiles in the 1940-1949 cohort, for men and women respectively.

In MISCAN-Lung, upon smoking initiation, an individual is randomly assigned to a quintile (with equal probabilities for each quintile) in which the individual will remain until smoking cessation or death. Similar to Anderson et al., smoking behavior was divided into a period of smoking uptake (ages under age 30) and smoking maintenance after the age of 30.[2] The number of CPD for persons younger than 30 are modeled by applying the uptake formulas described by Anderson et al., as implemented in the U.S. Smoking History Generator, to the average number of cigarettes per day at age 30 for that person's smoking quintile.[2,16]

$$UptakeMale(currentsmokingduration, calenderyear, currentage)$$
$$= -38.578 + 3.342 * \sqrt{currentsmokingduration} - 0.00168$$
$$* \max(79, calenderyear - 1900)^2 - 17.538 * \sqrt{currentage} + 44.967 * \ln(currentage)$$
$$UptakeFemale(currentsmokingduration, calenderyear, currentage)$$
$$= -56.751 + 0.700 * currentsmokingduration - 0.00163$$
$$* \max(79, calenderyear - 1900)^2 - 3.473 * currentage + 32.8 * \sqrt{currentage}$$

The uptake formulas are scaled so that the number of CPD the person smokes at age 30 matches that of the average number of CPD in the quintile the person belongs to, regardless of the age of initiation.

Mortality by smoking behavior

Ontario specific all-cause mortality life tables by birth-year, age and gender, were obtained from the CHMD.[10] To account for competing risks, these life tables were corrected for lung cancer mortality through subtracting the probability of dying from lung cancer from the probability of dying from all causes for each age, as described by Rosenberg et al.[17]

Information on lung cancer mortality by five-year age-groups and gender for years 1990-2009 was obtained from the Ontario Cancer Registry through an electronic copy located on a secure server at the Institute for Clinical Evaluative Sciences.[18] For each available year, the probability of dying from lung cancer was calculated by age-group and gender.

The registry data indicated that dying from lung cancer rarely occurs before ages 20-24. Therefore, we assumed the probability of dying from lung cancer to be zero for ages 0-19 for all cohorts. For cohorts with missing data for age-groups before 1990, we assumed that the probability of dying from lung cancer in those age-groups in the years before 1990 was similar to that of their respective age-groups in 1990. Similarly, the probabilities of dying from lung cancer for age-groups in the years after 2009 were assumed to be similar to those of the respective age-groups in 2009.

The life tables corrected for lung cancer mortality were then further corrected for smoking behavior, for each birth-year and gender. First, it was assumed that smoking behavior influences non-lung cancer mortality from age 40 onwards, similar to Rosenberg et al.[17] Thus, before age 40, the non-lung cancer mortality probabilities for never -and ever-smokers with the same birth-year are assumed to be similar:

$$P(nonlungcancermortality_{neversmoker}, currentage, gender, birthyear)$$
$$= P\left(nonlungcancermortality_{eversmoker,CPDcategory}, currentage, gender, birthyear\right)$$
$$= P\left(nonlungcancermortality_{overallpopulation}, currentage, gender, birthyear\right)$$

From age 40 onward, the non-lung cancer mortality probabilities for never-smokers were assumed to be lower than those of the overall population with the same birth-year, as the overall population includes ever-smokers who have higher non-lung cancer mortality probabilities compared with never-smokers.[17,19] Therefore, after age 40, the non-lung cancer mortality probabilities of never-smokers were assumed to be similar to the non-lung cancer mortality probabilities of the overall population with the same birth-year, corrected for never-smoking:

$$P(nonlungcancermortality_{neversmoker}, currentage, gender, birthyear) = P\left(nonlungcancermortality_{overallpopulation}, currentage, gender, birthyear\right) * Correctionfactor_{neversmoker}$$

However, as ever-smokers have higher non-lung cancer mortality probabilities compared with never-smokers, the proportion of ever-smokers in the overall population with the same birth-year is expected to decrease at higher ages. Thus, at higher ages, never-smokers will represent a higher proportion of the overall population. As a result, the non-lung cancer mortality probabilities for never-smokers will converge to those of the overall population of that birth-year at higher ages. This convergence is assumed to start from age 70 onward and therefore, the non-lung cancer mortality probabilities of never-smokers from that age onward were assumed to be:

$$P(nonlungcancermortality_{neversmoker}, currentage, gender, birthyear) = P\left(nonlungcancermortality_{overallpopulation}, currentage, gender, birthyear\right) * \left(Correctionfactor_{neversmoker} + \left(currentage - 69 * \left(\frac{(1 - Correctionfactor_{neversmoker})}{30}\right)\right)\right)$$

As indicated previously, ever-smokers have higher non-lung cancer mortality probabilities compared with never-smokers.[17,19] However, the non-lung cancer mortality probabilities for ever-smokers are also influenced by the average number of CPD smoked by a person[17,19] Therefore, it was assumed that non-lung cancer mortality probabilities increased with higher average numbers of CPD. Four categories of average numbers of CPD were defined, similar to Thun et al.: < 10 CPD, 10-19 CPD, 20-39 CPD and ≥ 40 CPD.[19] Furthermore, longer durations of smoking have been indicated to increase non-lung cancer mortality probabilities.[19] Therefore, the increase in non-lung cancer mortality probabilities for current-smokers compared with never-smokers was also assumed to increase with age (as a substitute for smoking duration). In addition, this increase was assumed to differ by smoking quintile, to reflect differences in the average CPD over longer periods of time. Thus, the non-lung cancer mortality probabilities for current-smokers from age 40 onward were assumed to be:

$$P\left(nonlungcancermortality_{currentsmoker_{CPDcategory,CPDquintile}}, currentage, gender, birthyear\right) = P(nonlungcancermortality_{neversmoker}, currentage, gender, birthyear) + smokingmortalityincrease(age)_{CPDcategory,CPDquintile}$$

Where

$$smokingmortalityincrease(age)_{CPDcategory,CPDquintile} = smokingmortalityincrease_{CPDcategory,CPDquintile} * \left(Smokingagecorrection_{CPDquintile} + ((Currentage - 40) * \left(\frac{\left(1 - Smokingagecorrection_{CPDquintile}\right)}{59} \right)) \right)$$

Previous research indicates that the age of smoking cessation and years since smoking cessation influence the excess risk of mortality due to past smoking behavior.[17,19] Overall, the excess risk of mortality decreases for a younger age of smoking cessation and a higher number of years since smoking cessation.[17,19] Therefore, the excess risk of non-lung cancer mortality was assumed to decrease over time for former smokers, similarly to the formula described by Rosenberg et al.[17]:

$$P\left(nonlungcancermortality_{formersmoker_{CPDcategory}}, currentage, gender, birthyear\right) = P(nonlungcancermortality_{neversmoker}, currentage, gender, birthyear) + ((mortalityincrease(age)_{CPDcategory,CPDquintile}) * \exp\left((-0.1711 + (0.00102 * averageCPDoverlifetime) + (0.00171 * QuitAge)) * YearsQuit^{1.08}\right))$$

The non-lung cancer mortality probabilities for never- and ever-smokers were estimated simultaneously with the smoking cessation probabilities to match the life tables previously corrected for lung cancer mortality for each birth-year and gender, using R package.[9] Figures S3 and S4 show the estimated all-cause mortality probabilities of the overall population for men and women born in 1955 compared to those of the CHMD life tables as examples. Overall, the estimated all-cause mortality probabilities match those of the CHMD lifetables. Figures S5 and S6 show the estimated cumulative mortality probabilities of dying from causes other than lung cancer (up to age 85) for never-smokers and current smokers (by smoking quintile) for men and women born in 1955 as examples.

Table S1: Observed and estimated ever-smoking prevalence at age 30 for the investigated cohorts

Men		
Cohort	**Observed ever-smoking prevalence at age 30**	**Estimated ever-smoking prevalence at age 30**
1940-1949	61.92%	61.84%
1950-1959	57.78%	57.75%
1960-1969	50.78%	50.85%
Women		
Cohort	**Observed ever-smoking prevalence at age 30**	**Estimated ever-smoking prevalence at age 30**
1940-1949	44.91%	44.90%
1950-1959	46.85%	46.82%
1960-1969	41.01%	41.03%

Table S2: Observed and estimated current-, former- and never-smoking prevalences in 2000

Men						
Cohort	**Observed current-smoking prevalence in 2000**	**Estimated current-smoking prevalence in 2000**	**Observed former-smoking prevalence in 2000**	**Estimated former-smoking prevalence in 2000**	**Observed never-smoking prevalence in 2000**	**Estimated never-smoking prevalence in 2000**
1940-1949	19.13%	19.62%	42.50%	41.55%	38.37%	38.83%
1950-1959	24.59%	25.02%	32.56%	32.61%	42.85%	42.37%
1960-1969	28.67%	30.33%	20.71%	20.50%	50.62%	49.17%
Women						
Cohort	**Observed current-smoking prevalence in 2000**	**Estimated current-smoking prevalence in 2000**	**Observed former-smoking prevalence in 2000**	**Estimated former-smoking prevalence in 2000**	**Observed never-smoking prevalence in 2000**	**Estimated never-smoking prevalence in 2000**
1940-1949	14,62%	15.18%	28.89%	29.27%	56.49%	55.55%
1950-1959	20.85%	20.65%	24.98%	26.10%	54.17%	53.26%
1960-1969	21.37%	24.14%	18.07%	16.89%	60.56%	58.97%

Figure S1: Average number of cigarettes per day for ages over 30, by smoking quintile for men born between 1940-1949

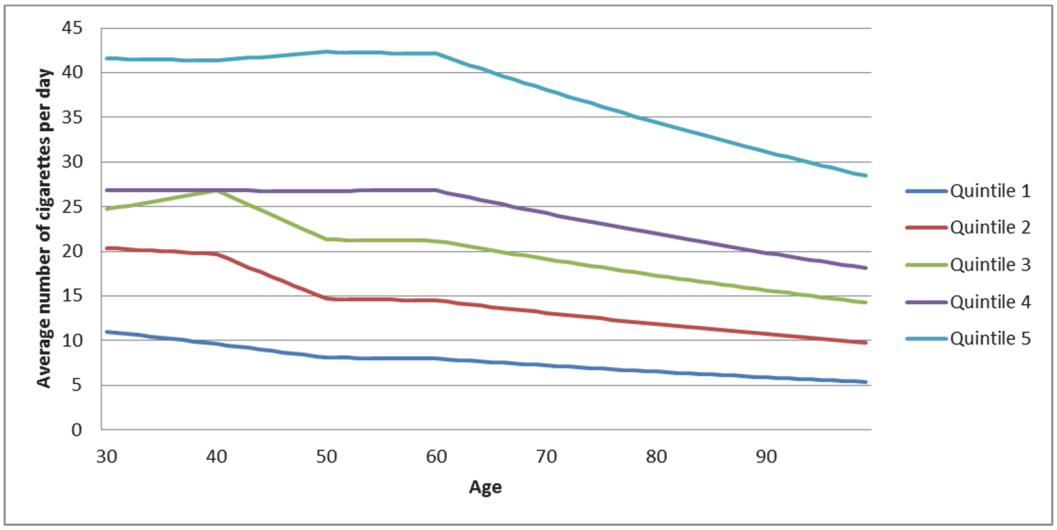


Figure S2: Average number of cigarettes per day for ages over 30, by smoking quintile for women born between 1940-1949

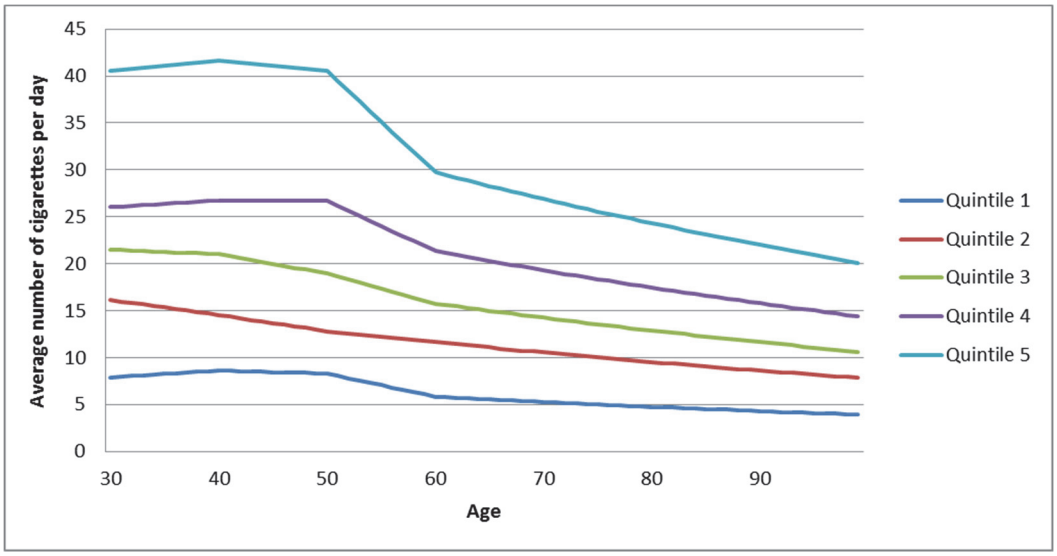


Figure S3: Annual probability of dying from all causes for men born in 1955

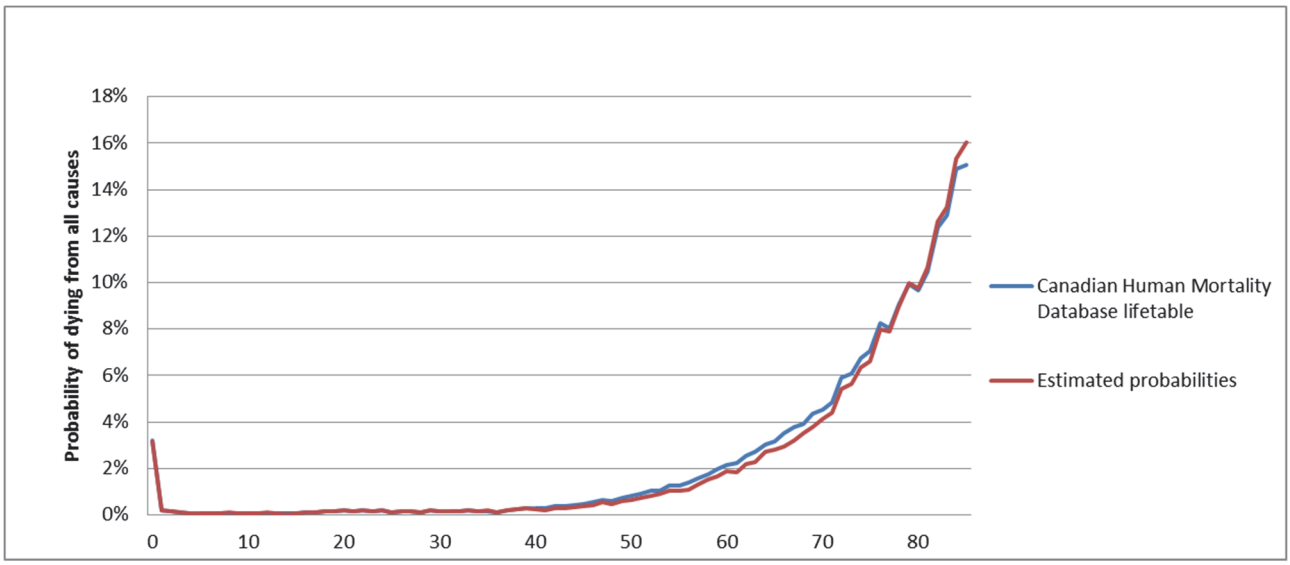


Figure S4: Annual probability of dying from all causes for women born in 1955

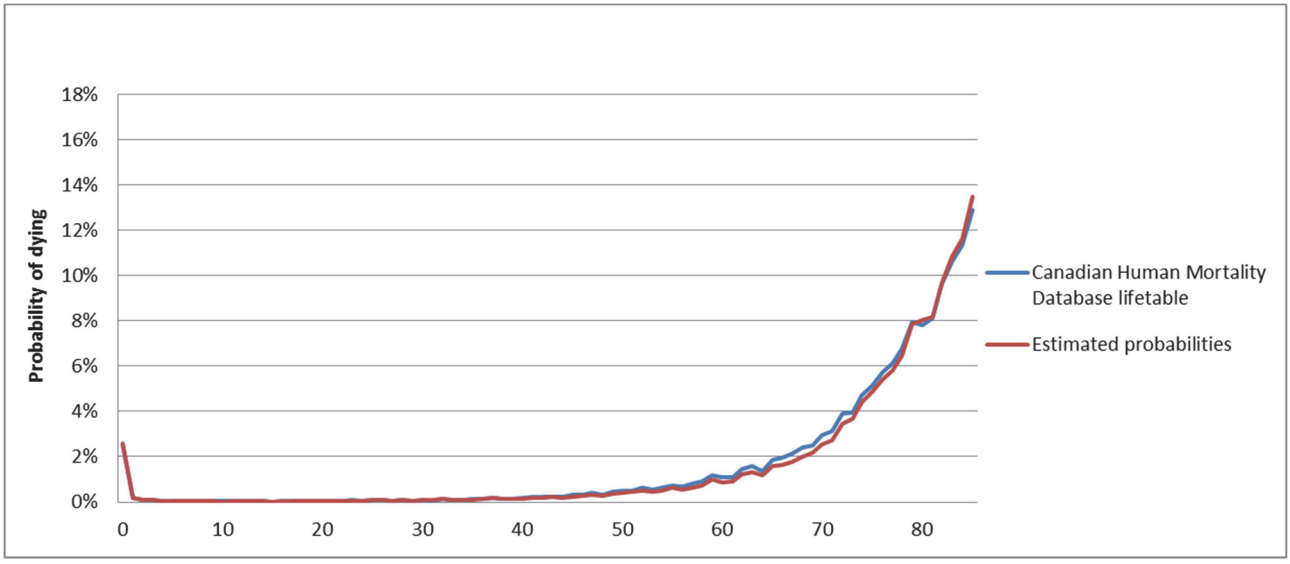


Figure S5: Cumulative probability of dying from causes other than lung cancer for never-smokers and current smokers (by smoking quintile) for men born in 1955

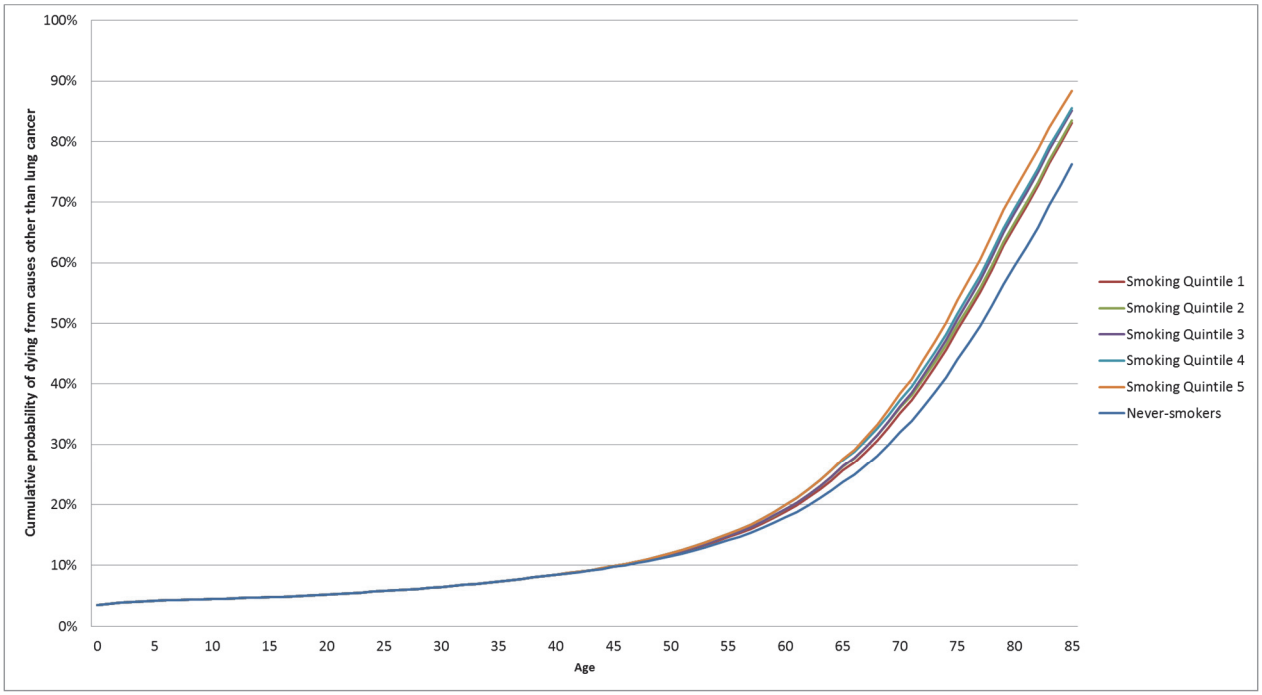
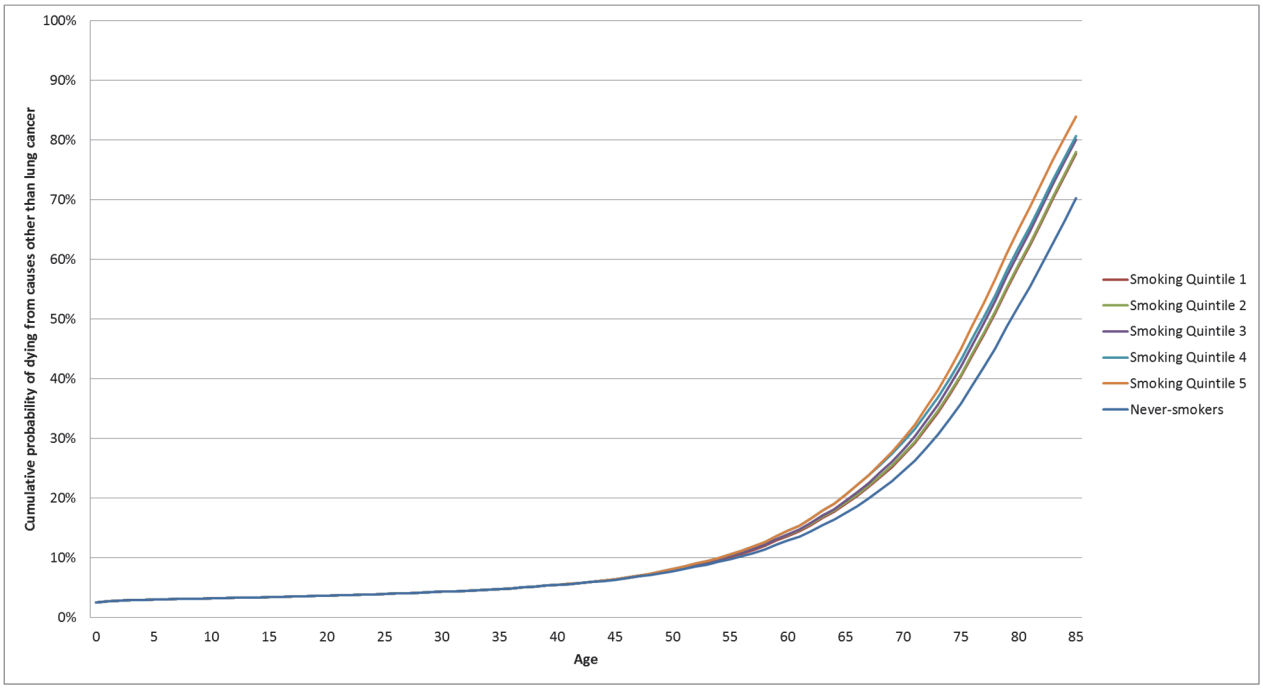


Figure S6: Cumulative probability of dying from causes other than lung cancer for never-smokers and current smokers (by smoking quintile) for women born in 1955



Lung cancer incidence in Ontario

The following sections detail the data and methods used to investigate whether the MISCAN-Lung model can reproduce the observed lung cancer incidence in Ontario, which allows the extrapolation of future lung cancer incidence and investigation of the effects of lung cancer screening.

Lung cancer incidence

Data on the incidence of lung cancer (by gender, stage and histology) in the province of Ontario, Canada, was obtained from the Ontario Cancer Registry through an electronic copy located on a secure server at the Institute for Clinical Evaluative Sciences (Toronto, Ontario), for years 2007-2009, for ages 40-74.[18] Only cancers with a known histology were taken into account, as cancers with an unknown or unspecified histological type could potentially be misdiagnosed metastases from other cancer sites (i.e. 7.74% of cases were excluded).

Lung cancer stages

Four tumor stages are distinguished in the Ontario Cancer Registry data, based on the American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 6th edition: stages I, II, III and IV.[20] In the MISCAN-Lung model, six AJCC tumor stages are distinguished, namely: IA, IB, II, IIIA, IIIB and IV. To allow comparisons between the observed lung cancer incidence in the Ontario registry data and the MISCAN-Lung estimates, model outputs for stages IA and IB were combined into stage I and model outputs for stages IIIA and IIIB were combined into stage III.

Lung cancer histology

MISCAN-Lung incorporates four histological types, based on the International Classification of Diseases for Oncology Third Edition (ICD-O-3) codes, namely: adenocarcinoma/large cell carcinoma/bronchioloalveolar carcinoma (AD), squamous cell carcinoma (SQ), other non-small cell carcinoma (OTH) and small cell carcinoma (SM).[21,22] The data from the Ontario Cancer Registry were matched to the definitions used by MISCAN-Lung, using the ICD-O-3 codes.

Five-year survival data by histology, stage and gender were available from the Ontario Cancer Registry and compared to survival data (by stage and histology) obtained from the U.S. Surveillance, Epidemiology, and End Results (SEER) program.[23] The overall survival rates in Ontario and SEER were similar (data not shown). However, due to known differences in survival between stages IA and IB, the survival data from SEER were used, because of availability of more detailed data on survival by stage. [24]

Comparison of MISCAN-Lung estimates to observed data

Figures S7 and S8 compare the proportions of histological types observed in the Ontario Cancer Registry data for ages 40-74, by gender, to the proportions estimated by MISCAN-Lung. Overall, MISCAN-Lung reproduces the observed proportions of histological types for both genders.

Figures S9 and S10 compare the proportions of the clinical stages observed for ages 40-74 for lung cancers with a known stage in the Ontario Cancer Registry data, by gender, to the proportions estimated by MISCAN-Lung. Overall, MISCAN-Lung reproduces the observed proportions of clinical stages for each gender, though it somewhat overestimates the proportion of stage I cancers and underestimates the proportion of stage IV cancers for men.

Figures S11 and S12 compare the incidence per 100,000 persons by age-group and gender, observed in the Ontario registry data to the incidence estimated by MISCAN-Lung. Overall, MISCAN-Lung reproduces the observed incidence well for men, though it somewhat underestimates the incidence at ages 70-74. MISCAN-lung somewhat underestimates the incidence for women across ages 45-69, while it somewhat overestimates the incidence for ages 70-74.

Overall, the MISCAN-Lung model adequately reproduced the overall lung cancer incidence in Ontario for 2007-2009, allowing extrapolation of future lung cancer incidence and investigation of the effects of lung cancer screening.

Figure S7: Lung cancer histology distributions estimated for men ages 40-74 by the MISCAN-Lung model compared to the observed lung cancer histology distributions in Ontario in 2007-2009

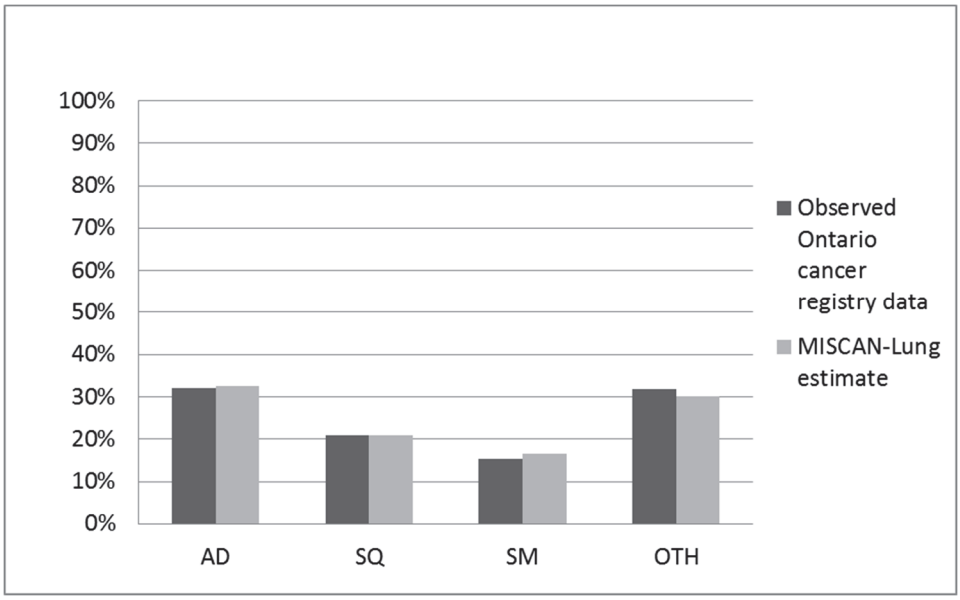
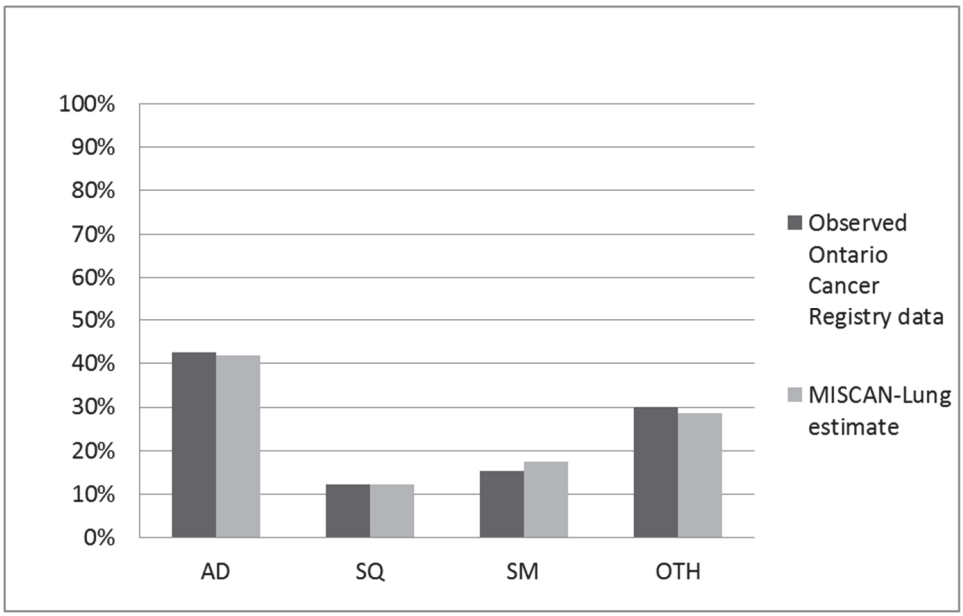


Figure S8: Lung cancer histology distributions estimated for women ages 40-74 by the MISCAN-Lung model compared to the observed lung cancer histology distributions in Ontario in 2007-2009



Abbreviations: adenocarcinoma/large cell carcinoma/ bronchioloalveolar carcinoma (AD), squamous cell carcinoma (SQ), small cell carcinoma (SM), other non–small cell carcinoma (OTH).

Figure S9: Lung cancer stage distributions estimated for men ages 40-74 by the MISCAN-Lung model compared to the observed lung cancer stage distributions in Ontario in 2007-2009

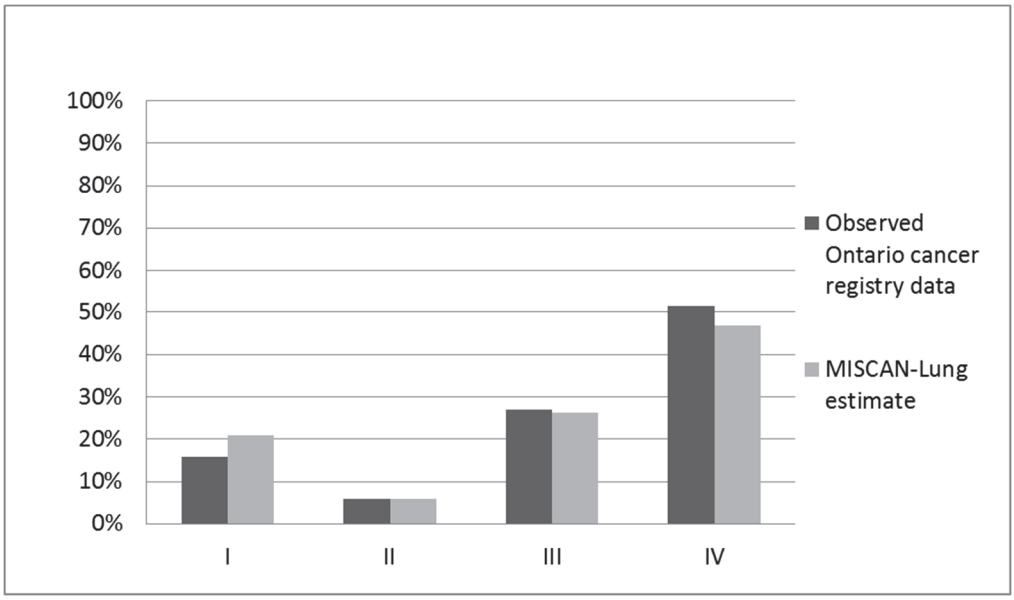


Figure S10: Lung cancer stage distributions estimated for women ages 40-74 by the MISCAN-Lung model compared to the observed lung cancer stage distributions in Ontario in 2007-2009

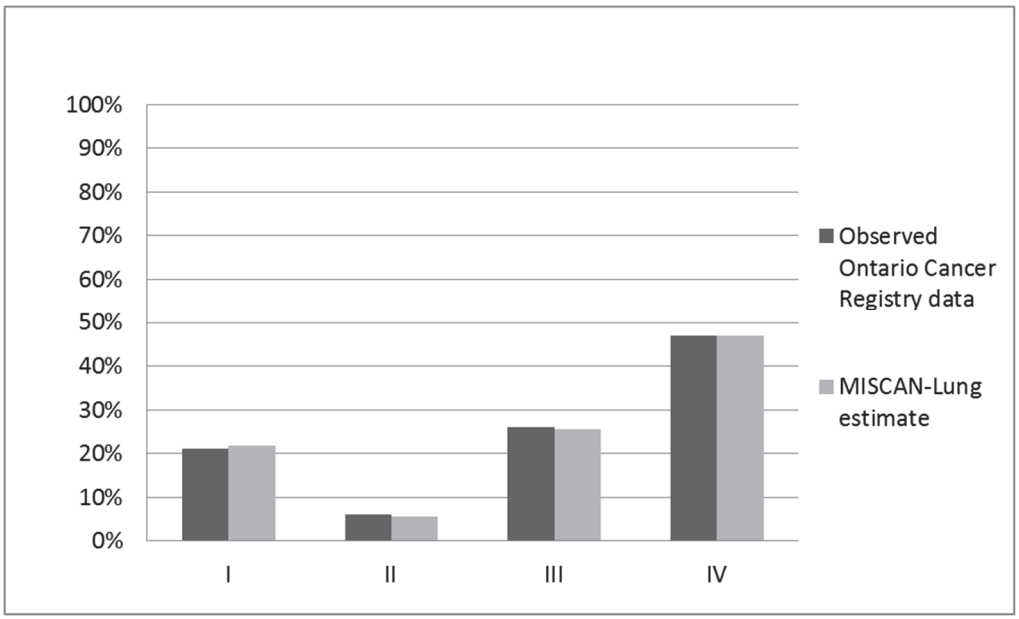


Figure S11: Lung cancer incidence per 100,000 estimated for men ages 40-74 by the MISCAN-Lung model compared to the observed lung cancer incidence in Ontario in 2007-2009

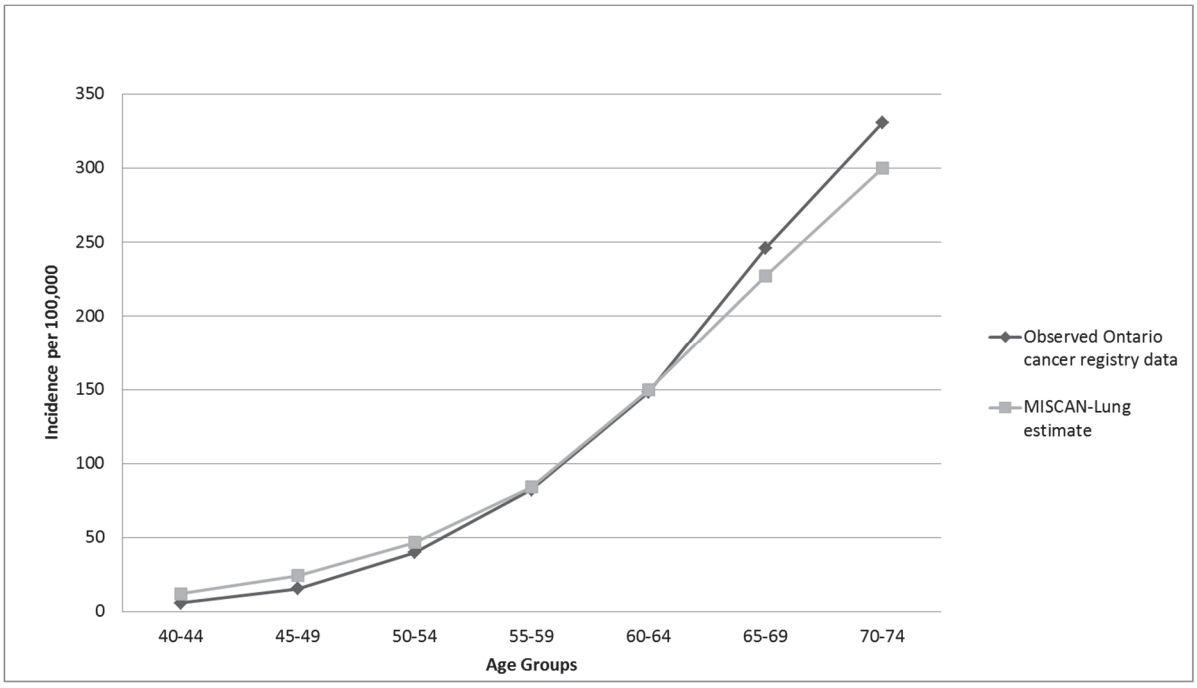
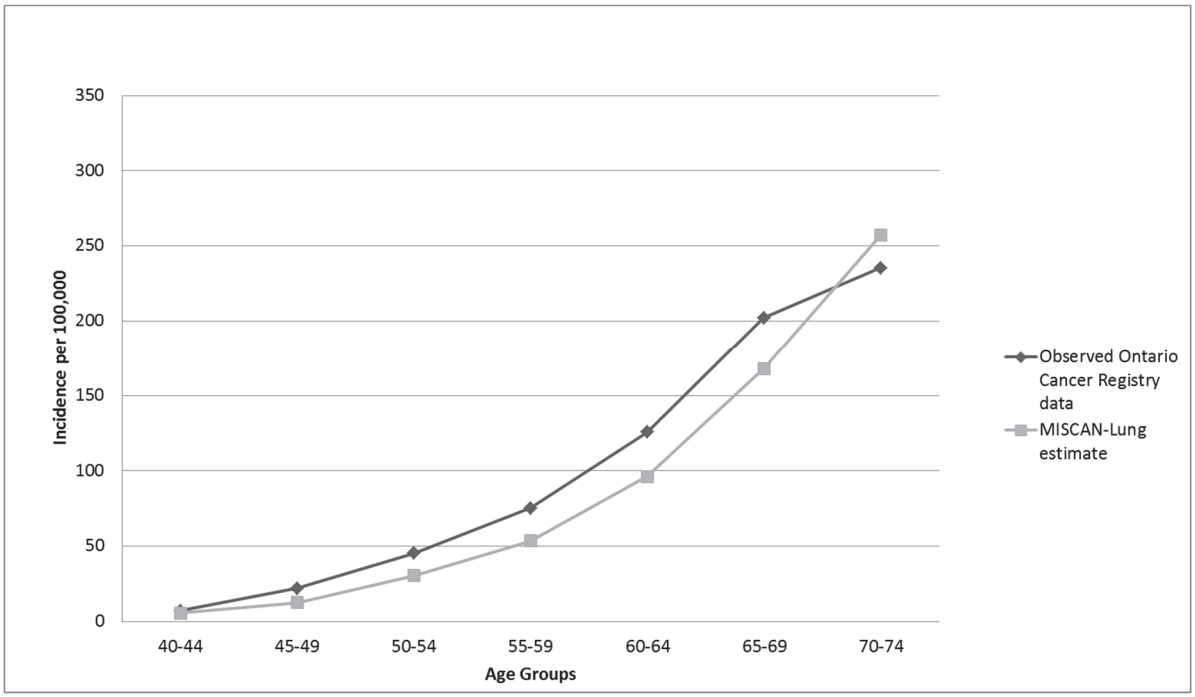


Figure S12: Lung cancer incidence per 100,000 estimated for women ages 40-74 by the MISCAN-Lung model compared to the observed lung cancer incidence in Ontario in 2007-2009



Screening related follow-up procedures

The MISCAN-Lung model incorporates the following screening related follow-up procedures: false-positive results, follow-up CT examinations, biopsies and surgeries not related to lung cancer as a result of screening. The following sections detail the methods and assumptions used to incorporate these screening related follow-up procedures in the MISCAN-Lung model.

False-positive results

Each individual in the model who attends a round of screening, in whom lung cancer is not detected by screening in that round, has a chance to receive a false-positive test result. The probability of receiving a false-positive test result was based on individual-level data from the computed tomography (CT) arm of the National Lung Screening Trial (NLST). For each screening round in the CT arm of the NLST, the number of positive screening test results and the number of persons with screen-detected lung cancer were determined. The number of false-positive test results in each round of the CT arm of the NLST was defined as the number of positive screening test results minus the number of persons with screen-detected lung cancer in that round. The probability to receive a false-positive screening test result in each round of the CT arm of the NLST was then defined as follows:

$$P(False\ positive\ test\ result\ in\ round\ X | Lung\ cancer\ not\ detected\ by\ screening\ in\ round\ X)$$
$$= \frac{(Number\ of\ positive\ test\ results\ in\ round\ X - Number\ of\ persons\ with\ screen\ detected\ lung\ cancer\ in\ round\ X)}{(Number\ of\ screens\ in\ round\ X - Number\ of\ persons\ with\ screen\ detected\ lung\ cancer\ in\ round\ X)}$$

The probability that an individual in the model who attends screening, in whom lung cancer is not detected by screening in that round, receives a false-positive test result is assumed to be equal to the probability observed in the corresponding round of the CT arm of the NLST. For every screening round attended after the third, this probability is assumed to be equal to that of the third screening round of the NLST. The corresponding probabilities for each round are noted in Table S3.

Follow-up CT-examinations

Each individual in the model in whom lung cancer is detected (through screening or clinical presentation of symptoms) receives a CT examination (in addition to a biopsy, as detailed in the section on biopsies) to confirm the diagnosis of cancer, which is incorporated into the attributable costs. In addition, each individual who attends a screening round, in whom lung cancer is not detected by screening in that round, has a chance to receive a number of follow-up CT examinations as a result of screening. The probability in the model to receive at least one follow-up CT examination and the average number of follow-up CT examinations received by persons without screen-detected lung cancer were based on individual-level data from the CT arm of the NLST. Table S4 lists the follow-up procedures in the NLST dataset classified as follow-up CT examinations for this investigation. For each screening round in the CT arm of the NLST, the number of persons without a screen-detected lung cancer in that round who received at least one screen-related follow-up CT examination was

determined. The probability that an individual without screen-detected lung cancer in that round receives at least one screen-related follow-up CT examination was then defined as follows for each round of the CT arm of the NLST:

$$P(At\ least\ one\ followup\ CT\ examination\ in\ round\ X | Lung\ cancer\ not\ detected\ by\ screening\ in\ round\ X) = \frac{(Number\ of\ persons\ with\ at\ least\ one\ followup\ CT\ examination\ without\ screen\ detected\ lung\ cancer\ in\ round\ X)}{(Number\ of\ screens\ in\ round\ X - Number\ of\ persons\ with\ screen\ detected\ lung\ cancer\ in\ round\ X)}$$

The average number of screen-related follow-up CT examinations per person, in whom lung cancer was not detected by screening, who received at least one follow-up CT examination was then defined as follows for each round of the CT arm of the NLST:

$$\left(\begin{array}{c} Average\ number\ of\ followup\ CT\ examinations \\ per\ person\ who\ received\ at\ least\ one \\ CT\ followup\ examination \\ in\ round\ X \end{array}\right. \left| Lung\ cancer\ not\ detected\ by\ screening\ in\ round\ X \right)$$

$$= \frac{(Number\ of\ followup\ CT\ examinations\ for\ persons\ without\ screen\ detected\ lung\ cancer\ in\ round\ X)}{(Number\ of\ persons\ with\ at\ least\ one\ followup\ CT\ examination\ without\ screen\ detected\ lung\ cancer\ in\ round\ X)}$$

The probability that an individual in the model who attends screening, in whom lung cancer is not detected by screening in that round, receives at least one follow-up CT examination is assumed to be equal to the probability observed in the corresponding round of the CT arm of the NLST. The average number of follow-up CT examinations per person, in whom lung cancer is not detected by screening, who receives at least one CT examination is assumed to be equal to the average number of CT examinations per person examined in the corresponding round of the CT arm of the NLST. For every screening round attended after the third, the probability to receive a follow-up CT examination and the average number of follow-up CT examinations were assumed to be equal to those of the third screening round of NLST. The corresponding probabilities and average number of follow-up CT examinations per person examined are described in Table S3.

Biopsies/bronchoscopies

Every person in the model in whom lung cancer is detected (through screening or clinical presentation of symptoms) receives a biopsy/bronchoscopy (in addition to a CT examination, as detailed in the section on CT examinations) to confirm the diagnosis of cancer, which is incorporated into the attributable costs. However, individuals who attend screening (in whom lung cancer is not detected by screening) also have a chance to receive a biopsy/bronchoscopy as a result of screening. The probability of receiving a biopsy/bronchoscopy was based on individual-level data from the CT arm of the NLST. Table S5 lists the follow-up procedures in the NLST dataset classified as biopsies for this investigation. For each screening round in the CT arm of the NLST, the number of persons, in whom lung cancer was not detected by screening, that received at least one biopsy related to the screening round was determined. Biopsies/bronchoscopies performed on days on which a surgical procedure was performed and cytology procedures performed on the same day as another

biopsy/bronchoscopy procedure were not counted as separate events. The probability for a person, in whom lung cancer was not detected by screening, to receive a screen-related biopsy/bronchoscopy was then defined as follows for each round of the CT arm of the NLST:

$$P(Screenrelated\ biopsy/bronchoscopy\ in\ round\ X|Lung\ cancer\ not\ detected\ by\ screening\ in\ round\ X) = \frac{(Number\ of\ persons\ with\ at\ least\ one\ screenrelated\ biopsy/bronchoscopy\ without\ screen\ detected\ lung\ cancer\ in\ round\ X)}{(Number\ of\ screens\ in\ round\ X - Number\ of\ persons\ with\ screen\ detected\ lung\ cancer\ in\ round\ X)}$$

The probability that an individual in the model who attends screening, in whom lung cancer is not detected by screening in that round, receives a biopsy/bronchoscopy related to screening is assumed to be equal to the probability observed in the corresponding round of the CT arm of the NLST. For every screening round attended after the third, this probability is assumed to be equal to that of the third screening round of NLST. The corresponding probabilities for each round are noted in Table S3.

Surgeries not related to lung cancer as a result of screening

Individuals in the model who attend screening (in whom lung cancer is not detected by screening) have a chance to receive a surgery not related to lung cancer as a result of the screening**.** The probability of receiving a surgery not related to lung cancer as a result of screening was based on individual-level data from the CT arm of the NLST. Table S7 lists the follow-up procedures in the NLST dataset classified as surgical procedures for this investigation. For each screening round in the CT arm of the NLST, the number of persons who received at least one surgical procedure not related to a diagnosis of lung cancer as a result of screening was determined.

The probability for a person, in whom lung cancer was not detected by screening, to receive a screen-related surgery not related to lung cancer was then defined as follows for each round of the CT arm of the NLST:

$$P(Screen\ related\ surgery\ not\ related\ to\ lung\ cancer\ in\ round\ X|Lung\ cancer\ not\ detected\ by\ screening\ in\ round\ X) = \frac{(Number\ of\ persons\ with\ at\ least\ one\ screen\ related\ surgery\ without\ screen\ detected\ lung\ cancer\ in\ round\ X)}{(Number\ of\ screens\ in\ round\ X - Number\ of\ persons\ with\ screen\ detected\ lung\ cancer\ in\ round\ X)}$$

The probability that an individual in the model who attends screening, in whom lung cancer is not detected by screening in that round, receives a surgical procedure not related to lung cancer is assumed to be equal to the probability observed in the corresponding round of the CT arm of the NLST. For every screening round attended after the third, this probability assumed to be equal to that of the third screening round of NLST. The corresponding probabilities for each round are noted in Table S3.

