Aus dem Deutschen Krebsforschungszentrum Heidelberg

(Wissenschaftlicher Stiftungsvorstand: Prof. Dr. Michael Baumann)

Abteilung Epidemiologie von Krebserkrankungen

(Leiter: Prof. Dr. Rudolf Kaaks)

Evaluation of Risk Models and Biomarkers for the Optimization of Lung Cancer Screening

Inauguraldissertation

zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)

an der

Medizinischen Fakultät Heidelberg

Der

Ruprecht-Karls-Universität

vorgelegt von

Sandra González Maldonado

aus

Mexiko-Stadt, Mexiko

2021

Dekan: Prof. Dr. Hans-Georg Kräusslich
Doktorvater: Prof. Dr. Rudolf Kaaks

To my grandparents Marcos Maldonado Gómez and América González Carranza

Papá Marcos, abuelita América: a ustedes éste y todos mis logros, que son tan míos
como suyos. Con cada paso en mi camino busco honrar su memoria.
Todo mi amor hasta ese bello lugar en el que descansan juntos.

To my parents Víctor and María Concepción

To my husband Lukas
Danke für alles!

TABLE OF CONTENTS

ABBREVIATIONS

AIC	Akaike Information Criterion
AIS	adenocarcinoma *in situ*
AME	AME Thoracic Surgery Collaborative Group trial
AUC	Area Under the Curve
BAC	bronchiolo-alveolar carcinoma
BC	baseline control group
BCCA	British Columbia Cancer Agency
BIO-MILD	Biological Randomized Multicentre Italian Lung Detection Trial
BMI	body mass index
BS	Brier Score
c/d	cigarettes per day
CAD	computer-aided detection
CARET	Carotene and Retinol Efficacy Trial
CI	confidence interval
CISNET	Cancer Intervention and Surveillance Modeling Network
COPD	chronic obstructive pulmonary disease
COSMOS	Continuous Observation of SMOking Subjects
CPS	Cancer Prevention Studies
CT	Computed Tomography
CTFPHC	Canadian Task Force on Preventive Health Care
CXR	chest X-ray
D	diameter
DANTE	Detection And screening of early lung cancer by Novel imaging Technology and molecular Essays
DCO	death certificate only
DKFZ	Deutsches Krebsforschungszentrum (German Cancer Research Center)
DLCST	Danish Lung Cancer Screening Trial
Dmax	longest diameter
Dmean	average diameter
Dmin	shortest diameter
DNA	deoxyribonucleic acid
ECLS	Early Cancer detection test Lung cancer Scotland
EI	emphysema index
I-ELCAP	International Early Lung Cancer Action Program
ELISA	Enzyme-Linked Immunosorbent Assay
EPIC	European Prospective Investigation into Cancer and nutrition cohort study
ERL	expected remaining lifetime
ERS	European Respiratory Society
EV	emphysema volume
FEV$_1$	1-second forced expiratory volume

FP	false positive
FPR	false positive rate
FU	Follow-up
FVC	forced vital capacity
GEE	Generalized Estimating Equations
GGO	ground glass opacity
HL	Hosmer-Lemeshow Test
HPFS	Health Professionals Follow-up Study
IC	interval cancer
ICER	incremental cost-efficiency ratio
ICD-O-3	International Classification of Diseases for Oncology, Third Edition
IQR	interquartile range
IR	incidence round
ITALUNG	Italian Lung Cancer Screening Trial
LC	lung cancer
LCDRAT	Lung Cancer Death Risk Assessment Tool
LCRAT	Lung Cancer Risk Assessment Tool
LCS	lung cancer screening
LDCT	low-dose computed tomography
LLP	Liverpool Lung Project
LLPv2	Liverpool Lung Project v2 model
LR+	Positive Likelihood Ratio
LSS	Lung Screening Study
LUSI	German Lung Cancer Screening Intervention Trial
LV	total lung volume in inspiration
LYG	life years gained
MD	mean diameter
MED	mean effective dose
MIA	Minimally Invasive Adenocarcinoma
MILD	Multicentre Italian Lung Detection
MLD	mean lung density
MLE	Maximum Likelihood Estimate
MPST	mean pre-clinical sojourn time
MSC	miRNA signature classifier
NCCN	National Comprehensive Cancer Network
NCN	non-calcified nodule
NELSON	NEderlands-Leuvens Longkanker Screenings Onderzoek (Dutch-Belgian Randomized Lung Cancer Screening Trial)
NHIS	National Health Interview Survey
NHS	Nurses' Health Study
NLST	US National Lung Screening Trial

NNS	number needed to screen
NPV	negative predictive value
NSCLC	non-small cell lung cancer
OR	odds ratio
PA	excess incidence (as proportion of all lung cancer cases)
PDP	pre-clinical detectable phase
PET	Positron Emission Tomography
PKUPH	Peking University People´s Hospital
PLCO	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial
PLCOM	PLCO Model
PPV	positive predictive value
PR	prevalence round
PS	excess incidence (as proportion of screen-detected cases)
py	pack-years
QALYS	quality-adjusted life years
QIC	quasi-Akaike Information Criterion
RNA	ribonucleic acid
mRNA	messenger RNA
miRNA	micro RNA
ROC	receiver operating characteristic
RR	relative risk, risk ratio
SCLC	small-cell lung cancer
SD	standard deviation
SDC	screen-detected cases
SNC	suspicious nodules control group
Sp. Z-Stat	Spiegelhalter Z-Statistic
Sp. Test p	Spiegelhalter Z-Test p-value
TAAb	tumor-associated autoantibodies
TPR	true positive rate
TSCE	two-stage clonal expansion
UICC	Union for International Cancer Control
UK	United Kingdom
UKLS	United Kingdom Lung Cancer Screening Trial
USA	United States of America
USPSTF	United States Preventive Services Task Force
V	volume
VA	Veterans Affairs
VATS	video-assisted thoracoscopic surgery
VDT	volume doubling time
Y	years

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTARY TABLES AND FIGURES

1 INTRODUCTION

1.1 Epidemiology of lung cancer

1.1.1 Incidence and mortality

Lung cancer has remained amongst the most commonly diagnosed types of cancer for at least the last two decades (Schabath and Cote 2019). In 2020, it was the second most common cancer type worldwide (2,206,771 new diagnoses [11.4%]), only after breast cancer (2,261,419 [11.7%]), and was also the deadliest (1,796,144 cancer-related deaths [18%]) (Ferlay J 2020b).

In Germany, according to the most recent report from the Federal Statistical Office (German Centre for Cancer Registry Data - Robert Koch Institute 2021), in 2017 lung cancer caused almost as many deaths (15.7%) as breast cancer (17.7%) among German women. Among German men, it was the first cause of cancer-related death (23.4%), surpassing both prostate (11.7%) and colorectal cancer (10.5%) combined (Figure 1) (German Centre for Cancer Registry Data - Robert Koch Institute 2021).

Although no data is publicly available for more recent years, the latest projections of German national rates reported by the World Health Organization (WHO) for the year 2020 estimated lung cancer to be the second most common cancer diagnosis among men (38,436 [11.2%]), and the third one among women (26,368 [9.3%]) (Figure 2) (Ferlay J 2020a). Regarding mortality, and considering both sexes, the projections estimated lung cancer to remain as the first cause of cancer related death (19.9% (50,282) (Figure 3) (Ferlay J 2020a).

Currently, organized screening programs for breast, cervical and colorectal cancer are offered in Germany (Bundesministerium für Gesundheit 2021; Gemeinsamer Bundesausschuss 2020; Gemeinsamer Bundesausschuss 2021b). Additionally, voluntary screening examinations for skin and prostate cancer are covered by health insurance for eligible subjects, though not in the form of structured programs (Bundesministerium für Gesundheit 2021; Gemeinsamer Bundesausschuss 2021a). However, so far, no official recommendations have been issued regarding strategies for the early detection of lung cancer.
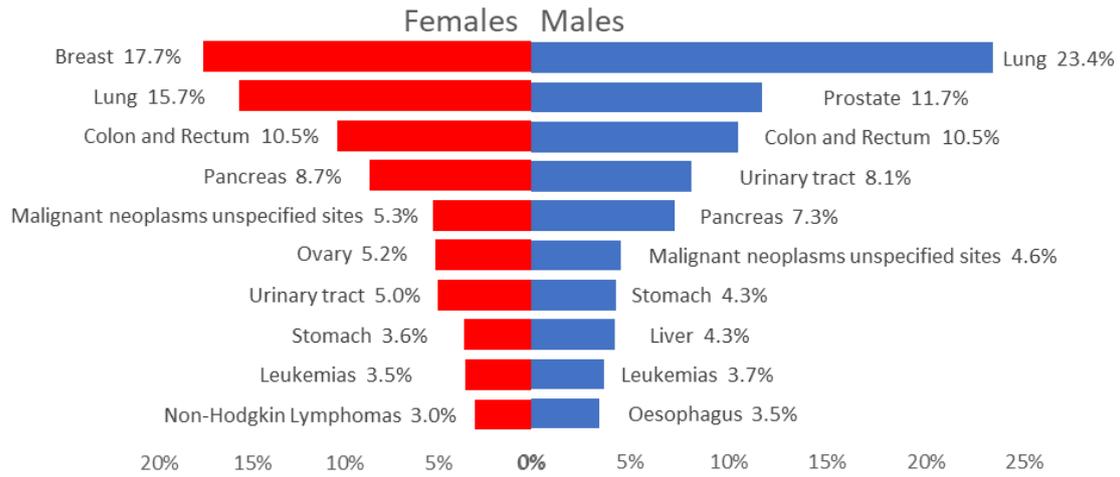
Figure 1. Distribution of cancer-related deaths in Germany 2017 across cancer types by sex.

Data source: (German Centre for Cancer Registry Data - Robert Koch Institute 2021).
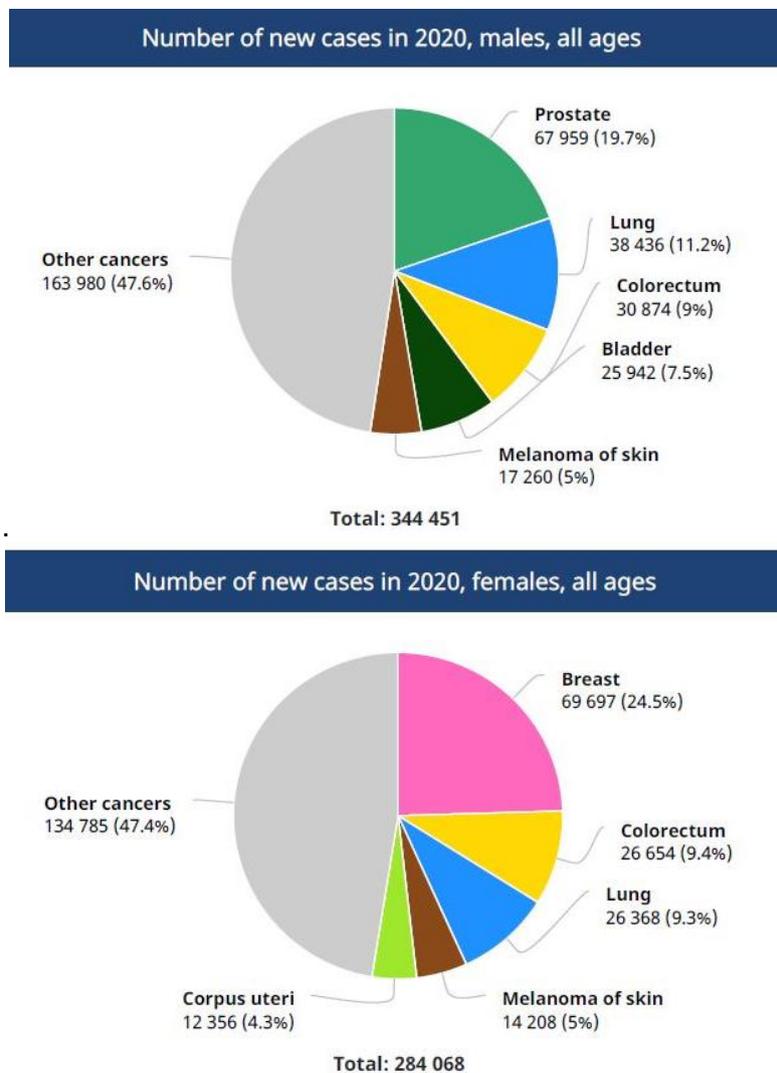
Figure 2. Distribution of new cancer cases in 2020 in Germany across cancer types by sex.

As published in (Ferlay J 2020a), reprinted with permission.
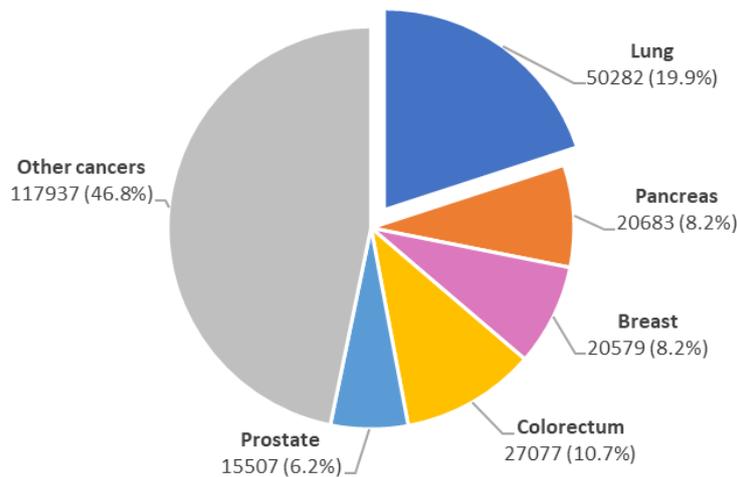
Figure 3. Distribution of cancer-related deaths in Germany in 2020 across cancer types.

Data source: (Ferlay J 2020a).

1.1.2 Histological classification

Lung tumors are classified into two major histological groups: non-small cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). Up to 85% of lung cancers are NSCLC, which can be further classified into adenocarcinomas (40%), squamous cell carcinomas (25%), large cell carcinomas (15%), and other less common subtypes (20%) (Figure 4) (Chansky et al. 2017; Rami-Porta et al. 2014; Schabath and Cote 2019; Wahbah et al. 2007).

The distribution of histologic subtypes varies by gender and also across countries (Youlden et al. 2008). In recent years, the proportion of adenocarcinomas has increased in both men and women in various countries in Europe, North America and Oceania (Travis et al. 2011). Among men, squamous cell carcinoma used to be the most common histologic subtype worldwide; however, since the mid-1990s it has been surpassed by adenocarcinoma in the USA, Canada, some European countries, and Japan (Toyoda et al. 2008), a trend that has not been observed in countries such as the Netherlands and Spain (Lortet-Tieulent et al. 2014). The increase in the number of adenocarcinomas, usually located in the peripheries of the lung, has been attributed to deeper tobacco smoke inhalation promoted by the marketing of filtered cigarettes as less harmful, mostly in western countries (Lewis et al. 2014; Warren et al. 2014). Among women, the gap between the proportion of adenocarcinomas and the proportion of other subtypes is larger than in men, which might be attributed to the preference of female smokers for filtered and "light" versions of cigarettes (Travis et al. 2011).
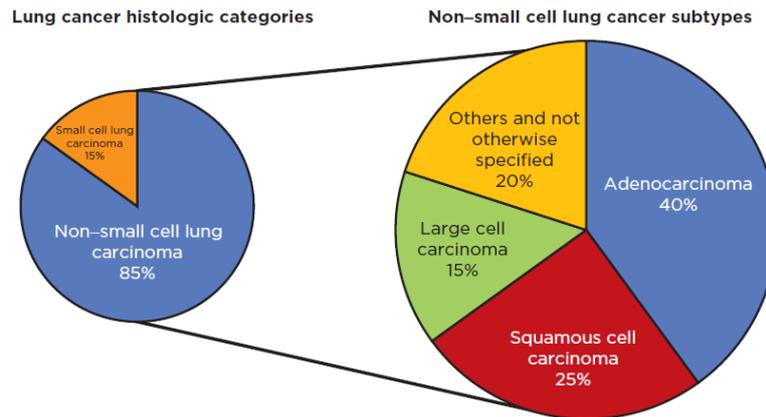
Figure 4. Histologic classification of lung cancer

As published in (Schabath and Cote 2019), reprinted with permission.

1.1.3 Risk factors

1.1.3.1 Tobacco smoking

Tobacco smoking has been identified as the main risk factor for lung cancer (Thun 2010). In western populations, it has been estimated to account for over 80% of all lung cancer diagnoses (Bray et al. 2018). Worldwide, about 75% of all lung cancer deaths among men, and 50% among women are attributed to it (American Cancer Society 2018). Not only smoking, but also indirect exposure, also known as *secondhand smoke* has been causally linked to lung cancer among never smokers, increasing the risk of exposed individuals by up to 28% compared to those unexposed (Hori et al. 2016; Kim et al. 2018).

1.1.3.2 Other risk factors

Besides tobacco smoking and secondhand tobacco smoke, other risk factors have been identified, such as environmental exposure to radon and polluted air (indoors and outdoors) (Garzillo et al. 2017; Lubin and Boice 1997), occupational exposure to asbestos, arsenic, cadmium and various substances present in the production of rubber, paving, mining, painting, as well as exposure to diesel exhaust and radiation (Alberg et al. 2013; American Cancer Society 2018; Dela Cruz et al. 2011; Driscoll et al. 2005).

Additionally, respiratory diseases associated with chronic lung tissue inflammation, such as chronic obstructive pulmonary disease (COPD), emphysema (these two being most commonly, though not always attributed to smoking), chronic bronchitis, asthma, pneumonia and tuberculosis have also been linked to lung cancer risk to various degrees and through various possible mechanisms (Brenner et al. 2011; Zhang et al. 2017). Other epidemiological risk factors are genetic susceptibility (Bailey-Wilson et al. 2004; Bosse and Amos 2018; McKay

et al. 2017) advanced age, malnutrition (Duan et al. 2015; Smith et al. 2012; Sun et al. 2016; Yang et al. 2013), and a compromised or suppressed immune system (Engels et al. 2006).

1.1.3.3 Lung cancer in non-smokers

About 25% of all lung cancer cases worldwide are found among never smokers, a high proportion of these among women (Parkin et al. 2005). Lung tumors among non-smokers are most commonly adenocarcinomas with genetic and mutation profiles significantly different from those detected among smokers (Rudin et al. 2009).

Among non-smokers, advanced age has been identified as the most significant risk factor (McCarthy et al. 2012). Therefore, with aging populations, especially in western countries such as Germany, the number of new lung cancer cases is expected to increase, regardless of the efficacy of any anti-tobacco interventions.

1.1.4 Primary prevention: tobacco control

Increased taxation, banning of smoking in public places, restrictions and regulations on marketing, public awareness and smoking cessation campaigns are some examples of tobacco control interventions (American Cancer Society 2018; Arenberg 2019; Bray et al. 2018)

In countries such as the United States, these interventions have contributed to a decrease in smoking rates (American Cancer Society 2018), a trend that is starting to emerge in other developed countries. However, smoking cessation interventions are not always successful and even if they are, their effects are noticeable only in the mid to long term since former smokers remain at elevated risk for lung cancer and lung cancer-related death over a period of several years (Halpern et al. 1993). Therefore, lung cancer is expected to remain a public health issue in the 21st century.

In the short term and for cases in which primary prevention fails and lung cancer develops, secondary prevention focuses on reducing the risk of lung-cancer related death by means of early detection, timely and accurate diagnosis and appropriate treatment assignment.

1.1.5 Secondary prevention: early detection

Given that early-stage lung cancer is usually asymptomatic, the disease is most commonly diagnosed at advanced stages of development (stages III –IV) at which survival rates are much lower compared to early stages (stage I) (Barnes et al. 2016; National Cancer Institute 2020). According to worldwide and German statistics, around 70% of lung cancers are diagnosed in advanced stages (stages III to IV) (Figure 5), at which tumors are no longer resectable (Barnes

et al. 2016; Travis et al. 2013), and in more than half of the cases at stages at which the disease has become metastatic (Sharma et al. 2015).
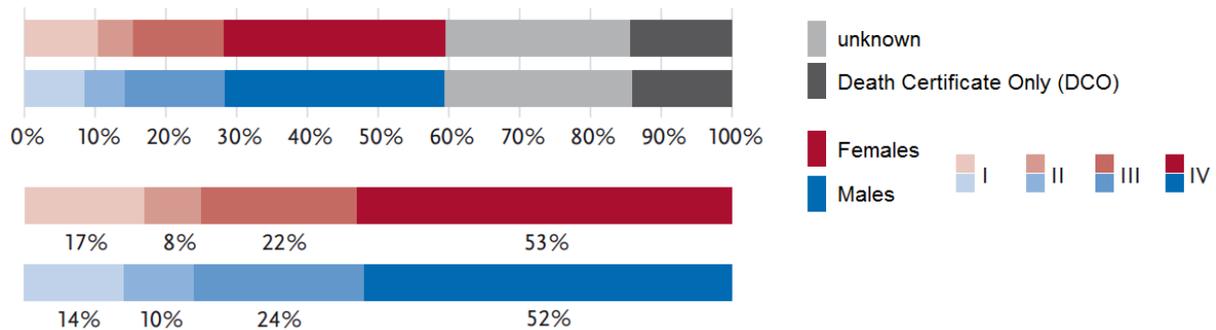
Figure 5. Distribution of lung cancer stages at first diagnosis by sex, Germany 2015-2016.

Upper plot: including "death certificate only" cases and cases with missing stage information. Lower plot: only cases with available stage information.

Adapted from (Barnes et al. 2016).

In Germany, 5-year survival rates for metastatic (stage IV) disease have been estimated at around 3% for males and 5% for females. In contrast, survival rates for the same time span are estimated at 58% and 75% for male and female patients with stage I tumors respectively (Figure 6) (Barnes et al. 2016; Goldstraw et al. 2016; Sharma et al. 2015).

Analysis on pooled data from 16 countries reported similar 5-year rates (stage IV: 0%-10% vs. stage I: 68%-92%) for all non-small cell tumors. Moreover, specific analyses on the same data have shown how these survival rates vary depending on tumor size, even within early-stage categories and as the disease progresses, with estimated 2-year survival rates for stage I ranging from 67% to 97%. (Figure 7) (Goldstraw et al. 2016).

However, it is important to note that the very high survival rates observed for the most recent definition of tumor stages IA1 and IA2 (Goldstraw et al. 2016; Nicholson et al. 2016), might be in part attributed to the earlier detection of such very small tumors (T1a: ≤1 cm in largest diameter; T1b: >1 cm but ≤ 2 cm in largest diameter) made possible by the application of imaging techniques with higher resolution. While earlier detection might extend the time from detection to death, it might not truly reflect extended survival in terms of lifespan.

Finally, analyses by tumor histology have shown the association of earlier diagnosis on survival rates to be stronger for some lung cancer subtypes. In particular, for NSCLC, 2-year survival in advanced stages (IV) is less than 20%, compared to 87% and up to 97% 2-year survival for patients with stage I tumors (Goldstraw et al. 2016), and of about 8% (stage IV) vs. 67% and up to 93% for SCLC (Figure 8) (Goldstraw et al. 2016; Nicholson et al. 2016).
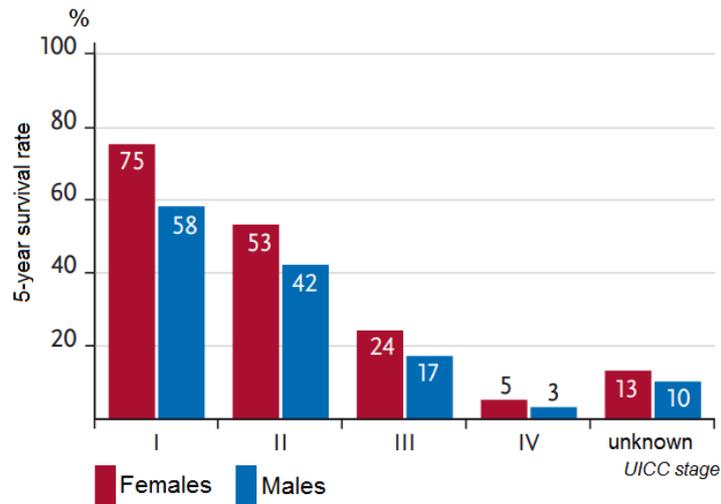
Figure 6. 5-year survival rates by lung cancer UICC stage and sex, Germany 2015-2016

Abbreviation: UICC: Union for International Cancer Control

Adapted from (Barnes et al. 2016).
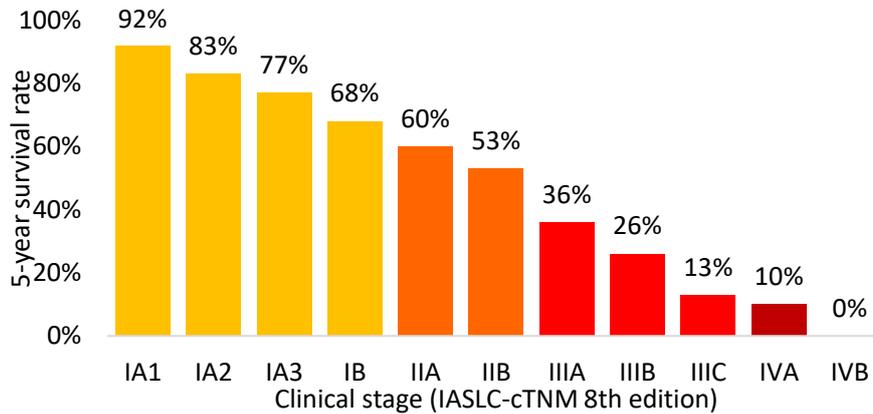
Figure 7. 5-year survival rates by lung cancer clinical stage (NSCLC)

Abbreviation: IASLC: International Association for the Study of Lung Cancer

Data source: (Goldstraw et al. 2016)
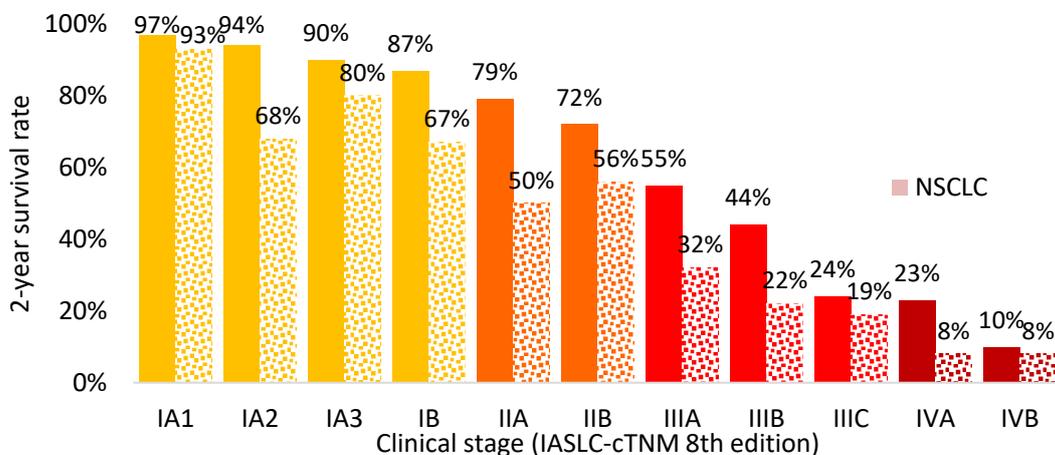
Figure 8. 2-year survival rates by lung cancer clinical stage and histology subtype

Abbreviation: IASLC: International Association for the Study of Lung Cancer

Data source: (Goldstraw et al. 2016) and (Nicholson et al. 2016)

These statistics suggest that early detection, and therefore a timely diagnosis, could improve the survival rates among lung cancer patients. Thus, with the aim of decreasing lung cancer-specific mortality, the design and implementation of effective and efficient screening programs is a topic of current discussion among the scientific community and policy makers worldwide.

1.2 Lung cancer screening with low-dose computed tomography

The ultimate goal of lung cancer screening is to reduce lung cancer specific mortality. The first step towards this goal is to find an examination method that can detect cancer at stages early enough for treatment to have a positive effect on the survival outcomes of screened subjects.

The following sections describe how low-dose computed tomography (LDCT) has come to be the detection method of choice for lung cancer screening and summarizes its risks and potential harms.

1.2.1 Early approaches with chest X-ray

Starting in the 1960s a number of clinical trials evaluated the effects of imaging-based lung cancer screening using chest X-ray (CXR) alone or in combination with other methods such as sputum cytology (Sharma et al. 2015). All such trials showed an increase in lung cancer detection through screening and some showed a shift towards detection at earlier stages, though none of them found a significant reduction in lung cancer mortality that could be attributed to CXR as a method for screening (Berlin et al. 1984; Brett 1968; Melamed et al. 1984; Tockman 1986). With the introduction of more advanced techniques with higher resolution such as computed tomography (CT), came a renewed interest amongst the scientific community for the investigation of alternative screening methods.

1.2.2 LDCT-based imaging as screening method

In the early 1990s low-dose computed tomography (LDCT), preferred to regular computed tomography (CT) due to its lower levels of radiation, was evaluated in a number of non-randomized clinical trials as an imaging method for identifying pulmonary lesions (see (Bach et al. 2003b; Sharma et al. 2015) for a summary). The largest of these trials, the *„International Early Lung Cancer Action Program (I-ELCAP)"* (N= 31,567 participants) showed that more than 80% of lung cancers identified by means of LDCT were stage I and reported a 92% 10-year survival rate for patients with disease in this stage who were surgically treated (Henschke et al. 2006).

However, neither a shift towards detection at earlier stages nor prolonged survival as measured from the time of diagnosis are enough to prove the efficacy of LDCT screening as a means to reduce lung cancer related mortality. In fact, a necessary condition to justify

screening is to prove that, thanks to early detection, the life span of screening participants becomes longer than it would have been, had the tumors been detected at a later point in time. This evidence, however, can only be obtained by analyzing data from randomized trials. Therefore, although all the non-randomized trials showed a shift in tumor stage and some an improvement in survival rates, the lack of a randomized control arm did not allow the researchers to test whether screening actually reduced mortality. The need for this evidence motivated the planning and conduction of a series of randomized clinical trials around the world (Table 1).

The first randomized trial to prove a mortality reduction benefit of LDCT-based compared to CXR-based screening was the National Lung Screening Trial (NLST) (Aberle et al. 2011), a study conducted in the USA that randomized over 53,000 former or current heavy tobacco smokers ages 55 to 75, to three annual rounds of either LDCT or CXR. The preliminary results of the trial, published after the first 9 years of follow-up, showed a 20% reduction of lung cancer mortality in favor of LDCT (relative risk (RR) = 0.80, 95% CI:[0.73,0.93] (Aberle et al. 2011). These results motivated expert societies in the USA such as the Preventive Service Task Force (USPSTF) and the National Comprehensive Cancer Network (NCCN) to issue formal recommendations in favor of LDCT-screening for early detection in high-risk individuals.

Though the results of the NLST increased the acceptance of LDCT screening, it was not clear whether the mortality reduction shown by the trial could be transferred to populations with other characteristics (Arenberg 2019). Also, the NLST produced very high false positive test rates, which could be largely attributed to the nodule evaluation criteria used during the trial. With the aim of validating the evidence in favor of LDCT as a screening method in other populations, and of testing alternative nodule evaluation and management criteria, further randomized trials in the USA, China and various European countries have been conducted (Table 1).

The largest European trial, with a total of 15,792 participants, was the Dutch–Belgian NELSON (Horeweg et al. 2014a; Ru Zhao et al. 2011; van Iersel et al. 2007; van Klaveren et al. 2009). After 10 years of follow-up since randomization, results from this trial showed an even stronger reduction in lung cancer mortality in favor of LDCT than that found after the first 9 years of follow-up in the NLST, that is: 24% among men (RR = 0.76, 95% CI:[0.61, 0.94]) and, from a smaller subgroup analysis, 33% among women (RR = 0.67, 95% CI:[0.38, 1.14]) (de Koning et al. 2020). Thanks to its sample size, the NELSON had enough power to produce statistically significant results for its main target population (male former and current smokers), thus being the first trial to confirm the benefits of LDCT-based screening in populations outside the USA.

Smaller randomized trials (Table 1) were conducted in Denmark (DLCST, N=4,104) (Wille et al. 2016), Italy (DANTE, N=2,532 (Infante et al. 2009); ITALUNG, N = 3,206 (Paci et al. 2017); and MILD, N = 4,099 (Pastorino et al. 2012)), Germany (LUSI N=4,052 (Becker et al. 2015;

Becker et al. 2012)), and the USA (LSS, N=3,318) (Doroudi et al. 2018; Gohagan et al. 2004; Gohagan et al. 2005). More recent ones are the UK Lung Cancer Screening Trial (UKLS, N=4,055) (Baldwin et al. 2011; Field et al. 2016) and the trial conducted by the AME Thoracic Surgery Collaborative Group in China (N=6,717) (Yang et al. 2018). A key difference between the USA trials (NLST and LSS) and those conducted in other countries is that instead of having a control arm without intervention, they compared LDCT to CXR-based screening.

By 2020, results from almost all smaller trials (except UKLS), as well as an update from the NLST after longer follow-up had been published. Initial results of the MILD (Pastorino et al. 2012), DANTE (Infante et al. 2015), DLCST (Wille et al. 2016) LSS (Croswell et al. 2009; Doroudi et al. 2018; Gohagan et al. 2004; Gohagan et al. 2005) showed no indication of a favorable effect of LDCT screening on lung cancer-related mortality. In contrast, the AME (Yang et al. 2018), LUSI (Becker et al. 2020) and ITALUNG trials (Paci et al. 2017), as well as further analyses on data from the MILD trial (Pastorino et al. 2019a) that corrected for incorrect randomization, reported point estimates of risk ratios for lung cancer mortality in favor of LDCT (Figure 9). However, none of these smaller trials were sufficiently powered to produce statistically significant results for the comparison of lung cancer mortality rates between the two arms. What is more, some of the trials have been evaluated as being at substantial risk of bias (Brodersen et al. 2020; Hoffman et al. 2020). In the AME as well as in DANTE, the sample sizes for the screening and control arms were not balanced and the MILD trial failed to randomize participants simultaneously; additionally, it reported a higher proportion of current smokers in the control versus the intervention group). Finally, after an extended follow-up of 12.3 years, further analyses on data from the NLST showed a lower reduction in lung-cancer mortality compared to the initial results (8%; RR = 0.92, 95% CI:[0.85, 1.00] and 11% dilution-adjusted; RR = 0.89, 95% CI:[0.80, 0.997]) (National Lung Screening Trial Research Team 2019).

Although the conclusions from individual studies seem partially conflicting, a recent meta-analysis (Hoffman et al. 2020) of the results of all such trials confirmed a reduction in lung cancer mortality attributed to LDCT-imaging as a screening method (RR=0.84, 95% CI:[0.75, 0.93]) (Figure 9).Differences in the effect of LDCT-screening between female and male participants have been reported by some randomized trials. Analysis on data from the NLST showed a stronger reduction in lung-cancer mortality among women (RR=0.73, 95 % CI:[0.6, 0.9], RR=0.86 after extended follow-up) than among men (RR=0.93 [95 % CI:[0.8, 1.08], RR = 0.97 after extended follow-up). Similar results were reported by the NELSON (RR=0.69 for women and 0.86 for men) and the LUSI.

Table 1. Characteristics of randomized control trials comparing LDCT with CXR or with a control arm without screening

Trial Country Recruitment Years	Trial size Intervention / control arm	Inclusion criteria		Screening interval in months / number of rounds	Evaluation criteria for nodules detected via LDCT		False Positive Rate [95% CI]
		Age	Smoking		Prevalence screen	Incidence screens	
LSS **U.S.** **2000-2001**	3,318 (1,660/1,658)	55-74	≥30 py, cessation <10 y	12 / 2	P: D > 3mm or abnormal appearance (spiculation, major atelectasis, etc.)	P: D ≥ 4mm or abnormal appearance (spiculation, major atelectasis, etc.)	0.16 [0.01, 0.01]
NLST* **U.S.** **2002–2004**	53,452 (26,722/26,730)	55-74	≥30 py, cessation <15 y	12 / 3	P: D ≥ 4mm		0.23 [0.23, 0.24]
DANTE **Italy** **2001-2006**	2,532 (1,300/1,232)	60-74	≥20 py, cessation <10 y	12 / 5	P: D ≥ 5mm	P: D > 5mm or any growth N: no new; no growth; NCN≤5mm	0.24 [0.22, 0.27]
NELSON **Netherlands/Belgium** **2003-2006**	15,792 (13,195/2,594)	50-74	>15 c/d, >25 y or >10 c/d, >30 y, cessation ≤10 y	12, 24, 30 / 4	P: V>500mm3; VDT < 400d I: V =50-500mm3, VDT 400-600d N: V <50mm3; VDT > 600d	P: new > 500mm3, VDT < 400d I: new 50-500mm3, VDT 400-600d N: none, new < 50mm3, VDT >600d	0.01 [0.01, 0.01]
DLCST **Denmark** **2004-2006**	4,104 (2,052/2,052)	50-70	≥20 py, cessation <10 y	12 / 5	P: D > 15mm I: D = 5-15mm N: D < 5mm	P: any new; VDT <400d I: VDT 400-600d N: no new; VDT > 600d	0.03 [0.03, 0.03]
ITALUNG **Italy** **2004-2006**	3,206 (1,613/1,593)	55-69	≥20 py, cessation <10 y	12 / 4	P: D ≥ 5mm N: D < 5mm	P: new > 3mm; any growth N: no new; no growth	0.19 [0.18, 0.20]

11

Table 1 (continued). Characteristics of randomized control trials comparing LDCT with CXR or with a control arm without screening

MILD **Italy** **2005-2011**	4099 (2376/1723)	49-75	≥20 py, cessation <10 y	12 or 24 / 3 or 5	P: V > 250mm3 I: V = 60–250mm3 N: V <60mm3	P: new V > 250mm3 I: new V 0–250mm3 N: no new	
LUSI **Germany** **2007-2011**	4,052 (2,029/2,023)	50–69	NELSON eligibility criteria	12 / 5	P: D >10mm or VDT < 400d I: D =5-19mm or VDT 400-600d N: D <5mm or VDT >600d	P: new > 10mm; VDT <400d I: VDT 400-600d N: no new; no growth; VDT >600d	0.08 [0.08, 0.09]
UKLS pilot **United Kingdom** **2011-2012**	4,055 (2,028/2027)	50-75	5-year lung cancer risk of ≥5%, based on the LLPv2 risk pre-diction model,	12 / 2	P: Solid: V >500mm3 Solid: Dmin >10mm; Part solid, solid component: V >500mm3 I: Solid: V 15-500mm3 Solid: Dmax 3-9.9mm Part solid, non-solid component: Dmean≥5mm. Part solid, solid component: V <500mm3 N: benign characteristics or D<3mm or V<15mm3	P: new solid: V>500mm3 or New solid, pleural based: Dmin >10mm New part solid, solid component: V >500mm3 or VDT <400 days or new solid component of non-solid nodule	0.036 [not reported] or 0.232 [not reported]
AME **China** **2013-2014**	6,717 (3,550/3,145)	45-70	≥20 py cessation <15 y	24 / 2	P: D ≥ 4mm		0.22 [0.20, 0.23]

* The NLST and the LSS compared LDCT vs CXR. The other trials had a control arm without screening (usual care or annual clinical assessment). In the DANTE trial, participants of the control arm were offered a baseline CXR and sputum cytology.

Abbreviations: LDCT: low-dose computed tomography; CXR: chest X-ray; AME: AME Thoracic Surgery Collaborative Group trial; UKLS: United Kingdom Lung Cancer Screening Trial ; DANTE: Detection and Screening of Early Lung Cancer by Novel Imaging Technology and Molecular Essays; DLCST: Danish Lung Cancer Screening Trial; ITALUNG: Italian Lung Cancer Screening Trial; LDCT: low-dose computed tomography; LSS: Lung Screening Study; LUSI: German Lung Cancer Screening Intervention Trial; NLST: National Lung Screening Trial; NR: not reported, y: years, py: pack-years, c/d: cigarettes per day; V: volume, D: diameter, Dmax: longest diameter, Dmean: average diameter, Dmin: shortest diameter; VDT: volume doubling time; LLPv2 = Liverpool Lung Project v2 model; P: positive; I: indeterminate; N: negative.
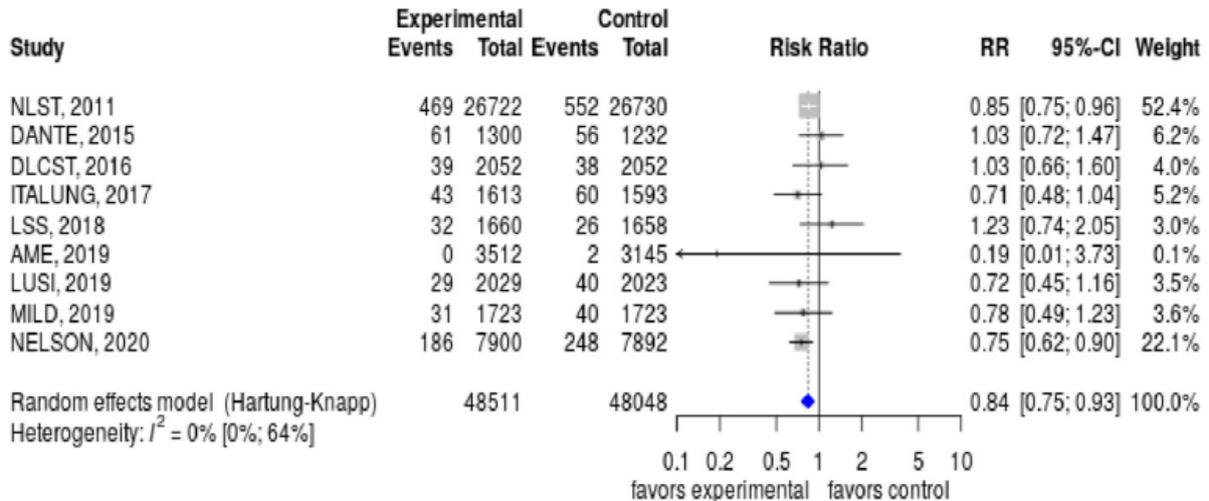
Figure 9. Forest plot of risk ratios for lung cancer mortality for all randomized screening trials until 2020

As published in (Hoffman et al. 2020), reprinted with permission

However, none of these trials showed a statistically significant interaction between treatment arm and sex. A hypothesis to explain these differences is the distribution of tumor histological types across sexes, (Becker et al. 2019; Pinsky et al. 2013), though this has not been proven.

1.2.3 Risks and potential harms of LDCT-based lung cancer screening

As mentioned in the previous section, the mortality reduction benefit provided by LDCT screening has been proven in populations in the USA and Europe. However, this benefit must be balanced against the various risks to which participants are exposed and against the related financial costs. Some of the main risks associated with LDCT screening are exposure to potentially harmful radiation, unnecessary and potentially invasive follow-up testing procedures or unnecessary treatment as a result of false positive screening test results, and overdiagnosis (Bach et al. 2012; Mazzone et al. 2018). These risks are discussed in more detail in the following sections.

1.2.3.1 Radiation exposure

Each LDCT scan exposes participants to potentially harmful radiation. Estimates of the radiation dose per LDCT scan have been reported at about 1.5 mSv with values in the range from 0.65 mSv to 2.36 mSv (see (Jonas et al. 2021) for review) a factor of 15 compared to that of CXR, but only 20-25% of that received as a result of full-dose chest CT (Oudkerk et al. 2021; Sharma et al. 2015).

Even though the radiation of single LDCT scans seems negligible, its effect is cumulative. Repeated exposure as a result of regular, follow-up or confirmatory appointments, the latter

usually by means of CT or PET with higher radiation doses (~ 8 and 14 mSv respectively), can therefore contribute to a long term increase in cancer risk (Sharma et al. 2015).

A study conducted on data from the NLST estimated that LDCT-screened participants were exposed to a mean effective dose (MED) of 8 mSv as a consequence of three annual screening rounds and all further imaging examinations, and predicted 4 radiation-induced cancer deaths for each 10,000 subjects (Bach et al. 2012). Analyses on data from the ITALUNG trial estimated a MED of 6.2 to 6.8 mSv after 4 annual LDCT rounds and further imaging examinations, with predictions of 1.2 to 1.3 and 3.1 to 3.3 radiation-induced cancers per 10,000 men and women respectively (Mascalchi et al. 2006). Based on data from the COSMOS trial (Rampinelli et al. 2017), researchers estimated a MED of 9.3 mSv for men and 13.0 mSv for women after 10 years of annual LDCT screening and additional examinations, and predicted 3 radiation-induced lung- and 5 major cancers per 10,000 screened subjects. According to these results, one radiation induced lung cancer would be expected for every 173 diagnosed lung cancers and one major radiation-induced cancer for every 108 diagnosed lung cancers.

Regarding long-term screening, a systematic review combining results from the COSMOS (Rampinelli et al. 2017) and the ITALUNG (Mascalchi et al. 2006) trials estimated a cumulative exposure of 20.8 mSv and 32.5 mSv after 25 years of screening (in line with the 2013 USPSTF recommendations to start at 55 and stop at 80 years). For the German population in particular, analyses from the Federal Office for Radiation Protection (Bundesamt für Strahlenschutz) (Nekolla E. Bundesamt für Strahlenschutz 2020) indicate that the lifelong risk of developing a lethal cancer attributed to radiation to be 0.07% among women and 0.03% among men, assuming annual screening between ages 50 to 54 and an effective radiation dose of 1.5 mSv per LDCT scan.

From these results it can be concluded that the cancer risk attributed to the radiation exposure of screening-eligible subjects (heavy current or former smokers, ages 50/55 to 80) is acceptable, given the mortality reduction benefit of LDCT-based screening in this high-risk population. There are however, some additional factors to consider, such as the effects of gender and age, with women and younger subjects being at higher risk of radiation-induced cancer and cancer-related death. Screening eligibility criteria are therefore important, since the balance between benefits and harms would not hold for low-risk subjects, for example for the young or the non-smokers, as shown by recent modeling studies (Berrington de Gonzalez et al. 2008).

Two strategies to further improve the balance between the risks of radiation exposure and the benefits of screening are to lower the effective dose per scan and to reduce the number of unnecessary examinations. The first one can be tackled through technological advances in the

field of imaging; the second through the improvement of evaluation criteria for LDCT scan results, and the optimal assignment of screening frequency and duration.

1.2.3.2 False positive findings

False positive (FP) test results are a major concern in LDCT screening since they lead to unnecessary additional radiation exposure from follow-up imaging procedures such as full-dose CT or positron emission tomography (PET) and in some cases to further, more invasive diagnostic verification procedures such as surgery or biopsy.

The high resolution provided by LDCT imaging allows for the detection of very small nodules, most of which turn out to be benign lesions. Analysis on data from the NLST indicate that 36% of participants received at least one FP result over three rounds of screening and that 96.4% of the nodules identified in the course of the study were benign lesions (Aberle et al. 2011). The high rates of FP results in the NLST can be attributed to the fact that the nodules were evaluated based only on their size (largest diameter) and with a very low threshold (4mm). Following these observations, criteria for nodule evaluation were adapted, culminating in more accurate systems such as the Lung CT Screening Reporting and Data System (Lung-RADS®) classification which combined nodule size (largest diameter and volume), with nodule growth based on short-term follow-up LDCT examinations, and nodule appearance as criteria for identifying potentially malignant lung nodules. Analyses showed (Pinsky et al. 2015) that had the NLST used the Lung-RADS® criteria, the FP rates would have decreased by up to 52% for the baseline and up to 76% for follow-up rounds. Other minor trials either used criteria similar to those recommended by the Lung-RADS® (DANTE, DLCST, ITALUNG and LUSI) or changed their definition of nodule size from longitudinal to volumetric (MILD, NELSON) (Table 1). Additionally, the LUSI and NELSON trials included a volumetric measure of nodule growth (volume doubling time, VDT) in their evaluation protocols (Table 1).

Further to simplified criteria, an approach that can potentially improve LDCT-based screening is the use of statistical models for nodule malignancy prediction. This alternative has the advantages of more accurately reflecting the continuous increase in risk associated with changes in the predictive variables, compared to simplified criteria, and of combining several subject and nodule characteristics based on LDCT scan images, and potentially other data sources, into a single measurement of risk. More details about the modeling approach, its advantages and areas for improvement will be discussed in a later section.

1.2.3.3 Overdiagnosis

Parts of this chapter have been published previously (González Maldonado et al. 2020b).

Overdiagnosis refers to the identification of tumors in cases in which they would not have become manifest in the absence of screening (Welch and Black 2010). It occurs when tumors

are detected within the so-called *lead-time,* the time by which diagnosis is brought forward thanks to early detection, and a period within which a proportion of patients with screen-detected lung cancer might have died of other causes before the disease would have become symptomatic.

Because a lung cancer diagnosis generally implies referral to aggressive treatment, overdiagnosis may cause serious and unnecessary losses in quality of life and unnecessary health-care costs (Welch and Black 2010).

Moreover, given that treatment changes the natural course of disease, overdiagnosis is a phenomenon that cannot be directly observed for individual patients. Therefore, the extent of overdiagnosis can only be approximated at a group level.

Using data from randomized trials, the extent of overdiagnosis has been most commonly estimated by obtaining the difference in cumulative incidence between the screening arm and the control arm. Depending on the trial, controls were subject to no screening (Becker et al. 2019; de Koning et al. 2020; Heleno et al. 2018; Paci et al. 2017) or to screening with CXR (Aberle et al. 2011; Patz et al. 2014). Using data from the NLST, the initially estimated excess incidence rate of LDCT relative to CXR after a median post-screening follow-up of 4.5 years was 18.5% (95% CI:[5.4%, 30.6%]) (Patz et al. 2014). Initial estimates from trials with a control arm without screening, and all at about 5 years post-screening follow-up, varied from zero excess in the ITALUNG (Paci et al. 2017) to 67.2% (95% CI:[37.1%, 95.4%]) in the DLCST (Heleno et al. 2018). In the Dutch-Belgian NELSON study, at 10 years post-randomization and 4.5 years post-screening follow-up since last screening participation, the estimated excess incidence among men was 19.7% of screen-detected cases (95% CI:[–5.2%, 41.6%]) (de Koning et al. 2020). In all these studies, however, follow-up times after screening cessation were likely too short to cover the longest possible tumor lead-times, a necessary condition for obtaining unbiased estimators of overdiagnosis.

In fact, excess incidence as an estimator of the magnitude of overdiagnosis depends strongly on the length of the post-screening follow-up. A good example of this association is the drop in estimated excess incidence calculated on data from the NLST after extended follow-up (3.1% at a median post-screening follow-up of 9.3 years). Another weakness of the excess incidence method is its sensitivity to deviations from perfect randomization between the screening arms. This might in part explain the large variability in estimates across trials, as reported in the previous paragraph. Furthermore, even if trials are well randomized and the post-screening follow-up period is long enough, excess incidence estimates are very specific to the data on which they are calculated and are therefore not transferable to other populations.

Besides the excess-incidence method, estimates of overdiagnosis have been generated via mathematical modeling, extending the screening scenarios of the original trials. Patz et al.

(Patz et al. 2014) calculated a lifetime excess incidence of 11% of non-small cell lung cancers (NSCLCs) after three annual screenings for a cohort of 60-year-old men and women in the USA. This result was based on the estimates of tumor mean pre-clinical sojourn time (MPST) and LDCT screening test sensitivity obtained by fitting convolution models to data from the NLST.

An alternative approach was used by Schultz et al. and ten Haaf et al. (Schultz et al. 2012; Ten Haaf and de Koning 2015) who, through microsimulation, estimated that overdiagnosis in the NLST trial cohort would decrease to 8.6% of all screen-detected cancers after life-time follow-up.

These model-based approaches have two main advantages compared to the excess-incidence method. First, they produce estimates of the magnitude of overdiagnosis that are not strongly dependent on the particular screening scenario in which data was collected, which makes such estimates more transferable. Second, by modeling the biological (MPST) and technological (detection sensitivity of LDCT) factors that influence the risk for overdiagnosis, they provide a better understanding of the phenomenon and therefore more transparency in the decision-making process.

1.3 Optimization of LDCT-based screening

Beyond the identification of an appropriate lung cancer detection method (a question that LDCT-imaging has already answered), designing the optimal screening program requires the consideration of multiple aspects in order to balance benefits, harms and costs. As with every optimization problem, this in turn requires the definition of clear and measurable objective functions. In the case of screening programs, there are two main criteria to maximize: net clinical benefit and cost-effectiveness.

Net benefit, also called net clinical benefit, is a decision analytic measure that allows for a direct comparison between benefits and harms of an intervention, and that depends on the definition of a "trade-off rate" (Vickers et al. 2016). In the case of lung cancer screening, the trade-off rate is most commonly defined in terms of mortality reduction or life years gained (LYG) and the risks and harms to which it exposes screening participants. Concrete examples are the number needed to screen, number of overdiagnosis, or the number of unnecessary invasive confirmatory procedures following false positive test results, per each lung cancer death avoided or per each LYG (de Koning et al. 2014; Meza et al. 2021). It follows then, that strategies which target the right candidates or that reduce the number of unnecessary invasive examinations can help in increasing the net clinical benefit of screening programs.

Cost-effectiveness, on the other hand, also defined in relative terms, refers to the balance between benefits and monetary costs. It is most commonly expressed as the incremental cost-

effectiveness ratio (ICER), with a denominator that measures the health gain of an intervention, such as LYG or alternatively quality life years gained (QALY), relative to the monetary costs associated with that health gain (Gold 1996). Simulation studies considering a wide variety of screening scenarios (Cressman et al. 2017) have been used for assessing how changes in screening strategies affect the ICER. Canadian and European studies (Hofer et al. 2018; Tomonaga et al. 2018) have shown how lung cancer screening can become cost-effective through the selection of appropriate eligibility criteria, screening frequency, algorithms for the evaluation of LDCT imaging findings, and by including tobacco cessation interventions (Black et al. 2014; Cressman et al. 2017; Mazzone et al. 2018; McMahon et al. 2011; Ten Haaf et al. 2017b). Given that CT and, to a lower degree, LDCT examinations are the main drivers of increases in costs, strategies to better target the right candidates for screening and to reduce unnecessary LDCT screening appointments and confirmatory CT examinations of indeterminate nodules can increase cost-effectiveness.

The following sections describe some strategies for the optimization of LDCT-based screening: 1) better targeting through the definition of appropriate eligibility criteria, 2) reduction of unnecessary examinations by assigning participants to optimal screening frequencies, 3) reduction of harms associated with unnecessary confirmatory procedures by appropriately evaluating LDCT findings, 4) how the accurate assessment of overdiagnosis risk can help reduce the potential harms of screening on an individual basis, and 4) how biomarkers can potentially be of help in all previous points.

1.3.1 Eligibility

Even though official recommendations have been issued in countries like the USA, there is no universal consensus regarding the optimal strategy for the selection of screening participants. Defining appropriate eligibility criteria to identify candidates with adequately long life expectancy and at sufficiently elevated short-term risk of developing lung cancer can help maximize the net clinical benefit and cost-effectiveness of screening programs.

