5-7-2016

A Retrospective Study Assessing the Predictive Performance of a Lung Cancer Screening Risk Prediction Model in a Clinical Lung Cancer Screening Program

Andrea K. Borondy Kitts
University of Connecticut, borondy@msn.com

A Retrospective Study Assessing the Predictive Performance of a Lung Cancer Screening Risk Prediction Model in a Clinical Lung Cancer Screening Program

Andrea K. Borondy Kitts

B.S. The University of Vermont, 1978

M.S. The Massachusetts Institute of Technology, 1996

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Public Health

At the

University of Connecticut

2016

APPROVAL PAGE

Master of Public Health Thesis

A Retrospective Study Assessing the Predictive Performance of a Lung Cancer Screening Risk Prediction Model in a Clinical Lung Cancer Screening Program

Presented by

Andrea K. Borondy Kitts, M.S

Major Advisor

_____

Julie Robison, Ph.D.

Associate Advisor

_____

Miranda Lynch, Ph.D.

Associate Advisor

_____

Brady McKee, M.D.

University of Connecticut

2016

ACKNOWLEDGEMENTS

Appreciation to my thesis advisors, Julie Robison, PhD (UConn Health) , Miranda Lynch, PhD (UConn Health) and Brady McKee, MD (Lahey Hospital & Medical Center) for their technical assistance, advice, comments on my thesis drafts and support for the study.

Thank you to Shawn Regis, PhD (Lahey Hospital & Medical Center) for his help with IRB study approval, data management and download and his continued encouragement.

Thank you to Martin Tammemagi, PhD, Brock University, for providing a reparameterized PLCOm2012 model for the study.

My thesis is dedicated to my late husband Dan, his late sister Bobbie Lynn and the millions of others who have been lost to lung cancer. My hope is that lung cancer screening will be quickly adopted giving people a chance to find their cancer early when a cure is more likely.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Background: United States Preventive Services Task Force (USPSTF) and Centers for Medicare & Medicaid Services (CMS) recommendations for annual screening for lung cancer with low dose CT (LDCT) scans rely on age and smoking history to identify those at high risk for lung cancer. The Tammemagi et al. six year lung cancer risk prediction model, PLCOm2012, developed and validated in large lung cancer screening clinical trials, demonstrated good predictive performance in screening selection. However, the model has not been validated in clinical practice. Validating the model in clinical practice would increase confidence in its ability to provide information for shared decision making discussions in the near term and would potentially allow for selection of other high risk groups, not currently recommended to be screened, in the future.

Methods: Retrospective evaluation of the predictive performance of the Tammemagi et al. six year lung cancer risk prediction model in the Lahey Hospital & Medical Center, Lahey physician referred patients enrolled in the lung cancer screening program between January 1, 2012 and November 30, 2015 (n=2302). Predictor variable data were gathered from the program clinical data base and program participant clinic medical records. All patients met the National Comprehensive Cancer Network (NCCN) Lung Cancer Screening Guidelines Group 1 or Group 2 high-risk criteria.

Results: The model six year mean risk for lung cancer was higher for participants with lung cancer, 4.56%, as compared to those without lung cancer, 3.55% (p=0.0265). Area under the curve (AUC) of the receiver operator characteristics (ROC) was 0.63 (95% CI 0.57 – 0.69). The mean absolute difference between observed and predicted risk was 0.013 or less for the first 9 deciles. At the 1.51% predicted risk recommended screening threshold; sensitivity = 85.7%, specificity = 29.7%, and PPV = 3.7%. In sub-group analysis, for NCCN Group 2 (younger,

lighter smoking history, no limit on time quit and one additional risk factor) the mean predicted risk for participants with lung cancer was 2.39% as compared to 1.83% for those without lung cancer but the difference was not statistically significant; p=0.2507. However, the incidence of lung cancer was the same for NCCN Group 2 as for the complete sample. NCCN Group 2 model AUC was 0.634 (95% CI 0.522 – 0.746), the sensitivity and specificity of the model at the recommended screening threshold were 64.7% and 56.0%, respectively and PPV was 4.2%.

Conclusions: Lung cancer risk prediction model, PLCOm2012noEd, predictive performance in a clinical lung cancer screening program was adequate to help patients and their physicians assess individual risk of lung cancer relative to the recommended model risk screening threshold (1.51%) and to supplement USPSTF and CMS screening program entry criteria for shared decision making discussions. Model risk predictive capability for the NCCN Group 2 subgroup did not match actual screening program lung cancer results.

Chapter 1: Introduction

The burden of lung cancer on US society is high. The cost of medical care for lung cancer is projected to be $14.8 billion per year by 2020; 8.5% of the total cost of medical care for all cancers.[1] Lost productivity contributes another $38.9 billion per year[2]; 27% of the total for all cancers. Lung cancer claims almost 160,000 lives a year; more than colon, breast, pancreatic, and prostate cancer combined.[3] Almost twice as many women die of lung cancer as die of breast cancer.[3] There are 221,000 newly diagnosed cases every year; about 65% of them in never or former smokers.[3]

The $250 million National Lung Screening Trial (NLST) was the first to show that annual screening for lung cancer with low dose CT (LDCT) scans reduced lung cancer mortality.[4] The NLST was a large randomized controlled clinical trial comparing annual screening with LDCT scan to chest x-ray (CXR). 53,456 current and former heavy smokers were enrolled between 2002 and 2004 and were randomized to either 3 annual screenings with LDCT or CXR. After 6.5 years of follow-up, the LDCT group had a 20% lower mortality than the CXR group.[4]

The NLST was stopped in November 2010 after an interim review showed the 20% reduction in mortality for the LDCT arm. The results of the trial were published on-line by the New England Journal of Medicine in June of 2011.[4] The primary endpoint was lung cancer mortality. Secondary endpoints included all-cause mortality, lung cancer incidence, and screening and treatment related morbidity.[5] Shortly afterwards, in the fall of 2011, the National Comprehensive Cancer Network (NCCN) published guidelines for lung cancer screening.[6] The guidelines recommended screening 2 groups; NCCN Group 1 was essentially the NLST group; ages 55 to 74, 30 or more pack year smoking history, current or former smokers quit within the last 15 years. People in NCCN Group 2 were > 50 years, 20 or more pack year history of smoking, current or former smokers with no limit on time quit and with at least one additional risk factor; a

personal history of a smoking related malignancy, a history of lung cancer in a first degree relative, having COPD or pulmonary fibrosis, or exposure to one or more of several known carcinogens including radon, asbestos, silica, cadmium, and diesel fumes. A 30 pack year history would be smoking a pack a day for 30 years or 2 packs a day for 15 years or essentially any combination of packs per day and years of smoking that add up to 30.

Lung cancer screening programs were started at several medical institutions across the US after the publication of the NLST results and the NCCN guidelines. Screening criteria were not uniform across the sites. Some screened only the NLST population (same as NCCN Group 1) some screened both NCCN Group 1 and 2. The Lahey Hospital & Medical Center (Lahey) started their Rescue Lung, Rescue Life community benefit lung cancer screening program in Jan 2012 and offered screening to both NCCN Group 1 and 2.[7,8] Lahey has the largest clinical lung cancer screening program in the US having screened over 3600 patients to date and performed nearly 8000 LDCT scans.

Lung cancer screening with LDCT scan is now recommended by the United States Preventive Services Task Force (USPSTF) and Centers for Medicare & Medicaid Services (CMS) and covered by private insurance and Medicare/Medicaid for a group of high risk current and former smokers.[9,10] This group includes people from age 55 to 77 (55-80 for USPSTF), current and former smokers quit within the last 15 years, with a 30 or greater pack year smoking history. There are estimated to be approximately eight million people in the United States that are eligible for lung cancer screening.[9] Screening this recommended high risk group is predicted to save 722,000 to 1,625,000 quality adjusted life years (years of life left adjusted for health related quality of life) over a 15 year period and to save 12,500 lives per year.[9] The criteria for screening are not inclusive of most patients meeting the NCCN Group 2 recommended to screen guidelines.

The CMS guidelines for screening include a required separate shared decision making (SDM) visit with a qualified health care practitioner, essentially a physician or a nurse practitioner, including use of a decision aid (DA) prior to entering a lung cancer screening program.[10]

The decision to undergo lung cancer screening involves tradeoffs between early detection and the potential for false positives, the risk of a surgical intervention for benign disease, the risk for treatment of indolent disease, and exposure to radiation (similar to a mammogram) every year.[11] Use of a risk prediction model as part of the shared decision making discussion helps to individualize the probability of lung cancer and provides patients and physicians with better information for shared decision making.

Chapter 2: The Study

The study is a retrospective IRB approved evaluation of the predictive performance of a six year lung cancer risk prediction model using the Lahey Rescue Lung, Rescue Life lung cancer screening program clinical data base and program participant clinic medical records for Lahey physician referred patients enrolled in the lung cancer screening program between January 1, 2012 and November 30, 2015. The model is an independently developed risk prediction model, developed using the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial data,[2,13] and validated with both PLCO and NLST data, for the six year probability of developing lung cancer.

Goal

The goal of the study is to evaluate the predictive capability of a reparameterized Tammemagi et al.[13] PLCOm2012 lung cancer screening risk prediction model in a clinical lung cancer screening program.

Hypothesis: The PLCOm2012 validated 6 year lung cancer risk prediction model, reparametrized without the education predictor variable, will show significant correlation of predicted risk to lung cancers found in a large US clinical LDCT scan lung cancer screening program.

Primary objective: Evaluate the predictive performance of the reparametrized Tammemagi PLCOm2012 six year lung cancer risk prediction model in a clinical lung cancer screening program.

Secondary objective: Evaluate the predictive performance of the reparametrized Tammemagi PLCOm2012 six year lung cancer risk prediction model, for the subgroup of patients meeting the NCCN Group 2 criteria for lung cancer screening, in a clinical lung cancer screening program.

Chapter 3: Background

The CMS requirement for a shared decision making visit using one or more decision aides prior to entering a lung cancer screening program is the first time reimbursement has been tied to the use of decision aids and shared decision making.

Despite many recommendations to use decision aids,[14-16] shared decision making is not widely used in clinical practice.[17] Surveys of primary care physicians and pulmonologists, those most likely to recommend LDCT screening for patients, show a lack of awareness of LDCT screening and a lack of understanding of screening guidelines and requirements.[18,19] Although DAs and SDM are sometimes used in prostate, colon and breast cancer screening, [20-23] they are not a requirement for CMS reimbursement as they are for lung cancer screening.

A recent Cochrane update of decision aids concluded that there was high quality evidence to show use of decision aides improved patient knowledge about options and reduced decisional conflict as compared to standard practice.[20,24] There was medium quality evidence that use of decision aides increased patient involvement in decision making and improved risk perceptions

when probabilities were included in the decision aid.* All of the included studies were randomized clinical trials.

Since shared decision making and the use of a decision aid is required for CMS reimbursement for lung cancer screening, it is important to assess the performance of lung cancer risk models in the clinical setting. Having an individualized predicted risk of lung cancer during the shared decision making discussion can help the patient and the physician in the decision to enter or not enter a lung cancer screening program.

A study of existing cancer risk prediction models found that many are being made available in the primary care setting.[25] The study identified several challenges for implementing these models including, choosing the "right" risk model and setting the threshold for intervention. Choosing a risk threshold for cancer screening is a trade between sensitivity and specificity, or in other words, a trade between finding more cancers with an increased possibility of a false positive (low risk selection threshold) and finding fewer cancers but having fewer false positives (higher selection risk threshold).

Although there are several lung cancer risk models that have been developed over the last 10 to 15 years,[26,27] the only one validated with the NLST data was the Tammemagi et al. six year lung cancer risk prediction model, PLCOm2012.[13] The model predicts an individual's probability of developing lung cancer in the next six years based on their individual risk factors. This model was initially developed and validated in the PLCO Cancer Screening Trial,[12] and subsequently modified for NLST criteria and validated with NLST data.[13] The results showed good predictive value for the validation data set and the model was more sensitive than the NLST criteria for lung cancer screening selection.[13] Both the USPSTF and CMS lung cancer screening guidelines

* High quality of evidence indicated research is very unlikely to change the confidence in the estimate of the effect and medium quality that further research is likely to have an impact on the confidence of the estimate of the effect and may change it.

were based on the NLST.[9,10] The PLCOm2012 was the only model validated with the NLST data and also had the best predictive performance for risk of lung cancer,[26,27] thus it was the model chosen for this study. Additionally, some of the other models had predictor variables that were not available for the study clinical lung screening program sample.[27]

The PLCOm2012 model is available for free download in Excel from Brock University.[28] The model is also used as part of the web-based, University of Michigan-developed, lung cancer screening decision aid.[29] However, the predictive capability of the PLCOm2012 model has not been evaluated in a clinical lung cancer screening program.