Table S3: **Probabilities of adverse outcomes due to screening in the MISCAN-Lung model given that lung cancer is not detected by screening**

Screening round	Probability to receive a false-positive result	Probability to receive a follow-up CT examination	Average number of follow-up CT examinations received per person receiving follow-up CT examinations	Probability to receive a biopsy/bronchoscopy	Probability to receive a surgery not related to lung as a result of screening
First attended screening	26.58%	19.27%	1.45	0.76%	0.35%
Second attended screening	27.44%	7.83%	1.32	0.39%	0.20%
Third attended screening and subsequent screens	16.05%	6.37%	1.69	0.39%	0.18%

Table S4: **Procedures in the NLST dataset classified as follow-up CT scans**

Procedures	
CT- Abdomen and pelvis	CT - Chest, limited thin section of nodule
CT - Diagnostic chest	CT - Chest and abdomen
CT - Brain	CT - Abdomen (or liver)
CT - Chest, low dose spiral	CT - Other (specify)
CT - Chest, abdomen, and pelvis	CT - Chest, plus nodule densitometry
CT- Chest limited thin section of entire lung	Radionuclide scan - Fusion PET/CT scan

Table S5: Procedures classified as biopsies/bronchoscopies in the NLST dataset

Procedures	
Biopsy - Endobronchial	Biopsy – Other
Biopsy - Percutaneous adrenal	Cytology - Bronchoscopic
Biopsy - Percutaneous liver	Biopsy - Open Surgical
Biopsy - Percutaneous transthoracic yielding histology	Cytology -Percutaneous transthoracic
Biopsy - Lymph node - other	Biopsy - Transbronchial
Bronchoscopy without biopsy or cytology	Cytology – Other
Biopsy - Lymph node - scalene nodes	Biopsy - Thoracoscopic
Cytology - Sputum	

Table S6: **Procedures classified as non-lung cancer related surgical procedures (if not linked to the diagnosis of lung cancer)**

Procedures	
Mediastinoscopy/Mediastinotomy	Thoracotomy
Lymphadenectomy/lymph node sampling	Resection
Thoracentesis	Thoracoscopy without biopsy
Thoracoscopy	

Figure S13: Cost-effectiveness of lung cancer screening scenarios by screening starting age in the base-case analysis

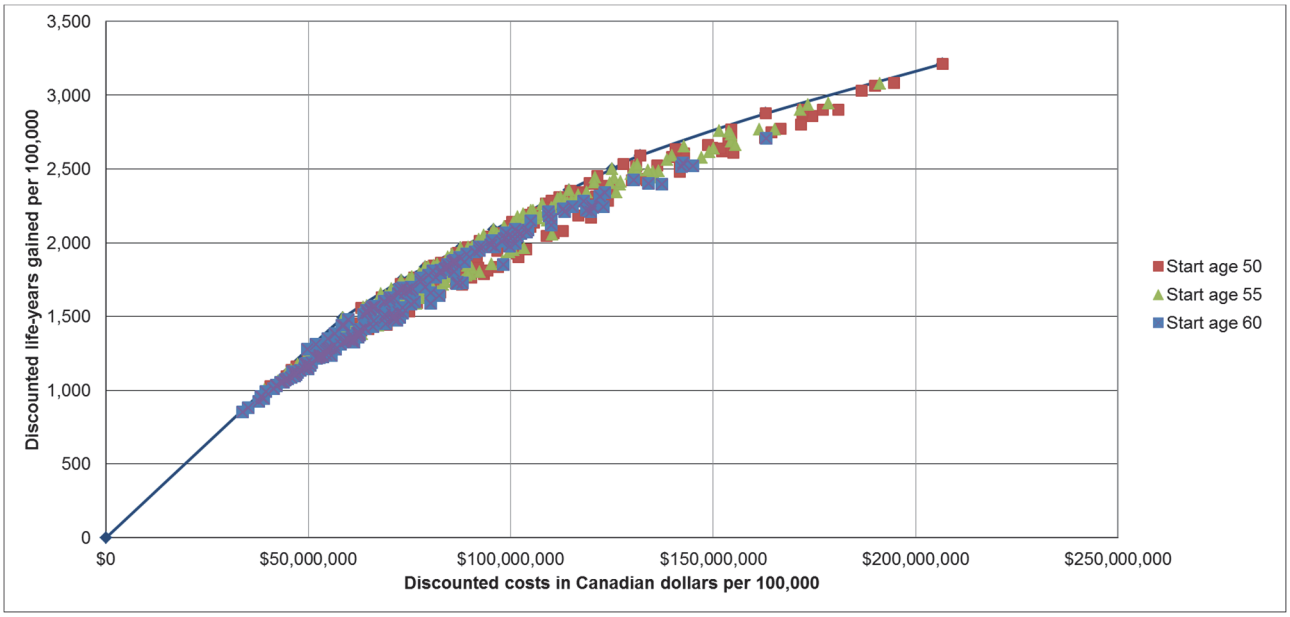


Figure S14: Cost-effectiveness of lung cancer screening scenarios by screening stopping age in the base-case analysis

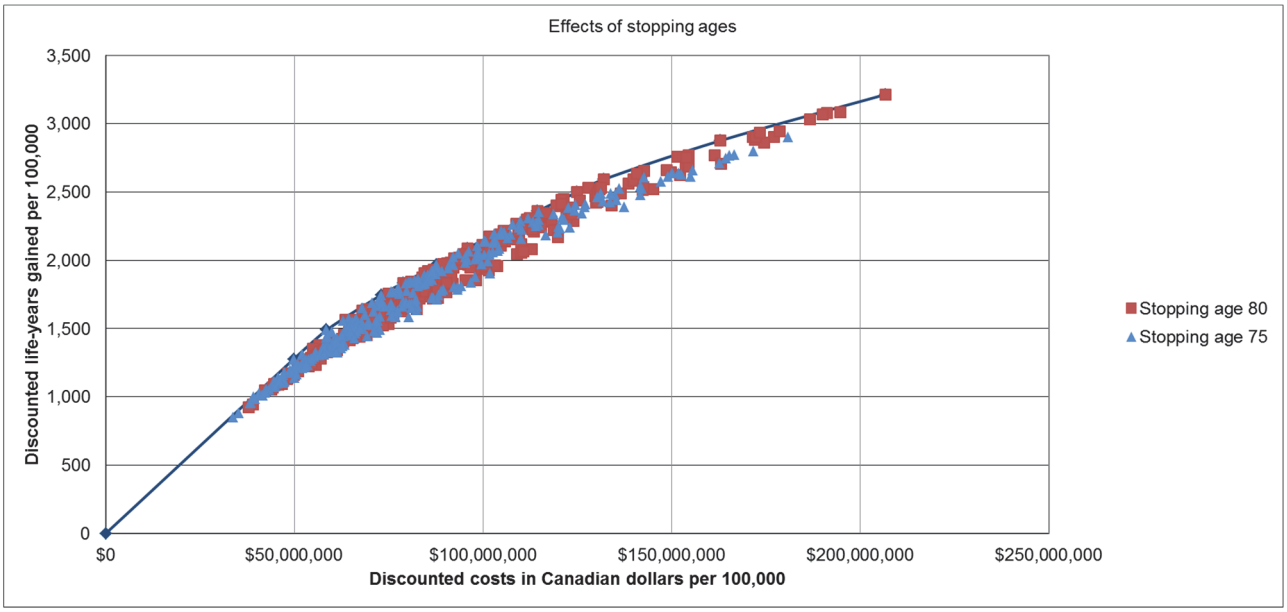


Figure S15: Cost-effectiveness of lung cancer screening scenarios by type of aggregated smoking criteria in the base-case analysis

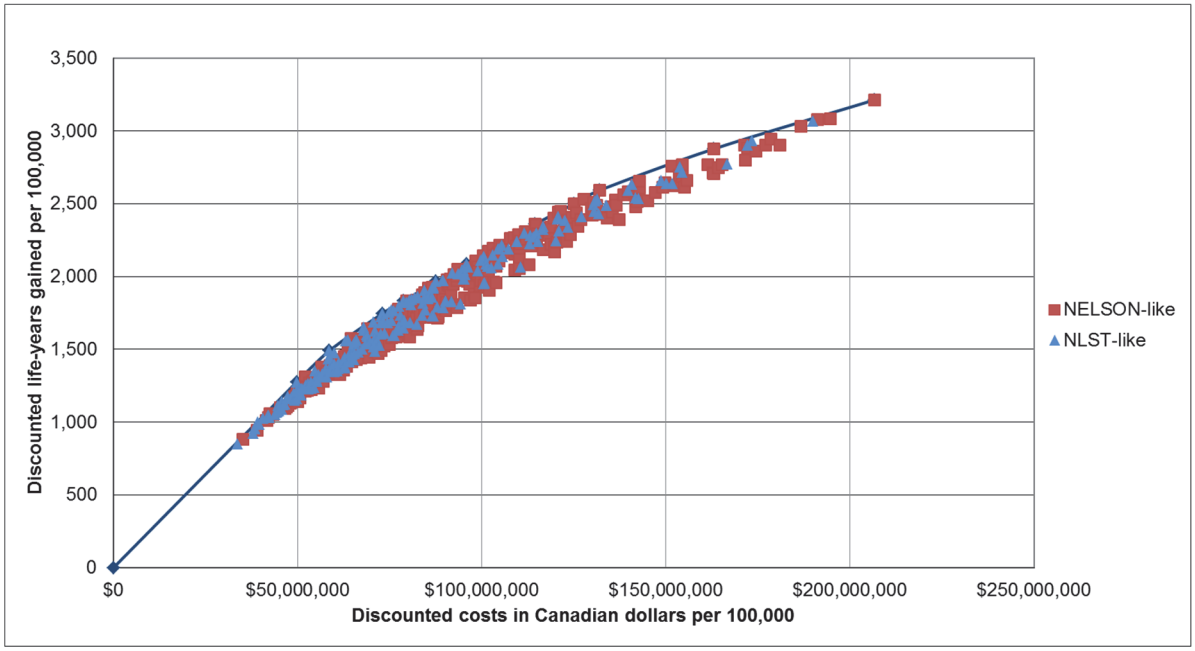
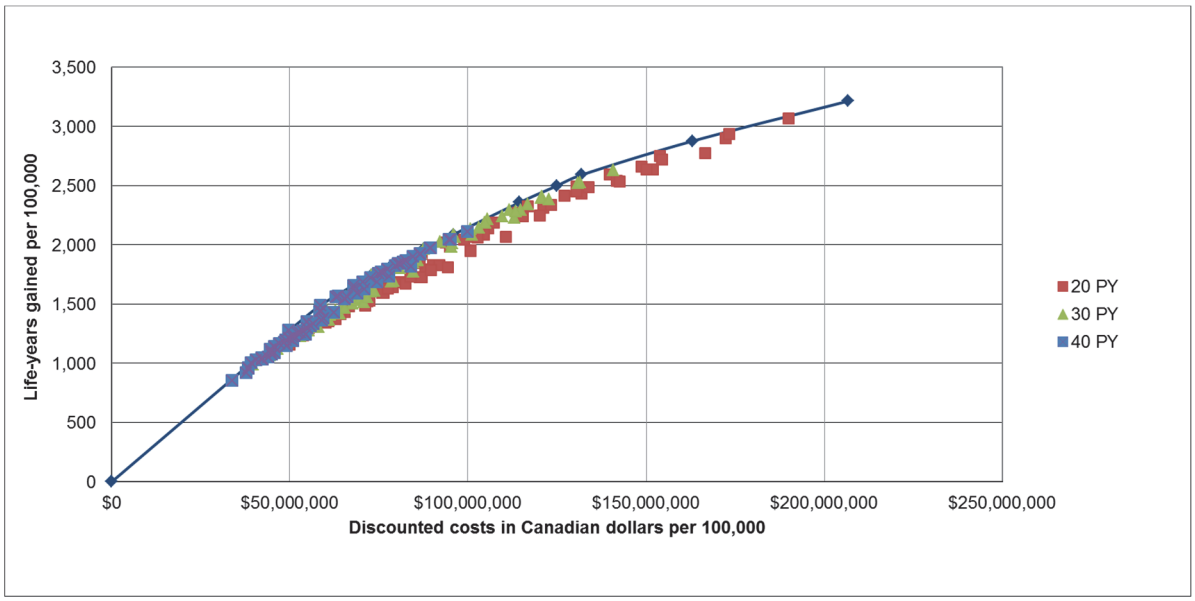
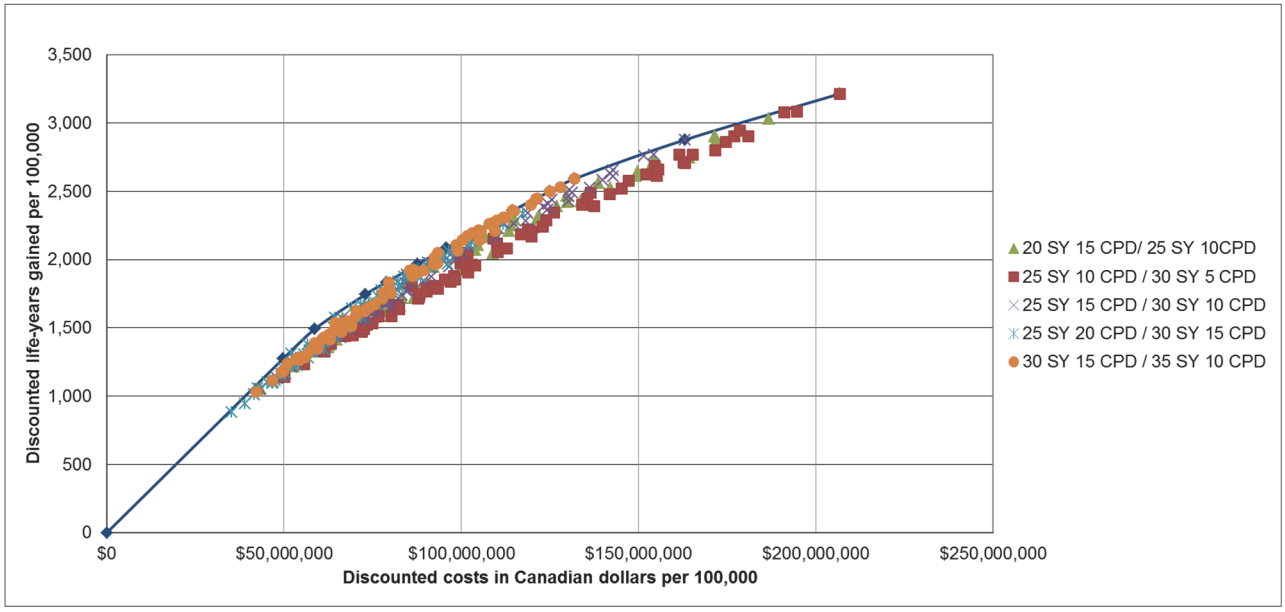


Figure S16: Cost-effectiveness of lung cancer screening scenarios by aggregated smoking criteria (efficient frontier and NLST-like scenarios only) in the base-case analysis



Abbreviations: pack-years (PY)

Figure S17: Cost-effectiveness of lung cancer screening scenarios by aggregated smoking criteria (efficient frontier and NELSON-like scenarios only) in the base-case analysis



Abbreviations: smoking duration in years (SY), cigarettes per day (CPD)

Figure S18: Cost-effectiveness of lung cancer screening scenarios by years since cessation in the base-case analysis

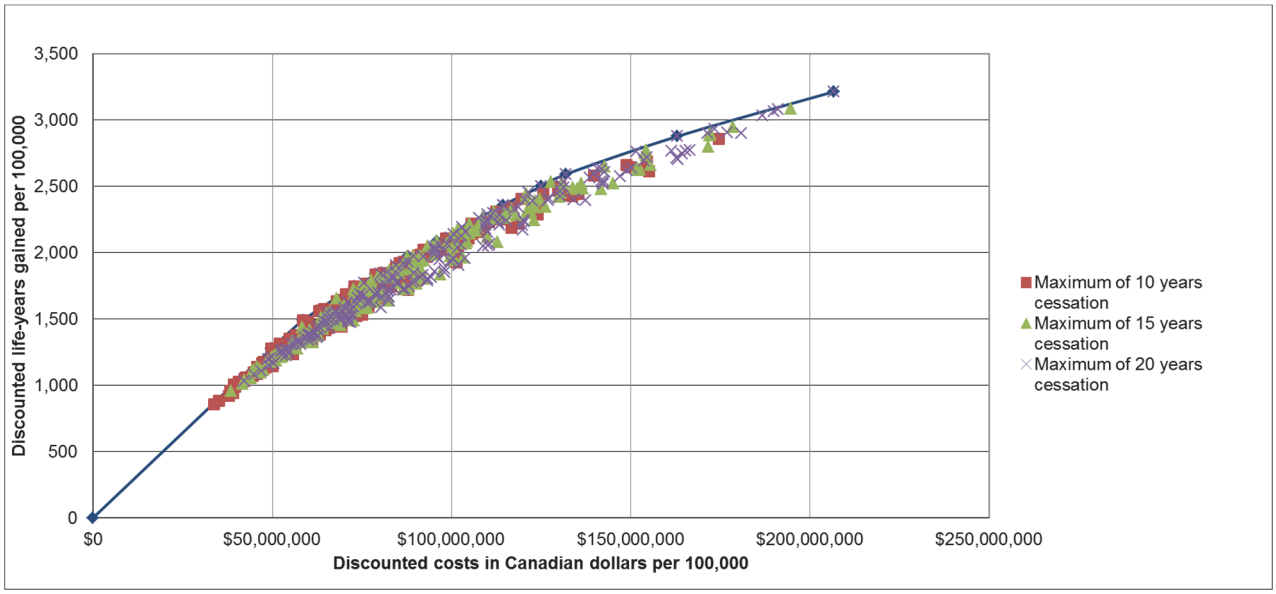
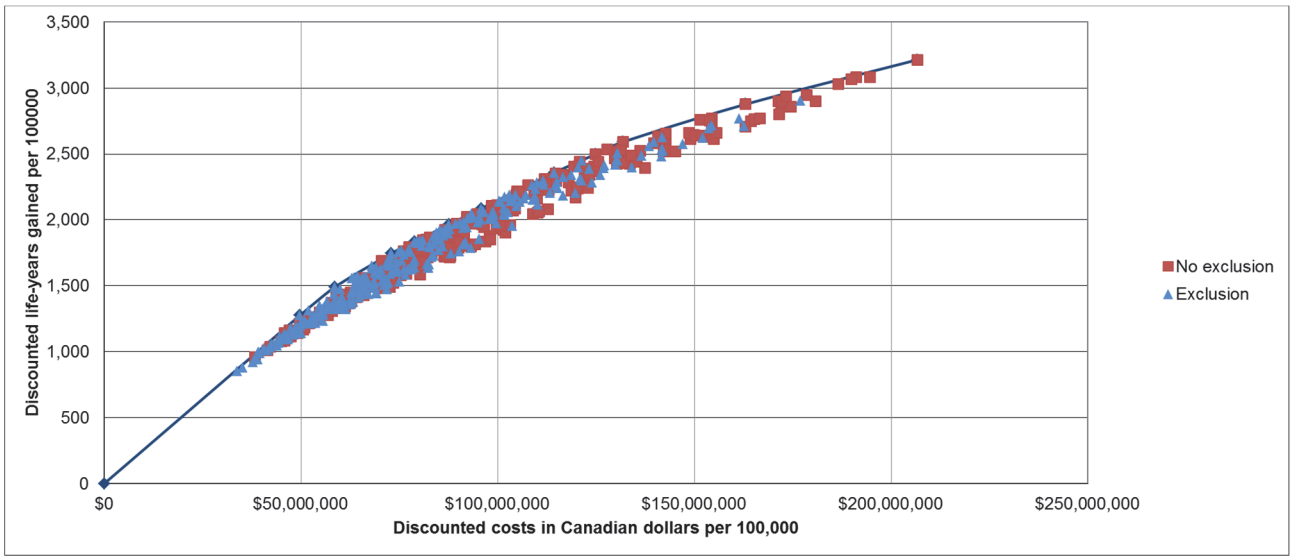
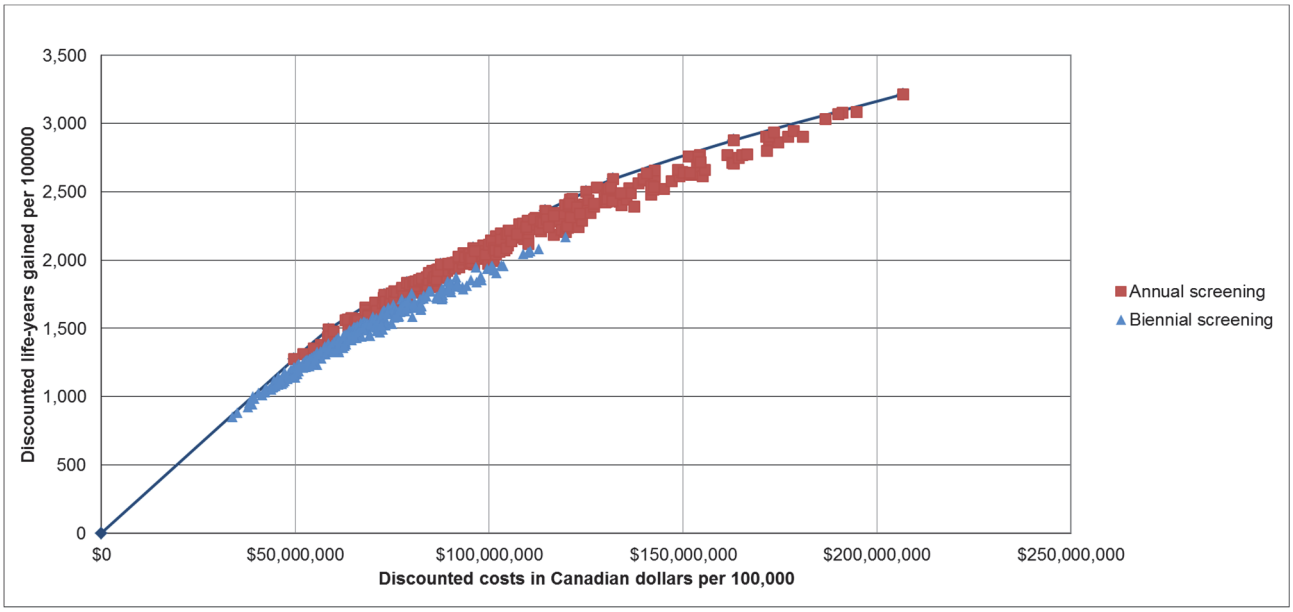


Figure S19: Cost-effectiveness of lung cancer screening scenarios: exclusion of individuals after reaching the maximum number of years since cessation compared with no exclusion in the base-case analysis



Supplemental Figure S20: Cost-effectiveness of lung cancer screening scenarios by intervals between screenings in the base-case analysis



Impact of sensitivity analyses

The following section details the characteristics, costs, life-years gained and cost-effectiveness of the scenarios on the efficient frontier of each sensitivity analysis. Tables S7-S9 show the effects of varying the attendance rates. Tables S10 and S11 show the effects of halving or doubling the attributable costs. Finally, Tables S12 and S13 show the effects of halving or doubling the CT examination costs.

Table S14 shows the presence of the scenarios that were on the efficient frontier in the base-case analysis across the efficient frontiers of the sensitivity analyses. Table S14 shows that the cost-effectiveness estimates for scenarios #2, #3 and #9-11 in the base-case were the most insensitive to changes in assumptions, as they were part of the efficient frontier across 85.71-100% of the sensitivity analyses.

Table S7: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier when assuming high attendance rates (64% overall compliance)

Scenario # on current efficient frontier	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs per 100,000	Discounted life-years gained per 100,000	Costs per life-year gained (discounted) / Average cost-effectiveness ratio (ACER) compared to no screening	Incremental cost-effectiveness ratio (ICER) compared to the previous efficient scenario on the current frontier	Scenario present on efficient frontier in the base case analyses?
#1	55	75	Annual	10	40 pack-years (NLST-like)	Yes	$42,753,778	1,049	$40,757	-	Yes (Scenario #2)
#2	55	75	Annual	10	30 pack-years (NLST-like)	Yes	$52,196,709	1,224	$42,630	$53,830	Yes (Scenario #3)
#3	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$75,079,668	1,596	$47,035	$61,544	No
#4	50	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$79,653,298	1,667	$47,779	$64,518	No
#5	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$91,761,957	1,845	$49,739	$68,122	Yes (Scenario #9)
#6	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$111,914,998	2,059	$54,352	$94,081	Yes (Scenario #10)
#7	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$140,231,037	2,297	$61,039	$118,824	Yes (Scenario # 11)

Table S8: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier when assuming average attendance rates (55% overall compliance)

Scenario # on current efficient frontier	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs per 100,000	Discounted life-years gained per 100,000	Costs per life-year gained (discounted) / Average cost-effectiveness ratio (ACER) compared to no screening	Incremental cost-effectiveness ratio (ICER) compared to the previous efficient scenario on the current frontier	Scenario present on efficient frontier in the base case analyses?
#1	55	75	Annual	10	40 pack-years (NLST-like)	Yes	$38,598,562	911	$42,373	-	Yes (Scenario # 2)
#2	55	75	Annual	10	30 pack-years (NLST-like)	Yes	$46,793,945	1,064	$43,971	$53,470	Yes (Scenario #3)
#3	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$64,034,640	1,345	$47,594	$61,304	No
#4	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$66,780,467	1,390	$48,049	$61,835	No
#5	50	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$70,788,400	1,450	$48,824	$66,768	No
#6	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$81,603,480	1,606	$50,797	$69,058	Yes (Scenario #9)
#7	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$99,166,673	1,791	$55,369	$95,172	Yes (Scenario #10)
#8	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$123,825,644	2,003	$61,828	$116,460	Yes (Scenario #11)

Table S9: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier when assuming low attendance rates (33% overall compliance)

Scenario # on current efficient frontier	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs per 100,000	Discounted life-years gained per 100,000	Costs per life-year gained (discounted) / Average cost-effectiveness ratio (ACER) compared to no screening	Incremental cost-effectiveness ratio (ICER) compared to the previous efficient scenario on the current frontier	Scenario present on efficient frontier in the base case analyses?
#1	55	75	Annual	10	30 pack-years (NLST-like)	Yes	$31,451,475	605	$52,023	-	Yes (Scenario #3)
#2	55	75	Annual	10	30 pack-years (NLST-like)	No	$38,205,270	717	$53,306	$ 60,224	No
#3	55	75	Annual	15	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$42,416,487	785	$54,043	$ 61,793	No
#4	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$43,306,130	799	$54,185	$ 61,961	No
#5	50	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$45,749,850	838	$54,620	$ 63,688	No
#6	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$52,709,093	935	$56,386	$ 71,596	Yes (Scenario #9)
#7	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$63,053,644	1,043	$60,483	$ 96,041	Yes (Scenario #10)
#8	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$77,514,316	1,166	$66,455	$ 116,703	Yes (Scenario #11)

Table S10: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier when assuming 50% lower attributable costs

Scenario # on current efficient frontier	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs per 100,000	Discounted life-years gained per 100,000	Costs per life-year gained (discounted) / Average cost-effectiveness ratio (ACER) compared to no screening	Incremental cost-effectiveness ratio (ICER) compared to the previous efficient scenario on the current frontier	Scenario present on efficient frontier in the base case analyses?
#1	55	75	Biennial	10	40 pack-years (NLST-like)	Yes	$33,992,471	1,002	$33,910	-	No
#2	60	75	Annual	10	40 pack-years (NLST-like)	Yes	$43,525,926	1,276	$34,113	$34,858	Yes (Scenario #1)
#3	55	75	Annual	10	40 pack-years (NLST-like)	Yes	$51,688,904	1,489	$34,705	$38,240	Yes (Scenario #2)
#4	55	75	Annual	10	30 pack-years (NLST-like)	Yes	$65,098,514	1,746	$37,288	$52,289	Yes (Scenario #3)
#5	55	80	Annual	10	30 pack-years (NLST-like)	Yes	$69,954,321	1,834	$38,146	$55,166	Yes (Scenario #4)
#6	55	80	Annual	15	30 pack-years (NLST-like)	Yes	$85,678,445	2,088	$41,039	$61,940	Yes (Scenario #6)
#7	55	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$102,913,644	2,359	$43,633	$63,627	Yes (Scenario #7)
#8	55	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$112,455,785	2,500	$44,989	$67,654	Yes (Scenario #8)
#9	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$119,148,287	2,592	$45,970	$72,560	Yes (Scenario #9)
#10	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$149,041,280	2,877	$51,810	$104,968	Yes (Scenario #10)
#11	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$191,165,730	3,214	$59,471	$124,705	Yes (Scenario #11)

Table S11: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier when assuming 50% higher attributable costs

Scenario # on current efficient frontier	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs per 100,000	Discounted life-years gained per 100,000	Costs per life-year gained (discounted) / Average cost-effectiveness ratio (ACER) compared to no screening	Incremental cost-effectiveness ratio (ICER) compared to the previous efficient scenario on the current frontier	Scenario present on efficient frontier in the base case analyses?
#1	55	75	Annual	10	40 pack-years (NLST-like)	Yes	$72,272,006	1,489	$48,525	-	Yes (Scenario #2)
#2	55	75	Annual	10	30 pack-years (NLST-like)	Yes	$88,738,235	1,746	$50,828	$64,208	Yes (Scenario #3)
#3	55	75	Annual	15	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$111,723,824	2,050	$54,487	$75,455	No
#4	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$122,665,084	2,192	$55,952	$77,138	No
#5	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$128,075,204	2,261	$56,657	$79,299	No
#6	50	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$135,555,605	2,354	$57,594	$80,335	No
#7	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$157,493,361	2,592	$60,764	$92,085	Yes (Scenario #9)
#8	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$190,901,753	2,877	$66,362	$117,312	Yes (Scenario #10)
#9	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$237,777,957	3,214	$73,971	$138,772	Yes (Scenario #11)

Table S12: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier when assuming 50% lower CT examination costs

Scenario # on current efficient frontier	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs per 100,000	Discounted life-years gained per 100,000	Costs per life-year gained (discounted) / Average cost-effectiveness ratio (ACER) compared to no screening	Incremental cost-effectiveness ratio (ICER) compared to the previous efficient scenario on the current frontier	Scenario present on efficient frontier in the base case analyses?
#1	55	75	Annual	10	40 pack-years (NLST-like)	Yes	$42,641,746	1,489	$28,630	-	Yes (Scenario #2)
#2	55	75	Annual	10	30 pack-years (NLST-like)	Yes	$51,297,788	1,746	$29,383	$33,753	Yes (Scenario #3)
#3	55	75	Annual	15	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$63,374,225	2,050	$30,907	$39,644	No
#4	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$69,111,248	2,192	$31,524	$40,447	No
#5	55	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$71,942,244	2,261	$31,825	$41,495	No
#6	50	75	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$75,927,044	2,354	$32,259	$42,794	No
#7	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day **or** 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$87,643,184	2,592	$33,814	$49,179	Yes (Scenario #9)
#8	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day **or** 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$105,283,781	2,877	$36,599	$61,944	Yes (Scenario #10)
#9	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day **or** 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$130,052,038	3,214	$40,458	$73,324	Yes (Scenario #11)

Table S13: Cost-effectiveness estimates for lung cancer screening scenarios on the efficient frontier when assuming 50% higher CT examination costs

Scenario # on current efficient frontier	Starting age of screening	Stopping age of screening	Screening interval	Maximum number of years since cessation	Cumulative smoking criteria	Exclusion from further screening invitations after reaching the maximum number of years since cessation	Discounted costs per 100,000	Discounted life-years gained per 100,000	Costs per life-year gained (discounted) / Average cost-effectiveness ratio (ACER) compared to no screening	Incremental cost-effectiveness ratio (ICER) compared to the previous efficient scenario on the current frontier	Scenario present on efficient frontier in the base case analyses?
#1	60	75	Biennial	10	40 pack-years (NLST-like)	Yes	$46,837,102	853	$54,909	-	No
#2	55	75	Biennial	10	40 pack-years (NLST-like)	Yes	$55,977,990	1,002	$55,842	$61,167	No
#3	60	75	Annual	10	40 pack-years (NLST-like)	Yes	$74,675,267	1,276	$58,526	$68,364	Yes (Scenario #1)
#4	55	75	Annual	10	40 pack-years (NLST-like)	Yes	$90,366,322	1,489	$60,673	$73,507	Yes (Scenario #2)
#5	55	80	Annual	10	40 pack-years (NLST-like)	Yes	$97,714,197	1,564	$62,474	$98,372	No
#6	55	80	Annual	10	30 pack-years (NLST-like)	Yes	$125,086,809	1,834	$68,209	$101,463	Yes (Scenario #4)
#7	55	80	Annual	15	30 pack-years (NLST-like)	Yes	$155,585,056	2,088	$74,524	$120,138	Yes (Scenario #6)
#8	55	80	Annual	20	30 years of smoking at least 15 cigarettes per day or 35 years of smoking at least 10 cigarettes per day (NELSON-like)	Yes	$189,053,363	2,359	$80,155	$123,555	Yes (Scenario #7)
#9	55	80	Annual	20	30 years of smoking at least 15 cigarettes per day or 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$207,607,764	2,500	$83,055	$131,551	Yes (Scenario #8)
#10	50	80	Annual	20	30 years of smoking at least 15 cigarettes per day or 35 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$220,503,567	2,592	$85,075	$139,816	Yes (Scenario #9)
#11	50	80	Annual	20	25 years of smoking at least 15 cigarettes per day or 30 years of smoking at least 10 cigarettes per day (NELSON-like)	No	$278,416,750	2,877	$96,785	$203,360	Yes (Scenario #10)
#12	50	80	Annual	20	25 years of smoking at least 10 cigarettes per day or 30 years of smoking at least 5 cigarettes per day (NELSON-like)	No	$360,005,341	3,214	$111,996	$241,533	Yes (Scenario #11)

Table S14: Presence of the scenarios on the efficient frontier in the base-case analysis across the efficient frontiers of the sensitivity analyses

Scenario number on the efficient frontier in the base-case analysis	Scenario present on efficient frontier when assuming high attendance rates?	Scenario present on efficient frontier when assuming average attendance rates?	Scenario present on efficient frontier when assuming low attendance rates?	Scenario present on efficient frontier when assuming 50% lower attributable costs?	Scenario present on efficient frontier when assuming 50% higher attributable costs?	Scenario present on efficient frontier when assuming 50% lower CT examination costs?	Scenario present on efficient frontier when assuming 50% higher CT examination costs?	Presence of scenario across the efficient frontiers of the sensitivity analyses (in %)
Scenario #1	No	No	No	Yes	No	No	Yes	28.57%
Scenario #2	Yes	Yes	No	Yes	Yes	Yes	Yes	85.71%
Scenario #3	Yes	Yes	Yes	Yes	Yes	Yes	No	85.71%
Scenario #4	No	No	No	Yes	No	No	Yes	28.57%
Scenario #5	No	No	No	No	No	No	No	0.00%
Scenario #6	No	No	No	Yes	No	No	Yes	28.57%
Scenario #7	No	No	No	Yes	No	No	Yes	28.57%
Scenario #8	No	No	No	Yes	No	No	Yes	28.57%
Scenario #9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	100.00%
Scenario #10	Yes	Yes	Yes	Yes	Yes	Yes	Yes	100.00%
Scenario #11	Yes	Yes	Yes	Yes	Yes	Yes	Yes	100.00%

Chapter 9 | Supplementary material

References

1. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort–Specific Smoking Histories, 1965–2009. *American Journal of Preventive Medicine* 2014; **46**(2): e31-e7.
2. Anderson CM, Burns DM, Dodd KW, Feuer EJ. Chapter 2: Birth-Cohort-Specific Estimates of Smoking Behaviors for the U.S. Population. *Risk Analysis* 2012; **32**: S14-S24.
3. Statistics Canada 2011. Canadian Community Health Survey (CCHS), Cycle 1.1, 2.1, 3.1, 4.1 and 2008, 2009, 2010 Annual Component surveys [Microdata]. Ottawa Ontario: Statistics Canada Health Statistics Division [producer and distributor]: Ottawa Ontario: Data Liberation Initiative [distributor].
4. Statistics Canada 2012. National Population Health Survey (NPHS), Cycle 1-9: Household Component - Longitudinal [Canada] [Microdata]. Ottawa Ontario: Statistics Canada Health Statistics Division [producer and distributor]: Ottawa Ontario: Data Liberation Initiative [distributor].
5. Statistics Canada & Health and Welfare Canada.Canada health survey, 1978-1979 [computer file]. Ottawa, Ont.: Statistics Canada [producer]; Statistics Canada. Data Liberation Initiative [distributor], 1983.
6. Statistics Canada. 1992. General Social Survey,1991 [Canada]: Cycle 6, Health [public use microdata file]. Ottawa, Ontario: Statistics Canada [producer and distributor]. .
7. Statistics Canada. 1986. General Social Survey, Cycle 1, 1985 [Canada]: Health and Social Support [public use microdata file]. Ottawa, Ontario: Statistics Canada [producer and distributor].
8. Statistics Canada. Survey of Smoking Habits, 1971-1986 [Canada] [Public-use microdata file(s)]. Ottawa, Ontario: Special Surveys Division, Statistics Canada [producer and distributor]. Ottawa, Ontario: Data Liberation Initiative [distributor].
9. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
10. Canadian Human Mortality Database. Department of Demography, Université de Montréal (Canada). Available at www.demo.umontreal.ca/chmd/ (data downloaded on 14-03-2014). [Internet].
11. Levy DT, Romano E, Mumford E. The relationship of smoking cessation to sociodemographic characteristics, smoking intensity, and tobacco control policies. *Nicotine & Tobacco Research* 2005; **7**(3): 387-96.
12. Fiore MC, Novotny TE, Pierce JP, et al. Methods used to quit smoking in the United States. Do cessation programs help? *Journal of the American Medical Association* 1990; **263**(20): 2760-5.
13. Warner KE. Possible Increases in the Underreporting of Cigarette Consumption. *Journal of the American Statistical Association* 1978; **73**(362): 314-8.
14. Hatziandreu EJ, Pierce JP, Fiore MC, Grise V, Novotny TE, Davis RM. The reliability of self-reported cigarette consumption in the United States. *American Journal of Public Health* 1989; **79**(8): 1020-3.
15. Burns D, Major JM, Shanks TG. Chapter 7: Changes in number of cigarettes smoked per day: cross-sectional and birht cohort analyses using NHIS. 2003. In: Those Who Continue to Smoke Smoking and Tobacco Control Monograph No 15 NIH Pub No 03-5370, September 2003 [Internet].
16. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Analysis* 2012; **32 Suppl 1**: S51-68.
17. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Analysis* 2012; **32 Suppl 1**: S25-38.
18. https://www.cancercare.on.ca/cms/one.aspx?objectId=121939&contextId=1377.
19. Thun MJ, Carter BD, Feskanich D, et al. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine* 2013; **368**(4): 351-64.
20. Greene FL, Page DL, Fleming ID, Fritz AG, Balch CM, Haller DG, et al. eds, for the American joint committee on cancer. AJCC Cancer Staging Manual. 6th ed. New York, NY: Springer-Verlag; 2002.
21. ICD-O-3 SEER Site/Histology Validation List - 12/04/2009 - updated from 2/9/2001 [cited Accessed Jan 12 2012]. Available from: http://seer.cancer.gov/icd-o-3/.
22. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al. International classification of diseases for oncology. 3rd ed. Geneva: World Health Organization; 2000.

23. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence – SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2012 Sub (1973–2010 varying) – Linked To County Attributes – Total U.S., 1969–2011 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2013, based on the November 2012 submission. Accessed March 29, 2014.

24. Groome PA, Bolejack V, Crowley JJ, et al. The IASLC Lung Cancer Staging Project: Validation of the Proposals for Revision of the T, N, and M Descriptors and Consequent Stage-groupings in the Forthcoming (Seventh) Edition of the TNM Classification of Malignant Tumours. *Journal of Thoracic Oncology* 2007; **2**(8): 694-705.

Chapter 10

Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study

Kevin ten Haaf, Jihyoun Jeon, Martin C. Tammemägi, Summer S. Han, Chung Yin Kong, Sylvia K. Plevritis, Eric J. Feuer, Harry J. de Koning, Ewout W. Steyerberg, Rafael Meza

Published as:

ten Haaf K, Jeon J, Tammemägi, MC et al

PLOS Medicine 2017*;* **14**(4): e1002277.

Abstract

Background

Selection of candidates for lung cancer screening based on individual risk has been proposed as an alternative to criteria based on age and cumulative smoking exposure (pack-years). Nine previously established risk models were assessed for their ability to identify those most likely to develop or die from lung cancer. All models considered age and various aspects of smoking exposure (smoking status, smoking duration, cigarettes per day, pack-years smoked, time since smoking cessation) as risk predictors. In addition, some models considered factors such as gender, race, ethnicity, education, body mass index, chronic obstructive pulmonary disease, emphysema, personal history of cancer, personal history of pneumonia, and family history of lung cancer.

Methods and findings

Retrospective analyses were performed on 53,452 National Lung Screening Trial (NLST) participants (1,925 lung cancer cases and 884 lung cancer deaths) and 80,672 Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) ever-smoking participants (1,463 lung cancer cases and 915 lung cancer deaths). Six-year lung cancer incidence and mortality risk predictions were assessed for (1) calibration (graphically) by comparing the agreement between the predicted and the observed risks, (2) discrimination (area under the receiver operating characteristic curve [AUC]) between individuals with and without lung cancer (death), and (3) clinical usefulness (net benefit in decision curve analysis) by identifying risk thresholds at which applying risk-based eligibility would improve lung cancer screening efficacy. To further assess performance, risk model sensitivities and specificities in the PLCO were compared to those based on the NLST eligibility criteria. Calibration was satisfactory, but discrimination ranged widely (AUCs from 0.61 to 0.81). The models outperformed the NLST eligibility criteria over a substantial range of risk thresholds in decision curve analysis, with a higher sensitivity for all models and a slightly higher specificity for some models. The PLCOm2012, Bach, and Two-Stage Clonal Expansion incidence models had the best overall performance, with AUCs >0.68 in the NLST and >0.77 in the PLCO. These three models had the highest sensitivity and specificity for predicting 6-year lung cancer incidence in the PLCO chest radiography arm, with sensitivities >79.8% and specificities >62.3%. In contrast, the NLST eligibility criteria yielded a sensitivity of 71.4% and a specificity of 62.2%. Limitations of

this study include the lack of identification of optimal risk thresholds, as this requires additional information on the long-term benefits (e.g., life-years gained and mortality reduction) and harms (e.g., overdiagnosis) of risk-based screening strategies using these models. In addition, information on some predictor variables included in the risk prediction models was not available.

Conclusions

Selection of individuals for lung cancer screening using individual risk is superior to selection criteria based on age and pack-years alone. The benefits, harms, and feasibility of implementing lung cancer screening policies based on risk prediction models should be assessed and compared with those of current recommendations.

Introduction

The National Lung Screening Trial (NLST) found that screening with low-dose computed tomography (CT) can reduce lung cancer mortality by 20%.[1] Based on an evidence review, including the results of the NLST and a comparative microsimulation modeling study, the United States Preventive Services Task Force (USPSTF) recommended lung cancer screening for current and former smokers aged 55 through 80 years who smoked at least 30 pack-years and, if quit, quit less than 15 years ago.[2-4] To our knowledge, only the United States has implemented lung cancer screening policies. Although the province of Ontario, Canada, recommends screening individuals at high risk for lung cancer through an organized program, no program has yet been established.[5] Cancer Care Ontario (the provincial cancer agency of Ontario) is currently evaluating the feasibility of implementing such a program.[6] European countries have not yet made any recommendations on lung cancer screening, as the final results of the Dutch-Belgian Lung Cancer Screening Trial (Nederlands-Leuvens Longkanker Screenings Onderzoek [NELSON] trial), potentially pooled with high-quality data from other trials, are still awaited.[7-9]

The screening eligibility criteria used in the current USPSTF recommendations are based on age and pack-years, a measure of cumulative smoking exposure. Thus, these recommendations do not take other important risk factors into account, such as family history, nor other relevant aspects of smoking, such as smoking duration or intensity. Recently, a number of investigations have suggested that determining screening eligibility using an individual's risk based on age, more detailed smoking history, and other risk factors, such as ethnicity and family history of lung cancer, could lead to more effective screening programs compared with the USPSTF recommendations.[10-13] Indeed, some lung cancer screening guidelines already encourage assessment of an individual's risk to determine screening eligibility.[14] While various lung cancer risk prediction models have been developed, external validation and direct comparisons between models have been limited due to insufficient numbers of events or methodological limitations.[15-21] Such validations are essential, as risk prediction models generally have optimistic performance within their development dataset.[15-17] This study aims to externally validate and directly compare the

performance of nine currently available lung cancer risk prediction models for stratifying lung cancer risk groups and determining screening eligibility.

Methods

Ethics statement

No identifiable information was used; therefore, no institutional review board (IRB) approval was needed. Nonetheless, a determination of exempt was given by the University of Michigan IRB (HUM00054750), and a determination of this not being human subjects research was given by the Fred Hutchinson Cancer Research Center (former affiliation of J. J.) IRB (6007-680).

Study population

We used data from two large randomized controlled screening trials: the NLST and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).[1,22-24] All participants in the CT arm (n = 26,722) and chest radiography (CXR) arm (n = 26,730) of the NLST and ever-smoking participants in the CXR arm (n = 40,600) and control arm (n = 40,072) of the PLCO were included in the analysis. Never-smokers in the PLCO were not considered, as (1) not all lung cancer risk prediction models can be applied to never-smokers and (2) never-smokers are unlikely to reach levels of risk that allow them to benefit from screening.[13,25] Data on the predictor variables in each trial were collected through epidemiologic questionnaires administered at study entry and harmonized across both trials. Reported average numbers of cigarettes smoked per day above 100 were considered implausible and recoded as 100 cigarettes per day (n = 11). Furthermore, body mass index values less than 14 and over 60 kg/ m2 were considered implausible for enrollment in both trials and recoded as 14 (n = 5) and 60 kg/m2 (n = 18), respectively. Lung cancer diagnoses (1,925 in the NLST and 1,463 in the PLCO) and lung cancer deaths (884 in the NLST and 915 in the PLCO) that occurred between study entry and 6 years of follow-up were included in the final dataset and were considered as binary outcomes.

Lung cancer risk prediction models

Our study includes nine risk prediction models for lung cancer incidence or death that have been used frequently in the literature. Risk prediction models were not considered for this investigation, if they (1) were developed for specific ethnicities and are therefore not broadly applicable, (2) used information on biomarkers or lung nodules and are therefore not readily applicable for the prescreening selection of individuals, (3) were developed for identifying symptomatic patients, (4) did not incorporate smoking behavior, (5) did not provide information on parameter estimates (e.g., baseline risk parameters) necessary to allow replication of the model, or (6) had poor discriminative ability in their development dataset.[11,12,26-37] Nine models remained and were investigated: the Bach model, the Liverpool Lung Project (LLP) model, the PLCOm2012 model, the Two-Stage Clonal Expansion (TSCE) model for lung cancer incidence, the Knoke model, two versions of the TSCE model for lung cancer death, and simplified versions of the PLCOm2012 and LLP models.[10,38-44] The characteristics of these models are shown in Table 1.

The TSCE and Knoke models consider only age, gender, and smoking-related characteristics as risk factors.[40-43] The Bach model considers asbestos exposure as an additional risk factor, while the LLP and PLCOm2012 models consider multiple additional risk factors.[10,38,39] The simplified versions of the PLCOm2012 and LLP models considered only age, gender, and smoking variables. A detailed description of each model can be found in the supplementary material of this Chapter, in the section: “Lung cancer risk prediction model descriptions”. Data on frequency and intensity of asbestos exposure, used in the LLP and Bach models, was not available for the PLCO participants and could not be accurately derived for the NLST participants.[38,39] Therefore, we assumed that none of the participants were exposed to asbestos, even though this assumption may lead to biased estimates.[45] However, as the potential number of individuals with asbestos exposure was low (less than 5% of the NLST participants reported ever working with asbestos), this bias is expected to be minor.[46] The LLP model incorporates age at lung cancer diagnosis of a first-degree relative: early age (60 years or younger) versus late age (older than 60 years).[38] However, while both the PLCO and the NLST had information about the occurrence of family history of lung cancer (yes/no), neither had information on the age of diagnosis for the affected relative(s). Since the median age of lung cancer diagnosis in the United States is 70 years and the majority of lung

cancers occur after the age of 65 years (68.6%), we assumed that lung cancer in first-degree relatives in the PLCO and the NLST always occurred after the age of 60 years.[47,48] In addition, the LLP model incorporates a history of pneumonia as a risk factor.[38] While information on this risk factor was available in the NLST, it was not available in the PLCO. Therefore, we assumed that none of the PLCO participants had a history of pneumonia for the complete case analyses. While 22.1% of NLST participants had a history of pneumonia (Table 2), the association of a history of pneumonia with a lung cancer diagnosis within 6 years was not clear (p = 0.338 in the CT arm and p = 0.004 in the CXR arm). Missing history of pneumonia for PLCO participants was imputed by using information from the NLST participants.[49]

Statistical analyses

To assess the performance of the risk prediction models, several metrics were employed: calibration, discrimination, and clinical usefulness (net benefit over a range of risk thresholds).[50] The performance of the investigated risk prediction models was assessed in each trial arm separately, for both lung cancer incidence and lung cancer mortality. We assessed both lung cancer incidence and mortality in both arms of both trials for all investigated risk models, as these outcomes may be influenced differently by screening. Screening may affect the predictive performance for lung cancer incidence, due to the advance in time of detection due to screening (lead-time) and the detection of cancers that would never have been detected if screening had not occurred (overdiagnosis).[51-53] Furthermore, CT screening reduces lung cancer mortality compared to CXR screening, which may influence the predictive performance of models for lung cancer mortality in the CT arm of the NLST.[1] Furthermore, the sensitivity and specificity of each model in the PLCO cohorts were compared to the sensitivity and specificity of the NLST/USPSTF smoking eligibility criteria (being a current or former smoker who smoked at least 30 pack-years and, if quit, quit less than 15 years ago). Model performance was assessed by varying follow-up duration and outcome (5- and 6-year lung cancer incidence or mortality) to investigate the effect of follow-up duration on the discrimination performance of each model.[54] The 5- and 6-year time frames were chosen because the LLP and PLCOm2012 models were calibrated to these respective time frames, and complete follow-up of NLST participants was limited to 6 year.[10,38] Since performance was similar for 5- and 6-year outcomes, only the results of the 6-year outcomes are presented. Performance was evaluated for the risk prediction models

as presented in their original publication, without any recalibration or reparameterization to the NLST and the PLCO. The only exception is the PLCOm2012 model, which was originally developed based on data from the control arm of the PLCO.[10] All analyses were performed in R (version 3.3.0).[55]

Aspects of calibration performance

Calibration plots were constructed for the observed proportions of outcome events against the predicted risks for individuals grouped by similar ranges of predicted risk.[56] Perfect predictions should show an ideal 45-degree line that can be described by an intercept of 0 and a slope of 1 in the calibration plot.[57] The calibration intercept quantifies the extent to which a model systematically under- or overestimates a person's risk; an intercept value of 0 represents perfect calibration in the large. The calibration slope was estimated by logistic regression analysis, using the log odds of the predictions for the single predictor of the binary outcome.[50] For a (near-)perfect calibration in the large, a calibration slope less than 1 reflects that predictions for individuals with low risk are too low and predictions for individuals with high risk are too high.[50] The calibration plots, calibration in the large, and calibration slopes for each model were obtained using the R package rms.[58]

Discrimination

Discrimination reflects the capability of a model to distinguish individuals with the event from those without the event; the risk predicted by the model should be higher for individuals with the event compared with those without the event.[59] The area under the receiver operating characteristic curve (AUC) was used to assess discrimination, which ranges between 0.5 and 1.0 for sensible models. The AUCs for each model were obtained using the R package rms.[58]

Clinical usefulness

While discrimination and calibration are important statistical properties of a risk prediction model, they do not assess its clinical usefulness.[50,54,59] For example, if a false-negative result causes greater harm than a false-positive result, one would prefer a model with a higher sensitivity over a model that has a greater specificity but a slightly lower sensitivity, even

though the latter might have a higher AUC.[60] In the context of selecting individuals for lung cancer screening, a model is clinically useful if applying that model to determine screening eligibility yields a better ratio of benefits to harms than not applying it. Decision curve analysis has been proposed to assess the net benefit of using a risk prediction model.[60,61] Decision curve analysis evaluates the net benefit of a model over a range of risk thresholds, i.e., the level of risk used to classify predictions as positive or negative for the predicted outcome. For example, for the PLCOm2012 model, a risk threshold of 1.51% has been suggested, meaning that individuals with an estimated risk of 1.51% or higher are classified as positive (and thus eligible for screening) and individuals with an estimated risk lower than 1.51% as negative (and thus ineligible for screening).[13] The net benefit is defined as:

$$\text{Net benefit} = \frac{\text{True positive count} - (\text{false positive count} * \text{weighting factor})}{\text{Number of individuals assessed for screening eligibility}}$$

where the weighting factor is defined as:

$$\text{Weighting factor} = \frac{\text{Risk threshold}}{(1 - \text{Risk threshold})}$$

This weighting factor represents how the relative harms of false-positive (classifying a person as eligible for screening who does not develop, or die from, lung cancer) and false-negative (classifying a person as ineligible for screening who develops, or dies from, lung cancer) results are valued at a given risk threshold, i.e., the ratio of harm to benefit, and is estimated by the threshold odds. For example, a risk threshold of 2.5% yields the following weighting factor:

$$\text{Weighting factor} = \frac{0.025}{(1 - 0.025)} = \frac{1}{39}$$

This weighting factor implies that missing one case of lung cancer that could be detected through screening is valued as 39 times worse than unnecessarily screening one person, or that one case should be detected per 40 screened persons. Consequently, the less relative weight one gives to detecting a lung cancer case, the higher the risk threshold one will favor. The net benefit can then be interpreted as follows: if the net benefit at a risk threshold of 2.5% is 0.002 greater compared with screening all persons eligible according to the NLST criteria, taking the weighing factor into account, this is equivalent to a net improvement in

true-positive results of 0.002 * 1,000 = 2 per 1,000 persons assessed for screening eligibility, or a net reduction in false-positive results of 0.002 * 1,000/(0.025/0.975) = 78 per 1,000 persons assessed for screening eligibility.[60] Thus, if the risk model has a positive net benefit at the preferred risk threshold, this indicates that applying the model at this risk threshold provides a better ratio of benefits to harms than current screening guidelines based on pack-years. Decision curves visualize the net benefit over a range of risk thresholds, allowing one to discern whether and at which risk thresholds applying the risk model can be clinically useful.[61] Decision curves were used to determine at which range of risk thresholds applying the models provides a net benefit over using the NLST eligibility criteria for selecting individuals for lung cancer screening. Finally, we identified the risk threshold for each model in the PLCO cohorts that selected a similar number of individuals for screening as the NLST eligibility criteria, on which most lung cancer screening recommendations are currently based. We then assessed the sensitivity (the number of individuals with lung cancer incidence or death classified as eligible for screening divided by the total number of individuals with lung cancer incidence or death) and specificity (the number of individuals without lung cancer incidence or death classified as ineligible for screening divided by the total number of individuals without lung cancer incidence or death) for each model compared to the NLST criteria at the chosen risk threshold, as reported before by Tammemägi et al.[13]

Multiple imputation of missing values

Multiple imputation of missing data for all considered risk factors was performed through the method of chained equations using the R package MICE.[62] History of pneumonia was not measured in the PLCO but was measured in the NLST; therefore, data from the NLST were used to impute history of pneumonia for PLCO participants.[49] Analyses were performed using 20 imputations, and the results were pooled through applying Rubin's rules.[63] The results of the analyses with imputation of missing variables were similar to those obtained from complete case analyses. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines suggest applying multiple imputation when missing data are present, as complete case analyses can lead to inefficient estimates.[64,65] Therefore, all analyses reported here were performed with multiple imputation of missing values.

Results

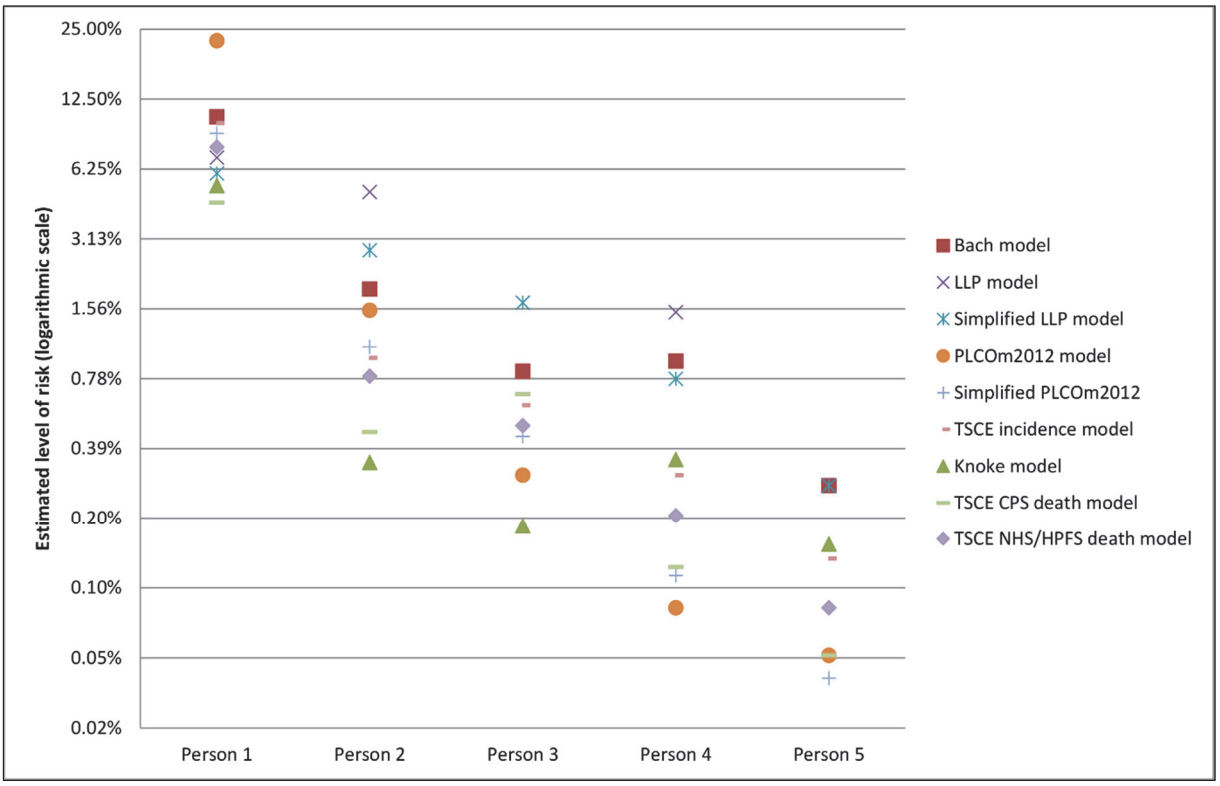
Characteristics of study populations

An overview of the characteristics of the four study cohorts (two trial arms in each trial) is given in Table 2, stratified by 6-year lung cancer incidence. A similar table stratifying participants by 6-year lung cancer mortality is provided in Table S6 in the supplementary material of this Chapter. An overview of the proportion of individuals with complete information on all risk factors, stratified by trial arm and 6-year outcome, is given in Table S7 in the supplementary material of this Chapter. Overall, approximately 93% of the study population had complete information for all considered risk factors.

Differences in levels of absolute risk

The risk prediction models included in this study were developed in different populations (Table 1) and incorporate risk factors, specifically smoking behavior, in different ways (shown in the supplementary material of this Chapter, in the section: "Lung cancer risk prediction model descriptions"). In addition, some models predict lung cancer incidence, while others predict lung cancer mortality. Therefore, the estimated absolute risk for the same individual varies between models.[66] Figure 1 shows the estimated 6-year risk of lung cancer incidence or mortality (depending on the target outcome of the model) across the models for five individuals with different risk factor profiles. This difference in estimated absolute risk between models suggests that specific risk thresholds might be needed for each model.

Figure 1: Examples of projected absolute risk for individuals with different risk factor profiles by model



Abbreviations: BMI, body mass index; COPD, chronic obstructive pulmonary disease; CPS, Cancer Prevention Study; HPFS, Health Professionals Follow-up Study; LLP, Liverpool Lung Project; NHS, Nurses' Health Study; TSCE, Two-Stage Clonal Expansion.

Person 1: 70-year-old, high-school graduated white male, current smoker, who smoked 30 cigarettes per day for 55 years, has a BMI of 28 kg/m2, has COPD, no asbestos exposure, no personal history of cancer, no personal history of pneumonia, but has a family history of lung cancer (relative was diagnosed at age > 60 years).

Person 2: 63-year-old, college-graduated black woman, former smoker who quit 10 years ago, who smoked 15 cigarettes per day for 40 years, has a BMI of 25 kg/m2, does not have COPD, no asbestos exposure, no personal history of cancer, has a personal history of pneumonia, and no family history of lung cancer.

Person 3: 65-year-old Asian male with some college education, former smoker who quit 14 years ago, who smoked 10 cigarettes per day for 30 years, has a BMI of 24 kg/m2, does not have COPD, has asbestos exposure, no personal history of cancer, no personal history of pneumonia, and no family history of lung cancer.

Person 4: 58-year-old, post-graduate-educated Hispanic woman, current smoker, who smoked 5 cigarettes per day for 38 years, has a BMI of 22 kg/m2, does not have COPD, no asbestos exposure, has a personal history of cancer, no personal history of pneumonia, and no family history of lung cancer.

Person 5: 50-year-old, college-educated white woman, current smoker, who smoked 5 cigarettes per day for 30 years, has a BMI of 22 kg/m2, does not have COPD, no asbestos exposure, no personal history of cancer, no personal history of pneumonia, and no family history of lung cancer.