In the coming sections two approaches for the definition of screening eligibility will be described: simplified eligibility criteria, and eligibility based on predicted lung cancer risk.

1.3.1.1 Simplified eligibility criteria

Expert organizations in the USA (United States Preventive Services Taskforce - USPSTF) (Moyer and USPSTF 2014) and Canada (Canadian Task Force on Preventive Health Care – CTFPHC) (Wood 2015) recommend screening eligibility based on lower and upper age limits and smoking history and/or smoking behavior (minimum lifetime cumulative smoking exposure and maximum time since quitting for ex-smokers). These recommendations are similar to the criteria used in the NLST trial (Patz et al. 2016) (55 to 75 years of age, >30 pack-years of

smoking, <15 years since smoking cessation), with the only difference being that the USPSTF allows participation until the age of 80. The relaxation of the upper age limit was motivated by the results of a modeling study in the USA (de Koning et al. 2014) that estimated a stronger reduction of cancer mortality relative to the number of screenings performed, compared to criteria with 75 years as upper-limit.

Most screening trials conducted after the NLST also applied age and smoking-based criteria, although with different age limits and different parameters related to smoking. In the Dutch-Belgian NELSON (Horeweg et al. 2014a; van Klaveren et al. 2009) and the German LUSI (Becker et al. 2012; Becker et al. 2019) trials, for example, age at entry and stopping were set at 50 and 70 years respectively for subjects with a smoking history of at least 10 cigarettes a day (cig/day) for at least 30 years (equivalent to 15 pack years) or at least 15 cig/day for 25 years (approximately 19 pack years). Criteria used by other minor European trials can be seen in Table 1.

The effects of various established or hypothetical selection criteria have been evaluated mainly through simulation studies (de Koning et al. 2014; Han et al. 2017; Ten Haaf et al. 2020; Ten Haaf et al. 2017b; Tomonaga et al. 2018; Treskova et al. 2017). In the German population, it has been estimated that the NLST and USPSTF criteria would select about 3 to 3.2 million individuals amongst which 40% to 45% of all yearly new lung cancer cases would occur. In contrast, the more inclusive criteria used in the NELSON trial would identify about 47% (11% to 17% more) incident cases, but at the cost of screening about 5.5 million (about 45% more) individuals (Hüsing and Kaaks 2020).

Simulations considering higher upper age limits (e.g., 80 instead of 75 years) predicted a higher proportion of avoided lung cancer deaths (de Koning et al. 2014). However, studies defining screening benefit as life years gained relative to the number of overdiagnosed cases, found an upper limit of 75 years to be more efficient than one of 80 years (Han et al. 2017).

As mentioned in section 1.2.3.1, in high-risk populations, the lung cancer mortality reduction of LDCT screening outweighs the long-term lung cancer risk attributed to radiation exposure. However, this does not hold for subjects with low short-term lung cancer risk and increased risk of radiation-induced cancer (for example, the young). Studies (Bach and Gould 2012; Hüsing and Kaaks 2020; Katki et al. 2016) have shown that criteria used in the NLST and NELSON trials, as well as those recommended by the USPSTF are prone to defining subjects from this group with low lung cancer risk as eligible, exposing them to potential harms though they do not benefit from screening. For the German population ages 50-54, a 5-year lung cancer risk threshold above which exposure to radiation would be justified has been estimated at around 0.5%, assuming 80% detection sensitivity and at least 20% lung-cancer mortality reduction for LDCT-based screening (Nekolla E. Bundesamt für Strahlenschutz 2020).

Additionally, more caution should be taken when assessing the eligibility of individuals with major comorbidities. For them, it is particularly important to weigh their predicted lung cancer risks against the harms of confirmatory or diagnostic invasive procedures, given that such procedures represent higher risks in this population compared to otherwise healthy individuals (Schneider and Arenberg 2015).

In terms of false-positive test results, a study showed (Pinsky et al. 2018) that amongst NLST-eligible subjects, the ratio of lung cancer diagnoses to invasive diagnostic workup caused by false positive tests went from 1.35 for the 10% of subjects with the lowest 5-year lung cancer risk, to about 5.25 in the upper 10%.

From the previous evidence it can be concluded that the benefit of screening increases together with lung cancer risk and that having accurate estimates of individual risk might help select the right candidates and in doing so, increase screening efficiency.

In summary, although simplified eligibility criteria have advantages such as their ease of use and their interpretability, they present some drawbacks. Due to their simplicity, they are not accurate enough to predict increases in risk related to changes in continuous risk factors such as age and time since quitting. Furthermore, they are prone to missing high-risk individuals, as well as to including individuals at low-risk for lung cancer or with major comorbidities, who are unlikely to benefit from screening.

1.3.1.2 Eligibility based on predicted lung cancer risk

An alternative to simplified criteria is the definition of screening eligibility in terms of absolute pre-test risk for lung cancer or lung cancer death as predicted by statistical models. Using this selection method, screening candidates are those individuals with estimated risks high enough to outweigh the associated harms (Hüsing and Kaaks 2020; Katki et al. 2018; Tammemägi et al. 2013; Ten Haaf et al. 2017a). Evidence in favor of this approach comes from retrospective analyses on data from the NLST trial that reported only 1% of all averted lung cancer deaths amongst 20% of participants with the lowest lung cancer risks, but 88% amongst 60% of participants at highest risks (Black et al. 2014; Kovalchik et al. 2013).

So far, around 22 statistical models for the prediction of short-term (1-6 years) risk for lung cancer and/or lung cancer death had been published (see (Jonas et al. 2021; Kauczor et al. 2020; Veronesi et al. 2020) for reviews of these models). Though they vary in complexity and in the weighting of predictors, all models are based on subsets or combinations of subject characteristics (age, BMI and socio-economic indicators), current or past smoking behavior, family or personal health status or history (cancer and/or lung diseases such as COPD or emphysema) and exposure to carcinogens.

External validations of some of these models have been conducted in cohorts from the USA (Katki et al. 2016; Kovalchik et al. 2013; Landy et al. 2019; Tammemägi et al. 2013; Ten Haaf et al. 2017a), Germany (Hüsing and Kaaks 2020; Li et al. 2015) and Australia (Weber et al. 2017). Across studies, the Bach (Bach et al. 2003a), LCRAT (Katki et al. 2016), and PLCOm2012 full and simplified models (Tammemägi et al. 2013) have shown good discrimination and fair calibration (see (Hüsing and Kaaks 2020; Jonas et al. 2021) for reviews). In terms of screening benefits, evaluations of the LCRAT and PLCOm2012 models reported higher numbers of screen-preventable deaths compared to the NLST criteria (see (Jonas et al. 2021) for review). In some studies, the PLCOm2012 showed also higher cost effectiveness (Katki et al. 2016; Kovalchik et al. 2013; Tammemägi et al. 2014; Tammemägi et al. 2013).

In general, risk-based eligibility criteria have been shown to improve screening efficiency and cost-effectiveness by identifying 10% to 20% more lung cancer cases amongst equal numbers of screening candidates compared to simplified criteria (Hüsing and Kaaks 2020; Katki et al. 2018; Li et al. 2015; Tammemägi et al. 2013; Ten Haaf et al. 2017a) and by lowering the number needed to screen (NNS) per averted lung cancer death (see (Jonas et al. 2021) for review). Similar results come from analyses on survey, cohort and lung cancer incidence data from Germany (Hüsing and Kaaks 2020; Li et al. 2015) which found the PLCOm2012 model to select a higher number of future lung cancer cases among the upper quantiles of risk compared to the USPSTF and NLST criteria.

The main advantage of models compared to criteria, is that they can predict continuous changes in risk associated with changes in risk factors also defined in a continuous scale, such as age, duration of lifetime smoking and for ex-smokers the time since smoking cessation, which makes risk estimates more accurate. Also, besides providing individual risk estimates, prediction models are able to make better use of data compared to simplified criteria, since they can include a larger number of variables, assign them different weights and model their potential interactions. Additionally, they open the possibility of including predictors that go beyond demographic and smoking behavior indicators, such as biomarkers (Guida et al. 2018; Hanash et al. 2018).

However, there are some concerns regarding a model-based approach. The major one relates to the risk of overdiagnosis. Compared to simplified criteria, risk-based eligibility tends to select individuals with a longer history of smoking, who are more likely to die from competing causes (Hüsing and Kaaks 2020; Katki et al. 2018; Li et al. 2015; Tammemägi 2018; Tammemägi et al. 2013; Ten Haaf et al. 2017a) and therefore receive only moderate benefit in terms of life years gained (LYG) and quality-adjusted life years (QALY) (Kumar et al. 2018; Ten Haaf et al. 2019). An additional concern is the lack of a concrete risk threshold above which to recommend screening and the need for it to be revisited over time and across populations. Regarding

absolute risks, the use of predictors difficult to quantify, standardize or based on self-reported variables (e.g., exposure to carcinogens and history or presence of emphysema or COPD) might introduce additional error and hinder model comparability and transferability. Finally, even if prediction models would prove to be superior to simplified eligibility criteria, clinical practitioners might be discouraged by complex models difficult to use in the absence of appropriate information technology (IT) tools.

An informed decision on whether eligibility should be based on thresholds of lung cancer risk predicted by statistical models requires more evidence from screening trials. So far, only one such trial, the UKLS, has used a prediction model to evaluate eligibility based on individual estimated risks. Results from the pilot study reported 2.1% lung cancer cases identified at the first screening round, a higher percentage than that seen in the NLST, by setting a threshold of 5% 5-year risk as estimated by the LLPv2 model (Field et al. 2016).

1.3.2 Screening frequency

Parts of this chapter have been published previously (González Maldonado et al. 2021a).

Currently, most trials and screening programs recommend annual screening rounds for all eligible subjects. Identifying subjects who would get an equivalent or even higher net benefit from less frequent screening could help minimize costs, cumulative radiation exposure and the risk for false positive findings, in exchange for a null to minimal reduction in health gains.

Analyses on data from the NELSON and MILD trials showed that screening at intervals longer than a year did not decrease the mortality reduction benefit of LDCT screening (de Koning et al. 2018; Pastorino et al. 2019b). However, decreasing screening frequency comes at the cost of higher numbers of interval cancers and delayed detections (Yousaf-Khan et al. 2017a). Therefore, it is important to weigh in the lung cancer risk of participants into the decision-making process. Based on these observations, it has been hypothesized that individuals with comparatively low short-term risk for lung cancer (e.g., within one to two years following an LDCT examination) could still benefit from early detection, by attending subsequent screening appointments at intervals longer than a year.

Further to deciding on the appropriate frequency based on an individual's lung cancer risk as estimated by models such as LCRAT or PLCOm2012, research groups have suggested that these estimates of pre-screening risk could be updated by findings from LDCT scan images from the baseline and/or further screening rounds in order to increase the accuracy of the assignment to longer screening intervals (Maisonneuve et al. 2011; Robbins et al. 2019; Silva et al. 2019; Tammemägi et al. 2019b; Yousaf-Khan et al. 2017b). Initial evidence in favor of this approach comes from a study by Patz et al. (Patz et al. 2016) which showed that, on data from the NLST, the average risk for lung cancer detection at the first annual follow-up screen

was 0.35% for screening participants with no pulmonary nodules of at least 4 mm in largest diameter at their initial screen (N=19,066, 73%); and 1.02% among all screening participants (N=26,231). Similar results were found in the NELSON trial (Yousaf-Khan et al. 2017b).

Along these same lines, statistical models have been developed based on subject-characteristics combined with the presence and characteristics of pulmonary nodules (Schreuder et al. 2018) or other radiologic indicators of pulmonary health (emphysema, consolidation) (Robbins et al. 2019; Schreuder et al. 2018), as well as general lung cancer risk factors such as smoking behavior.

Schreuder et al. (Schreuder et al. 2018) developed a polynomial model with linear and 2nd-degree terms for a total of 11 selected risk factors, including age, sex, smoking history, personal and family history of cancer, and LDCT scan findings at the initial prevalence screen such as pulmonary nodules and emphysema (Polynomial model).

A different model was developed for use among individuals with a negative LDCT screen (no nodules ≥4mm) by Robbins et al. (Robbins et al. 2019). It extends a pre-existing lung cancer risk prediction model (Lung Cancer Risk Assessment Tool [LCRAT]) (Katki et al. 2016) based on age, smoking history, family history of lung cancer, BMI and education level, by adding LDCT data on pulmonary emphysema and consolidation (LCRAT+CT).

Compared to LDCT imaging data only, these models considerably improved discrimination of screening participants by their likelihood of receiving a lung cancer diagnosis either at, or in the year following the next screening appointment. Based on the Polynomial and LCRAT+CT models it was further estimated that, in the NLST, up to about 45% of annual screenings in the second round, and 58% of all annual follow-up (incidence) screenings could have been skipped at the cost of delayed diagnosis for a comparatively small proportion of 10% to 24% of screen-detected cancers (Robbins et al. 2019; Schreuder et al. 2018).

While promising, both models (LCRAT+CT and Polynomial) were developed and tested exclusively on the basis of NLST data and, so far, have not been externally validated on independent screening data.

Finally, just as in the case of risk-based eligibility criteria, the inclusion of biomarkers in prediction models for short-term lung cancer risk may contribute to optimize the personalized assignment of screening intervals. This approach, however, has not yet been sufficiently investigated.

1.3.3 Assessment of LDCT results

Parts of this chapter have been published previously (González Maldonado et al. 2020a).

1.3.3.1 LDCT-based criteria for lung cancer prediction

Reducing the number of false positive (FP) test results in order to minimize unnecessary radiation exposure and invasive diagnostic procedures while preserving the mortality reduction benefit is one of the main goals when designing a screening program.

Analyses have shown that improvements in evaluation criteria, such as the use of higher thresholds for nodule size (longitudinal and volumetric) and the incorporation of information on nodule growth, among other nodule criteria, can drastically decrease FP rates without hampering the mortality reduction (Pinsky et al. 2015).

Based on these findings, radiologic and oncologic societies (American College of Radiology Committee on Lung-RADS® 2014; Baldwin and Callister 2015; Kanne et al. 2013; MacMahon et al. 2017; Wood et al. 2015) have issued recommendations and guidelines to improve the assessment of nodule malignancy based on features extracted from their LDCT-scan images.

More recent studies have investigated the use of statistical models for the prediction of nodule malignancy as an alternative to established simplified criteria. The following section describes this approach.

1.3.3.2 LDCT-based models for nodule malignancy prediction

Various statistical models have been developed that estimate the probability of malignancy for nodules seen in LDCT-scan images. Predictor variables in these models are based on radiologic features extracted from imaging findings of the lung and on subject characteristics such as age, sex, previous and/or current smoking behavior, presence and/or history of lung diseases, and personal and/or family history of cancer.

Early modeling studies were based on clinical data, e.g. at the Mayo (Swensen et al. 1997) and Veterans Affairs (VA) clinics (Gould et al. 2007) and Peking University People's Hospital (PKUPH)(Li et al. 2011). However, given that in clinical practice subjects are usually already symptomatic and therefore more likely to have cancer at advanced developmental stages, those models were trained on images of larger and mostly solid nodules, which makes them prone to overestimate the risk of smaller nodules, such as those detected during screening (Nair et al. 2018).

The first malignancy prediction models trained on data from screening contexts were developed at Brock University (Toronto, Canada). Data from participants of the Pan-Canadian Early Detection of Lung Cancer Study (PanCan) (McWilliams et al. 2013) and the LDCT images from their first screening appointment were used to fit four models (PanCan models)

(McWilliams et al. 2013) using either parsimonious (PanCan-1) or comprehensive variable selection strategies including (PanCan-1b, PanCan-2b) or excluding (PanCan-1a, PanCan-2a) nodule spiculation as a predictor. The PanCan-2b model, the most complex one, was recently used as basis for further versions using multi-dimensional measurements of nodule size (mean diameter or volume) instead of largest nodule diameter (Tammemägi et al. 2019a). A more recent model was trained independently on data from participants of the UKLS trial, including volume as a measure of nodule size (Marcus et al. 2019).

External validations of the PanCan models have been carried out on data from the NLST (Tammemägi et al. 2019a; White et al. 2017) and of the DLCST (Winkler Wille et al. 2015) in terms of discrimination but not of calibration. So far, the UKLS model (Marcus et al. 2019) has not been externally validated.

Further steps towards the improvement of prediction models for nodule malignancy include the use of pattern recognition and machine learning methods applied directly to the LDCT scan images, instead of relying on models based on pre-extracted nodule features (e.g., nodule size or nodule location).

1.3.4 Overdiagnosis risk and stop-screening criteria

Defining who should enter a screening program is not enough to guarantee an optimal net clinical benefit for participants. Additionally, deciding if and when screening participants should stop being examined is also of great relevance.

Participants who were once eligible might no longer benefit from screening if the quality life years they can gain are not significant compared to the harms to which they are exposed; that is, if they are at increased risk of overdiagnosis.

Besides net clinical benefit, seen from the cost-effectiveness perspective, overdiagnosis is also a factor to consider when aiming at designing a program for screening. Overdiagnoses that lead to overtreatment, represent unnecessary financial costs. Reducing the risk of overdiagnosis could also translate in a higher benefit to cost ratio.

As mentioned in section 1.3.1.1, studies that defined the trade-off between benefits and harms in terms of LYG or QALY relative to the number of overdiagnosed cases, found an upper age limit of 75 to be more efficient than one of 80 years,, since the latter includes participants with limited remaining life expectancies (RLE) (Han et al. 2017).

However, fixed upper age-limits that apply for all screening participants might not guarantee a positive net benefit at an individual level. Instead, accurate personalized estimates of overdiagnosis risk tied to indicators of comorbidities, frailty and RLE of participants, preferably

monitored over time, could be a better strategy for the optimization of lung cancer screening programs.

1.3.5 Potential of biomarkers as a complementary tool for screening

Parts of this chapter have been published previously (González Maldonado et al. 2021b).

Aiming at a better balance between benefits and harms, an alternative line of research focuses on identifying less invasive, radiation-free and cost-effective tools for lung cancer detection or risk-prediction, that can complement imaging-based screening. Among the most promising candidates are molecular biomarkers (Califf 2018) (in the following referred to as *biomarkers*) measured in blood, sputum, and exhaled breath condensates (Chu et al. 2018; Hasan et al. 2014; Hulbert et al. 2017; Lam et al. 2011; Liang et al. 2018; Lopez-Sanchez et al. 2017; Massion et al. 2017; Rodríguez et al. 2021).

Biomarkers for detection or risk prediction tackle two different problems (Califf 2018). In the latter case, as the name indicates, the goal is to accurately predict the risk for cancer-free subjects to develop the disease within a certain period of time. An example of a biomarker which has shown good performance at predicting short-term lung cancer risk consists of a panel of four circulating proteins and one protein precursor measured in blood samples (Guida et al. 2018). This panel, combined with information on smoking, was able to discriminate subjects developing lung cancer within one year of sample collection from those who remained cancer-free (Guida et al. 2018). In the context of screening, risk prediction markers could help identify individuals whose risk of developing lung cancer is high enough to justify their eligibility. In contrast, detection markers aim at identifying individuals who already have lung cancer, ideally at early, still curable stages, with high sensitivity and specificity. This type of markers could be used, for example, as a filter for the referral to invasive confirmatory procedures in the presence of suspicious or indeterminate nodules detected via LDCT-screening. Markers that have shown good detection accuracy are miRNA signatures such as the miR-Test (Bianchi et al. 2011; Montani et al. 2015) and the miRNA signature classifier (MSC) (Boeri et al. 2011; Sozzi et al. 2014) , as well as autoantibodies to tumor-related antigens (TAAbs) (see (Broodman et al. 2017; Du et al. 2018; Qin et al. 2018; Rodríguez et al. 2021) for reviews).

Most of the biomarker candidates so far identified, have been evaluated only in clinical settings in which cancer-free subjects are compared with patients harboring clinically manifest tumors (Seijo et al. 2019). An exception are TAAbs, which have been widely tested in clinical and screening settings (Broodman et al. 2017; Chapman et al. 2008; Du et al. 2018), and which were recently evaluated in a large randomized screening trial (Sullivan and Schembri 2019; Sullivan et al. 2017; Sullivan et al. 2021).

Higher levels of TAAbs, produced as an immune response to tumor antigens, have been observed in cancer patients compared to cancer-free controls (Kaaks et al. 2018; Macdonald et al. 2017; Tang et al. 2017). Furthermore, it has been hypothesized that TAAbs are produced at very early stages of tumor development. If this hypothesis holds true, TAAbs could be suitable markers for early detection (Anderson and LaBaer 2005).

The ability of individual TAAbs and multi-TAAb panels to discriminate lung cancer patients from cancer-free individuals has been compared in a series of studies (Broodman et al. 2017; Chapman et al. 2008; Du et al. 2018) with panels showing a better performance compared to individual ones. However, while most of the panels showed good specificity, their detection sensitivity was only modest (Broodman et al. 2017; Qin et al. 2018; Yang et al. 2019).

One such panel, measured by the commercially available EarlyCDT®-Lung test (Oncimmune Ltd, Nottingham, United Kingdom), is comprised of 7 TAABs (CAGE, GBU4-5, HuD, MAGEA4, NY-ESO-1, p53 and SOX2). EarlyCDT®-Lung has been proposed as a "rule-in" diagnostic test for lung cancer, with about 90% specificity at 40% sensitivity in clinical studies (Chapman et al. 2012; Healey et al. 2017; Jett et al. 2014; Massion et al. 2017). In addition, it has been evaluated as a confirmatory test in clinical settings, with the purpose of deciding on further, more invasive interventions for subjects with incidentally observed pulmonary nodules (Healey et al. 2017; Jett et al. 2014; Massion et al. 2017). Finally, in the context of screening, EarlyCDT®-Lung has been recently tested in the Early Cancer detection test Lung cancer Scotland (ECLS) (Sullivan and Schembri 2019; Sullivan et al. 2017; Sullivan et al. 2021). In the ECLS the panel was tested for its ability to identify high-risk subjects who could benefit from LDCT-based screening.

In summary, although a lot of research has been conducted and a number of promising biomarkers have been identified, there is still not enough evidence in favor of the added benefit of including biomarkers in screening contexts. Despite the good performance at detection and risk-prediction shown by some of the candidates, none have been included in lung cancer screening protocols.

What is more, even if biomarkers would prove highly accurate at detecting the presence of lung cancer, imaging methods would still be required for the localization and treatment of tumors. Therefore, even though biomarkers have the potential to complement screening, LDCT-based imaging is likely to remain as the gold standard for lung cancer early detection in the near future.

1.4 Aims and objectives

The work presented in this thesis analyzes the following key aspects of lung cancer screening, using data and blood samples collected as part of the Lung Cancer Screening Intervention Trial (LUSI):

1. *Personalized assignment of screening intervals for screening participants.* Studies have shown that screening eligibility criteria based on lung cancer risk, as predicted by models such as LCRAT (Katki et al. 2016) or PLCOM2012, could prevent more lung cancer related deaths and lead to more life years gained at equivalent numbers of screened subjects, compared to the use of simplified criteria (Katki et al. 2018; Tammemägi et al. 2013; Ten Haaf et al. 2017a). However, lung cancer risk is quite heterogeneous even among screening-eligible subjects. Thus, further to selecting candidates at risks high enough to justify at least low-frequency (e.g., biennial) screening, risk stratification could be used to assign them to varying screening intervals that optimize net clinical benefits and/or financial costs. For this purpose, models for the prediction of short-term lung cancer risk have been developed that combine subject-specific risk factors with LDCT imaging findings. Two such models are LCRAT+CT (an extension of the LCRAT) (Robbins et al. 2019) and the Polynomial model (Schreuder et al. 2018). While promising, these models have not been externally validated. The first aim of this study is to evaluate the LCRAT+CT and Polynomial models in terms of discrimination when assigning participants to annual vs. biennial screening, and to assess the calibration of their predicted risks.

2. *Application of risk prediction models to improve malignancy prediction of nodules detected by LDCT screening.* The efficiency of LDCT-based screening depends in part on the selection of nodule evaluation criteria. For this purpose, several research groups have developed statistical models for malignancy prediction of screen-detected nodules, based on radiological features and subject characteristics. The second aim of this study is to evaluate four of a total of six models originally trained on data from the Pan-Canadian Early Detection of Lung Cancer Study (PanCan) (McWilliams et al. 2013; Tammemägi et al. 2019a), as well as a recent model the UKLS model (all of them originally developed on data from screening trials) in terms of discrimination, calibration, and operational performance (e.g. sensitivity, specificity, and positive predictive value). Additionally, this study also presents findings for a selection of models originally trained on data from clinical contexts (Mayo Clinic (Swensen et al. 1997) USA Department of Veterans Affairs (VA) clinics (Gould et al. 2007), or Peking University People's Hospital (PKUPH) (Li et al. 2011)).

3. *Risk of overdiagnosis and its relationship with age limits for LDCT screening participation.* Recent analyses (Han et al. 2017; Ten Haaf et al. 2020) suggest that the balance between life years gained and the risk of overdiagnosis becomes less favorable when screening is offered to individuals with short remaining life expectancy. Thus, accurate estimation of overdiagnosis risk is relevant for the defining optimal eligibility criteria and for deciding when screening should stop. In particular, choosing appropriate upper age limits may help improve the balance between the benefits of early detection and the harms of irrelevant diagnoses and unnecessary treatment. The third aim of this study is to estimate the extent of overdiagnosis in the LUSI trial and to investigate its relationship with upper age limits for eligibility, based on estimates of the mean sojourn time of lung cancer tumors and of the sensitivity of LDCT screening.

4. *The use of biomarkers as a complementary method for the selection of candidates for LDCT-based screening.* Tumor-associated autoantibodies (TAAb) are considered promising markers for early detection of lung cancer. So far, however, their detection ability has been tested only in clinical contexts, in which tumors are often detected at advanced stages, and it has not yet been investigated whether blood TAAb concentrations are elevated in patients with small malignant nodules (< 10 mm in diameter), and whether antibody tests such as EarlyCDT®-Lung can detect tumors in equally early stages as LDCT-based screening. The fourth aim of this study is to address these questions, by evaluating the early detection accuracy of EarlyCDT®-Lung, in terms of sensitivity and specificity.

Finally, in the discussion, findings made in this study are put in perspective in terms of how to inform policy-making for the implementation of lung cancer screening programs in Germany and the rest of Europe.

2 MATERIALS AND METHODS

2.1 The Lung Cancer Screening Intervention Trial (LUSI)

Parts of this chapter have been published previously (Becker et al. 2015; Becker et al. 2012; González Maldonado et al. 2020a; González Maldonado et al. 2021a; González Maldonado et al. 2021b; González Maldonado et al. 2020b)

2.1.1 Trial design and eligibility criteria

The German Lung Cancer Screening Intervention Trial (LUSI) (Becker et al. 2012; Becker et al. 2019) is a randomized trial (ISRCTN30604390 2007) approved by the Medical Ethics Committee of the University of Heidelberg (073/2001), and by the German Federal Office for Radiation Protection (BfS). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). All participants enrolled provided written informed consent.

Details about trial design and mortality-reduction results have been published previously (Becker et al. 2015; Becker et al. 2012; Becker et al. 2019). In brief: the recruitment phase of the LUSI trial started in October 23, 2007 and stopped in April 11, 2011. Eligible subjects were those 50–69 years of age with a history of heavy smoking (≥25 years of smoking of ≥15 cigarettes per day, or ≥30 years smoking of ≥10 cigarettes per day; ≤10 years since smoking cessation). A total of 4,052 men and women, whose records were obtained as a random sample from population registries in Heidelberg (Germany) and surroundings, were recruited for the study. Participants were randomly assigned to either the screening intervention arm (N=2,029), comprised of a LDCT screening at time of randomization plus four annual follow-up screenings, or a control arm (N=2,023) with no intervention. Active screening was conducted between October 2007 and May 2016.

2.1.2 Image acquisition and reading

LDCT scans were performed by trained staff at the Radiology Department of the German Cancer Research Center (DKFZ) Heidelberg using either a Toshiba 16 row scanner (2007-Dec 31st 2009), or a Siemens 128 row scanner (March 18th 2010 onwards) with a maximum of 1.6–2 mSv radiation exposure(Becker et al. 2012).

LDCT images were read on a server for computer-aided detection (CAD). Nodule outlines generated via automatic segmentation (MEDIAN software, France) were manually corrected by the operator to include all parts of the nodule and exclude any adjacent structures. Perifissural nodules with oval or triangular shape and/or smooth delineations were excluded and identified as lymphatic nodules. Other characteristics collected from the nodules were: identifier, location, type (solid, subsolid), shape and border (spiculated, "clear", other) and

presence of calcification. Nodule size measurements (longest and transverse diameters, and volume) were automatically derived from the 3D segmented nodule. The lower limit for nodule detection was 1mm in diameter (Becker et al. 2012).

The presence of emphysema was determined at the Department of Diagnostic and Interventional Radiology with Nuclear Medicine of the Thoraxklinik at the University of Heidelberg via densitometry based on screening LDCT imaging findings, as performed by the YACTA software (Jobst et al. 2018; Weinheimer et al. 2011; Wielputz et al. 2014; Wielputz et al. 2013). Individual lung voxels with densities ≤ -950 Hounsfield Units were assigned to emphysema. Additional software-computed parameters were: total lung volume in inspiration (LV), emphysema volume (EV), emphysema index (EI), mean lung density (MLD) and the 15th percentile of lung density (15TH) (González Maldonado et al. 2020a; Jobst et al. 2018).

Further details about image acquisition and reading have been previously published (Becker et al. 2012; González Maldonado et al. 2020a; Jobst et al. 2018).

2.1.3 Nodule evaluation and management protocol

Nodules detected in participants of the screening arm were evaluated by two trained chest radiologists. The further management of nodules was decided based on their size and, for those previously detected, based on their growth (Table 2), applying criteria similar to those of Henschke et al. (Henschke et al. 1999) and to the Lung CT Screening Reporting and Data System (Lung-RADS®) assessment guidelines (American College of Radiology Committee on Lung-RADS® 2014; American College of Radiology Committee on Lung-RADS® 2019). According to such evaluations, screening participants were: returned to regular annual screening, invited for follow-up LDCT at 3- or 6-month intervals or recommended immediate diagnostic work-up. Immediate diagnostic work-up was carried out by a cooperating pulmonologist, who then decided about further confirmatory procedures or treatment (X-ray, full-dose CT, PET, bronchoscopy, video-assisted thoracoscopic surgery (VATS), biopsy, antibiotic treatment or short-term follow-up).

Malignant nodules were identified by communicating the location of the suspicious nodules to the thoracic surgeon previous to the resection surgery (via VATS under general anesthesia).

Further details about evaluation of LDCT findings, management of screen-detected nodules and additional diagnostic work-up have been published previously (Becker et al. 2015; Becker et al. 2012; Becker et al. 2019).

Table 2. LDCT evaluation algorithm applied in the German randomized lung cancer screening trial LUSI.

Newly observed nodules (First screening round or new in subsequent rounds)		Known nodules (Early recalls or subsequent screening rounds)	
Outcome by nodule size	Action	Outcome by nodule growth	Action
without abnormality or nodules < 5mm	back to routine screening (12 months)	-	-
nodules ≥5 and <8 mm	early recall (6 months)	> 600 VDT	back to routine screening
		400 – 600 VDT D < 7.5 mm	early recall (6 months)
nodules ≥8 and ≤10 mm	early recall (3 months)	D ≥ 7.5 mm – 10 mm	early recall (3 months)
nodules > 10 mm / not highly suspicious		≤ 400 VDT or D > 10 mm	immediate recall
highly suspicious	immediate recall	malignant	treatment
		non-malignant	back to routine screening

Abbreviations: VDT: Volume doubling time; D: diameter

As published in (Becker et al. 2015; González Maldonado et al. 2020a; González Maldonado et al. 2021a; González Maldonado et al. 2021b), reprinted with permission

2.1.4 Evaluation of lung function via spirometry

Spirometry was performed in all participants of the LDCT arm at their baseline examination using MasterScreen IOS (VIASYS Healthcare) to determine 1-second forced expiratory volume (FEV_1) and forced vital capacity (FVC). The ratio FEV_1/FVC was calculated from the largest FEV_1 and FVC values recorded in any one of two repeated assessments.

Further details about lung function evaluation via spirometry in the LUSI trial have been published previously (González Maldonado et al. 2020a; Jobst et al. 2018).

2.1.5 Blood sample collection protocol

Blood samples were collected by trained staff of the LUSI study center at the German Cancer Research Center (DKFZ) Heidelberg. A total of 1,576 participants of the screening arm provided a baseline blood sample either at the time of their first screening participation (N=1,362, round 1 (T_0)) or, due to technical issues, at the time they came back for subsequent screening rounds (206 in round 2 (T_1); 8 in rounds 3 (T_2) to 5 (T_4)). Such samples were taken at the time of non-suspicious findings. Further samples were collected from individuals presenting with suspicious LDCT scan results at any screening round, that is, all study participants referred to the local hospital clinic for immediate confirmatory diagnostic work-up

(N=111), and those returning for 3- (N=132) or 6-month (N=408) follow-up LDCT scans. All blood samples were processed within 2 hours of the blood draw. Serum was allowed to clot for 30 minutes, followed by centrifugation and aliquoting, before long term storage at -80°C.

Further details about the blood sample collection protocol of the LUSI trial have been published previously (Becker et al. 2012; González Maldonado et al. 2021b).

2.1.6 End-point definition

Screen-detected cases were defined as those detected (having suspicious screens that eventually led to diagnosis work-up with which diagnosis confirmation was made) at any of the 5 screening rounds of the study, independently of the time between detection and diagnosis. Incident cases were defined as those diagnosed either between two screens (interval cases) or in the years following the last screen (Becker et al. 2015; Becker et al. 2019; González Maldonado et al. 2020b).

2.1.7 Prospective case ascertainment

The prospective incidence of lung cancer in both study arms was ascertained by a combination of annual follow-up questionnaires (self-reports) and linkage to cancer and mortality registries. Detailed medical record information (pathology reports, medical letters on diagnosis, treatment and radiology reports) was obtained for all lung cancer cases by contacting the treating clinics (Becker et al. 2012; Becker et al. 2019; González Maldonado et al. 2020b).

2.1.8 Tumor histology coding

Tumor histology was coded according to ICD-O-3 version 2003 or 2013 depending of the date of diagnosis. Morphology codes were classified as follows: small-cell lung cancers (ICD-O-3: 8041/3, 8042/3, 8045/3, 8044/3); non–small-cell lung cancers, subdivided into squamous cell carcinomas (8070/3, 8072/3, 8071/3, 8083/3, 8076/3, 8078/3), adenocarcinomas (8140/3, 8255/3, 8250/3, 8480/3, 8550/3, 8260/3, 8310/3, 8490/3, 8046/3), bronchiolo-alveolar adenocarcinomas (8250/3, 8253/3) large cell carcinomas (8013/3), carcinoids (8240/3, 8246/3, 8249/3), unspecified carcinomas (8010/3) and malignant neoplasms (8000/3) (González Maldonado et al. 2021b; González Maldonado et al. 2020b).

2.2 Participant selection and data collection

2.2.1 For the validation of prediction models for the assignment of screening intervals

Parts of this chapter have been published previously (González Maldonado et al. 2021a).

The LCRAT+CT and Polynomial models and, for comparison, the criterion by Patz et al. were validated on data from participants of the screening arm of the LUSI trial who fulfilled the corresponding eligibility criteria (Patz et al. 2016; Robbins et al. 2019; Schreuder et al. 2018).

Participants eligible for the validation of the LCRAT+CT model (Robbins et al. 2019) were those with at least one negative LDCT scan as defined by the NLST criteria (no nodules ≥4mm in longest diameter) and who were at risk for lung cancer detection at the next screening appointment (N=1,194 at time point T_0, and 1,220, 1,262 and 1,228 at the three following incidence screens, at time points T_1-T_3) (Figure 10).
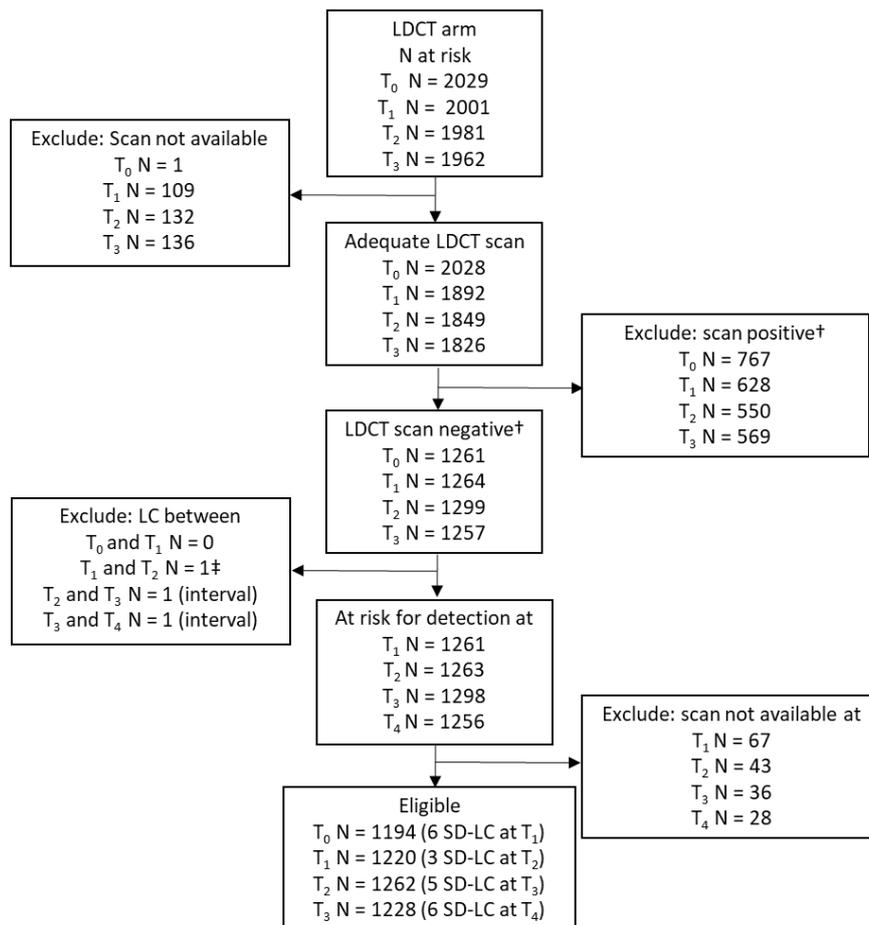
Figure 10. Participant selection flow chart for the validation of the LCRAT+CT model

Abbreviations: LDCT: low-dose computed tomography; LC: lung cancer
† NLST positive: at least one non-calcified nodule ≥4mm in longest diameter; negative: absence of non-calcified nodules ≥4mm in longest diameter. ‡ Excluded due to a lung cancer diagnosis based on additional LDCT findings in the absence of nodules.
As published in (González Maldonado et al. 2021a), reprinted with permission.

Interval cancers occurring between screening appointments (N=1 in the year between T_2 and T_3 and N=1 in the year between T_3 and T_4) were excluded (Figure 10).

The Polynomial model (Schreuder et al. 2018), and the Patz criterion (Patz et al. 2016) were validated on data from participants with available LDCT scan images at the first screening appointment (baseline screen) and at risk for lung cancer detection at the second annual screening appointment (N= 1,889). This excluded interval cancers occurring in the year between T_0-T_1 (N=1, Figure 11).

Two sensitivity analyses were conducted. First, the Polynomial model was applied to data from eligible subjects at all incidence rounds T_1-T_4 (Figure 11), and the risk of receiving a lung cancer diagnosis in subsequent years was estimated based on CT images obtained at the annual follow-up (incidence) screens. Second, though the differences in eligibility criteria and in the end-points they predict make the two models not directly comparable, the LCRAT+CT and Polynomial models were applied to the data set of subjects used for the validation of LCRAT+CT (i.e., showing no nodules ≥4mm in longest diameter).
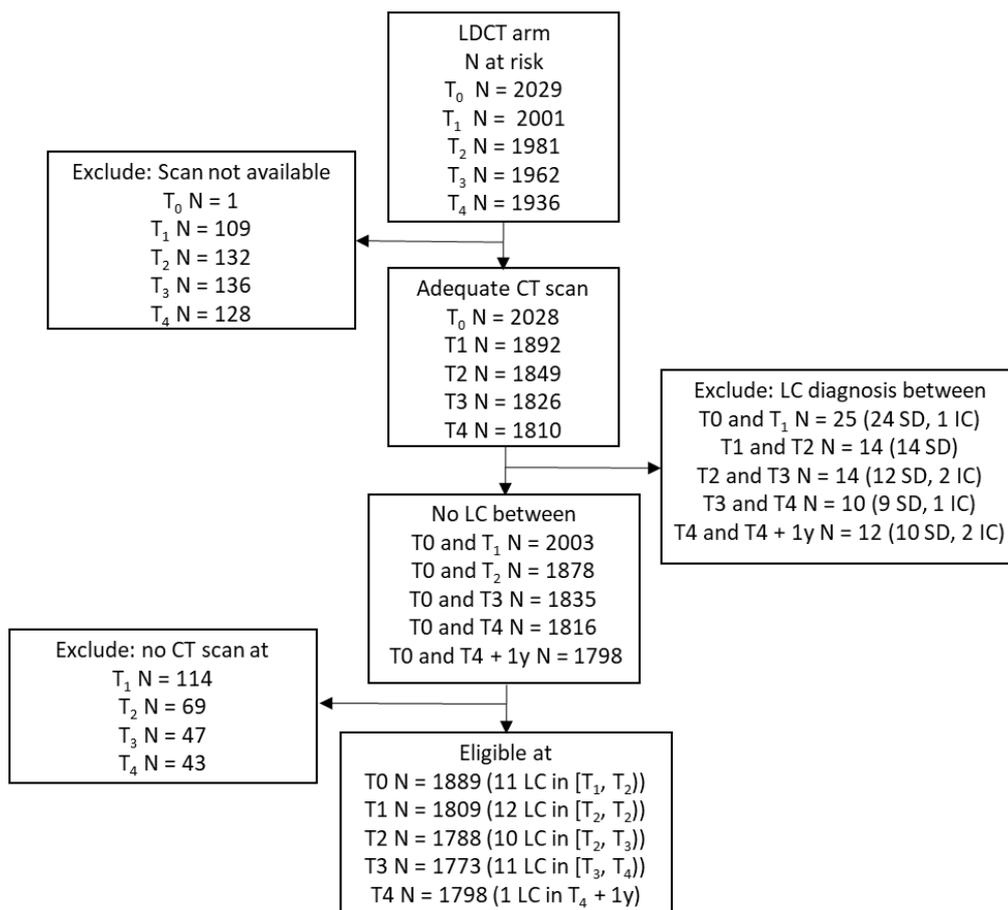
Figure 11. Participant selection flow chart for the validation of the Polynomial model

Abbreviations: LDCT: low-dose computed tomography; LC: lung cancer; SD: screen-detected; IC: interval cancer; y: year

As published in (González Maldonado et al. 2021a), reprinted with permission.

2.2.2 For the validation of nodule malignancy prediction models

Parts of this chapter have been published previously (González Maldonado et al. 2020a).

From all 2,029 participants of the screening arm of the LUSI trial, those without any non-calcified nodules in any of the CT scans (N=847) and those with screen-detected lung cancers but without unique identification of malignant nodules (N=23) were excluded for this validation study (Figure 12).

The selected nodule malignancy prediction models were evaluated on data from the LDCT scan image in which the individual nodules were first seen.
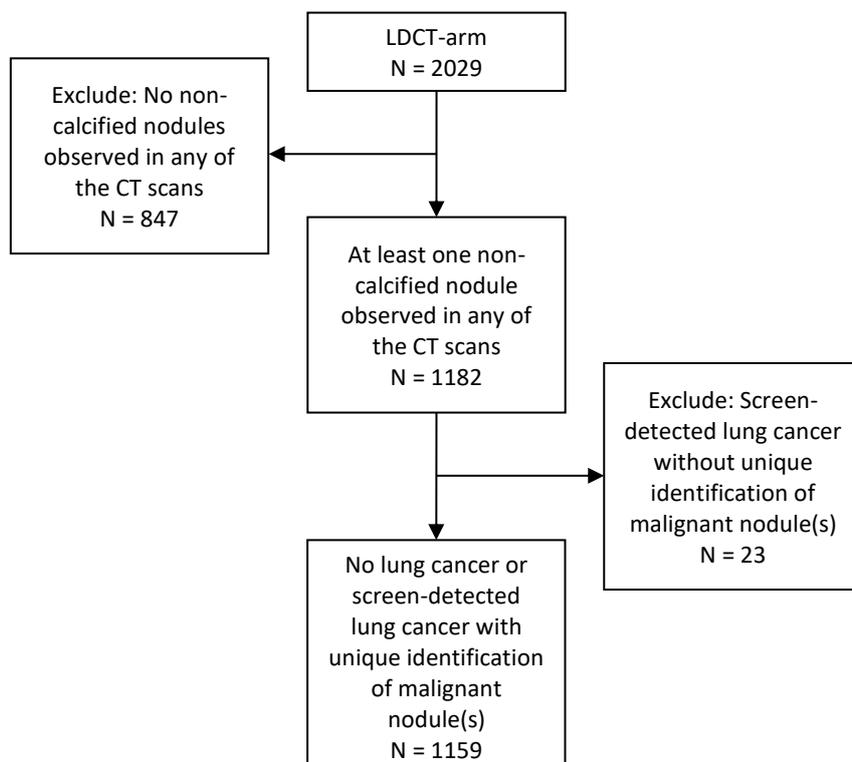
Figure 12. Flow chart illustrating inclusion and exclusion criteria applied to LUSI participants for the validation of the selected nodule malignancy prediction models.

2.2.3 For the estimation of overdiagnosis

Parts of this chapter have been published previously (González Maldonado et al. 2020b).

For the estimation of overdiagnosis in the LUSI trial, the end of follow-up for overall and lung cancer mortality was fixed on July 2nd 2019, the date of the most recent linkage between the trial and the mortality registers. Up until that date, more than 8 years had passed since the last trial participant was recruited (April 2011). Regarding lung cancer incidence, data was considered complete only until April 30th, 2019. This was decided in order to allow an approximate 1-year lag-time between diagnoses and their acknowledgment in the corresponding databases (i.e., in the LUSI trial or in the cancer registries).

2.2.4 For the evaluation of the TAAb panel EarlyCDT®-Lung

Parts of this chapter have been published previously (González Maldonado et al. 2021b).

A nested case-control design was used for the evaluation of EarlyCDT®-Lung on data from participants of the LUSI trial. All 63 screen-detected lung cancer cases and two sets of controls were selected. The first control set was a random selection of participants who had remained cancer-free until the end of follow up (April 30th 2019), and who provided a baseline blood sample (baseline control (BC) group, N=90). The second control set was a random selection of participants returning for follow-up scans of suspicious nodules found during the screening period but who were not diagnosed with lung cancer (suspicious nodule control (SNC) group, N=90) (Figure 13).

Only 46 of the 63 participants with screen-detected lung cancers and available blood samples taken at the time of the suspicious LDCT scan that led to further diagnostic work-up (X-ray, full-dose CT, PET, bronchoscopy, VATS, biopsy, antibiotic treatment and short-term follow-up) were selected for the intended analyses. Participants in the BC group were represented by the blood sample taken at the baseline examination, and those in the SNC group by the sample taken at the time of their first suspicious LDCT scan.

2.2.4.1 Sample processing and laboratory assays

Autoantibodies to seven tumor-associated antigens (CAGE, GBU4-5, HuD, MAGEA4, NY-ESO-1, p53, SOX2) were measured with the EarlyCDT®-Lung enzyme-linked immunosorbent assay (ELISA) kit (Oncimmune Ltd., Nottingham, United Kingdom). Measurements were carried out by trained specialists at the immunoassay laboratory of the division of cancer epidemiology of the German Cancer Research Center (DKFZ) Heidelberg, blinded with regards to any additional clinical information about the participants from which the samples were taken. All assays were performed on serum samples thawed for the first time for the purpose of the TAAb measurements, and using a two-plate set-up, according to the manufacturer's instructions.

The resulting optical density values were entered into the calculation table provided by Oncimmune. This calculation table indicated the control Pass/Fail status as well as the test result for the patient samples (No Significant Level, Moderate or High).

2.2.4.2 EarlyCDT®-Lung: interpretation and application

EarlyCDT®-Lung classifies test results using a proprietary scoring algorithm and autoantibody-specific cut-off values. "Moderate Level" (M) results are reported if the levels of one or more autoantibodies in the panel are above the low cut-off value but all below the high cut-off value, and reported as "High Level" (H) if the levels of one or more are above the high cut-off value.
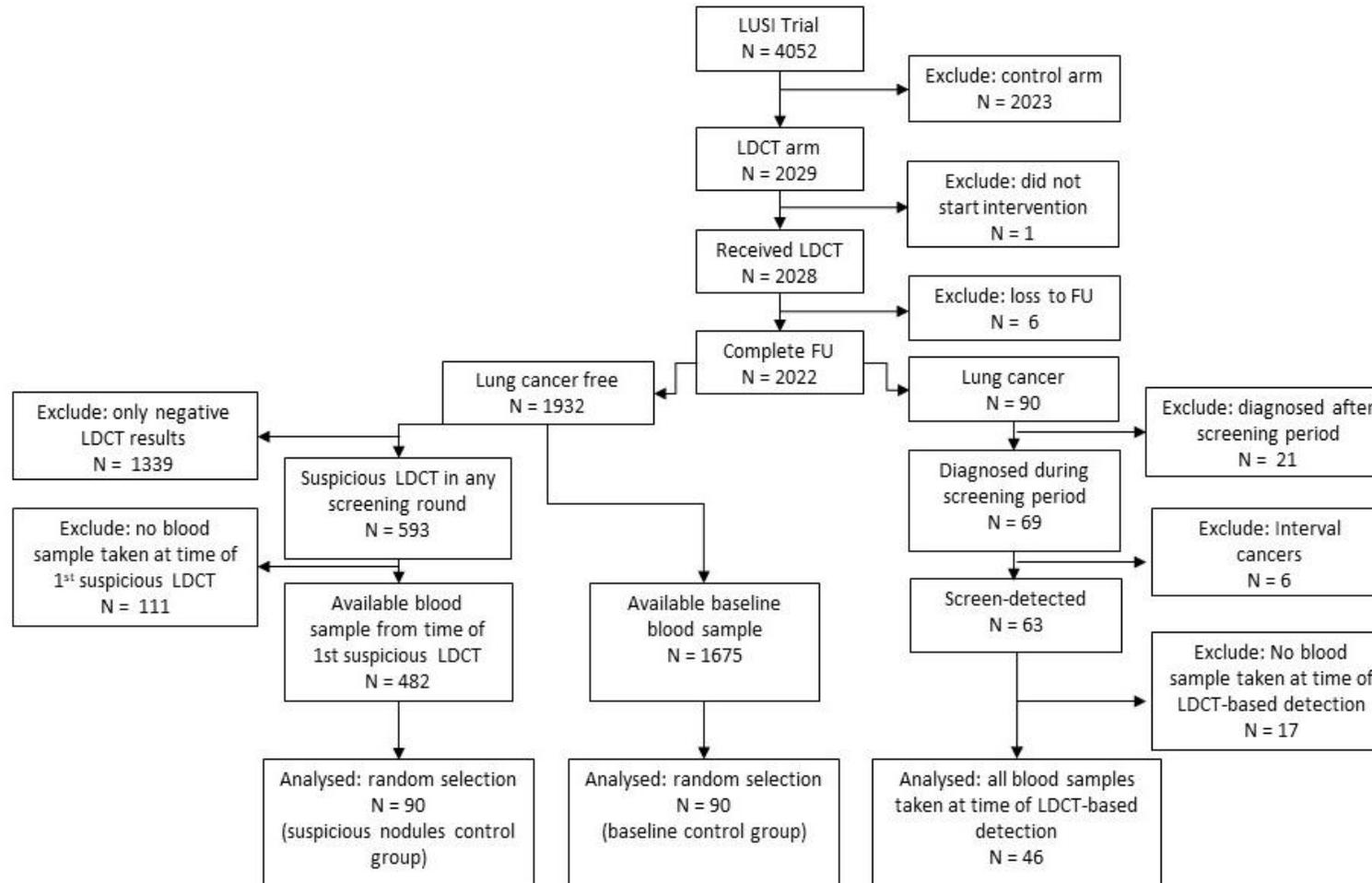
Figure 13. Flow chart illustrating inclusion and exclusion criteria applied to LUSI participants for the evaluation of EarlyCDT®-Lung.

As published in (González Maldonado et al. 2021b), reprinted with permission.

Both M and H results are considered positive results. "No Significant Level" (NS) is reported whenever the levels of all autoantibodies in the panel are below the low cut-off value.

Oncimmune recommends combining the test result with the estimated risk of a given nodule being malignant from a model based on age, sex and radiologic (CT) data, as developed in the Mayo clinic (Swensen et al. 1997). This recommendation was based on studies for lung cancer detection among patients with incidentally observed pulmonary nodules in clinical settings (Jett et al. 2014; Massion et al. 2017).

Early®CDT-Lung is intended to update the risk of patients presenting indeterminate nodules i.e., those with 10% - 65% risk of being malignant. For patients having ≥10% malignancy risk according to the Mayo model and with H EarlyCDT®-Lung results, and for those with estimated risks ≥45% and M test results, further diagnostic work-up is recommended. Any other result, including NS, should not affect the clinical management plan.

So far, no equivalent guidelines have been issued for the use of EarlyCDT®-Lung in population screening settings. In particular, no recommendations have been made regarding the scenario in which Early®CDT-Lung is used as a case-finding method prior to further CT investigations. Given the lack of concrete guidelines, in this study and for the purpose of evaluating the sensitivity of Early®CDT-Lung in the context of a screening trial, H or M level results were considered positive tests, and NS results were considered negative. This classification is equivalent to the 90% specificity/40% sensitivity scenario observed in several clinical studies (Chapman et al. 2012; Healey et al. 2017; Jett et al. 2014; Massion et al. 2017). For comparison purposes, analyses were also conducted for the alternative classification, defining only H levels as positive and M or NS as negative.

2.3 Statistical analyses

2.3.1 Validation of prediction models for the assignment of screening intervals

Parts of this chapter have been published previously (González Maldonado et al. 2021a).

2.3.1.1 Description of the selected risk prediction models

The LCRAT+CT model (Robbins et al. 2019) predicts the risk of lung cancer detection at the next annual screening by updating the 1-year lung cancer risk estimates obtained by the LCRAT model (Katki et al. 2016) with information from CT scan images deemed negative according to NLST criteria (absence of pulmonary nodules with largest diameter ≥4 mm). The model includes age, smoking history, family history of lung cancer, BMI and education level, together with LDCT imaging findings regarding the presence of pulmonary emphysema and consolidation (Table 3).

The Polynomial model (Schreuder et al. 2018) predicts the risk for lung cancer to be detected at the first follow-up screening appointment (T_1), or else diagnosed outside screening in the year (T_1, T_2) based on information available at the first screening appointment (T_0). It includes linear and/or 2nd-degree terms for a total of 11 selected risk factors including age, sex, smoking history, personal and family history of cancer, as well as LDCT scan findings such as pulmonary nodules and emphysema (Table 3)

The criterion by Patz (Patz et al. 2016) classifies participants according to the evaluation results of their LDCT imaging findings. Participants with negative test results according to NLST criteria are considered eligible for skipping the next screening round.