The USPSTF and CMS criteria for selecting screening participants are based on age, smoking history and time since quit smoking for former smokers.[9,10] Tammemagi et al. have presented analyses showing selection of screening participants may be improved by using the PLCOm2012 six year probability of developing lung cancer model.[30] Comparing results in the PLCO smoker intervention group sample using the PLCOm2012 risk threshold of 1.51% vs. USPSTF criteria resulted in improved sensitivity, specificity, and PPV.[30] Lung cancer screening risk models are not currently approved to be used for selecting participants for lung cancer screening. Participants in a clinical lung cancer screening program must meet USPSTF/CMS criteria to be eligible for insurance or Medicare reimbursement.

Evidence that USPSTF/CMS screening selection may not include many at high risk of lung cancer includes a recent journal article that assessed the use of a reparameterized PLCOm2012 model (family history of lung cancer predictor variable removed) to predict risk of lung cancer using National Health and Nutrition Examination Survey (NHANES) data and compared the results to CMS screening selection criteria.[31] The results from the study showed that more African American men than men of other races/ethnicities had high model predicted six year risk of lung cancer without meeting the CMS screening selection criteria. More Hispanics had low model predicted six year risk of lung cancer but met the CMS screening

selection criteria. This would result in fewer high risk African Americans meeting screening criteria and getting screened; potentially increasing the current disparity in lung cancer mortality between African American men (75.7 deaths per 100,000 per year) and other demographic groups (average all groups 45.8 deaths per 100,000 per year).[31] It may also result in more Hispanic men with lower risk of lung cancer entering screening programs with less potential benefits and greater potential harms.

Another recent study did a secondary analysis of the PLCO lung cancer data and found that current smokers with a 20 to 29 pack year history of smoking have a similar risk of lung cancer as former smokers, quit 15 years or less, with a 30 or more pack year history of smoking.[32] The group of former smokers, quit 15 years or less with a 30 pack year history are eligible for screening while current smokers with a 20 to 29 pack year history are not.

Two other recent studies retrospectively evaluated whether people with a lung cancer diagnosis would have been eligible for screening based on USPSTF screening criteria. One study evaluated the predicted risk in lung cancer patients that had undergone resection and compared the risk for those meeting and those not meeting USPSTF screening criteria to recommended screening risk probability thresholds. The Tammemagi PLCOm2012 model was used to calculate predicted risk. Almost 50% of those not meeting USPSTF criteria had model predicted risk levels above the model screening risk threshold.[33] The other study found that two thirds of people with lung cancer in 2 cohorts of almost 6000 patients in Olmstead County, Minnesota did not meet USPSTF screening selection criteria.[34]

Identifying and expanding the lung cancer screening selection criteria to others at equivalent or higher risk without another large randomized study will likely require a model based approach. Validating a lung cancer risk model in clinical practice would be a first step in using the model for lung cancer screening selection potentially addressing high risk individuals not eligible for screening under the current guidelines. Development and validation of lung cancer screening

risk prediction models are also critical for refining and improving the selection of screening participants to maximize the benefits and minimize the harms of screening. Confidence in use of risk models for screening participant selection will improve with demonstrated model performance in clinical lung cancer screening programs.

To my knowledge, this study is the first to evaluate the predictive capability of any lung cancer risk prediction model in a clinical lung cancer screening program.

Chapter 4: Pedigree of the Screening Program Study Model

The Tammemagi et al. initial lung cancer risk prediction model, PLCOm2011, was developed and validated in the PLCO Cancer Screening Trial participants.[12] The PLCO was an NIH sponsored general population multi-center randomized controlled trial that enrolled 154,901 people ages 55 to 74, with no known cancers, between 1993 and 2001.[35] For lung cancer screening, the intervention was four screenings with chest x-ray, a baseline and 3 annual screens, with follow-up to year 13 or December 2009.[36] Never smokers were not screened after April 1995.[36] The control arm was usual medical care with follow-up for up to 13 years after enrollment or to December 2009.[36]

Two models predicting the risk of developing lung cancer in the next 9 years were developed; a general population model and an ever smoker model.[12] The general population model was developed using prospective data from 70,962 participants; the smoker model used a sub-group of 38,254 participants. Follow-up was for a median of 9.2 years (inter-quartile range 7.5 -10.7). The validation data set had 44,223 PLCO intervention arm participants.

Both models were binary logistic regression models. Logistic regression models have binary outcomes (dependent variables) and include predictor variables that can be both categorical and continuous. The outcome for both of these models was a binary indicator of a lung cancer diagnosis. Predictor variables for the second model, the ever smoker model, PLCOm2011, were

age, level of education, BMI, family history of lung cancer, COPD, having had a chest x-ray in the previous 3 years, current or former smoker status, pack year smoking history, duration of smoking and, for former smokers, time quit.

The Tammemagi et al. six year lung cancer risk prediction model, PLCOm2012, was a modification of the PLCOm2011 ever smoker model, described in the previous paragraph, to make it compatible with the shorter follow up time in the NLST study data and included lung cancers identified through the first 6 years of follow-up.[13]

Tammemagi et al. developed the PLCOm2012 model using the PLCO control group of current and former smokers with follow-up truncated at 6 years.[13] Their stated approach was to consider predictor variables for inclusion that were literature- recognized lung cancer risk factors and all of the PLCOm2011 predictor variables.[13] They validated the PLCOm2012 model in the PLCO intervention group of current and former smokers, the NLST group and the PLCO intervention group participants that met NLST entry criteria.[13] The primary difference for the latter group being in the NLST requirement for a 30 pack year or greater smoking history and less than 15 years quit while PLCO enrolled participants with any smoking history, including never smokers. Thus the PLCO subgroup of former and current smokers (used for model development) included long term quit light smokers while the NLST did not.

As compared to PLCOm2011, the PLCOm2012 model predictor variables included race/ethnic group and personal history of cancer and didn't include history of a chest x-ray in the last 3 years. Pack years smoking history was split into 2 variables, smoking intensity and smoking duration. Smoking intensity (average number of cigarettes smoked per day) had a non-linear association with the outcome variable. The treatment of the non-linear effect used multivariable fractional polynomials.[26] The treatment of non-linear variables varied between the 2 models. The PLCOm2011 used restricted cubic splines for treatment of non-linear variables.[26]

For this study, I retrospectively evaluated the PLCOm2012 model performance in the Lahey lung cancer screening participants using the screening program data base and Lahey clinical medical records. All of the predictor variable data were available for the Lahey screening program participants except for education. Dr. Martin Tammemagi provided a reparameterized six year lung cancer risk prediction model, PLCOm2012noEd, without the education parameter for this study (M. Tammemagi, PhD e-mail communication, September 4, 2015). The reparameterized model was developed using the same data set used for development of PLCOm2012.

Chapter 5: Methods

5.1 Study Design

This study is a review of the predictive capability of the PLCOm2012noEd, six year lung cancer risk prediction model using data collected retrospectively from Lahey physician referred patients enrolled in the Lahey Hospital & Medical Center lung cancer screening program between January 1, 2012 and November 30, 2015. All patients met National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology: Lung Cancer Screening v1.2012 (NCCN Guidelines®) high risk criteria for lung cancer (Table 1) and had a physician order for CT lung screening. Patients were not eligible for screening if they had known metastatic disease, had been diagnosed with lung cancer within the previous five years, or had symptoms concerning for lung cancer. The CT lung screening program staff confirmed all participant eligibility at program entry. The study was approved by both the Lahey Hospital & Medical Center and the University of Connecticut Heath Center IRBs with a waiver of individual patient consent.

Table 1 – National Comprehensive Cancer Network (NCCN) Group 1 and Group 2 Lung Cancer Screening Criteria

	NCCN Group 1 (n= 1715)	NCCN Group 2 (n=587)
Age	55 – 74	50 – 74*
Smoking History	≥ 30 pack years	≥ 20 pack years
Smoking Status	Current or former	Current or former
Quit Duration	< 15 years	Any time
Additional Risk Factors	None required	At least one of the following required: ❖ Personal history of smoking related cancer (excluding metastatic disease) ❖ Family history of lung cancer (parent, sibling, or child) ❖ Personal history of chronic lung disease ❖ Occupational exposure to known lung carcinogen(s)**

*>50 in NCCN Guidelines®; **Carcinogens include arsenic, asbestos, beryllium, cadmium, soot, chromium, diesel fumes, nickel, silica, coal smoke, and radon (occupational or documented residential).

5.2 Description of the Lung Cancer Screening Program Study Sample

The screening program sample included 2302 Lahey physician referred patients in the lung cancer screening program between Jan 1, 2012 and November 30, 2015. Thirty five patients in the program during this period were not included in the 2302 sample due to incomplete data. Of those 35, one was removed due to lack of information on height in the medical record (needed to calculate BMI) and 34 were removed due to declining to provide race and race not being available in the medical record. Table 2 describes the screening program sample demographics. The participants range from age 48 to 77 with a mean of 62.5 years. All participants underwent a baseline scan (except for 3 participants in the program with a long delay between program entry and their first scheduled screen). Mean duration in the program was 1.86 years with a standard deviation of 1.04 years. In the screening program sample, 2269 (98.6%) of participants were white, 1205 (52.3%) current smokers, 1028 (44.7%) female, 538 (23.3%) had a 1st line relative diagnosed with lung cancer, 651 (28.3%) had a personal history

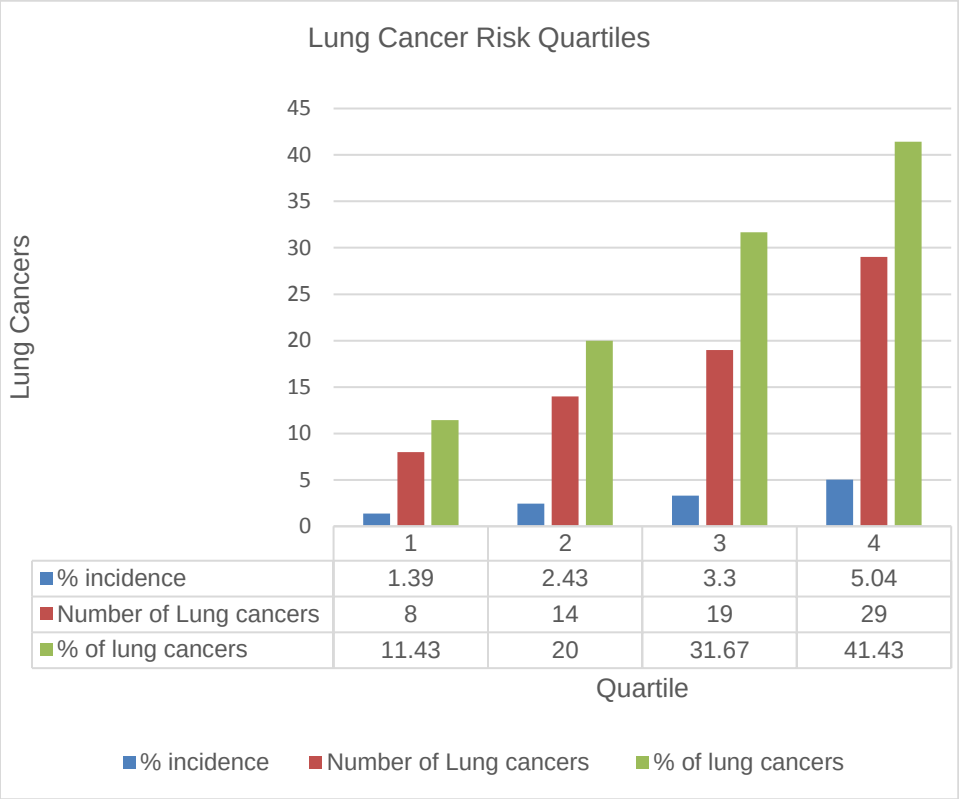
of cancer and 1690 (73.4%) had COPD, chronic bronchitis or emphysema. There were 70 lung cancers diagnosed. Eight (11.3%) lung cancers diagnosed were in participants in the first quartile of PLCOm2012noEd model predicted six year risk of developing lung cancer, 14 (20%) in participants in the second quartile, 19 (31.7%) in participants in the third quartile and 29 (41.43%) in participants in the fourth quartile (Figure1).

Table 2 – Lung Cancer Screening Program Study Sample Characteristics

Complete Group (n=2302)	
Age (years)	62.5 ± 6.22
BMI	29.13 ± 5.90
Duration in Program (years)	1.86 ± 1.04
White Race	2269 (98.6%)
Years Smoked	37.03 ± 9.51
Avg Cigarettes smoked per day	26.6 ± 11.4
Years Quit (former smokers)	10.6 ± 9.0
Smokers	1205 (52.3%)
Female	1028 (44.7%)
NCCN Group 2	587 (25.5%)
COPD/Emphysema/Chronic Bronchitis	1690 (73.4%)
Personal History Cancer	651 (28.3%)
Family History Lung Cancer	538 (23.3%)
Lung Cancers Diagnosed	70 (3.0%/Annualized 1.6%)
Deaths (all causes)	27 (1.17%/Annualized 0.63%)

Values are reported as mean ± standard deviation.