Table 1: Characteristics of investigated risk models

Model	Predicted outcome	Model prediction timeframe	Development dataset(s)	Risk factors incorporated in model	Reference
Bach model*	Lung cancer incidence	1-year (iterative)	Carotene and Retinol Efficacy Trial (CARET)	Age, gender, smoking duration, smoking intensity, years since cessation, asbestos exposure	39
Liverpool Lung Project (LLP) model†	Lung cancer incidence	5 years	Liverpool Lung Project (LLP) case-control study	Age, gender, smoking duration, personal history of cancer, family history of lung cancer, personal history of pneumonia, asbestos exposure	44
PLCOm2012 model†	Lung cancer incidence	6 years	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)	Age, race, education, BMI, COPD, personal history of cancer, family history of lung cancer, smoking status, smoking duration, smoking intensity, years since cessation	10
Two-Stage Clonal Expansion (TSCE) lung cancer incidence version	Lung cancer incidence	1-year (iterative)	Nurses' Health Study (NHS), Health Professionals Follow-up Study (HPFS)	Age, gender, smoking status, smoking duration, smoking intensity, years since cessation	43
Knoke model	Lung cancer death	1-year (iterative)	American Cancer Society's first Cancer Prevention Study (CPS-I)	Age, smoking status, smoking duration, smoking intensity, years since cessation	40
Two-Stage Clonal Expansion (TSCE) CPS lung cancer death model	Lung cancer death	1-year (iterative)	British Doctors' cohort, American Cancer Society's first Cancer Prevention Study (CPS-I), American Cancer Society's second Cancer Prevention Study (CPS-II)	Age, gender, smoking status, smoking duration, smoking intensity, years since cessation	41
Two-Stage Clonal Expansion (TSCE) NHS/HPFS lung cancer death model	Lung cancer death	1-year (iterative)	Nurses' Health Study (NHS), Health Professionals Follow-up Study (HPFS)	Age, gender, smoking status, smoking duration, smoking intensity, years since cessation	42

**Data on asbestos exposure was not available for PLCO participants and could not be accurately derived for NLST participants. Therefore, only age, gender and smoking related characteristics were considered for this model. †Simplified versions of these models, using only age, gender and smoking related characteristics, were considered as well.*

Table 2: Baseline characteristics of National Lung Screening Trial and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial participants according to 6-year lung cancer incidence

Characteristic	NLST (computed tomography-arm)			NLST (chest radiography arm)			PLCO (chest radiography arm)			PLCO (control arm)		
Number (percent) of participants	No lung cancer 25,692 (96.15%)	Lung cancer 1,030 (3.85%)	P-value	No lung cancer 25,835 (96.65%)	Lung cancer 895 (3.35%)	P-value	No lung cancer 39,846 (98.14%)	Lung cancer 754 (1.86%)	P-value	No lung cancer 39,363 (98.23%)	Lung cancer 709 (1.77%)	P-value
Age (years)												
Median (IQR)	60 (57-65)	63 (59-68)	<0.0001	60 (57-65)	64 (60-68)	<0.0001	62 (58-66)	65 (60.25-69)	<0.0001	62 (58-66)	65 (60-69)	<0.0001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Gender												
Male	15,148 (58.96%)	621 (60.29%)	0.401	15,235 (58.97%)	526 (58.77%)	0.917	23,228 (58.29%)	475 (63.00%)	0.010	22,775 (57.86%)	438 (61.78%)	0.038
Female	10,544 (41.04%)	409 (39.71%)		10,600 (41.03%)	369 (41.23%)		16,618 (41.71%)	279 (37.00%)		16,588 (42.14%)	271 (38.22%)	
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Hispanic ethnicity												
No	25,070 (97.58%)	1,008 (97.86%)	0.041	25,158 (97.38%)	881 (98.44%)	0.004	38,116 (95.66%)	731 (96.95%)	0.314	37,597 (95.51%)	690 (97.32%)	0.093
Yes	469 (1.83%)	10 (0.97%)		451 (1.75%)	5 (0.56%)		872 (2.19%)	12 (1.59%)		873 (2.22%)	9 (1.27%)	
Missing	153 (0.60%)	12 (1.17%)		226 (0.87%)	9 (1.01%)		858 (2.15%)	11 (1.46%)		893 (2.27%)	10 (1.41%)	
Race or ethnic group												
White	23,019 (89.60%)	933 (90.58%)	0.323	23,143 (89.58%)	806 (90.06%)	0.090	35,180 (88.29%)	644 (85.41%)	<0.001	34,787 (88.37%)	630 (88.86%)	0.007
Black	1,140 (4.44%)	47 (4.56%)		1,123 (4.35%)	51 (5.70%)		2,252 (5.65%)	73 (9.68%)		2,201 (5.59%)	53 (7.48%)	
Hispanic	337 (1.31%)	7 (0.68%)		314 (1.22%)	4 (0.45%)		810 (2.03%)	12 (1.59%)		807 (2.05%)	8 (1.13%)	
Asian	541 (2.11%)	18 (1.75%)		522 (2.02%)	14 (1.56%)		1232 (3.09%)	16 (2.12%)		1199 (3.05%)	13 (1.83%)	
Native Hawaiian or Pacific Islander	88 (0.34%)	3 (0.29%)		100 (0.39%)	2 (0.22%)		219 (0.55%)	8 (1.06%)		244 (0.62%)	1 (0.14%)	
American Indian or Alaskan Native	86 (0.33%)	6 (0.58%)		95 (0.37%)	3 (0.34%)		130 (0.33%)	1 (0.13%)		105 (0.27%)	4 (0.56%)	
Missing	481 (1.87%)	16 (1.55%)		538 (2.08%)	15 (1.68%)		23 (0.06%)	0 (0%)		20 (0.05%)	0 (0%)	
Education												
Less than high school graduate	1,552 (6.04%)	89 (8.64%)	<0.001	1,525 (5.90%)	83 (9.27%)	<0.001	3,424 (8.59%)	123 (16.31%)	<0.001	3,418 (8.68%)	96 (13.54%)	<0.001
High school graduate	5,989 (23.31%)	284 (27.57%)		6,177 (23.91%)	261 (29.16%)		8,775 (22.20%)	199 (26.39%)		8,647 (21.97%)	197 (27.79%)	
Post high school training	3,587 (13.96%)	146 (14.17%)		3,562 (13.79%)	139 (15.53%)		5,341 (13.40%)	92 (12.20%		5,393 (13.70%)	98 (13.82%)	
Some college	5,957 (23.19%)	232 (22.52%)		5,893 (22.81%)	195 (21.79%)		9,269 (23.26%)	165 (21.88%)		9,159 (23.27%)	161 (22.71%)	
College graduate	4,357 (16.96%)	148 (14.37%)		4,342 (16.81%)	99 (11.06%)		6,626 (16.63%)	97 (12.86%)		6,385 (16.22%)	94 (13.26%)	
Postgraduate/professional	3,679 (14.32%)	101 (9.81%)		3,718 (14.39%)	102 (11.40%)		6,352 (15.94%)	77 (10.21%)		6,218 (15.80%	59 (8.32%)	
Missing	571 (2.22%)	30 (2.91%)		618 (2.39%)	16 (1.79%)		59 (0.15%)	1 (0.13%)		143 (0.36%)	4 (0.56%)	
BMI (kg/m²)												
Median (IQR)	27.32 (24.46-30.73)	26.38 (23.94-29.28)	<0.001	27.38 (24.50-30.63)	26.22 (23.62-29.25)	<0.001	26.68 (24.19-30.02)	26.16 (23.48-28.77)	<0.001	26.68 (24.18-29.90)	25.88 (23.45-28.81)	<0.001
Missing	146 (0.57%)	13 (1.26%)		206 (0.80%)	7 (0.78%)		494 (1.24%)	10 (1.33%)		742 (1.89%)	15 (2.12%)	
COPD												
No	21,283 (82.84%)	765 (74.27%)	<0.001	21,435 (82.97%)	643 (71.84%)	<0.001	36,381 (91.30%)	602 (79.84%)	<0.001	35,899 (91.20%)	567 (79.97%)	<0.001
Yes	4,409 (17.16%)	265 (25.73%)		4,400 (17.03%)	252 (28.16%)		3,465 (8.70%)	152 (20.16%)		3,464 (8.80%)	142 (20.03%)	
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)
Emphysema												
No	23,661 (92.09%)	878 (85.24%)	<0.001	23,734 (91.87%)	766 (85.59%)	<0.001	38,011 (95.39%)	655 (86.87%)	<0.001	37,404 (95.02%)	599 (84.49%)	<0.001
Yes	1,910 (7.43%)	146 (14.17%)		1,915 (7.41%)	122 (13.63%)		1,650 (4.14%)	96 (12.73%)		1,612 (4.10%)	99 (13.96%)	
Missing	121 (0.47%)	6 (0.58%)		186 (0.72%)	7 (0.78%)		185 (0.46%)	3 (0.40%)		347 (0.88%)	11 (1.55%)	
Personal history of cancer												
No	24,588	956	<0.001	24,554	833	0.007	38,033	709	0.078	37,532	653	<0.001

	(95.70%)	(92.82%)		(95.04%)	(93.07%)		(95.45%)	(94.03%)		(95.35%)	(92.10%)	
Yes	1,028 (4.00%)	68 (6.60%)		1,154 (4.47%)	58 (6.48%)		1,813 (4.55%)	45 (5.97%)		1,831 (4.65%)	56 (7.90%)	
Missing	76 (0.30%)	6 (0.58%)		127 (0.49%)	4 (0.45%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Family history of lung cancer												
No	19,741 (76.84%)	746 (72.43%)	0.004	19,812 (76.69%)	640 (71.51%)	0.001	33,718 (84.62%)	565 (74.93%)	<0.001	33,485 (85.07%)	541 (76.30%)	<0.001
Yes	5,554 (21.62%)	261 (25.34%)		5,570 (21.56%)	236 (26.37%)		4,514 (11.33%)	139 (18.44%)		4,414 (11.21%)	130 (18.34%)	
Missing	397 (1.55%)	23 (2.23%)		453 (1.75%)	19 (2.22%)		1,614 (4.05%)	50 (6.63%)		1,464 (3.72%)	38 (5.36%)	
Personal history of pneumonia												
No	19,905 (77.48%)	781 (75.83%)	0.338	20,023 (77.50%)	657 (73.41%)	0.004	Not measured in PLCO		-	Not measured in PLCO		-
Yes	5,690 (22.15%)	240 (23.30%)		5,646 (21.85%)	233 (26.03%)		Not measured in PLCO			Not measured in PLCO		
Missing	97 (0.38%)	9 (0.87%)		166 (0.64%)	5 (0.56%)		Not measured in PLCO			Not measured in PLCO		
Smoking status												
Current smoker	12,183 (47.42%)	601 (58.35%)	<0.001	12,274 (47.51%)	558 (62.35%)	<0.001	7,744 (19.43%)	332 (44.04%)	<0.001	7,655 (19.45%)	324 (45.70%)	<0.001
Former smoker	13,509 (52.58%)	429 (41.65%)		13,561 (52.49%)	337 (37.65%)		32,102 (80.57%)	422 (55.97%)		31,708 (80.55%)	385 (54.30%)	
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Smoking duration (years)												
Median (IQR)	40 (35-44)	44 (40-49)	<0.001	40 (35-44)	44 (40-49)	<0.001	28 (16-39)	42 (35-48)	<0.001	28 (16-39)	42 (35-47)	<0.001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		766 (1.92%)	10 (1.33%)		877 (2.23%)	17 (2.40%)	
Smoking intensity (cigarettes per day)												
Median (IQR)	25 (20-35)	30 (20-40)	<0.001	25 (20-30.5)	25 (20-40)	0.029	20 (10-30)	30 (20-40)	<0.001	20 (10-30)	30 (20-40)	<0.001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		78 (0.20%)	4 (0.53%)		112 (0.28%)	2 (0.28%)	
Pack-years of smoking												
Median (IQR)	48 (39-66)	57 (45-82)	<0.001	48 (39-66)	55.5 (44-78)	<0.001	28.5 (14-48)	51 (38-74)	<0.001	29 (14-49)	54 (40-75)	<0.001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		827 (2.1%)	13 (1.7%)		957 (2.4%)	19 (2.7%)	
Years since cessation												
Median (IQR)	7 (3-11)	5 (2-10)	<0.001	7 (3-11)	6 (2-11)	0.067	20 (10-30)	10 (4-19)	<0.001	20 (10-30)	10 (4-18.25)	<0.001
Missing	219 (0.9%)	5 (0.5%)		216 (0.8%)	8 (0.9%)		561 (1.4%)	5 (0.7%)		679 (1.7%)	5 (0.7%)	

Abbreviations: BMI, body mass index; COPD, chronic obstructive pulmonary disease; IQR, interquartile range; NLST, National Lung Screening Trial; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

Data are given as n (percent) or median (IQR).

Aspects of calibration performance

Overall, all models showed satisfactory calibration performance (shown in Figures S2-S19 and Tables S8-S11 in the supplementary material of this Chapter). The models showed the best calibration performance when they were applied to their target outcome, i.e., lung cancer incidence rather than lung cancer mortality for lung cancer incidence models. The calibration was better for all models in the PLCO datasets than in the NLST datasets.

Discrimination

The discriminative performance of the models (Figures 2-5) was better in the PLCO datasets (AUCs ranging from 0.74 to 0.81) than in the NLST datasets (AUCs ranging from 0.61 to 0.73). The discriminative performance of most models was better for lung cancer mortality than for lung cancer incidence (i.e., the AUCs of most models were higher for lung cancer mortality than for lung cancer incidence) in all datasets, except for the PLCO control arm. The PLCOm2012 model (and its simplified version), the Bach model, and the TSCE incidence model showed the best discriminative performance across all datasets regardless of the type of predicted outcome. The discriminative performance of the models was similar for 5- and 6-year time frames, as shown in Figure S20 in the supplementary material of this Chapter.

Figure 2: Area under the receiver operating characteristic curve of the investigated risk models (with 95% confidence interval) in the National Lung Screening Trial computed tomography arm by predicted outcome

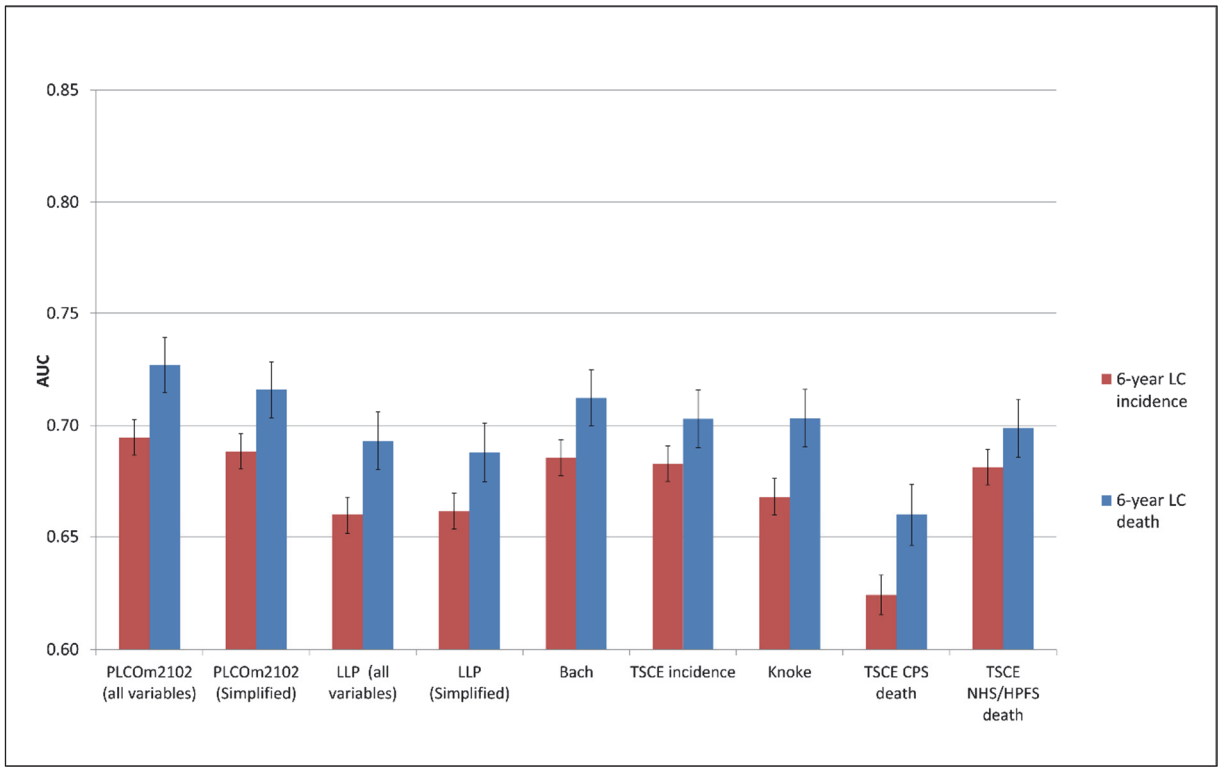


Figure 3: Area under the receiver operating characteristic curve of the investigated risk models (with 95% confidence interval) in the National Lung Screening Trial chest radiography arm by predicted outcome

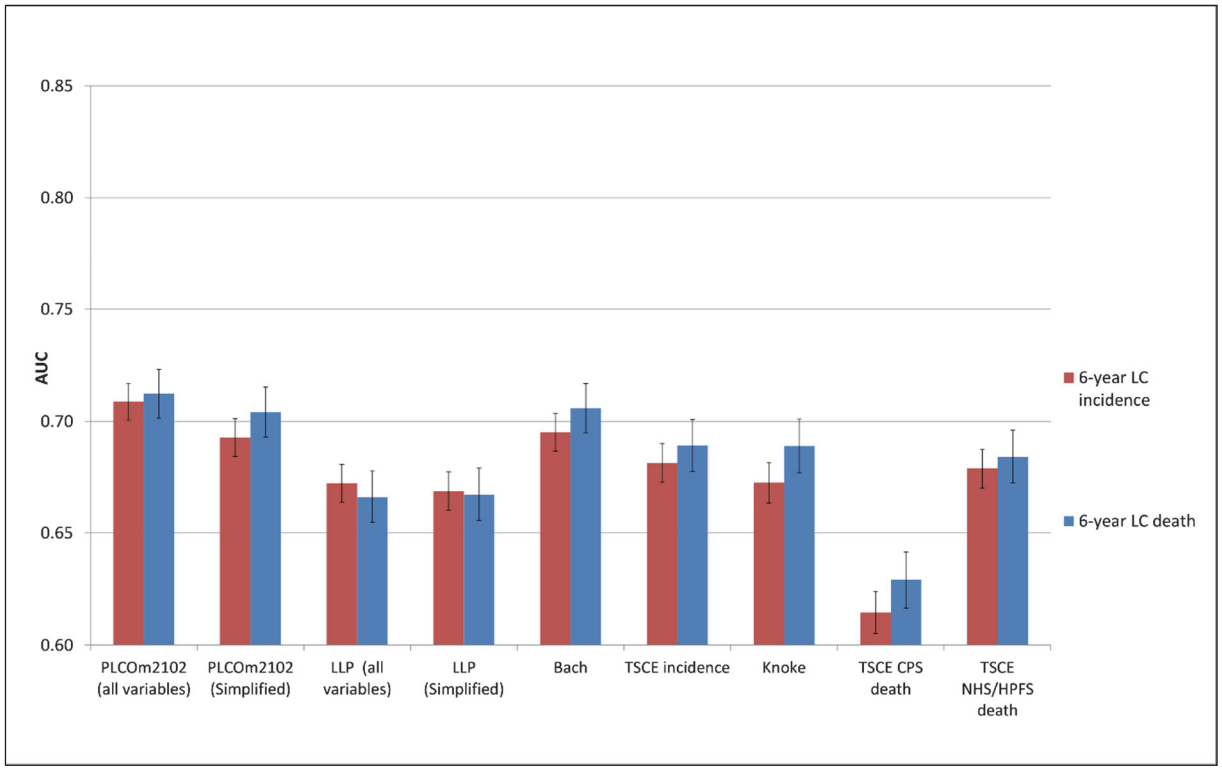


Figure 4: Area under the receiver operating characteristic curve of the investigated risk models (with 95% confidence interval) in the Prostate, Lung, Colorectal and Ovarian Cancer screening Trial chest radiography arm by predicted outcome

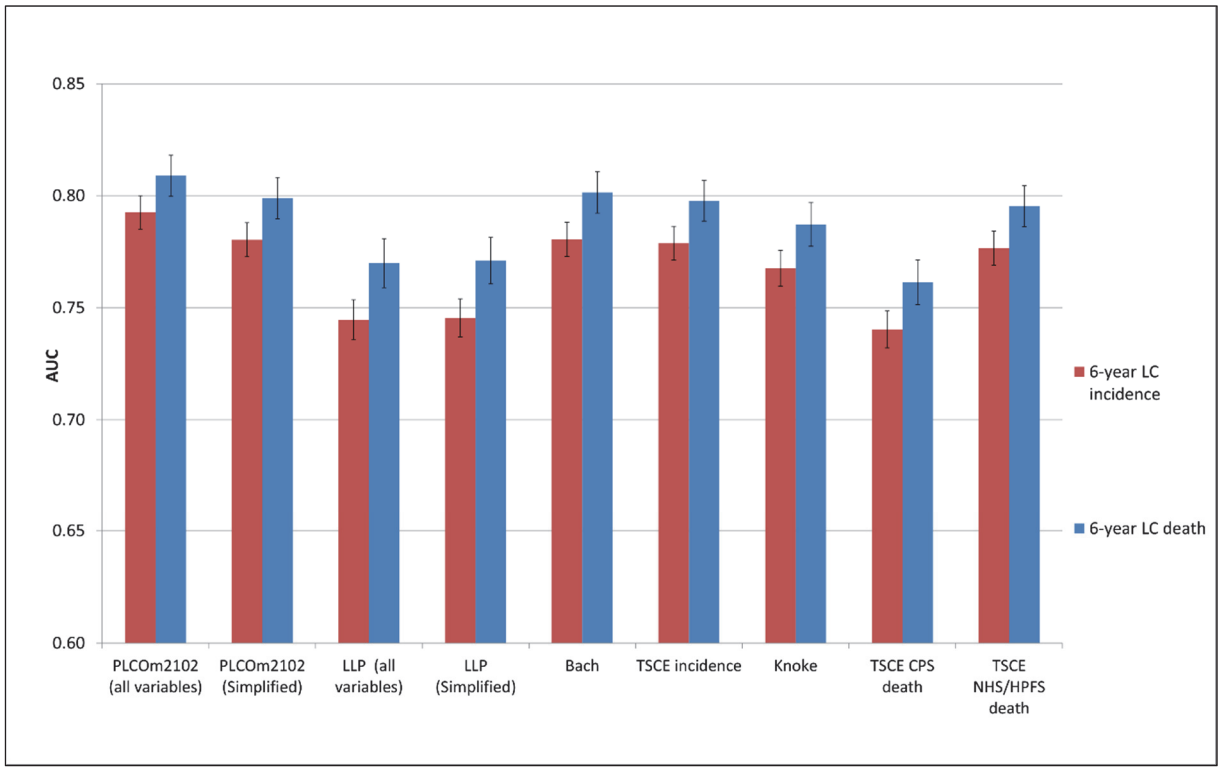
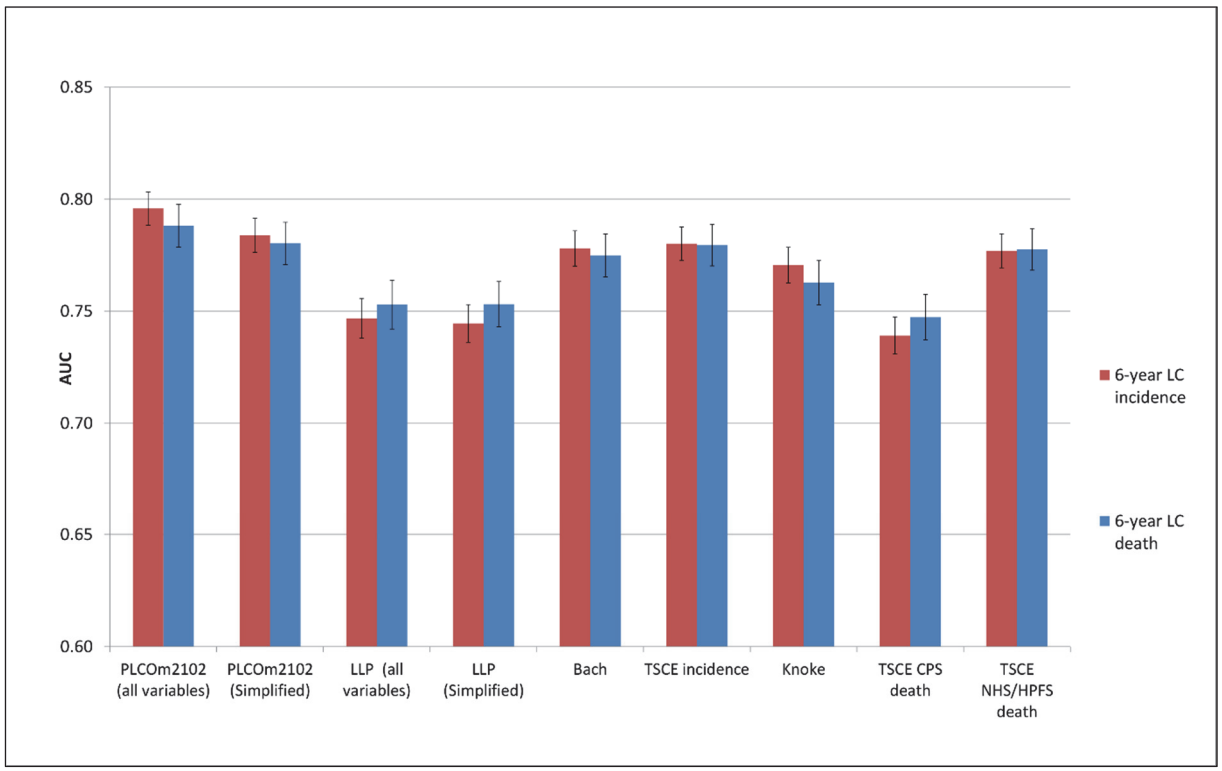


Figure 5: Area under the receiver operating characteristic curve of the investigated risk models (with 95% confidence interval) in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial control arm by predicted outcome



Clinical usefulness

Decision curve analysis for each risk prediction model provided a range of risk thresholds that yield a positive net benefit compared with the NLST eligibility criteria. Table 3 shows the lower and upper bounds for these ranges of risk thresholds for 6-year lung cancer incidence across all datasets. Overall, the lower and upper thresholds varied by model, but the ranges were roughly consistent across models, going from approximately 0.1% to 16.7%. This suggests that applying the models is useful for determining screening eligibility if missing one case of lung cancer that could be detected through screening is perceived as being between 999 and 5 times worse than unnecessarily screening one person. More detailed results for the decision curve analyses for both lung cancer incidence and mortality are shown in Figures S21-S37 and Table S12 in the supplementary material of this Chapter.

Comparison to National Lung Screening Trial eligibility criteria

Applying the NLST eligibility criteria yielded a sensitivity of 71.4% (95% confidence interval: 68.0-74.6%) and a specificity of 62.2% (95% confidence interval: 61.7-62.7%) for 6-year lung cancer incidence in the PLCO CXR arm (Figure 6; Table 4). The sensitivity and specificity of each of the risk prediction models were higher than those of the NLST eligibility criteria. The PLCOm2012 model, in particular, followed by the Bach model and the TSCE incidence model had the highest sensitivities (all three models >79.8%) and specificities (all three models >62.3%) among all evaluated models. Figure 6 also shows the risk thresholds for each model that select a similar number of individuals for screening as the NLST eligibility criteria. Similar results were found for the PLCO control arm and for using 6-year lung cancer death as the outcome measure (shown in Figures S38-S40 in the supplementary material of this Chapter).

Table 3: Lower and upper risk thresholds for which the risk prediction models have a positive net benefit compared with the National Lung Screening Trial criteria for 6-year lung cancer incidence

Model/dataset	NLST (computed tomography-arm)		NLST (chest radiography arm)		PLCO (chest radiography arm)		PLCO (control arm)	
	Lower risk threshold (WF*)	Upper risk threshold (WF*)	Lower risk threshold (WF*)	Upper risk threshold (WF*)	Lower risk threshold (WF*)	Upper risk threshold (WF*)	Lower risk threshold (WF*)	Upper risk threshold (WF*)
Bach model	0.5% (199.0)	12.7% (6.9)	0.9% (110.1)	13.5% (6.4)	0.3% (332.3)	10.4% (8.6)	0.2% (499.0)	8.9% (10.2)
LLP model	1.8% (54.6)	8.0% (11.5)	1.2% (82.3)	7.3% (12.7)	0.4% (249.0)	6.5% (14.4)	0.4% (249.0)	5.8% (16.2)
Simplified LLP model	1.9% (51.6)	8.5% (10.8)	1.9% (51.6)	8.5% (10.8)	0.4% (249.0)	8.5% (10.8)	0.3% (332.3)	5.3% (17.9)
PLCOm2012 model	0.9% (110.1)	16.1% (5.2)	0.1% (999.0)	10.5% (8.5)	0.2% (499.0)	9.0% (10.1)	0.1% (999.0)	11.0% (8.1)
Simplified PLCOm2012 model	0.7% (141.9)	13.6% (6.4)	0.7% (141.9)	12.0% (7.3)	0.3% (332.3)	9.3% (9.8)	0.2% (499.0)	8.5% (10.8)
TSCE lung cancer incidence model	2.0% (49.0)	12.1% (7.3)	1.1% (89.9)	8.0% (11.5)	0.3% (332.3)	7.9% (11.7)	0.2% (499.0)	6.9% (13.5)
Knoke model	3.5% (27.6)	13.3% (6.5)	2.8% (34.7)	8.8% (10.4)	3.0% (32.3)	7.7% (12.0)	2.9% (33.5)	7.2% (12.9)
TSCE CPS lung cancer death model	3.4% (28.4)	7.1% (13.1)	2.8% (34.7)	6.2% (15.1)	2.8% (34.7)	6.2% (15.1)	2.8% (34.7)	6.0% (15.7)
TSCE NHS/HPFS lung cancer death model	2.7% (36.0)	16.7% (5.0)	2.0% (49.0)	9.9% (9.1)	0.2% (499.0)	7.8% (11.8)	2.3% (42.5)	6.6% (14.2)

Abbreviations: CPS, Cancer Prevention Study; HPFS, Health Professionals Follow-up Study; LC, lung cancer; LLP, Liverpool Lung Project; NHS, Nurses' Health Study; NLST, National Lung Screening Trial; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; TSCE, Two-Stage Clonal Expansion; WF, weighting factor.

**Weighting factor corresponding to the risk threshold, i.e., the ratio of how much worse missing one case of lung cancer that could be detected through screening is valued compared to unnecessarily screening one person.*

Figure 6: Sensitivity, specificity, and risk thresholds for the investigated risk prediction models and the National Lung Screening Trial criteria for 6-year lung cancer incidence in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial chest radiography arm

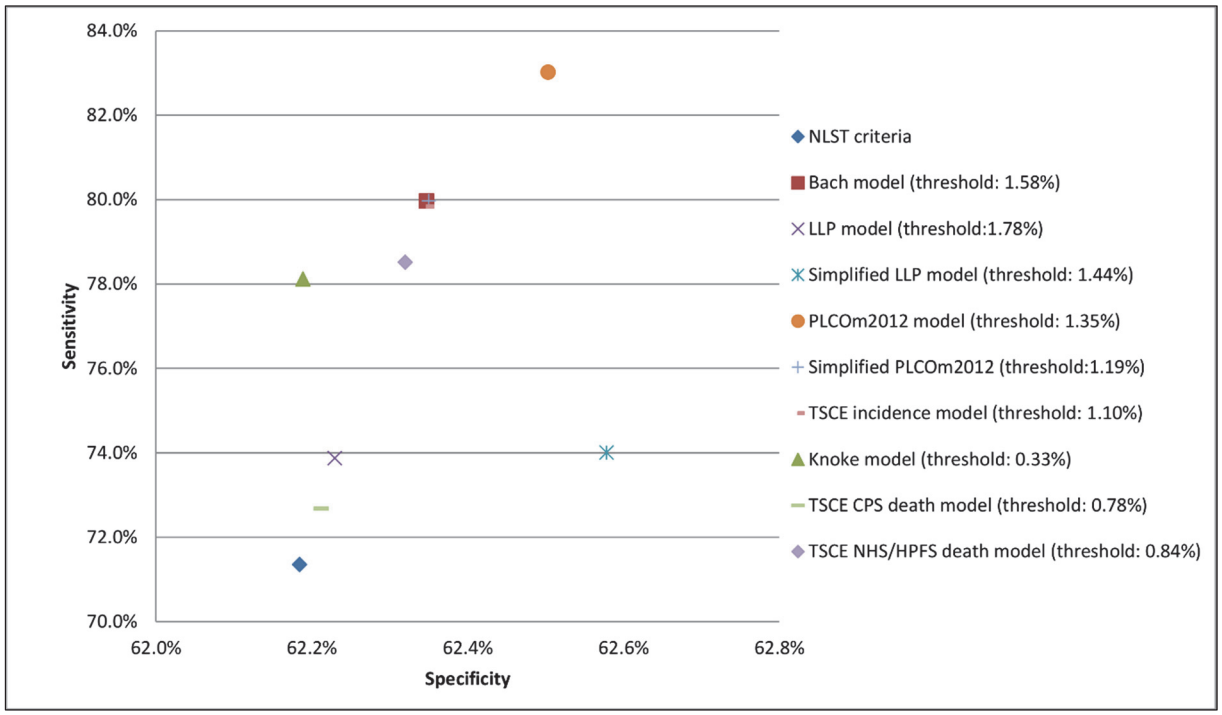


Table 4: Sensitivities and specificities corresponding to the suggested risk thresholds for the investigated risk prediction models and the National Lung Screening Trial criteria for 6-year lung cancer incidence in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial chest radiography arm

	NLST criteria	Bach model	LLP model	Simplified LLP model	PLCOm2012 model	Simplified PLCOm2012 model	TSCE lung cancer incidence model	Knoke model	TSCE CPS lung cancer death model	TSCE NHS/HPFS lung cancer death model
Sensitivity	71.4% (68.0-74.6%)	80.0% (76.9-82.8%)	73.9% (70.6-77.0%)	74.0% (70.7-77.1%)	83.0% (80.2-85.6%)	80.0% (76.9-82.8%)	79.8% (76.8-82.6%)	78.1% (75.0-81.0%)	72.7% (69.3-75.8%)	78.5% (75.4-81.4%)
Specificity	62.2% (61.7-62.7%)	62.3% (61.9-62.8%)	62.2% (61.8-62.7%)	62.6% (62.1-63.1%)	62.5% (61.9-62.8%)	62.4% (61.9-62.8%)	62.3% (61.9-62.8%)	62.2% (61.7-62.7%)	62.2% (61.7-62.7%)	62.3% (61.8-62.8%)

Abbreviations: CPS, Cancer Prevention Study; HPFS, Health Professionals Follow-up Study; LLP, Liverpool Lung Project; NHS, Nurses' Health Study; NLST, National Lung Screening Trial; TSCE, Two-Stage Clonal Expansion. Data given as percent (95% confidence interval).

Discussion

This study assessed the performance of nine lung cancer risk prediction models in two large randomized controlled trials: the NLST and the PLCO. The models had satisfactory calibration, had modest to good discrimination, and provided a substantial range of risk thresholds with a positive net benefit compared with the NLST eligibility criteria. Given appropriate model-specific risk thresholds, all risk prediction models had a better sensitivity and specificity than the NLST eligibility criteria. This implies that lung cancer risk prediction models, when coupled with model-specific risk thresholds, outperform currently recommended lung cancer screening eligibility criteria (Tables 3 and 4; Figure 6).

The risk prediction models considered in this study were developed in various cohorts for different outcome measures (lung cancer incidence versus mortality), with fundamental differences in model structures. Consequently, the absolute risk estimates differed between models, which led to differences in calibration performance between the models, specifically in the NLST cohorts. In addition, there were clear differences in discriminative ability between the models. The discriminative ability of all models was better in the PLCO cohorts than in the NLST cohorts, which may be caused by the higher heterogeneity in risk factor profiles among individuals in the PLCO compared with the NLST.[67,68] The NLST required individuals to have smoked at least 30 pack-years and included only current and former smokers (who quit less than 15 years ago), whereas the PLCO did not have any criteria for enrollment with regards to smoking history. In line with these criteria, the average NLST participant had a higher lung cancer risk than the average PLCO participant. The results of our investigation suggest that the discriminative ability of the evaluated models may be lower in groups at elevated risk, which may be due to the lower heterogeneity in risk among participants in these groups.[67,68]

However, randomized clinical trials suggest that the results of CT screening may provide an opportunity to improve risk stratification in these groups. In the NLST, participants with a negative prevalence screen had a substantially lower risk of developing lung cancer than participants with a positive prevalence screen.[69] Similarly, in the NELSON trial, the 2-year probability of developing lung cancer after a CT screen varied substantially by pulmonary

nodule size and the volume doubling time of these pulmonary nodules.[8] Therefore, incorporating the results of CT screening could improve the risk stratification in groups of individuals at elevated risk. Finally, while there was little difference in specificity between the models at risk thresholds similar to the NLST eligibility criteria, there was a clear difference in sensitivity. In particular, the PLCOm2012 model, followed by the Bach model and the TSCE incidence model, had the best performance across all aspects investigated in this study.

Previous studies have also compared the performance of different lung cancer risk prediction models.[20,21] D'Amelio et al. examined the discriminatory performance of three risk prediction models for lung cancer incidence in a case-control study and found modest differences between the models.[20] However, this study considered a limited number of participants (1,066 cases and 677 controls) and did not consider other aspects of model performance such as calibration or clinical usefulness. Li et al. examined four risk prediction models for lung cancer incidence in German participants of the European Prospective Investigation into Cancer and Nutrition cohort.[21] They found that while the differences between most of the evaluated models were modest, generally only the Bach and the PLCOm2012 models had similar or better sensitivity and specificity compared to the eligibility criteria used in the NLST and other eligibility criteria that were used in various European lung cancer screening trials (which applied less restrictive smoking eligibility criteria than the NLST). This cohort consisted of 20,700 individuals, but fewer than 100 lung cancer cases occurred, which limits statistical power for external validation.[18,19] In contrast to these previous studies, we performed a comprehensive validation, including aspects of calibration, discriminative ability, and clinical usefulness, for many models, in a large sample (n = 134,124) with 3,388 lung cancer cases and 1,799 lung cancer deaths. In addition, while our study supports earlier findings that risk prediction models outperform the NLST eligibility criteria, it also suggests that the PLCOm2012 model followed by the Bach and TSCE incidence models perform better than other models in all investigated aspects.

Our study has some limitations. While our results provide indications regarding at which risk thresholds the investigated risk models can be clinically useful, the optimal thresholds to apply remain uncertain. Determining optimal thresholds requires information on the long-

term benefits (such as life-years gained and mortality reduction) and harms (such as overdiagnosis) of applying these thresholds.[60] Natural history modeling may provide further information on the trade-off between the long-term benefits and harms for screening programs with different risk thresholds, similarly to how our previous study informed the USPSTF on its recommendations for lung cancer screening.[2]

Another limitation is that information on some of the predictor variables included in the evaluated risk prediction models was not available in the NLST and the PLCO, e.g., asbestos exposure was missing in both cohorts. However, only a few variables were unavailable. Furthermore, some of the evaluated models that used only age, gender, and smoking behavior, such as the TSCE models and the Knoke model, performed similarly to the other models that used additional information on risk factors, suggesting that age, gender, and smoking behavior are the most important risk factors for lung cancer. Thus, the improved performance of these models over the NLST eligibility criteria may primarily be due to the inclusion of detailed smoking behavior in these models. The NLST eligibility criteria use a dichotomized criterion for accumulated pack-years, e.g., an exposure of at least 30 pack-years, which leads to a loss of information for continuous variables.[70] Furthermore, pack-years are estimated by smoking duration and intensity (cigarettes per day), and previous studies indicate that both components contribute independently to an individual's risk for developing lung cancer; an aggregation of both may not fully capture the effects of smoking on lung cancer risk.[10,43,71] We chose to evaluate the models for varying follow-up lengths (5- and 6-year time frames) to investigate the effect of follow-up duration on the discrimination performance of each model.[54] Although the discriminative performance of the models was similar for 5- and 6-year time frames (shown in Figures S38-S40 in the supplementary material of this Chapter), this may not be the case for more disparate time frames.

A number of pertinent questions remain with regards to the implementation of lung cancer screening.[9] Current guidelines like the USPSTF recommendations suggest that individuals should be asked, at a minimum, about their age and smoking history.[3] A number of the models evaluated in our study use information on additional risk factors, such as personal history of cancer, which could be a potential barrier for implementing lung cancer screening based on risk prediction models. However, the LLP and PLCOm2012 models were

successfully used to recruit individuals for the UK Lung Cancer Screening Trial (UKLS) and the Pan-Canadian Early Detection of Lung Cancer Study (PanCan), respectively, through short questionnaires.[33,72] This suggests that acquiring information on the risk factors required for these models does not pose a major barrier for implementation. Furthermore, for some risk models, such as the Bach and PLCOm2012 models, online calculators are available, which provide opportunities for fast risk estimation in clinical practice.[73-76] For example, the PLCOm2012 model has been embedded in a lung cancer screening decision aid that has been widely adopted and that can be used to satisfy the Centers for Medicare & Medicaid Services reimbursement requirement for shared decision making.[75-77]

In conclusion, our study suggests that lung cancer screening selection criteria can be improved through the explicit application of risk prediction models rather than using criteria based on age and pack-years as a summary measure of smoking exposure. These models might also be helpful for improving the shared decision-making process for lung cancer screening recommended by the USPSTF and required in the U.S. by the Centers for Medicare & Medicaid Services.[3,75,78] However, recommendations for the implementation of risk-based lung cancer screening require a thorough evaluation of the benefits and harms of risk-based screening, as well as an assessment of the feasibility of implementing strategies based on risk models. Therefore, future studies need to evaluate the long-term benefits and harms of applying risk prediction models at different risk thresholds, while considering the potential challenges for implementation, and compare these with the expected benefits and harms of current guidelines.

Acknowledgments

We thank the National Cancer Institute (NCI) for access to NCI's data collected by the NLST and the PLCO. We thank the staff of Information Management Services for assistance with the harmonization of the NLST and PLCO datasets. Finally, we would like to thank the NLST and PLCO study participants for their contributions to these studies. The contents of the manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the NCI authors and do not represent or imply concurrence or endorsement by NCI.

References

1. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
2. 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. *Annals of Internal Medicine* 2014; **160**(5): 311-20.
3. Moyer VA, on behalf of the USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 2014; **160**(5): 330-8.
4. 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. *Annals of Internal Medicine* 2013; **159**(6): 411-20.
5. Roberts H, Walker-Dilks C, Sivjee K, et al. Screening High-Risk Populations for Lung Cancer: Guideline Recommendations. *Journal of Thoracic Oncology* 2013; **8**(10): 1232-7.
6. Tammemägi M, Hader J, Yu M, et al. P1.03-059 Organized High Risk Lung Cancer Screening in Ontario, Canada: A Multi-Centre Prospective Evaluation. *Journal of Thoracic Oncology*; **12**(1): S579.
7. Field JK, Hansell DM, Duffy SW, Baldwin DR. CT screening for lung cancer: countdown to implementation. *The Lancet Oncology* 2013; **14**(13): e591-e600.
8. Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *The Lancet Oncology* 2014; **15**(12): 1332-41.
9. van der Aalst CM, ten Haaf K, de Koning HJ. Lung cancer screening: latest developments and unanswered questions. *The Lancet Respiratory Medicine* 2016; **4**(9): 749-61.
10. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.
11. Kovalchik SA, Tammemägi M, Berg CD, et al. Targeting of Low-Dose CT Screening According to the Risk of Lung-Cancer Death. *New England Journal of Medicine* 2013; **369**(3): 245-54.
12. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and validation of risk models to select ever-smokers for ct lung cancer screening. *Journal of the American Medical Association* 2016; **315**(21): 2300-11.
13. Tammemägi MC, Church TR, Hocking WG, et al. Evaluation of the Lung Cancer Risks at Which to Screen Ever- and Never-Smokers: Screening Rules Applied to the PLCO and NLST Cohorts. *PLOS Medicine* 2014; **11**(12): e1001764.
14. Jaklitsch MT, Jacobson FL, Austin JHM, et al. The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. *The Journal of Thoracic and Cardiovascular Surgery* 2012; **144**(1): 33-8.
15. Collins GS, Moons KGM. Comparing risk prediction models. *British Medical Journal* 2012; **344**.
16. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *British Medical Journal* 2009; **338**.
17. Siontis GCM, Tzoulaki I, Siontis KC, Ioannidis JPA. Comparisons of established risk prediction models for cardiovascular disease: systematic review. *British Medical Journal* 2012; **344**.
18. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology* 2005; **58**(5): 475-83.
19. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Statistics in Medicine* 2016; **35**(2): 214-26.
20. D'Amelio AM, Jr., Cassidy A, Asomaning K, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. *British Journal of Cancer* 2010; **103**(3): 423-9.

21. Li K, Hüsing A, Sookthai D, et al. Selecting High-Risk Individuals for Lung Cancer Screening: A Prospective Evaluation of Existing Risk Models and Eligibility Criteria in the German EPIC Cohort. *Cancer Prevention Research* 2015; **8**(9): 777.
22. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and Study Design. *Radiology* 2011; **258**(1): 243-53.
23. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
24. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Controled Clinical Trials* 2000; **21**(6 Suppl): 273S-309S.
25. ten Haaf K, de Koning HJ. Should Never-Smokers at Increased Risk for Lung Cancer Be Screened? *Journal of Thoracic Oncology* 2015; **10**(9): 1285-91.
26. Etzel CJ, Kachroo S, Liu M, et al. Development and Validation of a Lung Cancer Risk Prediction Model for African-Americans. *Cancer Prevention Research* 2008; **1**(4): 255.
27. Li H, Yang L, Zhao X, et al. Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model. *BMC Medical Genetics* 2012; **13**(1): 118.
28. Park S, Nam B-H, Yang H-R, et al. Individualized Risk Prediction Model for Lung Cancer in Korean Men. *PLOS ONE* 2013; **8**(2): e54823.
29. El-Zein RA, Lopez MS, Amelio AM, et al. The Cytokinesis-Blocked Micronucleus Assay as a Strong Predictor of Lung Cancer: Extension of a Lung Cancer Risk Prediction Model. *Cancer Epidemiology Biomarkers & Prevention* 2014; **23**(11): 2462.
30. Maisonneuve P, Bagnardi V, Bellomi M, et al. Lung Cancer Risk Prediction to Select Smokers for Screening CT—a Model Based on the Italian COSMOS Trial. *Cancer Prevention Research* 2011; **4**(11): 1778.
31. Spitz MR, Etzel CJ, Dong Q, et al. An Expanded Risk Prediction Model for Lung Cancer. *Cancer Prevention Research* 2008; **1**(4): 250.
32. Hoggart C, Brennan P, Tjonneland A, et al. A Risk Model for Lung Cancer Incidence. *Cancer Prevention Research* 2012; **5**(6): 834.
33. McWilliams A, Tammemägi MC, Mayo JR, et al. Probability of Cancer in Pulmonary Nodules Detected on First Screening CT. *New England Journal of Medicine* 2013; **369**(10): 910-9.
34. Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. *British Journal of General Practice* 2011; **61**(592): e715.
35. Iyen-Omofoman B, Tata LJ, Baldwin DR, Smith CJP, Hubbard RB. Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. *Thorax* 2013; **68**(5): 451.
36. Young RP, Hopkins RJ, Hay BA, et al. A gene-based risk score for lung cancer susceptibility in smokers and ex-smokers. *Postgraduate Medical Journal* 2009; **85**(1008): 515.
37. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *Journal of the National Cancer Institute* 2007; **99**(9): 715-26.
38. 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. *Annals of Internal Medicine* 2012; **157**(4): 242-50.
39. Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *Journal of the National Cancer Institute* 2003; **95**(6): 470-8.
40. Knoke JD, Burns DM, Thun MJ. The change in excess risk of lung cancer attributable to smoking following smoking cessation: an examination of different analytic approaches using CPS-I data. *Cancer Causes & Control* 2008; **19**(2): 207-19.
41. Hazelton WD, Clements MS, Moolgavkar SH. Multistage Carcinogenesis and Lung Cancer Mortality in Three Cohorts. *Cancer Epidemiology Biomarkers & Prevention* 2005; **14**(5): 1171.
42. Hazelton WD, Jeon J, Meza R, Moolgavkar SH. Chapter 8: The FHCRC Lung Cancer Model. *Risk Analysis* 2012; **32**: S99-S116.

43. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.
44. Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *British Journal of Cancer* 2007; **98**(2): 270-6.
45. Gorelick MH. Bias arising from missing data in predictive models. *Journal of Clinical Epidemiology* 2006; **59**(10): 1115-23.
46. Aberle DR, Adams AM, Berg CD, et al. Baseline Characteristics of Participants in the Randomized National Lung Screening Trial. *Journal of the National Cancer Institute* 2010; **102**(23): 1771-9.
47. Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, et al., editors. SEER Cancer Statistics Review, 1975-2012. Bethesda (Maryland): National Cancer Institute; 2015 Apr [cited 2017 Mar 1]. http://seer.cancer.gov/csr/1975_2012/.
48. National Cancer Institute. Cancer stat facts: lung and bronchus cancer. Bethesda (Maryland): National Cancer Institute; 2016 [cited 2016 Mar 4]. http://seer.cancer.gov/statfacts/html/lungb.html.
49. Jolani S, Debray TP, Koffijberg H, van Buuren S, Moons KG. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Stat Med* 2015; **34**(11): 1841-63.
50. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European Heart Journal* 2014; **35**(29): 1925-31.
51. Etzioni R, Gulati R, Mallinger L, Mandelblatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Annals of Internal Medicine* 2013; **158**(11): 831-8.
52. Patz EF, Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Internal Medicine* 2014; **174**(2): 269-74.
53. ten Haaf K, de Koning HJ. Overdiagnosis in lung cancer screening: why modeling is essential. *Journal of Epidemiology and Community Health* 2015; **69**(11): 1035-39.
54. Vickers AJ, Cronin AM. Everything You Always Wanted to Know About Evaluating Prediction Models (But Were Too Afraid to Ask). *Urology* 2010; **76**(6): 1298-301.
55. R Core Team. R: a language and environment for statistical computing. Version 3.3.0. Vienna: R Foundation for Statistical Computing; 2016.
56. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *British Medical Journal* 2009; **338**.
57. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of Clinical Epidemiology*; **74**: 167-76.
58. Harrell FE Jr. rms: regression modeling strategies. Version 4.3-1. Comprehensive R Archive Network; 2014 [cited 2017 Mar 1]. http://CRAN.R-project.org/package=rms.
59. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2009.
60. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making* 2006; **26**(6): 565-74.
61. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Medical Informatics and Decision Making* 2008; **8**(1): 53.
62. van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011; **45**(3): 1-67.
63. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
64. Moons KM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): Explanation and elaboration. *Annals of Internal Medicine* 2015; **162**(1): W1-W73.

65. van der Heijden GJMG, T. Donders AR, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. *Journal of Clinical Epidemiology 2006*; **59**(10): 1102-9.
66. Steyerberg EW, Eijkemans MJC, Boersma E, Habbema JDF. Equally valid models gave divergent predictions for mortality in acute myocardial infarction patients in a comparison of logical regression models. *Journal of Clinical Epidemiology 2005*; **58**(4): 383-90.
67. Vergouwe Y, Moons KGM, Steyerberg EW. External Validity of Risk Models: Use of Benchmark Values to Disentangle a Case-Mix Effect From Incorrect Coefficients. *American Journal of Epidemiology* 2010; **172**(8): 971-80.
68. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of Clinical Epidemiology 2015*; **68**(3): 279-89.
69. Patz Jr EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomized, multicentre, diagnostic screening trial. *The Lancet Oncology* 2016; **17**(5): 590-9.
70. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 2006; **25**(1): 127-41.
71. Thun MJ, Carter BD, Feskanich D, et al. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine* 2013; **368**(4): 351-64.
72. McRonald FE, Yadegarfar G, Baldwin DR, et al. The UK Lung Screen (UKLS): Demographic Profile of First 88,897 Approaches Provides Recommendations for Population Screening. *Cancer Prevention Research* 2014; **7**(3): 362.
73. Brock University. Lung cancer risk calculators. St. Catharines (Ontario): Brock University; 2016 [cited 2016 Dec 19]. https://brocku.ca/lung-cancer-risk-calculator.
74. Memorial Sloan Kettering Cancer Center. Lung cancer screening decision tool. New York: Memorial Sloan Kettering Cancer Center; 2016 [cited 2016 Dec 19]. https://www.mskcc.org/cancer-care/types/lung/screening/lung-screening-decision-tool.
75. Lau YK, Caverly TJ, Cao P, et al. Evaluation of a Personalized, Web-Based Decision Aid for Lung Cancer Screening. *American Journal of Preventive Medicine* 2015; **49**(6): e125-9.
76. University of Michigan. Lung cancer CT screening [decision aid]. Ann Arbor: University of Michigan; 2016 [cited 2016 Dec 21]. http://www.shouldiscreen.com/.
77. American College of Radiology. Lung cancer screening resources. Reston (Virginia): American College of Radiology; 2016 [cited 2016 Dec 21]. https://www.acr.org/Quality-Safety/Resources/Lung-Imaging-Resources.
78. Jensen TS, Chin J, Ashby L, Hermansen J, Hutter JD. Decision memo for screening for lung cancer with low dose computed tomography (LDCT) (CAG-00439N). Baltimore: Centers for Medicare & Medicaid Services; 2015 [cited 2016 Aug 15]. https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274.