2.3.1.2 Model-based predictions and its evaluation

The scores of the LCRAT+CT and Polynomial models, as well as the Patz criterion were applied to data from eligible subjects as described in section (2.2.1).

For a few model variables, data were missing in the LUSI trial. This limitation was handled as follows. Race (for which no information was collected in the LUSI) was assumed Caucasian, reflecting the predominant demographic composition of the German population. The number of parents with lung cancer was assumed to be zero for all participants, given the low prevalence of the disease. History of emphysema or COPD (not explicitly asked in the recruitment or assessment questionnaires of the LUSI) was replaced by previous diagnosis of chronic bronchitis. Missing values (in <2% of participants for all variables) for education, BMI, smoking duration and time since quitting smoking were imputed by the median value recorded from participants within the same sex and age groups, and within the same smoking status group if applicable.

Participants without nodules were assigned values of zero for all nodule-related characteristics. Participants showing nodules, but for which nodule characteristics were missing (longest or perpendicular diameter, non-solid/solid, location, spiculation and/or nodule count were removed from the analysis (N=0 at T_0 and N=88 T_1 to T_4)). Positive LDCT test results were those triggering immediate referral for further diagnostic workup. Indeterminate test results were those triggering 3- or 6-months follow-up appointments.

The categories of the variable "highest education level" defined for the LCRAT+CT model according to the USA education system were linked to the available categories from the LUSI questionnaires, which were coded according to the German education system, as shown in Table 4.

Table 3. Coefficients of the LCRAT, LCRAT+CT and Polynomial models.

LCRAT		Polynomial model	
Predictor variable	β coefficient	Predictor variable	β coefficient
Sex (female)	-0.08057	Model constant	-28.15
Race (black/African-American)	0.217892	Age (y)	0.5845
Race (Hispanic)	-0.43413	Age^2 (y^2)	-0.004026
Race (other ethnicities)	-0.39556	Prior diagnosis of cancer	0.5555
Education (trend)	-0.07143	Smoking status (active)	0.5046
Number of parents with lung cancer	0.4183	Pack-years (y)	0.03922
Lung disease (COPD or emphysema)	0.563422	Pack-years^2 (y^2)	-0.0001632
BMI (≤18.5)	0.060925	Prior diagnosis of COPD	0.4144
Cigarettes per day (>20)	0.310609	Longest perpendicular diameter (mm)	0.09962
Pack years [30, 40)	0.491254	Longest perpendicular diameter^2 (mm^2)	-0.0006524
Pack years [40, 50)	0.562334	Presence of non-solid nodule	0.4217
Pack years (≥50)	0.715752	Presence of part solid nodule	0.9108
log(age)	4.386866	Presence of nodule in upper lobe	0.4685
log(BMI)	-0.72386	Presence of spiculated nodule	0.7512
log(years quit +1)	-0.3209	Nodule count per scan (per additional nodule)	0.5128
Years smoked	0.024002	Nodule count per scan^2 (per additional nodule^2)	-0.1947

LCRAT + CT	
Predictor variable	Exponent
Neither emphysema nor consolidation	1.08
Emphysema	0.96
Consolidation	0.77

Abbreviations: LCRAT: Lung Cancer Risk Assessment Tool; y: years; COPD: Chronic obstructive pulmonary disease; BMI: body mass index.

As published in (González Maldonado et al. 2021a), reprinted with permission.

Discrimination was evaluated via receiver operating characteristic (ROC) analysis. The stratified bootstrap method (B=10,000 repetitions) was used for the calculation of 95% confidence intervals (95% CI) for the area under the ROC curve (AUC). The method by DeLong et al. (DeLong et al. 1988) was used for testing the difference (inferiority) in AUC values of two models applied to the same data. Additionally, for all models, the numbers of participants who would have been candidates for skipping the next screening appointment were calculated using the deciles of predicted risk as thresholds. In addition, percentages of participants who would have had their diagnosis delayed if the screening round was skipped were estimated (point estimator and 95% CI), together with the percentages of false positive or indeterminate screen tests that would have been either avoided or delayed. The Wilson score method with continuity correction was used to obtain 95% CIs for the proportions of delayed diagnoses (Newcombe 1998).

Table 4. Equivalence between the education systems in the USA and Germany, as used for the validation of the selected prediction models on data from the LUSI Trial.

USA Education System	German Education System
1: <12 grade	kein Schulabschluss Volksschulabschluss / Hauptschulabschluss Mittlere Reife / Realschulabschluss keine berufliche Ausbildung und nicht in beruflicher Ausbildung
2: High school graduate	Noch in Ausbildung (Auszubildender, Student)
3: Post high school, no college	Fachhochschulreife / Fachoberschulabschluss Allgemeine Hochschulreife / Abitur Lehre (kaufmännisch) Lehre (gewerblich, technisch, landwirtschaftlich)
4: Associate degree/some college	Berufsfach-/Handelsschulabschluss Fach-/Meister-/Technikerschule, Berufs-/Fachakademie
5: Bachelor's degree	Not applicable
6: Graduate school	Fachhochschule (Ingenieurschule) Universität, Hochschule

As published in (González Maldonado et al. 2021a), reprinted with permission.

Calibration in-the-large (van Calster et al. 2019) was evaluated by comparing the mean predicted risk from the LCRAT, LCRAT+CT and Polynomial models to the observed incidence within the population of eligible subjects, either by screening round or separately for the prevalence and incidence rounds. Additionally, the models' calibration was evaluated via Brier Scores and Spiegelhalter Z-test (Brier 1950; Rufibach 2010; Spiegelhalter 1986). The Brier score is used for comparing the calibration of two prediction models, whereas the Spiegelhalter's z-statistic (z) is used for testing the null hypothesis of perfect calibration. Lower values of the Brier score indicate better calibration, while the null hypothesis of the Spiegelhalter's test is rejected at the significance level α if the absolute value of the z-score is larger than the α-quantile of the standard normal distribution (Rufibach 2010).

Statistical analyses were performed using the R language and environment for statistical computing version 3.4.4 (R Core Team 2018) and the lcrisk (Cheung et al. 2018), DescTools (Signorell 2020), rms (Harrell 2020), and pROC packages (Robin et al. 2011).

2.3.2 Validation of nodule malignancy prediction models

Parts of this chapter have been published previously (González Maldonado et al. 2020a)

2.3.2.1 Description of the selected nodule malignancy prediction models

Eight models from six published studies were selected for its validation on data from the LUSI trial. These models are: PanCan-1b (McWilliams et al. 2013) (parsimonious with spiculation), PanCan-2b (McWilliams et al. 2013) (full model including spiculation), PanCan-MD (Tammemägi et al. 2019a) (with the mean of the largest and perpendicular diameters as nodule size), PanCan-VOL (Tammemägi et al. 2019a) (with nodule volume as nodule size), the recently developed UKLS model (Marcus et al. 2019), and the models developed in the Veterans Affairs (VA) (Gould et al. 2007), Mayo (Swensen et al. 1997) and Peking University People´s Hospital (PKUPH) clinics (Li et al. 2011) (Table 5).

2.3.2.2 Model-based predictions and its evaluation

For participant-related characteristics, the Mann-Whitney-U-Test was used for investigating differences in continuous variables and the Chi-Squared test or Fisher's exact test for categorical variables. Differences in nodule-specific characteristics were analyzed using mixed-effects logistic regression including participant as random effect (Jiang 2007). Associations between risk of nodule malignancy and radiologic and/or participant-related parameters were explored by fitting multivariable logistic regression models via Generalized Estimating Equations (GEE) (Hardin and Hilbe 2003; Liang and Zeger 1986). The latter statistical method was selected in order to account for the correlation structure of multiple pulmonary nodules from any given trial participant. Model selection was based on the quasi-Akaike Information Criterion (QIC) (Pan 2001).

The ability of the selected models to discriminate malignant from non-malignant nodules was evaluated via cluster-adjusted ROC curves and AUCs (Obuchowski 1997a). Given the correlation structure of multiple nodules from any given trial participant, sensitivity, specificity, positive and negative predictive values were estimated using GEEs (Coughlin et al. 1992; Genders et al. 2012; Smith and Hadgu 1992).

The calibration of the selected models was assessed by examining observed vs. predicted nodule malignancy rates inside the categories of nodule size as defined in the LDCT evaluation algorithm of the LUSI trial (<5 mm, 5 to <8 mm, 8 to 10 mm, >10 mm), as well as by deciles of predicted risk. The Hosmer-Lemeshow goodness of fit test (Lemeshow and Hosmer 1982) was used to examine the fit between predicted and observed malignancy probabilities across deciles of predicted risk, and Brier scores and the Spiegelhalter's z-test were used to assess overall deviations of model risk predictions from observed rates (Rufibach 2010).

Table 5. Coefficients of the selected nodule malignancy prediction models

Predictors / variables	PanCan 1b[a]	PanCan 2b[a]	PanCan MD[b]	PanCan VOL[c]	UKLS	Mayo	PKUPH[d]	VA
Intercept	-6.6144	-6.7892	-6.5355	-6.4432	-2.2915	-6.8272	-4.496	-8.404
Age (years)		0.0287			-0.0257	0.0391	0.070	0.078
Sex (female vs male)	0.6467	0.6011	0.3749	0.3642	0.5105			
Family history of lung cancer[e] (yes/no)		0.2961			1.9985			2.061
Late onset (>60 y)					1.5724			
History of cancer[f] (excl. lung)					0.5305	1.3388		
Smoking (ever/never)						0.7917		2.061
Smoking cessation (years/10)								-0.567
Smoking duration (years)					0.0565			
Asbestos exposure (yes/no)					0.5884			
Asthma (yes/no)					-0.7777			
Bronchitis (yes/no)					1.7616			
Emphysema (yes/no)		0.2953	0.2879	0.2536				
FVC (L)					-1.1693			

Left side row group label: **Participant-Related Characteristics**

Table 5 (continued). Coefficients of the selected lung-cancer risk prediction models

Nodule-Related Characteristics	**Nodule size (mm, mm³)**								
	Largest diameter	-5.5537	-5.3854				0.1274	0.0676	0.112
	Mean diameter			-16.1232					
	Volume				-9.2285	0.00082			
	Nodule Type (ref: solid)								
	Non-solid, or GGO[g]		-0.1276	ref	ref	<u>1.6396</u>			
	Part-solid		0.3770	0.7005	0.7439	<u>0.4919</u>			
	Nodule location (upper/else)	0.6009	0.6581	0.5029	0.5012	<u>-0.1799</u>	0.7838		
	Nodule count per scan		-0.0824	-0.0853	-0.0865				
	Spiculation (yes/no)	0.9309	0.7729	0.9699	0.9502		1.0407	0.736	
	Border (clear/other)							- 1.408	

When applying the PanCan models, nodule size was transformed as follows:
    a For models 1b and 2b: largest diameter' = (((largest diameter)/10)^(-0.5))-1.58113883
    b For the mean diameter model: mean diameter' = ln(1/(largest diameter + perpendicular diameter)/2)- 0.6227482239
    c For the volume model: nodule volume' = (ln(nodule volume)/10)^(-0.5)- 1.619158938, and nodule count was centered: Nodule count'= nodule count – 4
    d In the PKUPH model nodule diameter was measured in centimeters
    e Family history of lung cancer was not available in the data from the LUSI trial and was therefore assigned a value of zero for all participants and all models.
    f For the Mayo model, history of cancer (excluding lung) refers to those diagnosed more than 5 years before randomization

Underlined coefficients indicate variables or variable levels excluded in the alternative version of the UKLS model
Abbreviations: PanCanMD: PanCan model with mean diameter; PanCanVOL: PanCan model with volume; UKLS: United Kingdom Lung Cancer Screening trial; PKUPH: Peking University People´s Hospital; VA: Veterans Affairs; FVC: Forced vital capacity in liters; GGO = Ground Glass Opacity
As published in (González Maldonado et al. 2020a), reprinted with permission.

For comparison, three additional multivariable logistic regression models were fitted on data from pulmonary nodules first observed in any screening round of the LUSI trial. Such models included a larger set of variables compared to the models selected for validation, and differed only in the measurement they used to define nodule size: largest diameter, mean diameter or nodule volume. The discrimination capacity of these newly fitted models was evaluated with the methods applied to the predictions of the models selected for validation. The variability of the model performance estimates was assessed using cluster bootstrapping (B=1,000) (Obuchowski 1997a). Model selection was done based on the QIC. Feature selection was done through backward elimination based on p-values (>0.05). At each elimination step, the reduced models were re-ranked according to their QIC as calculated by the function "model.sel", available in the MuMIn R package.

All analyses were carried out using the R language and environment for statistical computing version 3.5.1 (R Core Team 2018), with packages gee (Carey 2019), Hmisc (Harrell et al. 2020), lme4 (Bates et al. 2015), MuMIn (Barton 2020), rms, pROC (Robin et al. 2011) and ROCR (Sing et al. 2005), as well as the "clusteredROC" function (Obuchowski 1997a; Obuchowski 1997b).

2.3.3 Estimation of overdiagnosis and related parameters

2.3.3.1 Excess incidence

Parts of this chapter have been published previously (González Maldonado et al. 2020b)

The extent of overdiagnosis in the LUSI trial was estimated by calculating the excess incidence, that is, the difference in cumulative incidence of lung cancer in the LDCT and control arms, expressed as a ratio relative to the cumulative incidence of screen-detected lung cancers (PS) (Patz 2006; Patz et al. 2014; Welch and Black 2010). Based on an alternative definition, excess incidence was also calculated relative to all lung cancers in the LDCT arm of the LUSI trial (PA) regardless of whether they were screen-detected or not (Patz et al. 2014). End of follow-up for lung cancer diagnosis was set to April 30th 2019. The precision of the estimates was assessed by calculating their 95% confidence intervals via bootstrapping (5,000 repetitions).

2.3.3.2 Convolution model for clinical incidence and distribution of pre-clinical sojourn time

Besides the estimates of excess incidence, a convolution model for clinical incidence was used for the joint estimation of mean pre-clinical tumor sojourn time (MPST) and LDCT-based screen detection sensitivity in the screening arm (Straatman et al. 1997; Walter and Day 1983). A binomial distribution was assumed for the observed numbers of screen-detected cases and

cases diagnosed either in between screening appointments (interval cancers) or after the last screening appointment. Then, based on estimates of MPST, probabilities for screen-detected tumors to have remained in a pre-clinical (asymptomatic) phase were estimated, depending on the follow-up time after screen-detection (Paci et al. 2004; Pashayan et al. 2009). These probabilities were obtained for lung cancer overall and for subgroups defined by tumor histology. According to the convolution model (Walter and Day 1983), the clinical incidence of cancer (symptomatic cancers) at time t in the absence of screening is given by:

$$I(t) = \int_0^t g(s)f(t - s \mid s)\, ds$$

with g the incidence of cancers in the pre-clinical detectable phase (PDP) of length X in which tumors are detectable but asymptomatic, and f the density of the pre-clinical sojourn time of length Y (time between T_0 the start of the PDP and T_2 the tumor's clinical manifestation).

In this study, both X and Y were assumed to follow exponential distributions with $r > 0$ (pre-clinical incidence rate) and $\lambda > 0$ (transition rate from the pre-clinical into the clinical stage), such that $f_X(x) = re^{-rx}$ and $f_Y(y) = \lambda e^{-\lambda y}$ (Day and Walter 1984; Straatman et al. 1997; Walter and Day 1983). Under these assumptions, the expected value of the pre-clinical sojourn time (mean pre-clinical sojourn time) can be expressed as $E(Y) = 1/\lambda$.

Furthermore, both r and λ lambda were assumed to be constant over time, in agreement with the stable disease model. Also, assuming a low incidence of pre-clinical lung cancer, the probability density function of the duration of the PDP was approximated as $f_X(x) = re^{-rx} \approx r$. Under these assumptions, and as suggested by Straatman (Straatman et al. 1997) the clinical incidence at time t can be written as:

$$I(t) \approx \int_0^t r \left(\int_0^{t-x} \lambda e^{-\lambda y} dy \right) dx = rt - \frac{r}{\lambda}\left[1 - e^{-\lambda t}\right]$$

If screening takes place at t_1, $(T_0 < t_1)$, an existing tumor will be detected with probability ϕ = sensitivity of the detection method. If detection occurs at $T_0 < t_1 < T_2$, $T_2 - t_1$ is defined as the lead time. An interesting point to notice is that, under the exponential distribution assumption, the mean lead time equals mean pre-clinical sojourn time.

Likelihood of screen-detected and interval cases

Let screening occur at $t_1, t_2, \ldots, t_n$ years after age a (at which it can be assumed that no one has entered the PDP); and let the detection sensitivity be ϕ. Also, let i_0 be the number of subjects diagnosed before t_1; N_j the number of subjects screened at t_j; s_j the number of screen-detected cases at $t_j, 1 \leq j \leq k$; and i_j = number of interval (or incident) cancers in (t_j, t_{j+1}), $1 \leq j \leq k$. In the absence of loss to follow-up, as described in Straatman (Straatman et al. 1997), the vector $(i_0, s_1, i_1, \ldots, s_k, i_k)$ follows a multinomial distribution.

A difficulty when applying this model in the context of a screening trial, such as the LUSI, is that trials usually do not consider a pre-screening period. In the absence of data about diagnoses given before the occurrence of the first screening appointment (i_0), and in order to avoid potential bias through the estimation of the underlying cancer incidence based on data from cancer registries of from the control arm of such trials, Straatman (Straatman et al. 1997) suggests estimating the mean pre-clinical sojourn time (MPST) and the detection sensitivity by maximizing a partial likelihood for the vector of screen-detected and interval lung cancers $(s_1, i_1, \dots, s_k, i_k)$. The author further suggests describing such vector as the result of repeated binomial experiments, each with $s_j + i_j$ trials and s_j successes, with success probability $P[S_j]_\phi / (P[S_j]_\phi + P[I_j]_\phi)$ thus yielding the following partial likelihood:

$$\mathcal{L}_2 = \prod_{j=1}^{k} \binom{s_j + i_j}{s_j} \left(\frac{P[S_j]_\phi}{P[S_j]_\phi + P[I_j]_\phi} \right)^{s_j} \left(1 - \frac{P[S_j]_\phi}{P[S_j]_\phi + P[I_j]_\phi} \right)^{i_j}$$

where

$$P[I_0] \approx \int_0^{t_1} r \left(\int_0^{t_1 - x} \lambda e^{-\lambda y} dy \right) dx = r t_1 - \frac{r}{\lambda} \left[1 - e^{-\lambda t_1} \right]$$

$$P[S_1]_\phi = \phi P[S_1] = \phi \frac{r}{\lambda} \left[1 - e^{-\lambda t_1} \right]$$

$$P[I_1]_\phi = P[S_1](1 - \phi)\left(1 - e^{-\lambda(t_2 - t_1)} \right) + P[I_1]$$
$$= \frac{r}{\lambda} \left[1 - e^{-\lambda t_1} \right](1 - \phi)\left[1 - e^{-\lambda(t_2 - t_1)} \right] + r[t_2 - t_1] - \frac{r}{\lambda} \left[1 - e^{-\lambda(t_2 - t_1)} \right]$$

And in general, for $2 < j \le k$:

$$P[S_j]_\phi = P[S_{j-1}](1 - \phi)e^{-\lambda(t_j - t_{j-1})} + P[S_j]\phi$$

$$P[I_j]_\phi = P[S_j](1 - \phi)\left[1 - e^{-\lambda(t_{j+1} - t_j)} \right] + P[I_j]$$

Straatman further suggests the 95% joint confidence region for λ and ϕ to be defined as

$$\left\{ (\lambda, \phi) \mid 2 \left[\ln \left(\mathcal{L}_2(s_1, i_1, \dots, s_k, i_k; \hat{\lambda}, \hat{\phi}) \right) - \ln \left(\mathcal{L}_2(s_1, i_1, \dots, s_k, i_k; \lambda, \phi) \right) \right] < \chi_\alpha^2(2) \right\}$$

where $\ln$ is the natural logarithm, $\hat{\lambda}$ and $\hat{\phi}$ are the maximum likelihood estimates for (λ, ϕ), and $\chi_\alpha^2(2)$ is the quantile $100\%\alpha$ of $\chi_\alpha^2(2)$ distribution.

Estimation of mean pre-clinical sojourn time (MPST) and LDCT detection sensitivity in data from the LUSI trial

As mentioned in the previous section, the LUSI trial did not consider a pre-screening phase. Thus estimates of the mean pre-clinical sojourn time $(1/\lambda)$ and LDCT-based lung cancer

detection sensitivity (ϕ), were obtained by maximizing $\mathcal{L}_2$ under the constraints $\lambda > 0$ and ϕ in [0,1], as suggested by Straatman (Straatman et al. 1997).

The date of screening-related detection was defined as the date of the annual scan on which suspicious nodules were found, eventually leading to diagnostic work-up (independently of the date of diagnosis or histological confirmation). For all incident cases not detected via screening, the date of diagnosis was used as endpoint. Estimates for MPST and sensitivity were obtained for various values of a (age at which the PDP has not been reached) ranging from 0 to 30 years. Age at first screen was taken to be 57 years, which reflects the median age at first screen for the participants in the LUSI trial.

In order to better fit the characteristics of the LUSI trial, the model was further adjusted, as suggested by Straatman (Straatman et al. 1997): a) separate models by age at first screening (50-59 and 60+ years) were fitted in order to account for the heterogeneity in age at entry, b) no adjustment was applied regarding irregular attendance to screenings (>93.4% attendance in all rounds), c) loss to follow-up (death or migration) was adjusted by modifying the success probability in the partial likelihood including an estimate of the probability of loss to follow-up in between screens and after the last screening appointment (<0.6% for all yearly intervals up until year 7, 1.4% for years 7-8, 25.6% for years 8-9, 40.6% for years 9-10, 76.9% for years 10-11), d) increasing incidence with age was taken into account by including the adjustment factor suggested by Straatman (Straatman et al. 1997) to the probability of detection in the first screening appointment. Additionally, separate models were fitted by tumor histology subgroups (adenocarcinomas excluding bronchioalveolar adenocarcinomas (BAC), BAC, others).

Overall model fit was evaluated based on the observed and expected screen-detected cases, via the χ^2 goodness-of-fit test with 5 (number of rounds)–2=3 degrees of freedom.

Estimation of the probabilities of overdiagnosis and early detection

Under the assumption of an exponential distribution of pre-clinical sojourn time, which was made in this study, the probability of LDCT-based overdiagnosis (probability that a subject with an LDCT-detected lung cancer at time t_1 remains asymptomatic within her expected remaining lifetime (ERL)) is (Paci et al. 2004; Pashayan et al. 2009):

$$p(y > ERL \mid y > t_1) = \int_{ERL}^{\infty} \lambda e^{-\lambda y} dy = e^{-\lambda ERL}.$$

Similarly, the probability that the same screen-detected case becomes symptomatic within the next t years (for example, within the duration of post-screening follow-up and before death) is:

$$Pclin = p(y < t) = \int_{0}^{t} \lambda e^{-\lambda y} dy = 1 - e^{-\lambda t}$$

which can be interpreted as the probability that a screen-detected case was a product of early diagnosis rather than overdiagnosis.

These probabilities can be estimated by substituting $\hat{\lambda}$, the reciprocal of the estimated mean pre-clinical sojourn time and age-specific estimates of ERL.

Statistical analyses were performed using the R language and environment for statistical computing version 3.4.4 (R Core Team 2018) and the boot package (Canty and Ripley 2019; Davison and Hinkley 1997).

The R-code implemented and applied for the estimation of MPST, LDCT-detection sensitivity and proportion of tumors by lead time can be found in the Appendix (Section 10.1).

2.3.4 Evaluation of the TAAb panel EarlyCDT®-Lung

Parts of this chapter have been published previously (González Maldonado et al. 2021b).

The sensitivity of the EarlyCDT®-Lung test was assessed among participants with lung cancer detected through LDCT-screening. Additionally, the specificity of the test was assessed in each of the two control groups. Positive likelihood ratios (LR+) were calculated based on the estimates of sensitivity and specificity. Exact binomial confidence limits were calculated for sensitivity, specificity and LR+. Associations of EarlyCDT®-Lung test results with case-control status were assessed via logistic regression. Complementary statistical tests were performed to assess the relationship between positive test results and tumor characteristics such as size, stage or histology. The degree of association between test results and malignancy was evaluated among lung cancer patients showing nodules on their LDCT-scans and in the SNC group, overall and by categories of nodule size. For continuous variables, medians and ranges were reported if the variables were normally distributed, or medians and interquartile ranges (IQR) otherwise. Differences in central tendency parameters between groups were tested via the non-parametric Kruskal-Wallis Rank Sum Test (Mann-Whitney-U-Test when comparing two groups). Differences in distributions of categorical variables were tested via the Chi-Squared or Fisher's exact test as appropriate, depending on cell counts. All analyses were carried out using the R language and environment for statistical computing version 3.3.3 (R Core Team 2017) and the epiR package (Stevenson et al. 2020).

3 RESULTS

3.1 Validation of prediction models for the assignment of screening intervals

Parts of this chapter have been published previously (González Maldonado et al. 2021a).

There were 1,506 participants eligible for the validation of the LCRAT+CT model and 1,889 eligible for the validation of the Polynomial model.

For the LCRAT+CT model, some participants were eligible at multiple rounds, given that they remained at risk for lung cancer detection for an extended time period. These were selected as follows: 1,194 at the initial (prevalence) screen (T_0), and 1,220, 1,262 and 1,228 participants at the three following incidence screens (T_1-T_3) (Figure 10). The median age of these participants was 56.80 years range:[50.30, 71.80] at first screening participation. All were long-term smokers, and 960 of them were males (63.7%). Lung cancer was detected via LDCT for 24 of these eligible participants; 20 of these detections occurred at the annual screening appointment following a negative screening test result and were thus included for the validation of LCRAT+CT (Figure 10, Supplementary Table 1).

All 1,889 participants eligible for the validation of the Polynomial model (Figure 11) were long-term smokers with a median age of 56.80 years, range:[50.30, 71.90] at first screening participation, and 1,238 of them were males (65.5%) (Supplementary Table 1). Eleven out of these eligible participants received a lung cancer diagnosis either as a result of further work-up triggered by positive LDCT findings at T_1, or by other means outside screening in the year after T_1 (Supplementary Table 1).

3.1.1 Distribution of absolute risk estimates

Estimates from both models varied widely across participants. The LCRAT+CT estimated risks for lung cancer detection at the next annual screening appointment in a range of 0.009% to 2.76% (Figure 14). Estimated risks for lung cancer diagnosis (screen-detected or diagnosed outside screening) in the year [T_1, T_2) from the Polynomial model ranged from about 0.001% to 8.34% (Figure 15). Highest risks estimated by the LCRAT+CT model were obtained for individuals with LDCT-based indications of both consolidation and emphysema (N=9, 0.60% of eligible participants, contributing with 9 estimated risk values in rounds T_0-T_3), consolidation without emphysema (N=5, 0.33% of eligible participants, contributing with 5 estimated risk values in rounds T_0-T_3), and emphysema (N=786, 52.2% of eligible participants, contributing to 2,156 estimated risk values in rounds T_0-T_3).
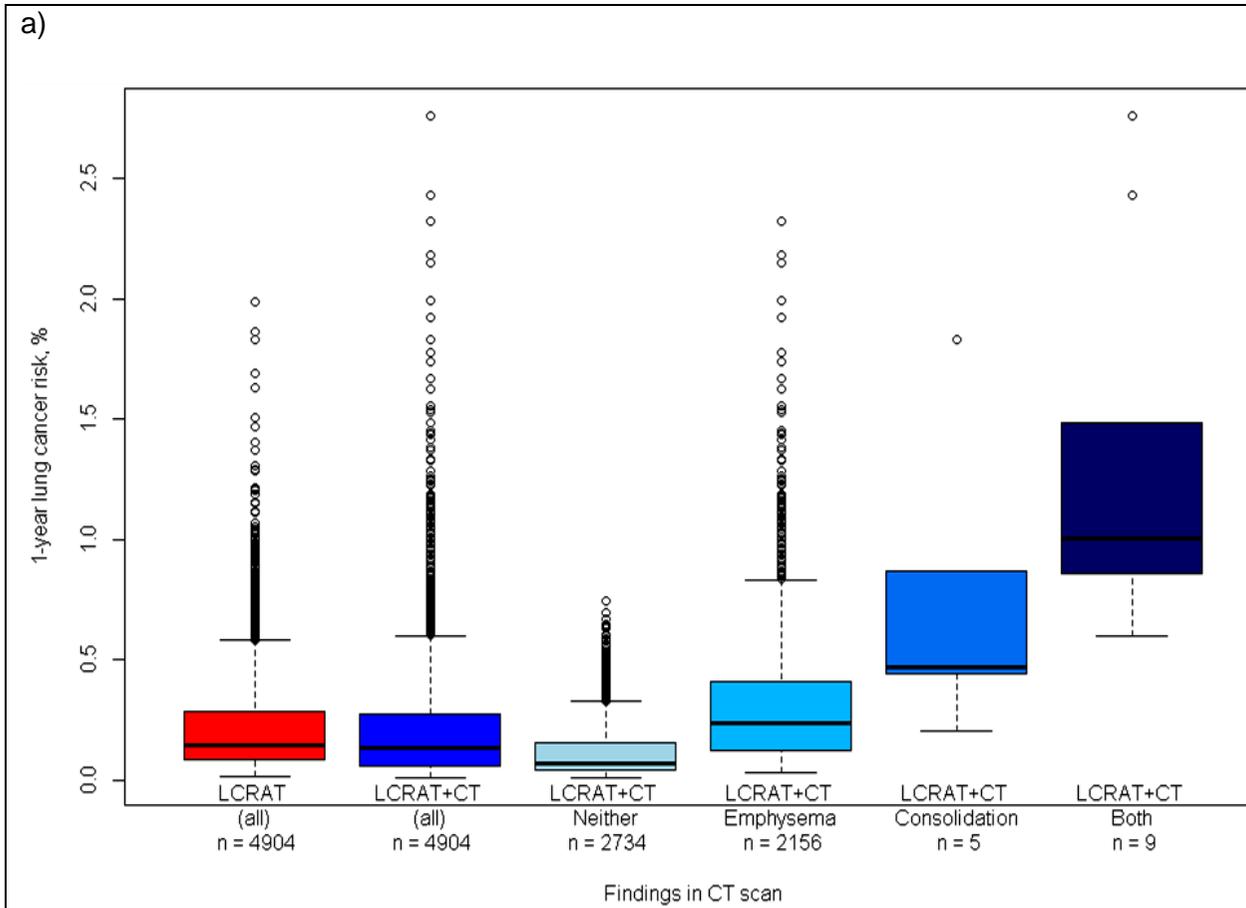
Figure 14. Distribution of predicted risks for the LCRAT, LCRAT+CT models.
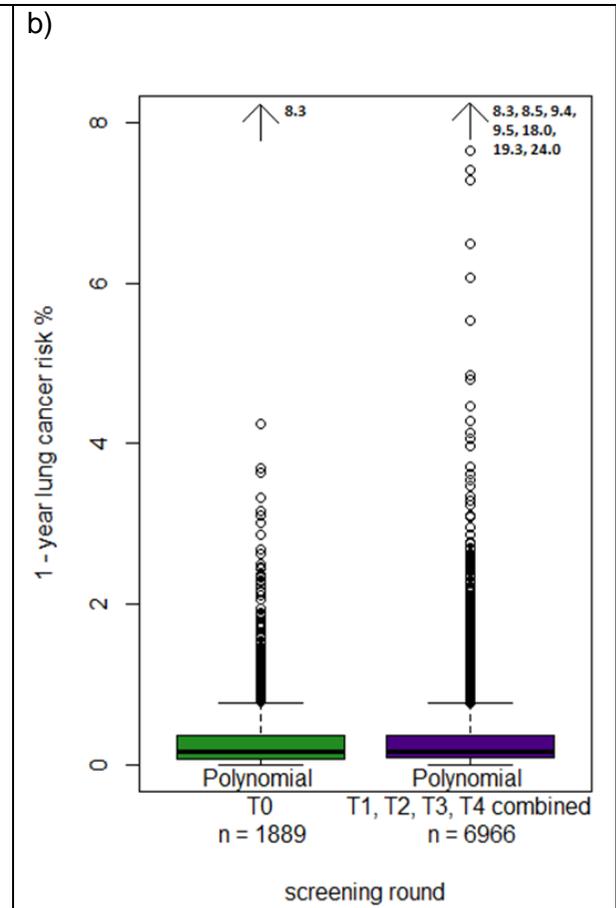
Figure 15. Distribution of predicted risks for the Polynomial model

As published in (González Maldonado et al. 2021a), reprinted with permission.

For the Polynomial model, highest risks were estimated for older participants with more pack-years, higher nodule-count per LDCT scan, higher nodule-count in the upper lobes of the lung, and higher count of nodules showing border spiculation.

3.1.2 Discrimination of participants with short-term lung cancer diagnoses

Analyzing data from T_0 to T_3, LCRAT+CT achieved an AUC of 0.73 95% CI:[0.63, 0.82] for the discrimination of participants with lung cancer detected at the next screening appointment. For comparison, the original LCRAT model without CT data (Katki et al. 2016) showed a lower AUC of 0.68 95% CI:[0.57, 0.78] (Supplementary Figure 1). However, the difference in AUC between LCRAT and LCRAT+CT was not statistically significant (z=-1.44, p=0.08).

The Polynomial model applied to data from the baseline (prevalence) screen showed an AUC of 0.75 95% CI:[0.67, 0.83] (Supplementary Figure 2) for the discrimination of participants who in the following year were diagnosed with lung cancer either through screening or independently of screening, from those who remained cancer-free. Applied to the combined data from the incidence screening rounds (T_1-T_4) the AUC was 0.74 95% CI:[0.65, 0.82].

For comparison purposes, the Polynomial model was applied to individuals presenting no nodules ≥4mm in diameter (i.e., eligible for the LCRAT+CT model). In this subset of data, the discrimination by the Polynomial model (AUC = 0.76 95% CI:[0.66, 0.87] at T_0, AUC = 0.72 95% CI:[0.62, 0.81] in T_0-T_3) was similar to that of the LCRAT+CT (AUC of 0.73 95% CI:[0.63, 0.82]) (Supplementary Figure 3).

Finally, the dichotomous Patz criterion produced an AUC of 0.56 95% CI:[0.53, 0.72] on baseline screen data (Supplementary Figure 4).

3.1.3 Risk-based assignment to biennial vs annual screening and its effect on delayed diagnosis

According to the risk estimates from LCRAT+CT, skipping about 40% to 50% of annual screenings (following a negative screen-test as of NLST criteria), that is, next-screening appointments for participants with estimated risks below 0.1% and 0.13% (fourth and fifth deciles of risk) respectively, would have avoided or delayed 1 (25% of all false positive test results 95% CI:[1.3%, 78.1%]) to 3 (75% 95% CI:[21.9%, 98.7%]) false positive screening tests and 3 (42.9% 95% CI:[11.8%, 79.8%]) indeterminate nodule findings, at the cost of 1 (5% 95% CI:[0.3%, 26.9%]) to 2 (10% 95% CI:[1.8%, 33.1%]) delayed lung cancer detections (Table 6). Using risk estimates from the LCRAT model, at equal proportions of annual screenings skipped, there were generally higher numbers of detections delayed compared to LCRAT+CT (Figure 16), and slightly higher numbers of false positive or indeterminate screening tests (data not shown).

Table 6. Potential effects of biennial screening based on risk thresholds from the LCRAT and LCRAT+CT models in eligible participants of the LUSI Trial

Percentile of risk	LCRAT+CT next-scan risk	Candidates for Longer Interval	Delayed cancer detections	False positives avoided / delayed	Indeterminates avoided / delayed
		N (%; 95% CI)	N (%; 95% CI)	N (%; 95% CI)	N (%; 95% CI)
10th	r ≤ 0.03%	491 (10)	1 (5; 0.3, 26.9)	0 (0; 0, 60.4)	1 (14.3; 0.8, 58)
20th	r ≤ 0.05%	981 (20)	1 (5; 0.3, 26.9)	0 (0; 0, 60.4)	3 (42.9; 11.8, 79.8)
30th	r ≤ 0.07%	1,471 (30)	1 (5; 0.3, 26.9)	0 (0; 0, 60.4)	3 (42.9; 11.8, 79.8)
40th	r ≤ 0.1%	1,962 (40)	1 (5; 0.3, 26.9)	1 (25; 1.3, 78.1)	3 (42.9; 11.8, 79.8)
50th	r ≤ 0.13%	2,452 (50)	2 (10; 1.8, 33.1)	3 (75; 21.9, 98.7)	3 (42.9; 11.8, 79.8)
60th	r ≤ 0.17%	2,942 (60)	5 (25; 9.6, 49.4)	3 (75; 21.9, 98.7)	4 (57.1; 20.2, 88.2)
70th	r ≤ 0.23%	3,433 (70)	7 (35; 16.3, 59.1)	3 (75; 21.9, 98.7)	4 (57.1; 20.2, 88.2)
80th	r ≤ 0.32%	3,923 (80)	12 (60; 36.4, 80)	3 (75; 21.9, 98.7)	5 (71.4; 30.3, 94.9)
90th	r ≤ 0.48%	4,413 (90)	16 (80; 55.7, 93.4)	4 (100; 39.6, 97.6)	6 (85.7; 42, 99.2)
100th	r ≤ 2.76%	4,904 (100)	20 (100; 80, 99.5)	4 (100; 39.6, 97.6)	7 (100; 56.1, 98.7)

As published in (González Maldonado et al. 2021a), reprinted with permission.

Based on risk estimates from the Polynomial model, skipping the second round (T_1) for 40% to 50% of participants, that is, those with model risks below 0.13% and 0.17% at T_0, would have avoided or delayed 10 (40% of all false positive test results 95% CI:[21.8%, 61.1%]) false positive screening tests and between 144 (38.8% 95% CI:[33.9%, 44%] and 173 (46.6% 95% CI:[41.5% to 51.8%]) indeterminate screenings without delaying any diagnosis (0 95% CI:[0%, 32.1%]) (Table 7, Figure 17). For comparison, the Patz criterion indicates that if all participants (N=1,194; 63.2% 95%CI:[61%, 65.4%]) with a negative T_0 scan would have skipped T_1, 1 (4% [0.2%, 22.3%]) false positive screen tests and 3 (0.8% [0.2%, 2.5%]) indeterminate scans could have been avoided, and 6 (54.5% [24.6%, 81.9%]) cancer diagnoses would have been delayed. For both the LCRAT+CT and Polynomial models (as applied to their respective eligible subsets) no statistically significant association was found between predicted model risks and the tumor stage of lung cancers detected upon next annual screening, although this analysis was hampered by small overall case-numbers (results not shown).
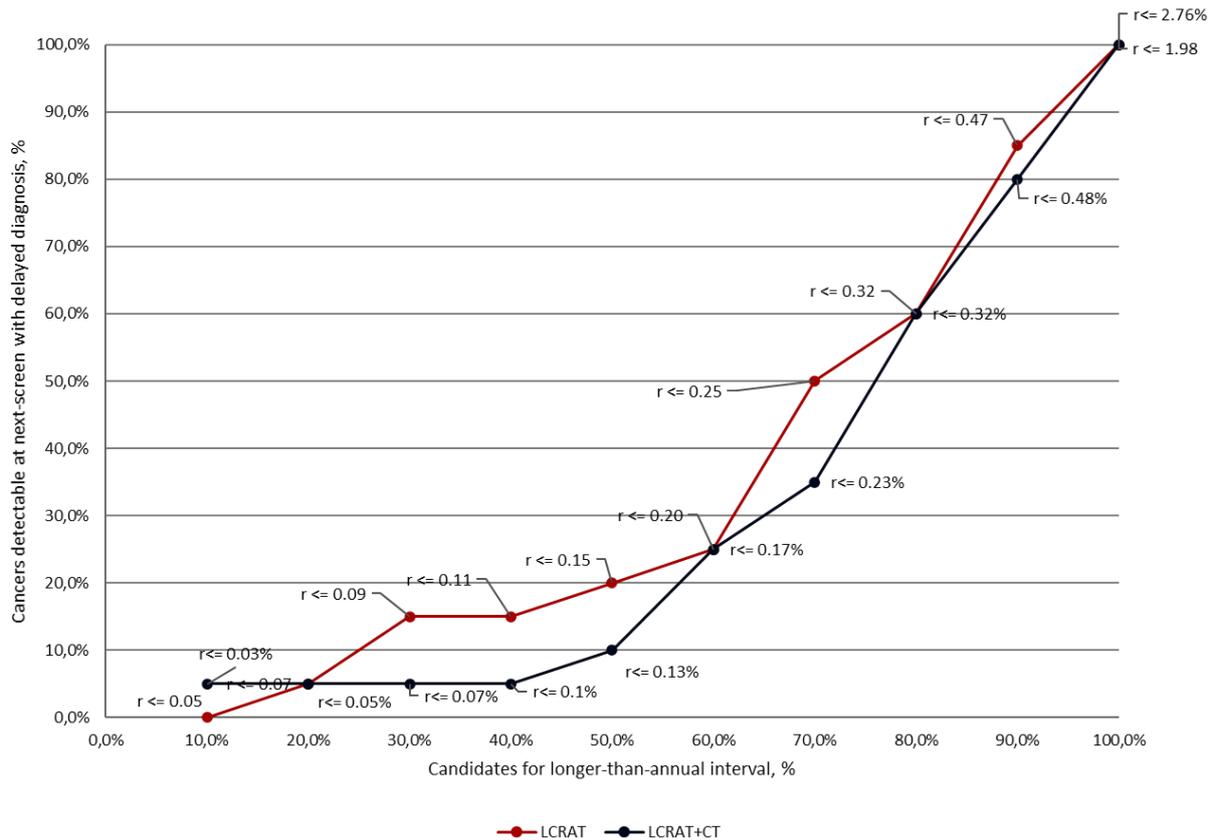
Figure 16. Potential effects of biennial screening based on risk thresholds from the LCRAT and LCRAT+CT models in eligible participants of the LUSI trial

As published in (González Maldonado et al. 2021a), reprinted with permission.

In the combined data from T_1 to T_4, the Polynomial model predicted 15 (18.8% 95% CI:[11.2%, 29.4%]) to 17 (21.2% 95% CI:[13.2%, 32.1%]) avoided false positive screen tests and 41 (18% 95% CI:[13.3%, 23.7%]) to 58 (25.4% [20%, 31.7%]) avoided indeterminate findings at the cost of delaying 4 (12.5% [4.1%, 29.9%]) to 6 (18.8% [7.9%, 37%]) lung cancer detections, by skipping 40% to 50% next-round screenings (those of participants with risks below 0.14% and 0.18%) (Table 7). For participants eligible for the LCRAT+CT model (i.e., those presenting no pulmonary nodules ≥ 4mm), the Polynomial model predicted 0 (0% 95% CI:[0%, 69%]) avoided false positives and 2 (33.3% 95% CI:[6%, 75.9%]) avoided indeterminant results. This was at the cost of delaying 3 (13.6% 95% CI:[3.6%, 36%]) lung cancer detections by skipping 50% of screenings (i.e., if those with a risk below 0.14% were recommended to skip the screening) (Supplementary Table 2, Supplementary Figure 5).
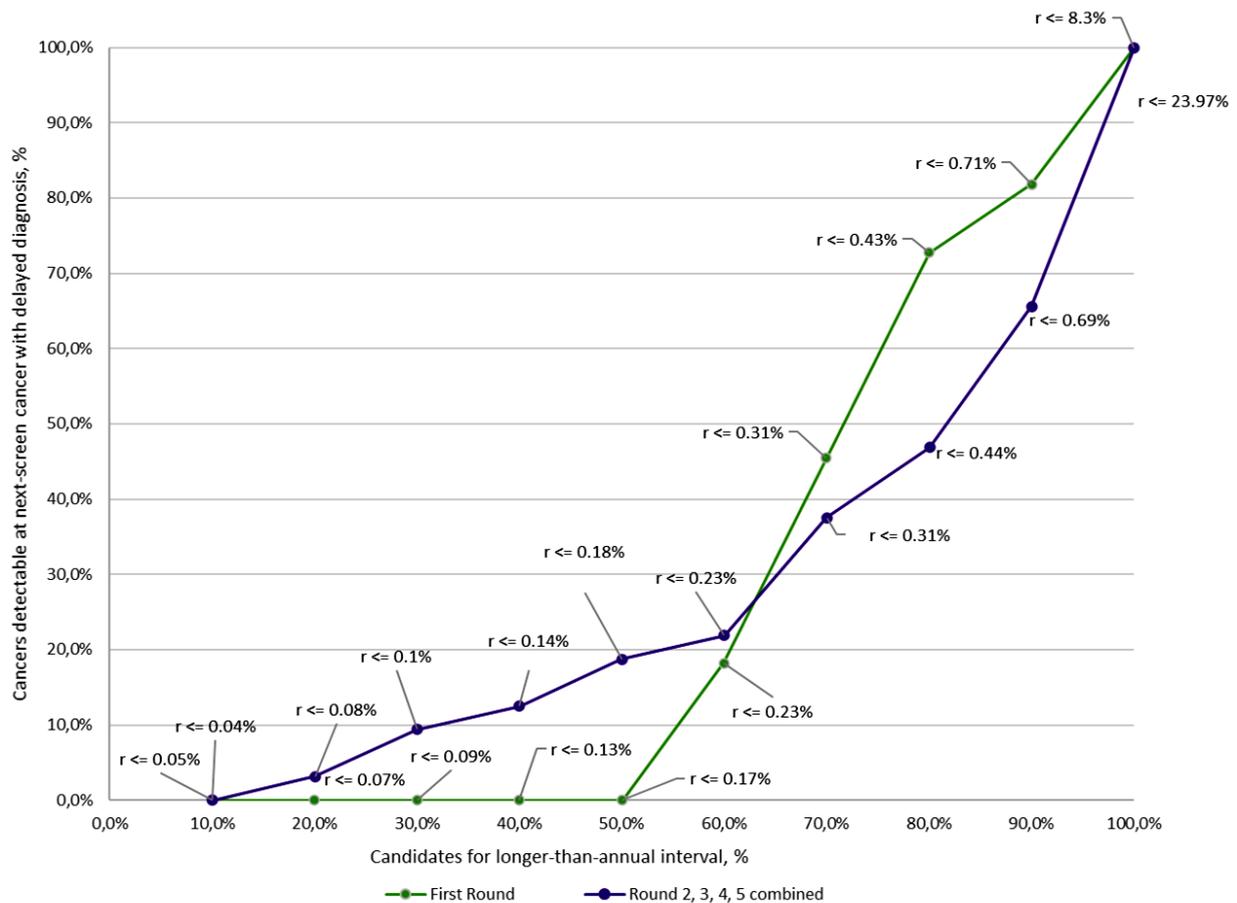
Figure 17. Potential effects of biennial screening based on risk thresholds the Polynomial model in eligible participants of the LUSI Trial

As published in (González Maldonado et al. 2021a), reprinted with permission

3.1.4 Calibration of absolute lung cancer risk

All models produced absolute risk estimates that were, on average, considerably lower than the observed lung cancer prevalence. Brier scores for the LCRAT, LCRAT+CT and Polynomial models were not significantly different from one another, indicating a similar calibration for the three models. For LCRAT and LCRAT+CT, the null hypothesis of calibration was rejected at $\alpha=0.05$ when applied to the combined data of screening rounds T_0 (prevalence round) to T_3 (3rd incidence screening) (p=0.004 for LCRAT, p=0.002 for LCRAT+CT), and also when applied to the data only from the incidence rounds T_1 to T_3 (p=0.049 for LCRAT and p=0.036 for LCRAT+CT). The same hypothesis was rejected at $\alpha=0.05$ when applied to the estimated risks from the Polynomial model from T_0 (p=0.032) and T_1 to T_4 (0.048) (Supplementary Table 3 Supplementary Table 4).

Table 7. Potential effects of biennial screening based on risk thresholds from the Polynomial model in eligible participants of the LUSI Trial

Percent ile of risk	Polyno mial risk (T0)	Candidates for Longer Interval	Delayed Cancers	False positives avoided	Indeterminates avoided
		N (%; 95% CI)	N (%; 95% CI)	N (%; 95% CI)	N (%; 95% CI)
10th	r ≤ 0.04%	189 (10)	0 (0; 0, 32.1)	8 (32; 15.7, 53.6)	83 (22.4; 18.3, 27)
20th	r ≤ 0.07%	378 (20)	0 (0; 0, 32.1)	9 (36; 18.7, 57.4)	108 (29.1; 24.6, 34.1)
30th	r ≤ 0.09%	567 (30)	0 (0; 0, 32.1)	9 (36; 18.7, 57.4)	128 (34.5; 29.7, 39.6)
40th	r ≤ 0.13%	756 (40)	0 (0; 0, 32.1)	10 (40; 21.8, 61.1)	144 (38.8; 33.9, 44)
50th	r ≤ 0.17%	945 (50)	0 (0; 0, 32.1)	10 (40; 21.8, 61.1)	173 (46.6; 41.5, 51.8)
60th	r ≤ 0.23%	1,133 (60)	2 (18.2; 3.2, 52.2)	11 (44; 25, 64.7)	195 (52.6; 47.3, 57.7)
70th	r ≤ 0.31%	1,322 (70)	5 (45.5; 18.1, 75.4)	14 (56; 35.3, 75)	237 (63.9; 58.7, 68.7)
80th	r ≤ 0.43%	1,511 (80)	8 (72.7; 39.3, 92.7)	17 (68; 46.4, 84.3)	267 (72; 67.1, 76.4)
90th	r ≤ 0.71%	1,700 (90)	9 (81.8; 47.8, 96.8)	20 (80; 58.7, 92.4)	314 (84.6; 80.5, 88.1)
100th	r ≤ 8.3%	1,889 (100)	11 (100; 67.9, 99.2)	25 (100; 83.4, 99.6)	371 (100; 98.7, 100)
Percent ile of risk	Polyno mial risk (T1-T4)	Candidates for Longer Interval	Delayed Cancers	False positives avoided	Indeterminates avoided
		N (%; 95% CI)	N (%; 95% CI)	N (%; 95% CI)	N (%; 95% CI)
10th	r ≤ 0.05%	699 (10)	0 (0; 0.3, 13.3)	13 (16.2; 9.3, 26.6)	16 (7; 4.2, 11.4)
20th	r ≤ 0.08%	1,395 (20)	1 (3.1; 0.2, 18)	14 (17.5; 10.2, 28)	26 (11.4; 7.7, 16.4)
30th	r ≤ 0.1%	2,090 (30)	3 (9.4; 2.5, 26.2)	15 (18.8; 11.2, 29.4)	31 (13.6; 9.6, 18.9)
40th	r ≤ 0.14%	2,787 (40)	4 (12.5; 4.1, 29.9)	15 (18.8; 11.2, 29.4)	41 (18; 13.3, 23.7)
50th	r ≤ 0.18%	3,483 (50)	6 (18.8; 7.9, 37)	17 (21.2; 13.2, 32.1)	58 (25.4; 20, 31.7)
60th	r ≤ 0.23%	4,181 (60)	7 (21.9; 9.9, 40.4)	20 (25; 16.3, 36.2)	76 (33.3; 27.3, 39.9)
70th	r ≤ 0.31%	4,876 (70)	12 (37.5; 21.7, 56.3)	25 (31.2; 21.6, 42.7)	110 (48.2; 41.6, 54.9)
80th	r ≤ 0.44%	5,576 (80)	15 (46.9; 29.5, 65)	36 (45; 34, 56.5)	134 (58.8; 52.1, 65.2)
90th	r ≤ 0.69%	6,269 (90)	21 (65.6; 46.8, 80.8)	46 (57.5; 46, 68.3)	162 (71.1; 64.6, 76.8)
100th	r ≤ 23.97%	6,966 (100)	32 (100; 86.7, 99.7)	80 (100; 94.3, 99.9)	228 (100; 97.9, 100)

As published in (González Maldonado et al. 2021a), reprinted with permission.

3.2 Validation of nodule malignancy prediction models

Parts of this chapter have been published previously (González Maldonado et al. 2020a).

A total of 62 participants were diagnosed with lung cancer up to 12 months after their last LDCT screening participation: 56 of them were screen-detected and 6 were interval cancers (Becker et al. 2019). For 54 out of the 56 screen-detected cancer cases, malignancy was linked to one (N=51) or more (N=3) nodules. For two screen-detected cases without nodules, the decision of referral to further diagnostic work-up was based on other pulmonary abnormalities. For two of the 6 interval cases no nodules were observed on LDCT screens; for the other 4, with non-suspicious nodules, no information was available to unequivocally link any of these to malignancy. Nodules with benign-appearing calcification and all lung cancer cases (N=8) for whom lung tumors could not be linked back to a specific nodule were excluded from statistical analyses.

Considering all 5 screening rounds, 1,182 participants of the LDCT arm of the LUSI trial showed at least one non-calcified pulmonary nodule. After removing participants with a lung cancer diagnosis given after more than one year following the last CT scan or for whom the malignant nodule(s) could not be identified in earlier CT images, data was available for 3,903 pulmonary nodules from 1,159 individuals (Figure 12). Detailed nodule counts by size (diameter), type (solid vs. sub-solid), malignancy status and screening round of first observation (prevalence or incidence) are shown in Supplementary Table 5. Of the 3,903 nodules observed, 2883 (32 of them identified as malignant) were first seen during the prevalence screen, whereas 1020 nodules (31 malignant), were first observed in one of the incidence screens. Irrespective of screening round, over 70% of malignant nodules (25 of 32 in prevalence round; 25 of 31 in incidence rounds) had a diameter ≥8mm. For all such nodules (≥8 mm), the average malignancy rate was higher in the prevalence round (12.8%) compared to the incidence round (8.6%); conversely, higher for nodules <8mm first observed in incidence rounds (0.8%) compared to the prevalence round (0.3%).

Irrespective of screening round of first detection, compared to benign nodules, the malignant ones were significantly more often located in the left upper lobe, were larger in terms of diameter and volume and more often had spiculated borders (Table 8). Participant related characteristics (Table 9) significantly associated with diagnosis of malignant nodules include age, presence of emphysema, and FEV_1; whereas sex, smoking status at randomization (current vs. former), smoking duration and intensity, self-reported history of extra-thoracic cancer, years since smoking cessation, presence of asthma, bronchitis, and FVC showed no significant association in the LUSI data. Family history of cancer was not available in data from the LUSI.