Figure 1 – Screening Program Diagnosed Lung Cancers by Quartile



5.3 Data Sources and Collection

Patient data for the model predictor variables and screening program sample characteristics for comparison to the model development sample was obtained from the sources shown in Table 3.

The data downloaded from the lung cancer screening program data base included the patient responses to the lung cancer screening eligibility questions, age, sex, medical record number (MRN), date of initial contact, dates of LDCT scans, LDCT scan results including indications for lung cancer, extent if any of emphysema (mild, moderate, marked) and incidental findings.

A paper intake form (Figure 2) was used by the screening program personnel to determine eligibility for screening when the patient called to schedule their CT scan. All patients were asked the first 4 questions; a) age, b) smoking history; packs per day smoked and number of years smoking, c) current smoking status and for former smokers time since quit, and d) if the

patient had a personal history of cancer including making sure no there was no history of lung cancer in the previous 5 years. Pack years were calculated using the Smoking Pack Years on-line calculator.[37] The data from the form was entered into the lung cancer screening data base. For smoking history only current smoking status, years quit for former smokers and the calculated pack years were entered into the data base. Average packs smoked per day and number of years smoked were collected from the paper record intake forms for this study.

Table 3 – Screening Program Study Predictor Variable Data Sources

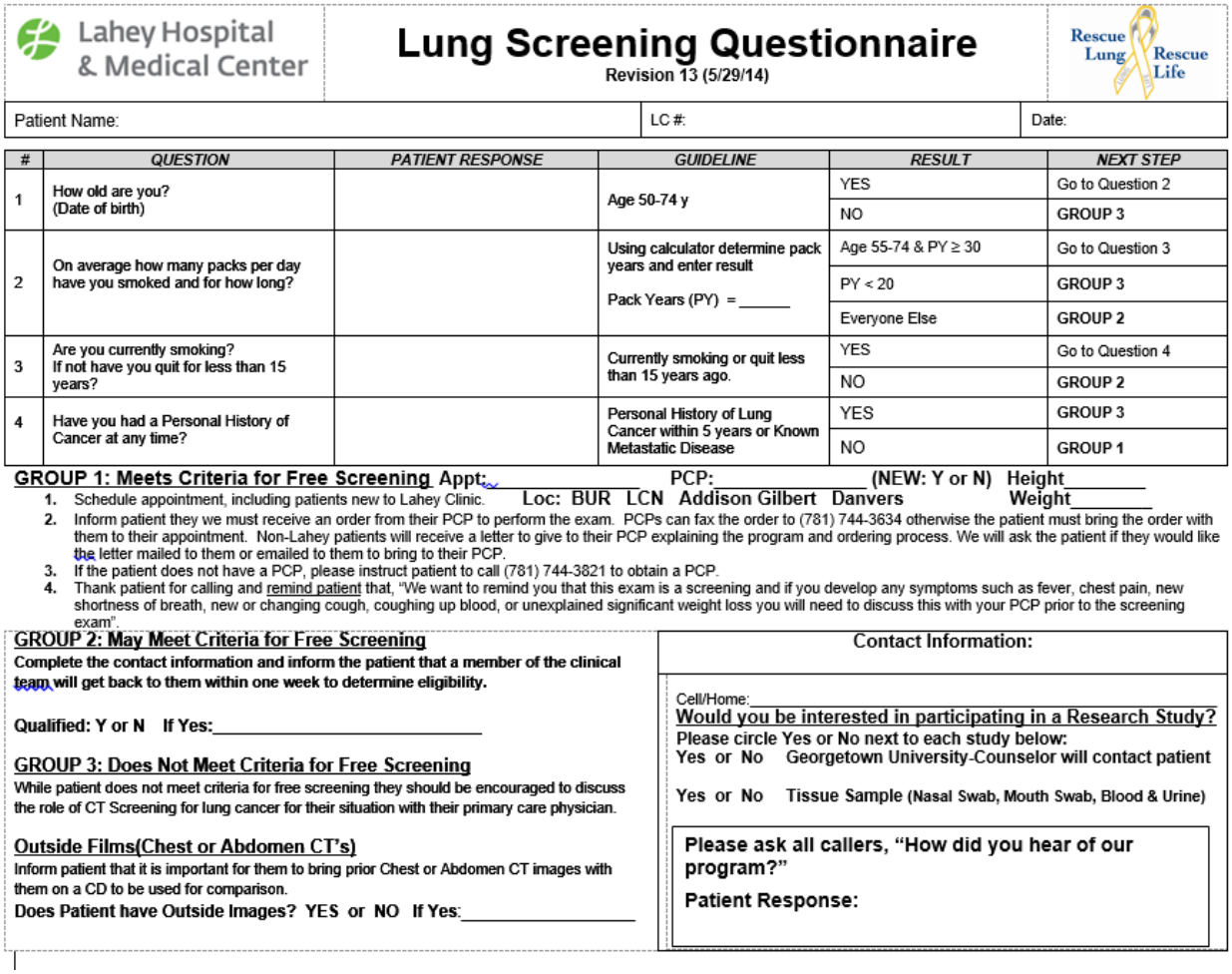
Predictor Variable	Description	Data Source
Age	Age in year at program entry	Screening Data Base
Race/Ethnicity	Self report: White/Black/Hispanic/Asian/Native Hawaiian or Pacific Islander	EPIC Medical Record
BMI	Kilograms/height in meters ^2 was calculated from the participant latest recorded height and weight in the medical record.*	EPIC Medical Record
Family history of lung cancer	Family history of lung cancer was coded yes if either the participant self-reported a 1st line relative with lung cancer or there was a note in any of the medical records of a family history of lung cancer in a 1st line relative	Screening Data Base or EPIC Medical Record
Personal history of cancer	Participant self-reported history of cancer, any cancer including basal cell carcinoma, or the medical record indicated any type of cancer diagnosis.	Screening Data Base or EPIC Medical Record
COPD	Participant self-reported COPD, emphysema or chronic bronchitis OR the medical record indicated a diagnosis of COPD, emphysema or chronic bronchitis OR emphysema was noted on the radiologist report for the baseline LDCT scan	Screening Data Base, EPIC Medical Record, or Radiologist note of emphysema on initial LDCT scan
Years smoked	Participant self-report	Paper intake forms
Years quit	Participant self-report	Screening Data Base
Duration	Time from the initial LDCT scan to November 30, 2015.**	Screening Data Base
Smoking status	Participant self-report	Screening Data Base
Average number of cigarettes smoked	Participant self-report	Paper intake forms
	* If significant weight changes were noted in the medical records since the initial LDCT scan, weight was obtained from a record closest in time to the initial LDCT scan (4 participants)	
	** Except for 3 participants in the program with first LDCT scans scheduled but not yet completed; duration was calculated from program entry	

If based on the answers to the first four questions, the patient met screening criteria for either NCCN Group 1 (age, smoking history, time quit) or Group 2 (age, smoking history and had a personal history of cancer), no further questions were asked. Patients who met the age and smoking history criteria for NCCN Group 2 but did not have a personal history of cancer were asked about the additional risk factors, one at a time until a risk was identified or until all the questions were asked and answered (Figure 3). Since not all participants were asked all of the questions, the data in the lung cancer screening data base for the predictor variables family history of lung cancer and patient history of COPD were incomplete. The protocol for intake screening was changed February 2, 2015 to ask all patients all of the questions and to enter the number of packs per day average and number of years smoked into the data base in addition to

pack years. Regardless of this change, the same protocol was used for data sources for all patients in this study.

For screening program variables not available from the screening data base, patient medical records in the Lahey EPIC system were examined (chart read). These records included demographic information, physician notes, pathology reports, and imaging reports.

Figure 2 – Screening Program Intake Form



Lahey Hospital & Medical Center

Lung Screening Questionnaire

Revision 13 (5/29/14)

Rescue Lung Rescue Life

Patient Name:	LC #:	Date:

#	QUESTION	PATIENT RESPONSE	GUIDELINE	RESULT	NEXT STEP
1	How old are you? (Date of birth)		Age 50-74 y	YES	Go to Question 2
				NO	GROUP 3
2	On average how many packs per day have you smoked and for how long?		Using calculator determine pack years and enter result Pack Years (PY) = _____	Age 55-74 & PY ≥ 30	Go to Question 3
				PY < 20	GROUP 3
				Everyone Else	GROUP 2
3	Are you currently smoking? If not have you quit for less than 15 years?		Currently smoking or quit less than 15 years ago.	YES	Go to Question 4
				NO	GROUP 2
4	Have you had a Personal History of Cancer at any time?		Personal History of Lung Cancer within 5 years or Known Metastatic Disease	YES	GROUP 3
				NO	GROUP 1

GROUP 1: Meets Criteria for Free Screening Appt:_____ PCP:_____ (NEW: Y or N) Height_____

Loc: BUR LCN Addison Gilbert Danvers Weight_____

1. Schedule appointment, including patients new to Lahey Clinic.
2. Inform patient they we must receive an order from their PCP to perform the exam. PCPs can fax the order to (781) 744-3634 otherwise the patient must bring the order with them to their appointment. Non-Lahey patients will receive a letter to give to their PCP explaining the program and ordering process. We will ask the patient if they would like the letter mailed to them or emailed to them to bring to their PCP.
3. If the patient does not have a PCP, please instruct patient to call (781) 744-3821 to obtain a PCP.
4. Thank patient for calling and remind patient that, "We want to remind you that this exam is a screening and if you develop any symptoms such as fever, chest pain, new shortness of breath, new or changing cough, coughing up blood, or unexplained significant weight loss you will need to discuss this with your PCP prior to the screening exam".

GROUP 2: May Meet Criteria for Free Screening

Complete the contact information and inform the patient that a member of the clinical team will get back to them within one week to determine eligibility.

Qualified: Y or N If Yes:_____

GROUP 3: Does Not Meet Criteria for Free Screening

While patient does not meet criteria for free screening they should be encouraged to discuss the role of CT Screening for lung cancer for their situation with their primary care physician.

Outside Films(Chest or Abdomen CT's)

Inform patient that it is important for them to bring prior Chest or Abdomen CT images with them on a CD to be used for comparison.

Does Patient have Outside Images? YES or NO If Yes:_____

Contact Information:

Cell/Home:_____

Would you be interested in participating in a Research Study?

Please circle Yes or No next to each study below:

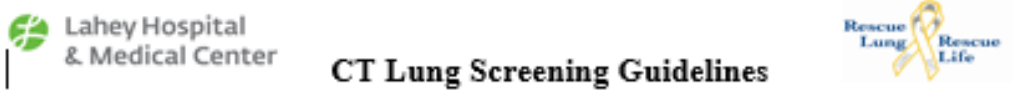
Yes or No Georgetown University-Counselor will contact patient

Yes or No Tissue Sample (Nasal Swab, Mouth Swab, Blood & Urine)

Please ask all callers, "How did you hear of our program?"

Patient Response:

Figure 3 – Screening Program – Follow-on Questions



All candidates for CT lung screening should be asymptomatic, have no known metastatic disease, and should not have a diagnosis of lung cancer within the past 5 years.

Pack-Years: _____	Smoking Status ○ Current ○ Former, quit _____years ago

NCCN High-Risk Groups Qualifying for CT Lung Screening

Group 1	Group 2
❑ 55-74 years old	❑ 50-74 years old
❑ Are currently a smoker or have quit within the past 15 years	❑ Have smoked at least a pack of cigarettes a day for 20+ years
❑ Have smoked at least a pack of cigarettes a day for 30+ years	❑ Have one additional lung cancer risk factor, not to include secondhand smoke exposure

NCCN Lung Cancer Risk Factors for Group 2 Qualification (one required)

1. Family history of lung cancer

○ Mother	○ Sibling
○ Father	○ Child

2. Personal history of chronic lung disease

○ COPD	○ Emphysema
○ Chronic bronchitis	○ Pulmonary fibrosis

3. Occupational exposure to 10 lung carcinogens

○ Arsenic	○ Chromium
○ Asbestos	○ Diesel Fumes
○ Beryllium	○ Nickel
○ Cadmium	○ Silica
○ Soot	○ Coal Smoke

4. Radon Exposure

○ **Documented Residential**

Occupational

○ **Mining**
○ **Firefighter**
○ **Military-Active Combat**

5. Personal history of cancer (excluding known metastatic disease)

○ Lung Cancer (greater than five years ago)	○ Colon
○ Lymphoma	○ Kidney
○ Head and neck	○ Pancreas
○ Esophageal	○ Stomach
○ Bladder	○ Other smoking related cancer (_____)
○ Cervix	

If the patient meets above criteria please order for "**CT Lung Screening Request**" **(IMG11242) in EPIC** and instruct the patient to call **781-744-5658** to see if they qualify for the program.