Chapter 10

Supplementary material

Lung cancer risk prediction model descriptions

Description of the Bach model

The Bach model was developed in the Carotene and Retinol Efficacy Trial (CARET) using 36,286 individuals (1,070 lung cancer cases).[1] The Bach model consists of two components, i.e. a model for lung cancer diagnosis and a model for death in the absence of lung cancer diagnosis, estimated through Cox proportional hazards regression, and together predict lung cancer incidence for a 1-year timeframe (as a binary outcome event). Applying the models iteratively allows for predictions over longer timeframes. The model predictors include: age, gender, asbestos exposure, smoking intensity (cigarettes per day), smoking duration and quit-time in former smokers. The component for estimating the one-year probability of death in the absence of lung cancer diagnosis is: $1 - S_0^{e(model)}$. With $S_0 = 0.9917663$ and $model$ being represented by the following equation, where CPD = cigarettes per day, SMK = duration of smoking, QUIT = duration of quitting, AGE = age, ASB = asbestos exposure, and GENDER = gender;

$$
\begin{aligned}
&-7.2036219 + (0.015490665 * \text{CPD}) - (0.00001737645 * (\text{CPD} - 15)^3) && \text{for all values CPD>15}\\
&\quad + (0.000021924149 * (\text{CPD} - 20.185718)^3) && \text{for all values CPD>20}\\
&\quad - (0.0000045476985 * (\text{CPD} - 40)^3) && \text{for all values CPD>40}\\
&+ (0.020041889 * \text{SMK}) + (0.0000065443781 * (\text{SMK} - 27.6577))^3 && \text{for all values SMK>27}\\
&\quad - (0.000013947696 * (\text{SMK} - 40)^3) && \text{for all values SMK>40}\\
&\quad + (0.0000074033175 * (\text{SMK} - 50.910335)^3) && \text{for all values SMK>50}\\
&- (0.023358962 * \text{QUIT}) + ((0.0019208669 * \text{QUIT})^3) && \text{for all values}\\
&\quad - (0.0020031611 * (\text{QUIT} - 0.50513347)^3) && \text{for all values QUIT>0}\\
&\quad + (0.000082294194 * (\text{QUIT} - 12.295688)^3) && \text{for all values QUIT>12}\\
&+ (0.099168033 * \text{AGE}) + (0.0000062174577 * (\text{AGE} - 53.459001))^3 && \text{for all values AGE>53}\\
&\quad - (0.000012115774 * (\text{AGE} - 61.954825)^3) && \text{for all values AGE>61}\\
&\quad + (0.0000058983164 * (\text{AGE} - 70.910335)^3) && \text{for all values AGE>70}\\
&\quad + (0.06084611) && \text{if ASB = yes}\\
&\quad - (0.49042298) && \text{if GENDER = female}
\end{aligned}
$$

The component for estimating the one-year probability diagnosis of lung cancer is:

$$1 - S_0^{e(model)}$$

With $S_0 = 0.99629$ and $model$ being represented by the following equation, where CPD = cigarettes per day, SMK = duration of smoking, QUIT = duration of quitting, AGE = age, ASB = asbestos exposure, and GENDER = gender;

$$
\begin{aligned}
&-9.7960571 + (0.060818386 * \text{CPD}) - (0.00014652216 * (\text{CPD} - 15)^3) && \text{for all values CPD>15}\\
&\qquad + (0.00018486938 * (\text{CPD} - 20.185718)^3) && \text{for all values CPD>20}\\
&\qquad - (0.000038347226 * (\text{CPD} - 40)^3) && \text{for all values CPD>40}\\
&+ (0.11425297 * \text{SMK} \quad + (0.000080091477 * (\text{SMK} - 27.6577))^3 && \text{for all values SMK>27}\\
&\qquad - (0.00017069483 * (\text{SMK} - 40)^3) && \text{for all values SMK>40}\\
&\qquad + (0.000090603358 * (\text{SMK} - 50.910335)^3) && \text{for all values SMK>50}\\
&- (0.085684793 * \text{QUIT} \quad + ((0.0065499693 * \text{QUIT})^3) && \text{for all values}\\
&\qquad - (0.0068305845 * (\text{QUIT} - 0.50513347)^3) && \text{for all values QUIT>0}\\
&\qquad + (0.00028061519 * (\text{QUIT} - 12.295688)^3) && \text{for all values QUIT>12}\\
&+ (0.070322812 * \text{AGE} \quad + (0.00009382122 * (\text{AGE} - 53.459001))^3 && \text{for all values AGE>53}\\
&\qquad - (0.00018282661 * (\text{AGE} - 61.954825)^3) && \text{for all values AGE>61}\\
&\qquad + (0.000089005389 * (\text{AGE} - 70.910335)^3) && \text{for all values AGE>70}\\
&\qquad + (0.2153936) && \text{if ASB = yes}\\
&\qquad - (0.0582726) && \text{if GENDER = female}
\end{aligned}
$$

The model has been externally by the authors in 6,239 smokers (with 333 lung cancer cases) from the placebo arm of the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study.[2]

Description of the Liverpool Lung Project (LLP) model

The Liverpool Lung Project (LLP) model was based on data from the Liverpool Lung Project case-control study.[3] The model was estimated through multivariate conditional logistic regression and predicts lung cancer incidence for a 5-year timeframe using 579 individuals with lung cancer and 1,157 age- and gender-matched population-based controls. The risk factors incorporated in the model are listed in Table S1, along with their log odds ratios and corresponding model coefficients.

As the model intercept could not be estimated using case-control data, the authors derived the age-group and gender-specific model intercepts through age- and gender-specific lung cancer incidence rates in the Liverpool area, as shown in Table S2.[3] The model intercept for an individual aged *x*+*y* years, where *x* is a multiple of 5 and *y* is 0, 1, 2, 3 or 4, can be calculated as follows:

$$Intercept_{Age\ x+y,sex} = \frac{(5-y-0.5)*Intercept_{Age\ x,sex}+(y+0.5)*Intercept_{Age\ x+5,sex}}{5}$$

The model has been externally validated twice by the authors: once in 1,066 cases and 677 controls treated at the Thoracic Surgery, Thoracic Oncology, or Pulmonary Units at the Massachusetts General Hospital (Boston, MA, USA) and once in 585 cases and 1,283 controls from the European Early Lung Cancer case–control study, 1,738 cases and 1,184 controls from the Harvard case–control study and 7,652 individuals (with 420 lung cancer cases) from the Liverpool Lung Project Cohort study.[4,5]

The simplified version of the LLP model uses the same parameter estimates as the original LLP model. However, when applying this model to a participant, it is assumed that only information on age and smoking history is known. Thus, the simplified model assumes that the participant had no prior diagnosis of pneumonia, no occupational exposure to asbestos, no prior diagnosis of a malignant tumor and no family history of lung cancer.

Table S1: Risk factors considered in the Liverpool Lung Project model

Risk factor	Log odds ratio	Model coefficient
Smoking duration		
Never	1.00 (reference)	0.000 (reference)
1-20 years	2.16	0.769
21-40 years	4.27	1.452
41-60 years	12.27	2.507
>60 years	15.25	2.724
Prior diagnosis of pneumonia		
No	1.00 (reference)	0.000 (reference)
Yes	1.83	0.602
Occupational exposure to asbestos		
No	1.00 (reference)	0.000 (reference)
Yes	1.89	0.634
Prior diagnosis of malignant tumor		
No	1.00 (reference)	0.000 (reference)
Yes	1.96	0.675
Family history of lung cancer		
No	1.00 (reference)	0.000 (reference)
Early onset (age < 60 years)	2.02	0.703
Late onset (age ≥ 60 years)	1.18	0.168

Table S2: Age- and gender-specific lung cancer incidence rates (per 100,000 person-years) in the Liverpool area (2002-2004)

	Men		**Women**	
Age-group	**Lung cancer incidence rate**	**Corresponding model intercept**	**Lung cancer incidence rate**	**Corresponding model intercept**
40-44	15.5	-9.06	5.97	-9.90
45-49	37.87	-8.16	37.34	-8.06
50-54	88.65	-7.31	68.14	-7.46
55-59	172.26	-6.63	175.24	-6.50
60-64	329.02	-5.97	230.60	-6.22
65-69	487.42	-5.56	288.06	-5.99
70-74	616.45	-5.31	464.99	-5.49
75-79	950.61	-4.83	594.19	-5.23
80-84	1096.42	-4.68	497.09	-5.42

Description of the PLCOm2012 model

The PLCOm2012 model was developed in the control arm of the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO), using 36,286 individuals (630 lung cancer cases).[6] The model was estimated through multivariate logistic regression and predicts lung cancer incidence for a 6-year timeframe. It was initially validated in 37,332 individuals (678 lung cancer cases) of the PLCO intervention arm (in which chest radiography screening occurred).

The model predictors include seven non-smoking variables: age, race/ethnicity, education (an estimator of socioeconomic circumstance), body mass index, personal history of cancer, family history of lung cancer and chronic obstructive pulmonary disease. The model includes four smoking variables: smoking status (former vs. current), smoking intensity (cigarettes per day), smoking duration and quit-time in former smokers. Using multivariable fractional polynomials, smoking intensity was shown to have a nonlinear relationship with lung cancer, and this nonlinear effect is incorporated into PLCOm2012. The risk factors incorporated in the model are listed in Table S3, along with their log odds ratios and corresponding model coefficients.

The initial predictive performance evaluation of the PLCOm2012 in the PLCO intervention arm demonstrated high discrimination (AUC = 0.80) and calibration (predicted probabilities / observed = 0.95). The model has also been externally validated by the authors in 51,033 (1,826 cases) participants of the National Lung Screening Trial.[6]

The simplified version of the PLCOm2012 model uses the same parameter estimates as the original PLCOm2012 model. However, similarly to the simplified LLP model, it is assumed that only information on age and smoking history is known. Thus, the simplified model assumes that the participant was white, had a body mass index of 27 (center value), some college education (center value), no chronic obstructive pulmonary disease, no personal history of cancer, and no family history of lung cancer.

Table S3: Risk factors considered in the PLCOm2012 model

Risk factor	Log odds ratio	Model coefficient
Age, per 1-year increase (centered on age 62)	1.081	0.0778868
Race or ethnic group (self-reported)		
White (non-Hispanic)	1.00 (reference)	0.000 (reference)
Black (non-Hispanic)	1.484	0.3944778
Hispanic	0.475	−0.7434744
Asian	0.627	−0.466585
Native Hawaiian or Pacific Islander	1.00	0.000
American Indian or Alaskan Native	2.793	1.027152
Education, per increase of 1 level. Education was centered on level 4*	0.922	−0.0812744
Body-mass index, per 1-unit increase (centered on 27)	0.973	−0.0274194
Chronic obstructive pulmonary disease		
No	1.00 (reference)	0.000 (reference)
Yes	1.427	0.3553063
Personal history of cancer		
No	1.00 (reference)	0.000 (reference)
Yes	1.582	0.4589971
Family history of lung cancer		
No	1.00 (reference)	0.000 (reference)
Yes	1.799	0.587185
Smoking status		
Former	1.00 (reference)	0.000 (reference)
Current	1.297	0.2597431
Smoking intensity**	This variable is nonlinear so no single odds ratio represents the entire association	−1.822606
Duration of smoking, per 1-year increase (centered on 27 years)	1.032	0.0317321
Smoking quit time, per 1-year increase (centered on 10 years)	0.970	−0.0308572
Model constant		−4.532506

** Education was measured in six ordinal levels: less than high-school graduate (level 1), high-school graduate (level 2), some training after high school (level 3), some college (level 4), college graduate (level 5), and postgraduate or professional degree (level 6).*

*** For smoking intensity, the contribution of the variable to the model should be calculated by dividing the number of cigarettes per day by 10, exponentiating by the power −1, centering by subtracting 0.4021541613, and multiplying this number by the beta coefficient of the variable.*

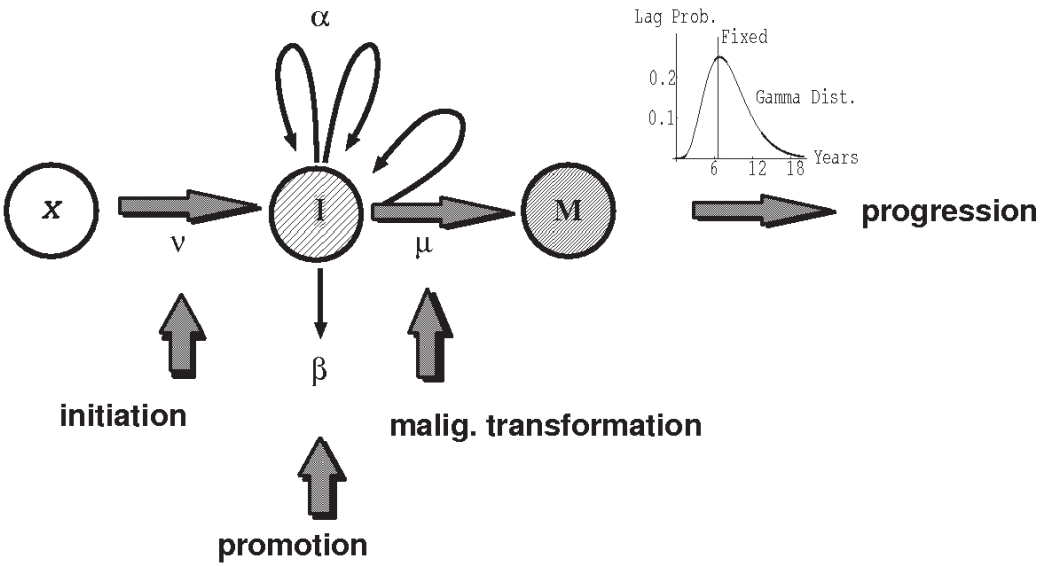
Description of the Two-Stage Clonal Expansion (TSCE) models

The Two-Stage Clonal Expansion (TSCE) model represents the process of carcinogenesis in three phases. In the first phase (initiation), a susceptible stem cell acquires one or more mutations resulting in an initiated cell, which has partially escaped growth control. In the second phase (promotion), initiated cells undergo clonal expansion, either spontaneously or in response to endogenous or exogenous promoters. Finally, in the third phase (malignant conversion), one of the initiated cells acquires further mutational changes leading to a malignant cell. The TSCE model assumes that normal stem cells become initiated according to a Poisson process with intensity "νX", where X is the number of normal stem cells (X) in the lungs and "ν" represents the initiation rate of normal stem cells. Once a cell is initiated, it undergoes a stochastic clonal expansion (promotion) with cell division rate "α" and cell death/differentiation rate "β". An initiated cell can also divide into one initiated and one malignant cell (malignant conversion) with rate "μ". A constant lag time, or lag time distribution, is commonly used to represent the time between the onset of the first malignant cell and lung cancer incidence or mortality depending on the model outcome (progression). Figure S1 shows the pictorial description of the TSCE model. To model the effects of smoking on lung cancer risk, the model initiation, promotion, and malignant transformation parameters are assumed to be altered during periods of smoking exposure through flexible dose-response relationships: $\theta(t) = \theta_0 \times \left(1 + \theta_1 \times d(t)^{\theta_2}\right)$, where θ represents identifiable biological parameters (i.e., combinations of ν, α, β, μ, and X), θ_0 the background rate, θ_1 the dose-response coefficient, θ_2 the non-linearity of the dose-response, and d(t) is smoking dose at time t. This dose-response relationship links the individual smoking history to the cell kinetic parameters in the TSCE model. The TSCE model was previously calibrated to lung cancer incidence or mortality data in several large prospective smoking cohorts: the Nurses' Health Study (NHS) for women, the Health Professionals Follow-up Study (HPFS) for men, the British doctors cohort, the American Cancer Society Cancer Prevention Studies I and II (CPS-I and CPS-II) cohorts.[7-9] Here we used the CPS-I, NHS, and HPFS versions of the model. Tables S4 and S5 present the structure and parameter values for each of the different TSCE model versions. To compute the 6-year probability to develop lung cancer (for incidence models) or die from lung cancer (for mortality models) given that the person survives by the age at the entry (say t_a), we assume that the smoking patterns at the entry remain the same for the next 6 years of follow-up. The conditional 6-year probability is computed by

$$\Pr(T \le t_a + 6 \mid T > t_a) = \frac{\Pr(t_a < T \le t_a + 6)}{\Pr(T > t_a)} = 1 - \frac{S(t_a + 6)}{S(t_a)},$$

where S(t) is the survival function from the TSCE model.[10]

Figure S1: Representation of the Two-Stage Clonal Expansion model



The Two-Stage Clonal Expansion (TSCE) model is a stochastic representation of the cell events during carcinogenesis. The carcinogenic process may be thought of as consisting of three phases: initiation, promotion, and malignant transformation. Normal stem cells (labeled X) may mutate at rate ν to create an initiated cell (labeled I). An initiated cell may divide at rate α, die or differentiate at rate β, and mutate at rate μ to create a malignant cell. A lag time, or lag time distribution, is used to represent the time from occurrence of the first malignant cell to lung cancer death.

Table S4: Model parameters for background rate and dose –response relationship

Background variables for CPS-I, NHS, and HPFS models	
$X=10^7$	Assume 10^7 normal stem cells in both lungs
α_0	Background cell division rate (per cell per year)
$g_0=\alpha_0-\beta_0-\mu_0$	Background net cell promotion rate (per cell per year)
$\nu_0=\mu_0$	Background initiation rate; Background malignant transformation rate (per cell per year)
t_{lag}**=5 years**	Fixed constant lag time
Dose-response variables	
CPS-I model	
$\nu_i=\nu_0(1+p_1)$**; p_1=0 for nonsmokers**	Initiation rate (per cell per year)
$g_i=g_0(1+p_2\times dose_i^{p3})$	Net initiated cell promotion rate (per cell per year)
$\alpha_i=\alpha_0(1+p_2\times dose_i^{p3})$	Initiated cell division rate (per cell per year)
$\mu_i=\mu_0$	Malignant transformation rate (per cell per year); No dose-response
NHS and HPFS models	
$\nu_i=\nu_0$	Initiation rate (per cell per year); No dose-response
$g_i=g_0(1+p_2\times dose_i^{p3})$	Net initiated cell promotion rate (per cell per year)
$\alpha_i=\alpha_0(1+p_2\times dose_i^{p3})$	Initiated cell division rate (per cell per year)
$\mu_i=\mu_0(1+p_4\times dose_i^{p5})$	Malignant transformation rate (per cell per year)

Table S5: Parameter estimates for CPS-I, NHS, and HPFS models

Model	α_0	g_0	ν_0 ($=\mu_0$)	p_1	p_2	p_3	p_4	p_5
Lung Cancer mortality models								
CPS-I females	71.56	0.086	8.93×10^{-8}	1.23	0.04	0.98	-	-
CPS-I males	22.65	0.075	1.40×10^{-7}	1.79	0.21	0.47	-	-
NHS	3.00	0.076	1.03×10^{-7}	-	0.20	0.50	0.05	0.60
HPFS	3.00	0.076	1.03×10^{-7}	-	0.33	0.35	0.21	0.18
Lung Cancer incidence models								
NHS	3.00	0.077	1.26×10^{-7}	-	0.17	0.53	0.10	0.62
HPFS	3.00	0.077	1.26×10^{-7}	-	0.26	0.39	0.33	0.25

Description of the Knoke model

The Knoke model was developed using lung cancer mortality data from the American Cancer Society Cancer Prevention Study I (CPS-I)[11] The model is specific for white males of ages 40-79. For never smokers, the absolute risk of death due to lung cancer was modeled as a two-parameter Poisson regression model on attained age in years:

$$R_{NS} = 9.21 * 10^{-13} * age^{4.6}$$

For continuing smokers, the excess risk of death due to lung cancer was modeled as a Poisson regression model with modified offset, assuming the mean value function to be a power function:

$$ER_S = 1.51 * 10^{-13} * age^{2.38} * age^{2.38} * CPD^{0.867} * dur^{2.87}$$

where *CPD* is the number of cigarettes per day, and *dur* the duration of smoking in years. The mean absolute risk of death due to lung cancer for continuing smokers is then given by

$$R_S = R_{NS} + ER_S$$

For former smokers, the absolute risk of death due to lung cancer was modeled as:

$$R_{FS} = R_{NS} + f(qt_{yrs}, qt_{age}) * ER_S$$

where R_{NS} is the plug-in absolute risk function for never smokers of the same age; ER_S is a plug-in excess risk function for continuing smokers of the same age, CPD, and duration; and *f* is a function of the decrease in excess risk for former smokers, which was assumed to be a non-increasing function of time in years since cessation (qt_{yrs}) and age in years at cessation (qt_{age}). A negative exponential function with a lag of two years was used to model the function *f*:

$$f(qt_{yrs}, qt_{age}) = exp[-(0.274 * qt_{age}) * (qt_{yrs} - 2)]$$

To compute the 6-year probability to die due to lung cancer in the follow-up given that the person hasn't died from cancer by the age at the entry (say t_a), we assume that the smoking patterns at age at entry remain the same for the next 6 years of follow-up. The conditional 6-year probability is computed by

$$Pr(T \leq t_a + 6 | T > t_a) = \frac{Pr(t_a < T \leq t_a + 6)}{Pr(T > t_a)} = 1 - \frac{S(t_a + 6)}{S(t_a)}$$

where S(t) is the survival function and computed by $exp\left[-\int_0^t R_i(u)du\right], i = NS, S, FS$ corresponding to never, continuing, former smokers.

Table S6: Characteristics of National Lung Screening Trial (NLST) and Prostate, Lung, Colorectal and Ovarian cancer screening trial (PLCO) participants according to 6-year lung cancer mortality

Characteristic	NLST (computed tomography-arm)			NLST (chest radiography arm)			PLCO (chest radiography arm)			PLCO (control arm)		
Number (percent) of participants												
Age (years)												
Median (IQR)	60 (57-65)	64 (60-68)	<0.0001	60 (57-65)	63 (59.5-68)	<0.0001	62 (58-66)	66 (61-69)	<0.0001	62 (58-66)	65 (61-69)	<0.0001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Gender												
Male	15,503 (58.89%)	266 (67.00%)	0.0012	15,465 (58.93%)	296 (60.78%)	0.4295	23,408 (58.29%)	295 (66.29%)	0.0007	22,906 (57.84%)	307 (65.32%)	0.0012
Female	10,822 (41.11%)	131 (33.00%)		10,778 (41.07%)	191 (39.22%)		16,747 (41.71%)	150 (33.71%)		16,696 (42.16%)	163 (34.68%)	
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Hispanic ethnicity												
No	25,689 (97.58%)	389 (97.98%)	0.0530	25,565 (97.42%)	474 (97.33%)	0.0725	38,418 (95.67%)	429 (96.40%)	0.8704	37,830 (95.53%)	457 (97.23%)	0.1113
Yes	477 (1.81%)	2 (0.50%)		453 (1.73%)	3 (0.62%)		874 (2.18%)	10 (2.25%)		877 (2.21%)	5 (1.06%)	
Missing	159 (0.60%)	6 (1.51%)		225 (0.86%)	10 (2.05%)		863 (2.15%)	6 (1.35%)		895 (2.26%)	8 (1.70%)	
Race or ethnic group												
White	23,593 (89.62%)	359 (90.43%)	0.0489	23,514 (89.60%)	435 (89.32%)	0.2235	35,449 (88.29%)	375 (84.27%)	0.0025	35,003 (88.39%)	414 (88.09%)	0.0440
Black	1,170 (4.44%)	17 (4.28%)		1,145 (4.36%)	29 (5.95%)		2,279 (5.68%)	46 (10.34%)		2,214 (5.59%)	40 (8.51%)	
Hispanic	343 (1.30%)	1 (0.25%)		316 (1.20%)	2 (0.41%)		812 (2.02%)	10 (2.25%)		810 (2.05%)	5 (1.06%)	
Asian	552 (2.10%)	7 (1.76%)		530 (2.02%)	6 (1.23%)		1239 (3.09%)	9 (2.02%)		1202 (3.04%)	10 (2.13%)	
Native Hawaiian or Pacific Islander	88 (0.33%)	3 (0.76%)		101 (0.38%)	1 (0.21%)		223 (0.56%)	4 (0.90%)		244 (0.62%)	1 (0.21%)	
American Indian or Alaskan Native	88 (0.33%)	4 (1.01%)		97 (0.37%)	1 (0.21%)		130 (0.32%)	1 (0.22%)		109 (0.28%)	0 (0%)	
Missing	491 (1.87%)	6 (1.51%)		540 (2.06%)	13 (2.67%)		23 (0.06%)	0 (0.0%)		20 (0.05%)	0 (0.0%)	
Education												
Less than high school graduate	1,597 (6.07%)	44 (11.08%)	<0.0001	1,563 (5.96%)	45 (9.24%)	<0.0001	3,470 (8.64%)	77 (17.30%)	<0.0001	3,445 (8.70%)	69 (14.68%)	<0.0001
High school graduate	6,161 (23.40%)	112 (28.21%)		6,299 (24.00%)	139 (28.54%)		8,860 (22.06%)	114 (25.62%)		8,704 (21.98%)	140 (29.79%)	
Post high school training	3,678 (13.97%)	55 (13.85%)		3,620 (13.79%)	81 (16.63%)		5,378 (13.39%)	55 (12.36%)		5,433 (13.72%)	58 (12.34%)	
Some college	6,099 (23.17%)	90 (22.67%)		5,982 (22.79%)	106 (21.77%)		9,336 (23.25%)	98 (22.02%)		9,223 (23.29%)	97 (20.64%)	
College graduate	4,455 (16.92%)	50 (12.59%)		4,385 (16.71%)	56 (11.50%)		6,667 (16.60%)	56 (12.58%)		6,421 (16.21%)	58 (12.34%)	
Postgraduate/professional	3,749 (14.24%)	31 (7.81%)		3,771 (14.37%)	49 (10.06%)		6,384 (15.90%)	45 (10.11%)		6,232 (15.74%)	45 (9.57%)	
Missing	586 (2.23%)	15 (3.78%)		623 (2.37%)	11 (2.26%)		60 (0.15%)	0 (0.0%		144 (0.36%)	3 (0.64%)	
BMI (kg/m²)												
Median (IQR)	27.32 (24.46-30.69)	26.22 (24.00-29.63	0.0002	27.38 (24.46-30.61)	25.90 (23.45-29.29)	<0.0001	26.68 (24.19-29.93)	26.16 (23.32-28.95)	0.0002	26,68 (24.18-29.89)	25.88 (23.44-28.94)	<0.0001
Missing	154 (0.58%)	5 (1.26%)		206 (0.78%)	7 (1.44%)		401 (1.00%)	3 (0.67%)		750 (1.89%)	7 (1.49%)	
COPD												
No	21,768 (82.69%)	280 (70.53%)	<0.0001	21,709 (82.72%)	369 (75.77%)	<0.0001	36,635 (91.23%)	348 (78.20%)	<0.0001	36,095 (91.14%)	371 (78.94%)	<0.0001
Yes	4,557 (17.31%)	117 (29.47%)		4,534 (17.28%)	118 (24.23%)		3,520 (8.77%)	97 (21.80%)		3,507 (8.86%)	99 (21.06%)	
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)
Emphysema												
No	24,217 (91.99%)	322 (81.11%)	<0.0001	24,088 (91.79%)	412 (84.60%)	<0.0001	38,287 (95.35%)	379 (85.17%)	<0.0001	37,607 (94.96%)	396 (84.26%)	<0.0001
Yes	1,985 (7.54%)	71 (17.88%)		1,967 (7.50%)	70 (14.37%)		1,681 (4.19%)	65 (14.61%)		1,641 (4.14%)	70 (14.89%)	

Missing	123 (0.47%)	4 (1.01%)		188 (0.72%)	5 (1.03%)		187 (0.47%)	1 (0.22%)		354 (0.89%_	4 (0.85%)	
Personal history of cancer												
No	25,183 (95.66%)	361 (90.93%)	0.0003	24,929 (94.99%)	458 (94.05%)	0.6584	38,321 (95.43%)	421 (94.61%)	0.4229	37,751 (95.33%)	434 (92.34%)	0.0042
Yes	1,064 (4.04%)	32 (8.06%)		1,188 (4.53%)	24 (4.53%)		1,834 (4.57%)	24 (5.39%)		1,851 (4.67%)	36 (7.66%)	
Missing	78 (0.30%)	4 (1.01%)		126 (0.48%)	5 (1.03%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Family history of lung cancer												
No	20,199 (76.73%)	288 (72.54%)	0.1224	20,103 (76.60%)	349 (71.66%)	0.0142	33,952 (84.55%)	331 (74.39%)	<0.0001	33,658 (84.99%)	368 (78.30%)	0.0006
Yes	5,717 (21.72%)	98 (24.69%)		5,678 (21.64%)	128 (26.28%)		4,568 (11.38%)	85 (19.10%)		4,467 (11.28%)	77 (16.38%)	
Missing	409 (1.55%)	11 (2.77%)		462 (1.76%)	10 (2.05%)		1,635 (4.07%)	29 (6.52%)		1,477 (3.73%)	25 (5.32%)	
Personal history of pneumonia												
No	20,402 (77.50%)	284 (71.54%)	0.0143	20,319 (77.43%)	361 (74.13%)	0.1209	**Not measured**			**Not measured**		
Yes	5,822 (22.12%)	108 (27.20%)		5,758 (21.94%)	121 (24.85%)		**Not measured**			**Not measured**		
Missing	101 (0.38%)	5 (1.26%)		166 (0.63%)	5 (1.03%)		**Not measured**			**Not measured**		
Smoking status												
Current smoker	13,794 (52.40%)	144 (36.27%)	<0.0001	13,736 (52.34%)	162 (33.26%)	<0.0001	32,285 (80.40%)	239 (53.71%)	<0.0001	31,824 (80.36%)	269 (57.23%)	<0.0001
Former smoker	12,531 (47.60%)	253 (63.73%)		12,507 (47.66%)	325 (66.74%)		7,870 (19.60%)	206 (46.29%)		7,778 (19.64%)	201 (42.77%)	
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)	
Smoking duration (years)												
Median (IQR)	40 (35-45)	45 (40-50)	<0.0001	40 (35-45)	44 (40-49)	<0.0001	28 (16-39)	43 (37-48)	<0.0001	28 (16-39)	42 (34-47)	<0.0001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		767 (1.91%)	9 (2.02%)		881 (2.22%)	13 (2.77%)	
Smoking intensity (cigarettes per day)												
Median (IQR)	25 (20-35)	30 (20-40)	0.0006	25 (20-32)	30 (20-40)	0.0029	20 (10-30)	30 (20-40)	<0.0001	20 (10-30)	30 (20-40)	<0.0001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		81 (0.20%)	1 (0.22%)		113 (0.29%)	1 (0.21%)	
Pack-years of smoking												
Median (IQR)	48 (39-66)	58 (46-86)	<0.0001	48 (39-66)	57 (45.5-80)	<0.0001	28.5 (14-48)	52 (40-75)	<0.0001	29 (14-49.5)	53 (40-75)	<0.0001
Missing	0 (0%)	0 (0%)		0 (0%)	0 (0%)		830 (2.1%)	10 (2.2%)		962 (2.4%)	14 (3.0%)	
Years since cessation												
Median (IQR)	7 (3-11)	5 (2-10)	0.0002	7 (3-11)	6 (2-11)	0.2844	20 (10-30)	9 (3-18.75)	<0.0001	20 (10-30)	10 (4-19)	<0.0001
Missing	223 (0.8%)	1 (0.3%)		220 (0.8%)	4 (0.8%)		561 (1.4%)	5 (1.1%)		680 (1.7%)	4 (0.9%)	

Abbreviations: BMI, body mass index; COPD, chronic obstructive pulmonary disease; IQR, interquartile range; NLST, National Lung Screening Trial; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

Data are given as n (percent) or median (IQR).

Table S7: Proportions of National Lung Screening Trial (NLST) and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) participants with complete information for all considered risk factors

6-year lung cancer incidence	All participants	All participants with complete information	Participants who were not diagnosed with lung cancer	Participants who were not diagnosed with lung cancer with complete information	Participants who were diagnosed with lung cancer	Participants who were diagnosed with lung cancer with complete information
NLST-CT	26,722	25,032 (93.68%)	25,692	24,076 (93.71%)	1,030	956 (92.82%)
NLST-chest radiography	26,730	24,998 (93.52%)	25,835	24,155 (93.50%)	895	843 (94.19%)
PLCO- chest radiography	40,600	37,657 (92.75%)	39,846	36,974 (92.79%)	754	683 (90.58%)
PLCO-Control	40,072	36,951 (92.21%)	39,363	36,313 (92.25%)	709	638 (89.99%)
6-year lung cancer mortality	**All participants**	**All participants with complete information**	**Participants who were not diagnosed with lung cancer**	**Participants who were not diagnosed with lung cancer with complete information**	**Participants who were diagnosed with lung cancer**	**Participants who were diagnosed with lung cancer with complete information**
NLST-CT	26,722	25,032 (93.68%)	26,325	24,661 (93.68%)	397	371 (93.45%)
NLST- chest radiography	26,730	24,998 (93.52%)	26,243	24,540 (93.51%)	487	458 (94.05%)
PLCO- chest radiography	40,600	37,657 (92.75%)	40,155	37,254 (92.78%)	445	403 (90.56%)
PLCO-Control	40,072	36,951 (92.21%)	39,602	36,527 (92.24%)	470	424 (90.21%)

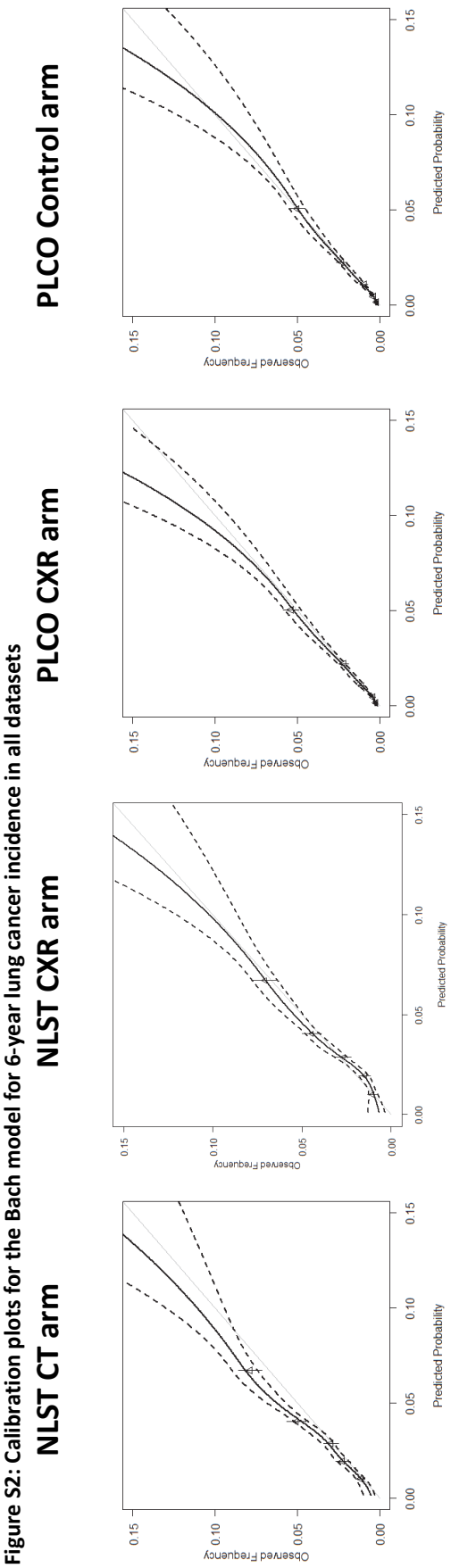


Figure S2: Calibration plots for the Bach model for 6-year lung cancer incidence in all datasets

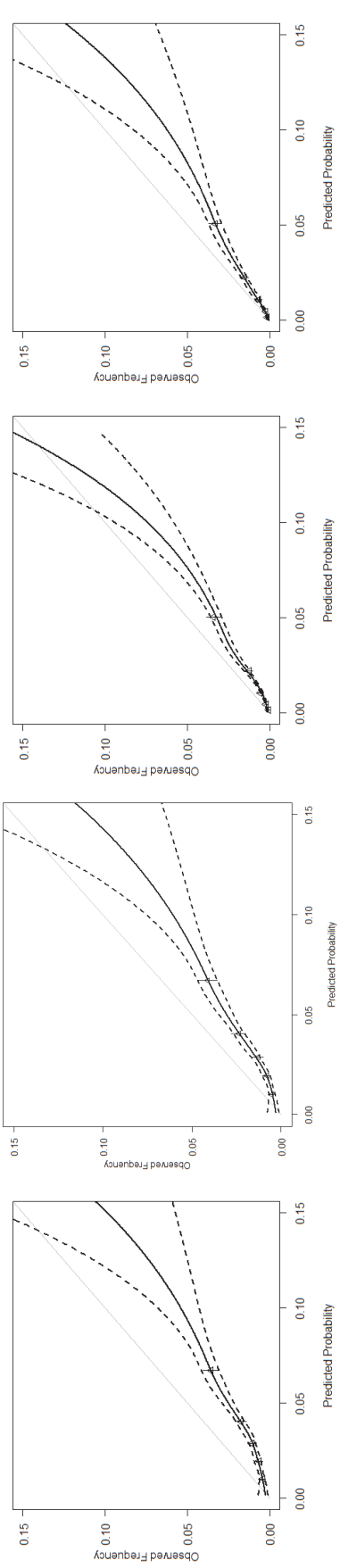


Figure S3: Calibration plots for the Bach model for 6-year lung cancer mortality in all datasets

Figure S4: Calibration plots for the LLP model for 6-year lung cancer incidence in all datasets

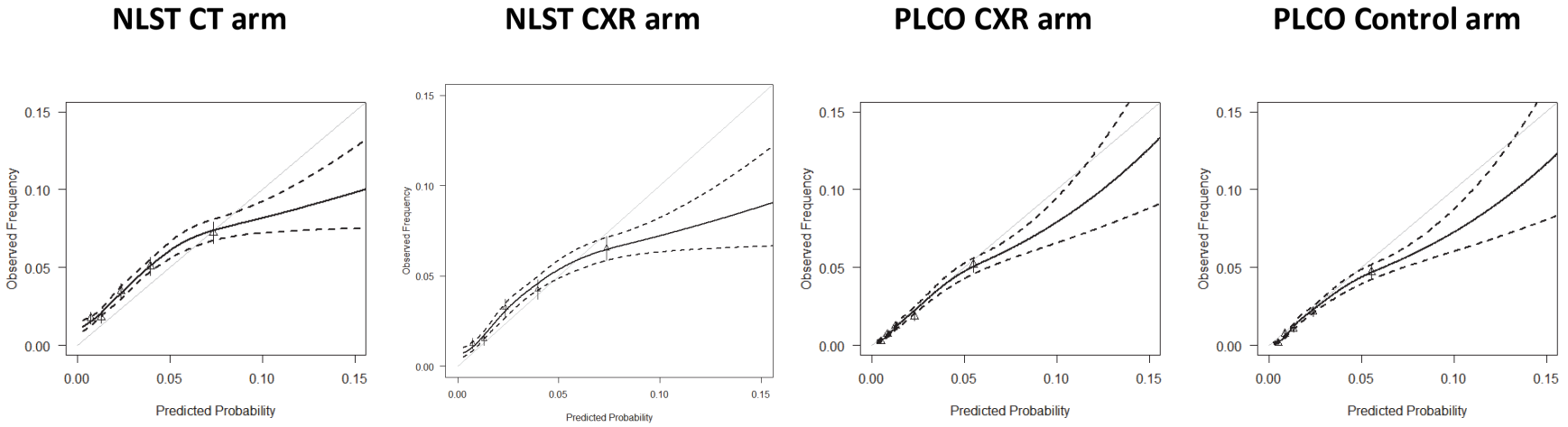


Figure S5: Calibration plots for the LLP model for 6-year lung cancer mortality in all datasets

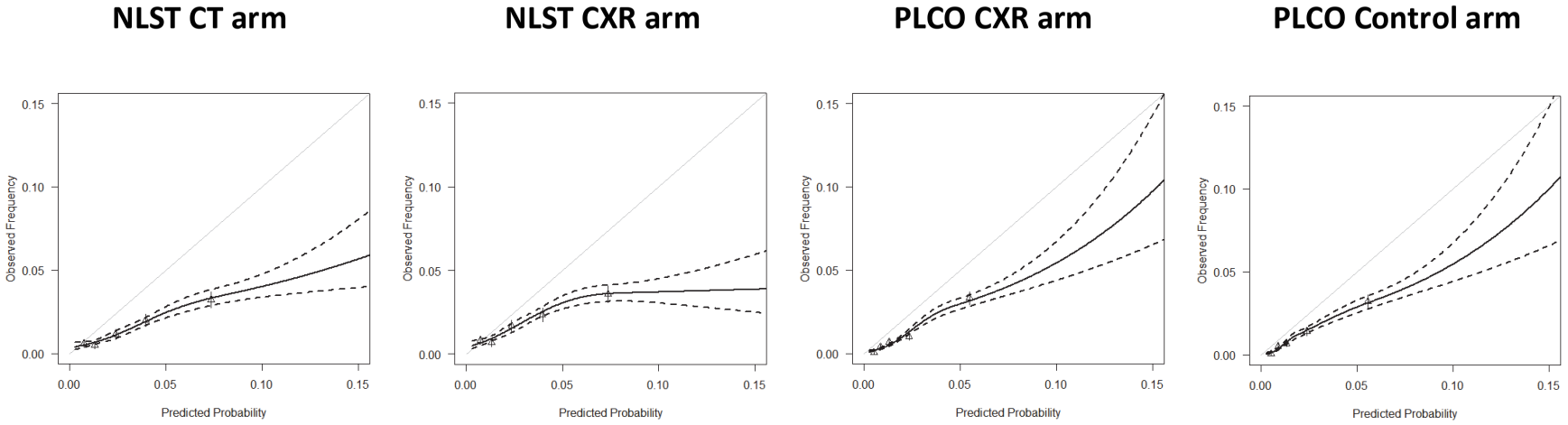


Figure S6: Calibration plots for the simplified LLP model for 6-year lung cancer incidence in all datasets

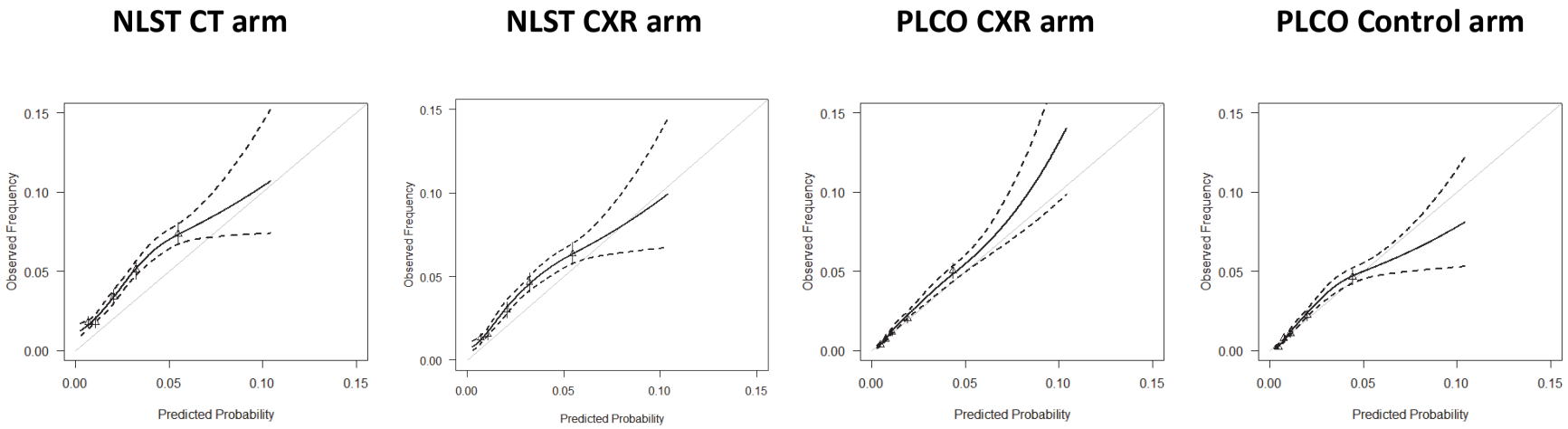


Figure S7: Calibration plots for the simplified LLP model for 6-year lung cancer mortality in all datasets

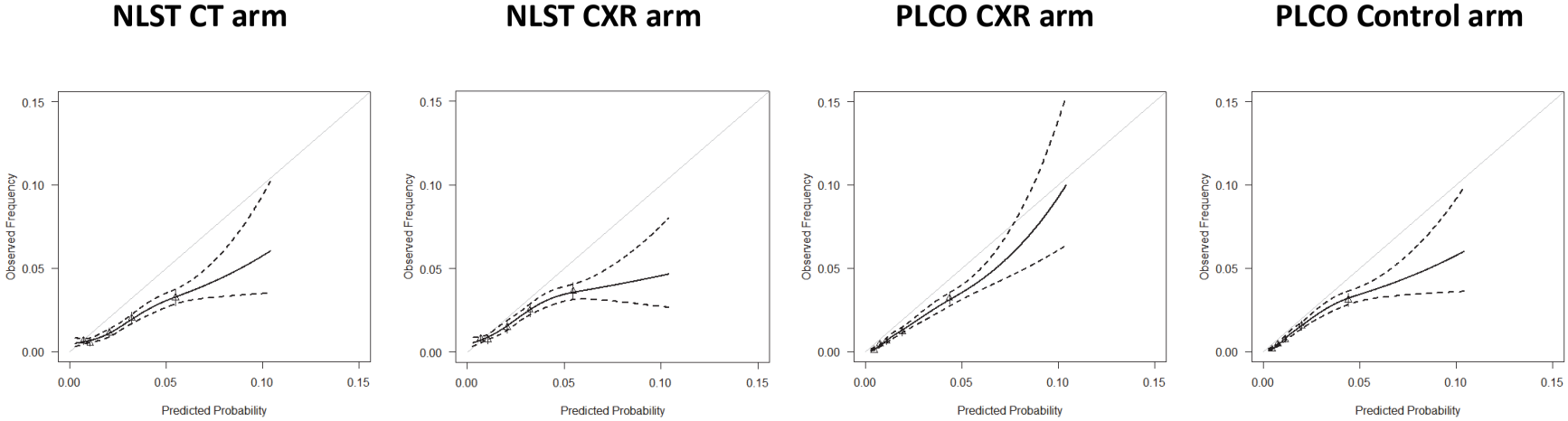


Figure S8: Calibration plots for the PLCOm2012 model for 6-year lung cancer incidence in all datasets

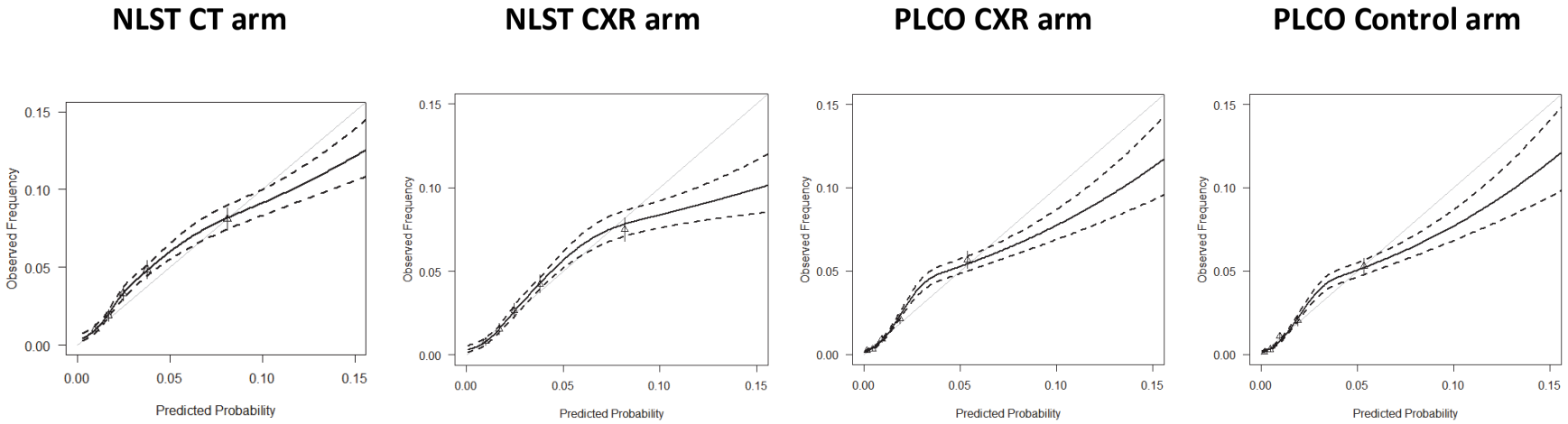


Figure S9: Calibration plots for the PLCOm2012 model for 6-year lung cancer mortality in all datasets

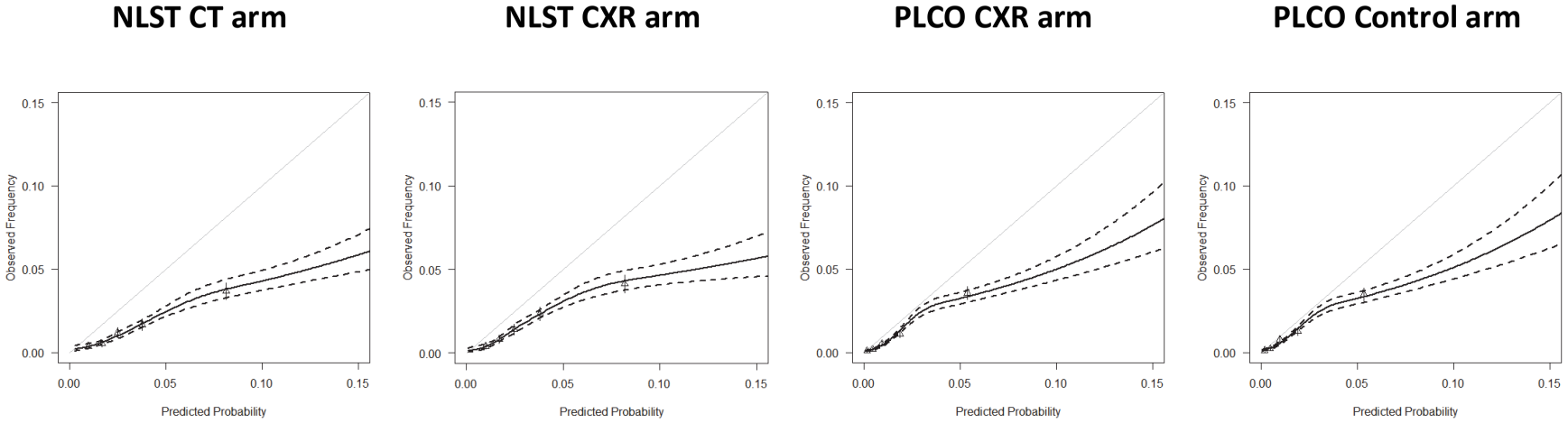


Figure S10: Calibration plots for the simplified PLCOm2012 model for 6-year lung cancer incidence in all datasets

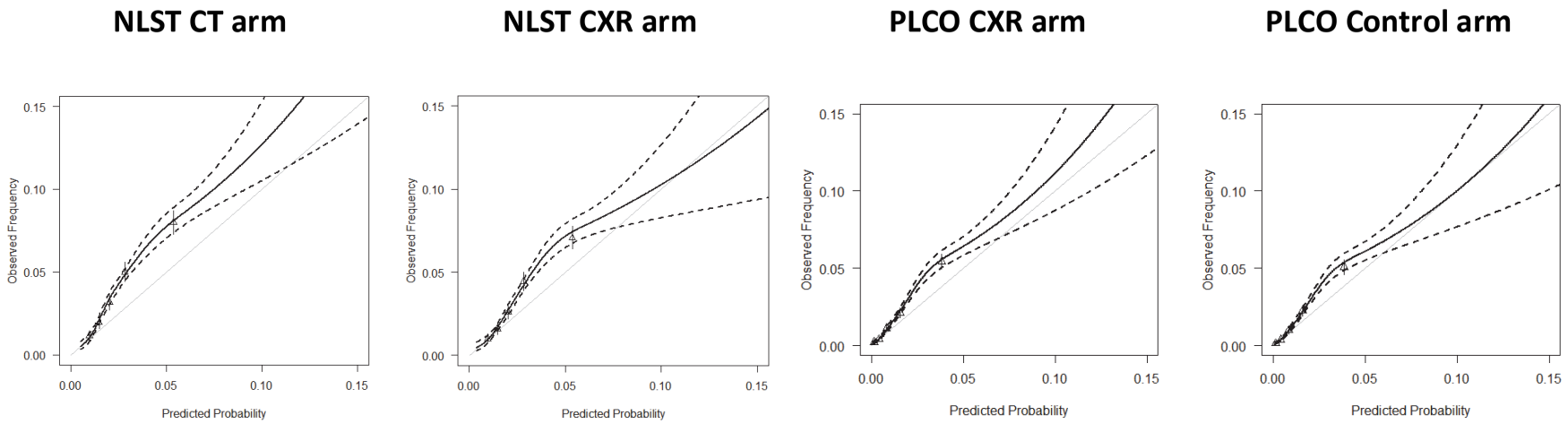


Figure S11: Calibration plots for the simplified PLCOm2012 model for 6-year lung cancer mortality in all datasets

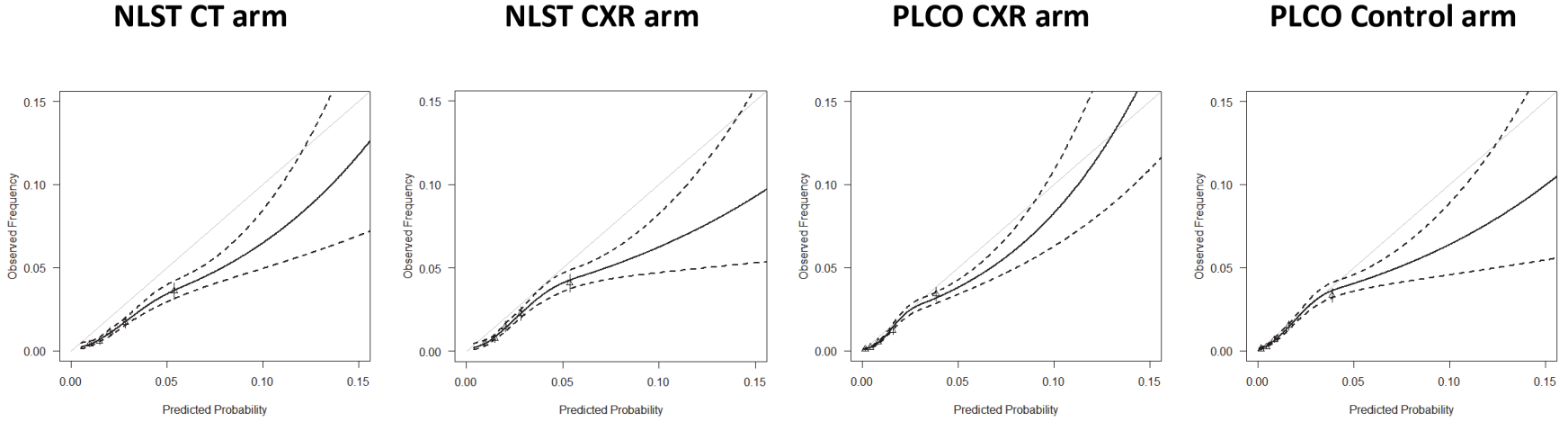


Figure S12: Calibration plots for the TSCE incidence model for 6-year lung cancer incidence in all datasets

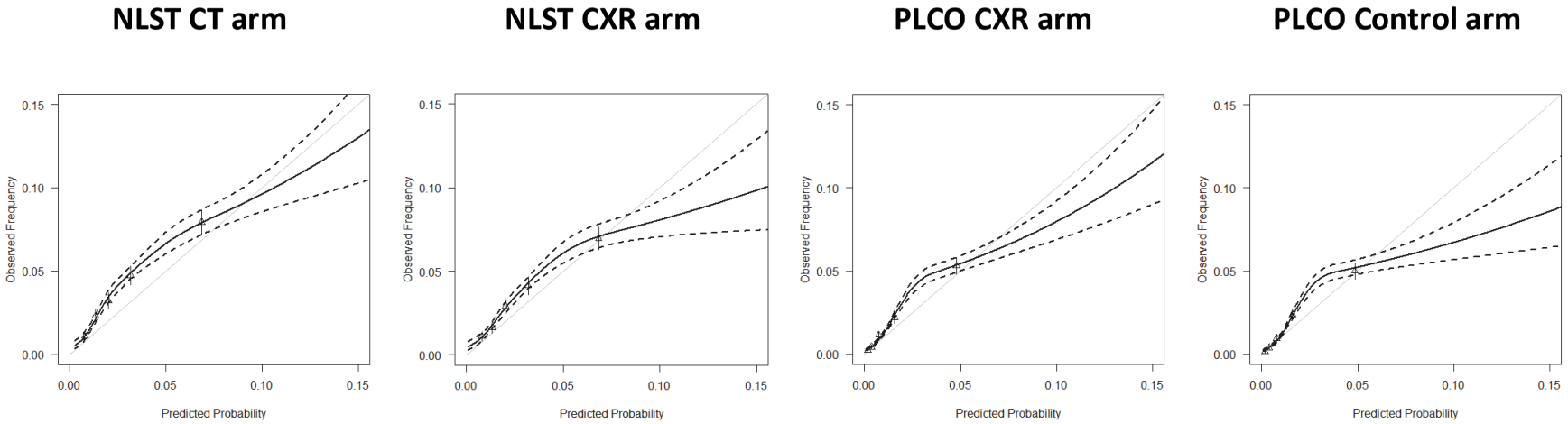
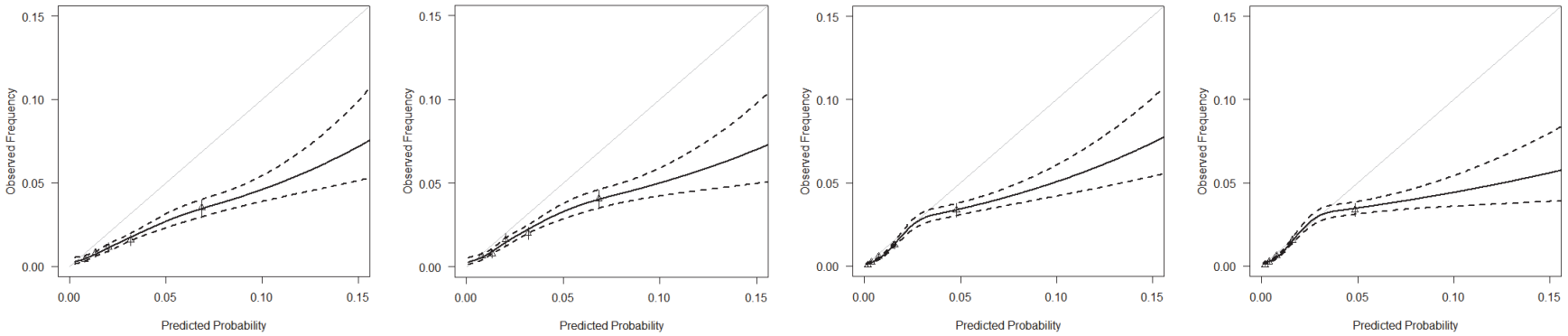