Table 8. Nodule characteristics by malignancy status and screening round

First seen on	Prevalence round			Any incidence round			Any round		
Characteristic (N (%) or median (range))	benign	malignant	P Value	benign	malignant	P Value	benign	malignant	P Value
Number of nodules	2,851 (98.9)	32 (1.1)		989 (97.0)	31 (3.0)		3840 (98.4)	63 (1.6)	
Diameter (mm)	4.3 [2.30, 64.2]	10.6 [3.9, 107.7]	<0.001	5.90 [2.2, 65.6]	11.60 [3.5, 38.7]		4.8 [2.2, 65.6]	11.6 [3.5, 107.7]	<0.001
Mean diameter (mm)	3.7 [1.9, 35.7]	9.3 [3.7, 93.7]	<0.001	4.95 [2, 51.9]	9.45 [2.7, 29.9]		4.0 [1.9, 51.9]	9.4 [2.75, 93.7]	<0.001
Volume (mm^3)	33.7 [5, 8960.3]	384.9 [19, 17466]	<0.001	63.2 [8, 28,947.9]	355.9 [21.1, 7,726.8]	0.09	41.4 [5.0, 28,947]	357.1 [19.3, 17,466]	<0.001
Type (Solid/Subsolid)	2,561/290 (89.8/10.2)	30/2 (93.8/6.2)	0.01	862/127 (87.2/12.8)	28/3 (90.3/9.7)	0.62	3423/417 (89.1/10.9)	58/5 (92.1/7.9)	0.09
Location									
Right upper	715 (25.1)	9 (28.1)		260 (26.3)	7 (22.6)		975 (25.4)	16 (25.4)	
Right middle	278 (9.8)	1 (3.1)		98 (9.9)	1 (3.2)		376 (9.8)	2 (3.2)	
Right lower	682 (23.9)	9 (28.1)	0.03	217 (21.9)	4 (12.9)	<0.001	899 (23.4)	13 (20.6)	<0.001
Left upper	438 (15.4)	9 (28.1)		123 (12.4)	14 (45.2)		561 (14.6)	23 (36.5)	
Left lower	596 (20.9)	4 (12.5)		251 (25.4)	4 (12.9)		847 (22.1)	8 (12.7)	
Lingula	111 (3.9)	0 (0.0)		33 (3.3)	1 (3.2)		144 (3.8)	1 (1.6)	
Unclear	31 (1.1)	0 (0.0)		7 (0.7)	0 (0.0)		38 (1.0)	0 (0.0)	
Nodule shape									
Spiculated	79 (2.8)	13 (40.6)		114 (11.5)	10 (32.3)		193 (5.0)	23 (36.5)	
Non-spiculated	2,277(79.9)	11 (34.4)	<0.001	619 (62.6)	14 (45.2)	0.74	2896 (75.4)	25 (39.7)	<0.001
Unclear	495 (17.4)	8 (25.0)		256 (25.9)	7 (22.6)		751 (19.6)	15 (23.8)	
Calcification									
Non-calcified	2,785 (97.7)	30 (96.9)	0.86	943 (95.2)	29 (93.5)	0.29	3728 (97.1)	60 (95.2)	0.83
Unclear	66 (2.3)	1 (3.1)		46 (4.7)	2 (6.5)		112 (2.9)	3 (4.8)	
Nodule count p/image	4 [1, 22]	2 [1, 16]	0.32	3 [1, 22]	2 [1, 20]	0.32	4 [1, 22]	2 [1, 20]	0.20

All p-values come from a chi-squared test for the difference in deviance of two mixed-effects logistic models: both with participant as random effect and one of them additionally having the corresponding nodule characteristic as fixed effect.

As published in (González Maldonado et al. 2020a), reprinted with permission.

Table 9. Participant-related characteristics at time of randomization

	Non-cases	Cases	P Value*	Overall
Number of participants	1,105	54		1,159
Age (median [range]	57.38 [50.34, 71.89]	59.88 [51.90, 69.98]	<0.001	57.63 [50.34, 71.89]
Sex (female / male (%))	379/726 (34.3/65.7)	17/37 (31.5/68.5)	0.78	396/763 (34.2/65.8)
Smoking status (current / former (%))	685/420 (62.0/38.0)	33/21 (61.1/38.9)	1	718/441 (61.9/38.1)
Smoking duration (years) (median [range])	37.50 [27.50, 52.50]	37.50 [27.50, 52.50]	0.06	37.50 [27.50, 52.50]
Smoking intensity (cigarettes/day) (median [range])	22.50 [12.50, 62.50]	22.50 [12.50, 62.50]	0.31	22.50 [12.50, 62.50]
History of cancer (excl. lung; > 5 years ago) (yes/no (%))	2/1,103 (0.2/99.8)	1/53 (1.0/98.1)	0.13	3/1,156 (0.3/99.7)
History of cancer (excl. lung) (yes/no (%))	103/1,002 (9.3/90.7)	6/48 (11.1/88.9)	0.63	109/1,050 (9.4/90.6)
Years since quitting (excluding current smokers) (median [range])	4 [0.04, 13]	8 [0.30, 13]	0.16	4 [0.04, 13.00]
Emphysema at randomization (yes/no (%))	474/631 (42.9/57.1)	35/19 (64.8/35.2)	0.002	509/650 (43.9/56.1)
Asthma (yes/no (%))	47/1058 (4.3/95.7)	0/54 (0/100)	0.16	47/1,112 (4.1/95.9)
Bronchitis (yes/no (%))	201/904 (18.2/81.8)	8/46 (14.8/85.2)	0.72	209/950 (18.0/82.0)
FEV$_1$ (median [range])	2.88 [0.66, 6.11]	2.66 [1.26, 4.24]	0.009	2.87 [0.66, 6.11]
FVC (median [range])	3.75 [0.33, 6.78]	3.45 [1.96, 5.69]	0.08	3.75 [0.33, 6.78]

Abbreviations: FEV$_1$: 1-second forced expiratory volume; FVC: Forced vital capacity.

As published in (González Maldonado et al. 2020a), reprinted with permission.

3.2.1 Discrimination of malignant nodules

For nodules detected at the participants' first screen (prevalence round), all PanCan models showed high discrimination accuracy, with AUCs ranging from 0.93 95% CI:[0.87, 0.99]) (PanCan-1b) to 0.94 (PanCan-2b 95% CI:[0.89, 0.99], PanCan-MD 95% CI:[0.91, 0.98], PanCan-VOL 95% CI:[0.90, 0.98]) whereas the discrimination by UKLS was poor (AUC=0.58 95% CI:[0.46, 0.70]). Models originally developed in clinical settings (Mayo, VA, PKUPH) showed moderately good discrimination, with AUCs between 0.84 95% CI:[0.76, 0.92] and 0.89 95% CI:[0.82, 0.97] (Figure 18). The comprehensive PanCan-2b model achieved only marginally better discrimination than the parsimonious model (PanCan-1b), and the use of volume or two-dimensional perpendicular mean diameter as measures for nodule size (PanCan-VOL, PanCan-MD) did not improve discrimination performance over that achieved by using largest diameter (PanCan-2b).

All models showed useful, though reduced discrimination when applied to nodules first noticed in any of the incidence rounds (e.g., for the PanCan models, AUCs between 0.81 95% CI:[0.73, 0.89] and 0.83 95% CI:[0.75, 0.88]) (Figure 19), which may be explained by the low variability in nodule size (Table 8**,** Supplementary Table 5).

Sensitivity, specificity, positive (PPV) and negative predicted values (NPV) were estimated at 2%, 5% or 10% risk thresholds for the PanCan and UKLS models (Table 10) applied on nodules observed in the prevalence screening round. The PanCan-1b model showed the highest sensitivity (80.6% 95% CI:[0.67, 0.94]) at 2% risk to 51.6% 95% CI:[0.32, 0.71] at 10%), but lower specificity compared to the other models (90.0%-98.6%). Highest specificity was observed for PanCan-MD (95.9% 95% CI:[0.95, 0.97] at 2% risk threshold to 99.1% 95% CI:[0.99, 1] at 10%) at the cost of decreased sensitivity (71.5% to 42.8%). At 2% or 5% risk thresholds, PPV ranged from, respectively, 8.3% 95% CI:[0.05, 0.12] and 19.3% 95% CI:[0.11, 0.27] for PanCan-1b, to 17.4% 95% CI:[0.10, 0.24] and 32.7% 95% CI:[0.20, 0.45] for PanCan-MD. Finally, the UKLS model showed inferior sensitivity and specificity compared to all PanCan models.
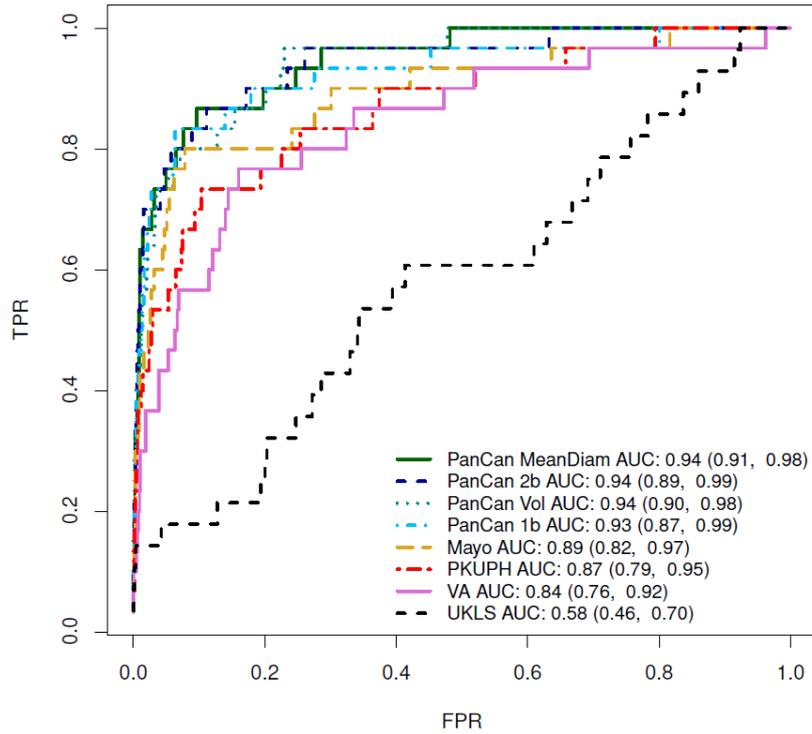
Figure 18. Receiver Operating Characteristic Curves and AUC with 95% confidence intervals (nodules first seen in the prevalence round)
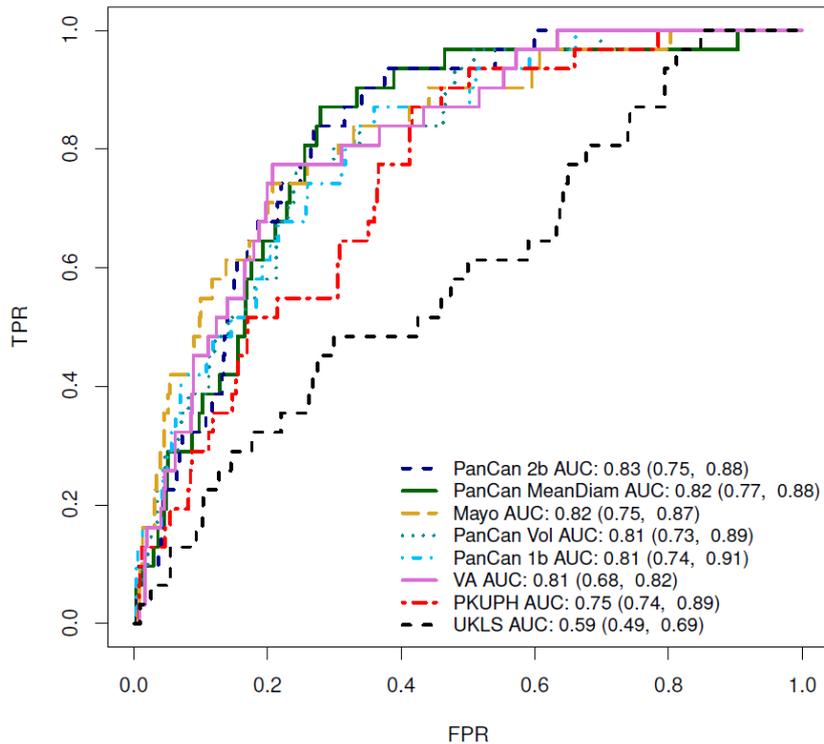
Figure 19. Receiver Operating Characteristic Curves and AUC with 95% confidence intervals (nodules first seen in any of the incidence rounds)

Areas under the curve are accompanied by 95% confidence intervals

Abbreviations: FPR: False Positive Rate; TPR: True Positive Rate; AUC: Area Under the Curve; MeanDiam: Mean diameter: = (largest nodule diameter + perpendicular diameter)/2; Vol: Volume; VA: Veterans Affairs; UKLS: United Kingdom Lung Cancer Screening trial.

As published in (González Maldonado et al. 2020a), reprinted with permission.

Table 10. Discrimination performance at fixed risk thresholds in LUSI (prevalence round) and other screening studies

Model, Dataset	Measure (95% Confidence Interval)	Threshold of Predicted Risk		
		0.02	0.05	0.10
PanCan 1b LUSI	Sensitivity	0.806 (0.67, 0.94)	0.715 (0.56, 0.87)	0.516 (0.32, 0.71)
	Specificity	0.900 (0.89, 0.92)	0.967 (0.96, 0.98)	0.986 (0.98, 0.99)
	PPV	0.083 (0.05, 0.12)	0.193 (0.11, 0.27)	0.287 (0.16, 0.42)
	NPV	0.998 (1, 1)	0.997 (0.99, 1)	0.994 (0.99, 1)
PanCan 2b LUSI	Sensitivity	0.785 (0.65, 0.92)	0.685 (0.51, 0.86)	0.497 (0.32, 0.67)
	Specificity	0.923 (0.91, 0.94)	0.972 (0.96, 0.98)	0.991 (0.99, 1)
	PPV	0.111 (0.07, 0.16)	0.233 (0.14, 0.32)	0.393 (0.24, 0.55)
	NPV	0.998 (1, 1)	0.996 (0.99, 1)	0.994 (0.99, 1)
PanCan MD LUSI	Sensitivity	0.715 (0.56, 0.86)	0.663 (0.48, 0.84)	0.428 (0.25, 0.61)
	Specificity	0.959 (0.95, 0.97)	0.984 (0.98, 0.99)	0.991 (0.99, 1)
	PPV	0.174 (0.1, 0.24)	0.327 (0.2, 0.45)	0.356 (0.19, 0.52)
	NPV	0.997 (0.99, 1)	0.996 (0.99, 1)	0.993 (0.99, 1)
PanCan VOL LUSI	Sensitivity	0.742 (0.6, 0.88)	0.557 (0.37, 0.74)	0.438 (0.26, 0.61)
	Specificity	0.941 (0.93, 0.95)	0.978 (0.97, 0.98)	0.991 (0.99, 0.99)
	PPV	0.136 (0.08, 0.19)	0.239 (0.13, 0.34)	0.351 (0.19, 0.51)
	NPV	0.997 (1, 1)	0.995 (0.99, 1)	0.993 (0.99, 1)
UKLS LUSI	Sensitivity	0.246 (0.11, 0.39)	0.197 (0.07, 0.33)	0.14 (0.01, 0.27)
	Specificity	0.829 (0.81, 0.85)	0.924 (0.91, 0.94)	0.972 (0.96, 0.98)
	PPV	0.015 (0, 0.03)	0.023 (0, 0.04)	0.042 (0, 0.08)
	NPV	0.990 (0.99, 0.99)	0.990 (0.99, 0.99)	0.990 (0.99, 0.99)

Abbreviations: PPV= Positive Predictive Value; NPV= Negative Predictive Value; PanCan MD= PanCan model with mean diameter; PanCan VOL= PanCan model with volume. [a] No confidence intervals provided in the original manuscripts.

As published in (González Maldonado et al. 2020a), reprinted with permission.

Table 10 (continued). Discrimination performance at fixed risk thresholds in LUSI (prevalence round) and other screening studies

Model, Dataset, Source	Measure (95% Confidence Interval)	Threshold of Predicted Risk		
		0.02	0.05	0.10
PanCan 1b PanCan McWilliams, 2013[a]	Sensitivity	0.847	0.714	0.602
	Specificity	0.896	0.955	0.975
	PPV	0.105	0.185	0.254
	NPV	0.998	0.996	0.994
PanCan-MD PanCan Tammemägi, 2018[a]	Sensitivity	0.897	0.838	0.701
	Specificity	0.869	0.928	0.962
	PPV	0.124	0.194	0.277
	NPV	0.998	0.996	0.994
PanCan-VOL PanCan Tammemägi, 2018[a]	Sensitivity	0.906	0.846	0.692
	Specificity	0.870	0.931	0.963
	PPV	0.126	0.201	0.279
	NPV	0.998	0.997	0.993
PanCan-2b NLST White, 2017[a]	Sensitivity	0.966	0.931	0.853
	Specificity	0.768	0.893	0.939
	PPV	0.101	0.189	0.274
	NPV	0.999	0.998	0.996

Abbreviations: PPV= Positive Predictive Value; NPV= Negative Predictive Value; MD= mean diameter; VOL= volume. a No confidence intervals provided in the original manuscripts.

As published in (González Maldonado et al. 2020a), reprinted with permission.

3.2.2 Calibration of absolute nodule malignancy risk

Absolute malignancy probabilities predicted by the PanCan-1b model were the closest to the observed proportions of malignant nodules detected at the prevalence screening round within categories defined by nodule size (Table 11). More detailed analyses within deciles of the predicted probability scores (Supplementary Table 6), combined with Hosmer-Lemeshow (HL) tests for deviance between predicted and observed rates, showed acceptable calibration for PanCan-1b, PanCan-2b and PanCan-VOL, but not for PanCan-MD and UKLS (HL test p<0.001). Brier scores and Spiegelhalter-z tests suggested acceptable overall calibration for PanCan-1b, PanCan-2b and (borderline; p=0.058) for PanCan-VOL, as well as for UKLS, but not PanCan-MD (Supplementary Table 8).

Malignancy probabilities estimated by models developed in clinical contexts (PKUPH, VA, Mayo) were all well above the observed rates (Table 11) thus showing poor calibration according to all tests performed.

None of the models showed acceptable calibration on nodules first observed in the incidence screening rounds (Table 12, Supplementary Table 7).

3.2.3 Nodule malignancy prediction models fitted to LUSI data

Among the three logistic regression models fitted via GEEs to LUSI data, the model including mean diameter as a measure of nodule size produced the best results (QIC=541.6 vs. 549.9 for largest diameter and 592.2 for nodule volume). With this model as starting point, backward feature elimination was performed resulting in a model including: age, years since quitting smoking, bronchitis, nodule mean diameter, nodule location, and spiculation (Supplementary Table 9). Sex, self-reported history of extra-thoracic cancer, smoking duration, emphysema (CT-based), FVC, nodule type and nodule count per scan showed no association with malignancy. The final model achieved an AUC=0.90 95% CI:[0.83, 0.93] (bootstrap AUC= 0.88 95% CI [0.84, 0.92]) for nodules detected in the prevalence round, AUC=0.81 95% CI [0.71, 0.90] (bootstrap AUC=0.81 95% CI:[0.73, 0.87]) for nodules first detected in the incidence round.

Table 11. Observed absolute risk for nodule malignancy versus predicted model estimates by nodule size (prevalence round)

Nodule size (mm)	Total nodule count	Malignant nodule count	Observed malignancy rate	Models fitted on screening data					Models fitted on data from a clinical setting		
				PanCan 1b	PanCan 2b	PanCan MD	PanCan VOL	UKLS	Mayo	PKUPH	VA
< 5	1,820	1	0.05%	0.2%	0.2%	0.02%	0.1%	1.5%	5.9%	24.5%	20.7%
5 to < 8	868	6	0.69%	1.3%	0.9%	0.4%	0.7%	1.8%	7.4%	27.6%	23.6%
8 to 10	110	7	6.36%	4.6%	3.5%	2.3%	3.1%	2.2%	10.6%	34.8%	29.7%
> 10	85	18	21.18%	18.6%	14.7%	12.9%	11.9%	6.8%	29.1%	52.3%	48.9%
Total	2,883	32	1.11%	1.2%	0.96%	0.6%	0.7%	1.8%	7.2%	26.6%	22.8%

Nodule size is defined as largest diameter in millimeters. Models were applied to the low-dose computed tomography image where nodules where first seen.

As published in (González Maldonado et al. 2020a), reprinted with permission.

Table 12. Observed absolute risk for nodule malignancy versus predicted model estimates by nodule size (incidence rounds)

Nodule size (mm)	Total nodule count	Malignant nodule count	Observed malignancy rate	Models fitted on screening data					Models fitted on data from a clinical setting		
				PanCan 1b	PanCan 2b	PanCan MD	PanCan VOL	UKLS	Mayo	PKUPH	VA
< 5	253	1	0.4%	0.3%	0.3%	0.03%	0.16%	1.1%	6.2%	25.9%	23.1%
5 to < 8	475	5	1.1%	1.4%	1.4%	0.5%	0.9%	1.7%	8.7%	35.4%	27.6%
8 to 10	103	5	4.9%	5.5%	4.9%	2.9%	3.4%	2.3%	14.6%	48.5%	34.3%
> 10	189	20	10.6%	21.7%	18.5%	15.9%	15.2%	10.3%	36.1%	68.6%	57.3%
Total	1,020	31	3.0%	5.3%	4.6%	3.5%	3.7%	3.2%	13.7%	40.5%	32.6%

Abbreviations: PanCanMD: PanCan model with mean diameter; PanCanVOL: PanCan model with volume; UKLS: United Kingdom Lung Cancer Screening trial; PKUPH: Peking University People´s Hospital; VA: Veterans Affairs

As published in (González Maldonado et al. 2020a), reprinted with permission.

3.3   Estimates of overdiagnosis and related parameters

Parts of this chapter have been published previously (González Maldonado et al. 2020b).

Data from all 4,052 participants of the LUSI trial were used for estimating the proportion of overdiagnosed lung cancer cases. All participants were long-term smokers and 2,622 of them were males (64.7%). The median age was 56.9 range:[50.3 - 71.9] years at first screening participation. Follow-up for lung cancer diagnosis was defined as the time until the first of the following events: lung cancer diagnosis, death, date of loss to follow-up or April 30th 2019. The median follow-up time for lung cancer diagnosis as endpoint was 9.77 (range:[0 – 11.5], IQR:[8.8 – 10.4], 10th percentile = 8.2) years post-randomization and 5.73 (range:[0 -11.4], IQR:[4.8 – 6.3]; 10th percentile = 4.0) years since last screening participation (Supplementary Table 10). Until the end of the follow-up period, there were 90 lung cancer cases in the LDCT arm, of which 63 were detected by LDCT screening, and 74 cases in the control arm (Figure 20). As of July 2nd 2019 (set as the end of the follow-up period) the median follow-up time for mortality as endpoint was 9.96 years IQR:[9, 10.6] (Supplementary Table 10).

3.3.1   Excess incidence

Considering all histologic subtypes, the excess cumulative incidence until April 30th 2019 was 25.4% 95% CI:-[11.3%, 64.3%], that is, 16 cases expressed as a proportion (PS) of screen-detected cases, and 17.8% 95% CI:[-7.4%, 44.7%] expressed as a proportion (PA) of all cases in the LDCT arm (Table 13, Supplementary Figure 6).

A total of 37 and 59 adenocarcinomas were diagnosed in the control and LDCT arms respectively (Supplementary Table 11). Of those in the LDCT arm, 44 were screen-detected. These values lead to estimates of PS=50.0% 95% CI:[14.0%, 88.4%] and PA=37.3% 95% CI:[11.5%, 65.4%]. To a large extent, the excess incidence of adenocarcinomas in the LDCT arm was influenced by tumors classified as bronchiolo-alveolar carcinomas (BAC). There were 10 BAC cases in the LDCT arm, 8 of which were screen-detected, and 1 in the control arm (PS = 112.5% 95% CI:[68.2%, 113.1%] and PA = 90% 95% CI:[54.3%–164.4%]).

After excluding BAC tumors, estimates of excess incidence were PS=36.1% 95% CI:[-8.4%, 84.8%] and PA=26.5% 95% CI:[-5.3, 61.8]. Regarding tumors of other (non-adenocarcinoma) histology, there were 31 cases in the LDCT (19 screen-detected) and 37 in the control arm, corresponding to estimates of PS=-31.6% 95% CI:[-130.8%, 83.0%] and PA=-19.4% 95% CI:[-76.8%, 45.6%] (Table 13, Supplementary Table 12).
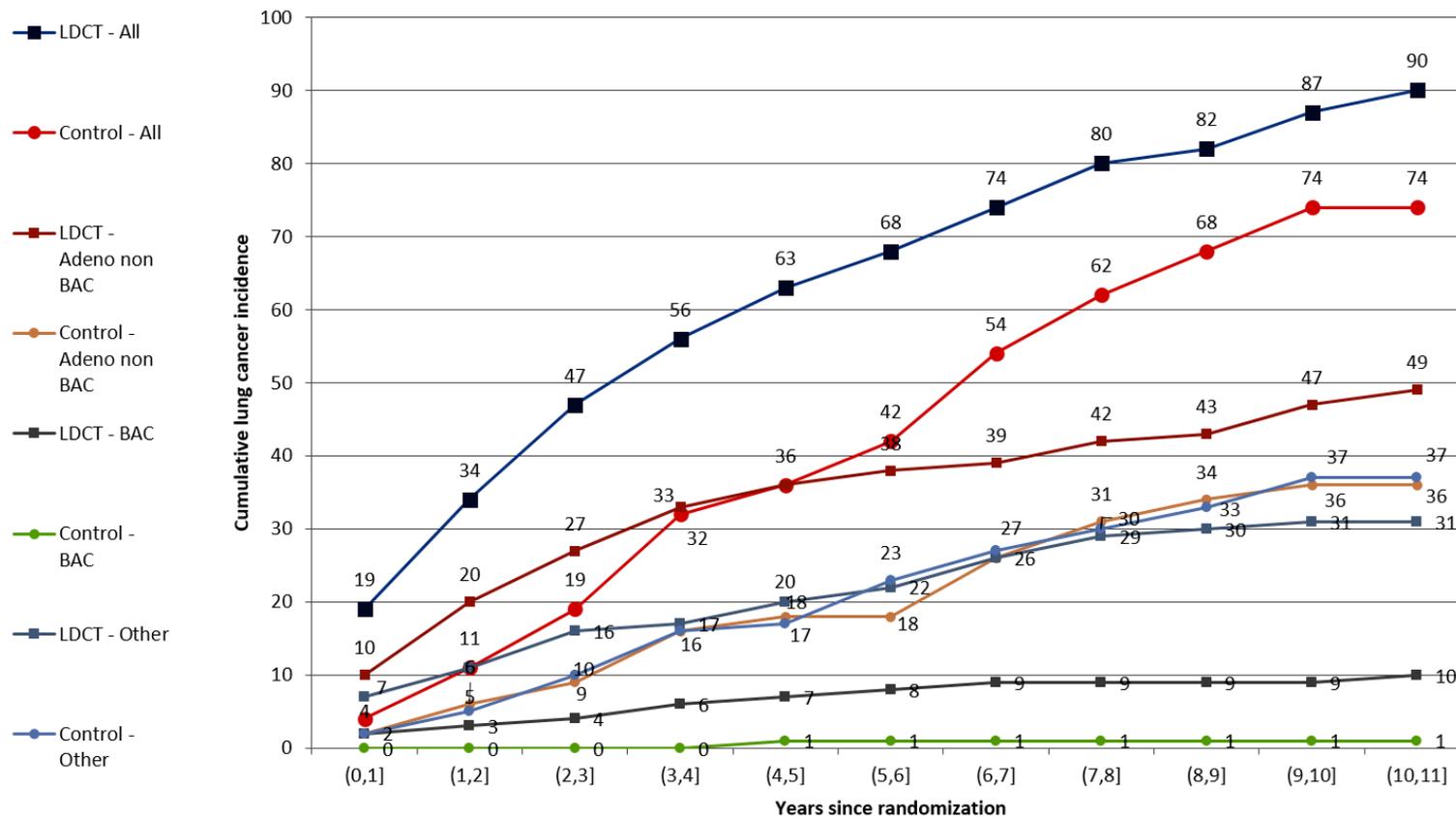
Figure 20. Cumulative lung cancer incidence by years from randomization until diagnosis, study arm and tumor histology

Abbreviations: LDCT: low-dose computed tomography, Adeno: adenocarcinoma, BAC: bronchiolo-alveolar carcinoma. After a median follow-up time of 9.77 years post-randomization there were 90 lung cancer cases in the low-dose computed tomography (LDCT) arm, of which 63 were detected by LDCT screening, whereas a total of 74 cases were observed in the control arm. A large proportion of lung cancer cases were adenocarcinomas, with 59 cases (10 of them bronchiolo-alveolar carcinomas (BAC)) observed in the LDCT arm of which 44 were screen-detected, and 37 (1 BAC) in the control arm.

As published in (González Maldonado et al. 2020b), reprinted with permission.

Table 13. Lung cancer cases by study arm and tumor histology, and excess incidence rates

	LDCT			Control	Estimated overdiagnosis rates[b] (95% CI)	
Tumor histology	Screen detected	Non-screen detected[a]	Subtotal	Subtotal	PS	PA
All	63	27	90	74	25.4 (-11.3, 64.3)	17.8 (-7.4, 44.7)
Adenocarcinoma (%)	44	15	59	37	50.0 (14.0, 88.4)	37.3 (11.5, 65.4)
BAC (%)	8	2	10	1	112.5 (68.2, 113.1)	90.0 (54.3, 164.4)
Adenocarcinoma non-BAC (%)	36	13	49	36	36.1 (-8.4, 84.8)	26.5 (-5.3, 61.8)
Other (non-adenocarcinoma) (%)	19	12	31	37	-31.6 (-130.8, 83.0)	-19.4 (-76.8, 45.6)

Abbreviations: LDCT: low-dose computed tomography; BAC: bronchiolo-alveolar carcinoma; PS: proportion of screen-detected lung cancers; PA: proportion of all lung cancers in the LDCT arm
a non-screen detected cases are all lung cancer cases diagnosed either between screening rounds or in the follow-up years after the last screening round.

As published in (González Maldonado et al. 2020b), reprinted with permission.

3.3.2 Model-based estimates of mean pre-clinical sojourn time and LDCT detection sensitivity

Based on the convolution model for clinical incidence described in section 2.3.3.2, the maximum likelihood estimate (MLE) of the mean pre-clinical sojourn time (MPST) was 5.38 years 95% CI:[4.76, 5.88] and the MLE of LDCT detection sensitivity was 81.6% 95% CI:[74.4%, 88.8%], all histologic subtypes combined. The histological subtype-specific MLEs for MPST and detection sensitivity were, respectively, 7.69 years 95% CI:[6.49, 8.77] and 69.6% 95% CI:[60.8%, 79.2%] for all adenocarcinomas, 7.69 years 95% CI:[6.49, 8.77] and 62.4% 95% CI:[53.6%, 72.8%] for adenocarcinomas excluding BAC, 8.77 years 95% CI:[6.49, 12.20] and 100% 95% CI:[92.8%, 100%] for BAC, and 2.89 years 95% CI:[2.49, 3.36] and 100% 95% CI:[94.4%, 100%] for all non-adenocarcinomas of the lung (Table 14). The model fit was good, with a χ^2-p=0.55 (Supplementary Table 13).

Additional models were fitted on subgroups defined by age at first screening (50-59 and 60+ years) in order to account for the heterogeneity in age at entry to the LUSI trial. This resulted in estimates of 5.15 years 95% CI:[4.27, 5.88] and 76% sensitivity 95% CI:[66.4%, 86.4%] in the 50–59 years group, and 5.38 years 95% CI:[4.42, 6.49], 93.6% sensitivity 95% CI:[80.8%, 100%] in the 60+ years group.

Table 14. Model-based estimates of mean pre-clinical sojourn time and LDCT detection sensitivity overall and by tumor histologic subtype

	Mean Pre-clinical Sojourn Time (years, 95% CI)	Sensitivity (%, 95% CI)
Overall	5.38 (4.76, 5.88)	81.6 (74.4, 88.8)
Adenocarcinoma	7.69 (6.49, 8.77)	69.6 (60.8, 79.2)
BAC	8.77 (6.49, 12.20)	100 (92.8, 100)
Adenocarcinoma non-BAC	7.69 (6.49, 8.77)	62.4 (53.6, 72.8)
Other (non- adenocarcinoma)	2.89 (2.49, 3.36)	100 (94.4, 100)

Abbreviations: LDCT: low-dose computed tomography; BAC: bronchiolo-alveolar carcinoma.

As published in (González Maldonado et al. 2020b), reprinted with permission.

3.3.3 Estimated proportion of tumors by lead time

Based on the estimates of MPST, it was further estimated that 47.5% 95% CI:[43.2%, 50.7%] of screen-detected tumors had a lead time ≥4 years, 32.8% 95% CI:[28.4%, 36.1%] a lead time ≥6 years, and 22.6% 95% CI:[18.6%, 25.7%] a lead time ≥8 years (Table 15), meaning that about 48%, 33%, and 23% of screen-detected tumors would have remained in a pre-clinical phase over respectively, 4, 6, and 8 further years in the absence of screening. For screening participants dying within these time periods (i.e., with remaining lifetimes shorter than the corresponding lead times), these proportions are equivalent to their probability of lung cancer overdiagnosis.

Table 15. Estimated proportions of screen-detected tumors by lead time (1-Pclin(t))

Histologic subtype	1- P_{clin} (95% CI)				
	4y	6y	8y	10y	12y
All tumors	47.5% (43.2%, 50.7%)	32.8% (28.4%, 36.1%)	22.6% (18.6%, 25.7%)	15,6% (12,2%, 18,3%)	10,7% (8,0%, 13,0%)
Adenocarcinoma	59.5% (54.0%, 63.4%)	45.8% (39.7%, 50.5%)	35.3% (29.2%, 40.2%)	27.3% (21.4%, 32.0%)	21.0% (15.8% 25.5%)
BAC	63.4% (54.0%, 72.0%)	50.5% (39.7%, 61.1%)	40.2% (29.2%, 51.9%)	32.0% (21.4%, 44.0%)	25.5% (15.8%, 37.4%)
Adenocarcinoma, non-BAC	59.5% (54.0%, 63.4%)	45.8% (39.7%, 50.5%)	35.3% (29.2%, 40.2%)	27.3% (21.4%, 32.0%)	21.0% (15.8%, 25.5%)
Other (non-adenocarcinoma)	25.1% (20.0%, 30.4%)	12.5% (9.0%, 16.7%)	6.3% (4.0%, 9.2%)	3.1% (1.8%, 5.1%)	1.6% (0.8%, 2.8%)

Abbreviations: BAC: bronchiolo-alveolar carcinoma

As published in (González Maldonado et al. 2020b), reprinted with permission.

3.4 Evaluation of the TAAb panel EarlyCDT®-Lung

Parts of this chapter have been published previously (González Maldonado et al. 2021b).

At time of blood collection, the 46 participants in the lung cancer group (32 of them males) were older (median: 63.0 years, range:[51.9, 74.5]) than those in the BC (56.8 95% CI:[50.9, 69.7], p<0.001) and those in the SNC (55.8 95% CI:[50.6, 70.0], p<0.001) (Table 16).

Lung cancer was detected for 19 (41.3%) participants at the prevalence screening round. For one of the participants, lung cancer was detected at the second screening round even in the absence of pulmonary nodules, due to the identification of atelectasis in the scan images. All tumors detected in the first round were deemed suspicious based on their size. A total of 21 (80.8%) of the remaining 26 detections in the incidence rounds were done in nodules already observed in previous rounds; with 7 of these immediate recall decisions based on nodule volume doubling time (VDT).

3.4.1 Lung cancer detection sensitivity

The EarlyCDT®-Lung test produced "High Level" (H) test results for 6 out of the 46 participants with LDCT-detected lung cancer and no "Moderate Level" (M) test results (Table 16). This resulted in an estimated detection sensitivity of 13.0% 95% CI:[4.9%, 26.3%].

Among participants with nodules <10 mm in diameter, the test produced H results for 1 out of 11 LDCT-detected lung cancer patients, for an estimated sensitivity of 9.1% 95% CI:[0.23%, 41.3%]. For participants with nodules ≥10mm, the estimated sensitivity was 14.7% 95% CI:[4.9%, 31.1%].

EarlyCDT®-Lung test results were positive for 5 of 6 tumors in stages IB or higher (sensitivity of 22.7%), and for 1 out of 24 stage IA tumors (sensitivity of 4.2%). An H test result was associated with a significant shift in tumor stage distribution (p=0.03) towards higher stages (83.3% in stages IB and above), compared to patients with negative test results (NS) who were predominantly (57.5%) stage IA (Table 17). No significant association of test results with malignant nodule size or histology was observed (Table 17).

3.4.2 Specificity and positive likelihood ratio

Considering both H and M results as positive, the false-positive detection rates were 11.1% in the BC group (specificity of 88.9% 95% CI:[80.5% - 94.5%]) and 8.9% (specificity of 91.1% 95% CI:[83.2%, 96.1%]) in the SNC group. Based on these estimates, the LR+ was 1.17 95% CI:[0.46, 3.03] in the BC group and at 1.47 95% CI:[0.54, 3.98] in the SNC group.

If only H results were considered positive, the false positive detection rate was 3.3% (specificity of 96.7% 95% CI:[90.6%, 99.3%]) in the BC group and 4.4% (specificity of 95.6% [89.0%, 98.8%]) in the SNC group. yielding estimates of LR+ of 3.91 95% CI:[1.03, 14.94] and 2.93 95% CI:[0.87, 9.88] respectively.

3.4.3 Association between positive EarlyCDT®-Lung test results and the presence of lung tumors

Logistic regression provided insufficient evidence to claim an increased risk of malignancy associated with positive test results (H and M) (Odds Ratio (OR): 1.20 95% CI:[0.41, 3.54], p=0.74 in the BC group, OR: 1.54 95% CI:[0.5, 4.73], p=0.45 in the SNC group). Considering only H results as positive, the same association was moderate, but significant in the BC group (OR: 4.35 95% CI:[1.04, 18.28], p=0.04), whereas not enough evidence was found in the SNC group (OR: 3.22 95% CI:[0.86, 12.07], p=0.08).

Analyses comparing data of lung cancer patients with nodules on their CT-Scans (N=45) (Table 17) to those of controls in the SNC group, showed no evidence of association between test results and malignancy, with ORs of 1.58 95% CI:[0.51, 4.86] and 3.31 95% CI:[0.88, 12.39] depending on the definition of positive test result (Supplementary Table 15). Similarly, there was not enough evidence of association when stratifying by nodule size (Supplementary Table 15).

Table 16. Subject characteristics and LDCT result by EarlyCDT®-Lung results for lung cancer cases and cancer-free controls.

Lung Cancer Status		Lung cancer			No lung cancer (BC)			p	No lung cancer (SNC)			p
N		46			90				90			
Subject characteristics	Gender (Male %)	32 (69.6)			56 (62.2)			0.51	64 (71.1)			1
	Age† (years, median [range])	63.0 [51.9, 74.5]			56.8 [50.9, 69.7]			0.001	55.8 [50.6, 70.0]			<0.001
	Smoking status‡ (Current/ Former, %)	24/22 (52.2/47.8)			54/36 (60.0/40.0)			0.49	54/36 (60.0/40)			0.49
EarlyCDT®-Lung Test Result		HIGH	NS	p	HIGH	MOD	NS	p	HIGH	MOD	NS	p
N		6	40		3	7	80		4	4	82	
LDCT result (%)	No nodules	0	0	1	3 (100.0)	1 (14.3)	48 (60.0)	0.15	0	0	0	0.11
	Non-suspicious	0	0		0	4 (57.1)	22 (27.5)		0	0	0	
	immediate recall	6 (100)	36 § (90.0)		0	0	2 (2.5)		0	1 (25.0)	4 (4.9)	
	3-month recall	0	0		0	0	1 (1.2)		2 (50.0)	0	10 (12.2)	
	6-month recall	0	4 (10.0)		0	2 (28.6)	7 (8.8)		2 (50.0)	3 (75.0)	68 (82.9)	

Abbreviations: BC: baseline control group; SNC: suspicious nodules control group; MOD: moderate level test result; NS: non-significant level test result
† At blood draw
‡ At randomization. Smokers: at least 25 years smoking of at least 15 cigarettes per day, or at least 30 years smoking of at least 10 cigarettes per day, including former smokers. Former smokers are those who had stopped smoking not more than 10 years before invitation to screening
§ For one subject, the CT scan evaluation at round 2 was deemed suspicious (with immediate recall) even in the absence of pulmonary nodules, due to the identification of atelectasis (collapsed lung) in the scan images.

As published in (González Maldonado et al. 2021b), reprinted with permission.

Table 17. Tumor characteristics by EarlyCDT®-Lung result in the lung cancer group.

EarlyCDT®-Lung Test Result			HIGH	NS	P	Total n (col %)
Detection / Diagnosis	**Round of detection (%)**	1st (Prevalence)	**3 (50.0)**	**16 (40.0)**	**0.67**	**19 (41.3)**
		2nd – 5th (Incidence)	**3 (50.0)**	**24 (60.0)**		**27 (58.7)**
	Time detection to diagnosis (months, median [range])		**2.28 [1.4, 7.8]**	**3.12 [0.5, 49.8]**	**0.35**	**3.00 [0.48, 49.8]**
Tumor characteristics	**Histology (%)**	Adenocarcinoma	5 (83.3)	27 (67.5)	1	32 (69.6)
		Carcinoid	0	2 (5.0)		2 (4.3)
		Large cell	0	1 (2.5)		1 (2.2)
		Small cell	0	3 (7.5)		3 (6.5)
		Squamous cell	1 (16.7)	7 (17.5)		8 (17.4)
	Stage (%)	IA	1 (16.7)	23 (57.5)	0.03	24 (52.2)
		IB	3 (50.0)	5 (12.5)		8 (17.4)
		IIA	1 (16.7)	2 (5.0)		3 (6.5)
		IIB	1 (16.7)	1 (2.5)		2 (4.3)
		IIIA	0	6 (15.0)		6 (13.0)
		IIIB	0	0		0
		IV	0	3 (7.5)		3 (6.5)
	Largest diameter (mm, median [range])		15.65 [8.7, 20.9]	12.00 [5.5, 54.1]	0.31	12.00 [5.5, 54.1]
	Largest diameter (mm, %)	No nodules§	0	1 (2.5)		1 (2.2)
		5 to <8	0	4 (10.0)	0.43	4 (8.7)
		8 to <10	1 (16.7)	6 (15.0)		7 (15.2)
		10 to <20	3 (50.0)	25 (62.5)		28 (60.9)
		>20	2 (33.3)	4 (10.0)		6 (13.0)

Abbreviations: NS: non-significant. § For one subject, the CT scan evaluation at round 2 was deemed suspicious (with immediate recall) even in the absence of pulmonary nodules, due to the identification of atelectasis (collapsed lung) in the scan images.

As published in (González Maldonado et al. 2021b), reprinted with permission.

4 DISCUSSION

The overarching goal of this thesis was to study four relevant aspects for the optimization of LDCT-based lung cancer screening programs, namely: the personalized risk-based assignment of screening intervals, the use of statistical models for improving malignancy prediction of screen-detected nodules, the estimation of overdiagnosis risk and its relationship to upper-age limits for screening participation, and the use of biomarkers as complementary tools for screening.

More concretely, data and blood samples collected as part of the German LUSI trial were used in order to:

1. Evaluate the discrimination and absolute risk calibration of two models (LCRAT+CT and Polynomial) that predict short-term lung cancer risk based on subject characteristics and LDCT imaging findings, and which have been proposed as tools for the identification of candidates for biennial vs annual screening.

2. Evaluate the discrimination and absolute risk calibration of models fitted to data from screening (PanCan and UKLS models) or clinical contexts (Mayo, VA, PKUPH), that predict the malignancy of screen-detected pulmonary nodules based on their radiological features and subject characteristics.

3. Estimate the extent of overdiagnosis in the LUSI trial via the observed excess incidence, and obtain estimates of mean pre-clinical sojourn time (MPST) of tumors and LDCT detection sensitivity, via mathematical modeling. Additionally, to study the effect of tumor MPST and remaining life expectancy (and therefore age) of screening participants on overdiagnosis risk.

4. Evaluate the TAAb panel measured by the EarlyCDT®-Lung test, in terms of its detection sensitivity and specificity on blood samples from screening participants with tumors observed in their LDCT screening images.

The main results of these sub-studies can be summarized as follows:

1. The LCRAT+CT and Polynomial models showed a good ability to discriminate participants diagnosed with lung cancer at the time of their next screening appointment or in the year after (AUC=0.73 and 0.75 respectively). The discrimination performance of these models was superior to that of the Patz criterion (based solely on nodule size; AUC=0.56) and the LCRAT model (based on subject characteristics; AUC= 0.68). Assignment of participants with the lowest risks (as predicted by either model) to biennial screening, would have caused few delayed diagnoses (e.g., LCRAT+CT: 10% delays; Polynomial: 0 delays, risk

< 5th decile). Absolute risk estimates from the two models were, on average, lower than the observed rates, indicating poor calibration.

2. The four selected PanCan models showed excellent discrimination of malignant nodules identified in the first (prevalence) screening round (AUC≥0.93 for all models). On the same data, the UKLS model showed poor discrimination (AUC=0.58), and the Mayo, VA and PKUPH models only moderate discrimination (AUC 0.84-0.89). The calibration of absolute malignancy probabilities for nodules detected on the prevalence screen was only acceptable for the PanCan-1b, PanCan-2b and Pan-Can-VOL models. None of the models showed acceptable calibration when applied to nodules first detected in later screening rounds.

3. The excess incidence observed in the LUSI trial, after 5.73 years post-screening follow-up, represented 24.5% of all screen-detected lung cancer cases. The highest excess incidence was observed for adenocarcinomas (50.0%) and, in particular, for bronchiolo-alveolar carcinomas (BAC) (112.5%). Overall, model-based estimates of mean pre-clinical sojourn time (MPST) and detection sensitivity of LDCT were 5.38 years and 81.6% respectively. These estimates also varied across histologic subtypes, with the longest MPST obtained for BAC tumors (8.77 years), compared to non-adenocarcinomas (2.89 years). Based on the overall MPST, it was further estimated that 47.5%, 32.8% and 22.6% of all screen detected tumors had lead times longer or equal to 4, 6 and 8 years respectively; proportions equivalent to the risk of overdiagnosis for patients with remaining life expectancies shorter than the respective lead times.

4. Positive EarlyCDT®-Lung test results were significantly associated with a shift towards advanced tumor stages (83.3% positives in stages ≥ IB; 57.5% negatives in stage IA; p=0.03). Specificity ranged from 88.9% in baseline blood samples to 91.1% in samples from participants with suspicious imaging findings. However, the TAAb panel showed an overall detection sensitivity of only 13.0% and of 9.1% among participants with small nodules (<10 mm in diameter) detected via LDCT.

A discussion on each of the sub-studies that comprise this thesis is presented in the following sections.

4.1 Validation of prediction models for the assignment of screening intervals

Parts of this chapter have been published previously (González Maldonado et al. 2021a).

4.1.1 Discrimination of participants with short-term lung cancer diagnoses

The good discrimination shown by the LCRAT and Polynomial models provides confirmatory evidence in favor of its application for the assignment of participants to biennial vs annual screening based on lung cancer risk predicted using the findings of the first screening examination.

On secondary analyses, the Polynomial model, though developed for its application on data from the prevalence screening round, also achieved good discrimination when evaluated on data from later (incidence) screening rounds (AUC=0.74, all incidence rounds combined and AUC=0.72 on individuals without any nodules ≥4mm, all rounds combined). Its good transferability suggests that the model could also support participant allocation to biennial screening further into the screening process.

Both models showed better discrimination than that of the Patz criterion and the LCRAT model. The minor difference in AUC observed between the LCRAT and the LCRAT+CT, suggests that a good amount of discrimination performance can be attributed to predictors related to risk factors, such as smoking. Although this difference in discrimination did not reach statistical significance, and the LCRAT+CT and Polynomial models were only compared to the criterion by Patz et al. and not to other models or criteria, these findings suggest that features from LDCT images may further improve model performance. Similar observations were made by Tammemägi et al. (Tammemägi et al. 2019b), who reported improved discrimination from a model adding screening results (PLCOm2012results) to the already well-established PLCOm2012 model for lung cancer risk prediction. Our observations also align with those of Schreuder et al. who reported superior discrimination from his Polynomial model compared to the criterion by Patz (AUC=0.79 vs 0.67) and to other models based solely on subject characteristics (Schreuder et al. 2018).

4.1.2 Risk-based assignment to biennial vs annual screening and its effects on delayed diagnosis

According to Lorenz curves based on predictions from the Polynomial and LCRAT+CT models (Figure 16, Figure 17), in populations similar to that of the LUSI trial, individuals who would have their diagnosis delayed as a result of biennial screening are likely to be those with the highest lung cancer risks, whereas a low proportion of delayed diagnosis is expected among

individuals with the lowest predicted risks. As an example, only 10% of all delayed diagnoses would have occurred among the 50% of participants with lowest risks estimated by the LCRAT+CT model (Figure 16). These results are in line with those reported by Robbins et al. (Robbins et al. 2019) who observed 15.2% delayed diagnoses among 40.9% of individuals with lowest risks predicted by the LCRAT+CT model. Similarly, 20% of all delayed diagnoses would have occurred among 60% of participants with the lowest risks, and 80% of delayed diagnoses among the remaining 40% with higher risks as estimated by the Polynomial model (Figure 17).

In contrast, if the second annual screening would have been skipped by participants selected according to the Patz criterion, about 54.5% of them would have had their lung cancer diagnoses delayed. Thus, assignment to biennial screening based solely on nodule size, applied to data from the LUSI trial, did not reach the levels of discrimination required to recommend alternative screening schedules. Similarly, Lorenz curves based on risk estimates from the LCRAT model showed higher numbers of delayed lung cancer detections compared to LCRAT+CT, at equal proportions of annual screenings skipped (Figure 16). It is, however, unclear whether the differences in delayed detections according to LCRAT vs LCRAT+CT are statistically significant, given the small sample size used for these analyses.

A limitation of this sub-study is its retrospective nature, which prevented an estimation of the actual harms or benefits of skipping a screening appointment (e.g., false-positive test results that might be permanently avoided or just postponed).

Taken together, these findings suggest that risk models that combine subject characteristics with LDCT imaging findings are useful tools for the identification of participants for whom biennial screening may be economically more efficient at only minor loss of life years gained (LYG), and potentially may also improve the trade-off between expected screening benefit (LYG) and the harms of overdiagnosis and/or false positive test results. In summary, models with similar performance to the ones validated in these analyses could identify those for whom biennial screening could provide a higher net clinical benefit and/or higher cost-efficiency compared to more frequent (annual) screening.

4.1.3 Calibration of absolute lung cancer risk

Absolute lung cancer risk estimates from the LCRAT+CT and Polynomial models were, on average, lower than the observed rates of lung cancer, indicating that these prediction models might need to be re-calibrated for their use on populations differing from that of the NLST, in which they were developed.

Additionally, predicted risks differed between the two models, suggesting the need for model-specific risk thresholds on the basis of which participants could be assigned to biennial

screening. In fact, when applied on data from eligible participants of the LUSI trial, approximately 10% delayed diagnoses were calculated when using a 0.20% lung cancer risk threshold for the Polynomial model, while the same proportion was reached at a 0.13% threshold for LCRAT+CT. However, it is important to mention that these two models differ on the risks they intend to estimate, the predictors they include, and the sub-populations of screening participants on which their application is intended. Due to these intrinsic differences, absolute risk estimates from the two models are not directly comparable.

Similarly, absolute risk thresholds calculated in data from the NLST, as reported by Robbins et al. (Robbins et al. 2019), differed from those calculated on data from the LUSI trial. This disagreement can be explained by differences in the distribution of lung cancer risks in the two populations. This in turn can be attributed in part to the less stringent eligibility criteria of the LUSI trial (Table 1) which included a larger proportion of participants at lower risks compared to the NLST.

A more refined assessment of absolute risk calibration was not possible due to the small sample size and low case numbers in the data from the LUSI trial. Additionally, a limitation of this calibration analysis was the missing data for a subset of predictors such as race and number of parents with lung cancer, which may have caused bias in the estimated risks.

4.2 Validation of nodule malignancy prediction models

Parts of this chapter have been published previously (González Maldonado et al. 2020a).

4.2.1 Discrimination of malignant nodules

The four selected PanCan models (PanCan-1b, PanCan-2b, PanCan-MD and PanCan-VOL) showed good performance for the discrimination of malignant pulmonary nodules observed at the first (prevalence) screening round. Compared to the initial versions, the more recent PanCan-VOL and PanCan-MD models showed good but no superior discrimination. By contrast, applied to data from the same screening round, the UKLS model showed poor discrimination.

The lack of improvement in discrimination shown by PanCan-VOL and PanCan-MD compared to the four original variants of the PanCan models confirms the results of evaluations on data from the PanCan study (Tammemägi et al. 2019a), but disagree with evaluations made on data from the NELSON trial (Horeweg et al. 2014b). These differences might be attributed to the nodule size estimation methods applied in each trial. In the LUSI trial, diameters were calculated via software, based on a 3-dimensional estimation, which possibly led to higher accuracy compared to manual measurements taken in other trials (Han et al. 2018).

In data from the LUSI trial, PanCan-1b showed very similar estimates of sensitivity, specificity and PPV at selected (2%, 5% or 10%) probability thresholds compared to those from the data in which the PanCan models were originally developed (McWilliams et al. 2013).

Compared to the values reported by the study in which they were first published (Tammemägi et al. 2019a), PanCan-VOL and PanCan-MD showed higher sensitivity and lower specificity, but similar PPV. In contrast, when validated on data from the NLST (White et al. 2017), the PanCan-2b model showed higher sensitivity, lower specificity and similar PPV compared to the estimates obtained in this external validation.

The selected models developed on data from clinical contexts (VA, Mayo, PKUPH) achieved only moderate discrimination (AUC 0.84-0.89) (Section 0). The performance and ranking of the PanCan, Mayo, VA and PKUPH models, in terms of discrimination, are in line with the results from validation studies on clinical data in context of incidentally or symptomatically detected pulmonary nodules, which have shown AUCs above 0.80 for these models (Al-Ameri et al. 2015; Chung et al. 2018; Deppen et al. 2014; Isbell et al. 2011; Li et al. 2011; Massion et al. 2017; Schultz et al. 2008; Xiao et al. 2013) or in context of LDCT screening (Marcus et al. 2019; Marshall et al. 2017; Nair et al. 2018; White et al. 2017; Winkler Wille et al. 2015; Zhao et al. 2016).

Finally, both the PanCan and UKLS models showed decreased discrimination performance when applied to nodules first observed in later (incidence) rounds.

Overall, these results suggest that, while the PanCan models can be recommended as tools for the identification of malignant nodules identified at the first screening round, none of the models here evaluated can be recommended for the identification of malignant nodules beyond the prevalence round.

4.2.2 Calibration of absolute nodule malignancy risks

Applied to data from the first (prevalence) screen of the LUSI trial, PanCan-1b, PanCan-2b, and PanCan-VOL but neither PanCan-MD nor UKLS, showed acceptable calibration of absolute malignancy probabilities.

The acceptable calibration observed from the PanCan-1b, PanCan-2b and PanCan-VOL, applied on prevalence screen data, is similar to that seen in evaluations carried out on data from a chemoprevention trial by the British Columbia Cancer Agency (BCCA)(McWilliams et al. 2013). The same study could only confirm such results for the PanCan-1b and PanCan-2b when applying them to NLST data (Nair et al. 2018; White et al. 2017). Independently, the latter two PanCan variants (1b, 2b) showed acceptable calibration in data from an Australian LDCT pilot screening trial (Marshall et al. 2017; Zhao et al. 2016), while an evaluation of data

from the NLST reported unsatisfactory calibration for the PanCan-MD and PanCan-VOL models (Tammemägi et al. 2019a).