5.4 Data Coding for Missing or Uncertain Data

Sensitivity analysis was performed to assess the impact of missing and uncertain data on model predictive capability as described in the next section (*5.5 Statistical Analysis*).

Participants with no response to family history of lung cancer in the lung cancer data base, or who reported being adopted, not knowing their family history, or not having any family history on file were coded as not having a family history of lung cancer. Participants with limited family history in the medical files and no information on family history of lung cancer were coded as not having a family history of lung cancer. These amounted to a total of 142 of 2302 participants:

- 22 adopted
- 75 NCCN Group 1 with limited family history
- 2 NCCN Group 2 with personal history of cancer and limited family history
- 43 with no family history on file in Group 1 or in Group 2 with a personal history of cancer

Of note, as part of the intake process protocol described in the previous section, participants in NCCN Group 2 who answered “no” to the question “Do you have a personal history of cancer?” were then asked if they had a family history of lung cancer. These participants were considered to be accurately coded as having no family history of lung cancer, even if they had limited or no family history on file, because they were asked, and answered, the family lung cancer history question. These participants are not included in the 142 described above. There are 17 of 2302 participants that met this criteria.

Participants with self-report of having a family history of lung cancer, contradicted by medical record review, were coded according to patient self-report. Participants not asked the question about family history of lung cancer with uncertainty around the family member with lung cancer based on chart review, were coded as having a family history of lung cancer. There were 10 of

2302 participants that met the first criteria and 13 of 2302 participants that met the second criteria.

Participants with a family history of unspecified cancer were coded as no family history of lung cancer. There were 71 of 2302 participants that met this criteria.

Participants with self-report of personal history of cancer, either yes or no, that was not confirmed or was contradicted by medical record review, were coded according to patient self-report. There were 14 of 2302 participants that met this criteria (4 coded as having a personal history and 10 coded as not having a personal history of cancer).

5.5 Statistical Analysis

As this study assessed an existing lung cancer risk prediction model in a different sample, it was important to delineate the differences between the samples used for model development and validation and the study sample. The model also evolved through time so understanding the differences in the models was important as well. As discussed in *Chapter 4: Pedigree of the Screening Program Study Model,* the baseline model was developed in the PLCO sample, then modified for the NLST sample criteria and then reparameterized for this study with the patient education predictor variable removed.

The differences in the PLCO, NLST and the screening program study design, demographics, screening scan frequency, data collection methods, and duration are described in *Chapter 6: Results*. Comparisons of the predictor variables and the odds ratios for the predictor variables for the PLCOm2011, PLCOm2012 and PLCOm2012noEd models are also described in *Chapter 6.* The PLCOm2012 and PLCOm2012noEd risk models were both run for the same individual with a high risk of lung cancer. The predicted risk of lung cancer for the individual was compared between the 2 models.

Demographic differences between the screening program sample, the model development sample and the NLST sample were assessed pairwise using Pearson's Chi-Square Test for categorical variables and the independent samples T test (2 sided) for continuous variables.

The reparameterized lung cancer risk prediction model (without the education predictor variable) provided by Dr. Martin Tammemagi (PLCOm2012noEd) was run for the complete Lahey sample.

Assessment of the predictive capability of a risk prediction model includes assessments of both discrimination and calibration. Discrimination is the ability of the model to identify the cases from the non-cases; in this study the ability of the PLCOm2012noEd model to identify participants in the screening program who will get lung cancer in the next six years from those that won't. Calibration is the ability of the model to accurately predict the level of risk that is observed. In this study, the accurate prediction of the six year risk of getting lung cancer, or in other words, how well did the number of observed lung cancers compare to the number of model predicted lung cancers. Gail et al. found that discriminatory performance was more important than good calibration in screening applications as compared to preventative interventions.[38] However, both calibration and discrimination are important for selecting the right individuals to screen. Particularly important is having good model calibration around the threshold for screening selection.[26] This is especially true for lung cancer screening models as the predicted absolute risk levels are small with peak frequency distributions close to the recommended screening threshold risk level.[30]

The model predictive capability was assessed for the complete sample and for the NCCN Group 1 and Group 2 subgroups. Discrimination was assessed using the area under the receiver operator curve (AUC). The receiver operator curve (ROC) is a plot of specificity and sensitivity characteristics across the full model predicted risk threshold levels; specifically a plot of 1-specificity (false positive rate) vs. sensitivity (true positive rate). An area under the curve (AUC)

of 1.0 represents perfect discrimination and an AUC of 0.5 is no better than random chance.[39] AUC's >0.5 and < 0.7 are considered poor discrimination, 0.7 to < 0.8 good discrimination and 0.8 to < 0.9 excellent discrimination.[26, 40] AUC's in samples with different characteristics and different follow-up times as compared to the development sample are likely to be lower than the AUC for the development samples.[26,39]

The Hosmer-Lemeshow goodness-of-fit test was used to assess calibration. The test forms sub-groups and compares predicted to observed risk. Typically 10 sub-groups are formed and the average observed and average predicted risk in each decile are compared to each other. The statistic follows a chi-squared distribution with degrees of freedom = number of subgroups – 2.[39] A small p value (<0.05) is an indication of statistically significant differences in the risks and thus poor calibration. In this study, the Hosmer-Lemeshow test compared the predicted risk of lung cancer to the observed risk of lung cancer for 10 sub-groups. There have been criticisms of the Hosmer-Lemeshow test over differences in test significance level for the same model/sample based on grouping of the data.[42] It is recommended to use the standard number of 10 groups for consistency.[42] For this study, in addition to the significance level of the test, the model calibration was assessed graphically for each decile, as described below, to evaluate calibration at the risk screening threshold, the area most important for good calibration in selection of individuals to screen.

Because the model outcome is a binary variable (lung cancer yes or no), plotting risk vs. outcome results in only 0 and 1 values for the outcome. A generally used graphical technique for assessing model calibration is to use a smoothing algorithm, typically the lowess algorithm, to estimate the probability of the predicted outcomes.[39] I decided not to assess calibration using the lowess algorithm for the screening program model outcomes because of the small number of cases (70 yes lung cancers in 2302 complete sample). Instead, graphical representations using the mean predicted risks and mean observed outcomes of the Hosmer-

Lemeshow test were used for graphical calibration assessments. These were plots of the absolute value of observed – predicted mean risk by decile, predicted vs. observed mean risk by decile and predicted vs. observed number of lung cancers by decile.

Additional assessments of model performance included evaluation of the difference and statistical significance of mean risks between participants with and without lung cancer, sensitivity and specificity in identifying lung cancer by risk percentile, and the lung cancer risk probability distributions for participants with lung cancer as compared to participants without lung cancer; especially around the screening risk threshold of 1.51%. Sensitivity, specificity and PPV at the recommended screening risk threshold were evaluated for the complete sample and for the NCCN Group 1 and NCCN Group 2 sub-groups. The difference and statistical significance in model predicted risks and observed lung cancers between NCCN Group 1 and Group 2 participants, the percentage of NCCN Group 1 and Group 2 participants that met the 1.51% screening threshold and the number of lung cancers identified in NCCN Group 1 and Group 2 participants using this threshold were assessed.

For the sensitivity analysis the PLCOm2012noEd model was run for 5 separate scenarios as detailed below to assess the impact on model performance of missing data, uncertain data, and differences in model development and validation coding of present/absent conditions for selected predictor variables.

The impacts on model predictive performance were assessed by comparing each of the PLCOMm2012noEd model scenario results to the complete sample (baseline) model results for the difference in mean risk and the statistical significance of the difference in mean risk for participants with lung cancer vs. without lung cancer. Additionally, the AUC for the complete sample model results and the model AUC for each of the five scenarios was compared.

1. Reduced sample; participants that were adopted or participants that were not asked the question “do you have a family history of lung cancer?” and had no or limited family history on file were deleted from the baseline sample – 142 of 2302 participants were deleted from the sample. Of these participants, one had lung cancer; the remaining 141 did not. This reduced the number of lung cancers in the sample from 70 to 69.
2. Complete sample with only patient self-reported history of cancer coded as patient history of cancer. In other words, participants coded as “1” in the baseline sample for personal history of cancer based solely on a medical record review indicating a diagnosis of cancer were recoded to “0”, or no personal history of cancer. 224 of 651 patients were recoded from “1”, yes personal history of cancer to “0” no personal history of cancer.
3. Complete sample with only patient self-reported history of cancer, excluding non-melanoma skin cancer, coded as patient history cancer. 342 of 651 patients were recoded from “1”, yes personal history of cancer to “0” no personal history of cancer.
4. Complete sample with only patient self-reported COPD, emphysema, or chronic bronchitis coded as patient history COPD. In other words participants coded as “1” in the baseline sample for history of COPD based only on; 1) a medical record review indicating a diagnosis of COPD, emphysema or chronic bronchitis and/or 2) a report of emphysema on the initial LDCT in the radiologist report were recoded to “0” no personal history of COPD. 1241 of 1690 participants were re-coded from “1” yes history of COPD to “0” no history of COPD.
5. Complete sample with family history of unspecified type of cancer, coded as “0”, no family history of lung cancer, in the baseline now coded as “1”, yes family history of lung cancer. 71 of 1764 participants were re-coded from “0” no family history of lung cancer to “1” yes family history of lung cancer.

The significance level was set at $p \leq 0.05$ for all statistical analysis. All reported P values for the independent T test are 2 sided. All data are reported as mean +/- standard deviation, range or percentage as appropriate. All statistical analysis was performed using the statistical software platform R version 3.1.2.[43]

Chapter 6: Results

6.1 Comparison Screening Program to PLCO and NLST

Table 4 summarizes the design for the screening program as compared to PLCO and NLST. The PLCO study was a randomized population study with entrance criteria consisting primarily of age and being cancer free. The NLST was a prospective randomized clinical trial while the screening program study was a retrospective review of clinical data. Both the screening program study population and the NLST required heavy smoking histories, minimum time since quit for most, as well as age. The screening program study sample differs from the NLST in including the NCCN Group 2 of younger, lighter smokers, with no minimum time quit and having an additional risk factor for lung cancer.

Additionally, the screening program sample only had 1.86 years mean duration as compared to eleven and six years for the PLCO and NLST samples (the PLCO sample for model development was limited to 6 years), respectively. Both the PLCO and NLST studies included non-screen detected lung cancers. Most of these were during the follow-up period after the planned rounds of screening were completed. All of the screening program lung cancers were screen detected. As discussed previously, in the screening program study, all participants had an initial screen, T0, (except for 3 participants still in the program with a long delay between program entry and their first scheduled screen). Not all screening program participants had follow-on scans; 248 (10.8%) only had an initial screen, 1239 (53.8%) had a first annual screen (T1), 635 (27.6%) had a second annual screen (T2) and 173 (7.5%) had a 3rd annual screen

(T3). Only 4 participants had a 4th annual screen, T4 (0.17%). In comparison, in the PLCO and NLST studies all participants completed the baseline, T0, and planned annual screens T1, T2, and for PLCO, T3. The annual screens were administered at one year intervals for all participants unless found to have lung cancer or other pathology that would preclude annual screening. For the screening program study, annual scans were administered one year from the last screening scan or from the last diagnostic scan follow-up for suspicious findings on an exam whichever occurred last. The average time between T0 and T1 was 430 days or 1.2 years

PLCO data were collected via a self-reported survey at the beginning of the study and lung cancer diagnoses were ascertained via annual study update forms and verified with medical records.[36] Similarly, for the NLST, the participants completed a questionnaire including demographic and smoking behavior characteristics. Medical records were examined to determine lung cancer diagnosis.[5]

Table 4 – Screening Program Comparison Summary

Study	Study Type	Arms	Entry Criteria	Number of Screens and follow-up duration	Number of Participants	Data sources
PLCO	Prospective, Randomized, Population Study, 10 sites	Chest X-Ray vs Usual Care	Ages 55-74, any smoking history including non-smokers, no history colon, prostate, lung, ovarian cancer	4 annual screens T0, T1, T2, T3 (only T0,T1, and T2 for never smokers), 11 year follow-up, self-report in survey: demographics and medical history, medical record confirmation lung cancer diagnosis.	77,464	Self-report in survey: demographics and medical history, medical record confirmation lung cancer diagnosis.
NLST	Prospective, Randomized, 33 sites	Low Dose CT scan vs Chest X-Ray	Ages 55-74, 30 or more pack years, less than 15 years quit, no history lung cancer within 5 years	3 annual screens T0, T1, T2, 6 year follow-up	53,454	Self-report survey: demographics and medical history, medical record confirmation lung cancer diagnosis
Screening Program	Retrospective review clinical records, 2 sites in close proximity geographically	N/A - all screens Low Dose CT scan in clinical practice	Ages 55 to 74; 30 or more pack years and less than 15 years quit AND ages 50 to 74; 20 or more pack years, no criteria time quit with one additional risk factor for lung cancer, no history lung cancer within 5 years	Baseline screen and continued annual screens, mean time in program 1.84 years	2,302	Self-report smoking history, medical records review and self-report demographics and medical history, medical record confirmation lung cancer diagnosis.