Figure S13: Calibration plots for the TSCE incidence model for 6-year lung cancer mortality in all datasets



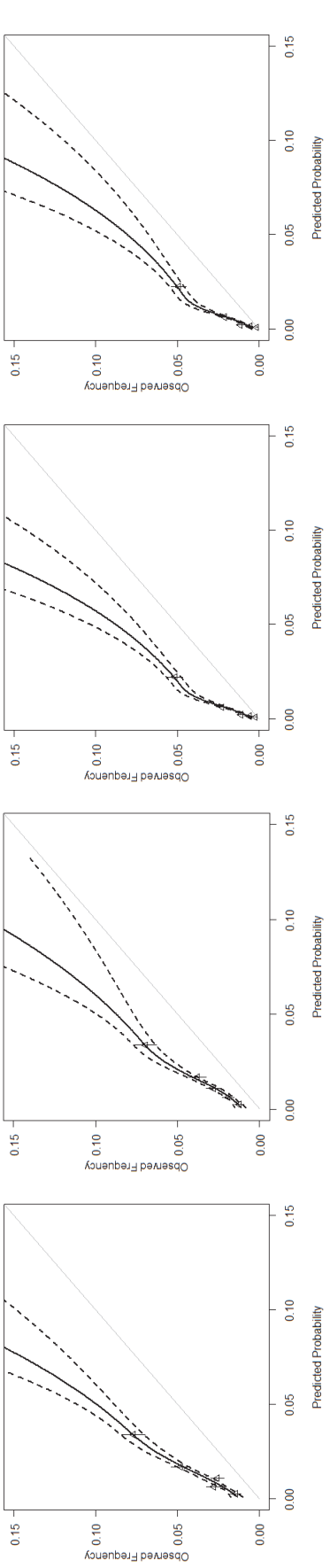


Figure S14: Calibration plots for the Knoke model for 6-year lung cancer incidence in all datasets

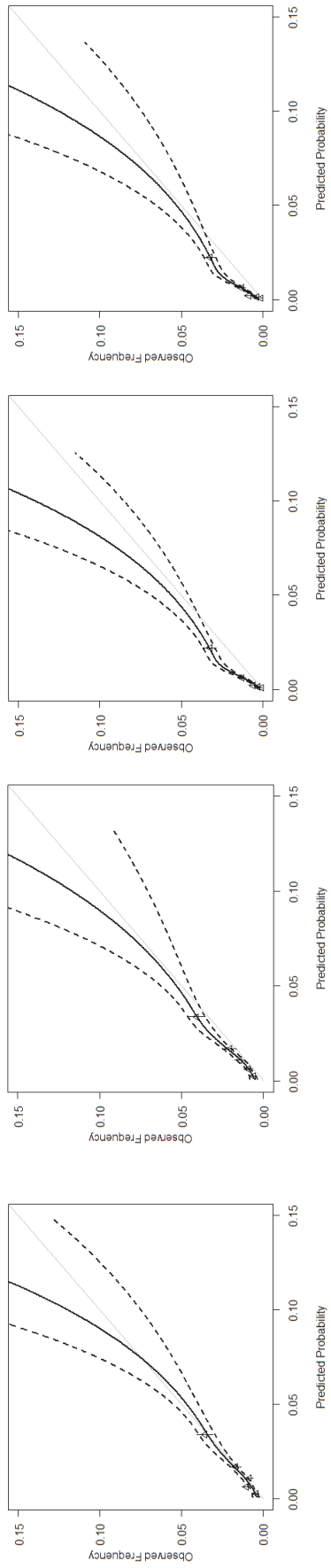


Figure S15: Calibration plots for the Knoke model for 6-year lung cancer mortality in all datasets

Figure S16: Calibration plots for the TSCE CPS lung cancer death model for 6-year lung cancer incidence in all datasets

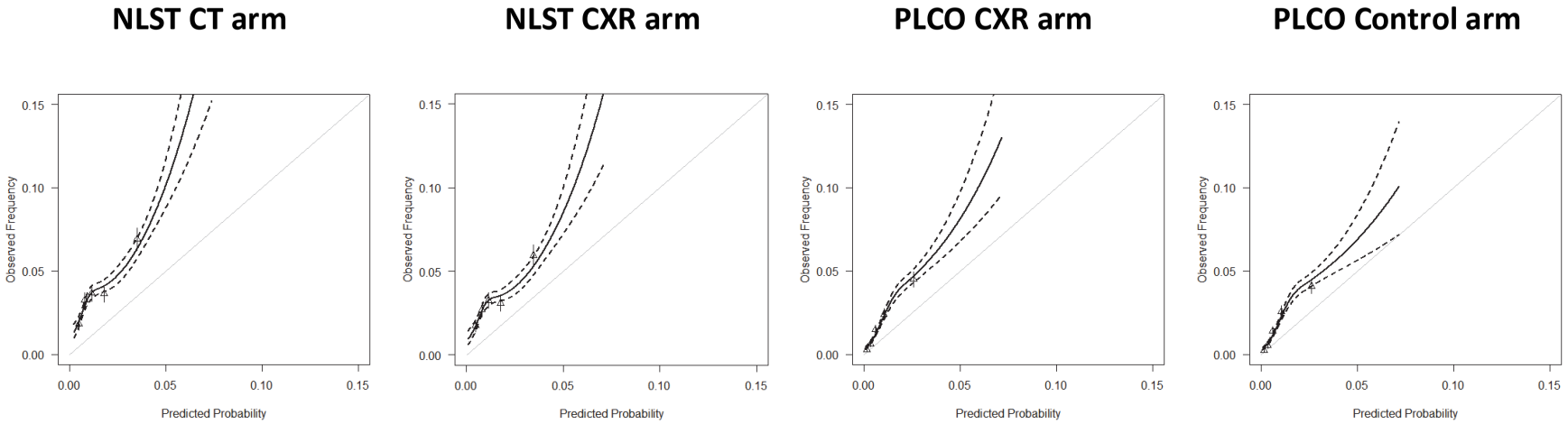


Figure S17: Calibration plots for the TSCE CPS lung cancer death model for 6-year lung cancer mortality in all datasets

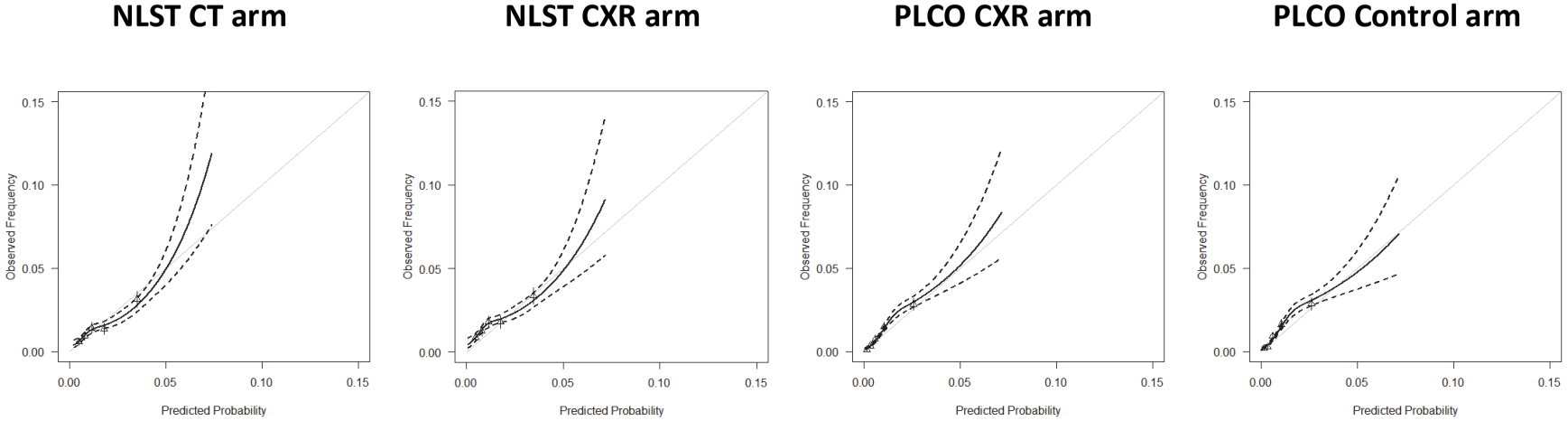


Figure S18: Calibration plots for the TSCE NHS/HPFS lung cancer death model for 6-year lung cancer incidence in all datasets

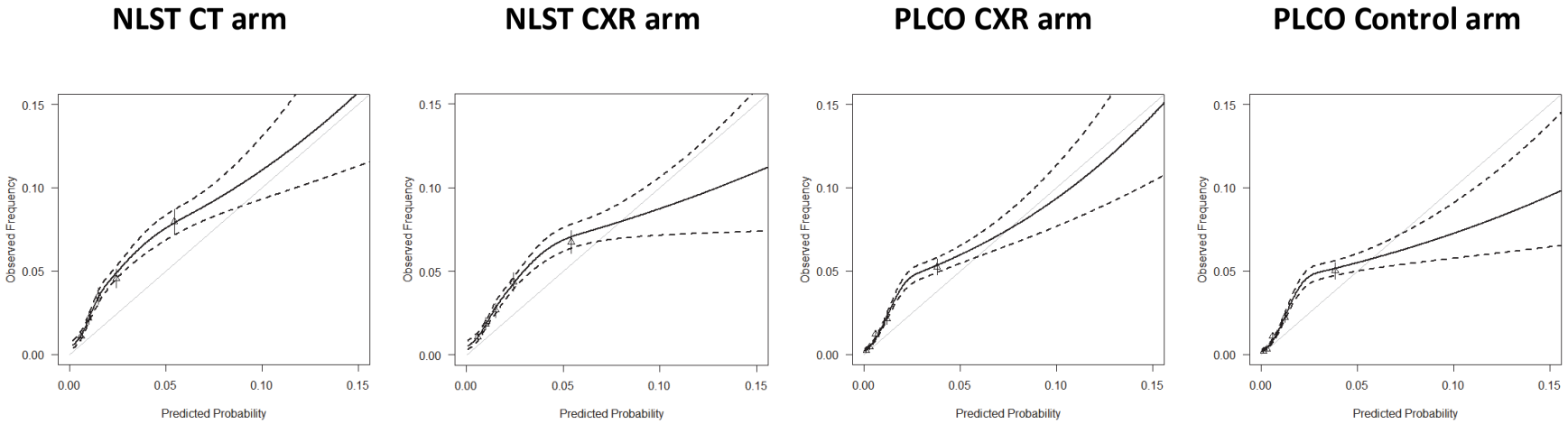


Figure S19: Calibration plots for the TSCE NHS/HPFS lung cancer death model for 6-year lung cancer mortality in all datasets

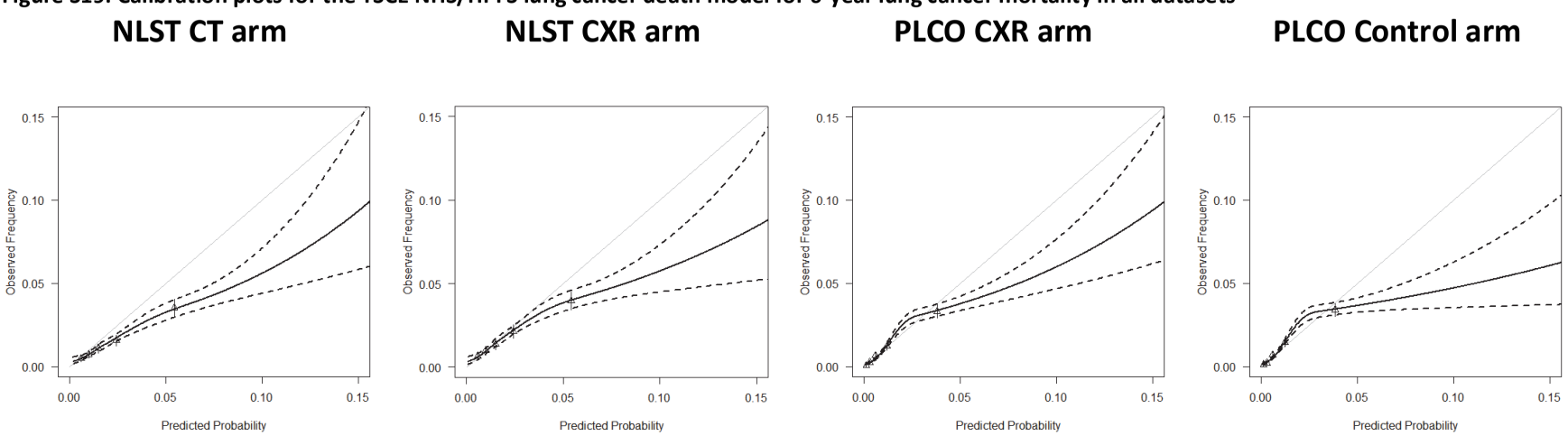


Table S8: Calibration intercepts of models for 6-year lung cancer incidence

Intercepts	NLST (CT-arm)	NLST (CXR arm)	PLCO (CXR arm)	PLCO (Control arm)
Bach model	0.16	0.01	0.06	-0.01
LLP model	0.22	0.07	-0.12	-0.19
Simplified LLP model	0.47	0.32	0.1	0.03
PLCOm2102 model	0.13	-0.02	0.06	0.00
Simplified PLCOm2102 model	0.45	0.30	0.32	0.26
TSCE incidence model	0.33	0.18	0.20	0.13
Knoke model	1.06	0.91	0.78	1.02
TSCE CPS death model	0.96	0.82	0.72	0.66
TSCE NHS/HPFS death model	0.60	0.45	0.47	0.40

Table S9: Calibration slopes of models for 6-year lung cancer incidence

Slopes	NLST (CT-arm)	NLST (CXR arm)	PLCO (CXR arm)	PLCO (Control arm)
Bach model	1.04	1.09	0.99	0.99
LLP model	0.68	0.73	1.05	1.05
Simplified LLP model	0.76	0.80	1.07	1.05
PLCOm2102 model	0.87	0.91	0.94	0.98
Simplified PLCOm2102 model	1.03	1.04	1.02	1.04
TSCE incidence model	0.80	0.79	0.87	0.87
Knoke model	0.70	0.72	1.09	0.79
TSCE CPS death model	0.63	0.59	0.90	0.89
TSCE NHS/HPFS death model	0.77	0.76	0.85	0.84

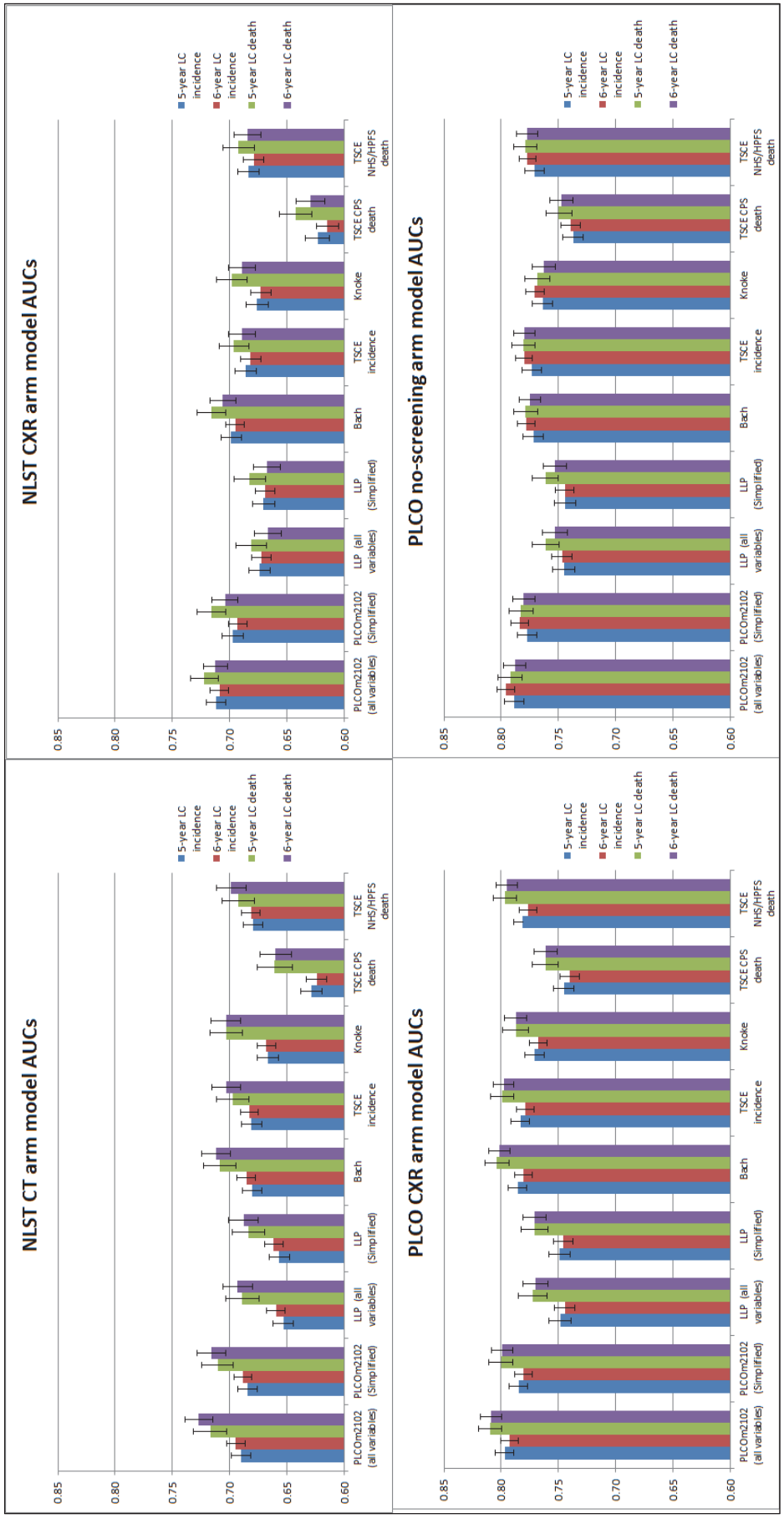
Table S10: Calibration intercepts of models for 6-year lung cancer mortality

Intercepts	NLST (CT-arm)	NLST (CXR arm)	PLCO (CXR arm)	PLCO (Control arm)
Bach model	-0.83	-0.62	-0.49	-0.44
LLP model	-0.77	-0.57	-0.67	-0.61
Simplified LLP model	-0.52	-0.31	-0.44	-0.39
PLCOm2102 model	-0.87	-0.66	-0.49	-0.42
Simplified PLCOm2102 model	-0.54	-0.33	-0.22	-0.17
TSCE incidence model	-0.67	-0.46	-0.35	-0.30
Knoke model	0.06	0.27	0.53	0.59
TSCE CPS death model	-0.03	0.19	0.18	0.23
TSCE NHS/HPFS death model	-0.40	-0.19	-0.08	-0.03

Table S11: Calibration slopes of models for 6-year lung cancer mortality

Slopes	NLST (CT-arm)	NLST (CXR arm)	PLCO (CXR arm)	PLCO (Control arm)
Bach model	1.21	1.17	1.12	0.97
LLP model	0.85	0.69	1.18	1.09
Simplified LLP model	0.91	0.79	1.21	1.09
PLCOm2102 model	1.01	0.92	1.01	0.95
Simplified PLCOm2102 model	1.19	1.10	1.11	1.02
TSCE incidence model	0.91	0.84	0.94	0.86
Knoke model	0.89	0.80	0.85	0.76
TSCE CPS death model	0.82	0.65	0.99	0.92
TSCE NHS/HPFS death model	0.87	0.80	0.92	0.84

Figure S20: Area under the receiver operator curve (AUC) of the investigated risk models (with 95% confidence interval), by dataset predicted outcome and 5- and 6-year timeframe



Abbreviations: National Lung Screening Trial (NLST); Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO); computed tomography (CT); chest radiography (CXR); Lung Cancer (LC).

Figure S21: Decision curve analyses for the Bach model for 6-year lung cancer incidence in all datasets

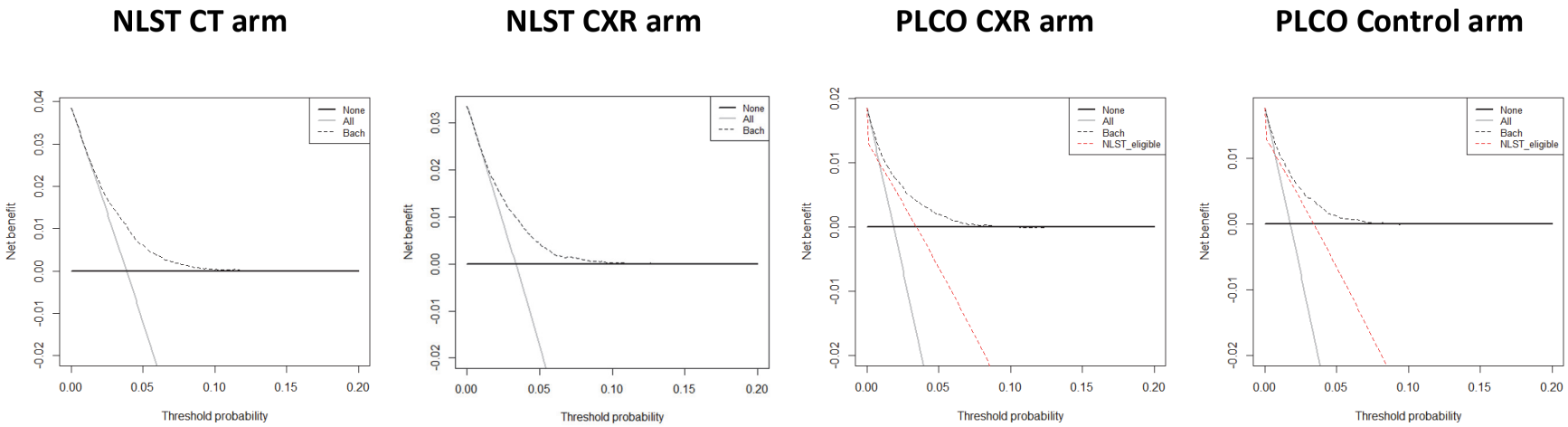
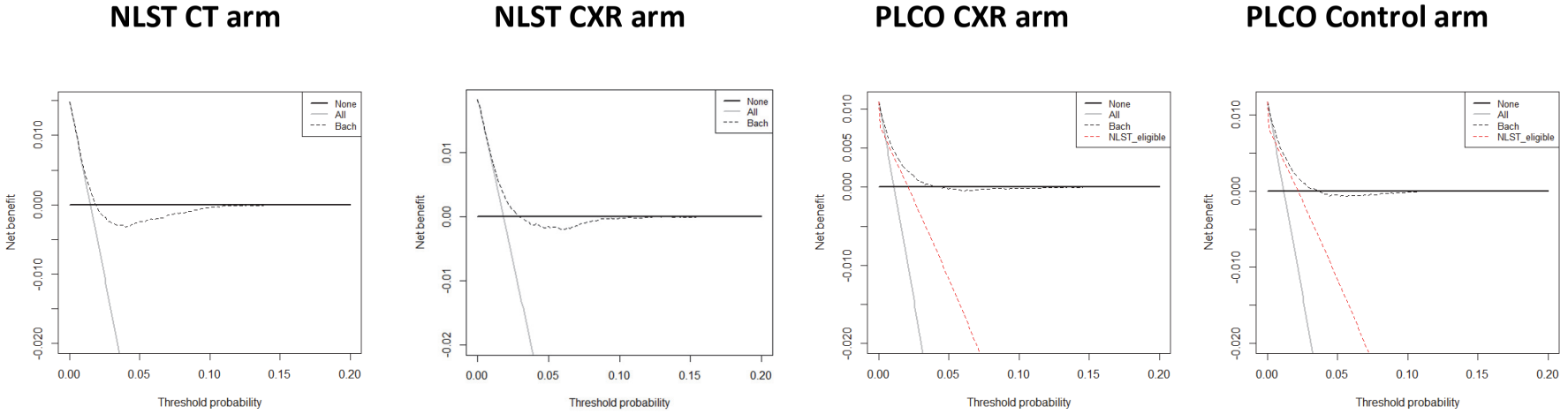


Figure S22: Decision curve analyses for the Bach model for 6-year lung cancer mortality in all datasets



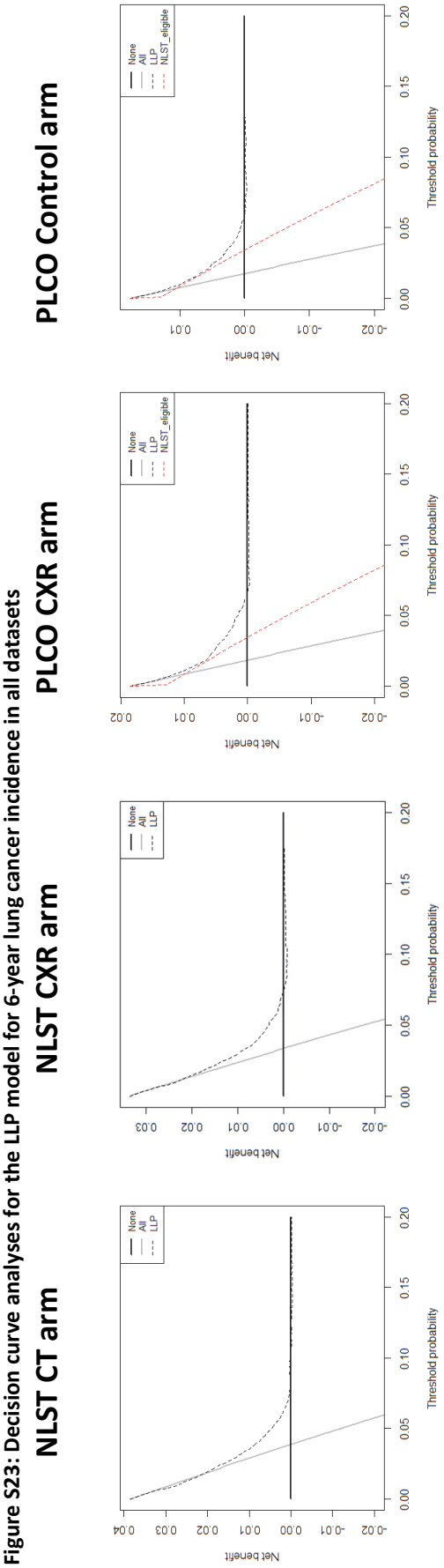


Figure S23: Decision curve analyses for the LLP model for 6-year lung cancer incidence in all datasets

Figure S24: Decision curve analyses for the LLP model for 6-year lung cancer mortality in all datasets

Figure S25: Decision curve analyses for the simplified LLP model for 6-year lung cancer incidence in all datasets

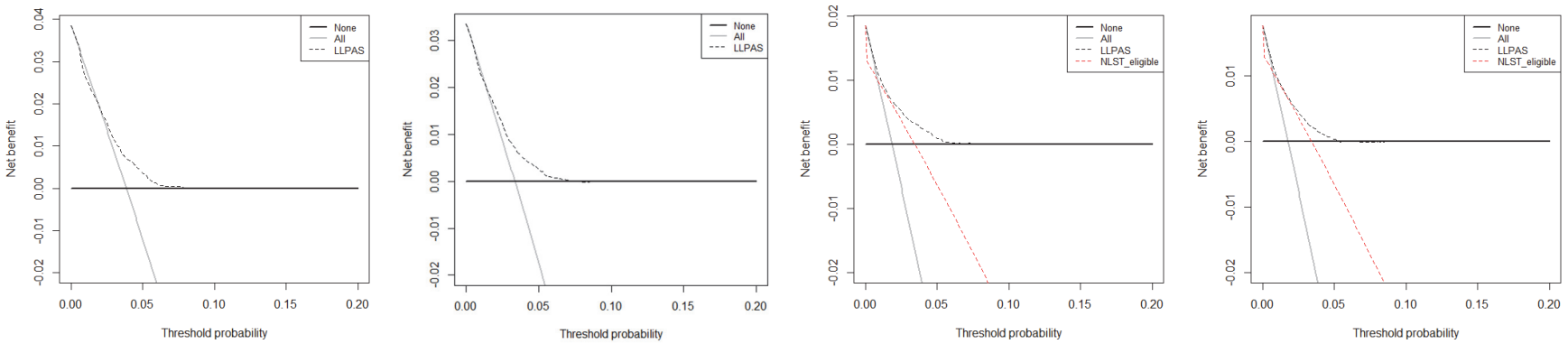
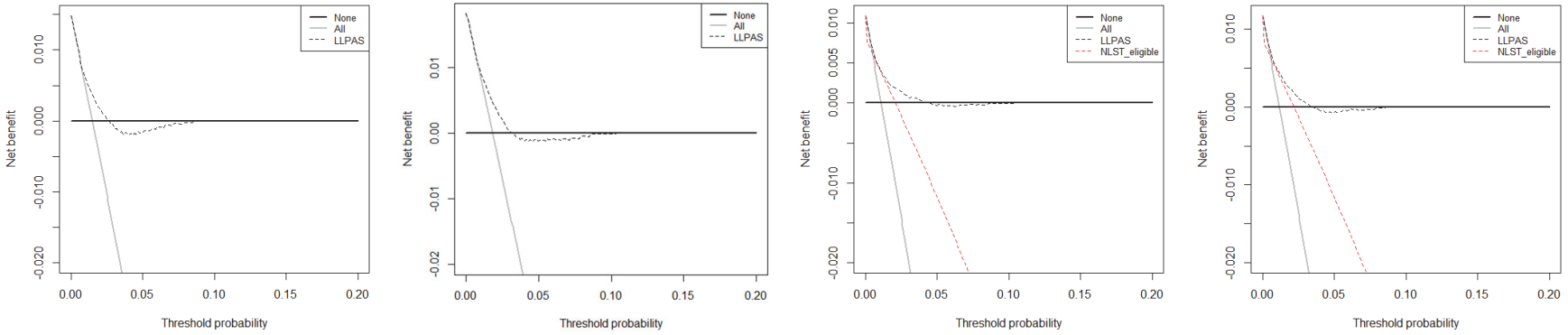


Figure S26: Decision curve analyses for the simplified LLP model for 6-year lung cancer mortality in all datasets



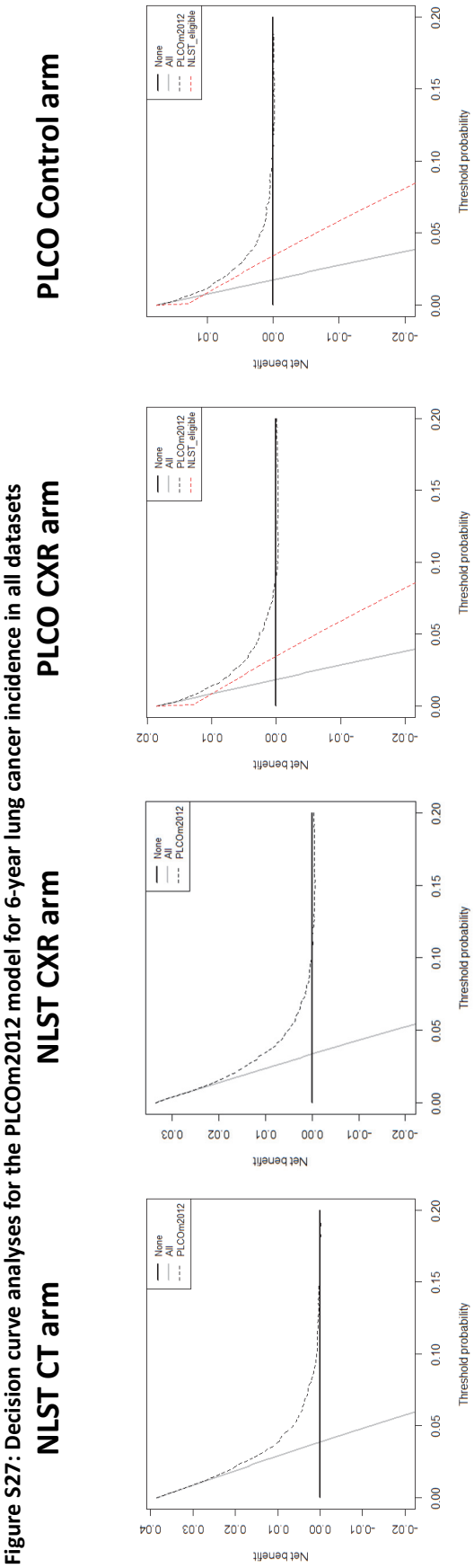


Figure S27: Decision curve analyses for the PLCOm2012 model for 6-year lung cancer incidence in all datasets

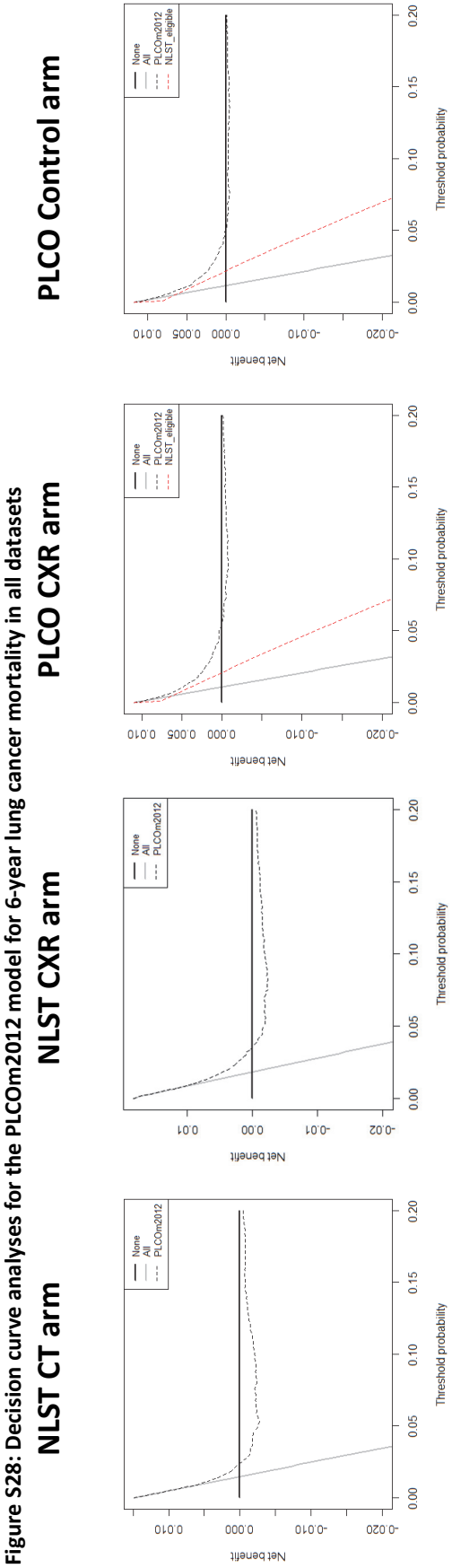


Figure S28: Decision curve analyses for the PLCOm2012 model for 6-year lung cancer mortality in all datasets

Figure S29: Decision curve analyses for the simplified PLCOm2012 model for 6-year lung cancer incidence in all datasets

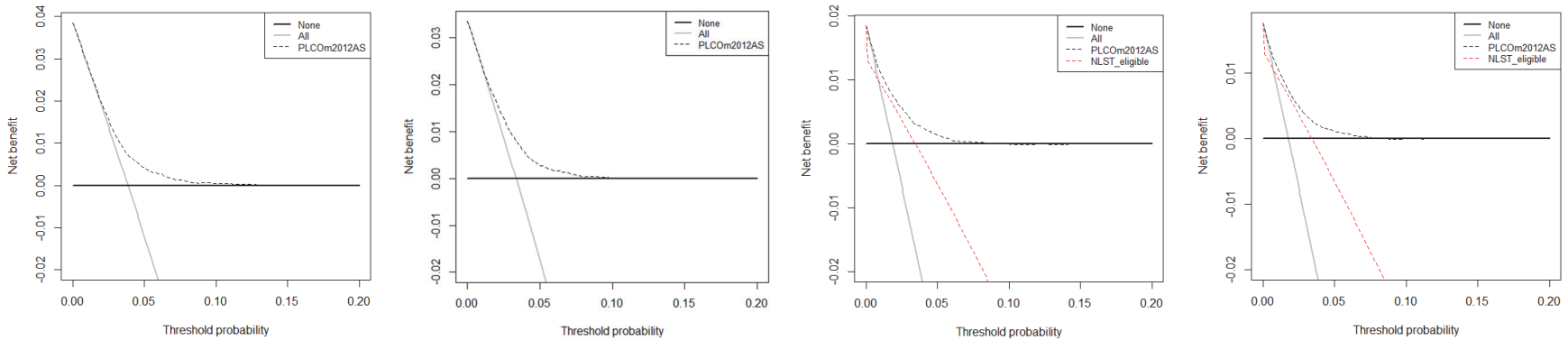
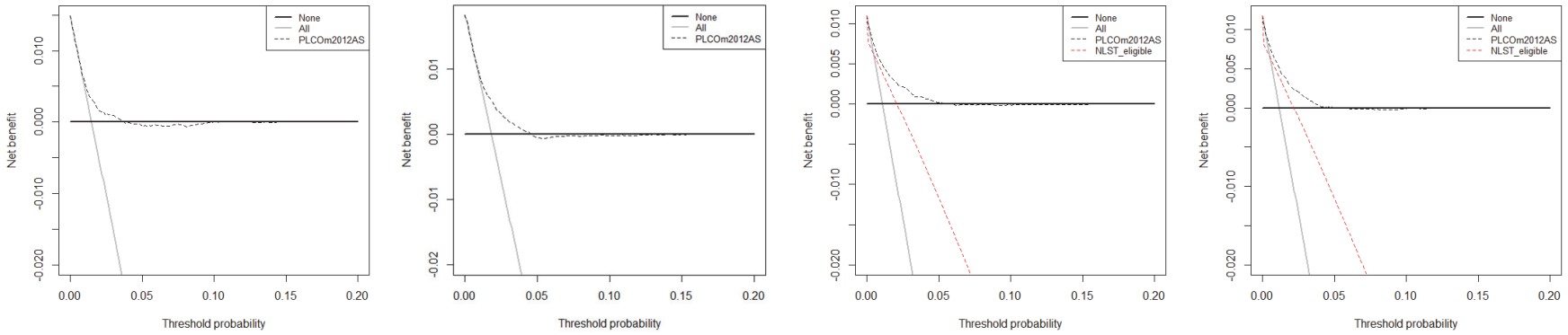


Figure S30: Decision curve analyses for the simplified PLCOm2012 model for 6-year lung cancer mortality in all datasets



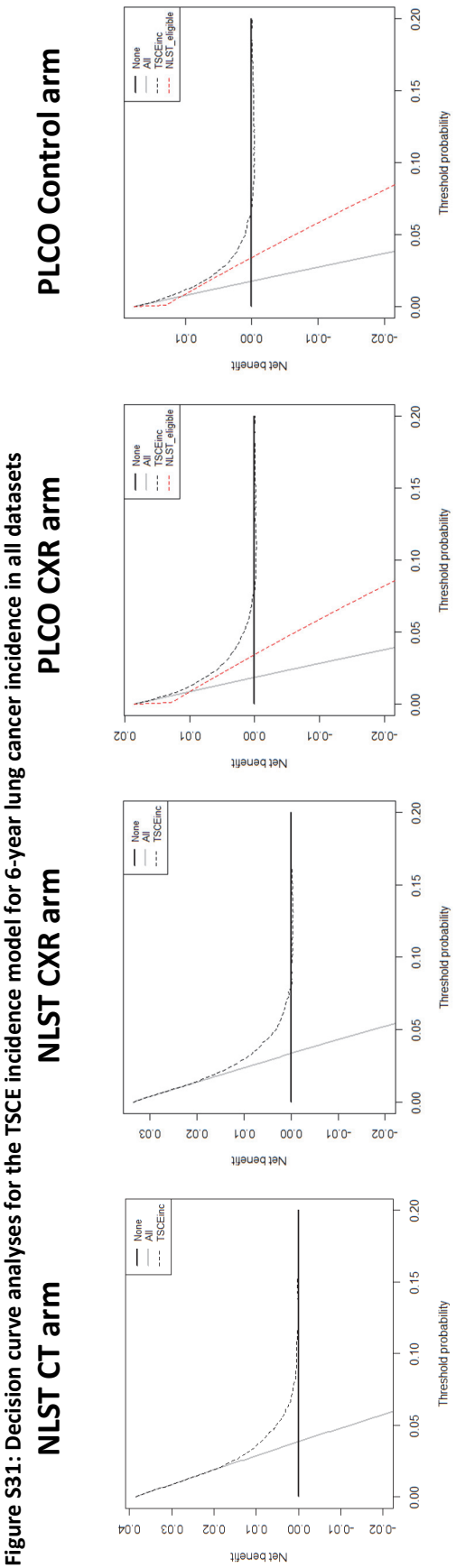


Figure S31: Decision curve analyses for the TSCE incidence model for 6-year lung cancer incidence in all datasets

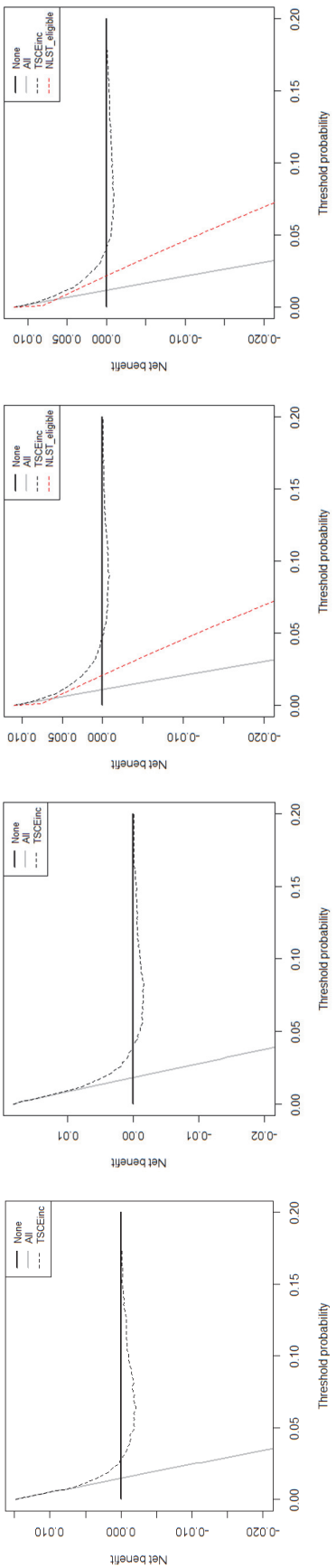


Figure S32: Decision curve analyses for the TSCE incidence model for 6-year lung cancer mortality in all datasets

Figure S33: Decision curve analyses for the Knoke model for 6-year lung cancer incidence in all datasets

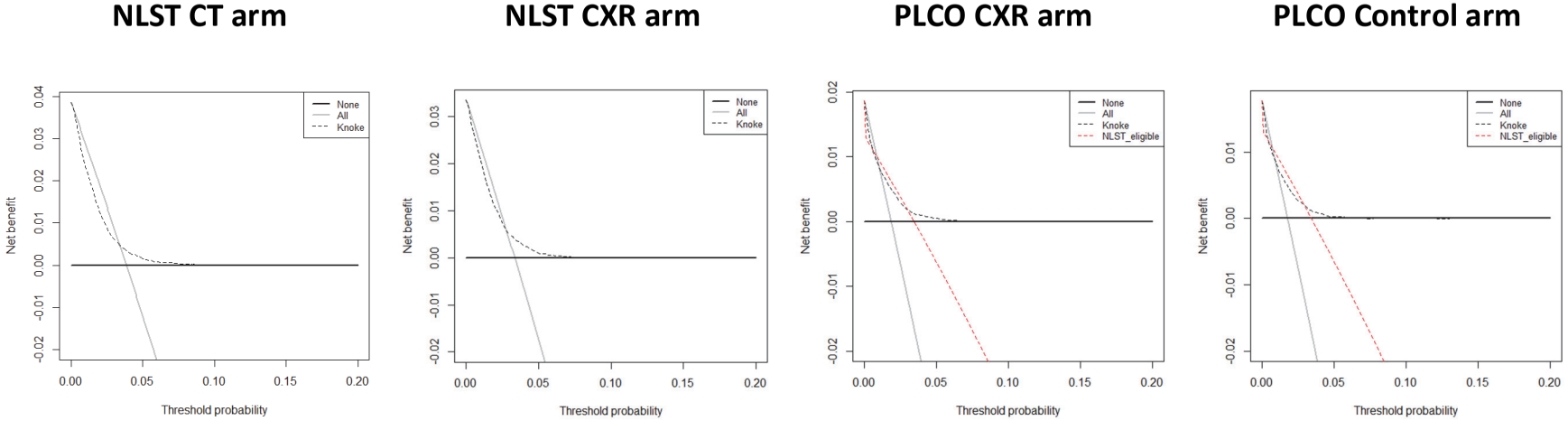


Figure S34: Decision curve analyses for the Knoke model for 6-year lung cancer mortality in all datasets

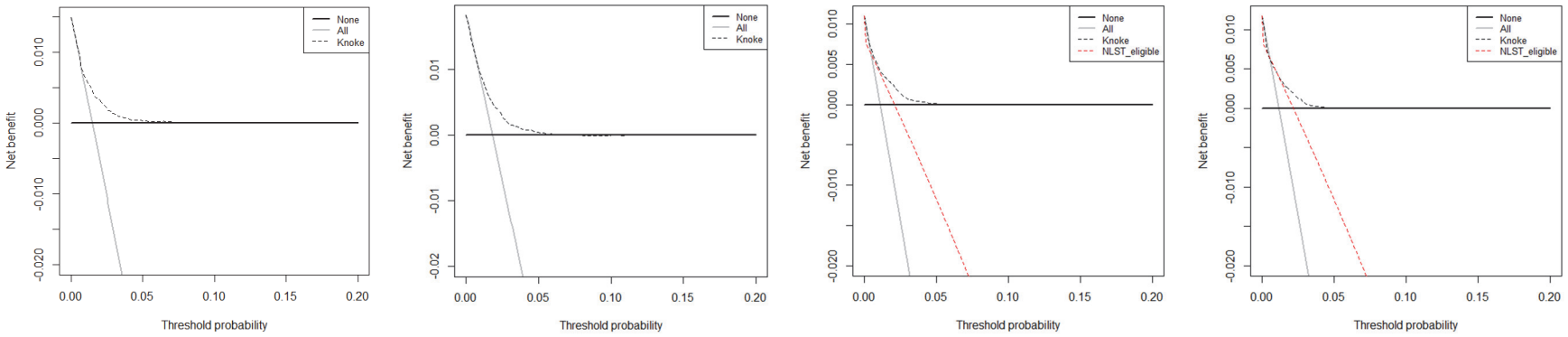


Figure S35: Decision curve analyses for the TSCE CPS lung cancer death model for 6-year lung cancer incidence in all datasets

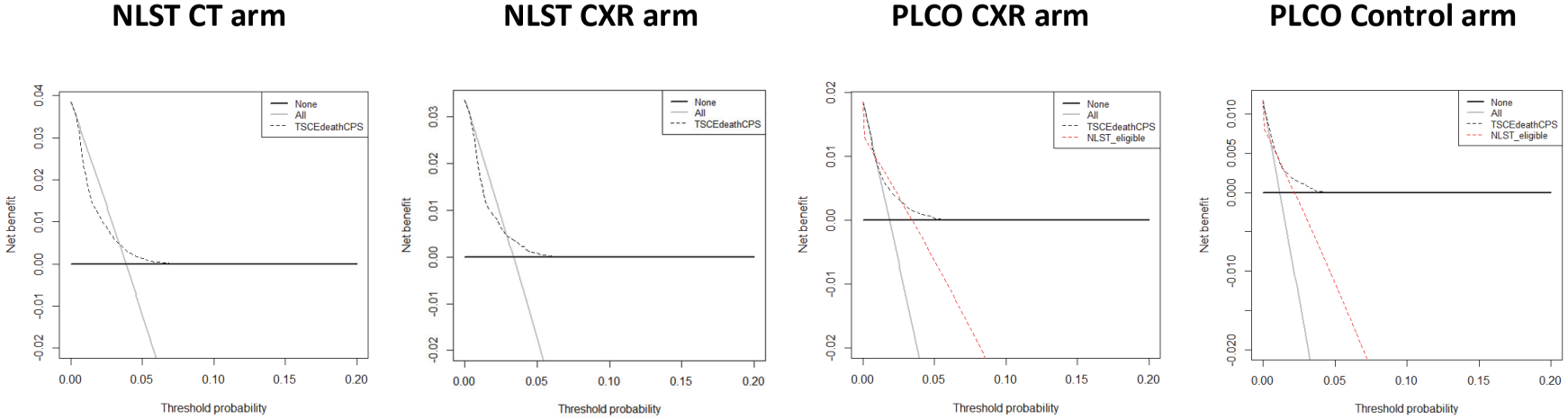


Figure S36: Decision curve analyses for the TSCE CPS lung cancer death model for 6-year lung cancer mortality in all datasets

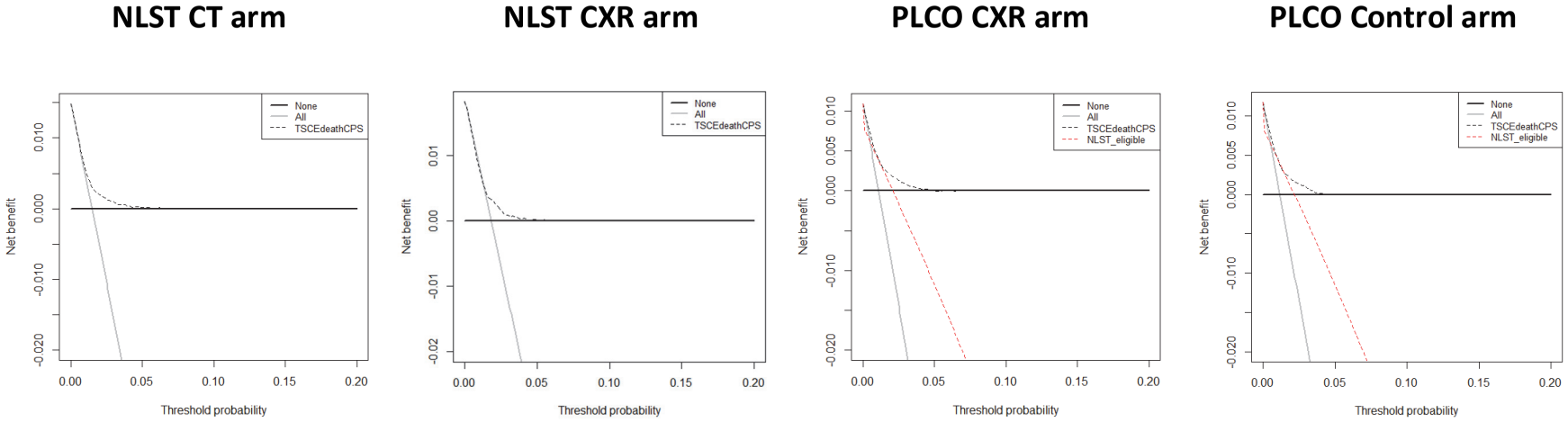


Figure S37: Decision curve analyses for the TSCE NHS/HPFS lung cancer death model for 6-year lung cancer incidence in all datasets

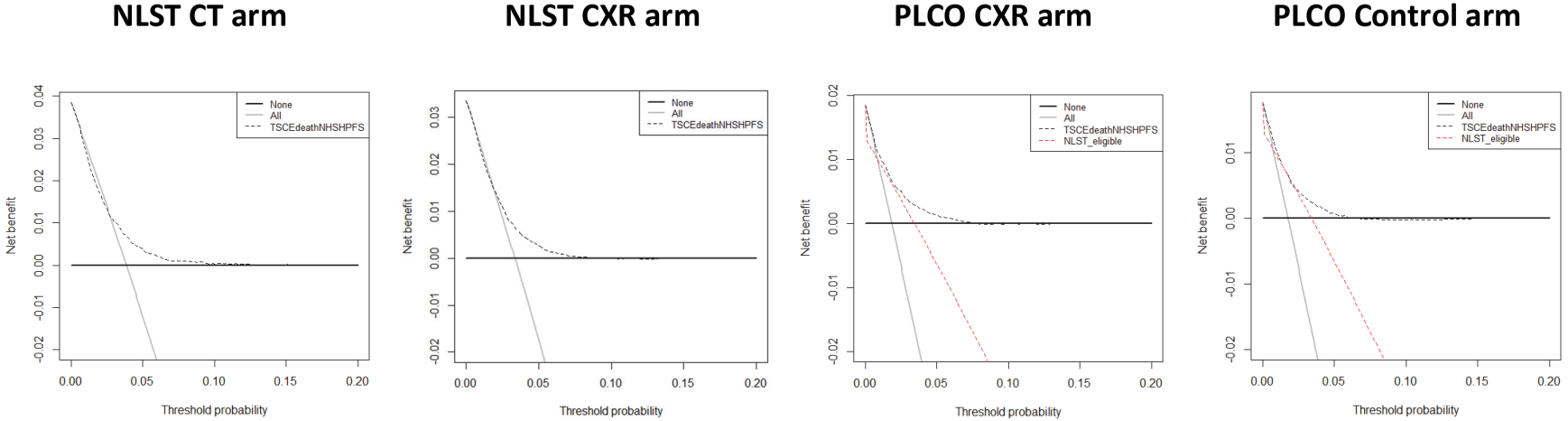


Figure S37: Decision curve analyses for the TSCE NHS/HPFS lung cancer death model for 6-year lung cancer mortality in all datasets

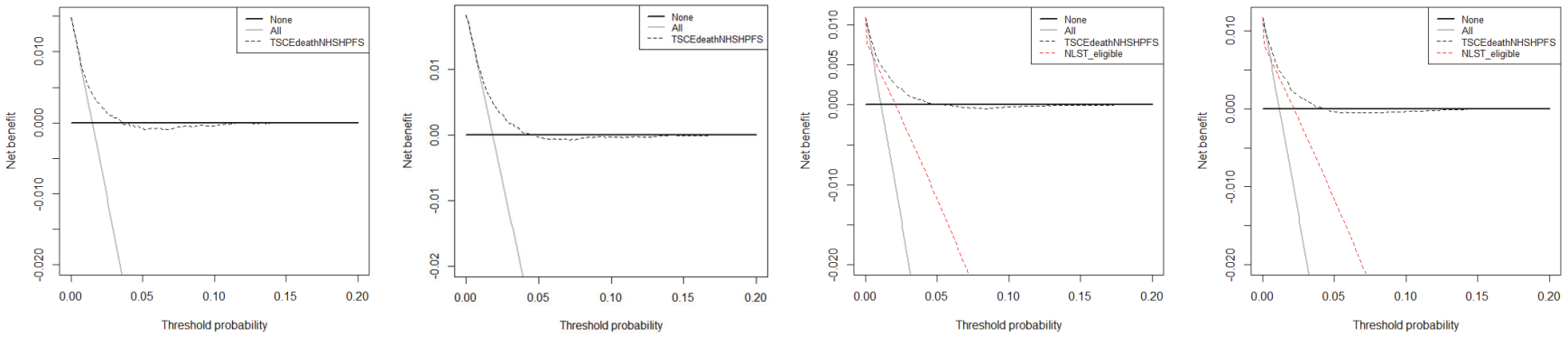


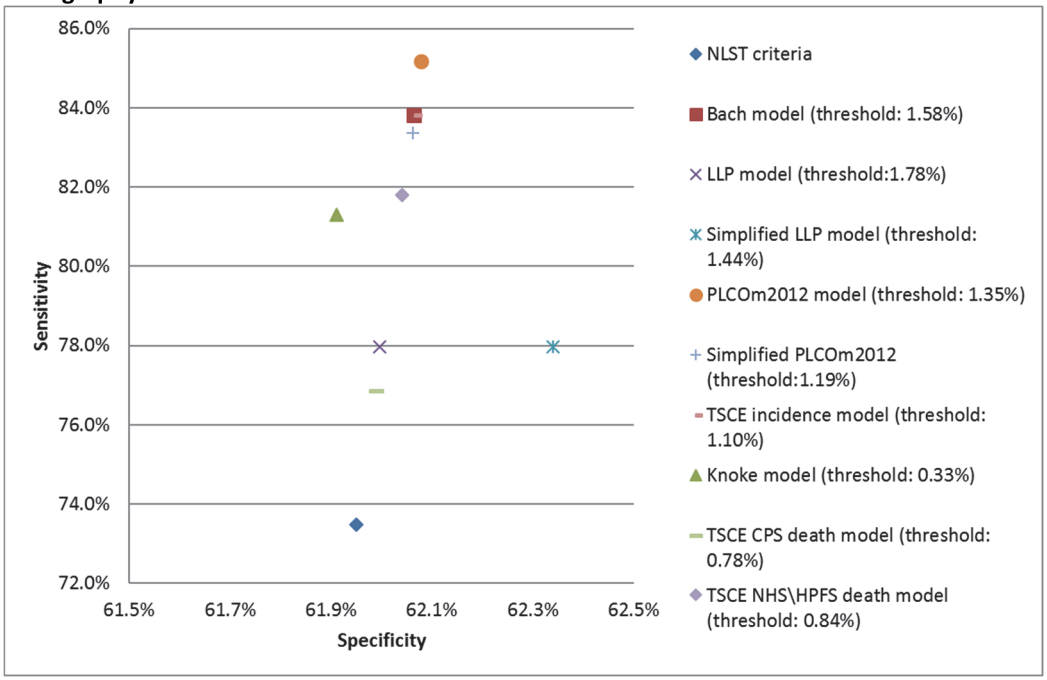
Table S12: Lower and upper risk thresholds for which the risk prediction models have a positive net-benefit compared to the National Lung Screening Trial (NLST) criteria for 6-year lung cancer death

Model/dataset	NLST (computed tomography-arm)		NLST (chest radiography arm)		PLCO (chest radiography arm)		PLCO (Control arm	
	Lower risk threshold (WF*)	Upper risk threshold (WF*)	Lower risk threshold (WF*)	Upper risk threshold (WF*)	Lower risk threshold (WF*)	Upper risk threshold (WF*)	Lower risk threshold (WF*)	Upper risk threshold (WF*)
Bach model	0.2% (499.0)	1.8% (54.6)	0.9% (110.1)	2.9% (33.5)	0.1% (999.0)	4.3% (22.3)	0.2% (499.0)	3.6% (26.8)
LLP model	0.7% (141.9)	2.3% (42.5)	0.9% (110.1)	2.9% (33.5)	1.3% (75.9)	3.3% (29.3)	1.2% (82.3)	2.9% (33.5)
Simplified LLP model	0.7% (141.9)	2.6% (37.5)	0.9% (110.1)	3.0% (32.3)	0.3% (332.3)	4.2% (22.8)	0.1% (999.0)	3.5% (27.6)
PLCOm2012 model	0.3% (332.3)	2.3% (42.5)	0.1% (999.0)	3.5% (27.6)	0.1% (999.0)	5.4% (17.5)	0.1% (999.0)	5.0% (19.0)
Simplified PLCOm2012 model	0.5% (199.0)	3.7% (26.0)	0.7% (141.9)	4.5% (21.2)	0.1% (999.0)	5.9% (16.0)	0.2% (499.0)	4.9% (19.4)
TSCE lung cancer incidence model	0.2% (499.0)	2.7% (36.0)	0.1% (999.0)	3.9% (24.6)	0.2% (499.0)	4.8% (19.8)	0.2% (499.0)	4.1% (23.4)
Knoke model	0.5% (199.0)	7.4% (12.5)	0.7% (141.9)	7.6% (12.2)	0.4% (249.0)	7.7% (12.0)	1.3% (75.9)	5.7% (16.5)
TSCE CPS lung cancer death model	0.6% (165.7)	7.1% (13.1)	1.3% (75.9)	6.2% (15.1)	0.1% (999.0)	5.0% (19.0)	1.5% (65.7)	4.2% (22.8)
TSCE NHS/HPFS lung cancer death model	0.2% (499.0)	3.5% (27.6)	0.1% (999.0)	4.7% (20.3)	0.1% (999.0)	5.5% (17.2)	0.2% (499.0)	4.1% (23.4)

Abbreviations: National Lung Screening Trial (NLST); Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO), Liverpool Lung Project (LLP), Two-Stage Clonal Expansion (TSCE), Cancer Prevention Study (CPS), Nurses' Health Study / Health Professionals Follow-up Study (NHS/HPFS).