None of the models showed acceptable calibration when applied to nodules first observed in any of the incidence rounds of the LUSI trial. These results make sense since, given that the PanCan and UKLS models were trained on data from prevalence screening rounds, the calibration of their estimated risks when applied on data from later rounds is not guaranteed (Tammemägi et al. 2019a). In the NLST, for example, the malignancy risk of newly seen nodules (≥4-6 or ≥6-8 mm) was higher than that of nodules found at the prevalence screening (Pinsky et al. 2017).

The VA, Mayo and PKUPH models (all developed in clinical settings) showed poor calibration when evaluated on data from the LUSI trial, which resembles the results of validation studies of the VA and Mayo models carried out using data from the NLST (Nair et al. 2018).

Thus, given the results previously discussed, only the PanCan models can be recommended for the prediction of absolute malignancy risk for nodules observed on the first screening round. However, none of the models here evaluated should be considered for malignancy risk estimation of nodules detected on later screening rounds.

4.2.3 Predictors of nodule malignancy in data from the LUSI and other trials

In multivariable models fitted to data from the LUSI trial, sex showed no predictive value and, similarly, age showed only borderline predictive ability. In a parsimonious model obtained via feature selection, and in the presence of LDCT features, age, bronchitis (despite their borderline significance), and time since cessation were kept as predictors.

Depending on whether other studies used training data from clinical settings or screening trials, predictive variables varied across models. Surprisingly, age, a well-known risk factor for lung cancer, was not included in most of the PanCan model variants (PanCan-1b, PanCan-MD, PanCan-VOL). However, age was a predictor in the UKLS model and, as mentioned in the previous paragraph, was also retained in the multivariable model trained in data from the LUSI trial. A similarity between these models is that they included indicators of smoking behavior. These findings resemble those of the analyses leading to the UKLS model (Marcus et al. 2019) showing that, in multivariable models including age and LDCT imaging features, additional smoking-related information (e.g., smoking duration, time since cessation) may improve malignancy prediction, even among screening eligible subjects.

The association between sex and malignancy risk is also heterogeneous across trials. Analysis on data from the PanCan and UKLS trials (those leading to the development of the PanCan and UKLS models) reported higher malignancy rates among females, while, on the contrary, lower malignancy risks were observed in data from female participants of the DLCST,

compared to males. This heterogeneity could be attributed to differences in lung cancer incidence between females and males across populations and age groups, which in turn might be the effect of sex-related differences in smoking habits and shifts in smoking trends over time (O'Keeffe et al. 2018). Another important factor to consider when incorporating sex in prediction models for tumor malignancy is the difference in distribution of histologic subtypes between males and females. Lung tumors found in females are more often adenocarcinomas, i.e., slow-growing tumors with longer lead-times, which are therefore more likely to be detected by means of screening at early developmental stages (Becker et al. 2019; International Association for the Study of Lung Cancer 2019; Pinsky et al. 2013).

Furthermore, the inclusion of predictors difficult to standardize across populations (definition and diagnosis of lung diseases, nodule type categories), may hinder model transferability. In fact, a reduced UKLS model (Table 5) ignoring variables with definitions markedly differing from those in LUSI, showed a better performance (e.g., AUC=0.79 95% CI:[0.68, 0.89] in the prevalence round) compared to the original version.

In summary, appropriate variable selection is crucial for the comparability, transferability and interpretability of prediction models. Thus, care should be taken when fitting models to data from different contexts and populations.

Due to the small sample size of the LUSI trial and the low numbers of malignant nodules observed, the calibration (and similarly, the discrimination) of the selected models for nodule malignancy prediction could not be evaluated within strata defined by histologic subtypes and nodule sizes. Another limitation was related to missing variables or variables with incomplete information (i.e., family history of lung or extra-thoracic cancers), which might have led to biased risk estimates from the PanCan-2b, UKLS and VA models.

4.3 Estimates of the extent of overdiagnosis and related parameters

Parts of this chapter have been published previously (González Maldonado et al. 2020b).

4.3.1 Excess incidence in the LUSI and other randomized screening trials

After a median post-screening follow-up of 5.73 years (range:[0, 11.4], IQR:[4.8, 6.3]), there was 24.5% (95% CI:[-11.3%, 64.3%]) excess incidence of screen-detected lung cancer cases.

The excess incidence observed in data from the LUSI trial was slightly higher than the values observed in the NELSON trial (19.7% 95% CI:[–5.2%, 41.6%])(de Koning et al. 2020) and in the NLST trial (18.5% 95% CI:[5.4%, 30.6%]) (Patz et al. 2014), both at an average follow-up of about 4.5 years after last screening. In contrast, after a 5-year follow-up period post-screening, the Italian ITALUNG study (Paci et al. 2017) reported no excess incidence, but

instead four <u>fewer</u> cases in the LDCT arm compared to the control arm; whereas also at 5 years after screening cessation, the DLCST (Heleno et al. 2018) reported an excess of 67.2% 95% CI:[37.1%, 95.4%] of screen-detected cases. Heterogeneity in excess incidence estimates across trials may be related to various factors. Among these are: differences in the duration of post-screening follow-up, in the rates of death from competing causes, and differences in screening frequency protocols (1, 2 and 2.5 years in the NELSON trial, versus annual screening in the other studies) (Supplementary Table 16).

The excess incidence observed in screening trials is mostly an effect of longer lead times as a result of earlier tumor detection, and does not necessarily reflect overdiagnosis. Overdiagnosis exclusively results when the lead time exceeds the remaining lifetime of participants with screen-detected tumors. As follow-up time in randomized stop-screening trials increases, the gap in cumulative incidence between screening and control arm becomes progressively narrower due to slowly-growing tumors gradually becoming manifest in the control arm. If the post-screening follow-up period does not include even the longest detection lead times, excess incidence will overestimate the amount of overdiagnosis. In data from the LUSI trial, although the median follow-up time after last screening participation was 5.73 years, 25% of participants had follow-up times below 4.8 years. Given that not all participants of the LUSI trial had follow-up times long enough compared to the estimated mean pre-clinical sojourn time of lung tumors, the reported excess incidence rates might be inaccurate estimates of overdiagnosis. This might have also been the case for other trials such as ITALUNG, DLCST and NELSON. The point is well-illustrated by reports from the NLST trial, in which an 18.5% excess incidence (LDCT vs CXR) was observed after a follow-up of 4.5 years post-screening (Patz et al. 2014), but only 3% was observed after about 9.3 years of extended follow-up (National Lung Screening Trial Research Team 2019). It is also important to note that, while excess incidence calculated on data from randomized screening trials may provide an overall estimate for the screened populations, these estimates might not apply for individuals at highest risk of being overdiagnosed, e.g., those with short residual life expectancies.

Moreover, as an estimator of overdiagnosis, excess incidence is at considerable risk of bias. First, bias may occur in the absence of a control arm without intervention. In the NLST, for example, in which all control arm participants were screened by (CXR), excess incidence in the LDCT arm may underestimate the true magnitude of overdiagnosis and is therefore not comparable with estimates from the European trials with control arms without screening and low reported contamination rates (Supplementary Table 16). Another potential source of bias is confounding, e.g., due to imbalances in pre-screening lung cancer risk between the trial arms as a result of imperfect randomization, or due to differences in post-randomization factors such as participation in smoking cessation programs. For the DLCST, the investigators reported imperfect randomization, resulting in significantly more participants with higher pack-

years of smoking and a higher proportion of participants with decreased lung function in the screening arm compared to the control arm, which may have increased excess lung cancer incidence in the screening arm independently of LDCT detection (Wille et al. 2016). In the ITALUNG trial, participants in the screening arm were reported to have higher smoking cessation rates, and lower rates of relapse into smoking among ex-smokers, as compared to participants in the control arm (Pistelli et al. 2019), which may have decreased the observed excess lung cancer incidence in the screening arm.

Furthermore, evidence suggests that excess incidence varies across tumor histologic subtypes. This was true in the LUSI trial, in which excess incidence was heterogeneous across histologic subtypes, with high excess incidence for all adenocarcinomas (50.0% [14.0%, 88.4%]), and separately for non-BAC adenocarcinomas (36.1% [-8.4%, 84.8%]) and BAC tumors (112.5% [68.2%, 113.1%]). This is in line with the results of analyses in data from the NLST with estimates of 22.5% 95% CI:[9.7%, 34.3%] for NSCLC, 78.9% 95% CI:[62.2%, 93.5%] for BAC, and 11.7 [-3.7%, 25.6%] for NSCLC excluding BAC.

Evidence suggests that overdiagnosis might also depend on sex. In data from the LUSI trial, the observed excess incidence of adenocarcinomas was much larger for women than for men (Supplementary Table 11). As mentioned in section 4.2, it has been previously reported (Becker et al. 2019) that the distribution of histologic subtypes in participants of the LUSI trial differed significantly between men and women, with women showing a higher proportion of adenocarcinomas, and a much smaller percentage of small cell tumors. Furthermore, there was an enrichment of adenocarcinomas in the LDCT screening arm as compared to the control arm, and it was more pronounced among women than among men. These observations suggest that women are at increased risk of overdiagnosis. The higher excess incidence of adenocarcinomas for women compared with men might be explained by the findings of independent analyses on data from the NLST and PLCO [Prostate, Lung, Colorectal Ovarian cancer screening) trials (Ten Haaf et al. 2015) that reported a longer mean preclinical sojourn time (MPST) for this tumor subtype among women. The observation that women are at higher risk of overdiagnosis is further supported by quantitative modeling analyses performed in context of the Cancer Intervention and Surveillance Modeling Network (CISNET) (Han et al. 2017). Additionally, these findings confirm of analyses on data the NELSON and NLST trials (de Koning et al. 2020; Pinsky et al. 2013) which suggest that compared to men, women might receive a greater lung cancer mortality reduction from LDCT screening.

4.3.2 Estimates of mean pre-clinical sojourn time and LDCT detection sensitivity

Further to the calculation of excess incidence, the application of statistical modeling led to estimates of 5.38 years 95% CI:[4.76, 5.88] for MPST, and of 81.6% 95% CI:[74.4%, 88.8%] for LDCT detection sensitivity, all histologic subtypes combined.

Longer estimates of MPST were obtained for BAC (8.77 years 95% CI:[6.49, 12.20], combined with 100% 95% CI:[92.8%, 100%] sensitivity) and non-BAC adenocarcinomas (7.69 years 95% CI:[6.49, 8.77], 62.4% 95% CI:[53.6%, 72.8%] sensitivity) and, and shorter ones (2.89 years 95% CI:[2.49, 3.36], 100% 95% CI:[94.4%, 100%] sensitivity) for all other subtypes combined. These estimates, though based on smaller numbers of cancer cases, resemble those obtained based on NLST data (Patz et al. 2014), namely: MPST of 32.1 95% CI:[17.3 – 270.7] years for BAC at 38% 95% CI:[7% - 62%] sensitivity, and 3.6 95% CI:[3.00-4.3] years at 83% 95% CI:[72%-94%] sensitivity for all other NSCLC tumors.

The very long estimated MPST for BAC may be particularly noteworthy. Bronchiolo-alveolar carcinoma (BAC), a term previously used for tumors characterized by a lepidic growth pattern, was recently reclassified as a form of lung adenocarcinoma (Gardiner et al. 2014). BAC tumors are characterized by slow growth, more frequently found among affect never-smokers, women, and young adults (Boffetta et al. 2011), and are associated with better survival outcomes compared to other NSCLC subtypes. Based on distinct radiologic appearances (Ten Haaf et al. 2018), identifying patients with adenocarcinomas *in situ* (AIS) or minimally invasive adenocarcinomas (MIA) (both previously included in the BAC subtype) and assigning them to less intensive treatment or just active surveillance, could reduce the burden caused by the early detection of these slowly growing tumors.

4.3.3 Estimated proportion of tumors by lead time

Predictions obtained on the basis of the estimated MPST, suggested high proportions of screen-detected tumors with relatively long lead times (approximately 48%, 33% and 23% with lead times longer or equal to 4, 6 and 8 years respectively). These proportions suggest than individuals with comparatively short remaining life expectancies are at increased risk of overdiagnosis. For comparison, proportions predicted on the basis of MPST estimates derived from NLST data (Patz et al. 2014) were somewhat lower (Supplementary Table 14).

Long-term smokers, a subset of which is eligible for screening, may have relatively low residual life expectancies at ages 75 or older. For example, for current or former smokers in the USA, men and women, age 75-79, remaining lifetime has been estimated at around 7.5 years and 10.5 years, respectively (Østbye and Taylor 2004). Actually, in the subset of current smokers

ages or older, shorter remaining lifetime estimates have been calculated. Thus, according to the estimates here reported, on average, 10-20% of screen-detected tumors could be overdiagnosed in this population.

The main strength of this sub-study is the application of mathematical modeling for the joint estimation of MPST and detection sensitivity of LDCT-based screening, which goes beyond the mere calculation of excess incidence. Moreover, the model used for the parameter estimation was chosen to best fit the characteristics of the LUSI trial (Straatman et al. 1997).

Finally, even though the model-based estimates reported here are subject to significant uncertainty, caused by the small sample size of the LUSI trial, they illustrate the dependency of overdiagnosis risk on factors that affect the residual life expectancy of screening participants, such as age at screen detection and other health indicators.

Taken together, the results of this modeling exercise confirm findings from previous studies regarding the increased risk of overdiagnosis for LDCT screening participants with reduced remaining life expectancies. More concretely, they confirm recently published results (Han et al. 2017; Ten Haaf et al. 2020) suggesting that the balance between life years gained and the risk of overdiagnosis may start becoming less favorable for participants older than 75 years.

4.4 Evaluation of the TAAb panel EarlyCDT®-Lung

Parts of this chapter have been published previously (González Maldonado et al. 2021b).

EarlyCDT®-Lung has been repeatedly evaluated as a test for deciding whether a biopsy or surgical intervention is necessary for subjects presenting with incidentally observed pulmonary nodules in clinical settings (Healey et al. 2017; Jett et al. 2014; Massion et al. 2017). However, there is little evidence about its performance in the context of population screening. The first trial to include the EarlyCDT®-Lung in its protocol was the Early detection of Cancer of the Lung Scotland (ECLS). In the ECLS, EarlyCDT®-Lung was evaluated for its ability to identify subjects likely to harbor a lung tumor, who were then further examined by LDCT (Sullivan and Schembri 2019; Sullivan et al. 2017; Sullivan et al. 2021).

Measured in samples collected in context of the LUSI trial, there was a significant association between positive EarlyCDT®-Lung test results and case-control status, and with a shift towards advanced tumor stages. Although this confirmed, to a certain degree, its validity as a detection test, the estimated sensitivity of this TAAb panel (13.0% 95% CI:[4.9% - 26.3%]) was too low for it to be considered of practical diagnostic use in the context of screening.

Similarly, results from the ECLS showed a statistically significant shift towards earlier stages for tumors identified by the TAAb panel (and later confirmed by CXR and LDCT). However, the estimated sensitivity was low, with a positive EarlyCDT®-Lung signal found for only 12 out of

23 cancer patients in stages I or II (sensitivity of 52.2%, [95% CI 30.6% to 73.2%]) and for 6 out of 33 patients with Stage-II-IV tumors (sensitivity of 18.2%, [7.0% to 35.5%]) (Sullivan et al. 2021). These findings indicate that using the panel as a first filter before LDCT imaging might miss the presence of tumors in early stages of development. An observation from the authors is that the study was designed for the comparison of the panel in combination with LDCT versus standard clinical care, but it was not intended for the comparison of this combined strategy vs LDCT alone, a question in need for further research.

The results of this present evaluation also indicate that EarlyCDT®-Lung may have insufficient sensitivity for the identification of screening participants with LDCT-screen-detected indeterminate nodules for which invasive diagnostics such as biopsy or surgery should be recommended. As mentioned in Section 1.3.5, EarlyCDT®-Lung is marketed as a "rule-in" test, meaning that positive test result should help identify patients at increased risk of harboring pulmonary tumors, whereas negative results should not modify the clinical management plan, i.e., the panel should not be used as a "rule out" test. The low sensitivity of the panel observed on data from the LUSI trial confirms the recommendations from the providers. Similarly, previous studies (Healey et al. 2017; Jett et al. 2014; Massion et al. 2017) among clinical case series or patients with indeterminate nodules (Jett et al. 2014) indicated that the negative predictive value (NPV) of the EarlyCDT®-Lung may be too low for it to be used for ruling out the need for subjects with suspicious nodules to get more invasive diagnostic procedures such as biopsies. Those studies, however, reported higher sensitivities for EarlyCDT®-Lung as a lung cancer detection test.

Finally, regarding the association between a positive test result and the presence of malignant lung tumors among subjects with suspicious nodules seen on their LDCT-scan images, the positive likelihood ratio estimates reported in this study (1.5 to 1.9 depending on the definition of a positive test), despite being non-statistically significant due to the small sample size, are in line with those from previous studies (LR+: 2.3 95% CI:[1.3, 4.1], all nodules combined) (Healey et al. 2017; Massion et al. 2017).

Two limitations of this study are the small size of the LUSI trial, leading to wide confidence intervals for all reported estimates and the way in which blood samples were collected. Furthermore, according to the protocol, no blood samples were taken from participants with interval cancers since, by definition, they all received a lung cancer diagnosis without previous suspicious findings on their LDCT-scan appointments. Without such samples, it was not possible to evaluate whether the EarlyCDT®-Lung test could have detected malignant nodules that went undetected by LDCT.

4.5 Strengths and limitations

The four topics analyzed in this thesis are relevant for the design and implementation of lung cancer screening programs. These analyses were carried out at a time when discussions on how to optimally design such a program are very active in various countries around the world.

Furthermore, this work presents findings product of the first external validation of the LCRAT+CT and Polynomial models, proposed as tools for the risk-based selection of candidates for biennial vs. annual screening.

The analyses regarding nodule malignancy prediction are the first external validation of the UKLS, PanCan-MD and PanCan-VOL models for the prediction of malignancy of pulmonary nodules detected by LDCT screening. This was also the first evaluation of the PanCan, UKLS, Mayo, VA and PKUPH models applied to nodules detected in prevalence vs. incidence screening rounds.

For the estimation of overdiagnosis risk, the work presented in this thesis goes beyond the mere estimation of excess incidence. Instead, mathematical modeling was applied in order to estimate the biological (tumor mean preclinical sojourn time) and technological (detection sensitivity) that influence the risk of tumors being overdiagnosed by LDCT screening.

Additionally, this thesis shows the results of the first evaluation of the EarlyCDT®-Lung test for lung cancer detection in a screening context, as measured in blood samples taken as part of the LUSI trial.

Further strengths of this thesis are inherited from the quality of the data generated by the LUSI trial. The structured data collection and management throughout the trial allowed for the unique identification of nodules deemed malignant, contrary to other studies, in which malignancy was assumed to be always linked to the largest non-calcified nodule (NCN) (Nair et al. 2018; Tammemägi et al. 2019a; White et al. 2017), or to studies on biomarkers for lung cancer detection (Healey et al. 2017; Massion et al. 2017), in which participants were always represented by their largest NCN. Furthermore, in the LUSI trial nodule diameters were determined based on a software-based three-dimensional estimation, which was likely more precise compared to manual measurements applied in other studies (Han et al. 2018). This might have in turn improved the accuracy of estimates from prediction models that included diameter as the parameter of choice for nodule size.

The main limitations of this study are the absence of complete information for some model predictors, the small size of the LUSI trial, and the low numbers of cancer cases and of malignant nodules observed. Missing data might have led to bias in predicted risks, while the small size of the trial led to wide confidence intervals for all reported estimates and low or even insufficient power for statistical tests.

4.6 Conclusions and outlook

Lung cancer screening by means of LDCT imaging has been shown to reduce lung cancer mortality in randomized control trials in the USA and Europe. These findings have led to the institution of national screening programs in countries like the USA (Centers for Medicare & Medicaid Services 2015), South Korea (Lee et al. 2019) and Poland (Rzyman et al. 2019) and started discussions about their implementation in Canada as well as in some Asian and European countries, including Germany. However, designing a screening program to optimally balance its mortality reduction benefit and risks, potential harms and associated costs is a complex task, with various open questions still awaiting response.

Main questions are related to, among others: 1) the use of prediction models that can support the risk-based assignment of participants to optimal screening frequencies, 2) the use of malignancy prediction models to improve nodule evaluation and management, 3) the selection of methods that can accurately estimate overdiagnosis risk and how these estimates can improve the definition of upper age limits for starting or continuing screening, and, finally 4) the identification of biomarkers that could serve as complementary tools at various steps of the screening process.

Extensive research suggests that screening eligibility and personalized assignment of screening intervals based on estimates of lung cancer risk can increase the cost-effectiveness of screening programs. Similarly, sufficient evidence indicates that more accurate nodule evaluation could help reduce costs and harms related to unnecessary confirmatory procedures caused by false positive findings. Following these observations, statistical models have been developed for the prediction of lung cancer risk and of malignancy of screen-detected nodules. Statistical modeling has also been used to estimate the extent of overdiagnosis attributed to screening, and to estimate the biological and technological parameters that influence this risk. Independently, a separate line of research has focused on the identification of biomarkers to complement imaging-based screening, leading to a number of promising candidates. However, external validation of such models and candidate biomarkers in data from screening contexts is limited.

This thesis aimed at evaluating a selection of risk prediction models, methods for the estimation of overdiagnosis, as well as biomarkers that could help design an optimal screening program. Based on analyses carried on data from the German LUSI trial, this thesis reached four main conclusions:

First, the heterogeneity of lung cancer risks among screening-eligible subjects justifies the assignment of participants to screening intervals longer than a year. This assignment (e.g., annual vs. biennial screening) can be supported by prediction models that discriminate

participants who would get a higher net benefit from such a screening schedule on the basis of their estimated short-term lung cancer risks. Furthermore, models that incorporate predictors derived from LDCT images could enhance discrimination, compared to models based only on patient characteristics or only on the presence of pulmonary nodules.

Second, screening programs would benefit from a statistical modeling approach for nodule evaluation. Models for the prediction of nodule malignancy seem to be useful tools for optimizing nodule evaluation and management in population screening settings. More specifically, models with good discrimination performance, such as the PanCan variants validated in this thesis could help reduce the number of false positive test results. However, there is the need for well-calibrated models and appropriate risk thresholds for the evaluation of nodules detected at various stages of the screening process.

Third, excess incidence estimates obtained on the basis of data from randomized stop-screening trials (i.e., with a pre-fixed duration of the screening period) have so far not accurately reflected the genuine extent of overdiagnosis. In contrast, statistical modeling could be used to produce more transferable and interpretable estimates of overdiagnosis. The reasoning behind this hypothesis is that the risk of overdiagnosis strongly depends on both biological and technological factors. On the one hand, this risk depends on the relationship between the lead time of tumors detected via LDCT and the residual life expectancy (RLE) of screening participants. Screening participants presenting with tumors with long estimated lead times and comparatively short RLEs are therefore at increased risk of overdiagnosis. On the other hand, the lead time depends on the mean preclinical sojourn time (MPST) of different tumor subtypes and the sensitivity of the detection method, whereas the RLE depends on age at screen detection and indicators of general health. Therefore, a better approach would be to combine accurate estimates of such parameters in order to obtain better predictions of overdiagnosis risk. Two challenges of this approach are the identification and application of appropriate models with which to infer these parameters and the availability of sufficient data for their training and validation. Finally, with the right estimates, personalized decisions about whether to start or continuing screening could be taken on the basis of these predictions.

Fourth, there is the potential for biomarkers to improve various aspects of screening programs, including eligibility and nodule evaluation. However, even the so far most promising biomarker, EarlyCDT®-Lung, widely validated as a lung cancer detection tool in clinical settings, and recently evaluated as a pre-screening tool in a large prospective randomized trial (ECLS), has shown insufficient sensitivity for the identification of participants with lung tumors detected via LDCT in this and other recent studies (Borg et al. 2021; Sullivan et al. 2021). This biomarker is therefore not recommended for its use as a tool for pre-selection or for diagnostic triage in lung cancer screening programs. More research and more evidence coming from screening

studies, as well as cost-effectiveness analyses are required before biomarkers can be incorporated into screening programs.

4.6.1 Outlook and need for further research

Although statistical models for the assignment of screening intervals and for the prediction of nodule malignancy have shown good discrimination ability, evidence regarding their transferability to other populations and other timepoints along the screening process is limited. Thus, there is a need for their validation in data from screening contexts other than those in which they were developed. Furthermore, given that predicted risks depend on the demographics and risk-factors of a given population, there is the need for their calibration to be evaluated in independent populations and, if necessary, for the models to be re-calibrated (Su et al. 2018). Additional work is also needed for the definition of risk thresholds for these models that guarantee an optimal net clinical benefit for screening participants and/or improvements in cost-effectiveness of screening programs.

In particular, more research is needed regarding how to develop and apply models for the optimization of screening intervals, instead of being constrained to biennial vs annual. In fact, the personalized definition of screening intervals based on lung cancer risks and remaining life expectancy is currently being investigated (O'Mahony et al. 2015; Toumazis et al. 2019). Further attempts at filling the evidence gaps around personalized risk-based screening intervals are being made by a number of randomized clinical trials such as SUMMIT, BioMILD and 4-IN-THE-LUNG-RUN (Horst et al. 2020; Pastorino 2014; Ten Haaf et al. 2020).

In the case of malignancy prediction of nodules seen in previous rounds, algorithms are now being developed that integrate relevant nodule features on repeated CT screening examinations over time to predict the presence of lung cancer (Horeweg et al. 2014b; Huang et al. 2019). For individuals screened at regular intervals, models may be further developed to incorporate estimates of nodule volume doubling times, determined on early recall follow-up CT or directly for nodules already detected on earlier visits when individuals return for annual incidence screenings.

In general, models that rely on LDCT findings, independently of the outcome they intend to predict, might become more accurate if trained using advanced modeling techniques based on the images instead of on pre-extracted features such as nodule size, location and texture. Examples of such approaches are radiomics and machine learning algorithms. Also, any type of model intended for risk-based decision making along the screening process could become more accurate and better calibrated if they were trained and validated in larger and richer datasets, for example based on pooled data worldwide.

In terms of overdiagnosis, there is the need to develop multivariable models for predicting an individual's residual life expectancy based on indicators of general health, or even on findings about diseases detected in screening CT-scans (Heuvelmans et al. 2019; Yip et al. 2021). A further step, also in need of research, could be to incorporate such estimates of residual life expectancy into criteria for eligibility (Cheung et al. 2019; Katki et al. 2019) and for stopping screening. In fact, lung cancer risk prediction models have been suggested that include comorbidities as a proxy of remaining life expectancy (Cheung et al. 2019).

Biomarkers, on the other hand, if used as a stand-alone tool, need to be superior to current risk-prediction models, for example in terms of discrimination. Alternatively, as complementary tools, they should improve risk prediction when added to such models. They need to be cost-effective, easy to apply and minimally invasive. According to findings of recent studies, there is the potential for biomarkers to improve not only the eligibility criteria of screening programs but also the assignment of screening frequency and the evaluation of nodules. For example, a model based on smoking history and a panel of protein markers showed higher sensitivity at a fixed specificity for the discrimination of future lung cancer cases, than smoking history alone when evaluated on samples from two European cohorts (Guida et al. 2018). First reports from the Bio-MILD study indicate that CT screening findings combined with blood microRNA may help finding optimal screening intervals (Montani et al. 2015; Pastorino 2014; Sozzi et al. 2014). Furthermore, evaluations on data from the Pittsburgh Lung Screening Study suggests that blood-based biomarkers can improve the assessment of nodule malignancy (Fehlmann et al. 2020). Finally, the SUMMIT study aims at evaluating a cell-free nucleic acid blood test in a large cohort of subjects with high lung cancer risks as predicted by the PLCO model (Horst et al. 2020). However, more external validations in independent screening studies, as well as cost-effectiveness analyses are required before biomarkers can be incorporated into screening programs.

Additionally, modeling approaches for the optimization of lung cancer screening might profit from sex-specific analyses. There is strong evidence for sex disparities at all stages of screening: eligibility, recruitment, assignment of screening frequency, and regarding benefits and harms (e.g., mortality reduction rates and risk of overdiagnosis) (see TenHaaf et al. 2020 for review). Furthermore, evidence from clinical trials seems heterogeneous across different screening populations. However, there is limited research regarding a sex-specific risk-based approach for the design of screening programs. More research along these lines is urgently needed.

Finally, there are various other aspects of lung cancer screening in need for research, however they are outside the scope of this thesis. These include screening of never smokers, successful invitation and appropriate distribution of information, unbiased recruitment strategies,

screening uptake and adherence, the incorporation of smoking cessation programs, cost-effectiveness and the role of health insurance programs.

4.6.2 Implications for future lung cancer screening programs

For future screening programs, more reflection will be needed about how to combine risk-based approaches to identify individuals for initial lung cancer screening, to determine optimized time points for follow-up screenings, and to optimally evaluate their imaging findings.

For example, a two-step strategy could be implemented that uses lung cancer risk models such as LCRAT or PLCOM2012 to identify candidates for at least low-frequency (e.g., biennial) screening. On a second step, augmented models incorporating risk indicators from the baseline LDCT scan, such as LCRAT+CT or PLCOm2021results, could identify participants those who would benefit most from more frequent (e.g., annual) screening, or may be even combined with models that estimate the optimal length of the screening interval. Once the candidates are selected, the number of false positive tests, one of the major concerns of screening, could be reduced by predicting the malignancy risk of nodules through statistical models based on LDCT imaging findings. For later screening rounds, previous findings could be used to obtain more accurate malignancy predictions.

Regarding improved rules for stopping screening participation, randomized trials as well as the first population-wide screening programs have set fixed upper age limits for all screening participants. Instead, a recommendation for future screening programs based on the findings presented in this thesis, would be to rely on estimates of MPST of different tumor types and of RLE of participants for deciding, on an individual basis, when screening should stop. Screening participants could be regularly monitored, their remaining lifetime predicted based on their age and overall health and on a second step, their risk of being overdiagnosed could be assessed based on these estimates relative to the predicted lead times of different tumor histologic subtypes. Then, the decision on whether or not to continue screening would be based on pre-defined thresholds of overdiagnosis risk.

In general, whether it is related to assignment of screening intervals or decision making about stopping screening, the longitudinal assessment of lung cancer risk is relevant given that it does not remain static over a subjects' lifetime, but rather varies due to ageing and changes in lifestyle and behavior (e.g., smoking habits).

Finally, assuming the search for appropriate biomarkers is successful, future programs might become more precise, less invasive and risky, and even more cost-effective by incorporating them at various stages of the screening process.

Overall, these are exciting times. There is now enough evidence in favor of LDCT screening as a tool to reduce lung cancer mortality. This evidence has awakened the interest of the

scientific and medical community and has drawn the attention of governments and policy makers.

There is a significant amount of research being conducted. Three current major efforts towards optimizing screening are the SUMMIT study in the UK, the Italian BioMILD, and 4-IN-THE-LUNG-RUN, a trial conducted jointly by several European countries, including Germany (Horst et al. 2020; Pastorino 2014; Ten Haaf et al. 2020).

Various countries, in and outside Europe, are taking steps towards the implementation of screening programs. Soon, decisions will be made that can potentially save millions of lives. It is to be hoped that the findings presented and discussed in this thesis can contribute to inform the decision-making process that leads to optimal screening programs and in doing so, help in the fight against lung cancer.

5   SUMMARY

More deaths can be attributed to lung cancer, than to any other cancer type. Evidence collected over the last 10 years, from randomized trials in the USA and Europe, indicates that screening by means of low-dose computed tomography (LDCT) could reduce the number of lung cancer (LC) deaths by about 20%-24%. While these findings have led to the implementation of screening programs in the USA, South Korea and Poland, discussions on their optimal design and execution are still ongoing in various countries, including Germany.

Optimizing screening means finding the right balance between mortality reduction and risks, harms, and monetary costs. LDCT-scans are expensive, expose participants to radiation and put them at risk for overdiagnosis, as well as at risk for unnecessary invasive and expensive confirmatory procedures triggered by false positive (FP) results. Minimizing the number of unnecessary screening and confirmatory examinations should be prioritized. While risk-based eligibility has been shown to best target candidates, questions regarding optimal screening frequency, accurate nodule evaluation, stop-screening criteria to reduce overdiagnosis, and the use of complementary non-invasive diagnostic methods, remain open. Statistical models and biomarkers have been developed to help answer these questions. However, there is limited evidence of their validity in data from screening contexts and populations other than those in which they were developed.

The analyses presented in this thesis are based on data collected as part of the German Lung Cancer Screening Intervention (LUSI) trial in order to validate models that address the questions: 1) can candidates for biennial vs annual screening be identified on the basis of their LC risk? 2) can the number of FP test results be reduced by accurately estimating the malignancy of LDCT-detected nodules? 3) What was the extent of overdiagnosis in the LUSI trial and how does overdiagnosis risk relate to the age and remaining lifetime of participants? Additionally, blood samples from participants of the LUSI were measured to evaluate: 4) whether the well-validated diagnostic biomarker test EarlyCDT®-Lung is sensitive enough to detect tumors seen in LDCT images.

The LCRAT+CT and Polynomial models predict LC risk based on subject characteristics and LDCT imaging findings. Results of this first external validation confirmed their ability to identify participants with LC detected within 1-2 years after first screening. Discrimination was higher compared to a criterion based on nodule size and, to a lesser degree, compared to a model based on smoking and subject characteristics (LCRAT). This suggested that while LDCT findings can enhance models, most of their performance can could be attributed to information on smoking. Skipping 50% of annual LDCT examinations (i.e., for participants with estimated risks <5th decile) would have caused <10% delayed diagnoses, indicating that candidates for biennial screening could be identified based on their predicted LC risks without compromising

on early detection. Absolute risk estimates were, on average, below the observed LC rates, indicating poor calibration.

Models developed using data from the Canadian screening study PanCan showed excellent ability to differentiate between tumors and non-malignant nodules seen on LDCT scans taken at first screening participation and to accurately predict absolute malignancy risk. However, they showed lower performance when applied on data of nodules detected in later rounds. In contrast, a model developed on data from the UKLS trial and models developed on data from clinical settings did not perform as well in any screening round.

Excess incidence of screen-detected lung tumors, an estimator of overdiagnosis, was within the range of values reported by other trials after similar post-screening follow-up (ca. 5-6 years). Estimates of mean pre-clinical sojourn time (MPST) and LDCT detection sensitivity were obtained via mathematical modeling. The highest excess incidence and longest MPST estimates were found among adenocarcinomas. The proportion of tumors with long lead times predicted based on MPST estimates (e.g., 23% with lead times ≥8 years) suggested a substantial overdiagnosis risk for individuals with residual life expectancies shorter than these hypothetical lead times, for example for heavy smokers over the age of 75.

The tumor autoantibody panel measured by EarlyCDT®-Lung, a test widely validated as a diagnostic tool in clinical settings and recently tested as a pre-screening tool in a large randomized Scottish trial (ECLS), was found to have insufficient sensitivity for the identification of lung tumors detected via LDCT and of participants with screen-detected pulmonary nodules for whom more invasive diagnostic procedures should be recommended.

Overall, the findings presented in this thesis indicate that risk prediction models can help optimize LC screening by assigning participants to appropriate screening intervals, and by increasing the accuracy of nodule evaluation. However, there is a need for further external model validation and re-calibration. Additionally, while excess incidence can provide estimates of overdiagnosis risk at a population-level, a better approach would be to obtain model-based personalized estimates of tumor lead and residual lifetime. Better individualized decisions about whether to start or stop screening could be taken on the basis of the relationship between these estimates and the risk of overdiagnosis. Finally, although there is evidence for the potential of biomarkers to complement LC screening, the so far most promising candidate (EarlyCDT®-Lung) cannot be recommended as a pre-screening tool given its poor sensitivity for the identification of lung tumors detected via LDCT. In conclusion, while steps have been taken in the right direction, more research is required in order to answer all open questions regarding the optimal design of lung cancer screening programs.

6 ZUSAMMENFASSUNG

Auf Lungenkrebs können mehr Todesfälle zurückgeführt werden als auf jede andere Krebsart. Evidenz der letzten 10 Jahre zeigt, dass Screening mittels Niedrigdosis-Computertomografie (LDCT) die Anzahl der Todesfälle durch Lungenkrebs (LK) um 20% -24% reduzieren könnte. Während diese Erkenntnis zur Einführung von Screening-Programmen in den USA, Südkorea und Polen geführt hat, wird in verschiedenen Ländern, darunter auch in Deutschland, noch über die optimale Gestaltung und Durchführung diskutiert.

Screening-Programme zu optimieren bedeutet, die Balance zwischen Mortalitätsreduktion und Risiken, negativen Auswirkungen und Kosten, zu finden. LDCT-Scans sind teuer, setzen die Teilnehmer Strahlung sowie einem Risiko für Überdiagnose und unnötiger invasiver und teurer Abklärungsdiagnostik aus, verursacht durch falsch-positive (FP) Ergebnisse. Die Reduktion unnötiger Screenings und Abklärungsdiagnostiken senkt somit Risiken und Kosten. Während sich gezeigt hat, dass zu screenende Personen idealerweise risikobasiert ausgewählt werden sollten, bleiben Fragen zu optimalen Screening-Intervallen, zur Lungenknotenbewertung, zu Abbruch-Kriterien und zur Verwendung nicht-invasiver Diagnostikmethoden offen. Statistische Modelle und Biomarker wurden entwickelt, um diese Fragen zu beantworten. Es gibt jedoch nur begrenzte Beweise für ihre Gültigkeit in anderen Studienpopulationen.

Diese Arbeit stellt Analysen, basierend auf Daten aus der deutschen Lung Cancer Screening Intervention (LUSI) Studie vor, die Modelle validieren, welche sich mit den folgenden Fragen befassen: 1) Können Kandidaten für ein nur zweijährliches Screening anhand ihres LK-Risikos identifiziert werden? 2) Kann die Anzahl der FP-Ergebnisse reduziert werden, indem die Bösartigkeit von LDCT-detektierten Knoten genau geschätzt wird? 3) Wie hoch war das Ausmaß der Überdiagnose in der LUSI-Studie und wie hängt es mit Alter und verbleibender Lebensdauer zusammen? Zusätzlich wurden Blutproben von Probanden der Studie gemessen, um zu bewerten: 4) ob der diagnostische Biomarker-Test, EarlyCDT®-Lung, sensitiv genug ist, um Tumore zu entdecken, die in LDCT-Bildern erkannt wurden.

Die LCRAT+CT- und Polynomial-Modelle sagen das LK-Risiko basierend auf Merkmalen der Probanden und deren LDCT-Befunden voraus. Die Ergebnisse dieser ersten externen Validierung bestätigten die Fähigkeit der Modelle, Teilnehmer zu identifizieren, bei denen LK innerhalb von 1-2 Jahren nach dem ersten Screening entdeckt werden konnte. Im Vergleich zur Knotengröße als alleinigem Kriterium und zu einem Modell basierend auf Rauchverhalten und personenbezogener Risikoindikatoren (LCRAT) war die Diskriminierung der beiden erstgenannten Modelle etwas genauer. Dies deutet darauf hin, dass LDCT-Ergebnisse zwar Modelle verbessern können, der größte Teil ihrer Leistung jedoch auf Informationen über das Rauchverhalten zurückzuführen ist. Eine zweijährliche- anstelle von jährlicher LDCT-Untersuchung für Teilnehmer mit geschätzten Risiken <5. Dezil hätte zu <10% verzögerten Diagnosen geführt. Dies weist darauf hin, dass Kandidaten für ein zweijährliches Screening basierend auf ihren vorhergesagten LK-Risiken identifiziert werden könnten, ohne Kompromisse bei der Früherkennung einzugehen. Die absoluten Risikoschätzungen lagen im Durchschnitt unter den beobachteten LK-Raten, was auf eine schlechte Kalibrierung hindeutet.

Modelle, die anhand von Daten aus der kanadischen PanCan Screening-Studie entwickelt wurden, zeigten eine ausgezeichnete Fähigkeit, zwischen Tumoren und nicht-bösartigen Knoten zu unterscheiden, die in LDCT-Scans bei der ersten Screening-Teilnahme beobachtet wurden und konnten das absolute Risiko gut vorhersagen. Sie zeigten jedoch eine geringere Leistung, wenn sie auf Daten von Knoten angewendet wurden, die in späteren Runden entdeckt wurden. Im Gegensatz dazu zeigte ein Modell, das aufgrund von Daten aus der UKLS-Studie-, ebenso wie andere Modelle die anhand von klinischen Daten entwickelt wurden, in keiner Screening-Runde vergleichbar gute Leistung.

Die überhöhte Inzidenz (ÜI) von screen-detektierten (SD) Lungentumoren, lag innerhalb des Wertebereichs, der in anderen Studien nach ähnlichem Follow-up (ca. 5-6 Jahre post-Screening) berichtet wurde. Schätzungen der mittleren präklinischen Verweilzeit (MPST) und der LDCT-Sensitivität wurden mittels mathematischer Modellierung geschätzt. Die höchste ÜI und längste MPST wurden bei Adenokarzinomen gefunden. Der Anteil der Tumore mit langen Vorlaufzeiten, die auf der Grundlage von MPST-Schätzungen vorhergesagt wurden (z. B. 23% mit Vorlaufzeiten ≥8 Jahren), deutete auf ein erhebliches Überdiagnoserisiko für Personen mit einer Restlebenserwartung hin, die kürzer ist als diese Vorlaufzeiten, beispielsweise für Raucher über 75 Jahre.

Das Tumor Autoantikörper-Panel, gemessen mit EarlyCDT®-Lung, einem Test, der weithin als diagnostische Methode in klinischen Umgebungen validiert und kürzlich als Pre-Screening-Tool in einer großen randomisierten schottischen Studie (ECLS) getestet wurde, wies eine unzureichende Sensitivität für die Identifizierung von Lungentumoren auf, die über LDCT erkannt wurden. Auch die Sensitivität zur Identifikation von Teilnehmern mit SD-Lungenknoten, für die invasivere Diagnoseverfahren empfohlen werden sollten, war unzureichend.

Insgesamt deuten die Ergebnisse dieser Arbeit darauf hin, dass Risikovorhersagemodelle dazu beitragen können LK-Screening zu optimieren, indem sie Teilnehmer geeigneten Screening-Intervallen zuweisen und die Genauigkeit der Lungenknotenbewertung erhöhen. Es besteht jedoch Bedarf an weiterer externer Modellvalidierung und Kalibrierung. Während die ÜI das Überdiagnoserisiko auf Bevölkerungsebene schätzen kann, ist eine modellbasierte personalisierte Schätzung der MPST und der Restlebensdauer zu bevorzugen. Anhand dieser Schätzungen und dem Risiko einer Überdiagnose könnten besser individualisierte Entscheidungen darüber getroffen werden, ob das Screening gestartet oder abgebrochen werden sollte. Obwohl es Hinweise auf das Potenzial von Biomarkern gibt, LK-Screening zu ergänzen, kann der bisher vielversprechendste Test, EarlyCDT®-Lung, aufgrund seiner niedrigen Sensitivität, nicht als Pre-Screening-Tool empfohlen werden.

Zusammenfassend, lässt sich feststellen, dass zwar Schritte in die richtige Richtung unternommen wurden, jedoch mehr Forschung erforderlich ist, um alle offenen Fragen zur optimalen Gestaltung von LK-Screening-Programmen zu beantworten.

7 REFERENCES

Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., Gareen, I. F., Gatsonis, C., Marcus, P. M. and Sicks, J. D. (2011). **Reduced lung-cancer mortality with low-dose computed tomographic screening**. N Engl J Med *365 (5)*, 395-409, doi: 10.1056/NEJMoa1102873.

Al-Ameri, A., Malhotra, P., Thygesen, H., Plant, P. K., Vaidyanathan, S., Karthik, S., Scarsbrook, A. and Callister, M. E. (2015). **Risk of malignancy in pulmonary nodules: A validation study of four prediction models**. Lung Cancer *89 (1)*, 27-30, doi: 10.1016/j.lungcan.2015.03.018.

Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. and Spivack, S. D. (2013). **Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines**. Chest *143 (5 Suppl)*, e1S-e29S, doi: 10.1378/chest.12-2345.

American Cancer Society (2018). **Global Cancer Facts & Figures 4th Edition**. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/global-cancer-facts-and-figures/global-cancer-facts-and-figures-4th-edition.pdf [as of July 17, 2021].

American College of Radiology Committee on Lung-RADS® (2014). **Lung-RADS™ Version 1.0 Assessment Categories**. URL: https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADS_AssessmentCategories.pdf [as of January 1, 2021].

American College of Radiology Committee on Lung-RADS® (2019). **Lung-RADS® Version 1.1 Assessment Categories**. URL: https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf [as of October 19, 2020].

Anderson, K. S. and LaBaer, J. (2005). **The sentinel within: exploiting the immune system for cancer biomarkers**. J Proteome Res *4 (4)*, 1123-1133, doi: 10.1021/pr0500814.

Arenberg, D. (2019). **Update on screening for lung cancer**. Transl Lung Cancer Res *8 (Suppl 1)*, S77-S87, doi: 10.21037/tlcr.2019.03.01.

Bach, P. B. and Gould, M. K. (2012). **When the average applies to no one: personalized decision making about potential benefits of lung cancer screening**. Ann Intern Med *157 (8)*, 571-573, doi: 10.7326/0003-4819-157-8-201210160-00524.

Bach, P. B., Kattan, M. W., Thornquist, M. D., Kris, M. G., Tate, R. C., Barnett, M. J., Hsieh, L. J. and Begg, C. B. (2003a). **Variations in lung cancer risk among smokers**. J Natl Cancer Inst *95 (6)*, 470-478, doi: 10.1093/jnci/95.6.470.

Bach, P. B., Kelley, M. J., Tate, R. C. and McCrory, D. C. (2003b). **Screening for lung cancer: a review of the current literature**. Chest *123 (1 Suppl)*, 72S-82S, doi: 10.1378/chest.123.1_suppl.72s.

Bach, P. B., Mirkin, J. N., Oliver, T. K., Azzoli, C. G., Berry, D. A., Brawley, O. W., Byers, T., Colditz, G. A., Gould, M. K., Jett, J. R., Sabichi, A. L., Smith-Bindman, R., Wood, D. E., Qaseem, A. and Detterbeck, F. C. (2012). **Benefits and harms of CT screening for lung cancer: a systematic review**. Jama *307 (22)*, 2418-2429, doi: 10.1001/jama.2012.5521.

Bailey-Wilson, J. E., Amos, C. I., Pinney, S. M., Petersen, G. M., de Andrade, M., Wiest, J. S., Fain, P., Schwartz, A. G., You, M., Franklin, W., Klein, C., Gazdar, A., Rothschild, H., Mandal, D., Coons, T., Slusser, J., Lee, J., Gaba, C., Kupert, E., Perez, A., Zhou, X., Zeng, D., Liu, Q., Zhang, Q., Seminara, D., Minna, J. and Anderson, M. W. (2004). **A major lung cancer susceptibility locus maps to chromosome 6q23-25**. Am J Hum Genet *75 (3)*, 460-474, doi: 10.1086/423857.

Baldwin, D. R. and Callister, M. E. (2015). **The British Thoracic Society guidelines on the investigation and management of pulmonary nodules**. Thorax *70 (8)*, 794-798, doi: 10.1136/thoraxjnl-2015-207221.

Baldwin, D. R., Duffy, S. W., Wald, N. J., Page, R., Hansell, D. M. and Field, J. K. (2011). **UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer**. Thorax *66 (4)*, 308-313, doi: 10.1136/thx.2010.152066.

Barnes, B., Kraywinkel, K., Nowossadeck, E., Schönfeld, I., Starker, A., Wienecke, A. and Wolf, U. (2016). **Bericht zum Krebsgeschehen in Deutschland 2016** (Robert Koch-Institut %J Gesundheitsberichterstattung-Hefte, doi: 10.17886/rkipubl-2016-014, p.^pp.

Barton, K. (2020). **MuMIn: Multi-Model Inference. R package version 1.43.17a** (https://CRAN.R-project.org/package=MuMIn.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). **Fitting Linear Mixed-Effects Models Using lme4**. 2015 *67 (1)*, 48, doi: 10.18637/jss.v067.i01.

Becker, N., Motsch, E., Gross, M. L., Eigentopf, A., Heussel, C. P., Dienemann, H., Schnabel, P. A., Eichinger, M., Optazaite, D. E., Puderbach, M., Wielputz, M., Kauczor, H. U., Tremper, J. and Delorme, S. (2015). **Randomized Study on Early Detection of Lung Cancer with MSCT in Germany: Results of the First 3 Years of Follow-up After Randomization**. J Thorac Oncol *10 (6)*, 890-896, doi: 10.1097/jto.0000000000000530.

Becker, N., Motsch, E., Gross, M. L., Eigentopf, A., Heussel, C. P., Dienemann, H., Schnabel, P. A., Pilz, L., Eichinger, M., Optazaite, D. E., Puderbach, M., Tremper, J. and Delorme, S. (2012). **Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round**. J Cancer Res Clin Oncol *138 (9)*, 1475-1486, doi: 10.1007/s00432-012-1228-9.

Becker, N., Motsch, E., Trotter, A., Heussel, C. P., Dienemann, H., Schnabel, P. A., Kauczor, H. U., Gonzalez Maldonado, S., Miller, A. B., Kaaks, R. and Delorme, S. (2019). **Lung cancer mortality reduction by LDCT screening - results from the randomised German LUSI trial**. Int J Cancer, doi: 10.1002/ijc.32486.

Berlin, N. I., Buncher, C. R., Fontana, R. S., Frost, J. K. and Melamed, M. R. (1984). **The National Cancer Institute Cooperative Early Lung Cancer Detection Program. Results of the initial screen (prevalence). Early lung cancer detection: Introduction**. Am Rev Respir Dis *130 (4)*, 545-549, doi: 10.1164/arrd.1984.130.4.545.

Berrington de Gonzalez, A., Kim, K. P. and Berg, C. D. (2008). **Low-dose lung computed tomography screening before age 55: estimates of the mortality reduction required to outweigh the radiation-induced cancer risk**. J Med Screen *15 (3)*, 153-158, doi: 10.1258/jms.2008.008052.

Bianchi, F., Nicassio, F., Marzi, M., Belloni, E., Dall'olio, V., Bernard, L., Pelosi, G., Maisonneuve, P., Veronesi, G. and Di Fiore, P. P. (2011). **A serum circulating**

miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer. EMBO Mol Med *3 (8)*, 495-503, doi: 10.1002/emmm.201100154.

Black, W. C., Gareen, I. F., Soneji, S. S., Sicks, J. D., Keeler, E. B., Aberle, D. R., Naeim, A., Church, T. R., Silvestri, G. A., Gorelick, J. and Gatsonis, C. (2014). **Cost-effectiveness of CT screening in the National Lung Screening Trial**. N Engl J Med *371 (19)*, 1793-1802, doi: 10.1056/NEJMoa1312547.

Boeri, M., Verri, C., Conte, D., Roz, L., Modena, P., Facchinetti, F., Calabro, E., Croce, C. M., Pastorino, U. and Sozzi, G. (2011). **MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer**. Proc Natl Acad Sci U S A *108 (9)*, 3713-3718, doi: 10.1073/pnas.1100048108.

Boffetta, P., Jayaprakash, V., Yang, P., Asomaning, K., Muscat, J. E., Schwartz, A. G., Zhang, Z. F., Le Marchand, L., Cote, M. L., Stoddard, S. M., Morgenstern, H., Hung, R. J. and Christiani, D. C. (2011). **Tobacco smoking as a risk factor of bronchioloalveolar carcinoma of the lung: pooled analysis of seven case-control studies in the International Lung Cancer Consortium (ILCCO)**. Cancer Causes Control *22 (1)*, 73-79, doi: 10.1007/s10552-010-9676-5.

Borg, M., Wen, S. W. C., Nederby, L., Hansen, T. F., Jakobsen, A., Andersen, R. F., Weinreich, U. M. and Hilberg, O. (2021). **Performance of the EarlyCDT(R) Lung test in detection of lung cancer and pulmonary metastases in a high-risk cohort**. Lung Cancer *158*, 85-90, doi: 10.1016/j.lungcan.2021.06.010.

Bosse, Y. and Amos, C. I. (2018). **A Decade of GWAS Results in Lung Cancer**. Cancer Epidemiol Biomarkers Prev *27 (4)*, 363-379, doi: 10.1158/1055-9965.EPI-16-0794.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018). **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. CA Cancer J Clin *68 (6)*, 394-424, doi: 10.3322/caac.21492.

Brenner, D. R., McLaughlin, J. R. and Hung, R. J. (2011). **Previous lung diseases and lung cancer risk: a systematic review and meta-analysis**. PLoS One *6 (3)*, e17479, doi: 10.1371/journal.pone.0017479.

Brett, G. Z. (1968). **The value of lung cancer detection by six-monthly chest radiographs**. Thorax *23 (4)*, 414-420, doi: 10.1136/thx.23.4.414.

Brier, G. W. (1950). **Verification of Forecasts Expressed in Terms of Probability**. Monthly Weather Review *78 (1)*, 1-3, doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2 %J Monthly Weather Review.

Brodersen, J., Voss, T., Martiny, F., Siersma, V., Barratt, A. and Heleno, B. (2020). **Overdiagnosis of lung cancer with low-dose computed tomography screening: meta-analysis of the randomised clinical trials**. Breathe (Sheff) *16 (1)*, 200013, doi: 10.1183/20734735.0013-2020.

Broodman, I., Lindemans, J., van Sten, J., Bischoff, R. and Luider, T. (2017). **Serum Protein Markers for the Early Detection of Lung Cancer: A Focus on Autoantibodies**. J Proteome Res *16 (1)*, 3-13, doi: 10.1021/acs.jproteome.6b00559.

Bundesministerium für Gesundheit (2021). **Krebsfrüherkennung**. URL: https://www.bundesgesundheitsministerium.de/krebsfrueherkennung.html [as of August 21, 2021].

Califf, R. M. (2018). **Biomarker definitions and their applications**. Experimental biology and medicine (Maywood, N.J.) *243 (3)*, 213-221, doi: 10.1177/1535370217750088.

Canty, A. and Ripley, B. D. (2019). **boot: Bootstrap R (S-Plus) Functions. R package version 1.3-24.** (

Carey, V. J. (2019). **gee: Generalized Estimation Equation Solver. R package version 4.13-20** (https://CRAN.R-project.org/package=gee.

Centers for Medicare & Medicaid Services (2015). **Proposed Decision Memo on Screening for Lung Cancer with Low Dose Computed Tomography (LDCT)**. URL: https://www.cms.gov/medicare-coverage-database/view/ncacal-decision-memo.aspx?proposed=Y&NCAId=274 [as of August 26, 2021].

Chansky, K., Detterbeck, F. C., Nicholson, A. G., Rusch, V. W., Vallieres, E., Groome, P., Kennedy, C., Krasnik, M., Peake, M., Shemanski, L., Bolejack, V., Crowley, J. J., Asamura, H., Rami-Porta, R., Staging, I., Prognostic Factors Committee, A. B. and Participating, I. (2017). **The IASLC Lung Cancer Staging Project: External Validation of the Revision of the TNM Stage Groupings in the Eighth Edition of the TNM Classification of Lung Cancer**. J Thorac Oncol *12 (7)*, 1109-1121, doi: 10.1016/j.jtho.2017.04.011.