6.2 Comparison of Lung Cancer Screening program Sample Characteristics to PLCO Model Development Sample and NLST Sample

Table 5 compares the screening program sample characteristics to the PLCO sample used to develop the PLCOm2012 model and to the NLST sample.[13] Age is similar across all 3 studies although the screening program sample is statistically older than the NLST sample ($p<0.0001$).

Family history of lung cancer is not statistically different between the screening program sample and NLST, 23.4% vs. 22.1%, respectively (p=0.156). All of the remaining characteristics are statistically significantly different between the screening program sample, and PLCO sample and the screening program sample and NLST sample (p<0.0001).

The screening program participants were primarily white, 98.6%, as compared to 88.5% for PLCO and 91.5% for NSLT. More than 25% of the screening program participants had a personal history of cancer as compared to 4.7% and 4.3% in PLCO and NLST, respectively. Almost three quarters of screening program participants had COPD as compared to 9.1% and 5.1% for PLCO and NLST, respectively. Smoking characteristics, although statistically significantly different between the screening program sample participants and NLST, are more similar to each other than to PLCO smoking characteristics. The screening program sample only had 1.86 years mean duration as compared to six years for the PLCO and NLST (the PLCO sample for model development was limited to first 6 years of the total 11 year follow-up).

The screening program coded any self-reported cancer, including basal cell skin cancer, as the participant having a personal history of cancer. Additionally, participants were coded as having a personal history of cancer if there was a cancer diagnosis in the medical record before the initial LDCT scan, regardless of patient self-report. The NLST study survey questions only asked about selected cancers; bladder, breast, cervical, colorectal, esophageal, kidney, larynx, lung, nasal, oral, pancreatic, pharynx, stomach, thyroid and transitional cell cancers.[44] PLCO entrance criteria specified no prostate, lung, colorectal or ovarian cancer history. Also, in the PLCO baseline survey, participants were asked to exclude basal cell carcinoma when answering the question about ever being diagnosed with cancer.[45]

Table 5 – Characteristics of Participants in the PLCO model development sample, the NLST sample and the screening program sample

Variables	PLCO Model Dev't Sample	NLST Sample	Lahey Sample	P Lahey vs PLCO	P Lahey vs NLST
Sample size	n=39,928	n=53,202	n=2302		
Number of Lung Cancers	709	1916	70		
Age (years)	62.5 (5.3)	61.4 (5.0)	62.5 (6.22)	1.00	<0.0001
Sex				0.0164	0.0005
Female	16816 (42.1%)	21811 (41%)	1028 (44.7%)		
Male	23112 (57.9%)	31391 (59%)	1274 (55.3%)		
Race/Ethnicity				<0.0001	<0.0001
White	35308 (88.5%)	47864 (91.5%)	2269 (98.6%)		
Black	2234 (5.6%)	2331 (4.5%)	9 (0.4%)		
Hispanic	814 (2.0%)	661 (1.3%)	7 (0.3%)		
Asian	1198 (3.0%)	1095 (2.1%)	15 (0.6%)		
Native American	242 (0.6%)	189 (0.4%)	2 (0.1%)		
Pacific Islander	108 (0.3%)	193 0.4%)	0 (0%)		
Body mass index (kg/m2)	27.4 (4.8)	28.0 (5.1)	29.1 (5.9)	<0.0001	<0.0001
Personal History Cancer				<0.0001	<0.0001
No	38038 (95.3%)	50895 (95.7%)	1651 (71.7%)		
Yes	1874 (4.7%)	2307 (4.3%)	651 (28.3%)		
Family history lung cancer				<0.0001	0.1558
No	33906 (88.2%)	40880 (77.9%)	1764 (76.6%)		
Yes	4523 (11.8%)	11608 (22.1%)	538 (23.4%)		
COPD				<0.0001	<0.0001
No	35944 (90.9%)	50494 (94.9%)	612 (26.6%)		
Yes	3593 (9.1%)	2689 (5.1%)	1690 (73.4%)		
Smoking Status				<0.0001	0.0001
Former	31985 (80.1%)	27590 (51.9%)	1097 (47.7%)		
Current	7924 (19.9%)	25612 (48.1%)	1205 (52.3%)		
Smoking Intensity (cig/day)	24.9 (14.7)	28.4 (11.5)	26.6 (11.4)	<0.0001	<0.0001
Smoking duration (years)	27.7 (13.8)	39.8 (7.3)	37.0 (9.5)	<0.0001	<0.0001
Smoking quit time - former smokers (years)	20.2 (12.1)	7.3 (4.8)	10.6 (9.0)	<0.0001	<0.0001
Follow-up/Duration	6.00	6.00	1.86 (1.04)	<0.0001	<0.0001

Table data reformatted from: Tammemagi MC, Katkiha HA, Hocking WG, et al. Selection criteria for lung-cancer screening, Supplement. *N Engl J Med.* 2013;368(8):728-236 DOI: 10.1056/NEJMoa1211776

In both the PLCO and NLST studies, COPD was self-reported in a baseline survey. Although the definition of COPD includes both emphysema and chronic bronchitis,[46] the NLST reported all three of these as separate items. Combining the NLST survey results for COPD, emphysema and chronic bronchitis results in 17.5% of the NLST sample having COPD.[47] The PLCO

baseline survey listed emphysema and chronic bronchitis as separate questions,[48] while the follow-on supplemental survey only listed emphysema.[49] Participants in the screening program were coded as having COPD if any one of the following criteria were met: self-report of COPD during program intake interview, medical record review indicating a diagnosis of COPD, emphysema, or chronic bronchitis, or report of emphysema on the radiologist report for the baseline LDCT scan.[50] The program intake interview protocol (previously discussed in *Section 5.3 Data Sources and Collection*) did not include asking all participants about a COPD diagnosis, thus patient self-report of COPD data was biased and not feasible to use for model evaluation.

Table 6 is a comparison of the screening program sample to the model development sample and the NLST sample segregated by non-cases (no lung cancer) and cases (yes lung cancer). The statistical significance of differences (p value) between the cases and non-cases for each sample are shown. Both the model development and the NLST sample were statistically significantly different between the cases and non-cases for essentially all of the listed demographic and participant characteristics. In other words, in both the PLCO and NLST samples people with lung cancer were significantly different from people without lung cancer for all of the characteristics listed. Only sex was not significantly different for people with lung cancer as compared to people without lung cancer for the NLST sample. For the screening program sample, most characteristics between people with lung cancer and people without lung cancer were not statistically significantly different. Only BMI and COPD were significantly different between the cases and non-cases. People with lung cancer in the screening program sample had lower BMIs than people without lung cancer and people with lung cancer were more likely to have COPD than people without lung cancer.

6.3 Comparison of Reparameterized study model, PLCOm2012noEd, to PLCOm2012 model

Table 7 compares the PLCOm2011, PLCOm2012 and PLCOm2012noEd logistic regression models. The variables for PLCOm2012 and PLCOm2012noEd are the same except PLCOm2012noEd does not include the education predictor variable. In PLCOm2012, four variables are centered; age, education, smoking duration and smoking time quit. The 3 of these variables that are included in PLCOm2012noEd are not centered. The difference in the model constant accounts for differences due to centering. Comparison of odds ratios and p values between the 2 models show they are similar with 5 of the 11 predictor variables being the same to 2 significant digits (age, BMI, personal history of cancer, smoking duration, and smoking time quit) and the others (race, COPD, family history of lung cancer, smoking status, and smoking intensity) are the same between the models to the first significant digit.

The AUC for the reparameterized model, PLCOm2012noEd, in the development sample was 0.8007 (M. Tammemagi, PhD e-mail communication, September 4, 2015) as compared to an AUC of 0.803 for PLCOm2012 in the same development sample.[13]

Table 6 – Model Development, Screening Program, and NLST Sample Characteristics

Variables	PLCO Development data set				Lahey data set				NLST data set			
	No Lung Cancer	Yes LC	Total	p	No Lung Cancer	Yes LC	Total	p	No Lung Cancer	Yes LC	Total	p
	n= 39,219 (n=98.2%)	n=709 (n=1.8%)	n=39,928 (100%)		n=2232 (97.0%)	n=70 (3.0%)	n=2302		n=51,286 (96.4%)	n=1916 (3.6%)	53,202 (100%)	
Age (years)	62.4 (5.3)	64.6 (5.2)	62.5 (5.3)	<0.0001	62.5 (6.2)	63.8 (7.2)	62.5 (6.22)	0.1314	61.3 (5.0)	63.7 (5.3)	61.4 (5.0)	<0.0001
Sex												
Female	16545 (98.4%)	271 (1.6%)	16816 (42.1%)	0.035	996 (96.9%)	32 (3.1%)	1028 (44.7%)	0.9532	21037 (96.5%)	774 (3.5%)	21811 (41%)	0.6
Male	22674 (98.1%)	438 (1.9%)	23112 (57.9%)		1236 (97.0%)	38 (3.0%)	1274 (55.3%)		30249 (96.4%)	1142 (3.6%)	31391 (59%)	
Race/Ethnicity								0.902				
White	34678 (98.2%)	630 (1.8%)	35308 (88.5%)	0.01	2199 (96.9%)	70 (3.1%)	2269 (98.6%)		46125 (96.4%)	1739 (3.6%)	47864 (91.5%)	0.01
Black	2181 (97.6%)	53 (2.4%)	2234 (5.6%)		9 (100%)	0 (0%)	9 (0.4%)		2233 (95.8%)	98 (4.2%)	2331 (4.5%)	
Hispanic	806 (99.0%)	8 (1.0%)	814 (2.0%)		7 (100%)	0 (0%)	7 (0.3%)		650 (98.3%)	11 (1.7%)	661 (1.3%)	
Asian	1185 (98.9%)	13 (1.1%)	1198 (3.0%)		15 (100%)	0 (0%)	15 (0.6%)		1063 (97.1%)	32 (2.9%)	1095 (2.1%)	
Native American	241 (99.6%)	1 (0.4%)	242 (0.6%)		2 (100%)	0 (0%)	2 (0.1%)		180 (95.2%)	9 (4.8%)	189 (0.4%)	
Pacific Islander	104 (96.3%)	4 (3.7%)	108 (0.3%)		0 (0%)	0 (0%)	0 (0%)		188 (97.5%)	5 (2.6%)	193 0.4%)	
Body mass index (kg/m2)	27.4 (4.8)	26.5 (4.5)	27.4 (4.8)	<0.0001	29.2 (5.9)	27.2 (5.2)	29.1 (5.9)	**0.003**	28.0 (5.1)	26.9 (4.7)	28.0 (5.1)	<0.0001
Personal History Cancer												
No	37385 (98.3%)	653 (1.7%)	38038 (95.3%)	<0.001	1604 (97.2%)	47(2.8%)	1651 (71.7%)	0.466	49105 (96.5%)	1790 (3.5%)	50895 (95.7%)	<0.001
Yes	1818 (97.0%)	58 (3.0%)	1874 (4.7%)		628 (96.5%)	23 (3.5%)	651 (28.3%)		2181 (94.5%)	126 (5.5%)	2307 (4.3%)	
Family history lung cancer												
No	33365 (98.4%)	541 (1.6%)	33906 (88.2%)	<0.001	1713 (97.1%)	51 (2.9%)	1764 (76.6%)	0.539	39494 (96.6%)	1386 (3.4%)	40880 (77.9%)	<0.001
Yes	4393 (97.1%)	130 (2.9%)	4523 (11.8%)		519 (96.5%)	19 (3.5%)	538 (23.4%)		11111 (95.7%)	497 (4.3%)	11608 (22.1%)	
COPD												
No	35388 (98.5%)	556 (1.5%)	35944 (90.9%)	<0.001	606 (99.0%)	6 (1.0%)	612 (26.6%)	**0.00088**	48770 (96.6%)	1724 (3.4%)	50494 (94.9%)	<0.001
Yes	3451 (96.1%)	142 (3.9%)	3593 (9.1%)		1626 (96.2%)	64 (3.8%)	1690 (73.4%)		2497 (92.9%)	192 (7.1%)	2689 (5.1%)	
Smoking Status												
Former	31600 (98.8%)	385 (1.2%)	31985 (80.1%)	<0.001	1063 (96.9%)	34 (3.1%)	1097 (47.7%)	0.973	26826 (97.2%)	764 (2.8%)	27590 (51.9%)	<0.001
Current	7600 (95.9%)	324 (4.1%)	7924 (19.9%)		1169 (97.0%)	36 (3.0%)	1205 (52.3%)		24460 (95.5%)	1152 (4.5%)	25612 (48.1%)	
Smoking Intensity (cig/d	24.8 (14.7)	30.2 (15.2)	24.9 (14.7)	<0.0001	26.5 (11.3)	28.2 (12.1)	26.6 (11.4)	0.308	28.4 (11.5)	29.6 (11.7)	28.4 (11.5)	<0.0001
Smoking duration (years)	27.5 (13.8)	39.9 (10.5)	27.7 (13.8)	<0.0001	37.0 (9.5)	37.8 (9.1)	37.0 (9.5)	0.49	39.7 (7.3)	44.2 (7.0)	39.8 (7.3)	<0.0001
Smoking quit time - former smokers (years)	20.3 (12.1)	12.4 (10.5)	20.2 (12.1)	<0.0001	10.6 (9.0)	8.9 (8.0)	10.6 (9.0)	0.209	7.3 (4.8)	6.6 (4.8)	7.3 (4.8)	<0.0001
Predictive Performance												
Lung cancer incidence per 10,000 per year in former, current smokers			20.7, 72.4				169.4, 158.1				48.0, 79.9	
Mean lung cancer probability	0.0169	0.0453	0.0174	<0.0001	0.03552	0.04556	0.03582	**0.02648**	0.0316	0.0519	0.0323	<0.0001
AUC (95% CI)	0.803 (0.788 - 0.817)				0.63 (0.57 - 0.69)				0.701 (0.689 - 0.712)			
Follow-up duration (years)	6.00				1.86				6.00			