**Weighting factor corresponding to the risk threshold; i.e. the ratio of how much worse missing one case of lung cancer that could be detected through screening is valued compared to unnecessarily screening one person.*

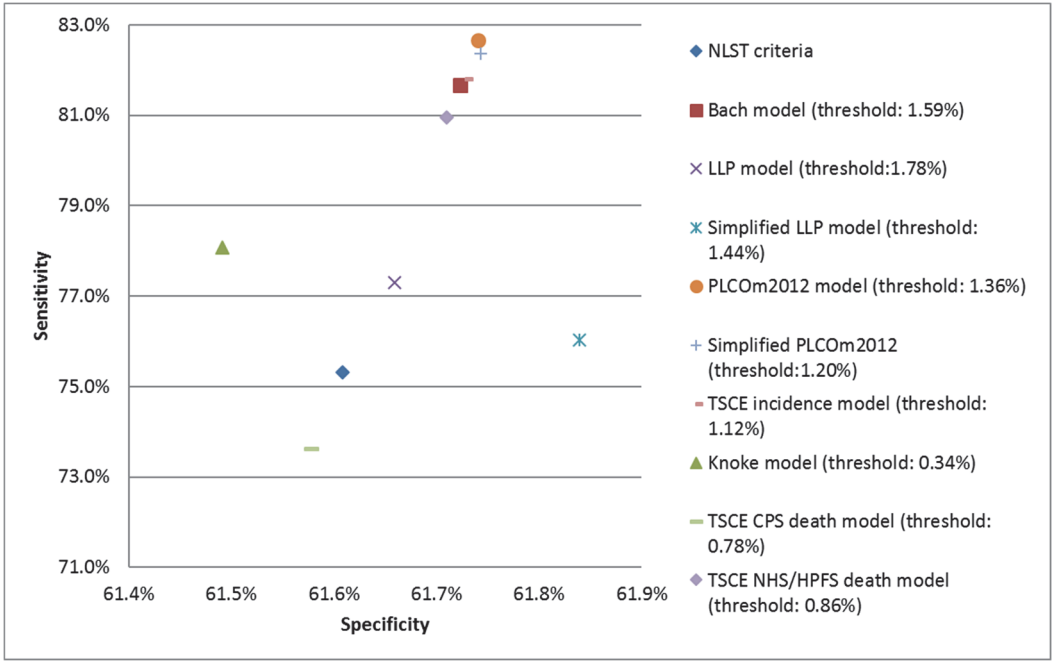
Figure S38: Sensitivity, specificity and risk thresholds for the investigated risk models and the NLST eligibility criteria for 6-year lung cancer mortality in the PLCO chest radiography arm



Abbreviations: National Lung Screening Trial (NLST); Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).

	NLST criteria	Bach model	LLP model	Simplified LLP model	PLCOm2012 model	Simplified PLCOm2012 model	TSCE incidence model	Knoke model	TSCE CPS death model	TSCE NHS/HPFS model
Sensitivity (95% confidence interval)	73.5% (69.1-77.5%)	83.8% (80.1-87.1%)	78.0% (73.8-81.7%)	78.0% (73.8-81.7%)	85.2% (81.5-88.3%)	83.4% (79.6-86.7%)	83.8% (80.1-87.1%)	81.3% (77.4-84.8%)	76.9% (72.7-80.7%)	81.8% (77.9-85.3%)
Specificity (95% confidence interval)	61.9% (61.5-62.4%)	62.1% (61.6-62.5%)	62.0% (61.5-62.5%)	62.3% (61.9-62.8%)	62.1% (61.6-62.6%)	62.1% (61.6-62.5%)	62.1% (61.6-62.5%)	61.7% (61.4-62.4%)	62.0% (61.5-62.5%)	62.0% (61.6-62.5%)

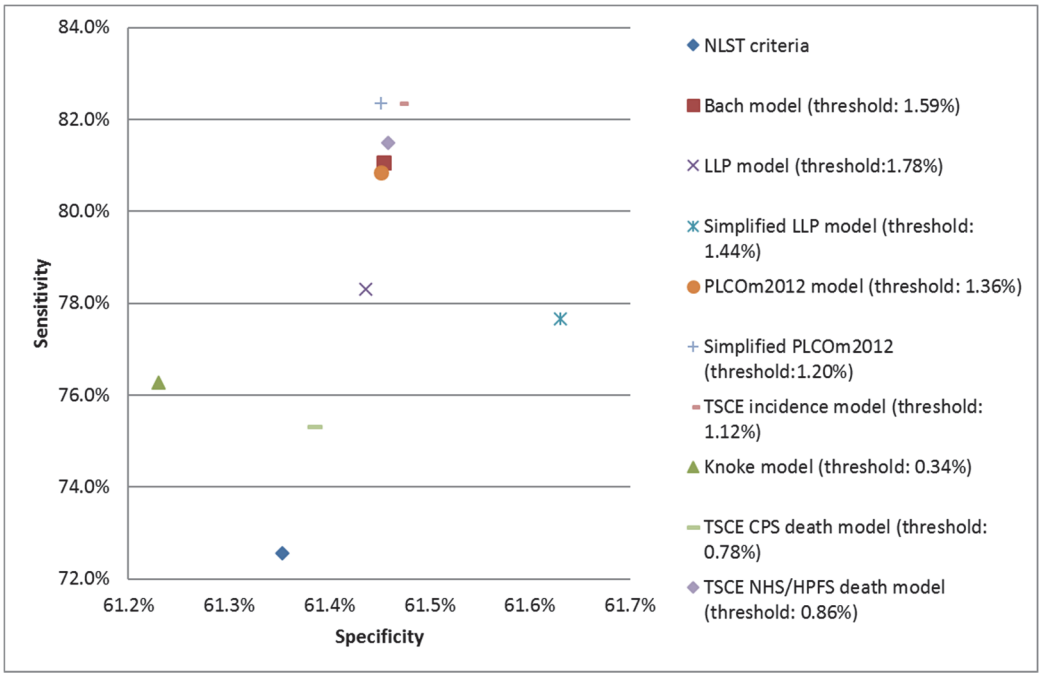
Figure S39: Sensitivity, specificity and risk thresholds for the investigated risk models and the NLST eligibility criteria for 6-year lung cancer incidence in the PLCO control arm



Abbreviations: National Lung Screening Trial (NLST); Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).

	NLST criteria	Bach model	LLP model	Simplified LLP model	PLCOm2012 model	Simplified PLCOm2012 model	TSCE incidence model	Knoke model	TSCE CPS death model	TSCE NHS/HPFS model
Sensitivity (95% confidence interval)	75.3% (72.0-78.5%)	81.7% (78.6-84.4%)	77.3% (74.0-80.3%)	76.0% (72.7-79.1%)	82.7% (79.7-85.4%)	82.4% (79.4-85.1%)	81.8% (78.8-84.6%)	78.1% (74.8-81.1%)	73.6% (70.2-76.8%)	81.0% (77.9-83.8%)
Specificity (95% confidence interval)	61.6% (61.1-62.1%)	61.7% (61.2-62.2%)	61.7% (61.2-62.1%)	61.8% (61.4-62.3%)	61.7% (61.3-62.2%)	61.7% (61.3-62.2%)	61.7% (61.2-62.2%)	61.5% (61.0-62.0%)	61.6% (61.1-62.1%)	61.7% (61.2-62.2%)

Figure S40: Sensitivity, specificity and risk thresholds for the investigated risk models and the NLST eligibility criteria for 6-year lung cancer mortality in the PLCO control arm



Abbreviations: National Lung Screening Trial (NLST); Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).

	NLST criteria	Bach model	LLP model	Simplified LLP model	PLCOm2012 model	Simplified PLCOm2012 model	TSCE incidence model	Knoke model	TSCE CPS death model	TSCE NHS/HPFS model
Sensitivity (95% confidence interval)	72.6% (68.3-76.5%)	81.1% (77.2-84.5%)	78.3% (74.3-81.9%)	77.7% (73.6-81.3%)	80.9% (77.0-84.3%	82.3% (78.6-85.7%)	82.3% (78.6-85.7%)	76.3% (72.2-80.1%)	75.3% (71.2-79.2%)	81.5% (77.7-84.9%)
Specificity (95% confidence interval)	61.4% (60.9-61.8%)	61.5% (61.0-61.9%)	61.4% (61.0-61.9%)	61.6% (61.1-62.1%)	61.5% (61.0-61.9%)	61.5% (61.0-62.0%)	61.5% (61.0-62.0%)	61.2% (60.7-61.7%)	61.4% (60.9-61.9%)	61.5% (61.0-61.9%)

References

1. Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *Journal of the National Cancer Institute* 2003; **95**(6): 470-8.

2. Cronin KA, Gail MH, Zou Z, Bach PB, Virtamo J, Albanes D. Validation of a model of lung cancer risk prediction among smokers. *Journal of the National Cancer Institute* 2006; **98**(9): 637-40.

3. Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *British Journal of Cancer* 2007; **98**(2): 270-6.

4. D'Amelio AM, Jr., Cassidy A, Asomaning K, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. *British Journal of Cancer* 2010; **103**(3): 423-9.

5. 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. *Annals of Internal Medicine* 2012; **157**(4): 242-50.

6. Tammemägi MC, Katki HA, Hocking WG, et al. Selection Criteria for Lung-Cancer Screening. *New England Journal of Medicine* 2013; **368**(8): 728-36.

7. Hazelton WD, Clements MS, Moolgavkar SH. Multistage Carcinogenesis and Lung Cancer Mortality in Three Cohorts. *Cancer Epidemiology Biomarkers & Prevention* 2005; **14**(5): 1171.

8. Hazelton WD, Jeon J, Meza R, Moolgavkar SH. Chapter 8: The FHCRC Lung Cancer Model. *Risk Analysis* 2012; **32**: S99-S116.

9. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. *Cancer Causes & Control* 2008; **19**(3): 317-28.

10. Heidenreich WF, Luebeck EG, Moolgavkar SH. Some Properties of the Hazard Function of the Two-Mutation Clonal Expansion Model. *Risk Analysis* 1997; **17**(3): 391-9.

11. Knoke JD, Burns DM, Thun MJ. The change in excess risk of lung cancer attributable to smoking following smoking cessation: an examination of different analytic approaches using CPS-I data. *Cancer Causes & Control* 2008; **19**(2): 207-19.

Chapter 11

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

Nanda Horeweg*, Joost van Rosmalen*, Marjolein A. Heuvelmans, Carlijn M. van der Aalst, Rozemarijn Vliegenthart, Ernst Th. Scholten, Kevin ten Haaf, Kristiaan Nackaerts, Jan-Willem J. Lammers, Carla Weenink, Harry J. Groen, Peter van Ooijen, Pim A. de Jong, Geertruida H. de Bock, Willem Mali, Harry J. de Koning*, Matthijs Oudkerk*

***Drs. Horeweg and van Rosmalen contributed equally to this manuscript. Drs. de Koning and Oudkerk contributed equally to this manuscript.**

Published as:

Horeweg N, van Rosmalen J, Heuvelmans MA, et al.

The Lancet Oncology 2014; **15**(12): 1332-41.

Abstract

Background

The main challenge in CT screening for lung cancer is the high prevalence of pulmonary nodules and the relatively low incidence of lung cancer. Management protocols use thresholds for nodule size and growth rate to determine which nodules require additional diagnostic procedures, but these should be based on individuals' probabilities of developing lung cancer. In this prespecified analysis, using data from the Dutch-Belgian randomized lung cancer screening trial (NELSON) CT screening trial, we aimed to quantify how nodule diameter, volume, and volume doubling time affect the probability of developing lung cancer within 2 years of a CT scan, and to propose and evaluate thresholds for management protocols.

Methods

Eligible participants in the NELSON trial were those aged 50–75 years, who have smoked 15 cigarettes or more per day for more than 25 years, or ten 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 whose results were available from the national cancer registry database. We calculated lung cancer probabilities, stratified 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. The NELSON trial is registered at www.trialregister.nl, number ISRCTN63545820.

Findings

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 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% confidence interval (CI) 0.4%–0.8%]) or maximum transverse diameter smaller than 5 mm (0.4% [95% CI 0.2%–0.7%]), and not significantly different from participants without nodules (0.4% [95% CI

0.3%–0.6%], $p = 0.17$ and $p = 1.00$, respectively). Lung cancer probability was intermediate (requiring follow-up CT) if nodules had a volume of 100–300 mm³ (2.4% [95% CI 1.7%–3.5%]) or a diameter 5–10 mm (1.3% [95% CI 1.0%–1.8%]). Volume doubling time further stratified the probabilities: 0.8% (95% CI 0.4%–1.7%) for volume doubling times 600 days or more, 4.0% (95% CI 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% [95% CI 12.7%–18.1%]). The simulated ACCP management protocol yielded a sensitivity and specificity of 90.9% (95% CI 81.2%–96.1%), and 87.2% (86.4%–87.9%), respectively. A diameter-based protocol with volumetry-based nodule diameter yielded a higher sensitivity (92.4% [95% CI 83.1%–97.1%]), and a higher specificity (90.0% [95% CI 89.3%–90.7%). 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 higher specificity (94.9% [95% CI 94.4%–95.4%]).

Interpretation

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.

Funding

Zorgonderzoek Nederland Medische Wetenschappen and Koningin Wilhelmina Fonds.

Introduction

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

When CT screening for lung cancer first began, the accepted standard of practice was to regard all noncalcified pulmonary nodules as potentially malignant lesions requiring follow-up screening until proven stable for a period of 2 years.[5-7] Later, the Fleischner Society recommended that nodules of 4 mm in diameter or smaller in high-risk individuals (i.e., history of smoking or other known risk factors) required no further follow-up if the nodule was unchanged at a 12-month follow-up examination, because the risk of the nodule being malignant was less than 1%.[8] However, people with nodules 4–8 mm in size were still recommended to undergo two to three follow-up examinations over a period of 2 years. Individuals with nodules larger than 8 mm were recommended to undergo diagnostic workup, which consisted of more invasive diagnostic procedures.[8]

Recently, the results of the Early Lung Cancer Action Project (ELCAP)—which suggested raising of the threshold for initiation of follow-up CT examinations to nodules of 8 mm or larger—were supported by data from the National Lung Screening Trial (NLST).[9,10] However, the ELCAP analyses were limited to screen-detected lung cancers, and only false-positive values and time to diagnosis were taken into account when assessing new thresholds for nodule diameter. Increasing the protocol-screening thresholds for nodule diameter to determine which patients should undergo diagnostic follow-up reduces the potential harms of diagnostic procedures, exposure to ionizing radiation, and costs.[11,12] However, it might also decrease the sensitivity for cancerous nodules, thus, in turn, increasing lung cancer mortality, and so it is important to balance these potential benefits and harms.[4] Therefore, thresholds for negative, indeterminate, and positive screening results should be based on individual participants' probability of developing lung cancer, and should be assessed in

terms of sensitivity, specificity, number of required CT examinations, and number of required invasive diagnostic procedures. Recommendations of the latest American College of Chest Physicians (ACCP) guidelines for management of individuals with pulmonary nodules with a volume of 8 mm or larger were based on the consensus statement of the Fleischner Society.[8] This statement has not been formally validated, and alternative management strategies might yield an improved performance in terms of sensitivity, specificity, and the number of required follow-up scans. The NELSON trial is a randomized trial to assess whether low-dose CT screening with an increasing length of screening interval (1, 2, and 2.5 years) compared with no screening reduces lung cancer mortality.[13] We used data from NELSON to quantify the probability of developing lung cancer within 2 years of CT screening, based on measurements of lung nodule diameters, volumes, and volume doubling times. We used lung cancer probabilities to assess the nodule management protocol recommended by the ACCP, and to propose improved management protocols.[8,14]

Methods

Study design and participants

Details about the design and conduct of the NELSON trial have been reported previously.[13,15] Briefly, participants from four centers in the Netherlands and Belgium were enrolled and randomly assigned to receive low-dose CT screening or no screening. Eligible participants were adults aged 50–75 years, who had smoked 15 or more cigarettes per day for more than 25 years or ten or more cigarettes per day for more than 30 years, and were still smoking or had stopped smoking less than 10 years previously. People with self-reported moderate or bad health (with a questionnaire adapted from the SF-36 questionnaire), inability to climb two flights of stairs, bodyweight of 140 kg or more, current or past renal cancer, melanoma, breast cancer, or lung cancer diagnosed less than 5 years ago, or a chest CT examination less than 1 year ago, were excluded. All participants who were diagnosed with lung cancer were identified from the national cancer registry of the Netherlands. We included all Dutch participants who were randomly assigned to the screening group, who had attended at least one round of screening in the first two screening rounds at 1 and 2 years after baseline screening. We excluded Belgian participants because no data were yet available from the Belgian cancer registry. The NELSON trial was approved by the Dutch

Minister of Health and ethics boards at each participating center. All participants gave written informed consent for participation and evaluation of personal data from hospital charts.

Procedures

The protocol describing how CT screening was done in the NELSON trial has been previously published, and is summarized in the supplementary material of this Chapter, in the section: "NELSON nodule management protocol".[15] Briefly, CT screening was done with 16-detector CT scanners in a low-dose setting (effective radiation dose <0.4 milliSievert (mSv), <0.8 mSv and <1.6 mSv, dependent on bodyweight).[15] Datasets were derived from images of the thorax (slice thickness 1 mm, interval 0.7 mm) and analyzed with software for semi-automated volume measurements (LungCARE, Siemens, version Somaris/5 VB 10A-W).[15] For any CT screen-detected non-calcified nodules, semi-automatic volumetric software independently then measured volume and maximum transverse diameter. Hence, the diameters used in this study were not measured manually. In cases in which no volume (V) could be assessed (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:

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

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:

$$\text{Volume doubling time} = \frac{\ln(2)\Delta \mathrm{t}}{\ln(V_2) - \ln(V_1)}$$

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:

$$\text{Volume doubling time} = \frac{\ln(2)\Delta t}{3 * \ln(\frac{\text{MaxDiamXY}_2}{\text{MaxDiamXY}_1})}$$

in which Δt represents time in days between scans, and MaxDiamXY_1 and MaxDiamXY_2 are maximum diameters on the X–Y axis at first and second assessment.[15] 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. 2-year probability was chosen because it is the recommended follow-up time for indeterminate nodules.[8,14]

We predicted lung cancer risk in the 2 years following each screening round using regression analysis stratified with nodule characteristics as potentially predictive variables. Based on these outcomes, we designed nodule management protocols for both nodule volume and diameter. Participants without nodules or with a lung cancer probability not significantly different from those without nodules were classified as negative, and were not recommended to undergo intensified CT surveillance besides screening.[8] Participants with a significantly increased lung cancer probability (but less than about 5%; adopted from ACCP guidelines) 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.[14] Participants with a lung cancer probability of more than 5% were directly classified as positive, and recommended to undergo additional diagnostic procedures immediately (adopted from ACCP guideline for nodules with a 5% to 65% risk of malignancy).[14] Furthermore, the ACCP management protocol (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, otherwise negative); and additional diagnostic procedures for nodules larger than 8 mm (classified as positive).[8,12,14]

Outcomes

The primary endpoint of the NELSON trial is reduction of lung cancer mortality by 25% or more at 10 years after randomization.[13,16] The aim of this prespecified analysis was to quantify the probability of developing lung cancer within 2 years after screening, stratified by measured nodule diameters, volumes, and volume doubling times. The secondary aims were to model lung cancer risk using predictive variables, and to propose and assess thresholds for nodule management protocols.

Statistical analysis

Probabilities of developing lung cancer stratified by different nodule variables were calculated by dividing the number of individuals with cancer by the total number of participants. Differences between lung cancer probabilities were tested using with Fisher's exact test; 95% confidence intervals (CIs) were calculated using the Agresti-Coull method. To predict lung cancer risk in the 2 years after each screening round, we did logistic regression analysis using diameter, volume, volume doubling time, and multinodularity as potential predictor variables. The model only included participants whose largest nodule measured 50–500 mm^3 and who had one nodule or more growing in this volume range, because volume doubling time was available only for this subgroup. We accounted for non-linear effects of the predictor variables using fractional polynomials. For each predictor variable, we included two terms of the form X^K, with the value of K chosen from the set (–2, –1, –0.5, 0, 0.5, 1, 2, 3); X^o denoted a logarithmic transformation. The predictor variables in the final model and the non-linear transformations were chosen with backward elimination with a significance level of 5%, on the basis of the multivariable fractional polynomials algorithm.[17] We used a closed-test procedure to control the family-wise type I error rate in a situation with multiple testing.[18] The calibration of the model was assessed with the Hosmer-Lemeshow test. We estimated test characteristics of all three nodule management protocols using the detection method with a 1-year interval plus all lung cancers detected in the same screening round (described in the supplementary material of this Chapter, in the section: "Framework for evaluating alternative nodule management protocols"). 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 (described in the supplementary material of this Chapter, in the section: “Methods for estimating screening test characteristics”). All statistical tests were two-sided, used a significance level of 5%, and were done with Stata (version 12), R (version 2.15), and Microsoft Excel (2010). This trial is registered as an International Standard Randomized Controlled Trial at www.trialregister.nl, number ISRCTN63545820.

Panel: 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 January 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

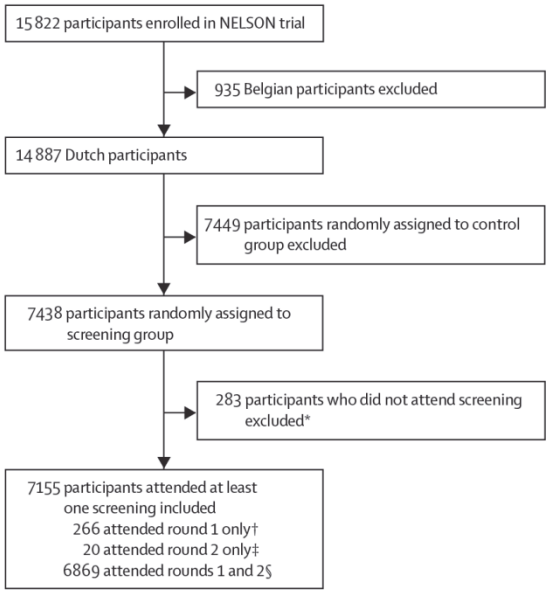
Published probabilities of lung cancer stratified by nodule size were similar to the probabilities estimated in our study. However, none of the published studies provided

estimates for such small ranges of diameter, as in our study. Moreover, no estimates of lung cancer probability were published for nodule volume and nodule volume doubling time. This prespecified 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.

Results

A total of 15,822 participants were enrolled in the NELSON trial between Dec 23, 2003, and July 6, 2006. Screening round one was conducted from January, 2004, to December, 2006, and screening round two from January, 2005, to September, 2008. For this study, we excluded 7,907 participants randomly assigned to the no screening group, 477 participants from Belgium (no data were yet available from the Belgian cancer registry), and 283 participants who did not attend their screening examinations (no screening test characteristics could be calculated in the absence of screening). Thus, we included 7,155 participants in our study (7,135 of whom received screening at the first screening round, and 6,889 of whom received screening at the second screening round; Figure 1).

Figure 1: Trial profile



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

Median length of available follow-up of the participants was 8.16 years (IQR 7.56–8.56). Median age was 58 years (IQR 50–66). 1,206 (16%) of 7,438 participants were women, 6,232 (84%) were men, 4,165 (56%) were current smokers, and their median number of pack-years at randomization was 38 (IQR 19–57). 2-year lung cancer probability for all included participants was 1.3% (95% CI 1.2%–1.5%; Table 1).

Table 1: Probability of lung cancer diagnosis within 2 years after a screening test, by nodule volume, diameter, and volume doubling time

	Round one	Round two	Rounds one and two	Lung cancer probability (95% CI)*	*p*-value†
All participants	109/7,135	79/6,889	188/14,024	1.3% (1.2%–1.5%)	<0.0001
No nodule detected	15/3,946	15/3,684	30/7,630	0.4% (0.3%–0.6%)	..
Nodule smaller or equal volume on second CT	3/476	3/430	6/906	0.7% (0.3%–1.5%)	0.27
Nodule resolved (i.e., not visible) on second CT	0/135	0/70	0/205	0.0% (0.0%–2.2%)	1.00
No follow-up CT, not referred to pulmonologist	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 50–500 mm³	33/1455	24/1,499	57/2,954	1.9% (1.5%–2.5%)	<0.0001
Volume of largest nodule (mm³)‡					
≥1000	36/137	26/104	62/241	25.7% (20.6%–31.6%)	<0.0001
750 to <1000	8/33	3/30	12/63	19.0% (11.1%–30.6%)	<0.0001
500 to <750	8/63	4/47	12/110	10.9% (6.2%–18.2%)	<0.0001
300 to <500	12/101	6/102	18/203	8.9% (5.6%–13.7%)	<0.0001
200 to <300	9/127	5/116	14/243	5.8% (3.4%–9.5%)	<0.0001
100 to <200	6/428	7/440	13/868	1.5% (0.9%–2.6%)	0.0002
50 to <100	6/800	6/843	12/1,643	0.7% (0.4%–1.3%)	0.07
25 to <50	6/961	4/1,008	10/19,69	0.5% (0.3%–0.9%)	0.44
<25	3/539	2/515	5/1,054	0.5% (0.2%–1.1%)	0.61
Maximum diameter of largest nodule (mm)§					
≥30	3/10	3/9	6/19	31.6% (15.2%–54.2%)	<0.0001
20 to <30	13/52	9/36	22/88	25.0% (17.1%–35.0%)	<0.0001
15 to <20	22/84	7/64	29/148	19.6% (14.0%–26.8%)	<0.0001
10 to <15	28/229	21/213	49/442	11.1% (8.5%–14.4%)	<0.0001
8 to <10	7/260	9/296	16/556	2.9% (1.7%–4.7%)	<0.0001
7 to <8	8/327	4/328	12/655	1.8% (1.0%–3.2%)	<0.0001
6 to <7	1/371	2/331	3/702	0.4% (0.1%–1.3%)	0.75
5 to <6	7/628	5/721	12/1,349	0.9% (0.5%–1.6%)	0.02669
4 to <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
Volume doubling time of fastest-growing nodule (days)¶					
<100	7/24	2/10	9/34	26.5% (14.4%–43.3%)	<0.0001
100 to <200	3/40	3/16	6/56	10.7% (4.7%–21.8%)	<0.0001
200 to <400	5/130	7/52	12/182	6.6% (3.7%–11.3%)	<0.0001
400 to <600	3/92	4/81	7/173	4.0% (1.8%–8.3%)	<0.0001
600 to <800	0/56	0/74	0/130	0.0% (0.0%–3.4%)	1.00
800 to <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%)	0.03450

*Unless otherwise stated data are n/N, where n refers to participants who were subsequently diagnosed with lung cancer and N refers to number of patients in category. *Probability of malignant disease within 2 years after a CT scan. †p-value refers to lung cancer risk for participants with nodules compared with participants without any nodules. ‡Volume of the largest non-calcified nodule; maximum diameter of the largest nodule in a participant. 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. ¶Maximum volume doubling time in participants whose largest nodule measured 50–500 mm³.*

Participants without any pulmonary nodule (7,630 [54%] of 14,024 screenings in rounds one and two combined) had a lung cancer probability of 0.4% (0.3%–0.6%). In all participants with CT-detected nodules, lung cancer probability was 2.5% (2.1%–2.9%), but individuals' probabilities depended strongly on nodule volume, diameter and 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% [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%]; $p = 1.00$), but was significantly increased for participants whose nodules measured 5–10 mm (1.3% [95% CI 1.0%–1.8%]; $p < 0.0001$), and participants whose nodules measured 10 mm or more (15.2% [95% CI 12.7%–18.1%]; $p < 0.0001$), who could be regarded as being at intermediate and high risk of developing lung cancer, respectively. The probability of being diagnosed with lung cancer within 2 years after CT scan 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 (i.e., disappeared by follow-up CT) nodules had a low probability of lung cancer (0.0% to 1.0%). Lung cancer probability was not significantly increased for participants with nodule volume doubling times of 600 days or more (0.8% [95% CI 0.4%–1.7%]; $p = 0.06$). Lung cancer probability was significantly increased for participants with nodule volume doubling times of 400–600 days (4.0% [1.8%–8.3%]; $p < 0.0001$), who could be regarded at low risk of developing lung cancer, and for participants with a nodule volume doubling time

of 400 days or fewer (9.9% [6.9%–14.1%]; *p* <0.0001), who could be regarded at high risk of developing lung cancer. Probabilities of developing lung cancer according to other categories of nodule volume and volume doubling time (e.g., volume doubling time ≤0, nodule resolved at follow-up, or no follow-up CT done and participants not referred for diagnostic workup) were done, but did not significantly differ from the probability for individuals without any pulmonary nodules (Table 1). We did logistic regression analyses to predict lung cancer probability; nodule diameter, volume, volume doubling time, and multinodularity were used as potential predictors. All four candidate predictors were significant univariate predictors (data not shown). Nodule volume, nodule volume doubling time, and multinodularity were also significant multivariate predictors (shown in Tables S1, S2a and S2b in the supplementary material of this Chapter). However, the relationship between multinodularity and lung cancer risk was ambiguous: for those participants whose nodules were growing and measured 50–500 mm^3, the relative proportion of participants with lung cancer decreased as the numbers of nodules per participant increased (shown Figure S1a in the supplementary material of this Chapter). However, in the total study population, the proportion of lung cancers varied as the amount of nodules per participant increased (shown in Figure S1b in the supplementary material of this Chapter). 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 analyze 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 ($p = 0.95$). In participants with nodules of 300 mm^3 in size or larger, the lung cancer probability was substantial (from 5.9% to >50%), even in case of slow nodule growth. In participants with nodules sized 100–300 mm^3, lung cancer probability ranged from less than 3% to 20%, dependent on the volume doubling time. Nodule management protocols, designed with either nodule volume or diameter thresholds based on lung cancer probability, or using the simulated ACCP management protocol, are presented in Table 2.

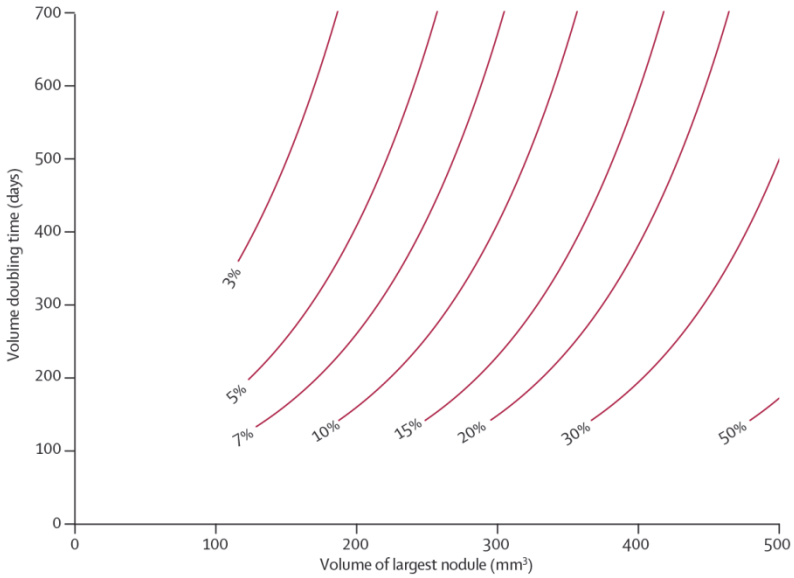
Table 2: Performance assessment of simulated nodule management protocols for CT-detected nodules at the first screening round

	Volume management protocol	Diameter management protocol†	ACCP simulated management protocol†
Screening result*			
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¶
Negative	Volume <100 mm^3	Diameter <5 mm	Diameter ≤4 mm
Screening test results			
Direct referral due to positive result (n = 7,135)	334 (5%)	375 (5%)	635 (9%)
Follow-up examination due to indeterminate result (n = 7,135)	555 (8%)	1,586 (22%)	2,125 (30%)
Positive results after follow-up examination (n = 7,135)	84 (1%)	394 (6%)	333 (5%)
Negative result after follow-up examination (n = 7,135)	471 (7%)	1,192 (17%)	1,792 (25%)
Detected lung cancers (n = 66)	60 (91%)	61 (92%)	60 (91%)
Screen test parameters			
Sensitivity	90.9%, 81.2%–96.1% (60/66)	92.4%, 83.1%–97.1% (61/66)	90.9%, 81.2%–96.1% (60/66)
Specificity	94.9%, 94.4%–95.4% (6,711/7,069)	90.0%, 89.3%–90.7% (6,361/7,069)	87.2%, 86.4%–87.9% (6,161/7,069)
Positive predicitve value	14.4%, 11.3%–18.1% (60/418)	7.9%, 6.2%–10.1% (61/769)	6.2%, 4.8%–7.9% (60/968)
Negative predictive value	99.9%, 99.8%–100.0% (6,711/6,717)	99.9%, 99.8%–100.0% (6,361/6,366)	99.9%, 99.8%–99.9% (6,161/6,167)

*Data are n (%) or %, 95% CI (n/N). ACCP = American College of Chest Physicians. The test characteristics were estimated with the detection method, with use of a 1-year interval plus all lung cancers detected in the same screening round (described in the supplementary material of this Chapter, in the section: "Methods for estimating screening test characteristics"). *In cases of multiple nodules, the size of the largest nodule determined the screening result. †Estimates based on diameters assessed with semi-automated volumetry. Although the lung cancer probability of nodules with volume doubling times of 400–600 days is intermediate (4.1% in 2 years), in this analysis we could not classify this subgroup as indeterminate because every participant needed to have definite screening test results (positive or negative) to determine whether lung cancer was detected by screening or not and to calculate the test characteristics of the screening protocol. §Participants with an indeterminate screening result should undergo a follow-up scan after 3 months to assess volume doubling time; a result of less than 600 days is a positive screening result and leads to referral for diagnostic work-up. ¶Participants with an indeterminate screening result should undergo a follow-up scan after 3 months to assess the volume doubling time; a result of less than 400 days is a positive screening result and leads to referral for diagnostic work-up, according to the ACCP guideline.[14]*

After the first screening round (for a 1-year interval), the protocol that used nodule volume had a sensitivity of 90.9% (95% CI 81.2%–96.1%), and a specificity of 94.9% (95% CI 94.4%–95.4%). Due to its high specificity, relatively few patients would have had follow up CT examinations (555 [8%] of 7,135) and additional diagnostic procedures (418 [6%]) compared to the other protocols. The protocol that used (volumetry-based) nodule diameter had a lower specificity than the volume protocol (90.0% [95% CI 89.3%–90.7%]), which would have led to more follow-up examinations (1,586 [22%]), and additional diagnostic procedures (769 [11%]), but had a slightly higher sensitivity for lung cancer (92.4% [95% CI 83.1%–97.1%]). The simulated ACCP protocol had a sensitivity of 90.9% (95% CI 81.2%–96.1%), and the lowest specificity of the three evaluated protocols (87.2% [95% CI 86.4%–87.9%], and would have led to the most follow-up CT examinations (2,125 [30%]) and additional diagnostic procedures (968 [14%]). Performance of the lung cancer probability-based volume and diameter protocols in the second screening round with the same thresholds is provided in Table S3 in the supplementary material of this Chapter.

Figure 2: Contour plot of the effect of the combined effect of nodule volume and volume doubling time on 2-year lung cancer probability



The risk isolines represent the percentage of NELSON participants that will be diagnosed with lung cancer within 2 years according to the volume of their largest nodule and volume doubling time of the fastest growing nodule in the 50–500 mm³ range.

Discussion

In this analysis, we used NELSON trial data to calculate the probability of developing lung cancer within 2 years for asymptomatic past or present smokers after a low-dose CT scan, and stratified this risk by nodule volume, diameter, and volume doubling time (panel). 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 with nodule volumes of 300 mm^3 or more, or diameters 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%.[7,14,19-23] Moreover, the 2-year probability of developing lung cancer in participants whose nodules measured 50–100 mm^3 or 4–5 mm was also low, and did not significantly differ from that in participants without nodules. At present, guidelines recommend two to four follow-up scans for such nodules.[8,14,24] Omission of these CT surveillance schedules for this patient population should be considered, because the risk of malignancy does not justify harms of ionizing radiation (effective dose estimated at 10 mSv per full-dose CT), psychological distress (clinically relevant increase in lung-cancer-specific distress as shown by van den Bergh and colleagues, and confusion, distress, and frustration as reported by Wiener and colleagues), and associated pressure on financial resources.[11,25-28]

Participants whose nodules measured 100–300 mm³ (or 5–10 mm in diameter) had a significantly higher 2-year lung cancer risk than did participants without nodules, which, according to current guidelines, justifies additional CT examinations.[8,24,25] Because lung cancer risk of participants with nodules between 5 mm and 8 mm is similar (0.9% to 1.8%), a uniform CT surveillance schedule could be applied, with volume doubling time assessed at CT surveillance used to reassess lung cancer probability.[21] 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 and return to regular screening.[1,2,8] 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 and additional diagnostic procedures.[1,8]

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 2 years. Hence, a follow-up CT scan at short notice (3–6 months depending on whether volumetry is available) 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 and 8.9% or higher for volumes of 300 mm³ to less than 500 mm³, or to 11.1% for diameters of 10 mm to less than 15 mm in this study. Risk for these large nodules remained high even when they grew slowly (Figure 2).[8,19,22,29] However, risk of developing lung cancer for participants with large nodules that had shrunken or resolved were diagnosed with lung cancer within 2 years was very low. Although classification of large slow-growing nodules as highly likely to be 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 individual increased. These findings contradict those of McWilliams and colleagues, 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.[23] 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 with protocols that use just one nodule evaluation (e.g., as used in the ELCAP and the NLST trials) is that a single low-dose CT examination is given at short notice (e.g., 3 months) for indeterminate nodules, instead of two to three CT scans in 2 years.[4,7,8] This approach allows for a better risk stratification by nodule volume doubling time, which is a strong lung cancer predictor.[3,5,14]

The protocol that used lung cancer probability-based diameter thresholds was more sensitive than the simulated ACCP protocol and would have led to fewer CT examinations and additional diagnostic procedures.[8,13] Nonetheless, these results imply that the simulated ACCP nodule management protocol performs well, but improvements are possible.

The protocol that used lung cancer probability-based thresholds for nodule volume had high specificity, and would have led to substantially fewer follow-up CT examinations and additional diagnostic procedures than would the simulated ACCP protocol. Moreover, this protocol was as sensitive as the simulated ACCP protocol. However, if manual diameter measurements had been used instead of volumetry-based measurements—as recommended in the ACCP protocol—it is unlikely that such high sensitivity values would have been reached due to the intrinsic unreliability of manual measurements.[30] We believe that the advantages of an increase in specificity of the volume protocol indicate that lung cancer screening should be done 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, and accounts for the fact that some interval cancers could not be matched to a nodule previously detected by screening.[14] 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 2-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 the 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 2 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%) and a comparable lung cancer risk (about 1.3% in 2 years).[3]

Lastly, presented lung cancer probabilities, volume doubling times, and nodule protocols were all estimated and evaluated using a data set of nodule measurements that were mainly assessed using volumetry. Evaluation of two nodule management protocols using diameter was done under the assumption that nodule diameters measured using semi-automatic volumetry software were comparable to manually measured nodule diameters. However, measurement error of manual measurement of nodule diameter is larger than measurement error of the volumetry-based diameters we used in this study.[30-34] Further, calculations of volume doubling time based on manually measured nodule diameters are less accurate than calculations of volume doubling time based on semi-automated volumetry. As a result, the relationship between nodule diameter and lung cancer probability may be weaker for manually measured nodule diameters. In addition, when results of this study are applied to manually measured diameters, presented sensitivities and specificities of protocols using diameter are likely to be too high, and the false-positive rate, number of follow-up CTs, and diagnostic work-ups are likely to be too low. These discrepancies could be reduced by using the mean transverse nodule diameter instead of maximal nodule diameter. Nonetheless, the aforementioned theoretical discrepancies in lung cancer probability and performance characteristics are probably limited in practice, as our estimates of lung cancer probability are comparable to the probabilities published by the ELCAP, NLST, and the Pan-Canadian Early Detection of Lung Cancer Study, which used manual measurements of nodule diameters for analyses.[9,10,22,23] 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 semi-automatic 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.[23] 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.

Acknowledgments

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

References

1. Detterbeck FC, Mazzone PJ, Naidich DP, Bach PB. Screening for Lung Cancer. *CHEST* 2013; **143**(5): e78S-e92S.
2. Wender R, Fontham ETH, Barrera E, et al. American Cancer Society lung cancer screening guidelines. *CA: A Cancer Journal for Clinicians* 2013; **63**(2): 106-17.
3. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of Lung Nodules Detected by Volume CT Scanning. *New England Journal of Medicine* 2009; **361**(23): 2221-9.
4. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
5. Ost D, Fein AM, Feinsilver SH. The Solitary Pulmonary Nodule. *New England Journal of Medicine* 2003; **348**(25): 2535-42.
6. Tan BB, Flaherty KR, Kazerooni EA, Iannettoni MD. The Solitary Pulmonary Nodule. *CHEST* 2004; **123**(1): 89S-96S.
7. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *The Lancet*; **354**(9173): 99-105.
8. MacMahon H, Austin JHM, 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.
9. Henschke CI, Yip R, Yankelevitz DF, Smith JP, for the International Early Lung Cancer Action Program I. Definition of a positive test result in computed tomography screening for lung cancer: A cohort study. *Annals of Internal Medicine* 2013; **158**(4): 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; **273**(2): 591-6.
11. Linet MS, Slovis TL, Miller DL, et al. Cancer risks associated with external radiation from diagnostic imaging procedures. *CA: A Cancer Journal for Clinicians* 2012; **62**(2): 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. *European Journal of Cardio-Thoracic Surgery* 2012; **42**(3): 420-9.
13. 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 randomized lung cancer multi-slice CT screening trial (NELSON). *International Journal of Cancer* 2007; **120**(4): 868-74.
14. Gould MK, Donington J, Lynch WR, et al. Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer? *CHEST* 2013; **143**(5): e93S-e120S.
15. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomized lung cancer screening trial. *Lung Cancer* 2006; **54**(2): 177-84.
16. Horeweg N, van Klaveren RJ, Groen HJM, et al. Blinded and uniform cause of death verification in a lung cancer CT screening trial. *Lung Cancer*; **77**(3): 522-5.
17. Royston P, Altman DG. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modeling. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 1994; **43**(3): 429-67.
18. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modeling continuous variables. Hoboken, NY: Wiley, 2008.
19. Midthun DE, Swensen SJ, Jett JR, Hartman TE. O-127 Evaluation of nodules detected by screening for lung cancer with low dose spiral computed tomography. *Lung Cancer* 2003; **41**: S40.
20. 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.
21. Swensen SJ, Jett JR, Hartman TE, et al. Lung Cancer Screening with CT: Mayo Clinic Experience. *Radiology* 2003; **226**(3): 756-61.

22. The National Lung Screening Trial Research Team. Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. *New England Journal of Medicine* 2013; **368**(21): 1980-91.
23. McWilliams A, Tammemägi MC, Mayo JR, et al. Probability of Cancer in Pulmonary Nodules Detected on First Screening CT. *New England Journal of Medicine* 2013; **369**(10): 910-9.
24. 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.
25. Wiener RS, Gould MK, Woloshin S, Schwartz LM, Clark JA. What Do You Mean, a Spot? *CHEST* 2013; **143**(3): 672-7.
26. van den Bergh KAM, Essink-Bot ML, Borsboom GJJM, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *British Journal of Cancer* 2009; **102**(1): 27-34.
27. 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 Affairs* 2012; **31**(4): 770-9.
28. Villanti AC, Jiang Y, Abrams DB, Pyenson BS. A Cost-Utility Analysis of Lung Cancer Screening and the Additional Benefits of Incorporating Smoking Cessation Interventions. *PLOS ONE* 2013; **8**(8): e71379.
29. Gohagan J, Marcus P, Fagerstrom R, Pinsky P, Kramer B, Prorok P. Baseline Findings of a Randomized Feasibility Trial of Lung Cancer Screening With Spiral CT Scan vs Chest Radiograph. *CHEST* 2004; **126**(1): 114-21.
30. Revel M-P, 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.
31. Gavrielides MA, Kinnard LM, Myers KJ, Petrick N. Noncalcified Lung Nodules: Volumetric Assessment with Thoracic CT. *Radiology* 2009; **251**(1): 26-37.
32. Reeves AP, Biancardi AM, Apanasovich TV, et al. The Lung Image Database Consortium (LIDC). *Academic Radiology* 2007; **14**(12): 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. *European Radiology* 2013; **23**(1): 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. *European Radiology* 2006; **16**(4): 781-90.

Chapter 11

Supplementary material

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 <50mm³
- Solid pleural based nodules <5mm in minimal diameter
- Part-solid nodules, non-solid component <8mm in mean diameter
- Part-solid nodules, solid component <50mm³
- Non-solid nodules <8mm in mean diameter

NODCAT 3:

- Solid nodules 50-500mm³
- Solid pleural based nodules 5-10mm in minimal diameter
- Part-solid nodules, non-solid component ≥8mm in mean diameter
- Part-solid nodules, solid component 50-500mm³
- Non-solid nodules ≥8mm in mean diameter

NODCAT 4:

- Solid nodules >500mm³
- Solid, pleural based nodules >10mm in minimal diameter
- Part-solid nodules, solid component >500mm³

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: <50mm³ was negative, 50-500mm³ was indeterminate, and >500mm³ was positive. For previously detected nodules, VDT was calculated and determined the screening result: >600 days was negative, 400-600 days was indeterminate and <400 days was positive. The protocol allowed radiologists to adjust the screening result in case of inaccurate measurements by LungCARE, high suspicion of malignancy (e.g. new solid component in non-solid nodule), or high suspicion of 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: $\mathrm{V} = \frac{1}{6}\pi D^3$).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.

1. 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.

2. 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.

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

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

1. For the evaluation of the simulated ACCP algorithm: participants with at least one indeterminate nodule with a VDT ≤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.

2. If the VDT of a nodule could not be calculated because the nodule had not grown or was not visible on the follow-up scan, this did not lead to a decision to refer the participant. If the VDT could not be calculated because no follow-up scan had been made in the NELSON trial, the decision to refer the patient was imputed using the referral decision made by the radiologists in the NELSON trial. This approach was necessary in approximately 15% of the subjects with the largest nodule in the 50-500mm³ range, e.g. due to manual adjustments of the screening result by the radiologists. The referral decisions made in the NELSON trial were based on the aforementioned formal NELSON protocol.

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:

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

- A lung cancer diagnosis was made during the period from the first CT examination of the screening round to either the next screening round or one year later, whichever came first (false negative screening results).

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

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

All screening test parameters were presented with 95% binomial confidence intervals (95% CI), which were calculated using the Agresti-Coull method.