Chapman, C. J., Healey, G. F., Murray, A., Boyle, P., Robertson, C., Peek, L. J., Allen, J., Thorpe, A. J., Hamilton-Fairley, G., Parsy-Kowalska, C. B., MacDonald, I. K., Jewell, W., Maddison, P. and Robertson, J. F. (2012). **EarlyCDT(R)-Lung test: improved clinical utility through additional autoantibody assays**. Tumour Biol *33 (5)*, 1319-1326, doi: 10.1007/s13277-012-0379-2.

Chapman, C. J., Murray, A., McElveen, J. E., Sahin, U., Luxemburger, U., Türeci, O., Wiewrodt, R., Barnes, A. C. and Robertson, J. F. (2008). **Autoantibodies in lung cancer: possibilities for early detection and subsequent cure**. Thorax *63 (3)*, 228-233, doi: 10.1136/thx.2007.083592.

Cheung, L. C., Berg, C. D., Castle, P. E., Katki, H. A. and Chaturvedi, A. K. (2019). **Life-Gained-Based Versus Risk-Based Selection of Smokers for Lung Cancer Screening**. Ann Intern Med *171 (9)*, 623-632, doi: 10.7326/m19-1263.

Cheung, L. C., Kovalchik, S. A. and Katki, H. A. (2018). **lcrisks: Lung Cancer Death Risk Predictor. R package version 4.0.0.** (https://dceg.cancer.gov/tools/risk-assessment/lcrisks.

Chu, G. C. W., Lazare, K. and Sullivan, F. (2018). **Serum and blood based biomarkers for lung cancer screening: a systematic review**. BMC Cancer *18 (1)*, 181, doi: 10.1186/s12885-018-4024-3.

Chung, K., Mets, O. M., Gerke, P. K., Jacobs, C., den Harder, A. M., Scholten, E. T., Prokop, M., de Jong, P. A., van Ginneken, B. and Schaefer-Prokop, C. M. (2018). **Brock malignancy risk calculator for pulmonary nodules: validation outside a lung cancer screening population**. Thorax *73 (9)*, 857-863, doi: 10.1136/thoraxjnl-2017-211372.

Coughlin, S. S., Trock, B., Criqui, M. H., Pickle, L. W., Browner, D. and Tefft, M. C. (1992). **The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test**. J Clin Epidemiol *45 (1)*, 1-7, doi: 10.1016/0895-4356(92)90180-u.

Cressman, S., Peacock, S. J., Tammemagi, M. C., Evans, W. K., Leighl, N. B., Goffin, J. R., Tremblay, A., Liu, G., Manos, D., MacEachern, P., Bhatia, R., Puksa, S., Nicholas, G., McWilliams, A., Mayo, J. R., Yee, J., English, J. C., Pataky, R., McPherson, E.,

Atkar-Khattra, S., Johnston, M. R., Schmidt, H., Shepherd, F. A., Soghrati, K., Amjadi, K., Burrowes, P., Couture, C., Sekhon, H. S., Yasufuku, K., Goss, G., Ionescu, D. N., Hwang, D. M., Martel, S., Sin, D. D., Tan, W. C., Urbanski, S., Xu, Z., Tsao, M. S. and Lam, S. (2017). **The Cost-Effectiveness of High-Risk Lung Cancer Screening and Drivers of Program Efficiency**. J Thorac Oncol *12 (8)*, 1210-1222, doi: 10.1016/j.jtho.2017.04.021.

Croswell, J. M., Kramer, B. S., Kreimer, A. R., Prorok, P. C., Xu, J. L., Baker, S. G., Fagerstrom, R., Riley, T. L., Clapp, J. D., Berg, C. D., Gohagan, J. K., Andriole, G. L., Chia, D., Church, T. R., Crawford, E. D., Fouad, M. N., Gelmann, E. P., Lamerato, L., Reding, D. J. and Schoen, R. E. (2009). **Cumulative incidence of false-positive results in repeated, multimodal cancer screening**. Ann Fam Med *7 (3)*, 212-222, doi: 10.1370/afm.942.

Davison, A. C. and Hinkley, D. V. (1997). **Bootstrap methods and their application**, Cambridge University Press, New York.

Day, N. E. and Walter, S. D. (1984). **Simplified models of screening for chronic disease: estimation procedures from mass screening programmes**. Biometrics *40 (1)*, 1-14.

de Koning, H., van Der Aalst, C., Ten Haaf, K. and Oudkerk, M. (2018). **PL02.05 Effects of Volume CT Lung Cancer Screening: Mortality Results of the NELSON Randomised-Controlled Population Based Trial**. Journal of Thoracic Oncology *13 (10)*, S185, doi: 10.1016/j.jtho.2018.08.012.

de Koning, H. J., Meza, R., Plevritis, S. K., ten Haaf, K., Munshi, V. N., Jeon, J., Erdogan, S. A., Kong, C. Y., Han, S. S., van Rosmalen, J., Choi, S. E., Pinsky, P. F., Berrington de Gonzalez, A., Berg, C. D., Black, W. C., Tammemagi, M. C., Hazelton, W. D., Feuer, E. J. and McMahon, P. M. (2014). **Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force**. Ann Intern Med *160 (5)*, 311-320, doi: 10.7326/m13-2316.

de Koning, H. J., van der Aalst, C. M., de Jong, P. A., Scholten, E. T., Nackaerts, K., Heuvelmans, M. A., Lammers, J. J., Weenink, C., Yousaf-Khan, U., Horeweg, N., van 't Westeinde, S., Prokop, M., Mali, W. P., Mohamed Hoesein, F. A. A., van Ooijen, P. M. A., Aerts, J., den Bakker, M. A., Thunnissen, E., Verschakelen, J., Vliegenthart, R., Walter, J. E., Ten Haaf, K., Groen, H. J. M. and Oudkerk, M. (2020). **Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial**. N Engl J Med *382 (6)*, 503-513, doi: 10.1056/NEJMoa1911793.

Dela Cruz, C. S., Tanoue, L. T. and Matthay, R. A. (2011). **Lung cancer: epidemiology, etiology, and prevention**. Clin Chest Med *32 (4)*, 605-644, doi: 10.1016/j.ccm.2011.09.001.

DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L. (1988). **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach**. Biometrics *44 (3)*, 837-845.

Deppen, S. A., Blume, J. D., Aldrich, M. C., Fletcher, S. A., Massion, P. P., Walker, R. C., Chen, H. C., Speroff, T., Degesys, C. A., Pinkerman, R., Lambright, E. S., Nesbitt, J. C., Putnam, J. B., Jr. and Grogan, E. L. (2014). **Predicting lung cancer prior to surgical resection in patients with lung nodules**. J Thorac Oncol *9 (10)*, 1477-1484, doi: 10.1097/jto.0000000000000287.

Doroudi, M., Pinsky, P. F. and Marcus, P. M. (2018). **Lung Cancer Mortality in the Lung Screening Study Feasibility Trial**. JNCI Cancer Spectr *2 (3)*, pky042, doi: 10.1093/jncics/pky042.

Driscoll, T., Nelson, D. I., Steenland, K., Leigh, J., Concha-Barrientos, M., Fingerhut, M. and Pruss-Ustun, A. (2005). **The global burden of disease due to occupational carcinogens**. Am J Ind Med *48 (6)*, 419-431, doi: 10.1002/ajim.20209.

Du, Q., Yu, R., Wang, H., Yan, D., Yuan, Q., Ma, Y., Slamon, D., Hou, D., Wang, H. and Wang, Q. (2018). **Significance of tumor-associated autoantibodies in the early diagnosis of lung cancer**. Clin Respir J *12 (6)*, 2020-2028, doi: 10.1111/crj.12769.

Duan, P., Hu, C., Quan, C., Yi, X., Zhou, W., Yuan, M., Yu, T., Kourouma, A. and Yang, K. (2015). **Body mass index and risk of lung cancer: Systematic review and dose-response meta-analysis**. Sci Rep *5*, 16938, doi: 10.1038/srep16938.

Engels, E. A., Brock, M. V., Chen, J., Hooker, C. M., Gillison, M. and Moore, R. D. (2006). **Elevated incidence of lung cancer among HIV-infected individuals**. J Clin Oncol *24 (9)*, 1383-1388, doi: 10.1200/JCO.2005.03.4413.

Fehlmann, T., Kahraman, M., Ludwig, N., Backes, C., Galata, V., Keller, V., Geffers, L., Mercaldo, N., Hornung, D., Weis, T., Kayvanpour, E., Abu-Halima, M., Deuschle, C., Schulte, C., Suenkel, U., von Thaler, A. K., Maetzler, W., Herr, C., Fahndrich, S., Vogelmeier, C., Guimaraes, P., Hecksteden, A., Meyer, T., Metzger, F., Diener, C., Deutscher, S., Abdul-Khaliq, H., Stehle, I., Haeusler, S., Meiser, A., Groesdonk, H. V., Volk, T., Lenhof, H. P., Katus, H., Balling, R., Meder, B., Kruger, R., Huwer, H., Bals, R., Meese, E. and Keller, A. (2020). **Evaluating the Use of Circulating MicroRNA Profiles for Lung Cancer Detection in Symptomatic Patients**. JAMA Oncol *6 (5)*, 714-723, doi: 10.1001/jamaoncol.2020.0001.

Ferlay J, E. M., Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2020a). **Global Cancer Observatory: Cancer Today. Germany fact sheet**. URL: https://gco.iarc.fr/today/data/factsheets/populations/276-germany-fact-sheets.pdf [as of April 24, 2021].

Ferlay J, E. M., Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2020b). **Global Cancer Observatory: Cancer Today. Lung fact sheet**. URL: https://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf [as of April 24, 2021].

Field, J. K., Duffy, S. W., Baldwin, D. R., Brain, K. E., Devaraj, A., Eisen, T., Green, B. A., Holemans, J. A., Kavanagh, T., Kerr, K. M., Ledson, M., Lifford, K. J., McRonald, F. E., Nair, A., Page, R. D., Parmar, M. K., Rintoul, R. C., Screaton, N., Wald, N. J., Weller, D., Whynes, D. K., Williamson, P. R., Yadegarfar, G. and Hansell, D. M. (2016). **The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer**. Health Technol Assess *20 (40)*, 1-146, doi: 10.3310/hta20400.

Gardiner, N., Jogai, S. and Wallis, A. (2014). **The revised lung adenocarcinoma classification-an imaging guide**. J Thorac Dis *6 (Suppl 5)*, S537-546, doi: 10.3978/j.issn.2072-1439.2014.04.05.

Garzillo, C., Pugliese, M., Loffredo, F. and Quarto, M. (2017). **Indoor radon exposure and lung cancer risk: a meta-analysis of case-control studies**. Translational Cancer Research, S934-S943.

Gemeinsamer Bundesausschuss (2020). **Programm zur Früherkennung von Gebärmutterhalskrebs**. URL: https://www.g-

ba.de/themen/methodenbewertung/ambulant/frueherkennung-krankheiten/erwachsene/krebsfrueherkennung/gebaermutterhalskrebs-screening/ [as of August 26, 2021].

Gemeinsamer Bundesausschuss (2021a). **Früherkennung von Krebserkrankungen**. URL: https://www.g-ba.de/themen/methodenbewertung/ambulant/frueherkennung-krankheiten/erwachsene/krebsfrueherkennung/ [as of August 26, 2021].

Gemeinsamer Bundesausschuss (2021b). **Richtlinie des Gemeinsamen Bundesausschusses für organisierte Krebsfrüherkennungsprogramme** URL: https://www.g-ba.de/downloads/62-492-2605/0fb293cafb8e3e67acaf7b0bdd2e6eea/oKFE-RL-2021-07-01-iK-2022-01-01.pdf [as of August 8, 2021].

Genders, T. S., Spronk, S., Stijnen, T., Steyerberg, E. W., Lesaffre, E. and Hunink, M. G. (2012). **Methods for calculating sensitivity and specificity of clustered data: a tutorial**. Radiology *265 (3)*, 910-916, doi: 10.1148/radiol.12120509.

German Centre for Cancer Registry Data - Robert Koch Institute (2021). **Mortality data provided by the Federal Statistical Office.** . URL: www.krebsdaten.de/database [as of July 15, 2021].

Gohagan, J., Marcus, P., Fagerstrom, R., Pinsky, P., Kramer, B., Prorok, P. and Writing Committee, L. S. S. R. G. (2004). **Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute**. Chest *126 (1)*, 114-121, doi: 10.1378/chest.126.1.114.

Gohagan, J. K., Marcus, P. M., Fagerstrom, R. M., Pinsky, P. F., Kramer, B. S., Prorok, P. C., Ascher, S., Bailey, W., Brewer, B., Church, T., Engelhard, D., Ford, M., Fouad, M., Freedman, M., Gelmann, E., Gierada, D., Hocking, W., Inampudi, S., Irons, B., Johnson, C. C., Jones, A., Kucera, G., Kvale, P., Lappe, K., Manor, W., Moore, A., Nath, H., Neff, S., Oken, M., Plunkett, M., Price, H., Reding, D., Riley, T., Schwartz, M., Spizarny, D., Yoffie, R., Zylak, C. and Lung Screening Study Research, G. (2005). **Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer**. Lung Cancer *47 (1)*, 9-15, doi: 10.1016/j.lungcan.2004.06.007.

Gold, M. R. (1996). **Cost-effectiveness in health and medicine**, Oxford University Press, New York.

Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., Nicholson, A. G., Groome, P., Mitchell, A. and Bolejack, V. (2016). **The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer**. J Thorac Oncol *11 (1)*, 39-51, doi: 10.1016/j.jtho.2015.09.009.

González Maldonado, S., Delorme, S., Hüsing, A., Motsch, E., Kauczor, H. U., Heussel, C. P. and Kaaks, R. (2020a). **Evaluation of Prediction Models for Identifying Malignancy in Pulmonary Nodules Detected via Low-Dose Computed Tomography**. JAMA Netw Open *3 (2)*, e1921221, doi: 10.1001/jamanetworkopen.2019.21221.

González Maldonado, S., Hynes, L. C., Motsch, E., Heussel, C. P., Kauczor, H. U., Robbins, H. A., Delorme, S. and Kaaks, R. (2021a). **Validation of multivariable lung cancer risk prediction models for the personalized assignment of optimal screening frequency: a retrospective analysis of data from the German Lung Cancer**

Screening Intervention Trial (LUSI). Transl Lung Cancer Res *10 (3)*, 1305-1317, doi: 10.21037/tlcr-20-1173.

González Maldonado, S., Johnson, T., Motsch, E., Delorme, S. and Kaaks, R. (2021b). **Can autoantibody tests enhance lung cancer screening?-an evaluation of EarlyCDT((R))-Lung in context of the German Lung Cancer Screening Intervention Trial (LUSI)**. Transl Lung Cancer Res *10 (1)*, 233-242, doi: 10.21037/tlcr-20-727.

González Maldonado, S., Motsch, E., Trotter, A., Kauczor, H. U., Heussel, C. P., Hermann, S., Zeissig, S. R., Delorme, S. and Kaaks, R. (2020b). **Overdiagnosis in lung cancer screening: Estimates from the German Lung Cancer Screening Intervention Trial**. Int J Cancer, doi: 10.1002/ijc.33295.

Gould, M. K., Ananth, L. and Barnett, P. G. (2007). **A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules**. Chest *131 (2)*, 383-388, doi: 10.1378/chest.06-1261.

Guida, F., Sun, N., Bantis, L. E., Muller, D. C., Li, P., Taguchi, A., Dhillon, D., Kundnani, D. L., Patel, N. J., Yan, Q., Byrnes, G., Moons, K. G. M., Tjonneland, A., Panico, S., Agnoli, C., Vineis, P., Palli, D., Bueno-de-Mesquita, B., Peeters, P. H., Agudo, A., Huerta, J. M., Dorronsoro, M., Barranco, M. R., Ardanaz, E., Travis, R. C., Byrne, K. S., Boeing, H., Steffen, A., Kaaks, R., Husing, A., Trichopoulou, A., Lagiou, P., La Vecchia, C., Severi, G., Boutron-Ruault, M. C., Sandanger, T. M., Weiderpass, E., Nost, T. H., Tsilidis, K., Riboli, E., Grankvist, K., Johansson, M., Goodman, G. E., Feng, Z., Brennan, P., Johansson, M. and Hanash, S. M. (2018). **Assessment of Lung Cancer Risk on the Basis of a Biomarker Panel of Circulating Proteins**. JAMA Oncol *4 (10)*, e182078, doi: 10.1001/jamaoncol.2018.2078.

Halpern, M. T., Gillespie, B. W. and Warner, K. E. (1993). **Patterns of absolute risk of lung cancer mortality in former smokers**. J Natl Cancer Inst *85 (6)*, 457-464, doi: 10.1093/jnci/85.6.457.

Han, D., Heuvelmans, M. A., Vliegenthart, R., Rook, M., Dorrius, M. D., de Jonge, G. J., Walter, J. E., van Ooijen, P. M. A., de Koning, H. J. and Oudkerk, M. (2018). **Influence of lung nodule margin on volume- and diameter-based reader variability in CT lung cancer screening**. Br J Radiol *91 (1090)*, 20170405, doi: 10.1259/bjr.20170405.

Han, S. S., Ten Haaf, K., Hazelton, W. D., Munshi, V. N., Jeon, J., Erdogan, S. A., Johanson, C., McMahon, P. M., Meza, R., Kong, C. Y., Feuer, E. J., de Koning, H. J. and Plevritis, S. K. (2017). **The impact of overdiagnosis on the selection of efficient lung cancer screening strategies**. Int J Cancer *140 (11)*, 2436-2443, doi: 10.1002/ijc.30602.

Hanash, S. M., Ostrin, E. J. and Fahrmann, J. F. (2018). **Blood based biomarkers beyond genomics for lung cancer screening**. Transl Lung Cancer Res *7 (3)*, 327-335, doi: 10.21037/tlcr.2018.05.13.

Hardin, J. W. and Hilbe, J. M. (2003). **Generalized estimating equations**, Chapman & Hall/CRC, Boca Raton, Fla.

Harrell, F. E. J. (2020). **rms: Regression Modeling Strategies. R package version 6.0-0.** (https://CRAN.R-project.org/package=rms.

Harrell, F. E. J., Dupont, C. and others, a. m. (2020). **Hmisc: Harrell Miscellaneous. R package version 4.5-0** (https://CRAN.R-project.org/package=Hmisc.

Hasan, N., Kumar, R. and Kavuru, M. S. (2014). **Lung cancer screening beyond low-dose computed tomography: the role of novel biomarkers**. Lung *192 (5)*, 639-648, doi: 10.1007/s00408-014-9636-z.

Healey, G. F., MacDonald, I. K., Reynolds, C., Allen, J. and Murray, A. (2017). **Tumor-Associated Autoantibodies: Re-Optimization of EarlyCDT-Lung Diagnostic Performance and Its Application to Indeterminate Pulmonary Nodules**. Journal of Cancer Therapy *8*, 506-517, doi: 10.4236/jct.2017.85043.

Heleno, B., Siersma, V. and Brodersen, J. (2018). **Estimation of Overdiagnosis of Lung Cancer in Low-Dose Computed Tomography Screening: A Secondary Analysis of the Danish Lung Cancer Screening Trial**. JAMA Intern Med *178 (10)*, 1420-1422, doi: 10.1001/jamainternmed.2018.3056.

Henschke, C. I., McCauley, D. I., Yankelevitz, D. F., Naidich, D. P., McGuinness, G., Miettinen, O. S., Libby, D. M., Pasmantier, M. W., Koizumi, J., Altorki, N. K. and Smith, J. P. (1999). **Early Lung Cancer Action Project: overall design and findings from baseline screening**. Lancet *354 (9173)*, 99-105, doi: 10.1016/S0140-6736(99)06093-6.

Henschke, C. I., Yip, R., Yankelevitz, D. F. and Miettinen, O. S. (2006). **Computed tomography screening for lung cancer: prospects of surviving competing causes of death**. Clin Lung Cancer *7 (5)*, 323-325, doi: 10.3816/CLC.2006.n.013.

Heuvelmans, M. A., Vonder, M., Rook, M., Groen, H. J. M., De Bock, G. H., Xie, X., Ijzerman, M. J., Vliegenthart, R. and Oudkerk, M. (2019). **Screening for Early Lung Cancer, Chronic Obstructive Pulmonary Disease, and Cardiovascular Disease (the Big-3) Using Low-dose Chest Computed Tomography: Current Evidence and Technical Considerations**. J Thorac Imaging *34 (3)*, 160-169, doi: 10.1097/RTI.0000000000000379.

Hofer, F., Kauczor, H. U. and Stargardt, T. (2018). **Cost-utility analysis of a potential lung cancer screening program for a high-risk population in Germany: A modelling approach**. Lung Cancer *124*, 189-198, doi: 10.1016/j.lungcan.2018.07.036.

Hoffman, R. M., Atallah, R. P., Struble, R. D. and Badgett, R. G. (2020). **Lung Cancer Screening with Low-Dose CT: a Meta-Analysis**. J Gen Intern Med *35 (10)*, 3015-3025, doi: 10.1007/s11606-020-05951-7.

Horeweg, N., Scholten, E. T., de Jong, P. A., van der Aalst, C. M., Weenink, C., Lammers, J. W., Nackaerts, K., Vliegenthart, R., ten Haaf, K., Yousaf-Khan, U. A., Heuvelmans, M. A., Thunnissen, E., Oudkerk, M., Mali, W. and de Koning, H. J. (2014a). **Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers**. Lancet Oncol *15 (12)*, 1342-1350, doi: 10.1016/S1470-2045(14)70387-0.

Horeweg, N., van Rosmalen, J., Heuvelmans, M. A., van der Aalst, C. M., Vliegenthart, R., Scholten, E. T., ten Haaf, K., Nackaerts, K., Lammers, J. W., Weenink, C., Groen, H. J., van Ooijen, P., de Jong, P. A., de Bock, G. H., Mali, W., de Koning, H. J. and Oudkerk, M. (2014b). **Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening**. Lancet Oncol *15 (12)*, 1332-1341, doi: 10.1016/s1470-2045(14)70389-4.

Hori, M., Tanaka, H., Wakai, K., Sasazuki, S. and Katanoda, K. (2016). **Secondhand smoke exposure and risk of lung cancer in Japan: a systematic review and meta-analysis of epidemiologic studies**. Jpn J Clin Oncol *46 (10)*, 942-951, doi: 10.1093/jjco/hyw091.

Horst, C., Dickson, J. L., Tisi, S., Ruparel, M., Nair, A., Devaraj, A. and Janes, S. M. (2020). **Delivering low-dose CT screening for lung cancer: a pragmatic approach**. Thorax *75 (10)*, 831-832, doi: 10.1136/thoraxjnl-2020-215131.

Huang, P., Lin, C. T., Li, Y., Tammemagi, M. C., Brock, M. V., Atkar-Khattra, S., Xu, Y., Hu, P., Mayo, J. R., Schmidt, H., Gingras, M., Pasian, S., Stewart, L., Tsai, S., Seely, J. M., Manos, D., Burrowes, P., Bhatia, R., Tsao, M.-S. and Lam, S. (2019). **Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method**. The Lancet Digital Health *1 (7)*, e353-e362, doi: https://doi.org/10.1016/S2589-7500(19)30159-1.

Hulbert, A., Jusue-Torres, I., Stark, A., Chen, C., Rodgers, K., Lee, B., Griffin, C., Yang, A., Huang, P., Wrangle, J., Belinsky, S. A., Wang, T. H., Yang, S. C., Baylin, S. B., Brock, M. V. and Herman, J. G. (2017). **Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum**. Clin Cancer Res *23 (8)*, 1998-2005, doi: 10.1158/1078-0432.Ccr-16-1371.

Hüsing, A. and Kaaks, R. (2020). **Risk prediction models versus simplified selection criteria to determine eligibility for lung cancer screening: an analysis of German federal-wide survey and incidence data**. Eur J Epidemiol, doi: 10.1007/s10654-020-00657-w.

Infante, M., Cavuto, S., Lutman, F. R., Brambilla, G., Chiesa, G., Ceresoli, G., Passera, E., Angeli, E., Chiarenza, M., Aranzulla, G., Cariboni, U., Errico, V., Inzirillo, F., Bottoni, E., Voulaz, E., Alloisio, M., Destro, A., Roncalli, M., Santoro, A. and Ravasi, G. (2009). **A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial**. Am J Respir Crit Care Med *180 (5)*, 445-453, doi: 10.1164/rccm.200901-0076OC.

Infante, M., Cavuto, S., Lutman, F. R., Passera, E., Chiarenza, M., Chiesa, G., Brambilla, G., Angeli, E., Aranzulla, G., Chiti, A., Scorsetti, M., Navarria, P., Cavina, R., Ciccarelli, M., Roncalli, M., Destro, A., Bottoni, E., Voulaz, E., Errico, V., Ferraroli, G., Finocchiaro, G., Toschi, L., Santoro, A. and Alloisio, M. (2015). **Long-Term Follow-up Results of the DANTE Trial, a Randomized Study of Lung Cancer Screening with Spiral Computed Tomography**. Am J Respir Crit Care Med *191 (10)*, 1166-1175, doi: 10.1164/rccm.201408-1475OC.

International Association for the Study of Lung Cancer (2019). **Nelson Study Shows CT Screening for Nodule Volume Management Reduces Lung Cancer Mortality by 26 Percent in Men**. URL: https://www.iaslc.org/iaslc-news/press-release/nelson-study [as of July 20, 2021].

Isbell, J. M., Deppen, S., Putnam, J. B., Jr., Nesbitt, J. C., Lambright, E. S., Dawes, A., Massion, P. P., Speroff, T., Jones, D. R. and Grogan, E. L. (2011). **Existing general population models inaccurately predict lung cancer risk in patients referred for surgical evaluation**. Ann Thorac Surg *91 (1)*, 227-233; discussion 233, doi: 10.1016/j.athoracsur.2010.08.054.

ISRCTN30604390 (2007). **Spiral computed tomography scanning for the early detection of lung cancer**. URL: http://www.isrctn.com/ISRCTN30604390 [as of March 2, 2020].

Jett, J. R., Peek, L. J., Fredericks, L., Jewell, W., Pingleton, W. W. and Robertson, J. F. (2014). **Audit of the autoantibody test, EarlyCDT(R)-lung, in 1600 patients: an evaluation of its performance in routine clinical practice**. Lung Cancer *83 (1)*, 51-55, doi: 10.1016/j.lungcan.2013.10.008.

Jiang, J. (2007). **Linear and Generalized Linear Mixed Models and Their Applications**, Springer-Verlag, New York.

Jobst, B. J., Weinheimer, O., Trauth, M., Becker, N., Motsch, E., Gross, M. L., Tremper, J., Delorme, S., Eigentopf, A., Eichinger, M., Kauczor, H. U. and Wielputz, M. O. (2018). **Effect of smoking cessation on quantitative computed tomography in smokers at risk in a lung cancer screening population**. Eur Radiol *28 (2)*, 807-815, doi: 10.1007/s00330-017-5030-6.

Jonas, D. E., Reuland, D. S., Reddy, S. M., Nagle, M., Clark, S. D., Weber, R. P., Enyioha, C., Malo, T. L., Brenner, A. T., Armstrong, C., Coker-Schwimmer, M., Middleton, J. C., Voisin, C. and Harris, R. P. (2021). **Screening for Lung Cancer With Low-Dose Computed Tomography: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force**. JAMA *325 (10)*, 971-987, doi: 10.1001/jama.2021.0377.

Kaaks, R., Fortner, R. T., Husing, A., Barrdahl, M., Hopper, M., Johnson, T., Tjonneland, A., Hansen, L., Overvad, K., Fournier, A., Boutron-Ruault, M. C., Kvaskoff, M., Dossus, L., Johansson, M., Boeing, H., Trichopoulou, A., Benetou, V., La Vecchia, C., Sieri, S., Mattiello, A., Palli, D., Tumino, R., Matullo, G., Onland-Moret, N. C., Gram, I. T., Weiderpass, E., Sanchez, M. J., Navarro Sanchez, C., Duell, E. J., Ardanaz, E., Larranaga, N., Lundin, E., Idahl, A., Jirstrom, K., Nodin, B., Travis, R. C., Riboli, E., Merritt, M., Aune, D., Terry, K., Cramer, D. W. and Anderson, K. S. (2018). **Tumor-associated autoantibodies as early detection markers for ovarian cancer? A prospective evaluation**. Int J Cancer *143 (3)*, 515-526, doi: 10.1002/ijc.31335.

Kanne, J. P., Jensen, L. E., Mohammed, T. L., Kirsch, J., Amorosa, J. K., Brown, K., Chung, J. H., Dyer, D. S., Ginsburg, M. E., Heitkamp, D. E., Kazerooni, E. A., Ketai, L. H., Parker, J. A., Ravenel, J. G., Saleh, A. G. and Shah, R. D. (2013). **ACR appropriateness Criteria(R) radiographically detected solitary pulmonary nodule**. J Thorac Imaging *28 (1)*, W1-3, doi: 10.1097/RTI.0b013e31827657c8.

Katki, H. A., Cheung, L. C. and Landy, R. (2019). **Basing eligibility for lung cancer screening on individualized risk calculators should save more lives, but life-expectancy matters**. J Natl Cancer Inst, doi: 10.1093/jnci/djz165.

Katki, H. A., Kovalchik, S. A., Berg, C. D., Cheung, L. C. and Chaturvedi, A. K. (2016). **Development and Validation of Risk Models to Select Ever-Smokers for CT Lung Cancer Screening**. Jama *315 (21)*, 2300-2311, doi: 10.1001/jama.2016.6255.

Katki, H. A., Kovalchik, S. A., Petito, L. C., Cheung, L. C., Jacobs, E., Jemal, A., Berg, C. D. and Chaturvedi, A. K. (2018). **Implications of Nine Risk Prediction Models for Selecting Ever-Smokers for Computed Tomography Lung Cancer Screening**. Ann Intern Med *169 (1)*, 10-19, doi: 10.7326/m17-2701.

Kauczor, H. U., Baird, A. M., Blum, T. G., Bonomo, L., Bostantzoglou, C., Burghuber, O., Cepicka, B., Comanescu, A., Couraud, S., Devaraj, A., Jespersen, V., Morozov, S., Agmon, I. N., Peled, N., Powell, P., Prosch, H., Ravara, S., Rawlinson, J., Revel, M. P., Silva, M., Snoeckx, A., van Ginneken, B., van Meerbeeck, J. P., Vardavas, C., von Stackelberg, O. and Gaga, M. (2020). **ESR/ERS statement paper on lung cancer screening**. Eur Radiol, doi: 10.1007/s00330-020-06727-7.

Kim, A. S., Ko, H. J., Kwon, J. H. and Lee, J. M. (2018). **Exposure to Secondhand Smoke and Risk of Cancer in Never Smokers: A Meta-Analysis of Epidemiologic Studies**. Int J Environ Res Public Health *15 (9)*, doi: 10.3390/ijerph15091981.

Kovalchik, S. A., Tammemagi, M., Berg, C. D., Caporaso, N. E., Riley, T. L., Korch, M., Silvestri, G. A., Chaturvedi, A. K. and Katki, H. A. (2013). **Targeting of low-dose CT screening according to the risk of lung-cancer death**. N Engl J Med *369 (3)*, 245-254, doi: 10.1056/NEJMoa1301851.

Kumar, V., Cohen, J. T., van Klaveren, D., Soeteman, D. I., Wong, J. B., Neumann, P. J. and Kent, D. M. (2018). **Risk-Targeted Lung Cancer Screening: A Cost-Effectiveness Analysis**. Ann Intern Med *168 (3)*, 161-169, doi: 10.7326/m17-1401.

Lam, S., Boyle, P., Healey, G. F., Maddison, P., Peek, L., Murray, A., Chapman, C. J., Allen, J., Wood, W. C., Sewell, H. F. and Robertson, J. F. (2011). **EarlyCDT-Lung: an immunobiomarker test as an aid to early detection of lung cancer**. Cancer Prev Res (Phila) *4 (7)*, 1126-1134, doi: 10.1158/1940-6207.Capr-10-0328.

Landy, R., Cheung, L. C., Berg, C. D., Chaturvedi, A. K., Robbins, H. A. and Katki, H. A. (2019). **Contemporary Implications of U.S. Preventive Services Task Force and Risk-Based Guidelines for Lung Cancer Screening Eligibility in the United States**. Ann Intern Med *171 (5)*, 384-386, doi: 10.7326/M18-3617.

Lee, J., Lim, J., Kim, Y., Kim, H. Y., Goo, J. M., Lee, C. T., Jang, S. H., Lee, W. C., Lee, C. W., An, J. Y., Ko, K. D., Lee, M. K., Choi, K. S., Park, B. and Lee, D. H. (2019). **Development of Protocol for Korean Lung Cancer Screening Project (K-LUCAS) to Evaluate Effectiveness and Feasibility to Implement National Cancer Screening Program**. Cancer Res Treat *51 (4)*, 1285-1294, doi: 10.4143/crt.2018.464.

Lemeshow, S. and Hosmer, D. W., Jr. (1982). **A review of goodness of fit statistics for use in the development of logistic regression models**. Am J Epidemiol *115 (1)*, 92-106, doi: 10.1093/oxfordjournals.aje.a113284.

Lewis, D. R., Check, D. P., Caporaso, N. E., Travis, W. D. and Devesa, S. S. (2014). **US lung cancer trends by histologic type**. Cancer *120 (18)*, 2883-2892, doi: 10.1002/cncr.28749.

Li, K., Husing, A., Sookthai, D., Bergmann, M., Boeing, H., Becker, N. and Kaaks, R. (2015). **Selecting High-Risk Individuals for Lung Cancer Screening: A Prospective Evaluation of Existing Risk Models and Eligibility Criteria in the German EPIC Cohort**. Cancer Prev Res (Phila) *8 (9)*, 777-785, doi: 10.1158/1940-6207.capr-14-0424.

Li, Y., Chen, K. Z. and Wang, J. (2011). **Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people**. Clin Lung Cancer *12 (5)*, 313-319, doi: 10.1016/j.cllc.2011.06.005.

Liang, K.-Y. and Zeger, S. L. (1986). **Longitudinal data analysis using generalized linear models**. Biometrika *73 (1)*, 13-22, doi: 10.1093/biomet/73.1.13.

Liang, W., Zhao, Y., Huang, W., Liang, H., Zeng, H. and He, J. (2018). **Liquid biopsy for early stage lung cancer**. J Thorac Dis *10 (Suppl 7)*, S876-s881, doi: 10.21037/jtd.2018.04.26.

Lopez-Sanchez, L. M., Jurado-Gamez, B., Feu-Collado, N., Valverde, A., Canas, A., Fernandez-Rueda, J. L., Aranda, E. and Rodriguez-Ariza, A. (2017). **Exhaled breath condensate biomarkers for the early diagnosis of lung cancer using proteomics**. Am J Physiol Lung Cell Mol Physiol *313 (4)*, L664-l676, doi: 10.1152/ajplung.00119.2017.

Lortet-Tieulent, J., Soerjomataram, I., Ferlay, J., Rutherford, M., Weiderpass, E. and Bray, F. (2014). **International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women**. Lung Cancer *84 (1)*, 13-22, doi: 10.1016/j.lungcan.2014.01.009.

Lubin, J. H. and Boice, J. D., Jr. (1997). **Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies**. J Natl Cancer Inst *89 (1)*, 49-57, doi: 10.1093/jnci/89.1.49.

Macdonald, I. K., Parsy-Kowalska, C. B. and Chapman, C. J. (2017). **Autoantibodies: Opportunities for Early Cancer Detection**. Trends Cancer *3 (3)*, 198-213, doi: 10.1016/j.trecan.2017.02.003.

MacMahon, H., Naidich, D. P., Goo, J. M., Lee, K. S., Leung, A. N. C., Mayo, J. R., Mehta, A. C., Ohno, Y., Powell, C. A., Prokop, M., Rubin, G. D., Schaefer-Prokop, C. M., Travis, W. D., Van Schil, P. E. and Bankier, A. A. (2017). **Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017**. Radiology *284 (1)*, 228-243, doi: 10.1148/radiol.2017161659.

Maisonneuve, P., Bagnardi, V., Bellomi, M., Spaggiari, L., Pelosi, G., Rampinelli, C., Bertolotti, R., Rotmensz, N., Field, J. K., Decensi, A. and Veronesi, G. (2011). **Lung cancer risk prediction to select smokers for screening CT--a model based on the Italian COSMOS trial**. Cancer Prev Res (Phila) *4 (11)*, 1778-1789, doi: 10.1158/1940-6207.Capr-11-0026.

Marcus, M. W., Duffy, S. W., Devaraj, A., Green, B. A., Oudkerk, M., Baldwin, D. and Field, J. (2019). **Probability of cancer in lung nodules using sequential volumetric screening up to 12 months: the UKLS trial**. Thorax, doi: 10.1136/thoraxjnl-2018-212263.

Marshall, H. M., Zhao, H., Bowman, R. V., Passmore, L. H., McCaul, E. M., Yang, I. A. and Fong, K. M. (2017). **The effect of different radiological models on diagnostic accuracy and lung cancer screening performance**. Thorax *72 (12)*, 1147-1150, doi: 10.1136/thoraxjnl-2016-209624.

Mascalchi, M., Belli, G., Zappa, M., Picozzi, G., Falchini, M., Della Nave, R., Allescia, G., Masi, A., Pegna, A. L., Villari, N. and Paci, E. (2006). **Risk-benefit analysis of X-ray exposure associated with lung cancer screening in the Italung-CT trial**. AJR Am J Roentgenol *187 (2)*, 421-429, doi: 10.2214/AJR.05.0088.

Massion, P. P., Healey, G. F., Peek, L. J., Fredericks, L., Sewell, H. F., Murray, A. and Robertson, J. F. (2017). **Autoantibody Signature Enhances the Positive Predictive Power of Computed Tomography and Nodule-Based Risk Models for Detection of Lung Cancer**. J Thorac Oncol *12 (3)*, 578-584, doi: 10.1016/j.jtho.2016.08.143.

Mazzone, P. J., Silvestri, G. A., Patel, S., Kanne, J. P., Kinsinger, L. S., Wiener, R. S., Soo Hoo, G. and Detterbeck, F. C. (2018). **Screening for Lung Cancer: CHEST Guideline and Expert Panel Report**. Chest *153 (4)*, 954-985, doi: 10.1016/j.chest.2018.01.016.

McCarthy, W. J., Meza, R., Jeon, J. and Moolgavkar, S. H. (2012). **Chapter 6: Lung cancer in never smokers: epidemiology and risk prediction models**. Risk Anal *32 Suppl 1*, S69-84, doi: 10.1111/j.1539-6924.2012.01768.x.

McKay, J. D., Hung, R. J., Han, Y., Zong, X., Carreras-Torres, R., Christiani, D. C., Caporaso, N. E., Johansson, M., Xiao, X., Li, Y., Byun, J., Dunning, A., Pooley, K. A., Qian, D. C., Ji, X., Liu, G., Timofeeva, M. N., Bojesen, S. E., Wu, X., Le Marchand, L., Albanes, D., Bickeboller, H., Aldrich, M. C., Bush, W. S., Tardon, A., Rennert, G., Teare, M. D., Field, J. K., Kiemeney, L. A., Lazarus, P., Haugen, A., Lam, S., Schabath, M. B., Andrew, A. S., Shen, H., Hong, Y. C., Yuan, J. M., Bertazzi, P. A., Pesatori, A. C., Ye, Y., Diao, N., Su, L., Zhang, R., Brhane, Y., Leighl, N., Johansen, J. S., Mellemgaard, A., Saliba, W., Haiman, C. A., Wilkens, L. R., Fernandez-

Somoano, A., Fernandez-Tardon, G., van der Heijden, H. F. M., Kim, J. H., Dai, J., Hu, Z., Davies, M. P. A., Marcus, M. W., Brunnstrom, H., Manjer, J., Melander, O., Muller, D. C., Overvad, K., Trichopoulou, A., Tumino, R., Doherty, J. A., Barnett, M. P., Chen, C., Goodman, G. E., Cox, A., Taylor, F., Woll, P., Bruske, I., Wichmann, H. E., Manz, J., Muley, T. R., Risch, A., Rosenberger, A., Grankvist, K., Johansson, M., Shepherd, F. A., Tsao, M. S., Arnold, S. M., Haura, E. B., Bolca, C., Holcatova, I., Janout, V., Kontic, M., Lissowska, J., Mukeria, A., Ognjanovic, S., Orlowski, T. M., Scelo, G., Swiatkowska, B., Zaridze, D., Bakke, P., Skaug, V., Zienolddiny, S., Duell, E. J., Butler, L. M., Koh, W. P., Gao, Y. T., Houlston, R. S., McLaughlin, J., Stevens, V. L., Joubert, P., Lamontagne, M., Nickle, D. C., Obeidat, M., Timens, W., Zhu, B., Song, L., Kachuri, L., Artigas, M. S., Tobin, M. D., Wain, L. V., SpiroMeta, C., Rafnar, T., Thorgeirsson, T. E., Reginsson, G. W., Stefansson, K., Hancock, D. B., Bierut, L. J., Spitz, M. R., Gaddis, N. C., Lutz, S. M., Gu, F., Johnson, E. O., Kamal, A., Pikielny, C., Zhu, D., Lindstroem, S., Jiang, X., Tyndale, R. F., Chenevix-Trench, G., Beesley, J., Bosse, Y., Chanock, S., Brennan, P., Landi, M. T. and Amos, C. I. (2017). **Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes**. Nat Genet *49 (7)*, 1126-1132, doi: 10.1038/ng.3892.

McMahon, P. M., Kong, C. Y., Bouzan, C., Weinstein, M. C., Cipriano, L. E., Tramontano, A. C., Johnson, B. E., Weeks, J. C. and Gazelle, G. S. (2011). **Cost-effectiveness of computed tomography screening for lung cancer in the United States**. J Thorac Oncol *6 (11)*, 1841-1848, doi: 10.1097/JTO.0b013e31822e59b3.

McWilliams, A., Tammemagi, M. C., Mayo, J. R., Roberts, H., Liu, G., Soghrati, K., Yasufuku, K., Martel, S., Laberge, F., Gingras, M., Atkar-Khattra, S., Berg, C. D., Evans, K., Finley, R., Yee, J., English, J., Nasute, P., Goffin, J., Puksa, S., Stewart, L., Tsai, S., Johnston, M. R., Manos, D., Nicholas, G., Goss, G. D., Seely, J. M., Amjadi, K., Tremblay, A., Burrowes, P., MacEachern, P., Bhatia, R., Tsao, M. S. and Lam, S. (2013). **Probability of cancer in pulmonary nodules detected on first screening CT**. N Engl J Med *369 (10)*, 910-919, doi: 10.1056/NEJMoa1214726.

Melamed, M. R., Flehinger, B. J., Zaman, M. B., Heelan, R. T., Perchick, W. A. and Martini, N. (1984). **Screening for early lung cancer. Results of the Memorial Sloan-Kettering study in New York**. Chest *86 (1)*, 44-53, doi: 10.1378/chest.86.1.44.

Meza, R., Jeon, J., Toumazis, I., Ten Haaf, K., Cao, P., Bastani, M., Han, S. S., Blom, E. F., Jonas, D. E., Feuer, E. J., Plevritis, S. K., de Koning, H. J. and Kong, C. Y. (2021). **Evaluation of the Benefits and Harms of Lung Cancer Screening With Low-Dose Computed Tomography: Modeling Study for the US Preventive Services Task Force**. JAMA *325 (10)*, 988-997, doi: 10.1001/jama.2021.1077.

Montani, F., Marzi, M., Dezi, F., Dama, E., Carletti, R. M., Bonizzi, G., Bertolotti, R., Bellomi, M., Rampinelli, C., Maisonneuve, P., Spaggiari, L., Veronesi, G., Nicassio, F., Di Fiore, P. P. and Bianchi, F. (2015). **miR-Test: A Blood Test for Lung Cancer Early Detection**. Journal of the National Cancer Institute *107*, doi: 10.1093/jnci/djv063.

Moyer, V. A. and USPSTF, U. S. P. S. T. F. (2014). **Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement**. Ann Intern Med *160 (5)*, 330-338, doi: 10.7326/m13-2771.

Nair, V. S., Sundaram, V., Desai, M. and Gould, M. K. (2018). **Accuracy of Models to Identify Lung Nodule Cancer Risk in the National Lung Screening Trial**. Am J Respir Crit Care Med *197 (9)*, 1220-1223, doi: 10.1164/rccm.201708-1632LE.

National Cancer Institute (2020). **Cancer stat facts: lung and bronchus cancer**. URL: https://seer.cancer.gov/statfacts/html/lungb.html [as of December 28, 2020].

National Lung Screening Trial Research Team (2019). **Lung Cancer Incidence and Mortality with Extended Follow-up in the National Lung Screening Trial**. J Thorac Oncol *14 (10)*, 1732-1742, doi: 10.1016/j.jtho.2019.05.044.

Nekolla E. Bundesamt für Strahlenschutz (2020). Personal communication.

Newcombe, R. G. (1998). **Two-sided confidence intervals for the single proportion: comparison of seven methods**. Stat Med *17 (8)*, 857-872, doi: 10.1002/(sici)1097-0258(19980430)17:8<857::aid-sim777>3.0.co;2-e.

Nicholson, A. G., Chansky, K., Crowley, J., Beyruti, R., Kubota, K., Turrisi, A., Eberhardt, W. E., van Meerbeeck, J. and Rami-Porta, R. (2016). **The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revision of the Clinical and Pathologic Staging of Small Cell Lung Cancer in the Forthcoming Eighth Edition of the TNM Classification for Lung Cancer**. J Thorac Oncol *11 (3)*, 300-311, doi: 10.1016/j.jtho.2015.10.008.

O'Keeffe, L. M., Taylor, G., Huxley, R. R., Mitchell, P., Woodward, M. and Peters, S. A. E. (2018). **Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis**. BMJ Open *8 (10)*, e021611, doi: 10.1136/bmjopen-2018-021611.

O'Mahony, J. F., van Rosmalen, J., Mushkudiani, N. A., Goudsmit, F. W., Eijkemans, M. J., Heijnsdijk, E. A., Steyerberg, E. W. and Habbema, J. D. (2015). **The influence of disease risk on the optimal time interval between screens for the early detection of cancer: a mathematical approach**. Med Decis Making *35 (2)*, 183-195, doi: 10.1177/0272989X14528380.

Obuchowski, N. A. (1997a). **Nonparametric analysis of clustered ROC curve data**. Biometrics *53 (2)*, 567-578.

Obuchowski, N. A. (1997b). **ROC analysis of clustered data with R**. URL: https://www.lerner.ccf.org/qhs/software/lib/funcs_clusteredROC.R [as of December 13, 2019].

Østbye, T. and Taylor, D. H. (2004). **The effect of smoking on years of healthy life (YHL) lost among middle-aged and older Americans**. Health Serv Res *39 (3)*, 531-552, doi: 10.1111/j.1475-6773.2004.00243.x.

Oudkerk, M., Liu, S., Heuvelmans, M. A., Walter, J. E. and Field, J. K. (2021). **Lung cancer LDCT screening and mortality reduction - evidence, pitfalls and future perspectives**. Nat Rev Clin Oncol *18 (3)*, 135-151, doi: 10.1038/s41571-020-00432-6.

Paci, E., Puliti, D., Lopes Pegna, A., Carrozzi, L., Picozzi, G., Falaschi, F., Pistelli, F., Aquilini, F., Ocello, C., Zappa, M., Carozzi, F. M., Mascalchi, M. and the, I. W. G. (2017). **Mortality, survival and incidence rates in the ITALUNG randomised lung cancer screening trial**. Thorax *72 (9)*, 825-831, doi: 10.1136/thoraxjnl-2016-209825.

Paci, E., Warwick, J., Falini, P. and Duffy, S. W. (2004). **Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern?** J Med Screen *11 (1)*, 23-27, doi: 10.1177/096914130301100106.

Pan, W. (2001). **Akaike's information criterion in generalized estimating equations**. Biometrics *57 (1)*, 120-125, doi: 10.1111/j.0006-341x.2001.00120.x.

Parkin, D. M., Bray, F., Ferlay, J. and Pisani, P. (2005). **Global cancer statistics, 2002**. CA Cancer J Clin *55 (2)*, 74-108.

Pashayan, N., Duffy, S. W., Pharoah, P., Greenberg, D., Donovan, J., Martin, R. M., Hamdy, F. and Neal, D. E. (2009). **Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: implications of sojourn time on screening**. Br J Cancer *100 (7)*, 1198-1204, doi: 10.1038/sj.bjc.6604973.

Pastorino, U. (2014). **Plasma microRNA Profiling as First Line Screening Test for Lung Cancer Detection: a Prospective Study (BIOMILD)**. URL: https://clinicaltrials.gov/ct2/show/study/NCT02247453 [as of September 21, 2021].

Pastorino, U., Rossi, M., Rosato, V., Marchiano, A., Sverzellati, N., Morosi, C., Fabbri, A., Galeone, C., Negri, E., Sozzi, G., Pelosi, G. and La Vecchia, C. (2012). **Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial**. Eur J Cancer Prev *21 (3)*, 308-315, doi: 10.1097/CEJ.0b013e328351e1b6.

Pastorino, U., Silva, M., Sestini, S., Sabia, F., Boeri, M., Cantarutti, A., Sverzellati, N., Sozzi, G., Corrao, G. and Marchiano, A. (2019a). **Prolonged Lung Cancer Screening Reduced 10-year Mortality in the MILD Trial**. Ann Oncol, doi: 10.1093/annonc/mdz117.

Pastorino, U., Silva, M., Sestini, S., Sabia, F., Boeri, M., Cantarutti, A., Sverzellati, N., Sozzi, G., Corrao, G. and Marchiano, A. (2019b). **Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy**. Ann Oncol *30 (10)*, 1672, doi: 10.1093/annonc/mdz169.

Patz, E. F., Jr. (2006). **Lung cancer screening, overdiagnosis bias, and reevaluation of the Mayo Lung Project**. J Natl Cancer Inst *98 (11)*, 724-725, doi: 10.1093/jnci/djj226.

Patz, E. F., Jr., Greco, E., Gatsonis, C., Pinsky, P., Kramer, B. S. and Aberle, D. R. (2016). **Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial**. Lancet Oncol *17 (5)*, 590-599, doi: 10.1016/s1470-2045(15)00621-x.

Patz, E. F., Jr., Pinsky, P., Gatsonis, C., Sicks, J. D., Kramer, B. S., Tammemagi, M. C., Chiles, C., Black, W. C. and Aberle, D. R. (2014). **Overdiagnosis in low-dose computed tomography screening for lung cancer**. JAMA Intern Med *174 (2)*, 269-274, doi: 10.1001/jamainternmed.2013.12738.

Pinsky, P. F., Bellinger, C. R. and Miller, D. P., Jr. (2018). **False-positive screens and lung cancer risk in the National Lung Screening Trial: Implications for shared decision-making**. J Med Screen *25 (2)*, 110-112, doi: 10.1177/0969141317727771.

Pinsky, P. F., Church, T. R., Izmirlian, G. and Kramer, B. S. (2013). **The National Lung Screening Trial: results stratified by demographics, smoking history, and lung cancer histology**. Cancer *119 (22)*, 3976-3983, doi: 10.1002/cncr.28326.

Pinsky, P. F., Gierada, D. S., Black, W., Munden, R., Nath, H., Aberle, D. and Kazerooni, E. (2015). **Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment**. Ann Intern Med *162 (7)*, 485-491, doi: 10.7326/m14-2086.

Pinsky, P. F., Gierada, D. S., Nath, P. H. and Munden, R. (2017). **Lung Cancer Risk Associated With New Solid Nodules in the National Lung Screening Trial**. AJR Am J Roentgenol *209 (5)*, 1009-1014, doi: 10.2214/ajr.17.18252.

Pistelli, F., Aquilini, F., Falaschi, F., Puliti, D., Ocello, C., Lopes Pegna, A., Carozzi, F. M., Picozzi, G., Zappa, M., Mascalchi, M., Paci, E. and Carrozzi, L. (2019). **Smoking cessation in the ITALUNG lung cancer screening: what does "teachable moment" mean?** Nicotine Tob Res, doi: 10.1093/ntr/ntz148.

Qin, J., Zeng, N., Yang, T., Wan, C., Chen, L., Shen, Y. and Wen, F. (2018). **Diagnostic Value of Autoantibodies in Lung Cancer: a Systematic Review and Meta-Analysis**. Cell Physiol Biochem *51 (6)*, 2631-2646, doi: 10.1159/000495935.

R Core Team (2017). **The R package for statistical computing: R: A language and environment for statistical computing.** (https://www.R-project.org/ Vienna, Austria).

R Core Team (2018). **The R package for statistical computing: R: A language and environment for statistical computing.** (https://www.R-project.org/ Vienna, Austria).

Rami-Porta, R., Bolejack, V., Giroux, D. J., Chansky, K., Crowley, J., Asamura, H., Goldstraw, P., International Association for the Study of Lung Cancer, S., Prognostic Factors Committee, A. B. M. and Participating, I. (2014). **The IASLC lung cancer staging project: the new database to inform the eighth edition of the TNM classification of lung cancer**. J Thorac Oncol *9 (11)*, 1618-1624, doi: 10.1097/JTO.0000000000000334.

Rampinelli, C., De Marco, P., Origgi, D., Maisonneuve, P., Casiraghi, M., Veronesi, G., Spaggiari, L. and Bellomi, M. (2017). **Exposure to low dose computed tomography for lung cancer screening and risk of cancer: secondary analysis of trial data and risk-benefit analysis**. BMJ *356*, j347, doi: 10.1136/bmj.j347.

Robbins, H. A., Berg, C. D., Cheung, L. C., Chaturvedi, A. K. and Katki, H. A. (2019). **Identification of Candidates for Longer Lung Cancer Screening Intervals Following a Negative Low-Dose Computed Tomography Result**. J Natl Cancer Inst *111 (9)*, 996-999, doi: 10.1093/jnci/djz041.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. and Müller, M. (2011). **pROC: an open-source package for R and S+ to analyze and compare ROC curves**. BMC Bioinformatics *12 (1)*, 77, doi: 10.1186/1471-2105-12-77.

Rodríguez, M., Ajona, D., Seijo, L. M., Sanz, J., Valencia, K., Corral, J., Mesa-Guzmán, M., Pío, R., Calvo, A., Lozano, M. D., Zulueta, J. J. and Montuenga, L. M. (2021). **Molecular biomarkers in early stage lung cancer**. Translational Lung Cancer Research *10 (2)*, 1165-1185.

Ru Zhao, Y., Xie, X., de Koning, H. J., Mali, W. P., Vliegenthart, R. and Oudkerk, M. (2011). **NELSON lung cancer screening study**. Cancer Imaging *11 Spec No A*, S79-84, doi: 10.1102/1470-7330.2011.9020.

Rudin, C. M., Avila-Tang, E., Harris, C. C., Herman, J. G., Hirsch, F. R., Pao, W., Schwartz, A. G., Vahakangas, K. H. and Samet, J. M. (2009). **Lung cancer in never smokers: molecular profiles and therapeutic implications**. Clin Cancer Res *15 (18)*, 5646-5661, doi: 10.1158/1078-0432.CCR-09-0377.