Table data reformatted from: Tammemagi MC, Katkiha HA, Hocking WG, et al. Selection criteria for lung-cancer screening, Supplement. *N Engl J Med.* 2013;368(8):728-236 DOI: 10.1056/NEJMoa1211776

Table 7 – Model Comparisons; PLCOm2011, PLCOm2012, PLCOm2012noEd

Predictor variables	PLCOm2011	p	PLCOm2012		p	PLCOm2012noEd		p
Age, per year			1.08 (1.057 to 1.105)	centered age 62	<0.001	1.080 (1.055 to 1.104)		<0.001
Age spline 1	1.245 (1.130 TO 1.3720	<0.001						
Age spline 2	0.705 (0.505 to 0.9840	0.04						
Age spline 3	2.205 (0.860 to 5.651)	0.1						
Education, per 1 of 7 levels change	0.928 (0.887 to 0.9710	0.001	0.922 (0.874 to 0.972)	centered on level 4 of 6 levels	0.003	n/a		
Race or ethnic group	n/a							
White			1.000	referent group		1.000	referent group	
Black			1.484 (1.083 to 2.033)		0.01	1.563 (1.144 to 2.134)		0.005
Hispanic			0.475 (0.195 to 1.160)		0.1	0.501 (0.205 to 1.221)		0.128
Asian			0.627 (0.332 to 1.185)		0.15	0.627 (0.332 to 1.184)		0.15
Native Hawaiian or Pacific Islander			2.793 (0.992 to 7.862)		0.05	2.763 (1.018 to 7.769)		0.054
BMI per 1 unit change	0.972 (0.956 TO 0.988)	0.001	0.973(0.955 to 0.991)		0.003	0.975 (0.957 to 0.993)		0.006
Family history of lung cancer, yes vs no	1.561 (1.313 TO 1.856)	0.001	1.799 (1.471 to 2.200)		<0.001	1.815 (1.484 to 2.219)		<0.001
Personal histoy of cancer	n/a		1.582 (1.172 to 2.128)		0.003	1.578 (1.168 to 2.131)		0.003
COPD, yes vs no	1.374 (1.145 TO 1.6480	0.001	1.427 (1.162 to 1.751)		0.001	1.451 (1.182 to 1.781)		<0.001
Chest x-ray in past 3 years per 1 of 3 levels	1.117 (1.019 to 1.225)	0.019	n/a			n/a		
Pack-years smoked per 1 pack year			n/a			n/a		
Pack- year spline 1	1.059 (1.044 to 1.074)	<0.001						
Pack-year spline 2	0.949 (0.935 to 0.964)	<0.001						
Smoking duration per 1 year increase	1.012 (0.995 to 1.029)	0.171	1.032 (1.014 to 1.051)	centered on 27 years	0.001	1.035 (1.016 to 1.054)		<0.001
Smoking quit time in smokers per 1 year quit			0.970 (0.950 to 0.990)	centered on 10 years	0.003	0.972 (0.951 to 0.992)		0.007
Quit time spline 1	0.945 (0.918 to 0.974)	<0.001						
Quit time spline 2	1.047 (1.011 to 1.085)	0.01						
Smoking status								
Former	referent			referent	0.02		referant	0.011
Current	1.356 (1.077 to 1.708)	0.01	1.297 (1.047 to 1.605)			1.319 (1.066 to 1.632)		
Smoking intensity			((avg# cigperday/10)^-1)-0.402154613	transformed variable		((avg# cigperday/10)^-1)-0.402154613	transformed variable	
Model constant			-4.533			-9.508		

Tammemagi CM, Pinsky PF, Caporaso PA, et al. Lung Cancer Risk Prediction: Prostate, Lung, Colorectql and Ovarian Cancer Screening Trial Models and Validation. J Natl Cancer Inst. 2011;103(13):1058-1068. Doi: 10.1093/jnci/djr173.
Tammemagi MC, Katkiha HA, Hocking WG, et al. Selection criteria for lung-cancer screening. N Engl J Med. 2013;368(8):728-236 DOI: 10.1056/NEJMoa1211776

Figures 4a and 4b show a comparison of the baseline model, PLCOm2012, (Figure 4a) to the PLCOm2012noEd model (Figure 4b) for a 69 year old former smoker, 11 years quit, with COPD, a family history of lung cancer and a heavy smoking history. The calculated risk of lung cancer is similar for the 2 models; 6.9% for the PLCOm2012 model and 6.15% for the PLCOm2012noEd model. The equivalent "entry" risk level to meet the USPSTF criteria for lung cancer screening is 1.3455%,[13] so this individual is at high risk for lung cancer. In fact, this individual is my late husband who was diagnosed with Stage 4 lung cancer in October 2011 and died in April 2013. He is a good example of how the model individual predicted six year risk of developing lung cancer would have provided additional information, beyond just meeting the USPSTF/CMS risk screening criteria, to help with the screening decision. His risk was well above the screening risk threshold indicative of a high risk of lung cancer and a higher likelihood of benefiting from lung cancer screening than someone who met the guidelines but with a model predicted six year risk of lung cancer closer to the screening risk threshold.

Figure 4a – PLCOm2012 Six Year Probability of Lung Cancer

	A	B	C	D	E	F
1	**Calculator predicting 6-year risk of lung cancer by the Tammemagi (PLCOm2012) model.***				*Ver2-24AUG13MT*	
2	**Instructions: For the characteristics described in Column A fill in the individuals values in Column B.**					
3	**Note: Values in columns C to F can be ignored and should not be changed.**					
4	**A**	**B**	**C**	**D**	**E**	**F**
5	**Characteristics to be entered**	**Enter Values**	**Centered or referent group**	**Coefficient**	**Contribution to estimate**	**ORs**
6	Age in years	69	62	0.0778868	0.5452076	1.08
7	Education (enter the number identifying the highest level obtained) 1 = Less than high school grad; 2 = High school grad; 3 = Post high school training; 4 = Some college; 5 = College grad; 6 = Postgraduate/professional.	5	4	-0.0812744	-0.0812744	0.92
8	Body Mass Index (BMI, weight in kilograms/height in meters^2)	32	27	-0.0274194	-0.137097	0.97
9	COPD, emphysema or chronic bronchitis (0=No; 1=Yes)	1		0.3553063	0.3553063	1.43
10	Personal history of cancer (0=No; 1=Yes)	0		0.4589971	0	1.58
11	Family history of lung cancer (0=No; 1=Yes)	1		0.587185	0.587185	1.80
12	Race/ethnicity (select only one from this category)					
13	White (referent group) (0=No; 1=Yes)	1		0	0	
14	Black (non-Hispanic) (0=No; 1=Yes)	0		0.3944778	0	1.48
15	Hispanic (0=No; 1=Yes)	0		-0.7434744	0	0.48
16	Asian (0=No; 1=Yes)	0		-0.466585	0	0.63
17	American Indian/Alaskan Native (0=No; 1=Yes)	0		0	0	
18	Native Hawaiian/Pacific Islander (0=No; 1=Yes)	0		1.027152	0	2.79
19	Smoking status, 0 = Former-smoker 1 = Current-smoker	0		0.2597431	0	1.30
20	Average number of cigarettes smoked per day**	40	-0.15215416	-1.822606	0.277317087	nonlinear
21	Duration smoked (years)	40	27	0.0317321	0.4125173	1.03
22	Years ago quit smoking. Enter zero for current smokers.	11	10	-0.0308572	-0.0308572	0.97
23	Model constant			-4.532506	-4.532506	
24				xb =	-2.604201313	
25				EXP(xb) =	0.0740	
26	**Probability of lung cancer in 6 years =**	**0.069**				
27						
28	* Reference: Tammemagi et al. *Selection Criteria for Lung-Cancer Screening* . **NEJM.** 2013;368(8):728-36.					
29	** Transformation of smoking intensity =(((Average number_cigarettes_smoked_per_day/10)^-1)-0.4021541613)					
30	Example: The 6-year risk of lung cancer in a white individual who is 55 years old, has some college education, a BMI of 28, who is a former smoker who quit 15 years ago and smoked on average 20 cigarettes per day for 30 years is estimated to be					
31	NOTE: This calculator was developed and tested in Microsoft® Excel® for Mac 2011 version 14.0.0.					

Figure 4 b – PLCOm2012noED Six Year Probability of Lung Cancer

Tammemagi reparameterizedPLCm2012 model without education predictor					
Characteristic	Value	Referet group/comment	Coefficient	Contribution to Estimate	Odds Ratios
Age in Years	69		0.076624	5.287056	1.08
Education - not in reparameterized model	NA				
BMI (kg/m2)	0		-0.025363	-0.811616	0.974
COPD (0=no; 1=yes)	1		0.3722556	0.3722556	1.45
Personal History Cancer (0=no, 1=yes)	0		0.4557664	0	1.58
Family history lung cancer (0=no, 1=yes)	1		0.5959728	0.5959728	1.82
Race/ethnicity (0=no, 1=yes)					
White (referrent group)	1			32.	
Black (non-hispanic)	0		0.4464319	0	1.56
Hispanic	0		-0.6915011	0	0.5
Asian	0		-0.466574	0	0.627
American Indian/Alaskan Native	0		0	0	
Native Hawiaan/Pacific Islander	0		1.016191	0	2.76
Smoking Status (0=former smoker, 1=current smoker	0		0.2766229	0	1.32
Average number cigarettes smoked per day - transformed ((#cig/10 ^-1)-0.4021541613)	40	-0.152154161	-1.806032	0.274795	
Duration smoked (years)	40		0.0345697	1.382788	1.035
Years ago quit smoking, enter zero for current smokers	11		-0.028896	-0.317856	0.972
Model constant			-9.508016	-9.508016	
				-2.7246	
				Exp (-2.7246) = 0.06557	
		Probability of lung cancer in 6 years		risk = 0.06557/(1+0.06557) = **0.06154**	

6.4 Model Predictive Performance in the Screening Program Sample

The PLCOm2012noEd model six year mean risk for lung cancer was higher for participants with lung cancer, 4.56%, as compared to those without lung cancer, 3.55% (p=0.0265) (Figures 5 and 6). Area under the curve (AUC) of the receiver operator characteristics curve (ROC) was 0.63 (95% CI 0.57 – 0.69) (Figure 7). The mean absolute difference between observed and predicted risk was 0.013 or less for the first 9 deciles (Figure 8).