Lung cancer diagnoses not confirmed by histological specimens

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

1. Tumor in the right upper lobe, volume 1,502mm³, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to cardiac impairment.
2. Tumor in left lower lobe, volume 2,687mm³, PET positive, cT1aN0M0, patient did not undergo thoracic surgery due to COPD stage IV.
3. Tumor in left lower lobe, volume 2,792mm³, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to COPD and renal failure.
4. Tumor in right upper lobe, volume 580mm³, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to metastasized prostate carcinoma.
5. Tumor in right lower lobe, volume 2,793mm³, PET-positive, cT1bN0M0, patient did not undergo thoracic surgery due to poor pulmonary function.
6. Tumor in right upper lobe, volume 891mm³, PET indeterminate, cT1aN0M0, patient died due to bowel ischemia just before intended thoracic surgery.
7. Tumor in right lower lobe, volume 731mm³, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery due to poor pulmonary function.
8. Tumor in left lower lobe, volume 108mm³, VDT 125 days, PET-positive, cT1aN0M0, patient did not undergo thoracic surgery because he also participated in another study and was randomized to the radiotherapy treatment arm.
9. Tumor in right upper lobe, volume 383mm³, VDT 289 days, PET indeterminate, cT1aN0M0, patient did not undergo thoracic surgery because he refused, he was treated with stereotactic radiotherapy instead.
10. Tumor in left lower lobe, volume 1,108mm³, PET positive, cT1aN1M0, patient did not undergo thoracic surgery due to poor pulmonary function.
11. Tumor in left lower lobe, diameter 10mm, PET positive, cT1aN0M0, patient did not undergo thoracic surgery due to poor pulmonary function.
12. Tumor in right upper lobe, diameter 13.2x11.6mm, PET positive, cT1aN0M0, patient did not undergo thoracic surgery due to poor pulmonary function and general condition
13. Tumor in right upper lobe, diameter 19.2x12.7mm, PET positive, cT1bN0M0, patient did not undergo thoracic surgery due to poor general condition

Table S1: 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)

Abbreviations: VDT = volume-doubling time, 95% CI = 95% confidence interval using the Agresti-Coull method
In this model, only the participants in whom the largest detected nodule had a volume of ≥50mm³ and <500mm³ 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.*
^ 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
** p-value < 0.001*

Table S2a: 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

Abbreviations: VDT = volume-doubling time, 95% CI = 95% confidence interval using the Agresti-Coull method
In this model, only the participants in whom the largest detected nodule had a volume of ≥50mm³ and <500mm³ 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.
^ 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 S2b: 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)	0.180

Abbreviations: VDT = volume-doubling time, 95% CI = 95% confidence interval using the Agresti-Coull method
In this model, only the participants in whom the largest detected nodule had a volume of ≥50mm³ and <500mm³ 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.
^ 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)*

Figure S1a: Relationship multi-nodularity and lung cancer probability in subjects whose largest measure 50-500mm³ and have a VDT>0

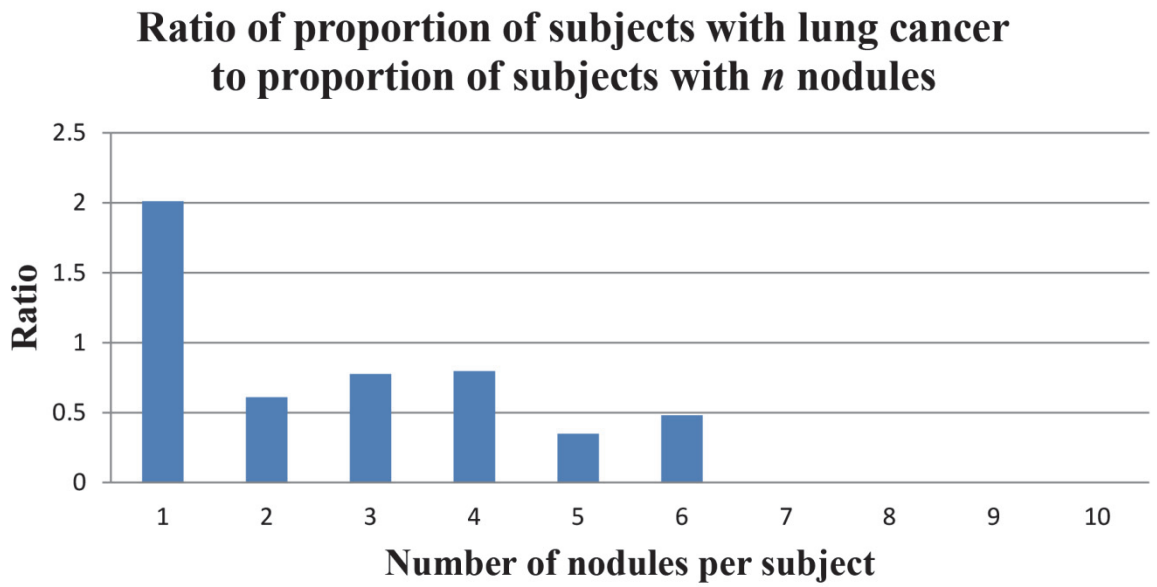


Figure 1b: Relationship multi-nodularity and lung cancer probability in all subjects with nodules

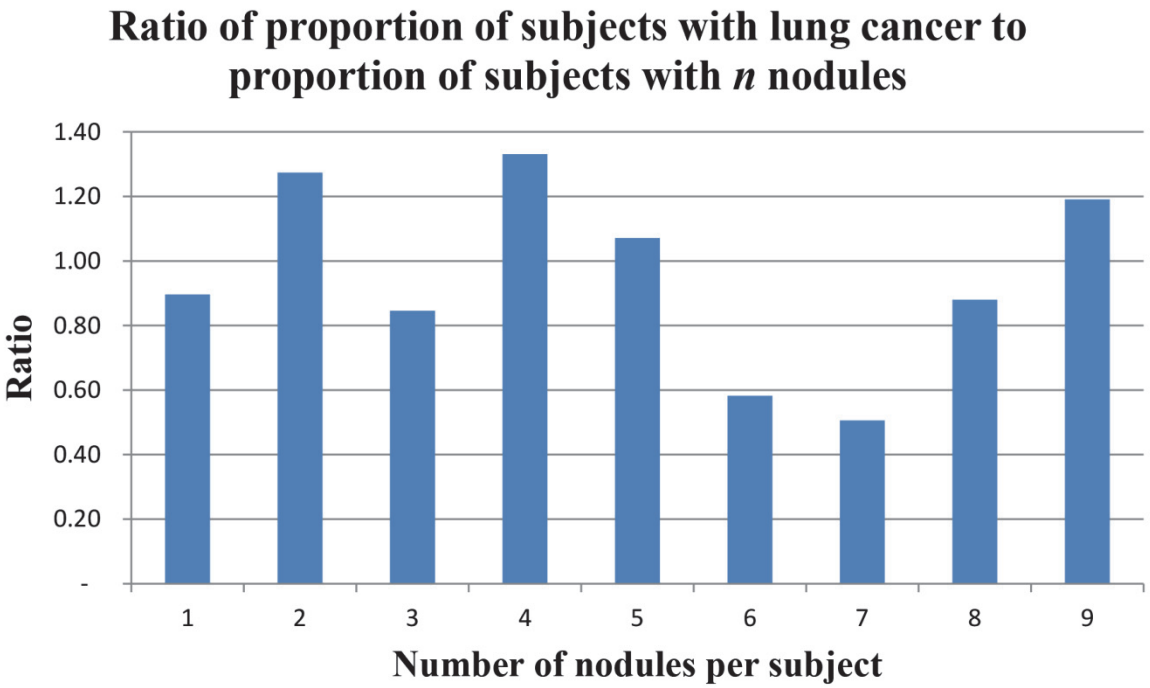


Table S3: Performance evaluation of simulated management algorithms for CT-detected nodules at the second screening round

	Volume management protocol	Diameter management protocol†	ACCP simulated management protocol†
Screening result*			
Positive	Volume ≥300 mm³	Diameter ≥10 mm	Diameter ≥8 mm
Indeterminate	Volume ≥100 to ≤300 mm³‡	Diameter ≥5 to <10 mm‡	Diameter >4 to <8 mm ‖
Negative	Volume <100 mm³	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)	
Follow-up examination due to indeterminate result	8.1 (556/6,889)	24.3 (1676/6,889)	
Positive results after follow-up examination	1.0 (72/6,889)	5.4 (370/6,889)	
Negative result after follow-up examination	7.0 (484/6,889)	19.0 (1,306/6,889)	
Detected lung cancers	83.1 (49/59)	86.4 (51/59)	
Screen test parameters	**Percentage [95%CI] (N/N)**	**Percentage [95%CI] (N/N)**	**Percentage [95%CI] (N/N)**
Sensitivity	83.1 [71.3–90.7] (49/59)	86.4 [75.2–93.2] (51/59)	88.1 [77.2–94.4] (52/59)
Specificity	95.5]95.0–96.0] (6,524/6,830)	90.6 [89.9–91.3] (6,830/6,189)	88.6 [87.8–89.3] (6,050/6,830)
Positive predictive value	13.8 [10.6–17.8] (49/355)	7.4 [5.6–9.6] (51/692)	6.3 [4.8–8.1] (52/832)
Negative predictive value	99.8 [99.7–99.9] (6,524/6,534)	99.9 [99.7–99.9] (6,189/6,197)	99.9 [99.8–99.9] (6,050/6,057)

Abbreviations:: 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.*
‖ 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 section: "Methods for estimating screening test characteristics")

General discussion

In this thesis, the effects of the implementation and optimization of lung cancer screening through risk stratification were investigated. This section will address the main findings pertaining to the research questions formulated in the Introduction of this thesis. This is followed by an assessment of the methodological considerations of the studies in this thesis. Then, directions for future research are provided. Finally, the main conclusions of this thesis and recommendations for further research and practice are provided.

Main Findings

Part 1: Development of a lung cancer screening model

In **Chapter 1**, five modeling groups independently developed lung cancer screening models based on information from the National Lung Screening Trial (NLST) and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).[1,2] While the models were based on different assumptions and had unique model structures, all were able to reproduce the results of the NLST and the PLCO. Initially, all models were first calibrated to the NLST population, which consisted of higher-risk individuals compared to those in the PLCO, due to the differences in criteria for enrollment based on smoking history between the two trials. However, while all models validated well to individuals in the PLCO who met the NLST eligibility criteria, further calibration to the full PLCO population was required to reproduce the lung cancer outcomes in never- and lighter smokers. This indicates that it is crucial to incorporate information from populations at different levels of lung cancer risk to allow extrapolation of the results of the NLST to the general population.

The results shown in **Chapter 2** provided further support for utilizing information from diverse sources. In this chapter, modeling was used to derive estimates for the natural history and the detectability of lung cancer based on information from the NLST and the PLCO. This information cannot be derived directly from observed data, as the onset and progression of preclinical disease is unobservable. In modeling, the estimates for the natural history (sojourn time) and the detectability (sensitivity) have a statistical dependence on each other. This causes uncertainty in the estimates of these parameters, but this uncertainty can be limited by calibrating a model to high quality data on disease outcomes

in the presence and absence of screening.[3] While the NLST provides information on the efficacy of computed tomography (CT) screening, it did not provide information on the natural history of lung cancer in the absence of screening, as chest radiography (CXR) screening occurred in the control arm. However, the PLCO had a non-screened control arm, which did provide information on the natural history of lung cancer in the absence of screening. Through jointly calibrating the model to data from the NLST and the PLCO, it was possible to derive natural history and screen-detectability estimates which were consistent with lung cancer outcomes in the presence and absence of screening. Through synthetization of the information from the NLST and the PLCO, the main differences in sensitivity between CT and CXR were found in the early stages, in particular stage IA. This could partially explain the difference in lung cancer mortality reduction between the CT and CXR arms of the NLST, as lung cancer survival drastically differs by stage; even between stages IA and IB.[4] Furthermore, while the overall sojourn time of lung cancer was estimated to be longer compared to previous research, the window of opportunity to detect lung cancer in stage IA was suggested to be more difficult with biennial or triennial screening intervals compared to annual screening intervals.[5-7] This suggests annual screening may yield higher mortality reductions compared to biennial or triennial screening.

An indication on how modeling can provide information on the occurrence of overdiagnosis beyond what was provided by the observed data of a randomized trial was shown in **Chapter 3**. Excess incidence analyses which compared the observed data of the NLST suggested that 18.5% of all cancers detected by CT were overdiagnosed.[5] However, the control arm of the NLST did not provide an unbiased baseline estimate for lung cancer in the absence of screening, as CXR screening occurred in this arm. Thus, an excess incidence analysis may underestimate the magnitude of overdiagnosis in the CT arm of the NLST. On the other hand, the median follow-up duration of NLST participants was limited to approximately 6.5 years, at which the incidence rates of the two arms had not yet converged. This suggests that the follow-up duration of the NLST was not of sufficient length to account for the difference in lead-time between CT and CXR screening. Modeling was used to extrapolate the follow-up period of the NLST to a lifetime period, using the estimates for the natural history and the detectability of lung cancer derived in Chapter 2. This modeling extrapolation estimated that 6.75% of all screen-detected cases in the CXR

arm and 8.62% of all screen-detected cases in the CT arm of the NLST were overdiagnosed. Given that 75% of the participants in the NLST were younger than 65 and over 50% of all cancers in the CT-arm were detected in this group, the potential for overdiagnosis in the NLST was relatively low. In contrast, the part of the U.S. population that met the NLST eligibility criteria was older and more likely to be a current smoker, suggesting that the potential for overdiagnosis by lung cancer screening in the general U.S. population may be higher.[8] However, the information derived from Chapters 1 and 2 allows extrapolation of the NLST's results to the general U.S. population and estimation of the potential for overdiagnosis by lung cancer screening in the general U.S. population.

The information derived from Chapters 1-3 thus provides the following answer to the research question:

What additional insights can models provide beyond observed trial data of randomized trials?

High quality models can synthesize information from populations with different levels of risk and different screening histories to derive estimates for unobservable processes, such as the natural history of a disease in the presence and absence of screening. This allows one to extrapolate the results of a randomized trial beyond its observed follow-up duration and to different designs and populations other than those considered in the trial.

Part 2: The long-term benefits and harms of lung cancer screening policies

In Chapter 1, five different models were developed based on information of the NLST and the PLCO. In **Chapter 4**, it was shown how these models can extrapolate the results of the NLST beyond the single design considered in the trial. The five models evaluated 576 different screening scenarios, which considered different variations of starting and stopping ages of screening, required smoking eligibility criteria, and different screening intervals. For each scenario, the required number of screening examinations and the number of lung cancer deaths averted were evaluated. While the models differed in their absolute estimates for the number of lung cancer deaths averted for a given scenario, they did rank the scenarios similarly with regards to their efficiency. In total, the models identified 120 'consensus efficient scenarios' which averted the greatest number of lung cancer deaths for

a given number of CT screens across all models. All models agreed that the scenario closest to the NLST criteria (annual screening from ages 55 through 75 for current and former smokers who smoked a minimum of 30 pack-years and quit smoking less than 15 years ago) was not among the consensus efficient scenarios, suggesting that the criteria used in the NLST may not be the most optimal for a population based screening program.

In **Chapter 5**, the long-term benefits and harms of 26 of the 'consensus efficient scenarios' identified in Chapter 4, along with a 27th scenario closest to the NLST criteria, were further evaluated to inform the United States Preventive Services Task Force (USPSTF) on their recommendations for lung cancer screening. The scenarios that considered annual screening provided substantially higher reductions in lung cancer mortality (11.0%-21.2%) compared to biennial (6.5%-9.6%) or triennial screening (4.6%-6%), similarly to what the estimates for the natural history and the detectability of lung cancer derived in Chapter 2 suggested. Overall, annual screening for individuals aged 55 through 80, who smoked at least 30 pack-years and currently smoke or quit for less than 15 years was judged to provide an advantageous balance between benefits and harms. This scenario is similar to the criteria used in the NLST, with the exception of extending the stopping age for screening to age 80. This scenario would reduce lung cancer mortality by 14%, avert 497 lung cancer deaths and yield 10.6 life-years per averted lung cancer death, but require 286,813 CT screening examinations and lead to 190 overdiagnosed cases (9.9% of all screen-detected cases) per 100,000 individuals. On average, each eligible individual would receive 14.9 screens, of which 3.5 would be false-positive (23.5%). The USPSTF currently recommends lung cancer screening in the U.S. using the eligibility criteria of this scenario.[9]

While various societies endorsed the implementation of lung cancer screening in the U.S., concerns were raised with regards to extending the stopping age for screening to 80, in part due to the potential for overdiagnosis in older individuals. For example, the Centers for Medicare & Medicaid Services (CMS) chose to cover CT lung cancer screening for individuals under Medicaid up to age 77, while the American Cancer Society recommended screening up to age 74.[10,11] Therefore, **Chapter 6** assessed how overdiagnosis affects the selection of efficient lung cancer screening strategies. Overdiagnosis was found to increase for shorter screening intervals, older starting and stopping ages for screening, and for increases in the

accumulated pack-years required for screening eligibility. The results from Chapters 2 and 5 indicated that annual screening yields a greater reduction in lung cancer mortality compared to biennial or triennial screening; the results of Chapter 6 indicate that this benefit should be weighed against an increase in overdiagnosis. Efficient scenarios which maximized the number of lung cancer deaths averted favored a stopping age of 80. In contrast, efficient scenarios which maximized the number of lung cancer deaths averted per overdiagnosed case favored a stopping age of 75. The scenario closest to the NLST criteria reduced lung cancer mortality by 16.9%, while 11.9% of all screen-detected cases would be overdiagnosed, averting 3.49 lung cancer deaths per overdiagnosed case. The scenario recommended by the USPSTF reduced lung cancer mortality by 14.7%, while 9.7% of all screen-detected cases would be overdiagnosed, averting 2.85 lung cancer deaths per overdiagnosed case. This suggests that raising the stopping age from 75 to 80 yields a higher lung cancer mortality reduction, at the cost of a proportionally higher increase in overdiagnosis. However, both of these scenarios compare favorably to the estimated ratios for prostate cancer screening (0.2 prostate cancer deaths averted per overdiagnosed case) and breast cancer screening (an independent U.K panel suggests that 0.3 breast cancer deaths were averted per overdiagnosed case; however, modeling analyses for the Netherlands suggests that 2.95 and 3.41 breast cancer deaths were averted per overdiagnosed case by film-based and digital mammography ,respectively).[12-14] This suggests that the negative impact of overdiagnosis is lower for lung cancer screening compared to prostate and breast cancer screening programs.

The research described in Chapters 4-6 provided estimates for the benefits and harms of lung cancer screening. The recommendations of the USPSTF and CMS led to the implementation of various lung cancer screening programs in the U.S..[9,15] However, various aspects with regards to the optimal implementation of lung cancer screening remained uncertain, which were discussed in **Chapter 7**. Firstly, several studies suggested that the selection of individuals based on risk prediction models could aid in further optimizing lung cancer screening. However, the implementation of a large scale screening program based on risk-stratification has had little to no consideration. It is yet unclear how to appropriately reach the population at highest risk for lung cancer, as this population may be harder to

reach and less likely to participate in screening.[16] In addition, the unintentional screening of low-risk individuals should be prevented to retain the efficiency of the screening program.

Secondly, while the NLST found a reduction in lung cancer mortality, the NLST investigators noted concerns about the high proportion of false-positive results.[2] Overall, 96.4% of all positive CT screens consisted of false-positive results, accounting for 23.3% of all CT screens.[2] While different nodule management protocols have been proposed that could reduce the number of false-positive results, this may come at a loss of sensitivity for detecting lung cancer.[17] Furthermore, currently only few evidence-based guidelines incorporate the suggestions of these newly proposed nodule management protocols, while this guidance is essential to aid radiologists in evaluating pulmonary nodules in a lung cancer screening program.[18-20]

The information derived from Chapters 4-7 provides the following answer to the research question:

What are the long-term benefits and harms of lung cancer screening policies, and the potential barriers for the implementation of these policies?

Annual screening for individuals aged 55 through 80, who smoked at least 30 pack-years and currently smoke or quit less than 15 years ago provides an advantageous balance between benefits and harms. This would reduce lung cancer mortality by 14.0% and yield 10.6 life-years per averted lung cancer death. However, 23.5% of all screens would be false-positive and 9.9% of all screen-detected cases would be overdiagnosed. The potential barriers for implementing lung cancer screening are: 1) how to implement a large scale screening program based on risk-stratification and 2) how to reduce the number of false-positive results while maintaining a high sensitivity.

Part 3: Optimization through risk stratification

Chapter 7 noted that several studies proposed using risk prediction models for the selection of individuals compared to the age and smoking related eligibility criteria recommended by the USPSTF. Some of these risk prediction models allow the estimation of lung cancer risk for never-smokers, in whom 10-25% of all lung cancers occur.[21,22] However, while the

majority of lung cancers occur in ever-smokers, lung cancer in never-smokers still ranks among the 10 most common causes of death due to cancer in the U.S. and worldwide.[22-24] This raises the question whether never-smokers at elevated risk for lung cancer may benefit from lung cancer screening, which was investigated in **Chapter 8**. MISCAN-Lung was used to evaluate the benefits and harms of lung cancer screening for cohorts of never-smokers at different levels of relative risk for lung cancer (relative risks of 2, 5, 10, 15, 20, 35) compared with average risk never-smokers. The ratio between benefits and harms for these cohorts of never-smokers were then compared to those for a cohort of ever-smokers eligible according to the USPSTF criteria. Overall, the relative reduction in lung cancer mortality was higher for the never-smoker cohorts than for the USPSTF eligible cohort (37% across all levels of relative risk compared with 32%). This may be due to the higher proportion of adenocarcinomas in never-smokers compared to ever-smokers.[25] Chapter 2 indicated that adenocarcinomas have a longer sojourn time, and thus a greater window of opportunity to be detected at an early stage by CT screening compared with other histological types of lung cancer. However, never-smokers also develop lung cancer at a later age compared to ever-smokers. As a result, the number of life-years gained per lung cancer death averted was lower (10.4 compared with 11.9) and the proportion of overdiagnosed cases was higher (9.6% compared with 8.4%) for the cohorts of never-smokers compared with the USPSTF eligible cohort. The overall level of relative risk required for never-smokers to have a similar trade-off between benefits and harms compared to those of the USPSTF eligible cohort was at least 15. However, the majority of the evaluated risk prediction models consider relative risks lower than 15 for never-smokers, suggesting that for the majority of never-smokers lung cancer screening is not beneficial.

The USPSTF and CMS recommendations for implementing lung cancer screening in the U.S., led to discussions on whether and how to implement lung cancer screening in other nations.[9,15,26-28] However, while the analyses shown in Chapters 4 and 5 aided in informing the USPSTF recommendations, they did not consider the cost-effectiveness of lung cancer screening. Although lung cancer screening in the NLST was reported to be cost-effective by U.S. standards, it was uncertain whether and how population-based lung cancer screening programs could be implemented in a cost-effective manner.[29-31] Therefore, **Chapter 9** evaluated the benefits, harms and costs of lung cancer screening in Ontario, Canada, to

inform Cancer Care Ontario on whether implementing lung cancer screening could be cost-effective. The MISCAN-Lung model developed in Chapters 1 and 2 was used to perform these evaluations, using Ontario-specific data on population characteristics, smoking behavior and costs to adapt the model to the population of Ontario. MISCAN-Lung was then used to examine 576 CT screening scenarios varying by age to start and end screening, smoking eligibility criteria, and screening interval, for persons born between 1940-1969. Overall, annual screening between ages 55-75 for persons who smoked at least 40 pack-years, and currently smoke or quit less than 10 years ago, was considered to be the optimal policy to implement. This policy yielded an incremental cost-effectiveness ratio (ICER) of $41,136 Canadian dollars per life-year gained. While Canada has no official cost-effectiveness threshold, this ICER is below the cost-effectiveness threshold of $50,000 Canadian dollars per life-year gained used in other studies.[32,33] If this policy were to be implemented, 9.56% (499,261 individuals) of the examined cohorts (ever- and never-smokers) would be screened at least once, which would require 4,788,523 CT examinations. However, this would reduce lung cancer mortality in these cohorts by 9.05% (preventing 13,108 lung cancer deaths). The ICER for this scenario ranged from $28,630-$73,507 Canadian dollars per life-year gained across different sensitivity analyses. The lower and upper bound of this range corresponded to assuming 50% lower and higher CT examination costs, respectively. Furthermore, major variations in cost-effectiveness were found between smoking eligibility criteria and screening intervals. Scenarios which applied stringent smoking eligibility criteria were more cost-effective than scenarios that applied less stringent smoking eligibility criteria. Chapter 6 indicated that while individuals with more extensive smoking histories are at higher risk for developing lung cancer, they are also at higher risk for overdiagnosis. The results of Chapter 9 indicate that while applying stringent smoking eligibility criteria may yield higher levels of overdiagnosis compared to less stringent criteria, this is essential to implement lung cancer screening in a cost-effective manner. Furthermore, annual screening intervals were found to be more cost-effective then biennial screening intervals, even though sensitivity analyses indicated that the costs of the CT examinations had the greatest impact on the cost-effectiveness of screening. This may be explained by the estimates for the sojourn time of lung cancer, explored in Chapter 2. Biennial screening intervals may not provide a sufficient window of opportunity to detect lung cancer in stage IA compared to stage IB. As mentioned previously, lung cancer survival

varies greatly by the stage at detection, even between stages IA and IB.[4] Thus, the greater window of opportunity provided by annual screening is suggested to outweigh the costs of the additional number of CT examinations compared with biennial screening.

Chapter 7 noted that various risk prediction models had been proposed to allow the selection of high-risk individuals for lung cancer screening. However, external validation and direct comparisons of these models had been limited. In addition, few models suggested risk thresholds (the level of risk used to classify predictions as positive or negative for the predicted outcome) to determine the eligibility of individuals, which is essential for implementation of such models in clinical practice. Therefore, nine risk prediction models were externally validated and directly compared in **Chapter 10**. The performance of each model to predict the risks for 5- and 6-year lung cancer incidence and lung cancer mortality was assessed in participants from the PLCO and the NLST. There was a satisfactory agreement between the risks predicted by the models and those observed in the participants of the PLCO and NLST (calibration). All models matched the overall risk observed in the PLCO and the NLST, as well as the levels of risk observed across different individuals. However, a wide range was found in the capability of the models to distinguish individuals with the predicted event from those without the event (discrimination), denoted by the area under the receiver operating characteristic curve (AUC). The discrimination performance of all models was better in the PLCO (AUCs ranging from 0.74 to 0.81) compared to the NLST (AUCs ranging from 0.61 to 0.73). This may be caused by the difference in risk factor profiles between individuals in the PLCO and the NLST, due to the difference in criteria for enrollment based on smoking history between the two trials. Therefore, the heterogeneity of risk factors in the PLCO participants was larger compared to the participants of the NLST, which influences the discriminative ability of a model.[34,35] Decision curve analyses over a wide range of risk thresholds indicated that all models provided a better ratio of benefits to harms compared to the NLST criteria over a substantial range of risk thresholds (0.1%-16.7%). Finally, the sensitivity and specificity of each model were determined for risk thresholds that selected a similar number of individuals as the NLST eligibility criteria in the PLCO. In the PLCO CXR arm, the NLST criteria yielded a sensitivity of 71.4% and a specificity of 62.2%; however, each of the risk prediction models had both a higher sensitivity and specificity. The three best performing models (the

PLCOm2012 model, the Bach model and the Two-Stage Clonal Expansion incidence model) yielded sensitivities >79.8% and specificities >62.3%. These results suggest that the selection of individuals based on personalized risk may improve the efficiency of lung cancer screening compared to current recommendations. Furthermore, Chapter 9 indicated that the cost-effectiveness of lung cancer screening is dependent on the selection of high-risk individuals for screening. Thus, improving the identification of high-risk individuals may not only increase the efficiency of lung cancer screening, but may also improve its cost-effectiveness.

Chapters 8-10 focused on optimizing lung cancer screening through selecting individuals for screening based on risk stratification. However, risk stratification may also be used to optimize the nodule management protocol. **Chapter 11** investigated how the presence and characteristics of pulmonary nodules detected through CT lung cancer screening can provide information on an individual's risk for developing lung cancer. Information from 7,155 individuals screened in the Dutch-Belgian randomized lung cancer screening trial (NELSON), in whom 9,681 non-calcified nodules were detected, was used to quantify the two-year risk for developing lung cancer after a CT scan based on nodule volume, volume doubling time, and volumetry-based diameter through logistic regression analysis. The overall two-year risk for developing lung cancer after a CT scan of all participants was 1.3%, but individuals without pulmonary nodules (54% of all individuals) had a lower risk compared to those with one or more pulmonary nodules (0.4% compared with 2.5%). The risk for individuals with pulmonary nodules could be further stratified by the volume or diameter of these nodules: individuals with a nodule volume of 100 mm^3 or smaller (0.6%) or a maximum transverse diameter smaller than 5 mm (0.4%) did not have a significantly different risk from individuals without nodules. However, individuals with nodules with a volume of 100-300 mm^3 or a diameter of 5-10 mm had an intermediate probability for lung cancer (2.4% and 1.3% respectively), while individuals with nodules with a volume of 300 mm^3 or larger, or a diameter of 10 mm or greater had a high probability for lung cancer (16.9% and 15.2% respectively). Nodule volume doubling time allowed further stratification of the lung cancer risk in individuals with nodules between 50-500 mm^3: the lung cancer probability was not significantly higher for individuals with nodule volume times of 600 days or more (0.8%), but significantly increased for individuals with nodule volume times of 400-

600 days (4.0%) and nodule volume times of 400 days or fewer days (9.9%). Nodule management based protocols based on volume and diameter were compared to a simulated protocol suggested by the American College of Chest Physicians (ACCP). The simulated AACP protocol had a sensitivity of 90.9% and a specificity of 87.2%, however the protocols based on nodule volume or nodule diameter had either similar or greater sensitivity (90.9% and 92.4% respectively) and greater specificity (94.9% and 90.0% respectively). In addition, while the AACP protocol would directly refer 9% of the individuals for immediate work-up, this was lower for the volume and diameter based protocols (both 5%). Furthermore, the proportion of follow-up examinations was lower for the volume and diameter based protocols (8% and 22% respectively) compared with the ACCP protocol (30%). Chapter 7 noted that one of the major barriers for the implementation of lung cancer screening was the high proportion of false-positive results while Chapter 9 showed that the costs of (follow-up) CT examinations had the greatest influence on its cost-effectiveness. Thus, information on nodule volume and diameter may allow optimization of the nodule management protocol. This could decrease the number of false-positive results and follow-up examinations, which would reduce the harms of lung cancer screening and improve its cost-effectiveness.

The information derived from Chapters 8-11 provides the following answer to the research question:

How can risk stratification be used to optimize lung cancer screening policies?
Risk stratification allows the identification of different risk-groups and assessment of the potential variation in risk within these groups. The identification and selection of high-risk individuals is essential for the cost-effective implementation of lung cancer screening. The identification of individuals at risk for lung cancer, and thus the efficiency of a lung cancer screening program, can be improved by applying risk prediction models rather than criteria based on age and pack-years smoked. Finally, risk stratification based on the results of CT screening examinations may aid in optimizing nodule management protocols, which could reduce the number of false-positive results and follow-up examinations.

Methodological Considerations

The majority of the results presented in this thesis are based on modeling analyses, primarily using the MISCAN-Lung model, of which the development has been described in Chapters 1 and 2. Thus, the majority of the methodological considerations of this thesis pertain to modeling, in particular the MISCAN-Lung model.

As stated in Chapter 1, models are simplifications of reality and may thus not capture the full influence of all endogenous and endogenous risk factors for lung cancer, such as family history or occupational exposures. Models with different assumptions and approaches to modeling the process of lung carcinogenesis managed to reproduce the results of the NLST and the PLCO. Furthermore, one requires assumptions on the preclinical progression of lung cancer to model this process, as it is unobservable.

Furthermore, the estimates effectiveness for CT lung cancer screening and the preclinical progression of lung cancer were solely based on the NLST and PLCO trials. Similarly, the risk for developing lung cancer based on the presence and characteristics of pulmonary nodules detected through CT lung cancer screening in Chapter 11 was solely based on information from the NELSON trial. Chapter 10 indicated that the characteristics of the population(s) used to develop a (risk prediction) model can influence the estimated level of risk for a given individual, which may affect the generalizability of the models evaluated in this thesis. Although the NLST, NELSON and PLCO trials are the largest studies on lung cancer screening with CT and CXR, further evaluation of the generalizability of the models and results presented in this thesis would be valuable.

The calibration of the MISCAN-Lung model, as well as the validation of the risk prediction models evaluated in Chapter 10 were based on self-reported smoking behavior in the NLST and PLCO. In addition, the estimates for the long-term benefits and harms of lung cancer screening for the U.S. and Ontario, Canada, presented in this thesis are based on extrapolations of future smoking behavior. These extrapolations are based on modeling estimates for historical cohort-specific smoking behavior derived from cross-sectional surveys of self-reported smoking. This information is subject to recall bias, social desirability

response bias and, in the case of reporting the number of cigarettes per day, digit bias.[36] Furthermore, the extrapolation of cohort-specific smoking behaviors based on historical behavior does not take the potential influence of future tobacco control policies into account. In addition, recent studies suggest that the use of e-cigarettes and other tobacco products such as (little) cigars has become more prevalent and these products are often used in combination with cigarettes.[37-39] In particular, the effects of the increasing prevalence of e-cigarette use are uncertain. While some advocate their use as a means of helping smokers quit the use of regular cigarettes, others argue that e-cigarettes could renormalize smoking behavior.[40,41] If smoking behavior is renormalized, it could lead to an increased uptake of smoking in adolescents and reduce smoking cessation rates in current cigarette users through promoting long-term dual of regular tobacco products and e-cigarettes. Moreover, both the short- and long-term health effects of e-cigarette use, including potential associations with the risk for developing lung cancer, have not yet been established.[42,43] Modeling analyses suggest that the net effects of e-cigarette use on population health are strongly dependent on their relative toxicity compared to cigarettes and their impact on cigarette smoking behavior patterns.[44,45]

Finally, the modeling extrapolations in this thesis assumed that an individual's smoking behavior was unaffected by attending lung cancer screening. Some argue that lung cancer screening can be used as a teachable moment, described by McBride et al as "*a naturally occurring health event thought to motivate individuals to spontaneously adopt risk-reducing health behaviors*", which may prompt participating current smokers to consider smoking cessation.[46,47] A number of single-arm studies suggest that smoking cessation rates were more favorable in individuals who attended lung cancer screening (14%-24%) compared with the general U.S. adult population (5%-7%).[48-52] However, the Danish Lung Cancer Screening Trial (DLCST) found comparable cessation rates between the screening and control arms, both of which received smoking cessation counseling.[53] Furthermore, results from the NELSON trial suggest that while the cessation rates in both the screening and control groups were higher compared with the general Dutch population (3%-7%), individuals in the screening group were less likely to quit (14.5%) compared to the control group (19.1%).[54,55] This could be an indication for the presence of a "health certificate effect"; smokers who receive a negative screening result may perceive this as a justification

to continue smoking.[56,57] Thus, it is currently uncertain how a lung cancer screening program will affect the smoking behavior of its participants, which in turn may affect the benefits and harms of such a program.

Future Research Directions

Reconciling the results of CT lung cancer screening trials

The NLST was a landmark study on the effectiveness of CT lung cancer screening, as it found a statistically significant reduction in lung cancer mortality of 20% compared to CXR screening.[2] However, the results of other CT lung cancer screening trials thus far have been less optimistic.

The DLCST found no significant reduction in lung cancer mortality for CT screening compared to no screening, with a relative risk for lung cancer mortality of 1.03 (95% confidence interval: 0.66–1.60).[58] The results of the Italian Detection And screening of early lung cancer with Novel imaging TEchnology (DANTE) trial were similar to those of the DLCST, as it found a relative risk for lung cancer mortality of 0.99 (95% confidence interval: 0.69–1.43) for CT screening compared to no screening.[59] Conversely, the Multicentric Italian Lung Detection (MILD) trial found a relative risk for lung cancer mortality of 1.99 (95% confidence interval: 0.80–4.96). In contrast, the Italian Lung Cancer Screening Trial (ITALUNG) found a relative risk for lung cancer mortality of 0.70 (95% confidence interval: 0.47–1.03).[60] However, none of these trials had a sufficiently large sample size to obtain adequate statistical power to show a reduction in lung cancer mortality. A number of trials are still ongoing: the German Lung Cancer Screening Intervention Trial (LUSI), the UK Lung Cancer Screening Trial (UKLS) and the NELSON trial.[61-64] However, of these trials, only the NELSON trial has sufficient statistical power to detect a reduction in lung cancer mortality lung cancer mortality reduction of at least 25%.

Although these trials have different and sometimes conflicting results, pooling information from these trials through comparative data analyses and modeling could provide further insights in the causes of these differences. Within the field of prostate cancer screening there is considerable debate on the effectiveness of prostate specific antigen (PSA)

screening.[65] This is mainly due to the conflicting results between the prostate component of the PLCO, which found a relative risk for prostate cancer mortality of 1.04 (95% confidence interval: 0.87-1.24) for PSA screening compared to no screening, and the European Randomized study of Screening for Prostate Cancer (ERSPC), which found a relative risk of 0.79 (95% confidence interval: 0.69-0.91) compared with no screening.[66,67] However, a comparative modeling analysis (with one of the models being informed by information from the ERSPC) attempted to reconcile the results of these trials and indicated that the conflicting results could potentially be caused due to extensive contamination in the control arm of the PLCO.[68,69] Chapter 1 of this thesis showed that despite the differences in design and populations, modeling was able to reconcile the results of the NLST and PLCO trials. Modeling may similarly be able to reconcile the results of the NLST and other CT lung cancer screening trials and provide valuable insights in the differences and similarities between the trials.

Effectively implementing risk-stratified lung cancer screening

The first of the major barriers for the implementation of lung cancer screening noted in Chapter 7 was that the implementation of a large scale screening program based on risk-stratification has had little to no consideration. Chapter 10 indicated that using risk prediction models to select individuals for lung cancer screening could lead to improvements in efficiency compared to current eligibility criteria. However, the long-term benefits and harms of personalized risk-based lung cancer screening have not yet been evaluated and should be compared to those of current guidelines.

Furthermore, the feasibility of recruiting high-risk individuals warrants further investigation. The UKLS recruited high-risk individuals using the Liverpool Lung Project model.[16] An evaluation of this recruitment process indicated that former smokers were more likely to participate than current smokers. In addition, while individuals in lower socioeconomic groups were more likely to be at higher risk for lung cancer compared to individuals from higher socioeconomic groups, they were less likely to participate. An early survey on attitudes toward lung cancer screening in the U.S. provided similar indications: 87.6% of never-smokers were willing to consider lung cancer screening, compared to 86.1% of former smokers and 71.7% of current smokers.[70] These investigations suggest that never-smokers

and other low-risk groups may be more willing to participate in lung cancer screening compared to high-risk groups. However, Chapter 8 indicated that never-smokers are unlikely to attain a sufficient level of risk to benefit from lung cancer screening. Furthermore, Chapter 9 indicated that lung cancer screening can only be cost-effectively implemented if eligibility is restricted to high-risk groups. Therefore, methods to appropriately reach and recruit high-risk individuals and discourage low-risk individuals from participating should be investigated to retain the (cost-)effectiveness of lung cancer screening. Currently, a number of pilot studies in Ontario (Canada), London (U.K.), Scotland and Queensland (Australia) are evaluating the implementation of lung cancer screening, as well as methods to identify and approach individuals at high-risk for lung cancer.[71-74]

It is uncertain how the implementation of lung cancer screening will affect the smoking behavior of individuals attending the screening examination. Analyses of the NLST participants suggest that former smokers in the CXR arm had a similar reduction in lung cancer mortality after 7 years of abstinence compared to the CT arm.[75] This reduction increased by 10% for former smokers in the CT arm, which suggests that integrating effective smoking cessation programs in lung cancer screening could improve the efficacy of lung cancer screening.[76] In addition, the number of individuals eligible for screening under the current USPSTF criteria is suggested to decline over time, in particular due to individuals exceeding the eligibility criteria for the maximum number of years since smoking cessation.[77] This is supported by modeling estimates, which suggests that while lung cancer screening would yield the highest benefits in the short-term, the long-term benefits diminish over time due the decreasing proportion of individuals that is eligible for screening.[78] Instead combining lung cancer screening with an effective smoking cessation program would yield the highest short- and long-term benefits. Furthermore, modeling suggest that the effect of a lung cancer screening program on smoking cessation rates is strongly linked to the cost-effectiveness of the program.[76] If screening increases smoking cessation rates, the costs per quality-adjusted-life-year (QALY) would decrease almost proportionally for men. However, if lung cancer screening reduces smoking cessation rates, the costs per QALY would increase disproportionally, as the harms of reduced smoking cessation rates would nearly offset the benefits of lung cancer screening.

As stated previously, it is currently unclear how a lung cancer screening program will affect the smoking behavior of its participants.[46-49,51-54,79] However, studies from NLST, NELSON, single-arm trials and implementation studies suggest that current smokers who received an abnormal screen result were more likely to quit smoking.[49,51,80,81] This suggests that an abnormal screen result may be a potential teachable moment to encourage smoking cessation. Thus, ideally one would combine lung cancer screening with smoking cessation programs that capitalize on the potential for a teachable moment of an abnormal screen and counteract the potential for a health certificate effect of a negative screen.[79] Thus, methods to tailor smoking cessation programs to a participant's screen result warrant further investigation.

Optimization, implementation and adherence to nodule management protocols

The second of the major barriers for the implementation of lung cancer screening noted in Chapter 7 was the absence of evidence-based guidelines that incorporate recent findings of nodule management studies. Chapter 11 provided recommendations on risk-based nodule management using information from the NELSON trial, which has been incorporated in the newest guidelines on nodule management of the British Thoracic Society.[18,20] However, other studies have contributed information which may allow further optimization of these protocols.

A risk prediction model developed in the Pan-Canadian Early Detection of Lung Cancer Study (PanCan) and validated in data from chemoprevention trials at the British Colombia Cancer Agency showed great promise in discriminating between benign and malignant nodules detected at baseline CT screening (AUCs of over 0.90).[82] In particular, the size and location of a nodule, as well as the presence of spiculation in the nodule were found to be important predictors for the malignancy of a pulmonary nodule. However, while some information has been provided on the sensitivity and specificity of the model at different risk thresholds, the optimal risk threshold has not yet been determined.[82]

The American College of Radiology recently proposed the Lung CT Screening Reporting and Data System (Lung-RADS) for the classification of nodules found during CT lung cancer screening.[19] The performance of Lung-Rads was evaluated by retrospectively applying the

Lung-Rads criteria to the screening results of the NLST.[17] The NLST defined a screening result as positive if a nodule of at least 4 mm was detected; Lung-RADS applies a 6 mm threshold for a positive screening result for baseline screenings and has further criteria for nodules detected at incidence screenings.[2,17] The application of Lung-RADS in the NLST would have yielded much lower false-positive rates for both the baseline screening (12.8% compared with 26.6%) and incidence screenings (5.3% compared with 21.8%). However, the decrease in false-positive rates would have come at the cost of a lower sensitivity: the application of Lung-RADS would have lowered the sensitivity for the baseline screening from 93.5% to 84.9%, and the sensitivity for the incidence screenings from 93.8% to 78.6%. The difference in sensitivity of Lung-RADS between the baseline and incidence screening rounds of the NLST suggests that nodule management protocols require further refinement. For example, the probability of malignancy for new solid pulmonary nodules detected at incidence screening rounds was recently assessed in the NELSON trial.[83] Overall, new solid nodules were detected in 5%-7% of individuals at each screening round, with 4% of all new solid nodules representing lung cancer. While new solid nodules smaller than 27 mm^3 had a 0.5% probability of being lung cancer, this increased to 3.1% for nodules with volumes between 27-206 mm^3 and to 16.9% for nodules with a volume of 206 mm^3 or greater. This information may be used to refine nodule management protocols for incidence screening rounds.

Furthermore, while it is important to further improve nodule management protocols, the implementation of and adherence to these updated protocols is equally important. An implementation study for lung cancer screening in the U.S. Veterans Health Administration indicated that 59.7% of all screened participants had a positive test result with 56.2% of all screened participants having nodules which required further investigation.[84] However, this implementation study recommended follow-up of nodules smaller than 4 mm (with 54.9% of screen-detected nodules being smaller than 5 mm), which is discouraged by recent studies and current protocols.[18-20,85] Furthermore, both the proportion of positive test results and proportion of screened participants having nodules which required further investigation ranged drastically across the evaluated centers (30.7%-85.0% and 28.1%-82.6% respectively), which suggests differences in interpretation and application of the nodule follow-up protocols between institutions.[84] Therefore, it is crucial that institutions provide

guidance to radiologists for implementing and uniformly adhering to these improved protocols.

Personalization of the screening regimen

Chapters 2 and 5 indicated that annual screening yields greater lung cancer mortality reductions compared to biennial screening. Chapter 9 concluded that despite the higher costs of annual screening policies, they were more cost-effective than biennial screening policies. Together, these investigations indicate that uniform biennial lung cancer screening policies may be (cost-)inefficient. However, Chapter 11 showed that the absence or presence of pulmonary nodules detected on CT screening, as well as the characteristics of these nodules, provides information on an individual's 2-year lung cancer risk. This suggests that the information derived from CT screenings may allow personalization of the screening regimen.

Investigations from the NELSON trial support this hypothesis.[86,87] In NELSON, the risk for screen-detected lung cancer over a 5.5 year period depended heavily on the result of the baseline screening.[86] Participants with a negative baseline screening had a 1.0%, risk, participants with an indeterminate baseline screening had a 5.7% risk and participants with a positive baseline screening had a 48.3% risk. Furthermore, the results of the first three screening rounds were found to be indicative of the risk for screen-detected cancer in the fourth round.[87] Those who solely had negative screening results in the first three rounds had a 0.6% risk for screen-detected lung cancer in the fourth round compared with 1.6% for those who had at least one indeterminate result (but never a positive result). Furthermore, the results of the third screening round itself were found to be indicative for the risk for screen-detected cancer in the fourth round. Participants with a negative screening result in the third round had a 0.6% risk for screen-detected lung cancer in the fourth round compared with 3.7% for those who had an indeterminate result. Retrospective analyses of the NLST provided similar indications.[88] Participants in the CT arm with a negative baseline CT-result had a subsequent lung cancer detection rate of 0.34% compared with 1.02% for all screened participants.

Chapter 9 showed that the costs of CT screening examinations had a major influence on the cost-effectiveness of lung cancer screening programs. Thus, personalization of the screening regimen may improve the cost-effectiveness of lung cancer screening, for example, by advising biennial screening intervals to low-risk individuals. If participants in the NLST CT arm with a negative baseline CT screen had not received the first incidence screening, the required number of CT screening examinations would have been reduced by 73%.[88] However, the lung cancer mortality rate would have increased by 14% in this group and by approximately 7% for all participants in the CT arm.[88] Thus, personalization of the screening regimen requires a balance between efficiency (reducing the number of screening examinations) and efficacy (the potential reduction in lung cancer mortality) which warrants further investigation.

Main Conclusions and Recommendations

The following conclusions can be drawn, based on the results of this thesis:

- Natural history modeling was able to reconcile the results of the NLST and PLCO trials, and synthesize information from both trials. Through this synthesis, information on the natural history of lung cancer in the presence and absence of screening could be derived for populations with different levels of risk. This allows extrapolation of the results of the NLST to different designs and populations than those considered in the NLST.

- Lung cancer screening has a favorable trade-off between benefits and harms and can be implemented in a cost-effective manner; provided that high-risk individuals are selected.

- Focusing on high-risk participants predominantly selects individuals who have a heightened risk of mortality due to their (past) smoking behavior and may increase the potential for overdiagnosis. However, despite the heighted risk of mortality and overdiagnosis, lung cancer screening still yields a substantial number of life-years for each lung cancer death averted.

- The selection of individuals for lung cancer screening based on lung cancer risk models rather than age and pack-years alone could improve the efficiency of lung cancer screening.

- The presence of pulmonary nodules on CT screening examinations, as well as the characteristics of these nodules, provide an indication on a person's lung cancer risk.

Furthermore, the following priorities for future research are suggested:

- Comparison and reconciliation of the results of the NLST, NELSON and other CT lung cancer screening trials, through pooled analyses and microsimulation modeling.

- Evaluation of the efficacy and efficiency of implementing lung cancer screening based on risk prediction models.

- Assessment of the feasibility to implement large-scale personalized lung cancer screening based on risk stratification.

- Optimization of the nodule management protocols and personalization of the screening regimen.

References

1. 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. *Journal of the American Medical Association* 2011; **306**(17): 1865-73.
2. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011; **365**(5): 395-409.
3. van Ballegooijen M, Rutter CM, Knudsen AB, et al. Clarifying Differences in Natural History between Models of Screening. *Medical Decision Making* 2011; **31**(4): 540-9.
4. Groome PA, Bolejack V, Crowley JJ, et al. The IASLC Lung Cancer Staging Project: Validation of the Proposals for Revision of the T, N, and M Descriptors and Consequent Stage-groupings in the Forthcoming (Seventh) Edition of the TNM Classification of Malignant Tumours. *Journal of Thoracic Oncology* 2007; **2**(8): 694-705.
5. Patz EF, Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Internal Medicine* 2014; **174**(2): 269-74.
6. Pinsky PF. An Early- and Late-Stage Convolution Model for Disease Natural History. *Biometrics* 2004; **60**(1): 191-8.
7. Chien C-R, Chen TH-H. Mean sojourn time and effectiveness of mortality reduction for lung cancer screening with computed tomography. *International Journal of Cancer* 2008; **122**(11): 2594-9.
8. Aberle DR, Adams AM, Berg CD, et al. Baseline Characteristics of Participants in the Randomized National Lung Screening Trial. *Journal of the National Cancer Institute* 2010; **102**(23): 1771-9.
9. Moyer VA, on behalf of the USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 2014; **160**(5): 330-8.
10. Wender R, Fontham ETH, Barrera E, et al. American Cancer Society lung cancer screening guidelines. *CA: A Cancer Journal for Clinicians* 2013; **63**(2): 106-17.
11. Parliament. *The Lancet* 1953; **261**(6761): 645.
12. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen–based prostate cancer screening strategies: Model estimates of potential benefits and harms. *Annals of Internal Medicine* 2013; **158**(3): 145-53.
13. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. 2012; *The Lancet*; **380**(9855): 1778-86.
14. de Gelder R, Fracheboud J, Heijnsdijk EAM, et al. Digital mammography screening: Weighing reduced mortality against increased overdiagnosis. *Preventive Medicine* 2011; **53**(3): 134-40.
15. Decision memo for screening for lung cancer with low dose computed tomography (LDCT) (CAG-00439N). Baltimore: Centers for Medicare & Medicaid Services; 2015. Available from: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274. Accessed April 24, 2017.
16. McRonald FE, Yadegarfar G, Baldwin DR, et al. The UK Lung Screen (UKLS): Demographic Profile of First 88,897 Approaches Provides Recommendations for Population Screening. *Cancer Prevention Research* 2014; **7**(3): 362.
17. Pinsky PF, Gierada DS, Black W, et al. Performance of Lung-RADS in the National Lung Screening Trial: A retrospective assessment. *Annals of Internal Medicine* 2015; **162**(7): 485-91.
18. Baldwin DR, Callister MEJ. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax* 2015; **70**(8): 794.
19. U.S. Department of Health, Education, and Welfare. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service. Washington: U.S. Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, 1964. PHS Publication No. 1103.
20. Callister MEJ, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules: accredited by NICE. *Thorax* 2015; **70**(Suppl 2): ii1.

21. Koo LC, Ho JHC. Worldwide Epidemiological Patterns of Lung Cancer in Nonsmokers. *International Journal of Epidemiology* 1990; **19**(Supplement_1): S14-S23.
22. Parkin DM, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. *CA: A Cancer Journal for Clinicians* 2005; **55**(2): 74-108.
23. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers-a different disease. *Nature Reviews Cancer* 2007; **7**(10): 778-90.
24. Thun MJ, Henley SJ, Burns D, Jemal A, Shanks TG, Calle EE. Lung Cancer Death Rates in Lifelong Nonsmokers. *Journal of the National Cancer Institute* 2006; **98**(10): 691-9.
25. Janakiraman S, Ramaswamy G. Lung Cancer in Never Smokers: A Review. *Journal of Clinical Oncology* 2007; **25**(5): 561-70.
26. Zhou Q-h, Fan Y-g, Bu H, et al. China national lung cancer screening guideline with low-dose computed tomography (2015 version). *Thoracic Cancer* 2015; **6**(6): 812-8.
27. Australian Governement Standing Committee on Screening: Community Care and Population Health Principal Committee. (2015). Position Statement: Lung Cancer Screening using Low-Dose Computed Tomography. Cancer Australia. Available from: http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/lung-cancer-screening, Accessed on: 24-06-2017.
28. Kauczor H-U, Bonomo L, Gaga M, et al. ESR/ERS white paper on lung cancer screening. *European Respiratory Journal* 2015; **46**(1): 28-39.
29. Black WC, Gareen IF, Soneji SS, et al. Cost-Effectiveness of CT Screening in the National Lung Screening Trial. *New England Journal of Medicine* 2014; **371**(19): 1793-802.
30. Sox HC. Better Evidence about Screening for Lung Cancer. *New England Journal of Medicine* 2011; **365**(5): 455-7.
31. Field JK, Oudkerk M, Pedersen JH, Duffy SW. Prospects for population screening and diagnosis of lung cancer. *The Lancet*; **382**(9893): 732-41.
32. Telford JJ, Levy AR, Sambrook JC, Zou D, Enns RA. The cost-effectiveness of screening for colorectal cancer. *Canadian Medical Association Journal* 2010; **182**(12): 1307-13.
33. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *CMAJ: Canadian Medical Association Journal* 1992; **146**(4): 473-81.
34. Vergouwe Y, Moons KGM, Steyerberg EW. External Validity of Risk Models: Use of Benchmark Values to Disentangle a Case-Mix Effect From Incorrect Coefficients. *American Journal of Epidemiology* 2010; **172**(8): 971-80.
35. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of Clinical Epidemiology* 2015; **68**(3): 279-89.
36. Hatziandreu EJ, Pierce JP, Fiore MC, Grise V, Novotny TE, Davis RM. The reliability of self-reported cigarette consumption in the United States. *American Journal of Public Health* 1989; **79**(8): 1020-3.
37. Benowitz NL, Goniewicz ML. The regulatory challenge of electronic cigarettes. *Journal of the American Medical Association* 2013; **310**(7): 685-6.
38. Richardson A, Xiao H, Vallone DM. Primary and Dual Users of Cigars and Cigarettes: Profiles, Tobacco Use Patterns and Relevance to Policy. *Nicotine & Tobacco Research* 2012; **14**(8): 927-32.
39. King BA, Patel R, Nguyen KH, Dube SR. Trends in Awareness and Use of Electronic Cigarettes Among U.S. Adults, 2010–2013. *Nicotine & Tobacco Research* 2015; **17**(2): 219-27.
40. Etter J-F. Should electronic cigarettes be as freely available as tobacco? Yes. *British Medical Journal* 2013; **346**.
41. Chapman S. Should electronic cigarettes be as freely available as tobacco cigarettes? No. *British Medical Journal* 2013; **346**.
42. Callahan-Lyon P. Electronic cigarettes: human health effects. *Tobacco Control* 2014; **23**(suppl 2): ii36.

43. Grana R, Benowitz N, Glantz SA. E-Cigarettes. *Circulation* 2014; **129**(19): 1972.
44. Kalkhoran S, Glantz SA. Modeling the health effects of expanding e-cigarette sales in the united states and united kingdom: A monte carlo analysis. *JAMA Internal Medicine* 2015; **175**(10): 1671-80.
45. Levy DT, Borland R, Villanti AC, et al. The Application of a Decision-Theoretic Model to Estimate the Public Health Impact of Vaporized Nicotine Product Initiation in the United States. *Nicotine & Tobacco Research* 2017; **19**(2): 149-59.
46. McBride CM, Emmons KM, Lipkus IM. Understanding the potential of teachable moments: the case of smoking cessation. *Health Education Research* 2003; **18**(2): 156-70.
47. Taylor KL, Cox LS, Zincke N, Mehta L, McGuire C, Gelmann E. Lung cancer screening as a teachable moment for smoking cessation. *Lung Cancer* 2007; **56**(1): 125-34.
48. Kitts AKB, McKee AB, Regis SM, Wald C, Flacke S, McKee BJ. Smoking cessation results in a clinical lung cancer screening program. *Journal of Thoracic Disease* 2016: S481-S7.
49. Townsend CO, Clark MM, Jett JR, et al. Relation between smoking cessation and receiving results from three annual spiral chest computed tomography scans for lung carcinoma screening. *Cancer* 2005; **103**(10): 2154-62.
50. Centers for Disease C, Prevention. Quitting smoking among adults--United States, 2001-2010. *Morbidity and Mortality Weekly Report* 2011; **60**(44): 1513-9.
51. Styn MA, Land SR, Perkins KA, Wilson DO, Romkes M, Weissfeld JL. Smoking Behavior 1 Year after Computed Tomography Screening for Lung Cancer: Effect of Physician Referral for Abnormal CT Findings. *Cancer Epidemiology Biomarkers & Prevention* 2009; **18**(12): 3484.
52. Cox LS, Clark MM, Jett JR, et al. Change in smoking status after spiral chest computed tomography scan screening. *Cancer* 2003; **98**(11): 2495-501.
53. Ashraf H, Saghir Z, Dirksen A, et al. Smoking habits in the randomized Danish Lung Cancer Screening Trial with low-dose CT: final results after a 5-year screening programme. *Thorax* 2014; **69**(6): 574.
54. van der Aalst CM, van den Bergh KAM, Willemsen MC, de Koning HJ, van Klaveren RJ. Lung cancer screening and smoking abstinence: 2 year follow-up data from the Dutch–Belgian randomized controlled lung cancer screening trial. *Thorax* 2010; **65**(7): 600.
55. Willemsen MC, Wagena EJ, van Schayck CP. [The efficacy of smoking cessation methods available in the Netherlands: a systematic review based on Cochrane data]. De effectiviteit van stoppen-met-rokenmethoden die in Nederland beschikbaar zijn: een systematische review op basis van Cochrane-gegevens. *Ned Tijdschr Geneeskd* 2003; **147**(19): 922-7.
56. Eddy DM. Screening for lung cancer. *Annals of Internal Medicine* 1990; **112**(1): 73-4.
57. Tymstra T, Bieleman B. The psychosocial impact of mass screening for cardiovascular risk factors. *Family Practice* 1987; **4**(4): 287-90.
58. Wille MM, Dirksen A, Ashraf H, et al. Results of the Randomized Danish Lung Cancer Screening Trial with Focus on High-Risk Profiling. *American Journal of Respiratory and Critical Care Medicine* 2016; **193**(5): 542-51.
59. Infante M, Cavuto S, Lutman FR, et al. Long-Term Follow-up Results of the DANTE Trial, a Randomized Study of Lung Cancer Screening with Spiral Computed Tomography. *American Journal of Respiratory and Critical Care Medicine* 2015; **191**(10): 1166-75.
60. Paci E, Puliti D, Lopes Pegna A, et al. Mortality, survival and incidence rates in the ITALUNG randomized lung cancer screening trial. *Thorax* 2017.
61. Becker N, Motsch E, Gross ML, et al. Randomized Study on Early Detection of Lung Cancer with MSCT in Germany: Results of the First 3 Years of Follow-up After Randomization. *Journal of Thoracic Oncology* 2015; **10**(6): 890-6.
62. Field JK, Duffy SW, Baldwin DR, et al. The UK Lung Cancer Screening Trial: a pilot randomized controlled trial of low-dose computed tomography screening for the early detection of lung cancer. *Health Technology Assessment* 2016; **20**(40): 1-146.