Rufibach, K. (2010). **Use of Brier score to assess binary predictions**. J Clin Epidemiol *63 (8)*, 938-939; author reply 939, doi: 10.1016/j.jclinepi.2009.11.009.

Rzyman, W., Szurowska, E. and Adamek, M. (2019). **Implementation of lung cancer screening at the national level: Polish example**. Transl Lung Cancer Res *8 (Suppl 1)*, S95-S105, doi: 10.21037/tlcr.2019.03.09.

Schabath, M. B. and Cote, M. L. (2019). **Cancer Progress and Priorities: Lung Cancer**. Cancer Epidemiol Biomarkers Prev *28 (10)*, 1563-1579, doi: 10.1158/1055-9965.EPI-19-0221.

Schneider, D. and Arenberg, D. (2015). **Competing Mortality in Cancer Screening: A Teachable Moment**. JAMA Internal Medicine *175 (6)*, 896-897, doi: 10.1001/jamainternmed.2015.1232.

Schreuder, A., Schaefer-Prokop, C. M., Scholten, E. T., Jacobs, C., Prokop, M. and van Ginneken, B. (2018). **Lung cancer risk to personalise annual and biennial follow-up computed tomography screening**. Thorax, doi: 10.1136/thoraxjnl-2017-211107.

Schultz, E. M., Sanders, G. D., Trotter, P. R., Patz, E. F., Jr., Silvestri, G. A., Owens, D. K. and Gould, M. K. (2008). **Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules**. Thorax *63 (4)*, 335-341, doi: 10.1136/thx.2007.084731.

Schultz, F. W., Boer, R. and de Koning, H. J. (2012). **Chapter 7: Description of MISCAN-lung, the Erasmus MC Lung Cancer microsimulation model for evaluating cancer control interventions**. Risk Anal *32 Suppl 1*, S85-98, doi: 10.1111/j.1539-6924.2011.01752.x.

Seijo, L. M., Peled, N., Ajona, D., Boeri, M., Field, J. K., Sozzi, G., Pio, R., Zulueta, J. J., Spira, A., Massion, P. P., Mazzone, P. J. and Montuenga, L. M. (2019). **Biomarkers in Lung Cancer Screening: Achievements, Promises, and Challenges**. Journal of Thoracic Oncology *14 (3)*, 343-357, doi: 10.1016/j.jtho.2018.11.023.

Sharma, D., Newman, T. G. and Aronow, W. S. (2015). **Lung cancer screening: history, current perspectives, and future directions**. Arch Med Sci *11 (5)*, 1033-1043, doi: 10.5114/aoms.2015.54859.

Signorell, A. (2020). **DescTools: Tools for Descriptive Statistics. R package version 0.99.38** (

Silva, M., Milanese, G., Pastorino, U. and Sverzellati, N. (2019). **Lung cancer screening: tell me more about post-test risk**. J Thorac Dis *11 (9)*, 3681-3688, doi: 10.21037/jtd.2019.09.28.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2005). **ROCR: visualizing classifier performance in R**. Bioinformatics *21 (20)*, 3940-3941, doi: 10.1093/bioinformatics/bti623.

Smith, L., Brinton, L. A., Spitz, M. R., Lam, T. K., Park, Y., Hollenbeck, A. R., Freedman, N. D. and Gierach, G. L. (2012). **Body mass index and risk of lung cancer among never, former, and current smokers**. J Natl Cancer Inst *104 (10)*, 778-789, doi: 10.1093/jnci/djs179.

Smith, P. J. and Hadgu, A. (1992). **Sensitivity and specificity for correlated observations**. Stat Med *11 (11)*, 1503-1509, doi: 10.1002/sim.4780111108.

Sozzi, G., Boeri, M., Rossi, M., Verri, C., Suatoni, P., Bravi, F., Roz, L., Conte, D., Grassi, M., Sverzellati, N., Marchiano, A., Negri, E., La Vecchia, C. and Pastorino, U. (2014). **Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study**. J Clin Oncol *32 (8)*, 768-773, doi: 10.1200/jco.2013.50.4357.

Spiegelhalter, D. J. (1986). **Probabilistic prediction in patient management and clinical trials**. Stat Med *5 (5)*, 421-433, doi: 10.1002/sim.4780050506.

Stevenson, M., with contributions from, Nunes, T., Heuer, C., Marshall, J., Sanchez, J., Thornton, R., Reiczigel, J., Robison-Cox, J., Sebastiani, P., Solymos, P., Yoshida, K., Jones, G., Pirikahu, S., Firestone, S., Kyle, R., Popp, J., Jay, M. and Reynard, C. (2020). **R Package epiR: Tools for the Analysis of Epidemiological Data** (https://CRAN.R-project.org/package=epiR.

Straatman, H., Peer, P. G. and Verbeek, A. L. (1997). **Estimating lead time and sensitivity in a screening program without estimating the incidence in the screened group**. Biometrics *53 (1)*, 217-229.

Su, T. L., Jaki, T., Hickey, G. L., Buchan, I. and Sperrin, M. (2018). **A review of statistical updating methods for clinical prediction models**. Stat Methods Med Res *27 (1)*, 185-197, doi: 10.1177/0962280215626466.

Sullivan, F. and Schembri, S. (2019). **Early detection of cancer of the lung Scotland (ECLS): trial results [abstract]**. Journal of Thoracic Oncology *14 (10(S))*, S5.

Sullivan, F. M., Farmer, E., Mair, F. S., Treweek, S., Kendrick, D., Jackson, C., Robertson, C., Briggs, A., McCowan, C., Bedford, L., Young, B., Vedhara, K., Gallant, S., Littleford, R., Robertson, J., Sewell, H., Dorward, A., Sarvesvaran, J. and Schembri, S. (2017). **Detection in blood of autoantibodies to tumour antigens as a case-finding method in lung cancer using the EarlyCDT(R)-Lung Test (ECLS): study protocol for a randomized controlled trial**. BMC Cancer *17 (1)*, 187, doi: 10.1186/s12885-017-3175-y.

Sullivan, F. M., Mair, F. S., Anderson, W., Armory, P., Briggs, A., Chew, C., Dorward, A., Haughney, J., Hogarth, F., Kendrick, D., Littleford, R., McConnachie, A., McCowan, C., McMeekin, N., Patel, M., Rauchhaus, P., Ritchie, L., Robertson, C., Robertson, J., Robles-Zurita, J., Sarvesvaran, J., Sewell, H., Sproule, M., Taylor, T., Tello, A., Treweek, S., Vedhara, K., Schembri, S. and Early Diagnosis of Lung Cancer Scotland, T. (2021). **Earlier diagnosis of lung cancer in a randomised trial of an autoantibody blood test followed by imaging**. Eur Respir J *57 (1)*, doi: 10.1183/13993003.00670-2020.

Sun, Y., Li, Z., Li, J., Li, Z. and Han, J. (2016). **A Healthy Dietary Pattern Reduces Lung Cancer Risk: A Systematic Review and Meta-Analysis**. Nutrients *8 (3)*, 134, doi: 10.3390/nu8030134.

Swensen, S. J., Silverstein, M. D., Ilstrup, D. M., Schleck, C. D. and Edell, E. S. (1997). **The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules**. Arch Intern Med *157 (8)*, 849-855.

Tammemägi, M., Ritchie, A. J., Atkar-Khattra, S., Dougherty, B., Sanghera, C., Mayo, J. R., Yuan, R., Manos, D., McWilliams, A. M., Schmidt, H., Gingras, M., Pasian, S., Stewart, L., Tsai, S., Seely, J. M., Burrowes, P., Bhatia, R., Haider, E. A., Boylan, C., Jacobs, C., van Ginneken, B., Tsao, M. S. and Lam, S. (2019a). **Predicting Malignancy Risk of Screen-Detected Lung Nodules-Mean Diameter or Volume**. J Thorac Oncol *14 (2)*, 203-211, doi: 10.1016/j.jtho.2018.10.006.

Tammemägi, M. C. (2018). **Selecting lung cancer screenees using risk prediction models-where do we go from here**. Transl Lung Cancer Res *7 (3)*, 243-253, doi: 10.21037/tlcr.2018.06.03.

Tammemägi, M. C., Church, T. R., Hocking, W. G., Silvestri, G. A., Kvale, P. A., Riley, T. L., Commins, J. and Berg, C. D. (2014). **Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts**. PLoS Med *11 (12)*, e1001764, doi: 10.1371/journal.pmed.1001764.

Tammemägi, M. C., Katki, H. A., Hocking, W. G., Church, T. R., Caporaso, N., Kvale, P. A., Chaturvedi, A. K., Silvestri, G. A., Riley, T. L., Commins, J. and Berg, C. D. (2013). **Selection criteria for lung-cancer screening**. N Engl J Med *368 (8),* 728-736, doi: 10.1056/NEJMoa1211776.

Tammemägi, M. C., Ten Haaf, K., Toumazis, I., Kong, C. Y., Han, S. S., Jeon, J., Commins, J., Riley, T. and Meza, R. (2019b). **Development and Validation of a Multivariable Lung Cancer Risk Prediction Model That Includes Low-Dose Computed Tomography Screening Results: A Secondary Analysis of Data From the National Lung Screening Trial**. JAMA Netw Open *2 (3),* e190204, doi: 10.1001/jamanetworkopen.2019.0204.

Tang, Z. M., Ling, Z. G., Wang, C. M., Wu, Y. B. and Kong, J. L. (2017). **Serum tumor-associated autoantibodies as diagnostic biomarkers for lung cancer: A systematic review and meta-analysis**. PLoS One *12 (7),* e0182117, doi: 10.1371/journal.pone.0182117.

Ten Haaf, K., Bastani, M., Cao, P., Jeon, J., Toumazis, I., Han, S. S., Plevritis, S. K., Blom, E. F., Kong, C. Y., Tammemägi, M. C., Feuer, E. J., Meza, R. and de Koning, H. J. (2020). **A Comparative Modeling Analysis of Risk-Based Lung Cancer Screening Strategies**. J Natl Cancer Inst *112 (5),* 466-479, doi: 10.1093/jnci/djz164.

Ten Haaf, K. and de Koning, H. J. (2015). **Overdiagnosis in lung cancer screening: why modelling is essential**. J Epidemiol Community Health *69 (11),* 1035-1039, doi: 10.1136/jech-2014-204079.

Ten Haaf, K., Jeon, J., Tammemagi, M. C., Han, S. S., Kong, C. Y., Plevritis, S. K., Feuer, E. J., de Koning, H. J., Steyerberg, E. W. and Meza, R. (2017a). **Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study**. PLoS Med *14 (4),* e1002277, doi: 10.1371/journal.pmed.1002277.

Ten Haaf, K., Tammemagi, M. C., Bondy, S. J., van der Aalst, C. M., Gu, S., McGregor, S. E., Nicholas, G., de Koning, H. J. and Paszat, L. F. (2017b). **Performance and Cost-Effectiveness of Computed Tomography Lung Cancer Screening Scenarios in a Population-Based Setting: A Microsimulation Modeling Analysis in Ontario, Canada**. PLoS Med *14 (2),* e1002225, doi: 10.1371/journal.pmed.1002225.

Ten Haaf, K., van der Aalst, C. M. and de Koning, H. J. (2018). **Clinically detected non-aggressive lung cancers: implications for overdiagnosis and overtreatment in lung cancer screening**. Thorax *73 (5),* 407-408, doi: 10.1136/thoraxjnl-2017-211149.

Ten Haaf, K., van Rosmalen, J. and de Koning, H. J. (2015). **Lung cancer detectability by test, histology, stage, and gender: estimates from the NLST and the PLCO trials**. Cancer Epidemiol Biomarkers Prev *24 (1),* 154-161, doi: 10.1158/1055-9965.Epi-14-0745.

Thun, M. J. (2010). **Early landmark studies of smoking and lung cancer**. Lancet Oncol *11 (12),* 1200, doi: 10.1016/S1470-2045(09)70401-2.

Tockman, M. S. (1986). **Survival and Mortality from Lung Cancer in a Screened Population: The Johns Hopkins Study**. CHEST *89 (4),* 324S-325S, doi: 10.1378/chest.89.4_Supplement.324S-a.

Tomonaga, Y., Ten Haaf, K., Frauenfelder, T., Kohler, M., Kouyos, R. D., Shilaih, M., Lorez, M., de Koning, H. J., Schwenkglenks, M. and Puhan, M. A. (2018). **Cost-effectiveness of low-dose CT screening for lung cancer in a European country**

with high prevalence of smoking-A modelling study. Lung Cancer *121*, 61-69, doi: 10.1016/j.lungcan.2018.05.008.

Toumazis, I., Alagoz, O., Leung, A. and Plevritis, S. (2019). **P2.11-02 Individualized Risk-Based Lung Cancer Screening Incorporating Past Screening Findings and Changes in Smoking Behaviors**. Journal of Thoracic Oncology *14 (10)*, S792, doi: 10.1016/j.jtho.2019.08.1702.

Toyoda, Y., Nakayama, T., Ioka, A. and Tsukuma, H. (2008). **Trends in Lung Cancer Incidence by Histological Type in Osaka, Japan**. Japanese Journal of Clinical Oncology *38 (8)*, 534-539, doi: 10.1093/jjco/hyn072.

Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K. R., Yatabe, Y., Beer, D. G., Powell, C. A., Riely, G. J., Van Schil, P. E., Garg, K., Austin, J. H., Asamura, H., Rusch, V. W., Hirsch, F. R., Scagliotti, G., Mitsudomi, T., Huber, R. M., Ishikawa, Y., Jett, J., Sanchez-Cespedes, M., Sculier, J. P., Takahashi, T., Tsuboi, M., Vansteenkiste, J., Wistuba, I., Yang, P. C., Aberle, D., Brambilla, C., Flieder, D., Franklin, W., Gazdar, A., Gould, M., Hasleton, P., Henderson, D., Johnson, B., Johnson, D., Kerr, K., Kuriyama, K., Lee, J. S., Miller, V. A., Petersen, I., Roggli, V., Rosell, R., Saijo, N., Thunnissen, E., Tsao, M. and Yankelewitz, D. (2011). **International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma**. J Thorac Oncol *6 (2)*, 244-285, doi: 10.1097/JTO.0b013e318206a221.

Travis, W. D., Brambilla, E. and Riely, G. J. (2013). **New pathologic classification of lung cancer: relevance for clinical practice and clinical trials**. J Clin Oncol *31 (8)*, 992-1001, doi: 10.1200/JCO.2012.46.9270.

Treskova, M., Aumann, I., Golpon, H., Vogel-Claussen, J., Welte, T. and Kuhlmann, A. (2017). **Trade-off between benefits, harms and economic efficiency of low-dose CT lung cancer screening: a microsimulation analysis of nodule management strategies in a population-based setting**. BMC Med *15 (1)*, 162, doi: 10.1186/s12916-017-0924-3.

van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L. and Steyerberg, E. W. (2019). **Calibration: the Achilles heel of predictive analytics**. BMC Med *17 (1)*, 230, doi: 10.1186/s12916-019-1466-7.

van Iersel, C. A., de Koning, H. J., Draisma, G., Mali, W. P., Scholten, E. T., Nackaerts, K., Prokop, M., Habbema, J. D., Oudkerk, M. and van Klaveren, R. J. (2007). **Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON)**. Int J Cancer *120 (4)*, 868-874, doi: 10.1002/ijc.22134.

van Klaveren, R. J., Oudkerk, M., Prokop, M., Scholten, E. T., Nackaerts, K., Vernhout, R., van Iersel, C. A., van den Bergh, K. A., van 't Westeinde, S., van der Aalst, C., Thunnissen, E., Xu, D. M., Wang, Y., Zhao, Y., Gietema, H. A., de Hoop, B. J., Groen, H. J., de Bock, G. H., van Ooijen, P., Weenink, C., Verschakelen, J., Lammers, J. W., Timens, W., Willebrand, D., Vink, A., Mali, W. and de Koning, H. J. (2009). **Management of lung nodules detected by volume CT scanning**. N Engl J Med *361 (23)*, 2221-2229, doi: 10.1056/NEJMoa0906085.

Veronesi, G., Baldwin, D. R., Henschke, C. I., Ghislandi, S., Iavicoli, S., Oudkerk, M., De Koning, H. J., Shemesh, J., Field, J. K., Zulueta, J. J., Horgan, D., Fiestas Navarrete, L., Valentino Infante, M., Novellis, P., Murray, R. L., Peled, N., Rampinelli, C., Rocco, G., Rzyman, W., Scagliotti, G. V., Tammemagi, M. C., Bertolaccini, L., Triphuridet, N.,

Yip, R., Rossi, A., Senan, S., Ferrante, G., Brain, K., van der Aalst, C., Bonomo, L., Consonni, D., Van Meerbeeck, J. P., Maisonneuve, P., Novello, S., Devaraj, A., Saghir, Z. and Pelosi, G. (2020). **Recommendations for Implementing Lung Cancer Screening with Low-Dose Computed Tomography in Europe**. Cancers (Basel) *12 (6)*, doi: 10.3390/cancers12061672.

Vickers, A. J., Van Calster, B. and Steyerberg, E. W. (2016). **Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests**. BMJ *352*, i6, doi: 10.1136/bmj.i6.

Wahbah, M., Boroumand, N., Castro, C., El-Zeky, F. and Eltorky, M. (2007). **Changing trends in the distribution of the histologic types of lung cancer: a review of 4,439 cases**. Ann Diagn Pathol *11 (2)*, 89-96, doi: 10.1016/j.anndiagpath.2006.04.006.

Walter, S. D. and Day, N. E. (1983). **Estimation of the duration of a pre-clinical disease state using screening data**. Am J Epidemiol *118 (6)*, 865-886, doi: 10.1093/oxfordjournals.aje.a113705.

Warren, G. W., Alberg, A. J., Kraft, A. S. and Cummings, K. M. (2014). **The 2014 Surgeon General's report: "The health consequences of smoking--50 years of progress": a paradigm shift in cancer care**. Cancer *120 (13)*, 1914-1916, doi: 10.1002/cncr.28695.

Weber, M., Yap, S., Goldsbury, D., Manners, D., Tammemagi, M., Marshall, H., Brims, F., McWilliams, A., Fong, K., Kang, Y. J., Caruana, M., Banks, E. and Canfell, K. (2017). **Identifying high risk individuals for targeted lung cancer screening: Independent validation of the PLCOm2012 risk prediction tool**. Int J Cancer *141 (2)*, 242-253, doi: 10.1002/ijc.30673.

Weinheimer, O., Achenbach, T., Heussel, C. and Düber, C. (2011). **Automatic Lung Segmentation in MDCT Images**,

Welch, H. G. and Black, W. C. (2010). **Overdiagnosis in cancer**. J Natl Cancer Inst *102 (9)*, 605-613, doi: 10.1093/jnci/djq099.

White, C. S., Dharaiya, E., Campbell, E. and Boroczky, L. (2017). **The Vancouver Lung Cancer Risk Prediction Model: Assessment by Using a Subset of the National Lung Screening Trial Cohort**. Radiology *283 (1)*, 264-272, doi: 10.1148/radiol.2016152627.

Wielputz, M. O., Bardarova, D., Weinheimer, O., Kauczor, H. U., Eichinger, M., Jobst, B. J., Eberhardt, R., Koenigkam-Santos, M., Puderbach, M. and Heussel, C. P. (2014). **Variation of densitometry on computed tomography in COPD--influence of different software tools**. PLoS One *9 (11)*, e112898, doi: 10.1371/journal.pone.0112898.Wielputz, M. O., Weinheimer, O., Eichinger, M., Wiebel, M., Biederer, J., Kauczor, H. U., Heussel, C. P., Mall, M. A. and Puderbach, M. (2013). **Pulmonary emphysema in cystic fibrosis detected by densitometry on chest multidetector computed tomography**. PLoS One *8 (8)*, e73142, doi: 10.1371/journal.pone.0073142.

Wille, M. M., Dirksen, A., Ashraf, H., Saghir, Z., Bach, K. S., Brodersen, J., Clementsen, P. F., Hansen, H., Larsen, K. R., Mortensen, J., Rasmussen, J. F., Seersholm, N., Skov, B. G., Thomsen, L. H., Tonnesen, P. and Pedersen, J. H. (2016). **Results of the Randomized Danish Lung Cancer Screening Trial with Focus on High-Risk Profiling**. Am J Respir Crit Care Med *193 (5)*, 542-551, doi: 10.1164/rccm.201505-1040OC.

Winkler Wille, M. M., van Riel, S. J., Saghir, Z., Dirksen, A., Pedersen, J. H., Jacobs, C., Thomsen, L. H., Scholten, E. T., Skovgaard, L. T. and van Ginneken, B. (2015). **Predictive Accuracy of the PanCan Lung Cancer Risk Prediction Model - External Validation based on CT from the Danish Lung Cancer Screening Trial**. Eur Radiol *25 (10)*, 3093-3099, doi: 10.1007/s00330-015-3689-0.

Wood, D. E. (2015). **National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines for Lung Cancer Screening**. Thorac Surg Clin *25 (2)*, 185-197, doi: 10.1016/j.thorsurg.2014.12.003.

Wood, D. E., Kazerooni, E., Baum, S. L., Dransfield, M. T., Eapen, G. A., Ettinger, D. S., Hou, L., Jackman, D. M., Klippenstein, D., Kumar, R., Lackner, R. P., Leard, L. E., Leung, A. N., Makani, S. S., Massion, P. P., Meyers, B. F., Otterson, G. A., Peairs, K., Pipavath, S., Pratt-Pozo, C., Reddy, C., Reid, M. E., Rotter, A. J., Sachs, P. B., Schabath, M. B., Sequist, L. V., Tong, B. C., Travis, W. D., Yang, S. C., Gregory, K. M., Hughes, M. and National comprehension cancer, n. (2015). **Lung cancer screening, version 1.2015: featured updates to the NCCN guidelines**. J Natl Compr Canc Netw *13 (1)*, 23-34.

Xiao, F., Liu, D., Guo, Y., Shi, B., Song, Z., Tian, Y. and Liang, C. (2013). **Novel and convenient method to evaluate the character of solitary pulmonary nodule-comparison of three mathematical prediction models and further stratification of risk factors**. PLoS One *8 (10)*, e78271, doi: 10.1371/journal.pone.0078271.

Yang, B., Li, X., Ren, T. and Yin, Y. (2019). **Autoantibodies as diagnostic biomarkers for lung cancer: A systematic review**. Cell Death Discov *5*, 126, doi: 10.1038/s41420-019-0207-1.

Yang, W., Qian, F., Teng, J., Wang, H., Manegold, C., Pilz, L. R., Voigt, W., Zhang, Y., Ye, J., Chen, Q., Han, B. and Written on behalf of the, A. M. E. T. S. C. G. (2018). **Community-based lung cancer screening with low-dose CT in China: Results of the baseline screening**. Lung Cancer *117*, 20-26, doi: 10.1016/j.lungcan.2018.01.003.

Yang, Y., Dong, J., Sun, K., Zhao, L., Zhao, F., Wang, L. and Jiao, Y. (2013). **Obesity and incidence of lung cancer: a meta-analysis**. Int J Cancer *132 (5)*, 1162-1169, doi: 10.1002/ijc.27719.

Yip, R., Jirapatnakul, A., Hu, M., Chen, X., Han, D., Ma, T., Zhu, Y., Salvatore, M. M., Margolies, L. R., Yankelevitz, D. F. and Henschke, C. I. (2021). **Added benefits of early detection of other diseases on low-dose CT screening**. Transl Lung Cancer Res *10 (2)*, 1141-1153, doi: 10.21037/tlcr-20-746.

Youlden, D. R., Cramb, S. M. and Baade, P. D. (2008). **The International Epidemiology of Lung Cancer: geographical distribution and secular trends**. J Thorac Oncol *3 (8)*, 819-831, doi: 10.1097/JTO.0b013e31818020eb.

Yousaf-Khan, U., van der Aalst, C., de Jong, P. A., Heuvelmans, M., Scholten, E., Lammers, J. W., van Ooijen, P., Nackaerts, K., Weenink, C., Groen, H., Vliegenthart, R., Ten Haaf, K., Oudkerk, M. and de Koning, H. (2017a). **Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval**. Thorax *72 (1)*, 48-56, doi: 10.1136/thoraxjnl-2016-208655.

Yousaf-Khan, U., van der Aalst, C., de Jong, P. A., Heuvelmans, M., Scholten, E., Walter, J., Nackaerts, K., Groen, H., Vliegenthart, R., Ten Haaf, K., Oudkerk, M. and de Koning, H. (2017b). **Risk stratification based on screening history: the NELSON lung cancer screening study**. Thorax, doi: 10.1136/thoraxjnl-2016-209892.

Zhang, X., Jiang, N., Wang, L., Liu, H. and He, R. (2017). **Chronic obstructive pulmonary disease and risk of lung cancer: a meta-analysis of prospective cohort studies**. Oncotarget *8 (44)*, 78044-78056, doi: 10.18632/oncotarget.20351.

Zhao, H., Marshall, H. M., Yang, I. A., Bowman, R. V., Ayres, J., Crossin, J., Lau, M., Slaughter, R. E., Redmond, S., Passmore, L., McCaul, E., Courtney, D., Leong, S. C., Windsor, M., Zimmerman, P. V. and Fong, K. M. (2016). **Screen-detected subsolid pulmonary nodules: long-term follow-up and application of the PanCan lung cancer risk prediction model**. Br J Radiol *89 (1060)*, 20160016, doi: 10.1259/bjr.20160016.

8 OWN CONTRIBUTION TO DATA COLLECTION, DATA EVALUATION AND PUBLICATIONS

This thesis presents the results of analyzing data and blood samples collected as part of the German Lung Cancer Screening Intervention (LUSI), a randomized trial (ISRCTN30604390 2007) approved by the Medical Ethics Committee of the University of Heidelberg (073/2001), and by the German Federal Office for Radiation Protection (BfS). The aim of the LUSI trial was to evaluate the lung cancer mortality reduction caused by LDCT-based screening in subjects from the population living in the areas of Heidelberg, Germany and surroundings. Details about the trial can be found in this thesis (Section 2.1) and in previous related publications (Becker et al. 2015; Becker et al. 2012; Becker et al. 2019).

Laboratory measurements of the tumor associated autoantibodies panel EarlyCDT®-Lung, as described in this thesis, were taken at the laboratory of the Division of Cancer Epidemiology of the German Cancer Research Center (DKFZ) by Dr. Theron Johnson and the laboratory staff.

I conducted all statistical analyses central to this thesis.

Note: some tables and figures shown in this thesis have been either adapted or reprinted from previous publications by other authors. These are correctly indicated and citations are included. Permission to reprint these materials has been granted as follows:

- Free to be reproduced for research, private and educational purposes with limited circulation:
    - (Ferlay J 2020a)

- Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) (https://creativecommons.org/licenses/by-nc-nd/4.0/):
    - (Barnes et al. 2016), (Becker et al. 2019), (González Maldonado et al. 2020a), (González Maldonado et al. 2020b), (González Maldonado et al. 2021a), (González Maldonado et al. 2021b)

- Licenses granted through the Copyright Clearance Center's RightsLink® service **(**https://www.copyright.com/publishers/rightslink-scientific/**)**
    - (Schabath and Cote 2019) (Order number: 5183630720782), (Hoffman et al. 2020) (Order number: 5183631344473), (Becker et al. 2015) (Order number: 5183640374458)

Note: the results of this thesis have been previously published in part in the following research articles

Publications related to this thesis

First-author publications

González Maldonado, S., Hynes, L. C., Motsch, E., Heussel, C. P., Kauczor, H. U., Robbins, H. A., Delorme, S. and Kaaks, R. (2021a). **Validation of multivariable lung cancer risk prediction models for the personalized assignment of optimal screening frequency: a retrospective analysis of data from the German Lung Cancer Screening Intervention Trial (LUSI)**. Transl Lung Cancer Res *10 (3)*, 1305-1317, doi: 10.21037/tlcr-20-1173.

González Maldonado, S., Delorme, S., Hüsing, A., Motsch, E., Kauczor, H. U., Heussel, C. P. and Kaaks, R. (2020a). **Evaluation of Prediction Models for Identifying Malignancy in Pulmonary Nodules Detected via Low-Dose Computed Tomography**. JAMA Netw Open *3 (2)*, e1921221, doi: 10.1001/jamanetworkopen.2019.21221.

González Maldonado, S., Motsch, E., Trotter, A., Kauczor, H. U., Heussel, C. P., Hermann, S., Zeissig, S. R., Delorme, S. and Kaaks, R. (2020b). **Overdiagnosis in lung cancer screening: Estimates from the German Lung Cancer Screening Intervention Trial**. Int J Cancer, doi: 10.1002/ijc.33295.

González Maldonado, S., Johnson, T., Motsch, E., Delorme, S. and Kaaks, R. (2021b). **Can autoantibody tests enhance lung cancer screening?-an evaluation of EarlyCDT((R))-Lung in context of the German Lung Cancer Screening Intervention Trial (LUSI)**. Transl Lung Cancer Res *10 (1)*, 233-242, doi: 10.21037/tlcr-20-727.

Co-authored publications

Becker, N., Motsch, E., Trotter, A., Heussel, C. P., Dienemann, H., Schnabel, P. A., Kauczor, H. U., González Maldonado, S., Miller, A. B., Kaaks, R. and Delorme, S. (2019). **Lung cancer mortality reduction by LDCT screening - results from the randomized German LUSI trial**. Int J Cancer, doi: 10.1002/ijc.32486.

Own contribution to these publications

1. This publication details the results of the external validation of selected multivariable lung cancer risk prediction models for the personalized assignment of optimal screening frequency on data from the LUSI trial, as presented in the Introduction (Section 1.3.2), Methods (Sections 2.2.1 and 2.3.1), Results (Section 3.1), and Discussion (Section 4.1) of this thesis. I conducted all statistical analyses and wrote the manuscript.

2. This publication details the results of the external validation of selected prediction models for malignancy detection in pulmonary nodules detected via LDCT in the LUSI trial, as presented in the Introduction (Section 1.3.3), Methods (Section 2.2.2, 2.3.2), Results

(Section 3.2), and Discussion (Section 4.2) of this thesis. I conducted all statistical analyses and wrote the manuscript.

3. This publication details the results on the estimation of overdiagnosis on data from the LUSI trial, as presented in the Introduction (Section 1.2.3.3), the Methods (Section 2.2.3, 2.3.3), Results (Section 3.3), and Discussion (Section 4.3) of this thesis. I conducted all statistical analyses and wrote the manuscript.

4. This publication details the results on the evaluation of the EarlyCDT®-Lung on data and blood samples collected as part of the LUSI trial, presented in the Introduction (Section 1.3.5), Methods (Section 2.2.4, 2.3.4), Results (Section 3.4), and discussion (Section 4.4) of this thesis. I conducted all statistical analyses and wrote the manuscript.

5. This publication details the results on lung cancer mortality reduction by LDCT screening observed in the LUSI trial, as presented in the, Methods (Section 2.1). I conducted part of the statistical analyses and contributed to writing the manuscript.

Other publications and presentations

Posters and presentations related to this thesis

González Maldonado S, Delorme S, Kauczor HU, Heussel CP, Becker N, Kaaks R.
Validation of risk models to predict malignancy of LDCT-detected pulmonary nodules on the German Lung Cancer Screening Trial (LUSI). Deutsches Zentrum für Lungenforschung (DZL) Annual Meeting 2019. Mannheim, 7 February 2019 (Poster).

Publications not related to this thesis

Grafetstätter, M., Hüsing, A., González Maldonado, S., Sookthai, D., Johnson, T., Pletsch-Borba, L., Katzke, V. A., Hoffmeister, M., Bugert, P., Kaaks, R. and Kühn, T. (2019). **Plasma Fibrinogen and sP-Selectin are Associated with the Risk of Lung Cancer in a Prospective Study**. Cancer Epidemiol Biomarkers Prev *28 (7)*, 1221-1227, doi: 10.1158/1055-9965.Epi-18-1285.

Idahl, A., Le Cornet, C., González Maldonado, S., Waterboer, T., Bender, N., Tjønneland, A., Hansen, L., Boutron-Ruault, M. C., Fournier, A., Kvaskoff, M., Boeing, H., Trichopoulou, A., Valanou, E., Peppa, E., Palli, D., Agnoli, C., Mattiello, A., Tumino, R., Sacerdote, C., Onland-Moret, N. C., Gram, I. T., Weiderpass, E., Quirós, J. R., Duell, E. J., Sánchez, M. J., Chirlaque, M. D., Barricarte, A., Gil, L., Brändstedt, J., Riesbeck, K., Lundin, E., Khaw, K. T., Perez-Cornago, A., Gunter, M. J., Dossus, L., Kaaks, R. and Fortner, R. T. (2020). **Serologic markers of Chlamydia trachomatis and other sexually transmitted infections and subsequent ovarian cancer risk: Results from the EPIC cohort**. Int J Cancer *147 (8)*, 2042-2052, doi: 10.1002/ijc.32999.

Jaskulski, S., Jung, A. Y., Huebner, M., Poschet, G., Hell, R., Hüsing, A., González - Maldonado, S., Behrens, S., Obi, N., Becher, H. and Chang-Claude, J. (2020). **Prognostic associations of circulating phytoestrogens and biomarker changes in long-term survivors of postmenopausal breast cancer**. Nutr Cancer *72 (7)*, 1155-1169, doi: 10.1080/01635581.2019.1672762.

Kamlage, B., Neuber, S., Bethan, B., González Maldonado, S., Wagner-Golbs, A., Peter, E., Schmitz, O. and Schatz, P. (2018). **Impact of Prolonged Blood Incubation and Extended Serum Storage at Room Temperature on the Human Serum Metabolome**. Metabolites *8 (1)*, doi: 10.3390/metabo8010006.

Le Cornet, C., Schildknecht, K., Rossello Chornet, A., Fortner, R. T., González Maldonado, S., Katzke, V. A., Kühn, T., Johnson, T., Olek, S. and Kaaks, R. (2020). **Circulating Immune Cell Composition and Cancer Risk: A Prospective Study Using Epigenetic Cell Count Measures**. Cancer Res *80 (9)*, 1885-1892, doi: 10.1158/0008-5472.Can-19-3178.

Mayerle, J., Kalthoff, H., Reszka, R., Kamlage, B., Peter, E., Schniewind, B., González Maldonado, S., Pilarsky, C., Heidecke, C. D., Schatz, P., Distler, M., Scheiber, J. A., Mahajan, U. M., Weiss, F. U., Grützmann, R. and Lerch, M. M. (2018). **Metabolic biomarker signature to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis**. Gut *67 (1)*, 128-137, doi: 10.1136/gutjnl-2016-312432.

Müller, O. J., Heckmann, M. B., Ding, L., Rapti, K., Rangrez, A. Y., Gerken, T., Christiansen, N., Rennefahrt, U. E. E., Witt, H., González Maldonado, S., Ternes, P., Schwab, D. M., Ruf, T., Hille, S., Remes, A., Jungmann, A., Weis, T. M., Kreußer, J. S., Gröne, H. J., Backs, J., Schatz, P., Katus, H. A. and Frey, N. (2019). **Comprehensive plasma and tissue profiling reveals systemic metabolic alterations in cardiac hypertrophy and failure**. Cardiovasc Res *115 (8)*, 1296-1305, doi: 10.1093/cvr/cvy274.

Padberg, I., Peter, E., González-Maldonado, S., Witt, H., Mueller, M., Weis, T., Bethan, B., Liebenberg, V., Wiemer, J., Katus, H. A., Rein, D. and Schatz, P. (2014). **A new metabolomic signature in type-2 diabetes mellitus and its pathophysiology**. PLoS One *9 (1)*, e85082, doi: 10.1371/journal.pone.0085082.

Pletsch-Borba, L., Grafetstätter, M., Hüsing, A., González Maldonado, S., Kloss, M., Groß, M. L., Johnson, T., Sookthai, D., Bugert, P., Kaaks, R. and Kühn, T. (2019a). **Author Correction: Biomarkers of vascular injury in relation to myocardial infarction risk: A population-based study**. Sci Rep *9 (1)*, 8037, doi: 10.1038/s41598-019-44430-w.

Pletsch-Borba, L., Grafetstätter, M., Hüsing, A., González Maldonado, S., Kloss, M., Groß, M. L., Johnson, T., Sookthai, D., Bugert, P., Kaaks, R. and Kühn, T. (2019b). **Biomarkers of vascular injury in relation to myocardial infarction risk: A population-based study**. Sci Rep *9 (1)*, 3004, doi: 10.1038/s41598-018-38259-y.

Pletsch-Borba, L., Grafetstätter, M., Hüsing, A., Johnson, T., González Maldonado, S., Groß, M. L., Kloss, M., Hoffmeister, M., Bugert, P., Kaaks, R. and Kühn, T. (2020). **Vascular injury biomarkers and stroke risk: A population-based study**. Neurology *94 (22)*, e2337-e2345, doi: 10.1212/wnl.0000000000009391.

Pletsch-Borba, L., Watzinger, C., Turzanski Fortner, R., Katzke, V., Schwingshackl, L., Sowah, S. A., Hüsing, A., Johnson, T., Groß, M. L., González Maldonado, S., Hoffmeister, M., Bugert, P., Kaaks, R., Grafetstätter, M. and Kühn, T. (2019c). **Biomarkers of Vascular Injury and Type 2 Diabetes: A Prospective Study, Systematic Review and Meta-Analysis**. J Clin Med *8 (12)*, doi: 10.3390/jcm8122075.

126

Rodrigues, G., Lukka, H., Warde, P., Brundage, M., Souhami, L., Crook, J., Cury, F., Catton, C., Mok, G., Martin, A. G., Vigneault, E., Morris, J., Warner, A., González Maldonado, S., Pickles, T. and Genitourinary Radiation Oncologists of, C. (2013b). **The prostate cancer risk stratification (ProCaRS) project: recursive partitioning risk stratification analysis**. Radiother Oncol *109 (2)*, 204-210, doi: 10.1016/j.radonc.2013.07.020.

Schubel, R., Nonnenmacher, T., Sookthai, D., González Maldonado, S., Sowah, S. A., von Stackelberg, O., Schlett, C. L., Grafetstatter, M., Nabers, D., Johnson, T., Kirsten, R., Ulrich, C. M., Kaaks, R., Kauczor, H. U., Kuhn, T. and Nattenmuller, J. (2019). **Similar Weight Loss Induces Greater Improvements in Insulin Sensitivity and Liver Function among Individuals with NAFLD Compared to Individuals without NAFLD**. Nutrients *11 (3)*, doi: 10.3390/nu11030544.

9   SUPPLEMENTARY MATERIAL

Supplementary Table 1. Baseline characteristics of eligible participants and their pulmonary nodules

	Participants eligible for the LCRAT+CT model				Participants eligible for the Polynomial model			
	No LC	LC in any round[†]	P	Total	No LC	LC at T_1[‡]	p	Total
N	1482	24		1506	1878	11		1889
Sex (M/F)	942/540 (63.6/36.4)	18/6 (75.0/25.0)	0.346	960/546 (63.7/36.3)	1229/649 (65.4/34.6)	9/2 (81.8/18.2)	0.411	1238/651 (65.5/34.5)
Age [median, range]	56.70 [50.30, 71.80]	60.20 [54.60, 70.00]	<0.001	56.80 [50.30, 71.80]	56.80 [50.30, 71.90]	60.20 [54.60, 69.40]	0.013	56.80 [50.30, 71.90]
BMI (mean, SD)	26.94 (4.20)	25.73 (3.69)	0.163	26.92 (4.19)	26.91 (4.14)	25.95 (3.28)	0.444	26.90 (4.14)
COPD or emphysema (no/yes)	1465/17 (98.9/1.1)	23/1 (95.8/4.2)	0.687	1488/18 (98.8/1.2)	1636/242 (87.1/12.9)	8/3 (72.7/27.3)	0.334	1644/245 (87.0/13.0)
Asbestos exposure	1482 (100.0)	24 (100.0)	-	1506 (100.0)	1819 (96.9)	11 (100.0)	1.000	1830 (96.9)
Education			0.989				0.916	
< 12th grade	119 (8.0)	2 (8.3)		121 (8.0)	146 (7.8)	1 (9.1)		147 (7.8)
High school graduate	1 (0.1)	0 (0.0)		1 (0.1)	2 (0.1)	0 (0.0)		2 (0.1)
Post high school, no college	795 (53.6)	14 (58.3)		809 (53.7)	1014 (54.0)	7 (63.6)		1021 (54.0)
Associate degree / some college	305 (20.6)	4 (16.7)		309 (20.5)	392 (20.9)	1 (9.1)		393 (20.8)
Graduate school	262 (17.7)	4 (16.7)		266 (17.7)	324 (17.3)	2 (18.2)		326 (17.3)
Emphysema in scan (no/yes)	943/539 (63.6/36.4)	11/13 (45.8/54.2)	0.114	954/552 (63.3/36.7)	1159/719 (61.7/38.3)	5/6 (45.5/54.5)	0.427	1164/725 (61.6/38.4)
Consolidation in scan (no/yes)	1475/7 (99.5/0.5)	24/0 (100.0/0.0)	1.000	1499/7 (99.5/0.5)	1867/11 (99.4/0.6)	11/0 (100.0/0.0)	1.000	1878/11 (99.4/0.6)

As published in (González Maldonado et al. 2021a), reprinted with permission.

Supplemental Table 1 (continued). Baseline characteristics of eligible participants and their pulmonary nodules.

	Participants eligible for the LCRAT+CT model				Participants eligible for the Polynomial model			
	No LC	LC in any round[†]	P	Total	No LC	LC at T_1[‡]	p	Total
Former smokers	574	6		580	714	3		717
Years of smoking (SD)	18.27 (4.37)	20.00 (2.74)	0.333	18.28 (4.36)	18.34 (4.40)	19.17 (2.89)	0.745	18.34 (4.39)
Years since quitting (SD)	4.90 (2.76)	4.33 (2.70)	0.620	4.89 (2.76)	4.85 (2.76)	5.50 (3.46)	0.684	4.85 (2.76)
Cigarettes per day (SD)	26.24 (12.13)	27.50 (17.32)	0.801	26.25 (12.17)	26.60 (12.29)	30.83 (7.64)	0.552	26.62 (12.27)
Pack-years (SD)	24.00 (13.07)	28.33 (20.81)	0.423	24.05 (13.15)	24.41 (13.14)	28.85 (3.77)	0.559	24.43 (13.11)
Current smokers	908	18		926	1164	8		1172
Years of smoking (SD)	34.87 (4.89)	37.50 (5.94)	0.025	34.92 (4.92)	35.03 (4.91)	38.75 (3.54)	0.033	35.05 (4.91)
Years since quitting (SD)	0.00 (0.00)	0.00 (0.00)	-	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-	0.00 (0.00)
Cigarettes per day (SD)	20.40 (8.42)	21.94 (12.11)	0.444	20.43 (8.50)	20.55 (8.60)	25.62 (13.35)	0.098	20.59 (8.64)
Pack-years (SD)	35.85 (16.02)	41.56 (27.02)	0.141	35.96 (16.30)	36.24 (16.39)	50.55 (29.97)	0.015	36.34 (16.54)
With at least one NCN					895	6		901
Longest diameter (SD)					5.75 (3.62)	6.15 (2.40)	0.788	5.75 (3.62)
Perpendicular diameter (SD)					3.85 (1.75)	4.47 (1.53)	0.386	3.85 (1.74)
In upper lobe (%)					524 (58.5)	2 (33.3)	0.405	526 (58.4)
Solid (%)					853 (95.3)	6 (100.0)	1.000	859 (95.3)
Spiculated (%)					37 (4.1)	0 (0.0)	1.000	37 (4.1)

† Lung cancer screen-detected at any annual screening appointment. Not necessarily screen-detected at next-screen following a negative one.
‡ ‡ Lung cancer screen-detected at round 2 (T1) or diagnosed in between round 2 (T1) and round 3(T2).
Abbreviations: SD: standard deviation; BMI: body mass index; COPD: chronic obstructive pulmonary disease; LC: lung cancer; NCN: **non-**calcified nodule.

As published in (González Maldonado et al. 2021a), reprinted with permission.

Supplementary Figure 1. ROC curves for the LCRAT (1-year), LCRAT+CT models†

Supplementary Figure 2. ROC curves for the Polynomial model‡

Supplementary Figure 3. ROC curves for the Polynomial model applied to data from participants eligible for the LCRAT+CT model

Supplementary Figure 4. ROC curves for the Polynomial model and the Patz criterion applied to T0 data
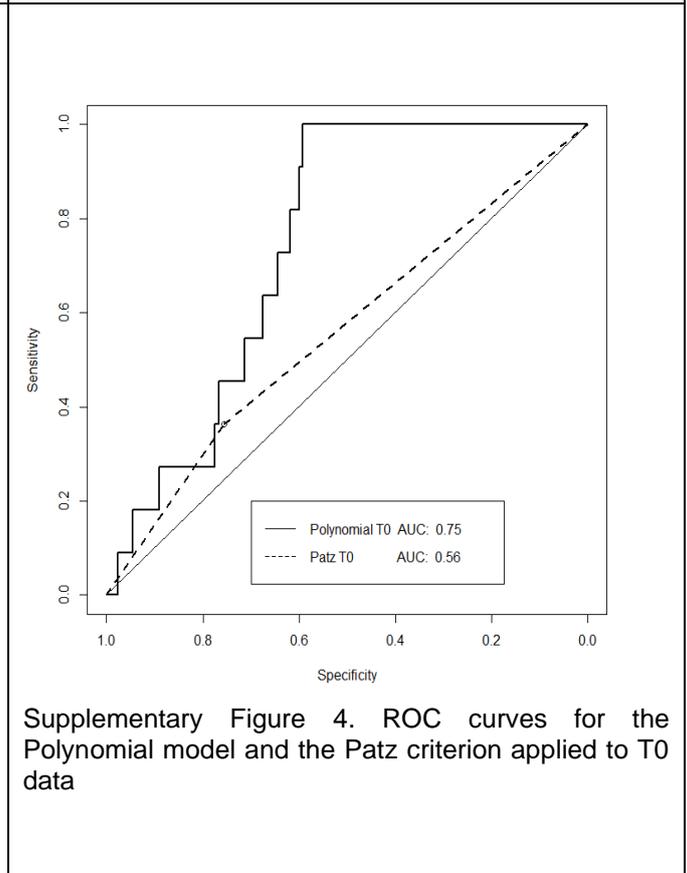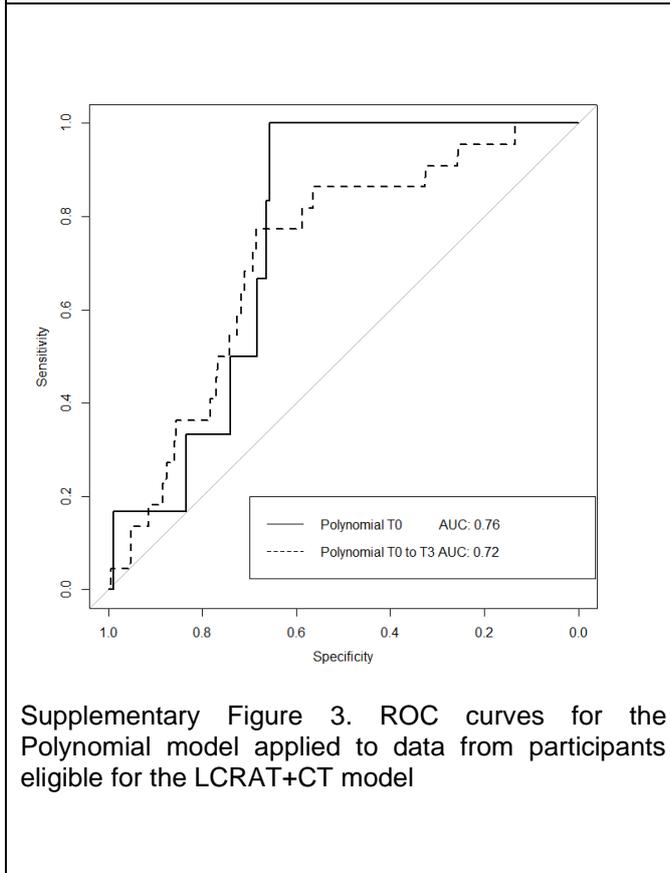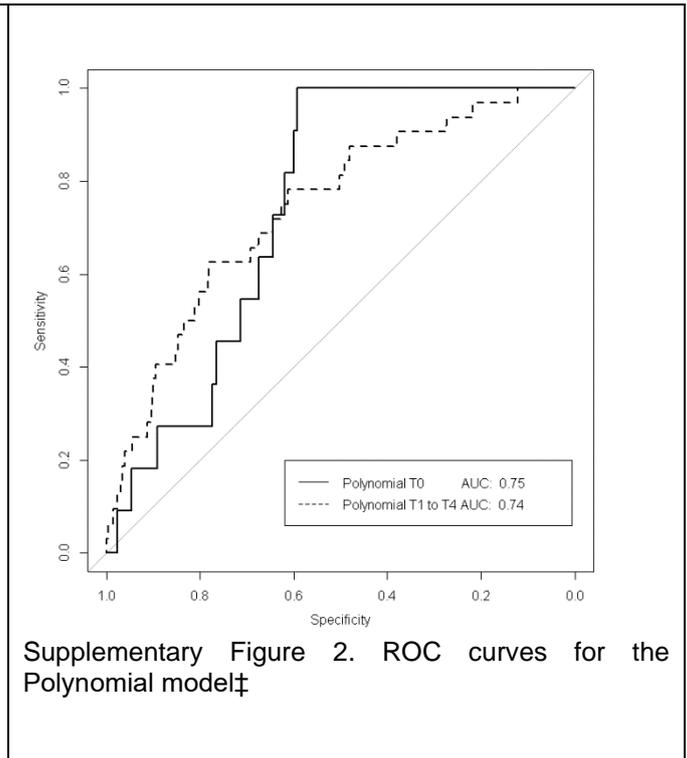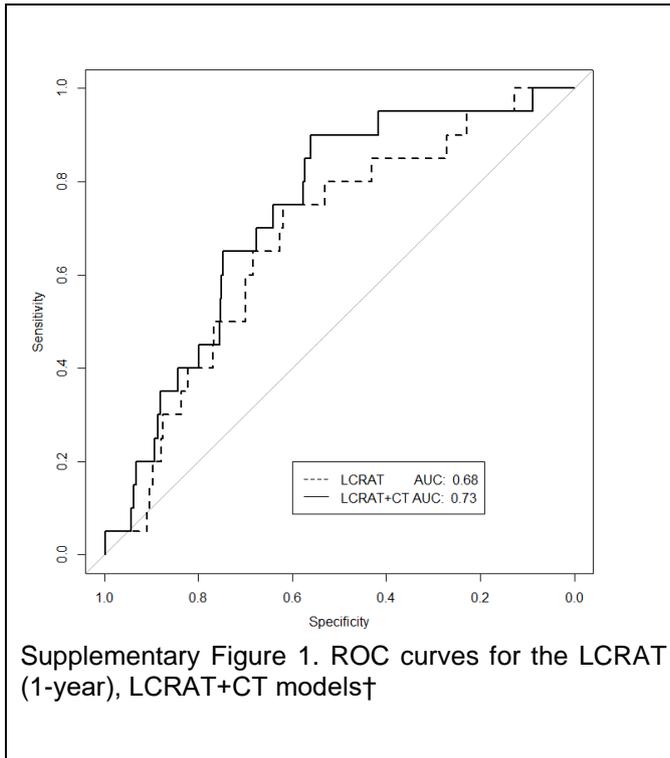
Abbreviations: LCRAT: Lung Cancer risk Assessment Tool; AUC: Area under the Curve.
† Applied to data from LCRAT+CT-eligible participants from T0 to T3.
‡ Applied to data from Polynomial model eligible participants from T0 or T1 to T4

As published in (González Maldonado et al. 2021a), reprinted with permission.