The Hosmer-Lemeshow goodness of fit test p value was less than 0.05 (p= 0.0053) indicating poor calibration (Figure 9). The large difference between the observed risk of lung cancer compared to the PLCOm2012noEd model predicted six year risk of lung cancer above the 9th decile drove the poor calibration result. The model predicted risk of lung cancer at the 10th decile was 12.1% as compared to an observed risk of 3%; the predicted number of lung cancers in the 10th decile was about 28 with only 7 lung cancers observed in this decile (Figures 10 and 11). Accurately capturing the extremes of a distribution is one of the limitations of the Hosmer-Lemeshow test and was evident in this analysis.[40]

Figure 5 - Model Predicted Lung Cancer Risk in Screening Program Sample vs Lung Cancers Diagnosed (p=0.0265)

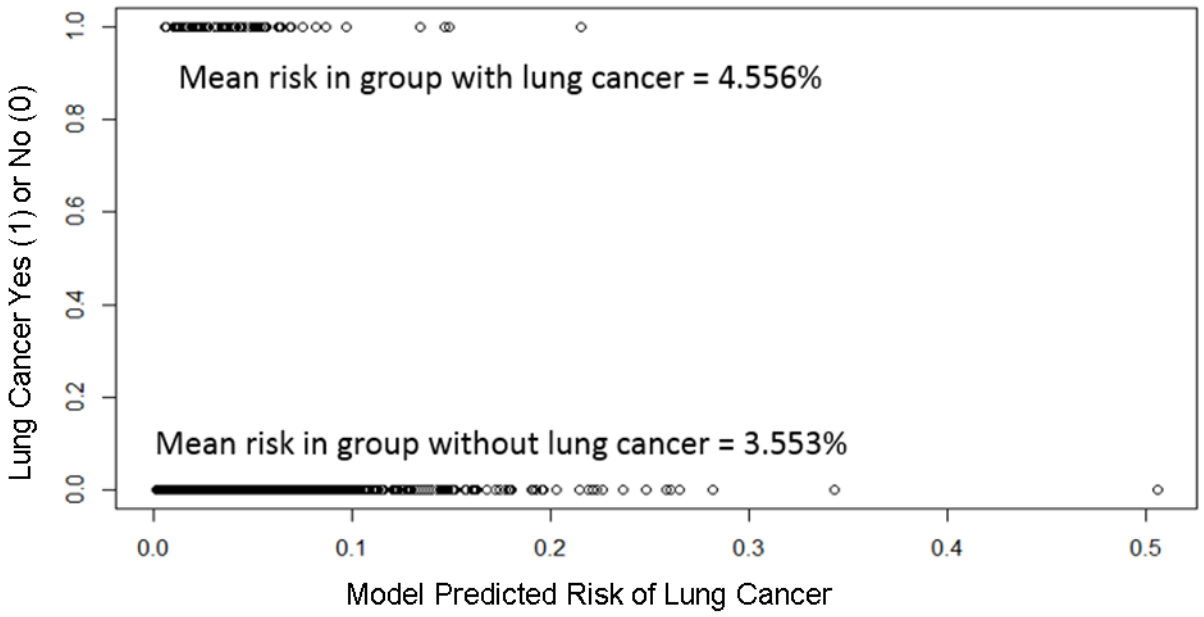


Figure 6 – Box Plot Model Predicted Risk for Group with Lung Cancer and Group without Lung Cancer (p=0.0265)

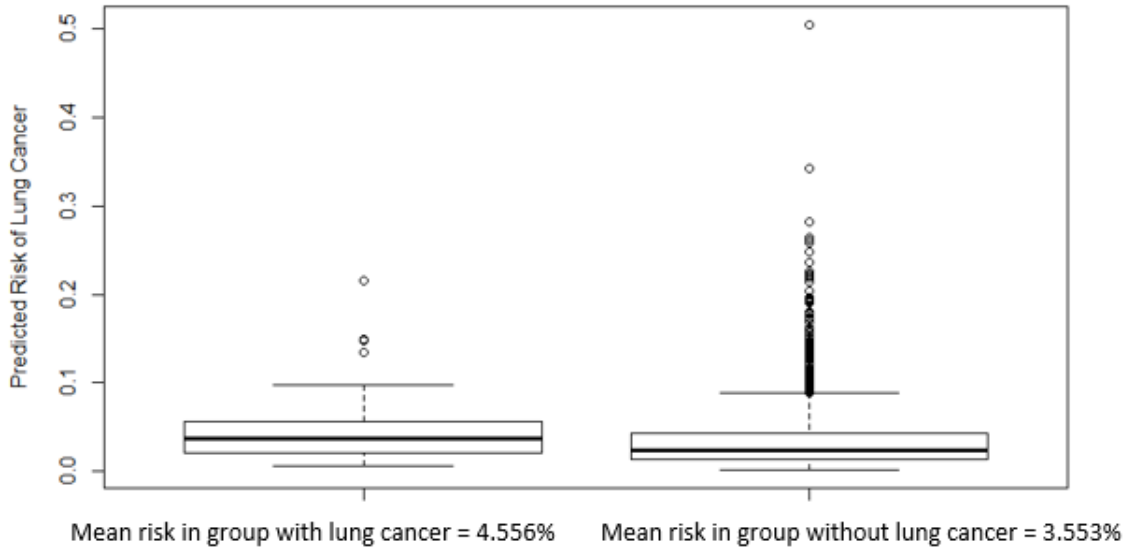


Figure 7 – Area Under the Receiver Operating Curve (AUC) is 0.63 (95% CI 0.57 – 0.69)

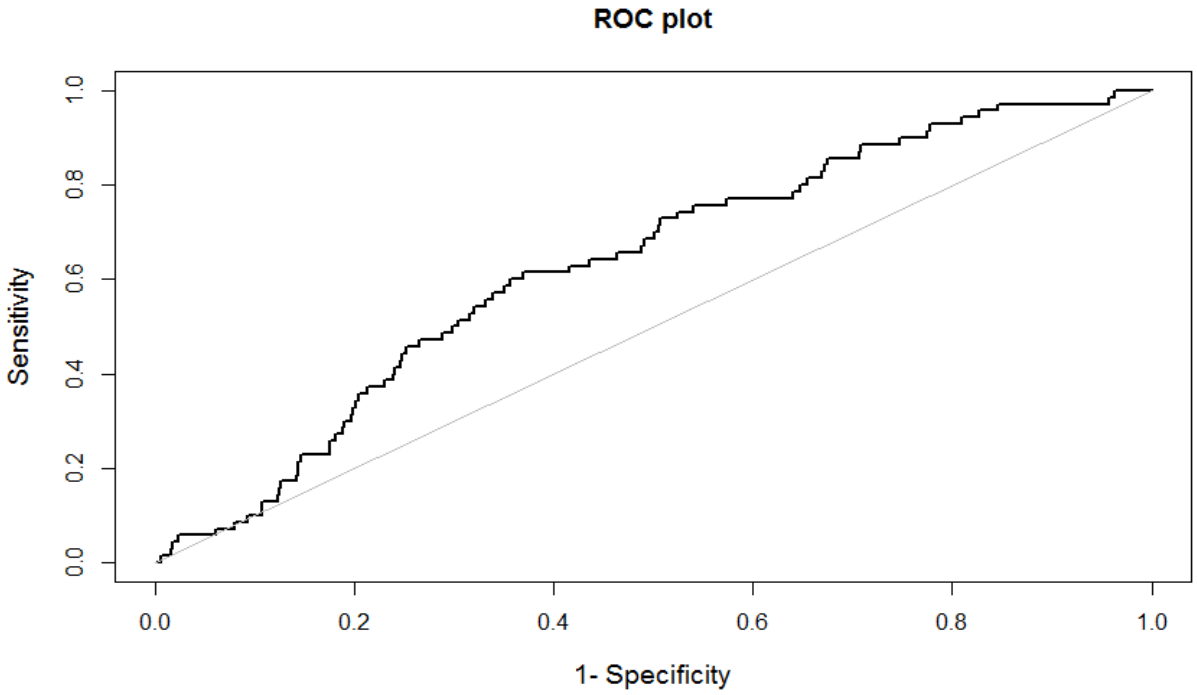
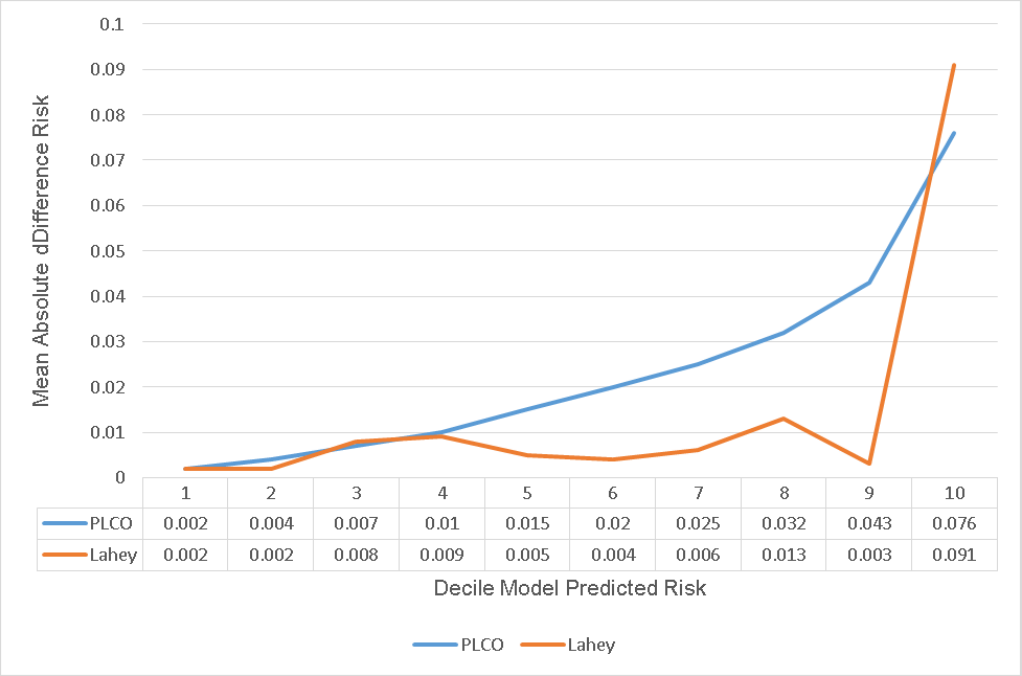


Figure 8 – Mean Absolute Difference Between Observed and Predicted Risk; 0.013 or Less in the First 9 Deciles of Screening Program Sample



	1	2	3	4	5	6	7	8	9	10
PLCO	0.002	0.004	0.007	0.01	0.015	0.02	0.025	0.032	0.043	0.076
Lahey	0.002	0.002	0.008	0.009	0.005	0.004	0.006	0.013	0.003	0.091

PLCO data reformatted from: Tammemagi MC, Katkiha HA, Hocking WG, et al. Selection criteria for lung-cancer screening, Supplement. *N Engl J Med.* 2013;368(8):728-236 DOI: 10.1056/NEJMoa1211776

Figure 9 – Hosmer-Lemeshow Test; Observed vs Predicted Risk – 10 Sub-Groups (p=0.0053)