63. Field JK, Duffy SW, Baldwin DR, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax* 2016; **71**(2): 161-70.
64. 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 randomized lung cancer multi-slice CT screening trial (NELSON). *International Journal of Cancer* 2007; **120**(4): 868-74.
65. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: A review of current evidence. *Journal of the American Medical Association* 2014; **311**(11): 1143-9.
66. Pinsky PF, Prorok PC, Yu K, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. *Cancer* 2017; **123**(4): 592-9.
67. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomized Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *The Lancet*; **384**(9959): 2027-35.
68. Gulati R, Tsodikov A, Wever EM, et al. The impact of PLCO control arm contamination on perceived PSA screening efficacy. *Cancer Causes & Control* 2012; **23**(6): 827-35.
69. Gulati, 2017, in press.
70. Silvestri GA, Nietert PJ, Zoller J, Carter C, Bradford D. Attitudes towards screening for lung cancer among smokers and their non-smoking counterparts. *Thorax* 2007; **62**(2): 126.
71. Tammemägi M, Hader J, Yu M, et al. P1.03-059 Organized High Risk Lung Cancer Screening in Ontario, Canada: A Multi-Centre Prospective Evaluation. *Journal of Thoracic Oncology*; **12**(1): S579.
72. Sullivan FM, Farmer E, Mair FS, et al. Detection in blood of autoantibodies to tumour antigens as a case-finding method in lung cancer using the EarlyCDT®-Lung Test (ECLS): study protocol for a randomized controlled trial. *BMC Cancer* 2017; **17**(1): 187.
73. Marshall HM, Bowman RV, Crossin J, et al. Queensland Lung Cancer Screening Study: rationale, design and methods. *Internal Medicine Journal* 2013; **43**(2): 174-82.
74. Marshall HM, Bowman RV, Ayres J, et al. Lung cancer screening feasibility in Australia. *European Respiratory Journal* 2015; **45**(6): 1734-1737.
75. Tanner NT, Kanodra NM, Gebregziabher M, et al. The Association between Smoking Abstinence and Mortality in the National Lung Screening Trial. *American Journal of Respiratory and Critical Care Medicine* 2016; **193**(5): 534-41.
76. McMahon PM, Kong CY, Bouzan C, et al. Cost-Effectiveness of Computed Tomography Screening for Lung Cancer in the United States. *Journal of Thoracic Oncology* 2011; **6**(11): 1841-8.
77. Wang Y, Midthun DE, Wampfler JA, et al. Trends in the proportion of patients with lung cancer meeting screening criteria. *Journal of the American Medical Association* 2015; **313**(8): 853-5.
78. Tramontano AC, Sheehan DF, McMahon PM, et al. Evaluating the impacts of screening and smoking cessation programmes on lung cancer in a high-burden region of the USA: a simulation modeling study. *BMJ Open* 2016; **6**(2).
79. Slatore CG, Baumann C, Pappas M, Humphrey LL. Smoking behaviors among patients receiving computed tomography for lung cancer screening. Systematic review in support of the U.S. preventive services task force. *Annals of the American Thoracic Society* 2014; **11**(4): 619-27.
80. van der Aalst CM, van Klaveren RJ, van den Bergh KAM, Willemsen MC, de Koning HJ. The impact of a lung cancer computed tomography screening result on smoking abstinence. *European Respiratory Journal* 2011; **37**(6): 1466.
81. Tammemägi MC, Berg CD, Riley TL, Cunningham CR, Taylor KL. Impact of Lung Cancer Screening Results on Smoking Cessation. *Journal of the National Cancer Institute* 2014; **106**(6): dju084-dju.
82. McWilliams A, Tammemägi MC, Mayo JR, et al. Probability of Cancer in Pulmonary Nodules Detected on First Screening CT. *New England Journal of Medicine* 2013; **369**(10): 910-9.

83. Walter JE, Heuvelmans MA, de Jong PA, et al. Occurrence and lung cancer probability of new solid nodules at incidence screening with low-dose CT: analysis of data from the randomized, controlled NELSON trial. *The Lancet Oncology* 2016; **17**(7): 907-16.
84. Kinsinger LS, Anderson C, Kim J, et al. Implementation of lung cancer screening in the veterans health administration. *JAMA Internal Medicine* 2017; **177**(3): 399-406.
85. Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *The Lancet Oncology* 2014; **15**(12): 1332-41.
86. Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *European Respiratory Journal* 2013; **42**(6): 1659.
87. Yousaf-Khan U, van der Aalst C, de Jong PA, et al. Risk stratification based on screening history: the NELSON lung cancer screening study. *Thorax* 2017.
88. Patz Jr EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomized, multicentre, diagnostic screening trial. *The Lancet Oncology* 2016; **17**(5): 590-9.

Summary

In 2012, 1.8 million new cases and 1.6 million lung cancer deaths were estimated to have occurred worldwide, accounting for 13% of all cancer cases and 19% of all cancer related deaths. This makes lung cancer the leading cause of cancer related mortality. Despite decreasing smoking prevalence, lung cancer is expected to remain a major public health problem in the coming decades. Lung cancer is a highly lethal disease; over half of all lung cancers are detected after having spread to another organ or part of the body, at which point the 5-year survival rate is less than 10%. However, if the cancer is still confined to its primary site, 5-year survival is over 50%. Thus, the detection of lung cancer at an earlier stage could improve the potential for curative treatment.

This was confirmed by the findings of the National Lung Screening Trial (NLST), a randomized controlled screening trial, which compared lung cancer screening with low-dose computed tomography (CT) to chest radiography (CXR) screening. The NLST found a significant relative reduction in lung cancer mortality of 20% for CT screening compared to chest radiography screening, as well as an relative reduction in all-cause mortality of 6.7%. However, many questions remained on whether and how to implement a lung cancer screening program, such as: 'which persons should be invited for screening?' and 'what is the optimal screening regimen?'.

Microsimulation models allow the extrapolation of the results of randomized clinical trials to different designs and populations than those considered in these trials. Furthermore, it allows one to evaluate the impact on the benefits and harms of screening for these different designs and populations. Therefore, modeling can provide information on whether lung cancer screening should be implemented and, if so, how it can be optimally implemented.

The first part of this thesis describes the development of a microsimulation model, MISCAN-Lung. The second part of this thesis evaluates the long-term benefits and harms of different lung cancer screening policies. The final part of this thesis describes how risk stratification can be used to optimize lung cancer screening policies.

Part 1: Development of a lung cancer screening model

The development of five independent lung cancer screening models was described in **Chapter 1**. The models were developed through using common inputs and calibration targets derived from the NLST and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO). The models all reproduced the results of both trials, however, calibration to both trials was required to reproduce lung cancer outcomes across all ranges of smoking. This indicated that incorporating information from populations at different levels of lung cancer risk is crucial to allow extrapolation of the results of these trials to the general population.

In **Chapter 2**, one of the models developed in Chapter 1, MISCAN-Lung, was used to derive estimates for the natural history and the detectability of lung cancer through synthetization of information from the NLST and the PLCO. Overall, the sensitivity of CT was higher compared to CXR across all stages and histologies. In particular, CT sensitivity for stage IA was estimated to be more than 3-fold higher compared with CXR, for all histologies. Lung cancer was found to have a longer sojourn time for women compared to men, in particular for adenocarcinoma.

Chapter 3 indicated how modeling can provide information on overdiagnosis beyond what was provided by the observed data of a randomized trial. Excess incidence analyses suggested that 18.5% of all cancers detected by CT in the NLST were overdiagnosed. However, this analysis was biased due to the occurrence of CXR screening in the control arm of the NLST and an insufficient follow-up duration. MISCAN-Lung was used to extrapolate the follow-up period of the NLST to a lifetime period, estimating that 6.75% of all screen-detected cases in the CXR arm and 8.62% of all screen-detected cases in the CT arm of the NLST were overdiagnosed. Furthermore, the model allows estimation of the potential for overdiagnosis by lung cancer screening in the general U.S. population.

Part 2: The long-term benefits and harms of lung cancer screening policies

In **Chapter 4**, the five models developed in Chapter 1 were used to extrapolate the results of the NLST beyond the single design considered in the trial. The models evaluated 576 different screening scenarios, which considered different variations of starting and stopping

ages of screening, required smoking eligibility criteria, and different screening intervals. The models identified 120 'consensus efficient scenarios' which averted the greatest number of lung cancer deaths for a given number of CT screens across all models. All models agreed that the scenario closest to the NLST criteria (annual screening from ages 55 through 75 for current and former smokers who smoked a minimum of 30 pack-years and quit smoking less than 15 years ago) was not among the consensus efficient scenarios. This suggests that the criteria used in the NLST may not be the most optimal for a population based screening program.

Chapter 5 evaluated the long-term benefits and harms of 26 of the 'consensus efficient scenarios' identified in Chapter 4, along with a 27th scenario closest to the NLST criteria, to inform the United States Preventive Services Task Force (USPSTF) on their recommendations for lung cancer screening. Overall, annual screening for individuals aged 55 through 80, who smoked at least 30 pack-years and currently smoke or quit for less than 15 years was judged to provide an advantageous balance between benefits and harms. This scenario would reduce lung cancer mortality by 14.0%, avert 497 lung cancer deaths and yield 10.6 life-years per averted lung cancer death, but require 286,813 CT screening examinations (of which 23.6% false-positive) and lead to 190 overdiagnosed cases (9.9% of all screen-detected cases) per 100,000 individuals. The USPSTF currently recommends lung cancer screening in the U.S. using the eligibility criteria of this scenario.

Chapter 6 assessed how overdiagnosis affects the selection of efficient lung cancer screening strategies through modeling. Overdiagnosis was found to increase for shorter screening intervals, older starting and stopping ages for screening, and for increases in the number of accumulated pack-years required for screening eligibility. Efficient scenarios which maximized the number of lung cancer deaths averted favored a stopping age of 80, while those which maximized the number of lung cancer deaths averted per overdiagnosed case favored a stopping age of 75. Overall, raising the stopping age from 75 to 80 yielded a higher lung cancer mortality reduction, at the cost of a proportionally higher increase in overdiagnosis.

Chapter 7 identified the main barriers to the optimal implementation of lung cancer screening. Firstly, the implementation of a large scale screening program based on risk stratification has had little to no consideration. It is yet unclear how to appropriately identify and approach the population at highest risk for lung cancer, as this population may be harder to reach and less likely to participate in screening. Secondly, the high proportion of false-positive results in lung cancer screening represents a barrier. In the NLST, 96.4% of all positive CT screens consisted of false-positive results, accounting for 23.3% of all CT screens. While different nodule management protocols have been proposed that could reduce the number of false-positive results, this may come at a loss of sensitivity for detecting lung cancer.

Part 3: Optimization through risk stratification

Chapter 8 investigated whether never-smokers at elevated risk for lung cancer may benefit from lung cancer screening. Modeling was used to evaluate the benefits and harms of lung cancer screening for cohorts of never-smokers at different levels of relative risk for lung cancer compared with never-smokers at average risk. The overall level of relative risk required for never-smokers to have a similar trade-off between benefits and harms compared to those of an USPSTF eligible cohort was at least 15. However, the majority of lung cancer risk prediction models that can be applied to never-smokers consider relative risks lower than 15 for never-smokers, suggesting that for the majority of never-smokers lung cancer screening is not beneficial.

Chapter 9 evaluated the cost-effectiveness of lung cancer screening in Ontario, Canada. Ontario-specific data on population characteristics, smoking behavior and costs were used to examine 576 CT screening scenarios varying by age to start and end screening, smoking eligibility criteria, and screening interval, for persons born between 1940-1969. Overall, annual screening between ages 55-75 for persons who smoked at least 40 pack-years, and currently smoke or quit less than 10 years ago, was considered to be the optimal policy to implement. This policy yielded an incremental cost-effectiveness ratio of $41,136 Canadian dollars per life-year gained. If this policy were to be implemented, 9.56% (499,261 individuals) of the examined population (ever- and never-smokers) would be screened at least once, which would require 4,788,523 CT examinations. However, this would reduce

lung cancer mortality in this population by 9.05% (preventing 13,108 lung cancer deaths). Scenarios which applied stringent smoking eligibility criteria were more cost-effective than scenarios that applied less stringent smoking eligibility criteria. Furthermore, annual screening intervals were found to be more cost-effective then biennial screening intervals.

Chapter 7 noted that one of the main barriers for the implementation of lung cancer screening was the identification of high-risk individuals. Therefore, nine established lung cancer risk prediction models were assessed for their ability to identify the participants in the NLST and PLCO who were most likely to develop or die from lung cancer in **Chapter 10**. All models matched the overall risk and the levels of risk across different groups of individuals observed in the two trials. However, a wide range was found in the capability of the models to distinguish individuals with the predicted event from those without the event (discrimination). The discrimination performance of all models was better in the PLCO compared to the NLST. The models provided a better ratio of benefits to harms compared to the NLST eligibility criteria over a substantial range of risk thresholds. Finally, all models yielded a higher sensitivity and specificity for risk thresholds that selected a similar number of individuals as the NLST eligibility criteria (sensitivity of 71.4% and a specificity of 62.2%) in the PLCO.

Chapter 11 showed how risk stratification may be used to optimize the nodule management protocol. Information from the Dutch-Belgian randomized lung cancer screening trial (NELSON) was used to quantify the two-year risk for developing lung cancer after a CT scan based on nodule volume, nodule volume doubling time, and volumetry-based nodule diameter. The analyses showed that individuals with a nodule volume of 100 mm^3 or smaller or a maximum transverse diameter smaller than 5 mm did not have a significantly different risk from individuals without nodules (0.6%, 0.4% and 0.4% respectively). Individuals with nodules with a volume of 100-300 mm^3 or a diameter of 5-10 mm had an intermediate risk for lung cancer (2.4% and 1.3% respectively), and individuals with nodules with a volume of 300 mm^3 or larger, or a diameter of 10 mm or greater had a high risk for lung cancer (16.9% and 15.2% respectively). Volume doubling time allowed for further risk stratification: nodules with a volume doubling time of 600 days or more had a 0.8% two-year risk for lung cancer compared to 4.0% for volume doubling times of 400–600 days, and 9.9% for volume

doubling times of 400 days or fewer. A nodule management protocol based on nodule volume and volume doubling-time yielded similar sensitivity, but greater specificity compared to a protocol suggested by the American College of Chest Physicians.

Discussion

The general discussion of this thesis answered and discussed the research questions stated in the introduction of this thesis. The analyses discussed in the chapters of this thesis showed that models can be used to synthesize information populations with different levels of risk and screening histories to derive the natural history of lung cancer. This information allows one to derive the long-term benefits and harms of lung cancer screening and identify screening policies with an advantageous balance between benefits and harms. Finally, the analyses described in this thesis showed that risk stratification can be used to optimize lung cancer screening in various ways. Firstly, risk stratification allows the identification of different risk-groups and the potential variation in risk within these groups. The analyses in this thesis showed that the selection of high-risk individuals is essential to implement lung cancer screening in a cost-effective manner. Finally, risk stratification based on the results of the screening examination may reduce the number of false-positive results and follow-up examinations. Future research should focus on reconciling the results the NLST with other CT lung cancer screening trials. In addition, the effective implementation of risk-stratified lung cancer screening warrants further investigation. Furthermore, to reduce the number of false-positive results, nodule management protocols need to be further optimized, as well as implemented and adhered to in clinical practice. Finally, it may be possible to personalize the screening regimen, which could reduce the number of screening examinations and improve the cost-effectiveness of lung cancer screening.

Samenvatting

Het aantal nieuwe diagnoses van en sterfgevallen aan longkanker die wereldwijd in 2012 plaatsvonden worden op respectievelijk 1.8 miljoen en 1.6 miljoen geschat. Longkanker was in 2012 verantwoordelijk voor 13% van alle kankerdiagnoses en 19% van alle aan kanker-gerelateerde sterfgevallen. Longkanker is hierdoor de voornaamste aan kanker gerelateerde doodsoorzaak en ondanks de afnemende prevalentie van roken wordt verwacht dat longkanker ook de komende decennia een groot volksgezondheidsprobleem zal vormen. Longkanker is een zeer dodelijke ziekte; meer dan de helft van alle longkankers worden gediagnosticeerd nadat de kanker zich heeft verspreid naar een ander gedeelte van het lichaam. Indien de kanker zich in dit vergevorderde stadium bevindt is de 5-jaarsoverleving minder dan 10%. Als de kanker zich echter nog niet verder verspreidt heeft, is de vijfjaarsoverleving meer dan 50%. Dit suggereert dat vroege opsporing van deze ziekte de mogelijkheden voor curatieve behandeling zou kunnen verbeteren.

De bevindingen van de National Lung Screening Trial (NLST), een gerandomizeerd onderzoek met een controle groep, dat screening naar longkanker tussen lage-dosis computertomografie (CT) en röntgenfoto's van de borstkas vergeleek bevestigen dit. De NLST gaf aan dat screening naar longkanker door middel van CT de relatieve sterfte aan longkanker met 20% kan reduceren ten opzichte van screening met röntgenfoto's. Bovendien werd er een relatieve reductie in sterfte aan alle oorzaken van 6.7% gevonden. Veel vragen bleven echter onbeantwoord, zoals: 'moet een longkankerscreeningsprogramma ingevoerd worden?', 'Welke personen moeten uitgenodigd worden voor screening?' en 'Welk screeningsprotocol is het meest optimaal om aan te houden?'.

Microsimulatie modellen kunnen de resultaten van gerandomizeerde onderzoeken extrapoleren naar andere studie-opzetten en populaties dan die onderzocht werden in deze gerandomizeerde onderzoeken. Bovendien kunnen microsimulatie modellen de voor- en nadelen van screening voor deze alternatieve studie-opzetten en populaties evalueren. Hierdoor kunnen modellen informatie geven over of een longkankerscreeningsprogramma ingevoerd moet worden, en zo ja, hoe dit programma optimaal geïmplementeerd kan worden.

Het eerste deel van dit proefschrift beschrijft de ontwikkeling van een microsimulatie model, MISCAN-Long. Het tweede deel van dit proefschrift evalueert de voor- en nadelen van het invoeren van screening naar longkanker op verscheidene manieren. Het laatste deel van dit proefschrift beschrijft hoe risico stratificatie gebruikt kan worden om longkankerscreeningsprogramma's te optimaliseren.

Deel 1: Het ontwikkelen van een longkanker screening model

De ontwikkeling van vijf onafhankelijke longkankerscreening modellen werd beschreven in **Hoofdstuk 1.** Deze modellen werden ontwikkeld met behulp van gemeenschappelijke invoerparameters en kalibratiedoelen die waren afgeleid van de NLST en de Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO). De modellen slaagden er allen in om de resultaten van beide gerandomiseerde onderzoeken te reproduceren, echter, kalibratie op beide gerandomiseerde onderzoeken was vereist om de uitkomstmaatregelen voor longkanker voor verschillende soorten rookgedrag te reproduceren. Dit geeft aan dat het cruciaal is om informatie van populaties met verschillende risiconiveaus voor het ontwikkelen van longkanker te evalueren, zodat de resultaten van deze gerandomiseerde onderzoeken geëxtrapoleerd kunnen worden naar de algemene bevolking.

In **Hoofdstuk 2**, werd een van de modellen beschreven in Hoofdstuk 1, MISCAN-Long, gebruikt om schattingen te geven voor het preklinische verloop en de opspoorbaarheid van longkanker. Hiervoor werd gebruik gemaakt van samengevoegde informatie van de NLST en de PLCO. Over het algemeen was de sensitiviteit van CT hoger dan die van röntgenfoto's voor alle stadia en types van longkanker. De sensitiviteit van CT voor stadium IA werd drie maal zo hoog geschat ten opzichte van röntgenfoto's voor alle types van longkanker. De modelschattingen gaven aan dat de preklinische duur van longkanker langer is voor vrouwen dan voor mannen, in het bijzonder voor adenocarcinomen.

Hoofdstuk 3 liet zien hoe modeleren additionele informatie kan geven over overdiagnose ten opzichte van de geobserveerde data van een gerandomiseerde onderzoek. Schattingen voor de surplus aan incidentie suggereerden dat 18.5% van alle door CT gedetecteerde longkankers in de NLST waren overgediagnosticeerd. Deze analyses waren echter beperkt doordat 1) screening met röntgenfoto's plaatsvond in de controlegroep van de NLST en 2)

door de beperkte vervolgtijd van de deelnemers van de NLST. De vervolgtijd van de NLST werd met behulp van MISCAN-long geëxtrapoleerd tot het levenseinde van alle deelnemers. Op basis van deze vervolgtijd schatte het model dat 6.75% van alle door screening gedetecteerde longkankers in de met röntgenfoto's gescreende groep en 8.62% van alle door screening gedetecteerde longkankers in de met CT gescreende groep overgediagnosticeerd waren. Bovendien is het met het model mogelijk om schattingen te geven voor de potentiele hoeveelheid overdiagnose indien een longkankerscreeningsprogramma voor de bevolking van de Verenigde Staten wordt ingevoerd.

Deel 2: De voor- en nadelen van longkankerscreeningsprogramma's op de lange termijn

In **Hoofdstuk 4** werden de vijf modellen die in Hoofdstuk 1 waren ontwikkeld gebruikt om de resultaten van de NLST te extrapoleren naar andere studie-opzetten dan degene die werd onderzocht in de NLST. De modellen evalueerden 576 longkankerscreeningsprogramma's met verschillende studie-opzetten, met betrekking tot de start- en stopleeftijden voor screening, het vereiste rookgedrag om in aanmerking te komen voor screening en verschillende tijdsintervallen tussen de screenings. Er was consensus tussen de modellen over 120 "efficiënte screeningsprogramma's", welke voor alle modellen het hoogste aantal sterfgevallen aan longkanker voorkwamen voor een gegeven aantal CT screenings. Er was overeenstemming tussen alle modellen dat het longkankerscreeningsprogramma dat de meeste overeenstemming had met de criteria die werden gebruikt in de NLST (jaarlijkse screening voor individuen tussen de 55 en 75 jaar, die tenminste 30 pak-jaren gerookt hebben en huidig roker zijn of minder dan 15 jaar geleden zijn gestopt met roken) niet tot de consensusprogramma's behoorde. Dit suggereert dat de criteria die gebruikt zijn voor de studie-opzet van de NLST mogelijk niet de meest optimale zijn om te gebruiken voor een longkankerscreeningsprogramma gericht op de algemene bevolking.

Hoofdstuk 5 beschrijft de evaluatie van de voor-en nadelen op de lange termijn van 26 van de consensusprogramma's van Hoofdstuk 4, samen met een 27ste programma welke het meeste overeenstemming had met de criteria die werden gebruikt in de NLST. Deze analyses werden gebruikt om de United States Preventive Services Task Force (USPSTF)

informatie te geven om aanbevelingen over het invoeren van longkankerscreening in de Verenigde Staten te formuleren. Het programma wat voorstelde om individuen tussen de 55 en 80 jaar, die tenminste 30 pak-jaren gerookt hebben en huidig roker zijn of minder dan 15 jaar geleden zijn gestopt met roken, jaarlijks te screenen werd beoordeeld als het programma wat de meest voordelige verhouding tussen de voor- en nadelen van screening gaf. Dit scenario zou de sterfte aan longkanker met 14.0% reduceren, 497 sterfgevallen aan longkanker voorkomen en per voorkomen longkankersterfgeval 10.6 gewonnen levensjaren opleveren. Er zouden echter 286,813 CT screenings benodigd zijn (waarvan 23.6% een vals-positief resultaat leveren) en 190 longkankers zouden worden overgediagnosticeerd (9.9% van alle door screening ontdekte gevallen) per 100,000 individuen. De USPSTF beveelt momenteel longkankerscreening in de Verenigde Staten aan op basis van de criteria van dit programma.

In **Hoofdstuk 6** werd met behulp van modelleren onderzocht hoe overdiagnose de selectie van efficiënte longkankerscreeningsprogramma's beïnvloed. De modelschattingen gaven aan dat overdiagnose toeneemt met kortere tijdsintervallen tussen screenings, oudere start- en stopleeftijden voor screening en voor toenames in het aantal pak-jaren vereist om in aanmerking te komen voor screening. Efficiënte scenario's welke het aantal voorkomen longkankersterfgevallen maximaliseerden gaven de voorkeur aan een stopleeftijd van 80, terwijl scenario's die het aantal voorkomen longkankersterfgevallen per overgediagnosticeerde longkanker maximaliseerden de voorkeur gaven aan een stopleeftijd van 75. Over het algemeen leverde een toename in de stopleeftijd voor screening van leeftijd 75 naar leeftijd 80 een hogere reductie voor longkankersterfte op ten koste van een proportioneel hogere toename van het aantal overgediagnosticeerde longkankers.

In **Hoofdstuk 7** werden de hoofdobstakels voor de optimale implementatie van longkankerscreeningsprogramma's benoemd. Ten eerste heeft de implementatie van een grootschalig screeningsprogramma op basis van risicostratificatie tot nu toe weinig overweging heeft gehad. Het is op dit moment onduidelijk op welke manier personen met een hoog risico op het ontwikkelen van longkanker het beste geïdentificeerd en benaderd kunnen worden. Bovendien zouden deze personen mogelijk moeilijker te benaderen en minder geneigd zijn om deel te nemen aan screeningsprogramma's dan individuen met een

lager risico. Het tweede hoofdobstakel is het hoge aantal vals-positieve resultaten bij longkankerscreeningsprogramma's. In de NLST waren 23.3% van alle CT screenings en 96.4% van alle CT screenings met een positieve uitslag fout-positief. Hoewel er verscheidene aanpassingen aan de richtlijnen voor het evalueren van long nodules zijn voorgesteld om het aantal fout-positieve resultaten terug te brengen, is het onduidelijk hoeveel deze aanpassingen de sensitiviteit voor het detecteren van longkanker zouden verminderen.

Deel 3: Optimalisatie door risico stratificatie

In **Hoofdstuk 8** werd onderzocht of de voordelen voor het screenen van nooit-rokers met een verhoogd risico op het ontwikkelen van longkanker opweegt tegen de nadelen. De voor- en nadelen van het screenen van cohorten van nooit-rokers met verschillende relatieve risiconiveaus voor het ontwikkelen van longkanker ten opzichte van nooit-rokers met een gemiddeld risiconiveau werd onderzocht door middel van modellering. De modelschattingen gaven aan dat nooit-rokers tenminste een 15 keer zo hoog risico voor het ontwikkelen van longkanker ten opzichte van nooit-rokers met een gemiddeld risiconiveau moesten hebben om een vergelijkbare balans tussen de voor- en nadelen van screening te hebben als personen die voldoen aan de door de USPSTF aanbevolen criteria voor screening. Het merendeel van de longkanker risico-predictie modellen die toepasbaar zijn op nooit-rokers beschouwen echter maximale relatieve risico's van minder dan 15 voor nooit-rokers. Dit suggereert dat voor de meerderheid van nooit-rokers de voordelen van longkankerscreening niet opwegen tegen de nadelen.

Hoofdstuk 9 onderzocht de kosteneffectiviteit van longkankerscreening in de Canadese provincie Ontario. Ontario-specifieke gegevens over de populatie, rookgedrag en kosten werden gebruikt om 576 longkankerscreeningsprogramma's te evalueren voor personen geboren tussen 1940 en 1969. De 576 onderzochte programma's evalueerden verschillende studie-opzetten, met verschillende start- en stopleeftijden voor screening, het vereiste rookgedrag om in aanmerking te komen voor screening en verschillende tijdsintervallen tussen de screenings. Jaarlijkse screening voor personen tussen de 55 en 75 jaar, die tenminste 40 pak-jaren gerookt hebben en huidig roker zijn of minder dan 10 jaar geleden zijn gestopt met roken, werd beoordeeld als het optimale programma om te implementeren. Dit programma had een incrementele kosteneffectiviteit ratio van $41,136

Canadese dollars per gewonnen levensjaar. Als dit programma geïmplementeerd zou worden, zou 9.56% (499,261 personen) van de personen in de onderzochte cohorten (ooit- en nooit-rokers) tenminste één keer onderzocht worden, wat 4,788,523 CT screenings zou vereisen. Dit zou echter de longkankersterfte in deze cohorten met 9.05% kunnen verminderen, waardoor 13,108 sterfgevallen aan longkanker voorkomen zouden kunnen worden. Programma's die personen selecteerden met zwaarder rookgedrag waren kosten-effectiever dan programma's die personen selecteerden met lichter rookgedrag. Tegen de verwachting in hadden programma's met jaarlijkse tijdsintervallen tussen screenings een gunstigere kosteneffectiviteit dan programma's met tweejaarlijkse tijdsintervallen tussen screenings.

Hoofdstuk 7 gaf aan dat een van de hoofdobstakels voor het implementeren van een longkankerscreeningsprogramma de identificatie van personen met een hoog risico op het ontwikkelen van longkanker is. In **Hoofdstuk 10** werd getracht dit obstakel te overkomen door negen eerder ontwikkelde risico-predictie modellen te evalueren op hun vermogen om deelnemers van de NLST en de PLCO met een hoog risico op het ontwikkelen van of sterven aan longkanker te identificeren. De risicoschattingen van alle modellen evenaarden het geobserveerde gemiddelde risico en de risiconiveaus voor verschillende groepen in deze twee onderzoeken. De modellen verschilden echter in hun vermogen om individuen die longkanker ontwikkelden te onderscheiden van individuen die geen longkanker ontwikkelden (discriminatie). Alle modellen hadden een groter vermogen om te discrimineren tussen deze twee groepen in de PLCO in vergelijking met de NLST. Door de modellen te gebruiken kon voor een wijde reeks aan risicodrempels een betere balans tussen voor- en nadelen behaald worden in vergelijking met de deelname criteria van de NLST. Tenslotte hadden alle modellen zowel een hogere sensitiviteit als specificiteit voor risicodrempels die een vergelijkbaar aantal individuen voor screening selecteerden in de PLCO als de deelname criteria van de NLST (welke een sensitiviteit had van 71.4% en een specificiteit van 62.2%).

Hoofdstuk 11 liet zien hoe risicostratificatie gebruikt kan worden om nodule management richtlijnen te optimaliseren. Gegevens van het Nederlands-Leuvens Longkanker Screenings Onderzoek (NELSON) werd gebruikt om het risico op het ontwikkelen van longkanker binnen

twee jaar na een CT scan te kwantificeren op basis van het volume van de gedetecteerde longnodule, de volume verdubbelingstijd en de door volume-metingen afgeleide diameter van de nodule. De analyses gaven aan dat het risico voor individuen met een nodule met een volume van 100 mm^3 of kleiner (0.6% risico) of een nodule met een maximale dwarsdoorsnede van kleiner dan 5 millimeter (0.4% risico) niet significant afweek van individuen bij wie geen nodules werden gedetecteerd (0.4% risico). Individuen met nodules met een volume van 100-300 mm^3 of met een diameter van 5-10 mm hadden een intermediair risico voor longkanker (respectievelijk 2.4% en 1.3% risico). Individuen met nodules met een volume van 300 mm^3 of meer, of een diameter van 10 mm of meer hadden een hoog risico op het ontwikkelen van longkanker (respectievelijk 16.9% en 15.2% risico). Door middel van de volume verdubbelingstijd konden deze risicogroepen verder gestratificeerd worden: nodules met een volume verdubbelingstijd van 600 dagen of meer hadden een risico op het ontwikkelen van longkanker van 0.8%, terwijl dit 4.0% was voor nodules met een volume verdubbelingstijd van 400-600 dagen en 9.9% voor nodules met een volume verdubbelingstijd van 400 dagen of minder. Een nodule management richtlijn op basis van volume en volume verdubbelingstijd had een vergelijkbare sensitiviteit, maar een hogere specificiteit in vergelijking met de voorgestelde richtlijnen van de American College of Chest Physicians.

Discussie

De algemene bespreking van dit proefschrift beantwoorde en ging dieper in op de onderzoeksvragen die gesteld werden in de introductie van dit proefschrift. De analyses die zijn beschreven in de hoofdstukken van dit proefschrift toonden aan dat modellen gebruikt kunnen worden om informatie van populaties met verschillende risiconiveaus en screeningsgeschiedenissen te synthetiseren om het preklinische verloop van longkanker te bestuderen. Deze informatie kan gebruikt worden om schattingen te geven voor de voor- en nadelen van longkankerscreeningsprogramma's op de lange termijn, waardoor programma's met een voordelige balans tussen de voor- en nadelen geïdentificeerd kunnen worden. Tenslotte tonen de analyses die in dit proefschrift beschreven zijn aan dat risicostratificatie op verschillende wijzen gebruikt kan worden om longkankerscreeningsprogramma's te optimaliseren. Ten eerste kunnen verschillende risicogroepen en de variatie in risiconiveaus tussen deze groepen geïdentificeerd en

geanalyseerd worden. De analyses van dit proefschrift lieten zien dat het selecteren van individuen met een hoog risico op het ontwikkelen van longkanker essentieel is om longkankerscreening kosteneffectief te implementeren. Tenslotte zou risico stratificatie door middel van de uitslagen van de screeningstest gebruikt kunnen worden om de hoeveelheid fout-positieve resultaten en vervolgonderzoeken te reduceren. Toekomstig onderzoek zou zich moeten richten op het verzoenen van de resultaten van de NLST en andere gerandomiseerde onderzoeken naar longkankerscreening met CT. Het effectief invoeren van longkankerscreening op basis van risicostratificatie vereist verder onderzoek op basis van de stappen die zijn gezet in dit proefschrift. Nodule management richtlijnen vereisen verdere optimalisatie om het aantal fout-positieve resultaten terug te dringen; bovendien zullen deze richtlijnen in de klinische praktijk geïmplementeerd en nagevolgd moeten worden. Tenslotte zou het mogelijk kunnen zijn om het screeningssysteem te personaliseren, vooral op het gebied van de tijdsinterval tussen screenings. Hierdoor zou de hoeveelheid screeningsonderzoeken gereduceerd kunnen worden, wat de kosteneffectiviteit van longkankerscreening zou kunnen verbeteren.

About the author

Kevin ten Haaf was born on January 31st 1988, in Breda, the Netherlands. In 2006 he completed his secondary education (pre-university education, VWO) at “de Nassau Scholengemeenschap” in Breda. In the same year, he started the study “Econometrics and Management Science” at the Erasmus University Rotterdam. He obtained his Master of Science degree in Econometrics in 2011, with a specialization in “Quantitative Marketing”. For his Master’s thesis he investigated whether the use of structural equation models could improve the estimate for the number of media campaign contacts an individual has with an advertising campaign. This research was based on an internship at Pointlogic Rotterdam. From May 2011 to May 2016, he was employed as a researcher at the department of Public Health at the Erasmus University Medical Center in Rotterdam. From May 2016 to July 2016 he briefly left the department of Public Health to work as a Research Associate at the department of Pharmacy at the National University of Singapore, to investigate the cost-effectiveness of Singapore’s national breast cancer screening program. In August 2016 he returned to the department of Public Health at Erasmus, where he will continue his activities as a researcher. He developed the MISCAN-Lung microsimulation model as part of his research at the department of Public Health at Erasmus. He used the MISCAN-Lung model to investigate the natural history of lung cancer and evaluate the effects of implementing lung cancer screening policies. The results of this research are described in this thesis.

List of peer-reviewed publications

In this thesis

1. Meza R, **ten Haaf K**, Kong CY, Erdogan A, Black WC, Tammemägi MC, Choi SE, Jeon J, Han SS, Munshi V, van Rosmalen J, Pinsky P, McMahon PM, de Koning HJ, Feuer EJ, Hazelton WD, Plevritis SK.
   Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials.
   Cancer 2014; **120**(11): 1713-24.

2. **ten Haaf K**, van Rosmalen J, de Koning HJ.
   Lung cancer detectability by test, histology, stage, and gender: estimates from the NLST and the PLCO trials.
   Cancer Epidemiology Biomarkers & Prevention 2015; **24**(1): 154-61.

3. **ten Haaf K**, de Koning HJ
   Overdiagnosis in lung cancer screening: why modeling is essential.
   Journal of Epidemiology and Community Health 2015; **69**(11): 1035-39.

4. McMahon PM, Meza R, Plevritis SK, Black WC, Tammemägi CM, Erdogan A, **ten Haaf K**, Hazelton W, Holford TR, Jeon J, Clarke L, Kong CY, Choi SE, Munshi VN, Han SS, van Rosmalen J, Pinsky PF, Moolgavkar S, de Koning HJ, Feuer EJ.
   Comparing benefits from many possible computed tomography lung cancer screening programs: extrapolating from the National Lung Screening Trial using comparative modeling.
   PLOS ONE 2014; **9**(6): e99978.

5. de Koning HJ, Meza R, Plevritis SK, **ten Haaf K**, Munshi VN, Jeon J, Erdogan SA, Kong CY, Han SS, van Rosmalen J, Choi SE, Pinsky PF, Berrington de Gonzalez A, Berg CD, Black WC, Tammemägi MC, Hazelton WD, Feuer EJ, McMahon PM.
   Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force.
   Annals of Internal Medicine 2014; **160**(5): 311-20.

6. Han SS, **ten Haaf K**, Hazelton WD, Munshi VN, Jeon J, Erdogan SA, Johanson C, McMahon PM, Meza R, Kong CY, Feuer EJ, de Koning HJ, Plevritis SK.
   The impact of overdiagnosis on the selection of efficient lung cancer screening strategies.
   International Journal of Cancer 2017; **140**(11): 2436-43.

7. van der Aalst CM, **ten Haaf K**, de Koning HJ.
   Lung cancer screening: latest developments and unanswered questions.
   The Lancet Respiratory Medicine 2016; **4**(9): 749-61.

8. **ten Haaf K**, de Koning HJ
   Should never-smokers at Increased risk for lung cancer be screened?
   Journal of Thoracic Oncology 2015; **10**(9): 1285-91.

9. **ten Haaf K**, Tammemägi MC, Bondy SJ, van der Aalst CM, Gu S, McGregor SE, Nicholas G, de Koning HJ, Paszat LF.
   Performance and cost-effectiveness of computed tomography lung cancer screening scenarios in a population-based setting: a microsimulation modeling analysis in Ontario, Canada.
   PLOS Medicine 2017; **14**(2): e1002225.

10. **ten Haaf K**, Jeon J, Tammemägi MC, Han SS, Kong CY, Plevritis SK, Feuer EJ, de Koning HJ, Steyerberg EW, Meza R.
    Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study.
    PLOS Medicine 2017; **14**(4): e1002277.

11. Horeweg N, van Rosmalen J, Heuvelmans MA, van der Aalst CM, Vliegenthart R, Scholten ET, **ten Haaf K**, Nackaerts K, Lammers JW, Weenink C, Groen HJ, van Ooijen P, de Jong PA, de Bock GH, Mali W, de Koning HJ, Oudkerk M.
    Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening.
    The Lancet Oncology 2014; **15**(12): 1332-41.

Other publications

12. **ten Haaf K**, de Koning HJ
    Clarifying assumptions and outcomes in cost-effectiveness analyses.
    JAMA Oncology 2016; **2**(2): 277-78.

13. **ten Haaf K**, Steyerberg EW.
    Methods for individualized assessment of absolute risk in case-control studies should be weighted carefully.
    European Journal of Epidemiology 2016; **31**(11): 1067-68.

14. Ripping TM, **ten Haaf K**, Verbeek ALM, van Ravesteyn NT, Broeders MJM
    Quantifying overdiagnosis in cancer screening: a systematic review to evaluate the methodology.
    Journal of the National Cancer Institute 2017; **109**(10)

15. Horeweg N, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers JW, Nackaerts K, Vliegenthart R, **ten Haaf K**, Yousaf-Khan UA, Heuvelmans MA, Thunnissen E, Oudkerk M, Mali W, de Koning HJ.
    Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers.
    The Lancet Oncology 2014; **15**(12): 1342-50.

16. Yousaf-Khan U, Horeweg N, van der Aalst C, **ten Haaf K**, Oudkerk M, de Koning H.
    Baseline characteristics and mortality outcomes of control group participants and eligible non-responders in the NELSON lung cancer screening study.
    Journal of Thoracic Oncology 2015; **10**(5): 747-53.

17. Ripping TM, Verbeek AL, **ten Haaf K,** van Ravesteyn NT, Broeders MJ.
    Extrapolation of pre-screening trends: impact of assumptions on overdiagnosis estimates by mammographic screening.
    Cancer Epidemiology 2016; **42**:147-53.

18. Walter JE, Heuvelmans MA, de Jong PA, Vliegenthart R, van Ooijen PM, Peters RB, **ten Haaf K**, Yousaf-Khan U, van der Aalst CM, de Bock GH, Mali W, Groen HJ, de Koning HJ, Oudkerk M.
    Occurrence and lung cancer probability of new solid nodules at incidence screening with low-dose CT: analysis of data from the randomized, controlled NELSON trial.
    The Lancet Oncology 2016; **17**(7): 907-16.

19. Yousaf-Khan U, van der Aalst C, de Jong PA, Heuvelmans M, Scholten E, Lammers JW, van Ooijen P, Nackaerts K, Weenink C, Groen H, Vliegenthart R, **ten Haaf K**, Oudkerk M, de Koning H.
    Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval.
    Thorax 2017;**72**(1): 48-56.

20. Yousaf-Khan U, van der Aalst C, de Jong PA, Heuvelmans M, Scholten E, Walter J, Nackaerts K, Groen H, Vliegenthart R, **ten Haaf K**, Oudkerk M, de Koning H.
    Risk stratification based on screening history: the NELSON lung cancer screening study.
    Thorax 2017; Epub ahead of print.

PhD portfolio

Name of PhD student: Kevin ten Haaf

Erasmus MC Department: Public Health

PhD period: 2011-2017

Promotor: Prof.dr. H.J. de Koning

PhD training	Period	Workload
Courses at the Netherlands Institute for Health Sciences (NIHES)		
Erasmus Summer Program		
Principles of Research in Medicine	2011	0.7 ECTS
Methods of Public Health Research	2011	0.7 ECTS
Advanced Short Courses		
Planning and Evaluation of Screening	2011	1.4 ECTS
Cancer Epidemiology	2012	1.4 ECTS
Seminars and symposia attended		
NELSON lung cancer screening symposium, Rotterdam, the Netherlands	2011	8 hours
Methodologie van patiëntgebonden-onderzoek en voorbereiding subsidieaanvragen, Erasmus MC, Rotterdam, the Netherlands	2012	8 hours
European meeting on lung cancer screening trials, Rotterdam, the Netherlands	2012	12 hours
Lung Cancer Summit, Ontario, Canada	2014	8 hours
Health Technology Assessment of CT Lung Screening meeting, Ontario, Canada	2014	8 hours
Breast Cancer Screening Symposium, National University of Singapore, Singapore	2015	8 hours
Workshop Scientific Integrity, Erasmus MC, Rotterdam, the Netherlands	2016	8 hours
Teach the Teacher, Erasmus MC Desiderius School, Rotterdam, the Netherlands	2016	16 hours
Starten met leidinggeven in de wetenschap, Erasmus MC, Rotterdam, the Netherlands	2016	16 hours
Seminars at the Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2011-2017	120 hours
(Inter-)national conferences attended		
15th World Conference on Lung Cancer, Sydney, Australia	2013	40 hours

International Cancer Screening Network Meeting, Rotterdam, the Netherlands	2015	20 hours
3rd Preventing Overdiagnosis Conference, Bethesda, United States	2015	20 hours
16th Annual Conference of the Society for Research on Nicotine & Tobacco (SRNT) Europe, Maastricht, the Netherlands	2015	20 hours
2016 World Cancer Congress, Paris, France	2016	12 hours
17th World Conference on Lung Cancer, Vienna, Austria	2016	40 hours
Presentations and workshops provided		
Methodology club, Department of Public Health, Erasmus MC, Rotterdam, the Netherlands. Oral presentation: "Modeling lung cancer mortality reduction by CT screening: upgrading and calibrating MISCAN-Lung to the National Lung Screening Trial"	2012	14 hours
European meeting on lung cancer screening trials, Rotterdam, the Netherlands. Two oral presentations: "Consequences of trial differences for modeling lung cancer incidence" and a "Confidential" presentation	2012	56 hours
15th World Conference on Lung Cancer, Sydney, Australia. Oral presentation: "The natural history and screen-detectability of lung cancer: estimates from NLST and the PLCO"	2013	14 hours
Health Technology Assessment of CT Lung Screening meeting, Ontario, Canada. Oral presentation: "Modeling Health Outcomes of CT Lung Screening in Ontario"	2014	28 hours
Research meeting department of Public Health Erasmus University Medical Center. Oral presentation: "Should never-smokers at increased risk for lung cancer be screened?"	2014	14 hours
International Cancer Screening Network Meeting, Rotterdam, the Netherlands. Workshop (invited): "Workshop on Microsimulation Modeling for Cancer Screening"	2015	28 hours
International Cancer Screening Network Meeting, Rotterdam, the Netherlands. Oral presentation: "The (cost-)effectiveness of lung cancer screening in Ontario, Canada"	2015	28 hours
Breast Cancer Screening Symposium, National University of Singapore, Singapore. Workshop (invited): "Interactive hands-on workshop on the Erasmus MC MISCAN-Fadia Model"	2015	28 hours
3rd Preventing Overdiagnosis Conference, Bethesda, United States. Workshop: "Quantifying overdiagnosis in cancer screening: evaluation of study designs"	2015	28 hours

2016 World Cancer Congress, Paris, France. Oral presentation (invited, provided on behalf of Martin Tammemägi): "The evidence on (risk stratification in) lung cancer screening, and next steps"	2016	28 hours
17th World Conference on Lung Cancer, Vienna, Austria. Oral presentation (invited): "The Potential of Microsimulation Modeling in Lung Cancer Screening"	2016	28 hours
Cancer Intervention and Surveillance Modeling Network (CISNET) meetings, National Cancer Institute, various locations in the United States	2012-2017	210 hours
Teaching activities		
Reviewing Bachelor essays of 3rd year medical students, Bachelor in Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands	2012-2014	84 hours
Teaching assistant: Choices in health care, Bachelor in Medicine, Erasmus MC University Medical Center, Rotterdam, Netherlands.	2016	28 hours
Lecturer: Planning and Evaluation of Screening, NIHES, Rotterdam, the Netherlands	2015-2017	84 hours
Other		
Peer-review for a number of international journals: BMC Cancer, British Medical Journal, European Journal of Epidemiology, JAMA Oncology, Journal of Thoracic Oncology, Risk Analysis	2015-2017	60 hours
Ad hoc scientific evaluator for the National Cancer Institute of the United States	2015	6 hours
Junior researcher representative; section Early Detection 1 Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2012-2015	60 hours
Total	2011-2017	46.7 ECTS*

* 28 working hours represents 1 ECTS

Dankwoord & Acknowledgements

Wat al die jaren zo ver weg leek te zijn, is nu eindelijk klaar: mijn proefschrift! Dit proefschrift had er echter niet kunnen zijn zonder de directe en indirecte bijdrages van vele personen. Daarom wil ik graag iedereen bedanken die heeft bijgedragen aan de totstandkoming van dit proefschrift.

Ten eerste wil ik mijn promotor bedanken: professor dr. Harry J. de Koning. Beste Harry, jij gaf mij de mogelijkheid om te kunnen promoveren. Door de jaren heen heb ik jouw leren kennen als iemand die haarfijn de sterke en zwakke punten van een onderzoek kan aanwijzen. Je inzichten in het formuleren van de hoofdboodschap van een onderzoek op een duidelijke en strategische manier hebben mij enorm gevormd in mijn wetenschappelijke ontwikkeling. Je hebt mij alle vrijheid en vertrouwen gegeven in mijn ontwikkeling als onderzoeker, waarvoor ik erg dankbaar ben. Ik heb onze samenwerking tijdens deze jaren als enorm prettig ervaren, mede door de ruimte voor humor tijdens onze discussies. Ik kijk er erg naar uit om deze samenwerking in de toekomst voort te zetten.

Een ander belangrijk persoon die mij gevormd heeft in mijn wetenschappelijke ontwikkeling is Dr. Joost van Rosmalen. Joost, jij hebt mij samen met Harry de eerste periode van mijn promotietraject begeleid. Je kennis op het gebied van (bio-)statistiek is bewonderingswaardig en ik heb veel geleerd van jouw inzichten. Ik waardeer je oprechtheid en humor tijdens onze (niet-)wetenschappelijke discussies. Ik hoop in de toekomst nog op vele vlakken met je van gedachten te kunnen wisselen.

Ik wil de leden van de kleine commissie, professor dr. Joachim Aerts, professor dr. Matthijs Oudkerk en professor dr. Karl Moons heel hartelijk bedanken voor het beoordelen van dit proefschrift. Daarnaast wil ik de overige commissieleden hartelijk bedanken voor hun interesse in dit proefschrift en de bereidheid om hierover met mij van gedachten te wisselen.

Het is een voorrecht geweest om tijdens het schrijven van dit proefschrift kennis te hebben mogen maken met mijn paranimfen Suzette Matthijsse en Carlijn van der Aalst. Suzette, wij zijn sinds het begin van mijn promotietraject kamergenoten geweest. Gedurende die tijd hebben we heel wat ups, en ook wat downs gedeeld. Ik kon altijd bij je terecht voor advies,

aanmoediging en een hele dosis gezelligheid. Ik had geen betere kamergenote kunnen treffen! Carlijn, jou heb ik beter leren kennen nadat ik meer bij NELSON betrokken raakte. Ik heb grote waardering voor je doorzettingsvermogen en oprechtheid. Ik ben enorm dankbaar dat je deur altijd voor me openstaat. Ik prijs mij zeer gelukkig om met jou te mogen samenwerken!

Ik wil alle (oud-)collega's van de afdeling MGZ, de screensectie en de MGZ-collega's binnen de NELSON trial (Marianne, Roel, Frank, Erik) bedanken voor de gezellige gesprekken de afgelopen jaren. In het bijzonder wil ik Tiago Marques, Lea Jabberian, Uraujh Yousaf-Khan, Reinier Meester, Nanda Horeweg en Arry de Bruijn bedanken. Tiago, het was een plezier om jou als kamergenoot te hebben. De vele avonden die ik op kantoor doorbracht waren een stuk aangenamer door jouw aanwezigheid! Lea, ik ben erg gesteld op je vrolijke persoonlijkheid en de wijze waarop je iemand een hart onder de riem weet te steken. Uraujh, de gezellige gesprekken (met een soms wat plagerige insteek) tijdens dit promotietraject had ik niet willen missen. Ik hoop je op zeer korte termijn jouw proefschrift te zien verdedigen! Reinier, ik kijk met veel plezier terug naar de afwisselend serieuze en niet-serieuze gesprekken die we hebben gehad. Ik hoop dat je tijd aan Stanford je goed gaat bevallen! Nanda, ik heb aan het begin van mijn promotietraject veel van je geleerd over de klinische aspecten van longkanker. Bovendien kon ik altijd bij je terecht voor een gezellig gesprek! Mevrouw de Bruijn, lieve Arry; hartelijk dank voor de hulp bij het inplannen van meetings met Harry, praktische vragen en de morele ondersteuning. De screensectie boft met jou!

During the past years I have had the opportunity to collaborate with a number of international groups, in particular the CISNET-Lung working group. Many of the chapters of this thesis were derived from the CISNET collaboration. Therefore I would like to thank all members of the CISNET-Lung working group for their contributions to this thesis, as well as the interesting discussions and enjoyable in-person meetings. In particular, I would like to thank Rafael Meza, Jihyoun Jeon, Pamela McMahon, Chung Yin Kong Bill Hazelton, Suresh Moolgavkar, Sylvia Plevritis, Summer Han and Eric Feuer (Rocky). I have also had the opportunity to collaborate with a number of Canadian partners, which resulted in Chapter 9 of this thesis. I would like to thank Dr. Garth Nicholas and Dr. Elizabeth McGregor for their

valuable insights during this collaboration. Furthermore, I would like to thank Susan Bondy for her aid in modeling Ontarian smoking patterns and the enjoyable anecdotes she shared. Finally, I am much obliged to Dr. Lawrence Paszat, who initiated this collaboration. Lawrence, you provided me a great deal of support and freedom in the research that resulted in Chapter 9. It was an absolute pleasure to collaborate with you and I hope to collaborate with you again in the near future! I have had the pleasure to collaborate with professor Martin Tammemägi within CISNET and the earlier mentioned Canadian collaboration. Martin, your keen insights and suggestions provide much food for thought and helped to improve many of the Chapters of this thesis. In addition to our research-related conversations, I greatly enjoy the non-research related conversations we have and always look forward to them when we meet in person. I would like to thank Dr. Hwee Lin Wee for providing the opportunity to perform research at the National University of Singapore during my PhD. I very much enjoyed my time in Singapore and our pleasant in-person meetings. I look forward to successfully completing our current collaboration and hope to collaborate with you on new projects in the future!

Ook binnen Nederland heb ik de mogelijkheid gehad om met een verscheidenheid aan mensen te mogen werken, zoals Professor Ewout Steyerberg en Dr. Dorien Hoeve-Ripping. Professor Steyerberg, beste Ewout, hartelijk dank voor de gezellige gesprekken waarin we van gedachten wisselden op het gebied van risico-predictie modellen, die uiteindelijk tot Hoofdstuk 10 van dit proefschrift hebben geleid. Dorien, het was een plezier om met je samen te werken aan de verschillende projecten over overdiagnose!

Mijn vrienden, de leden van de schermvereniging en de leden van de badmintonvereniging wil ik bedanken voor de gezellige etentjes, verjaardagen, spelletjesavonden en uitjes. In het bijzonder wil ik Jasper, Mathijs, Tirza, Michael, Sandy, Charlie, Ann, Yvonne en Martijn bedanken voor hun vriendschap en steun.

Mijn familie wil ik bedanken voor de steun die ze mij al deze jaren hebben gegeven. Mijn nicht Naomi wil ik in het bijzonder bedanken voor het ontwerpen van de cover van dit proefschrift! Ik mag mij gelukkig prijzen met de steun en aanmoedigingen van mijn broer Brian en schoonzus Esther. Mijn nichtje Evy en neefje Thijs heb ik tijdens dit promotietraject

mogen ontmoeten. Lieve Evy en Thijs, jullie brengen altijd een glimlach op mijn gezicht als ik jullie zie. Het is een enorm plezier om jullie oom te zijn!

Tot slot wil ik mijn ouders bedanken. Pa en Ma, ik ben enorm dankbaar dat jullie er altijd voor mij zijn. Jullie hebben mij geleerd om nooit op te geven en het uiterste uit mezelf te halen. Ik hou zielsveel van jullie.