Supplementary Table 2. Potential effect of risk thresholds from the Polynomial model in participants from all screening rounds of the LUSI Trial, eligible for the LCRAT+CT model

Percentile of risk	Polynomial risk (T0)	Candidates for Longer Interval N (%)	Delayed Cancers N (%; 95% CI)	False positives avoided N (%; 95% CI)	Indeterminates avoided N (%; 95% CI)
10th	r ≤≤ 0.048 %	120 (10)	0 (0; 1.5, 48.3)	0 (0; 10.8, 94.5)	1 (33.3; 1.8, 87.5)
20th	r ≤ 0.071%	239 (20)	0 (0; 1.5, 48.3)	0 (0; 10.8, 94.5)	1 (33.3; 1.8, 87.5)
30th	r ≤ 0.091%	358 (30)	0 (0; 1.5, 48.3)	0 (0; 10.8, 94.5)	1 (33.3; 1.8, 87.5)
40th	r ≤ 0.12%	478 (40)	0 (0; 1.5, 48.3)	0 (0; 10.8, 94.5)	1 (33.3; 1.8, 87.5)
50th	r ≤ 0.15%	597 (50)	0 (0; 1.5, 48.3)	0 (0; 10.8, 94.5)	1 (33.3; 1.8, 87.5)
60th	r ≤ 0.20%	716 (60)	0 (0; 1.5, 48.3)	0 (0; 10.8, 94.5)	1 (33.3; 1.8, 87.5)
70th	r ≤ 0.25%	836 (70)	3 (50; 13.9, 86.1)	0 (0; 10.8, 94.5)	2 (66.7; 12.5, 98.2)
80th	r ≤ 0.35%	955 (80)	4 (66.7; 24.1, 94)	0 (0; 10.8, 94.5)	2 (66.7; 12.5, 98.2)
90th	r ≤ 0.55%	1074 (90)	5 (83.3; 36.5, 99.1)	1 (100; 5.5, 100)	2 (66.7; 12.5, 98.2)
100th	r ≤ 3.1%	1194 (100)	6 (100; 51.7, 100)	1 (100; 5.5, 100)	3 (100; 31, 100)
Percentile of risk	Polynomial risk (T0-T3)	Candidates for Longer Interval N (%)	Delayed Cancers N (%; 95% CI)	False positives avoided N (%; 95% CI)	Indeterminates avoided N (%; 95% CI)
10th	r ≤ 0.048	476 (10)	0 (0; 0.4,18.5)	0 (0; 3.2, 69)	1 (16.7; 0.9, 63.5)
20th	r ≤ 0.071	950 (20)	1 (4.5; 0.2, 24.9)	0 (0; 3.2, 69)	1 (16.7; 0.9, 63.5)
30th	r ≤ 0.090	1428 (30)	2 (9.1; 1.6, 30.6)	0 (0; 3.2, 69)	1 (16.7; 0.9, 63.5)
40th	r ≤ 0.12	1899 (40)	3 (13.6; 3.6, 36)	0 (0; 3.2, 69)	2 (33.3; 6, 75.9)
50th	r ≤ 0.14	2378 (50)	3 (13.6; 3.6, 36)	0 (0; 3.2, 69)	2 (33.3; 6, 75.9)
60th	r ≤ 0.18	2851 (60)	5 (22.7; 8.7, 45.8)	1 (33.3; 1.8, 87.5)	3 (50; 13.9, 86.1)
70th	r ≤ 0.23	3325 (70)	7 (31.8; 14.7, 54.9)	1 (33.3; 1.8, 87.5)	4(66.7; 24.1, 94)
80th	r ≤ 0.32	3798 (80)	14 (63.6; 40.8, 82)	1 (33.3; 1.8, 87.5)	5 (83.3; 36.5, 99.1)
90th	r ≤ 0.49	4273 (90)	18 (81.8; 59, 94)	2 (66.7; 12.5, 98.2)	6 (100; 51.7, 100)
100th	r ≤ 7.6	4748 (100)	22(100; 81.5, 99.6)	3 (100; 31, 96.8)	6 (100; 51.7, 100)

Abbreviation: LCRAT: Lung Cancer risk Assessment Tool

As published in (González Maldonado et al. 2021a), reprinted with permission.
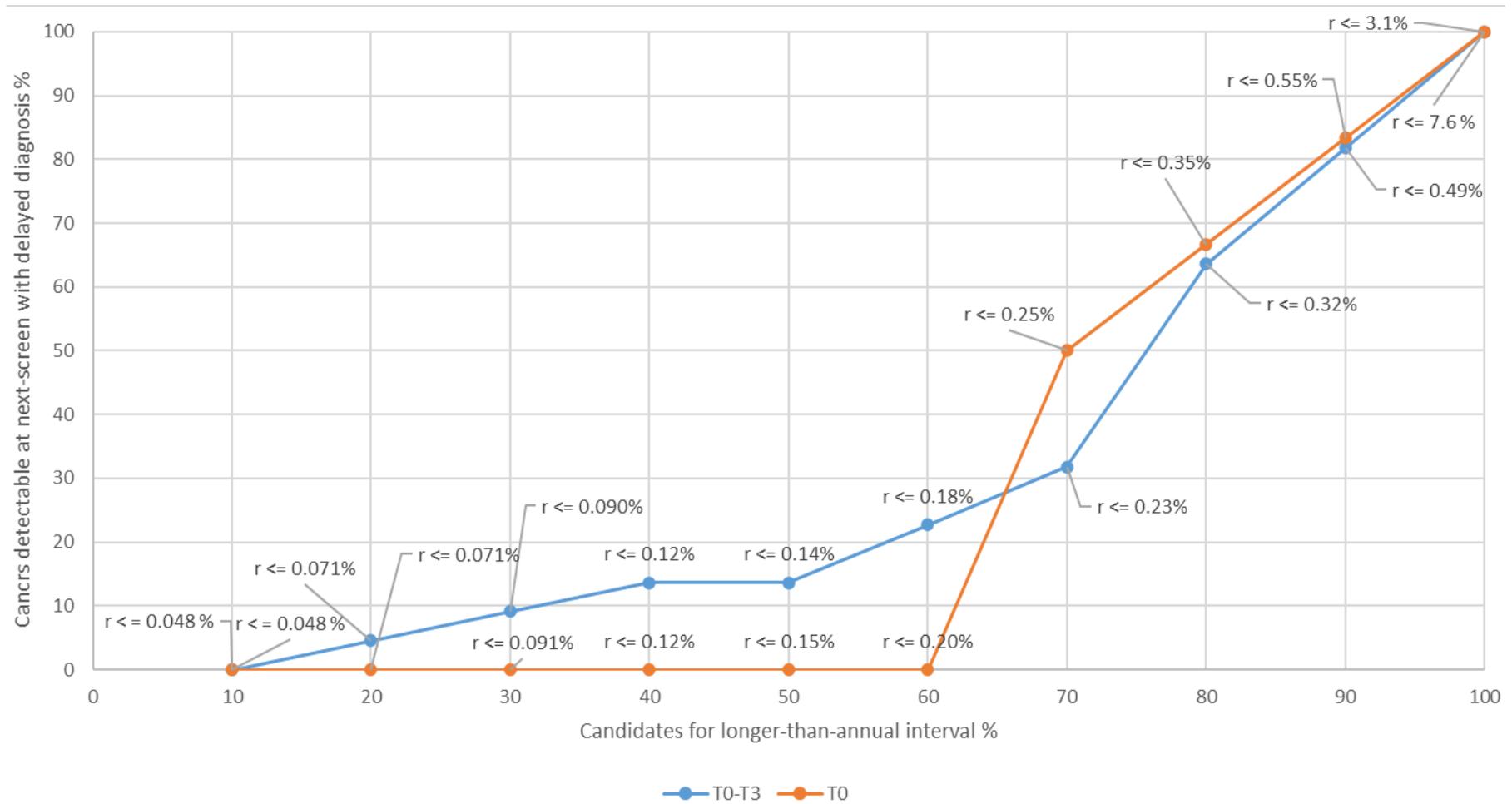
Supplementary Figure 5. Potential effect of risk thresholds from the Polynomial model in participants from all screening rounds of the LUSI Trial, eligible for the LCRAT+CT model

Abbreviations: LCRAT: Lung Cancer risk Assessment Tool.

As published in (González Maldonado et al. 2021a), reprinted with permission.

Supplementary Table 3. Observed incidence and mean predicted risk from the LCRAT, LCRAT+CT in subjects eligible for the LCRAT+CT model

Screening round	Observed incidence at next scan (%)	Mean predicted risk (%)		BS; Sp-Z (p-value) †	
		LCRAT	LCRAT + CT	LCRAT	LCRAT + CT
T0 to T3	20/4904 (0.41)	0.22	0.21	0.004;2.88 (0.004)	0.004; 3.09 (0.002)
First round (T0)	6/1194 (0.50)	0.19	0.18	0.005; 2.44 (0.015)	0.005; 2.68 (0.007)
Second round (T1)	3/1220 (0.25)	0.21	0.20	0.002; 0.27 (0.787)	0.002; 0.39 (0.693)
Third round (T2)	5/1262 (0.40)	0.23	0.22	0.004; 1.28 (0.201)	0.004; 1.36 (0.175)
Fourth round (T3)	6/1228 (0.49)	0.24	0.24	0.005; 1.82 (0.069)	0.005; 1.82 (0.069)
T1 to T3	14/3710 (0.38)	0.23	0.22	0.004; 1.97 (0.049)	0.004; 2.10 (0.036)

As published in (González Maldonado et al. 2021a), reprinted with permission.

Supplementary Table 4. Observed incidence and mean predicted risk from the Polynomial model

Screening round	Observed incidence in [Tn, Tn+1) ‡ (%)	Mean predicted risk (%)	BS; Sp-Z (p-value) †
		Polynomial model	
First round (T0)	11/1889 (0.58)	0.31	0.006; 2.14 (0.032)
Second round (T1)	11/1737 (0.63)	0.33	0.006; 2.28 (0.023)
Third round (T2)	10/1728 (0.57)	0.32	0.006; 1.91 (0.055)
Fourth round (T3)	10/1726 (0.58)	0.36	0.006; 1.71 (0.088)
Fifth round (T4)	1/1754(0.057)	0.32	0.006; -1.95 (0.052)
T1 to T4	32/6966 0.46	0.33	0.005; 1.98 (0.048)

†Brier Score, Spiegelhalter's Z-Score and Spiegelhalter's test p-value. ‡ For the purpose of evaluating model calibration, the observed lung cancer incidence in the subset of participants eligible for the Polynomial model was calculated only among those with a valid prediction from the model (i.e., after removing those with NA for risks).

As published in (González Maldonado et al. 2021a), reprinted with permission.

Supplementary Table 5. Nodule count by size, screening round (prevalent vs. incident), malignancy status, and nodule type

| Nodule count (%) | First seen on the prevalence round | | | | First seen on any incidence round | | | | First seen on any round | | | |
| | benign | | malignant | | benign | | malignant | | benign | | malignant | |
Nodule size (largest diameter in mm)	solid	sub-solid	solid	sub-solid	solid	sub-solid	solid	sub-solid	solid	sub-solid	solid	sub-solid
	2772	79	19	13	875	114	21	10	3647	193	40	23
≤ 3	213 (7.7)	3 (3.8)	0 (0.0)	0 (0.0)	16 (1.8)	0 (0.0)	0 (0.0)	0 (0.0)	229 (6.3)	3 (1.6)	0 (0.0)	0 (0.0)
> 3 and ≤ 4	1038 (37.4)	13 (16.5)	1 (5.3)	0 (0.0)	121 (13.8)	1 (0.9)	0 (0.0)	1 (10.0)	1159 (31.8)	14 (7.3)	1 (2.5)	1 (4.3)
> 4 and ≤ 5	649 (23.4)	5 (6.3)	1 (5.3)	0 (0.0)	136 (15.5)	10 (8.8)	0 (0.0)	0 (0.0)	785 (21.5)	15 (7.8)	1 (2.5)	0 (0.0)
> 5 and ≤ 6	366 (13.2)	14 (17.7)	2 (10.5)	0 (0.0)	221 (25.3)	19 (16.7)	1 (4.8)	0 (0.0)	587 (16.1)	33 (17.1)	3 (7.5)	0 (0.0)
> 6 and ≤ 8	368 (13.3)	24 (30.4)	5 (26.3)	0 (0.0)	190 (21.7)	16 (14.0)	3 (14.3)	1 (10.0)	558 (15.3)	40 (20.7)	8 (20.0)	1 (4.3)
> 8 and ≤ 10	83 (3.0)	8 (10.1)	2 (10.5)	3 (23.1)	73 (8.3)	17 (14.9)	4 (19.0)	1 (10.0)	156 (4.3)	25 (13.0)	6 (15.0)	4 (17.4)
> 10	55 (2.0)	12 (15.2)	8 (42.1)	10 (76.9)	118 (13.5)	51 (44.7)	13 (61.9)	7 (70.0)	173 (4.7)	63 (32.6)	21 (52.5)	17 (73.9)

As published in (González Maldonado et al. 2020a), reprinted with permission.

Supplementary Table 6. Observed versus predicted nodule malignancy rates by deciles of predicted risk (prevalence round)

Prediction Model	Observed and predicted values	Deciles of predicted nodule malignancy (prevalence round)									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
PanCan 1b	Decile cut-offs %	(0.01-0.07)	(0.07-0.12)	(0.12-0.16)	(0.16-0.25)	(0.25-0.32)	(0.32-0.47)	(0.47-0.68)	(0.68-1.1)	(1.1-2.25)	(2.25-79.87)
	Nodule count	318	275	289	281	284	316	256	290	285	289
	Malignant nodule count	0	1	0	0	0	1	0	1	3	25
	Observed malignancy %	0.00	0.36	0.00	0.00	0.00	0.32	0.00	0.34	1.05	8.00
	Predicted malignancy %	0.05	0.10	0.14	0.20	0.28	0.40	0.58	0.87	1.52	8.42
PanCan 2b	Decile cut-offs %	(0.01-0.05)	(0.05-0.09)	(0.09-0.12)	(0.12-0.17)	(0.17-0.24)	(0.24-0.34)	(0.34-0.51)	(0.51-0.83)	(0.83-1.54)	(1.54-68.07)
	Nodule count	289	288	290	286	289	288	288	288	288	289
	Malignant nodule count	0	0	0	1	0	0	0	2	3	26
	Observed malignancy %	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.69	1.04	9.00
	Predicted malignancy %	0.04	0.07	0.10	0.15	0.20	0.28	0.42	0.65	1.12	6.57
PanCan MD	Decile cut-offs %	(0-0)	(0-0)	(0-0)	(0-0.01)	(0.01-0.02)	(0.02-0.05)	(0.05-0.11)	(0.11-0.26)	(0.26-0.74)	(0.74-70.76)
	Nodule count	290	287	288	288	289	288	288	288	288	289
	Malignant nodule count	0	0	0	0	0	1	0	4	1	26
	Observed malignancy %	0.00	0.00	0.00	0.00	0.00	0.35	0.00	1.39	0.35	9.00
	Predicted malignancy %	0.00	0.00	0.00	0.01	0.02	0.03	0.07	0.18	0.45	5.16
PanCan VOL	Decile cut-offs %	(0-0.02)	(0.02-0.04)	(0.04-0.06)	(0.06-0.09)	(0.09-0.13)	(0.13-0.2)	(0.2-0.31)	(0.31-0.54)	(0.54-1.15)	(1.15-61.11)
	Nodule count	289	288	288	288	289	288	288	288	288	289
	Malignant nodule count	0	0	0	0	0	1	0	3	3	25
	Observed malignancy %	0.0	0.0	0.0	0.0	0.0	0.4	0.0	1.0	1.0	8.7
	Predicted malignancy %	0.0	0.0	0.1	0.1	0.1	0.2	0.3	0.4	0.8	5.5

As published in (González Maldonado et al. 2020a), reprinted with permission.

Supplemental Table 6 (continued). Observed versus predicted nodule malignancy rates by deciles of predicted risk (prevalence round)

Prediction Model	Observed and predicted values	Deciles of predicted nodule malignancy (prevalence round) - continued									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
UKLS	Decile cut-offs %	(0-0.06)	(0.06-0.1)	(0.1-0.15)	(0.15-0.22)	(0.22-0.36)	(0.36-0.61)	(0.61-1.14)	(1.14-1.94)	(1.94-4.36)	(4.36-100)
	Nodule count	281	280	281	280	281	280	280	281	280	281
	Malignant nodule count	2	2	4	5	0	1	4	4	3	5
	Observed malignancy %	0.71	0.71	1.42	1.79	0.00	0.36	1.43	1.42	1.07	1.78
	Predicted malignancy %	0.04	0.08	0.13	0.18	0.28	0.46	0.81	1.52	2.94	12.67
Mayo	Decile cut-offs %	(2.5-3.3)	(3.3-3.72)	(3.72-4.18)	(4.18-4.8)	(4.8-5.7)	(5.7-6.66)	(6.66-7.7)	(7.7-9.08)	(9.08-11.05)	(11.05-100)
	Nodule count	290	287	288	288	289	288	288	288	288	289
	Malignant nodule count	0	1	0	1	0	1	1	2	0	26
	Observed malignancy %	0.00	0.35	0.00	0.35	0.00	0.35	0.35	0.69	0.00	9.00
	Predicted malignancy %	3.02	3.52	3.96	4.49	5.20	6.19	7.17	8.36	9.90	20.22
PKUPH	Decile cut-offs %	(10.55-12.91)	(12.91-14.26)	(14.26-16.27)	(16.27-18.42)	(18.42-20.83)	(20.83-24.18)	(24.18-28.35)	(28.35-37.28)	(37.28-52.47)	(52.47-99.92)
	Nodule count	289	288	288	288	290	287	288	288	288	289
	Malignant nodule count	0	0	1	1	1	0	3	3	3	20
	Observed malignancy %	0.00	0.00	0.35	0.35	0.34	0.00	1.04	1.04	1.04	6.92
	Predicted malignancy %	12.17	13.60	15.21	17.33	19.67	22.43	26.29	32.03	44.00	63.63
VA	Decile cut-offs %	(8.24-13.79)	(13.79-15.32)	(15.32-16.74)	(16.74-18.43)	(18.43-20.3)	(20.3-22.67)	(22.67-25.79)	(25.79-29.32)	(29.32-34.62)	(34.62-100)
	Nodule count	289	288	288	289	289	287	288	288	290	287
	Malignant nodule count	1	0	0	1	1	1	3	1	6	18
	Observed malignancy %	0.35	0.00	0.00	0.35	0.35	0.35	1.04	0.35	2.07	6.27
	Predicted malignancy %	12.17	14.60	16.06	17.58	19.34	21.44	24.11	27.50	31.91	42.96

Models were applied to the low-dose computed tomography image where nodules were first seen.
Abbreviations: MD: mean diameter; VOL: volume; UKLS: United Kingdom Lung Cancer Screening trial; PKUPH: Peking University People´s Hospital; VA: Veterans Affairs

As published in (González Maldonado et al. 2020a), reprinted with permission.

Supplementary Table 7. Observed versus predicted nodule malignancy rates by deciles of predicted risk (incidence rounds)

Prediction Model	Observed and predicted values	Deciles of predicted nodule malignancy (incidence rounds)									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
PanCan 1b	Decile cut-offs %	(0.01-0.16)	(0.16-0.37)	(0.37-0.58)	(0.58-0.82)	(0.82-1.16)	(1.16-1.74)	(1.74-2.92)	(2.92-6.4)	(6.4-14.86)	(14.86-82.52)
	Nodule count	102	112	93	101	104	100	102	102	103	101
	Malignant nodule count	0	0	0	2	2	0	4	5	5	13
	Observed malignancy %	0.00%	0.00%	0.00%	1.98%	1.92%	0.00%	3.92%	4.90%	4.85%	12.87%
	Predicted malignancy %	0.09%	0.27%	0.47%	0.70%	0.98%	1.43%	2.24%	4.47%	9.66%	33.09%
PanCan 2b	Decile cut-offs %	(0-0.13)	(0.13-0.32)	(0.32-0.55)	(0.55-0.79)	(0.79-1.16)	(1.16-1.63)	(1.63-2.62)	(2.62-5.26)	(5.26-12.74)	(12.74-81.97)
	Nodule count	102	102	102	102	102	102	102	102	102	102
	Malignant nodule count	0	0	0	1	1	0	3	6	10	10
	Observed malignancy %	0.00%	0.00%	0.00%	0.98%	0.98%	0.00%	2.94%	5.88%	9.80%	9.80%
	Predicted malignancy %	0.07%	0.22%	0.43%	0.66%	0.95%	1.39%	2.12%	3.76%	8.17%	28.63%
PanCan MD	Decile cut-offs %	(0-0)	(0-0.04)	(0.04-0.09)	(0.09-0.17)	(0.17-0.31)	(0.31-0.53)	(0.53-1.26)	(1.26-3.23)	(3.23-9.77)	(9.77-78.07)
	Nodule count	102	102	102	102	102	102	102	102	102	102
	Malignant nodule count	1	0	0	0	0	2	1	8	9	10
	Observed malignancy %	0.98%	0.00%	0.00%	0.00%	0.00%	1.96%	0.98%	7.84%	8.82%	9.80%
	Predicted malignancy %	0.00%	0.02%	0.06%	0.13%	0.24%	0.41%	0.84%	2.04%	5.39%	25.89%
PanCan VOL	Decile cut-offs %	(0-0.05)	(0.05-0.14)	(0.14-0.26)	(0.26-0.42)	(0.42-0.61)	(0.61-0.94)	(0.94-1.71)	(1.71-3.44)	(3.44-9.76)	(9.76-71.91)
	Nodule count	102	102	102	102	102	102	102	102	102	102
	Malignant nodule count	0	0	1	0	1	3	2	6	6	12
	Observed malignancy %	0.00%	0.00%	0.98%	0.00%	0.98%	2.94%	1.96%	5.88%	5.88%	11.76%
	Predicted malignancy %	0.03%	0.09%	0.20%	0.34%	0.51%	0.75%	1.25%	2.45%	5.50%	25.09%

As published in (González Maldonado et al. 2020a), reprinted with permission.

Supplementary Table 7 (continued). Observed versus predicted nodule malignancy rates by deciles of predicted risk (incidence rounds)

Prediction Model	Observed and predicted values	Deciles of predicted nodule malignancy (incidence rounds) - continued									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
UKLS	Decile cut-offs %	(0.01-0.07)	(0.07-0.13)	(0.13-0.18)	(0.18-0.3)	(0.3-0.48)	(0.48-0.78)	(0.78-1.4)	(1.4-2.55)	(2.55-6.51)	(6.51-100)
	Nodule count	100	99	99	99	100	99	99	99	99	100
	Malignant nodule count	0	2	4	5	2	3	1	4	5	5
	Observed malignancy %	0.00%	2.02%	4.04%	5.05%	2.00%	3.03%	1.01%	4.04%	5.05%	5.00%
	Predicted malignancy %	0.04%	0.10%	0.15%	0.23%	0.37%	0.64%	1.03%	1.86%	3.89%	25.29%
Mayo	Decile cut-offs %	(2.71-4.02)	(4.02-4.8)	(4.8-5.77)	(5.77-6.95)	(6.95-8.28)	(8.28-9.79)	(9.79-12.4)	(12.4-16.72)	(16.72-29.15)	(29.15-99.87)
	Nodule count	102	102	101	102	101	102	101	102	101	102
	Malignant nodule count	0	1	0	2	0	2	2	4	6	14
	Observed malignancy %	0.00%	0.98%	0.00%	1.96%	0.00%	1.96%	1.98%	3.92%	5.94%	13.73%
	Predicted malignancy %	3.55%	4.37%	5.29%	6.32%	7.64%	9.04%	10.91%	14.16%	21.70%	54.07%
PKUPH	Decile cut-offs %	(11.54-15.81)	(15.81-17.93)	(17.93-21.67)	(21.67-25.69)	(25.69-32.11)	(32.11-40.21)	(40.21-55.03)	(55.03-66.37)	(66.37-77.76)	(77.76-99.46)
	Nodule count	102	102	102	102	102	102	102	102	102	102
	Malignant nodule count	0	0	1	1	1	4	7	1	7	9
	Observed malignancy %	0.00%	0.00%	0.98%	0.98%	0.98%	3.92%	6.86%	0.98%	6.86%	8.82%
	Predicted malignancy %	14.18%	16.73%	19.88%	23.47%	28.55%	36.30%	47.59%	60.99%	71.38%	86.04%
VA	Decile cut-offs %	(9.07-16.71)	(16.71-19.07)	(19.07-21.24)	(21.24-24.82)	(24.82-28.01)	(28.01-32.06)	(32.06-36.49)	(36.49-43.3)	(43.3-54.58)	(54.58-99.77)
	Nodule count	102	102	101	102	101	102	101	102	101	102
	Malignant nodule count	0	0	0	1	3	1	2	4	6	14
	Observed malignancy %	0.00%	0.00%	0.00%	0.98%	2.97%	0.98%	1.98%	3.92%	5.94%	13.73%
	Predicted malignancy %	14.22%	18.00%	20.06%	22.85%	26.30%	30.03%	34.23%	39.74%	48.35%	72.26%

Models were applied to the low-dose computed tomography image where nodules were first seen.
Abbreviations: MD: mean diameter; VOL: volume; UKLS: United Kingdom Lung Cancer Screening trial; PKUPH: Peking University People´s Hospital; VA: Veterans Affairs

As published in (González Maldonado et al. 2020a), reprinted with permission.

Supplementary Table 8. Evaluation of absolute risk calibration of the selected models when applied on nodules first observed in the prevalence or incidence rounds

	Estimate, Test Statistic or p-value	Risk Prediction Model							
		PanCan 1b	PanCan 2b	PanCan MD	PanCan VOL	UKLS	Mayo	PKUPH	VA
Prevalence round	**HL Stat**	7.71	7.23	30.53	10.89	158.99	162.32	1119.9	774.38
	HL Test p	0.56	0.61	<0.001	0.28	<0.001	<0.001	<0.001	<0.001
	BS all	0.009	0.009	0.009	0.009	0.012	0.014	0.094	0.063
	Sp. Z-Stat	-1.081	0.436	3.888	1.978	-1.076	-12.63	-19.35	-25.24
	Sp. Test p	0.28	0.67	<0.001	0.05	0.28	<0.001	<0.001	<0.001
Incidence rounds	**HL Stat**	20.67	12.72	71.08	17.62	283.03	114.56	992.27	478.43
	HL Test p	<0.001	0.01	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	BS all	0.034	0.034	0.033	0.032	0.041	0.053	0.214	0.134
	Sp. Z-Stat	-1.114	-0.277	2.123	1.273	7.09	-6.052	4.925	-9.303
	Sp. Test p	0.27	0.78	0.03	0.20	<0.001	<0.001	<0.001	<0.001

Models were applied to the low-dose computed tomography image where nodules were first seen.
Abbreviations: HL: Hosmer-Lemeshow; BS: Brier Score; Sp. Z-Stat: Spiegelhalter Z-Statistic; Sp.Test p: Spiegelhalter Z-Test p-value; MD: mean diameter; VOL: volume; UKLS: United Kingdom Lung Cancer Screening trial; PKUPH: Peking University People´s Hospital; VA: Veterans Affairs

As published in (González Maldonado et al. 2020a), reprinted with permission.

Supplementary Table 9. Coefficients of multivariable logistic regression models fitted on data from the LUSI trial

Covariate	All factors included			After variable selection		
	β-coef	95% CI	p-value	β-coef	95% CI	p-value
Intercept	-8.27	(-12.16, -4.38)	<0.001	-9.88	(-13.03, -6.73)	<0.001
Age	0.07	(0.003, 0.14)	0.05	0.06	(0.01, 0.11)	0.02
Sex	-0.23	(-1.07, 0.60)	0.59			
History of cancer (excl. thorax)	0.04	(-0.89, 0.97)	0.94			
Years since quit smoking	0.74	(-0.19, 1.68)	0.13	0.83	(-0.15, 1.82)	0.10
Smoking duration (years)	-0.03	(-0.11, 0.04)	0.39			
Emphysema	0.23	(-0.41, 0.88)	0.49			
Bronchitis	-1.22	(-2.34, -0.11)	0.04	-1.23	(-2.29, -0.17)	0.03
FVC	-0.18	(-0.60, 0.24)	0.41			
Nodule size (MD, mm)	0.14	(0.09, 0.19)	<0.001	0.14	(0.09, 0.19)	<0.001
Nodule type	-0.58	(-1.67, 0.50)	0.30			
Nodule location (upper vs middle-or-lower)	1.23	(0.39, 2.07)	0.004	1.23	(0.35, 2.11)	0.01
Nodule count per scan	-0.06	(-0.18, 0.06)	0.34			
Nodule spiculation	1.64	(0.93, 2.35)	<0.001	1.72	(1.02, 2.42)	<0.001

Abbreviation: MD = mean diameter = (largest diameter + perpendicular diameter)/2

As published in (González Maldonado et al. 2020a), reprinted with permission.

Supplementary Table 10. Basic characteristics of LUSI trial participants at randomization and at the end of follow-up

Characteristics (at time of randomization or recruitment)		Intervention arm	Control arm	p-value[a]	Total
		n (% by arm)	n (% by arm)		n (% overall)
Participants		2,029 (50.1)	2,023 (49.9)		4,052 (100)
Sex (male)		1,315 (50.1)	1,307 (49.9)	0.92	2,622 (64.7)
Age at first screening (years, median [range])		56.9 [50.3, 71.9]	56.9 [50.3, 71.8]		56.9 [50.3, 71.9]
Age at 1st screening (years, age group)	[50 – 55)	788 (39.0)	772 (38.2)	0.69	1560 (38.5)
	[55 – 60)	557 (27.5)	570 (28.2)		1127 (27.8)
	[60 – 65)	380 (18.8)	398 (19.7)		778 (19.2)
	>65	304 (15.0)	283 (14.0)		587 (13.6)
Current smokers		1,259 (50.2)	1,248 (49.8)	0.84	2,506 (61.9)
Former smokers		770 (49.8)	775 (50.2)		1,546 (38.1)
Observation time since randomization[b] (Median [IQR])	Incidence	9.76 [8.8 – 10.4]	9.77 [8.8 – 10.4]	0.52	9.77 [8.8 - 10.4]
	Mortality	9.96 [9.0 – 10.6]	9.96 [9.0 – 10.6]	0.91	9.96 [9.0 – 10.6]
Observation time after last screen[b] (Median [IQR])	Incidence	5.73 [4.8 – 6.3]	-		-
	Mortality	5.94 [5.0 – 6.6]	-		-
Participants lost to follow-up		6 (0.3)	9 (0.4)	0.60	15
Total number of lung cancers		90 (4.4)	74 (3.7)	0.24	164
Number of deaths[c]		183 (9.0)	174 (8.6)	0.69	357
Number of deaths from lung cancer		33 (1.6)	44 (2.2)	0.24	77
Self- or clinician-initiated X-ray or LDCT diagnostics for screening purposes	a. During the active screening period	12	98	< 0.001	110
	b. After the active screening period / after 5th questionnaire	132	166	0.05	298

[a] p-Value from a chi-squared test for categorical variables or from a Mann–Whitney U test for the difference in continuous variables between the two study arms.
[b] Until April 30th 2019 for lung cancer as endpoint (incidence) and until July 2nd 2019 for mortality.
[c] Until July 2nd 2019 (last vital status update from registry).

As published in (González Maldonado et al. 2020b), reprinted with permission.

	(0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,6]	(6,7]	(7,8]	(8,9]	(9,10]	(10,11]
LDCT - LC cases	19	15	13	9	7	5	6	6	2	5	3
LDCT - n censored	5	9	9	13	9	11	10	26	474	561	627
LDCT - PY at risk	2015.1	1990.1	1972.7	1946.9	1932.6	1915.9	1896.3	1875.3	1654.2	1155.9	408.7
Control - LC cases	4	7	8	13	4	6	12	8	6	6	0
Control - n censored	6	8	13	13	12	13	7	16	488	560	614
Control - PY at risk	2018.6	2007.7	1990.3	1966.1	1943.4	1927.1	1909.9	1887.1	1657.9	1161.6	418.4
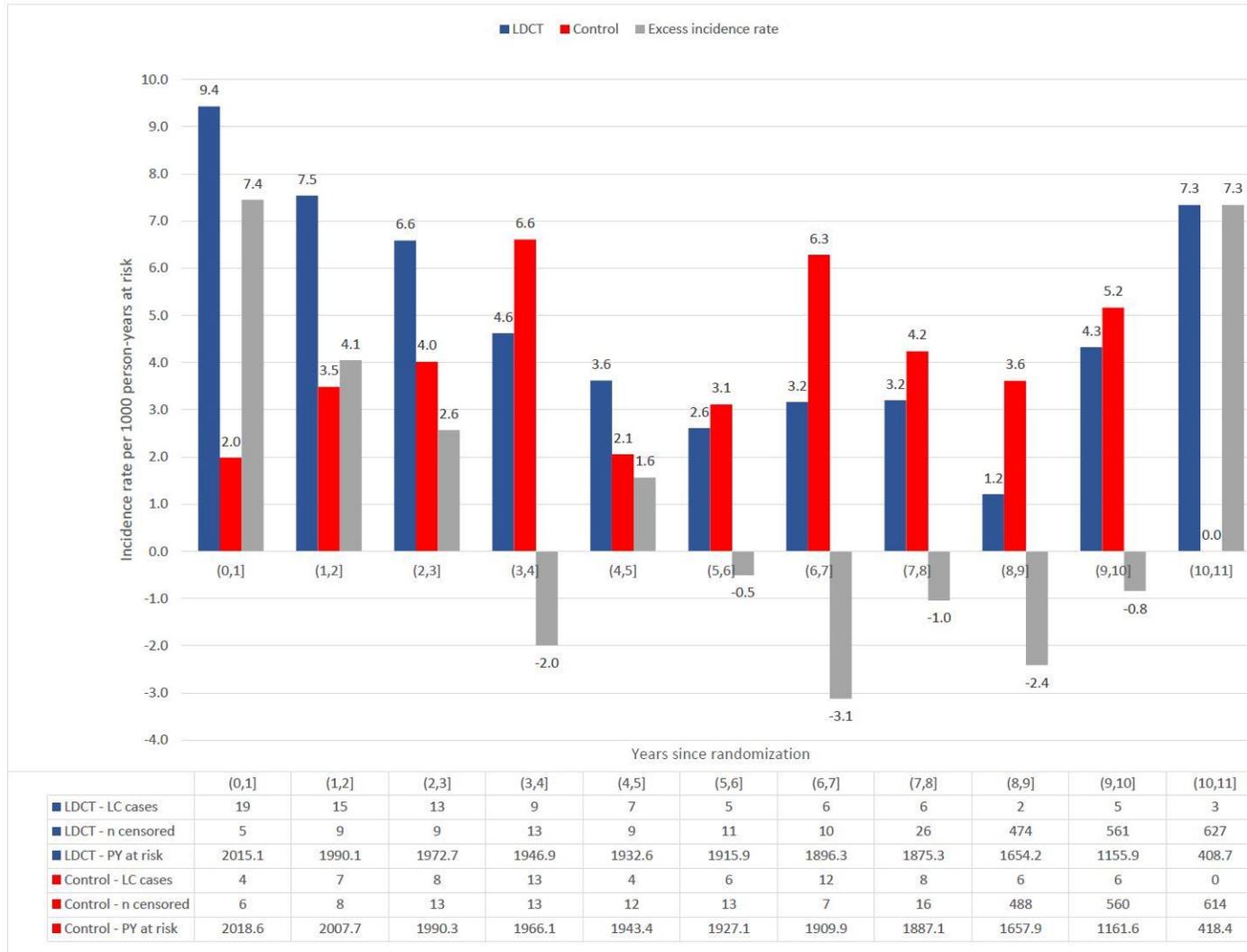
Supplementary Figure 6. Annual incidence rates by study arm and excess incidence rates between study arms by years since randomization

As published in (González Maldonado et al. 2020b), reprinted with permission.

142

Supplementary Table 11. Lung cancer cases by tumor histology, study arm and sex

Tumor histology	Control			LDCT			Total
	Male	Female	Subtotal	Male	Female	Subtotal	
nonBAC Adenocarcinoma	24	12	36	31	18	49	85
BAC Adenocarcinoma	0	1	1	4	6	10	11
BAC (8250/3)	0	1	1	3	6	9	10
Mucinous BAC (8253/3)	0	0	0	1	0	1	1
cTX cN0 cM1a	1	0	1	1	0	1	2
pT1 pN0 cM0	1	1	2	0	0	0	2
pT1a pN0 cM0	0	2	2	0	0	0	2
pT1b pN1 pM1b	0	1	1	0	0	0	1
pT2 pN0 cM0	0	1	1	0	0	0	1
pT2a pN0 cM0	2	0	2	0	0	0	2
pT3 pN0 cM0	0	1	1	0	0	0	1
Squamous cell	11	4	15	13	2	15	30
Small cell	11	6	17	9	1	10	27
Large cell	1	0	1	1	0	1	2
Carcinoid	1	0	1	2	0	2	3
Malignant neoplasm	0	1	1	1	0	1	2
Unspecified	2	0	2	2	0	2	4
Total	50	24	74	63	27	90	164

Tumor histology	Male			Female			Sum
	Control	LDCT	Subtotal	Control	LDCT	Subtotal	
nonBAC Adenocarcinoma	24	31	55	12	18	30	85
BAC Adenocarcinoma	0	4	4	1	6	7	11
BAC (8250/3)	0	3	3	1	6	7	10
Mucinous BAC (8253/3)	0	1	1	0	0	0	1
cTX cN0 cM1a	0	1	1	1	0	1	2
pT1 pN0 cM0	0	1	1	0	1	1	2
pT1a pN0 cM0	0	0	0	0	2	2	2
pT1b pN1 pM1b	0	0	0	0	1	1	1
pT2 pN0 cM0	0	0	0	0	1	1	1
pT2a pN0 cM0	0	2	2	0	0	0	2
pT3 pN0 cM0	0	0	0	0	1	1	1
Squamous cell	11	13	24	4	2	6	30
Small cell	11	9	20	6	1	7	27
Large cell	1	1	2	0	0	0	2
Carcinoid	1	2	3	0	0	0	3
Malignant neoplasm	0	1	1	1	0	1	2
Unspecified	2	2	4	0	0	0	4
Total	50	63	113	24	27	51	164

As published in (González Maldonado et al. 2020b), reprinted with permission.

Supplementary Table 12. Incidence by study arm and excess incidence by years since randomization

	Years since randomization	(0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,6]	(6,7]	(7,8]	(8,9]	(9,10]	(10,11]
LDCT	**Cases**	19	15	13	9	7	5	6	6	2	5	3
	nonBAC Adenocarcinoma (%)	10 (52.6)	10 (66.7)	7 (53.8)	6 (66.7)	3 (42.9)	2 (40.0)	1 (16.7)	3 (50.0)	1 (50.0)	4 (80.0)	2 (66.7)
	BAC (%)	2 (10.5)	1(6.7)	1 (7.7)	2 (22.2)	1 (14.3)	1 (20.0)	1 (16.7)	0	0	0	1 (33.3)
	Other (%)	7 (36.8)	4 (26.7)	5 (38.5)	1 (11.1)	3 (42.9)	2 (40.0)	4 (66.7)	3 (50.0)	1 (20.0)	1 (20.0)	0
	Screen-detected	18	15	11	8	5	4	2	0	0	0	0
	nonBAC Adenocarcinoma (%) [a]	9 (90)	10 (100)	7 (100)	5 (83.3)	3 (100)	2 (100)	0	0	0	0	0
	BAC (%) [a]	2 (100)	1 (100)	1 (100)	2 (100)	1 (100)	0	1 (100)	0	0	0	0
	Other (%) [a]	7 (100)	4 (100)	3 (60)	1 (100)	1 (33.3)	2 (100)	1 (25)	0	0	0	0
Control	**Cases**	4	7	8	13	4	6	12	8	6	6	0
	nonBAC Adenocarcinoma (%)	2 (50.0)	4 (57.1)	3 (37.5)	7 (53.8)	2 (50.0)	0	8 (66.7)	5 (62.5)	3 (50.0)	2 (33.3)	0
	BAC (%)	0	0	0	0	1 (25.0)	0	0	0	0	0	0
	Other (%)	2 (50.0)	3 (42.9)	5 (62.5)	6 (46.2)	1 (25.0)	6 (100)	4 (33.3)	3 (37.5)	3 (50.0)	4 (66.7)	0
Excess incidence	All subtypes	15	8	5	-4	3	-1	-6	-2	-4	-1	3
	nonBAC Adenocarcinoma	8	6	4	-1	1	2	-7	-2	-2	2	2
	BAC	2	1	1	2	0	1	1	0	0	0	1
	Other	5	1	0	-5	2	-4	0	0	-2	-3	0

[a] Percentage of screen-detected tumors relative to all tumors of the same histologic subtype in the LDCT arm.

As published in (González Maldonado et al. 2020b), reprinted with permission.

Supplementary Table 13. Model fit based on observed and expected screen-detected and non-screen detected lung cancer cases

Screening round or post-screening years	Screened	Screen-detected cases (SDC)	Non-screen detected[a] cases (IC)	Expected SDC	Expected IC	χ^2 statistic, p-value
R1	2028	24	1	23.55	1.45	
R2	1892	11	0	9.99	1.01	
R3	1849	10	2	10.48	1.52	1.95, p= 0.58
R4	1826	9	1	8.73	1.27	
0-1 y after R5	1810	9	2	9.85	1.15	
0-2 y after R5[b]		9	3	8.73	3.27	1.28, p= 0.73
0-3 y after R5[b]		9	7	9.12	6.88	1.25, p= 0.74
0-4 y after R5[b]		9	13	9.87	12.13	1.39, p= 0.71
0-5 y after R5[b]		9	15	9.32	14.68	1.27, p= 0.74
0-6 y after R5[b]		9	20	10.43	18.57	1.56, p= 0.67
0-7 y after R5[b]		9	23	11.51	20.49	2.11, p= 0.55

[a] Non-screen detected cases are interval cancer cases (diagnosed in between screening appointments or within one year of the last screening appointment, following a negative screening appointment), as well as lung cancer cases diagnosed later than one year after last screening.
[b] These numbers represent cumulative lung cancer incidence in the years following the last screening appointment (screen-detected cases after R5 remain constant in the absence of screening and non-screen detected cases continue to be diagnosed in the subsequent years). For example, in the period 0-4 years after R5, there were 2 interval cancers in the first year after screening and 11 diagnoses in the years 2 to 4 after R5 for a total of 13 non-screen detected cases.

As published in (González Maldonado et al. 2020b), reprinted with permission.

Supplementary Table 14. Estimated proportions of screen-detected tumors by lead time ((1 – Pclin(t))

	$1 - P_{clin}$ (95% CI)				
Histologic subtype	4y	6y	8y	10y	12y
Non-small cell lung cancers non- BAC	32.9% (26.4%, 39.4%)	18.9% (13.5%, 24.8%)	10.8% (6.9%, 15.6%)	6.2% (3.6%, 9.8%)	3.6% (1.8%, 6.1%)
BAC	88.3% (79.4%, 98.5%)	83.0% (70.7%, 97.8%)	77.9% (63.0%, 97.1%)	73.2% (56.1%, 96.4%)	68.8% (50.0%, 95.7%)

[a] Probabilities calculated based on the estimates of mean pre-clinical sojourn time and sensitivity published by Patz et al. 2018[8].

As published in (González Maldonado et al. 2020b), reprinted with permission.

Supplementary Table 15. Performance of EarlyCDT®-Lung amongst subjects with suspicious nodules by nodule size.

	EarlyCDT®-Lung result	HIGH		MOD		NS		OR [95% CI]		LR+	
	Lung cancer status	LC	No LC	LC	No LC	LC	No LC	Positive = High	Positive = High or Moderate	Positive = High	Positive = High or Moderate
Largest diameter	**< 10**	1	4	0	3	10	78	2.03 [0.20, 19.95]	1.11 [0.12, 10.02]	1.93 [0.24, 15.77]	1.10 [0.16, 7.53]
	≥ 10	5	0	0	1	29	4	---	0.69 [0.06, 7.51]	1.17 [1.02, 1.35)	0.95 [0.65, 1.39]
	Overall†	6	4	0	4	39	82	3.31 [0.88, 12.39]	1.58 [0.51, 4.86]	1.92 [1.09, 3.40]	1.50 [0.55, 4.06]

† For one subject, the CT scan evaluation at round 2 was deemed suspicious (with immediate recall) even in the absence of pulmonary nodules, due to the identification of atelectasis (collapsed lung) in the scan images. That subject was excluded for these analyses.

As published in (González Maldonado et al. 2020b), reprinted with permission.

Supplementary Table 16. Trial characteristics and corresponding overdiagnosis estimates defined as the fraction of screen-detected lung cancers

Trial name[a]	Age limits for eligibility (years)	FU after last screen (years)[b]	Participation rate	Contamination rate	Estimated Overdiagnosis, PS, (95% CI)
DLCST (Heleno 2018)	50–70	5	95.5%	20.3%	0.67 (0.37, 0.96)
ITALUNG (Paci 2017, Paci 2020)	55–69	5	81%	Not reported (NR)	−0.11 (NR)
		8.3[c]		NR	-0.24 (NR)
LUSI (this study)	50–69	5.71[d]	Lowest at R4 with 93.4% and highest at R1 with 99.9%	13.0%[e] (264/2023)	0.25 (-0.11, 0.64)
NELSON (deKoning 2020)	50–74	4.5	85.8% in total (lowest at R4 with 67.4%, highest at R1 with 95.8%)	NR	0.20 (–0.05, 0.42)
		5.5			0.09 (-0.18, 0.32)
NLST (Patz 2014; NLST Team 2019)	55-74	4.5[f]	95%	NR	0.19 (0.05, 0.31)
		9.3[f]		NR	0.03 (NR)

[a] Results shown in this table come from studies receiving a quality rating of 1: properly powered and conducted randomized clinical trial.
[b] Reported approximate post-screening follow up in the original publications.
[c] Approximate post-screening follow-up calculated as originally reported median post-randomization follow-up = 11.3 - 3 years corresponding to 4 screening rounds.
[d] Median follow-up time after an individual's last screening appointment and until 30.04.2019.
[e] Contamination rate: percentage of participants in the control group with at least one test (LDCT or X-rays) during the active phase or during follow-up for lung cancer screening purposes outside the LUSI trial.
[f] Median follow-up time after an individual's last screening appointment as reported by the NLST team.

As published in (González Maldonado et al. 2020b), reprinted with permission.

10 APPENDIX

10.1 R code (convolution model)

```
###############################################################################
# Author: Sandra Gonzalez Maldonado
# Goal: Estimate tumor mean pre-clinical sojourn time,
#       LDCT detection sensitivity, lead time and
#       the proportion of tumors with lead times below a certain value,
#       using a convolution model
# Method implemented using a multinomial distribution for the vector of
# screen-detected and interval cancers
# Literature: Straatman H et al 1997.
# Estimating Lead Time and Sensitivity in a Screening Program without
# Estimating the Incidence in the Screened Group. Biometrics
# Modification: P[I0] = 0 <- null incidence between recruitment and first screen


###############################################################################
#------        Model-based estimation : Convolution model    ----------------#
# --      assuming an exponential distribution of sojourn times   -----------#
###############################################################################
#---------                 ALL HISTOLOGIES, ALL AGES        ---------------#
   # Vector with the screening rounds
      screens
   # Age at entry
      age_entry
   # Years of follow-up since randomization
      yearsofFU
   # Observed prevalent cases at screening times
      S_vect
   # Observed interval cases
      I_vect_screenperiod
   # Incident cases post-screening
      (I_vect <- I_vect_screenperiod+ c(0,0,0,0,post_sc_cases))

estimate_results <- list()

# Consider different ages at which tumor genesis starts

  for (a in c(1:6))
  {
   age_tumor_genesis <- c(0, 10, 20, 30, 40, 45)

  print( t_vect <- age_entry - age_tumor_genesis[a] + c(0,1,2,3,4,yearsofFU))

###############################################################
# Estimate lambda (1/MST) and theta = sensitivity
# By maximizing the PARTIAL LOG-LIKELIHOOD of a multinomial distribution

partial_loglik_f <- function(x){

        l = x[1]
        theta = x[2]
        r <- 0.001 # This parameter disappears in the partial log likelihood;
                   # it can be set to any arbitrary value

  # Calculate probabilities of screen-detected and interval cases
  # separately as in Straatman 1997 (Biometrics)
```

148

```
# Ignore the pre-screening period (0 to t1), as in method 1 in Straatman 1997

 P_S <- c(rep(0,length(screens))) ; P_I <- c(rep(0,length(screens)))

 P_S[1] <- (r/l)*(1-(1/(l*t_vect[1]))+
                      exp(-l*t_vect[1])/(l*t_vect[1]))*(1-exp(-l*t_vect[1]))

 P_I[1] <- r*(t_vect[2]-t_vect[1]) - (r/l)*(1-exp(-l*(t_vect[2]-t_vect[1])))

 if (length(screens)>1){
  for (j in c(2:length(screens))){
    P_S[j] = (r/l)*(1-exp(-l*(t_vect[j]-t_vect[j-1])))
    P_I[j] = r*(t_vect[j+1]-t_vect[j]) -
        (r/l)*(1-exp(-l*(t_vect[j+1]-t_vect[j])))
       }}

# Considering sensitivity = theta in (0,1]

    P_S_theta <- c(rep(0,length(screens)))
    P_I_theta <- c(rep(0,length(screens)))

    P_S_theta[1] = P_S[1]*theta
    P_I_theta[1] = P_S[1]*(1-theta)*(1-exp(-l*(t_vect[2]-t_vect[1])))+P_I[1]

 if (length(screens)>1){
   for (j in c(2:length(screens))){
     P_S_theta[j] = P_S[j]*theta +
        P_S[j-1]*(1-theta)*(exp(-l*(t_vect[j]-t_vect[j-1])))

     P_I_theta[j] = P_I[j]+(1-theta)*P_S[j]*(1-exp(-l*(t_vect[j+1]-t_vect[j])))
       }}

 s_logps <- sum(S_vect*log(P_S_theta/(P_S_theta + (1-p_l/2)*P_I_theta)))
 i_logpi <- sum(I_vect*log(1-(P_S_theta/(P_S_theta + (1-p_l/2)*P_I_theta))))

 LL <-  s_logps + i_logpi

     return(LL)
}

 # Matrix defining the constrains on lambda >=0 and 0<=theta<=1

  ui <- matrix(c(1, 0,  0, 1,  0, -1), ncol = 2, byrow= T)
  ci <- c(0,0,-1)

 # Maximize the partial log-likelihood by applying constrained maximization

   maxima <- constrOptim(theta = c(0.3, 0.8), # initialize parameters
                      f= partial_loglik_f, grad = NULL,
                      ui = ui, ci = ci,
                      control = list(fnscale = -1))$par

 # Calculate the function on a grid, in order to find the confidence regions

   MST_SENS = expand.grid(x=seq(0.01, 1, 0.008),
                        y=seq(0.2,1, 0.008))

   results <- apply(as.matrix(MST_SENS), 1, partial_loglik_f)
 # get arg max of the likelihood function
```

149

```
  MST_SENS[which(results == max(results, na.rm = T)),]

  (l_hat <- MST_SENS[which(results == max(results, na.rm = T)),1])
  (theta_hat <- MST_SENS[which(results == max(results, na.rm = T)),2])

# Estimated mean sojourn time (1/lambda_hat), and sensitivity

  (mean_sojourn_time <- 1/l_hat)
  (sens_hat <- theta_hat)

# Calculate confidence regions

 # Get the quantiles of a chi-squared distribution with
 # two degrees of freedom as suggested in Walter and Day 1984

 chi2_90 <- qchisq(0.9, df = 2, ncp = 0, lower.tail = FALSE, log.p = FALSE)
 chi2_95 <- qchisq(0.95, df = 2, ncp = 0, lower.tail = FALSE, log.p = FALSE)

 # calculate the regions based on whether twice the distance to the maximum
 # is less than the chisquare quantile

 conf_region_90 <- MST_SENS[which((partial_loglik_f(maxima)  - results )
                                 < chi2_90/2),]
 (conf_MST_90  <- c(1/max(conf_region_90[,1], na.rm = T),
                     1/min(conf_region_90[,1], na.rm = T)))

 (conf_sens_90  <- c(min(conf_region_90[,2], na.rm = T),
                      max(conf_region_90[,2], na.rm = T)))

 conf_region_95 <- MST_SENS[which((partial_loglik_f(maxima)   - results )
                                 < chi2_95/2),]
 (conf_MST_95  <- c(1/max(conf_region_95[,1], na.rm = T),
                     1/min(conf_region_95[,1], na.rm = T)))

 (conf_sens_95  <- c(min(conf_region_95[,2], na.rm = T),
                      max(conf_region_95[,2], na.rm = T)))

    estimate_results[[a]] <- c(age_tumor_genesis[a], mean_sojourn_time,
                               conf_MST_90, conf_MST_95,
                               sens_hat, conf_sens_90, conf_sens_95)
 }

  estimate_results_df <- data.frame(do.call(rbind, estimate_results))

  names(estimate_results_df) <- c("age_tumorgenesis_start", "MST",
                                 "MST_90%CI_L", "MST_90%CI_U",
                                 "MST_95%CI_L", "MST_95%CI_U",
                                 "SENS", "SENS_90%CI_L", "SENS_90%CI_U",
                                 "SENS_95%CI_L", "SENS_95%CI_U")
  (estimate_results_df )

#############################################################################
#---------     CALCULATE PROPORTION OF TUMORS BY LEAD TIMES    ------------#
#############################################################################

  print("Proportion becoming clinical: 1- exp(-lambda*t) all")
     for (t in c(4, 6, 8, 10, 12))
     {print(exp(-t*(1/estimate_results[[1]][2])))
      print("CIs low")
      print( exp(-t*(1/estimate_results[[1]][6])))
      print("CIs upper")
```

```
      print( exp(-t*(1/estimate_results[[1]][5])))
    }

###########################################################################
#----- EXPECTED SCREEN DETECTED CASES (SDC) AND INTERVAL CASES (IC) -----#
###########################################################################

  t_vect <- age_entry  + c(0,1,2,3,4,yearsofFU)

  expected_SD_INT <- function(x){

    l = x[1]
    theta = x[2]

    P_S <- c(rep(0,length(screens))) ; P_I <- c(rep(0,length(screens)))

    P_S[1] <- (r/l)*(1-(1/(l*t_vect[1]))+
                        exp(-l*t_vect[1])/(l*t_vect[1]))*(1-exp(-l*t_vect[1]))
    P_I[1] <- r*(t_vect[2]-t_vect[1]) - (r/l)*(1-exp(-l*(t_vect[2]-t_vect[1])))

    if (length(screens)>1){
      for (j in c(2:length(screens))){
        P_S[j] = (r/l)*(1-exp(-l*(t_vect[j]-t_vect[j-1])))
        P_I[j] = r*(t_vect[j+1]-t_vect[j]) -
            (r/l)*(1-exp(-l*(t_vect[j+1]-t_vect[j])))
      }}

    #  Considering sensitivity = theta in (0,1]

      P_S_theta <- c(rep(0,length(screens)))
      P_I_theta <- c(rep(0,length(screens)))

      P_S_theta[1] = P_S[1]*theta
      P_I_theta[1] = P_S[1]*(1-theta)*(1-exp(-l*(t_vect[2]-t_vect[1])))+P_I[1]

     if (length(screens)>1){
      for (j in c(2:length(screens))){
      P_S_theta[j] = P_S[j]*theta +
            P_S[j-1]*(1-theta)*(exp(-l*(t_vect[j]-t_vect[j-1])))

      P_I_theta[j] = P_I[j]+
          (1-theta)*P_S[j]*(1-exp(-l*(t_vect[j+1]-t_vect[j])))
        }}

    return(c(P_S_theta, P_I_theta))

  }

  ##------  Calculate expected proportions of SDCs and ICs  -------------#

  #    All
    r<- 0.08 # This parameter disappears in the partial log likelihood
    probs <- expected_SD_INT(c(1/estimate_results_df$MST[1],
                            estimate_results_df$SENS[1]))
    probs_sd <- probs[1:5]
    probs_int  <- probs[6:10]
    exp_proportion_sd <- probs_sd/(probs_sd + (1-p_l/2)*probs_int)

###########################################################################
```

ACKNOWLEDGMENTS

Special thanks and recognition go to my supervisor Prof. Dr. Rudolf Kaaks for trusting me with this and other research endeavors. Thank you for your commitment and for sharing your knowledge and passion for science with me. I enjoyed our technical and personal discussions during which we sometimes forgot about the passing of time. I learnt a lot about Epidemiology and about myself thanks to the interaction with you. This helped me become a stronger researcher and more resilient person.

To my officemate at DKFZ and dear friend Dr. Anika Hüsing: no words can express my gratitude. You are a beautiful mix of knowledge and kindness and I am grateful to have worked and interacted at a personal level with you. Thank you for your advice and support in all statistics and life-related topics. Thanks for the occasional piece of chocolate as well.

Many thanks to all colleagues involved in the LUSI trial, in particular to Dr. Erna Motsch for answering all my medicine-related questions and to Anke Trotter for all her support with complicated database tasks. I also thank Dr. Theron Johnson and the laboratory staff for having performed the measurements of the EarlyCDT®-Lung test as described in this thesis. To Lucas Hynes and all other co-authors of my publications: thank you for your support and feedback.

Without all of you, this thesis would not have been possible.

I thank all colleagues from the C020 division at DKFZ for being by my side all throughout this interesting journey. Petra and Heike: you are our guardian angels! Without you, I wouldn't have enjoyed my time at DKFZ as much as I did!

All my gratitude goes to my grandparents for they raised me with love and forged my character. To my parents: thank you for all the freedom that allowed me to follow my own path in life. To my sister for teaching me how to read! To my brother and extended family for always cheering me on.

To my friends in Mexico, Germany and many other corners of the world: thank you for being part of my life and for making it so colorful! In particular to the Berlin-Crew: Mats, Mark and Marco who were very close to me in the last years despite the geographical distance and who are always in my mind and heart.

And last but definitely not least, to my husband Lukas: thank you for your love, patience, support and motivation, in particular through rough patches. You are my strength and the best reason to smile everyday. I thank God for crossing our paths.

EIDESSTATTLICHE VERSICHERUNG

1. Bei der eingereichten Dissertation zu dem Thema "**Evaluation of Risk Models and Biomarkers for the Optimization of Lung Cancer Screening"** handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Ort und Datum   Unterschrift

Heidelberg, 10.11.2021