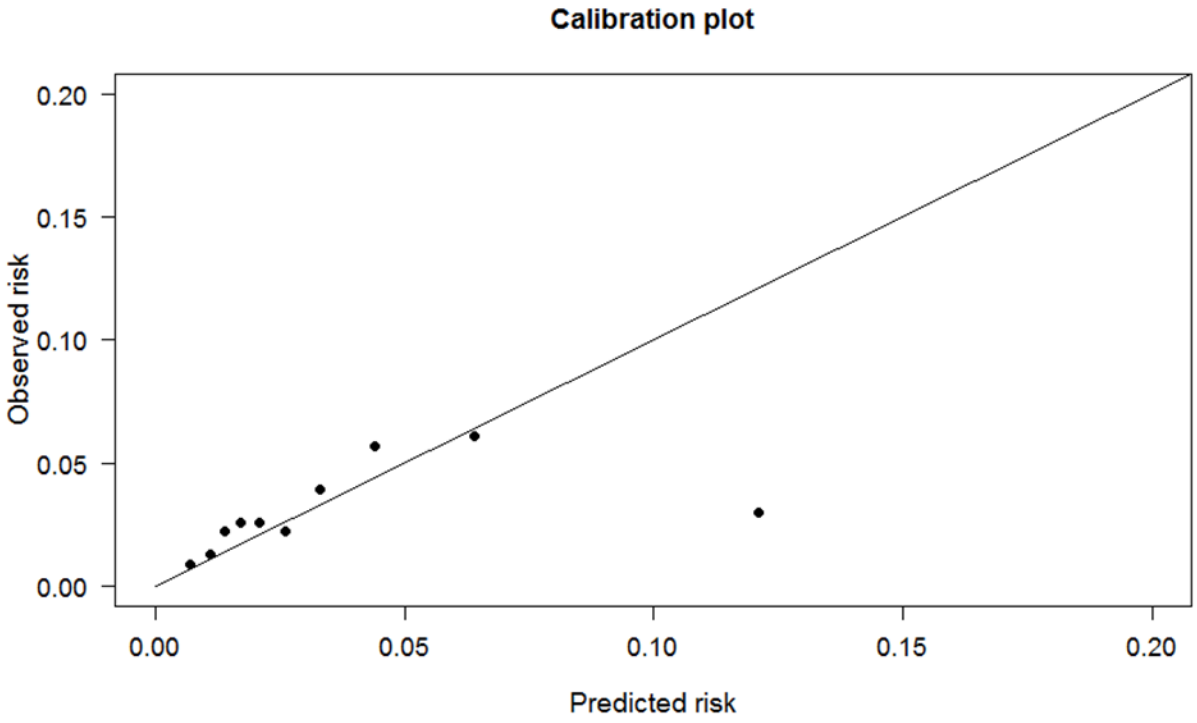


Figure 10 - Predicted and Observed Lung Cancer Risk by Percentile

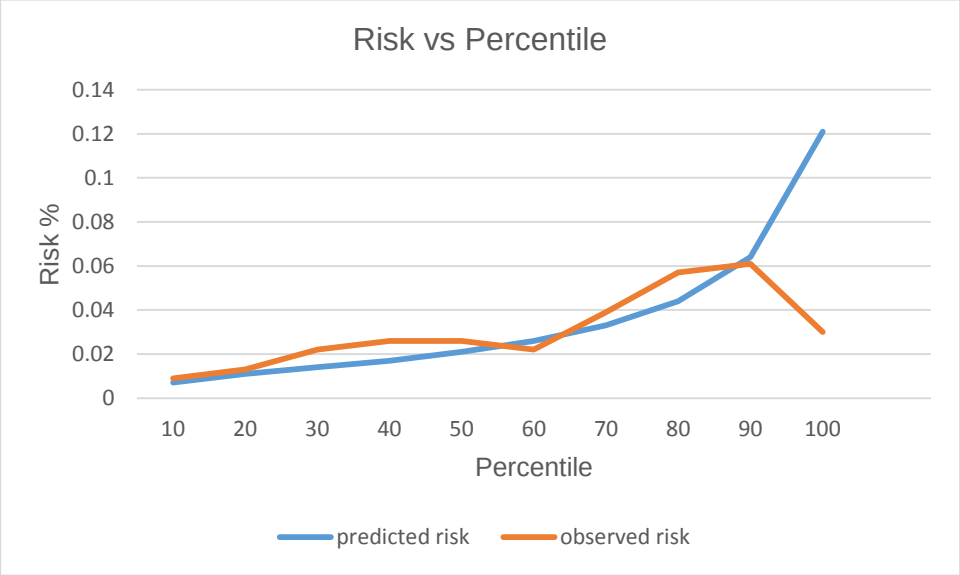


Figure 11- Number of Lung Cancers Predicted vs Observed by Percentile

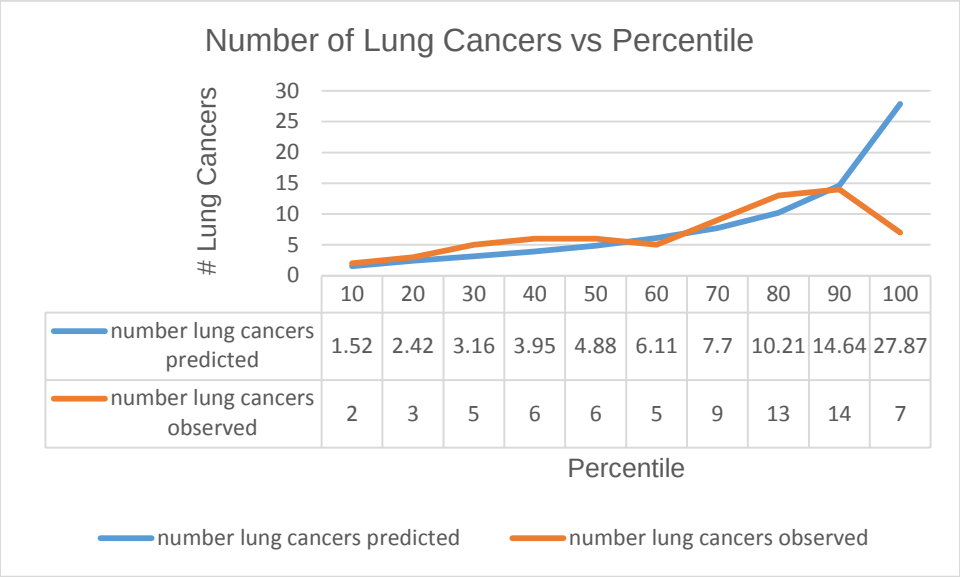


Table 8 shows PLCOm2012noEd model sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), number of true positives, false positives, true negatives and false negatives for a range of cut-off model predicted screening risk thresholds. At the lowest levels of model predicted risk, the sensitivity was high because all lung cancers were identified. However, the model specificity was the lowest as all of the non-lung cancers were false positives. At the higher levels of model predicted risk, sensitivity was lower, not all lung cancers were found, but the specificity was higher and there were fewer false positives. Tammemagi et al. recommended a model predicted risk threshold to screen of 1.51% (based on NLST mortality outcomes).[30] At this recommended risk threshold, the PLCOm2012noEd model sensitivity in the screening program sample was 85.7% with a specificity of 29.7%; the model PPV was 3.7% and the NPV was 98.5%. The sensitivity of 85.7% indicated that almost 86% of people who developed lung cancer met or exceeded the model screening predicted risk threshold or the corollary, about 15% of people who developed lung cancer were below the model predicted screening risk threshold. The low specificity of 29.7%, a result of the large number of false positives (1569), indicated the selected model predicted screening risk threshold was poor at

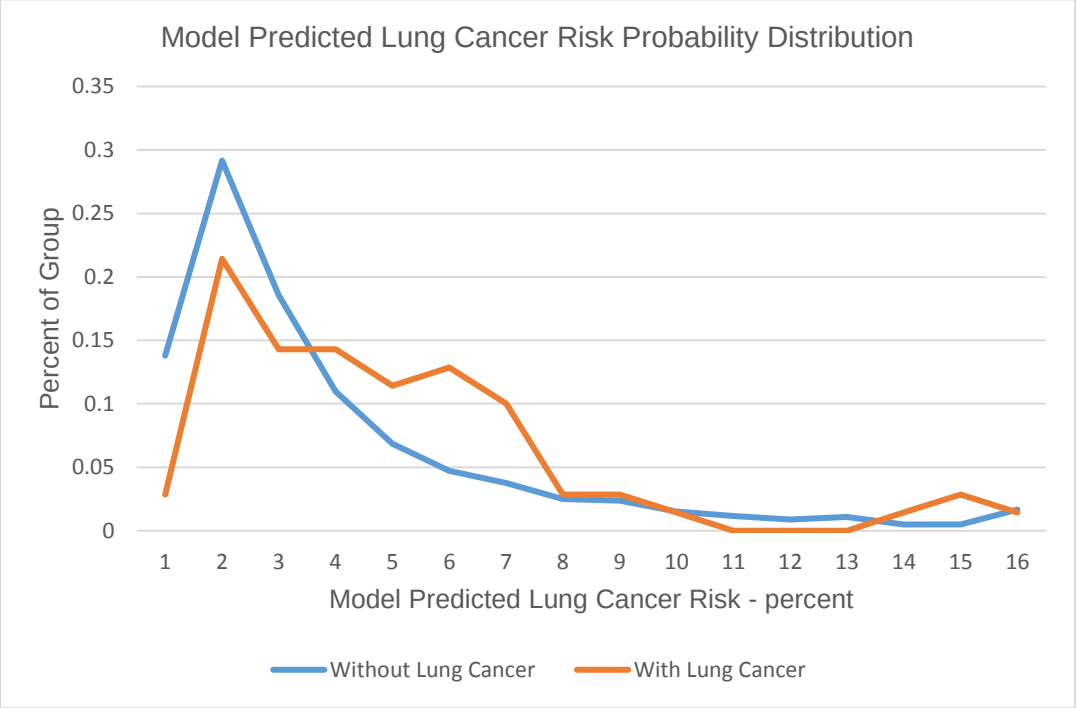
identifying participants that did not develop lung cancer. The low PPV indicated only 3.7% of the participants that met the 1.51% risk screening threshold developed lung cancer despite having a model predicted risk that exceeded the screening threshold. This is similar to the PPV of 4.2% for the PLCOm2012 model in the validation sample. Although the PPV and specificity are low, in the case of lung cancer screening, having a small percentage of false negatives (people who develop lung cancer who are not identified as high risk) is important to reduce lung cancer mortality.

Table 8 – Sensitivity and Specificity at Screening Risk Threshold (1.51%) 85.7% and 29.7%, Respectively

Cut off risk value	Percentile of risk	Patients at risk	True positives	False positives	False negatives	True negatives	Sensitivity	Specificity	PPV	NPV
0.00%	0.00	2302	70	2232	0	0	1.000	0.000	0.030	-
0.89%	10.00	2071	68	2003	2	229	0.971	0.103	0.033	0.991
1.00%	13.51	1991	68	1923	2	309	0.971	0.138	0.034	0.994
1.21%	20.00	1841	65	1776	5	456	0.929	0.204	0.035	0.989
1.37%	25.00	1726	62	1664	8	568	0.886	0.254	0.036	0.986
1.51%	29.20	1629	60	1569	10	663	0.857	0.297	0.037	0.985
1.89%	40.00	1381	54	1327	16	905	0.771	0.405	0.039	0.983
2.38%	50.00	1151	48	1103	22	1129	0.686	0.506	0.042	0.981
3.53%	60.00	920	43	877	27	1355	0.614	0.607	0.047	0.980
4.41%	75.00	575	29	546	41	1686	0.414	0.755	0.050	0.976
7.88%	90.00	231	7	224	63	2008	0.100	0.900	0.030	0.970
50.53%	100.00	0	0	0	70	2232	0.000	1.000	-	0.970

Comparison of the risk frequency distribution (percent of group) for participants with lung cancer as compared to participants without lung cancer (Figure 12), shows the highest frequency for both groups was at approximately 2%. This is above the 1.51% screening risk threshold. The frequency decreased quickly for participants without lung cancer, to only 0.05% of the group by 6% predicted risk. The frequency for participants with lung cancer only decreased to 1.3% at this same predicted risk. Between 4% and 8% predicted risk, participants with lung cancer were a higher percent of their group as compared to participants without lung cancer.

Figure 12 - Lung Cancer Risk Frequency Probability Distribution as Percent of Group



Note: Lung cancer risk percent of "16" includes all model predicted risks above 15%

6.5 Model Predictive Performance NCCN Group 1 vs Group 2

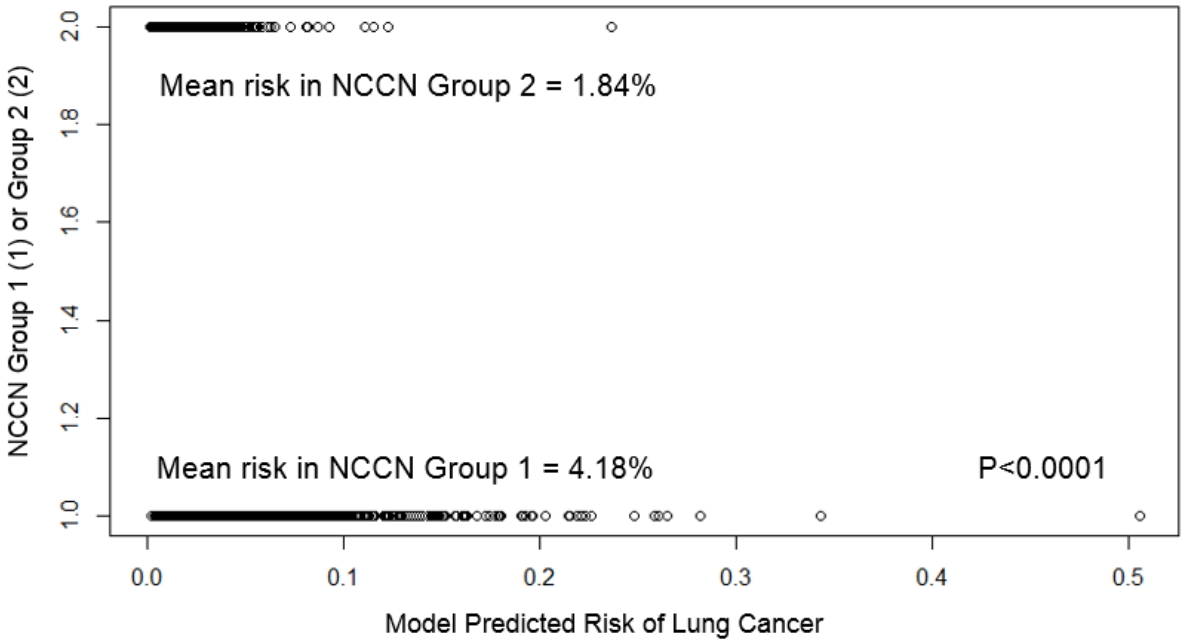
Lung cancer mean risk for the subgroup of NCCN Group 2 was significantly lower ($p<0.0001$) than for NCCN Group 1; 1.84% vs. 4.18% (Figure 13). However, the percentages of lung cancer were the same in both groups. Out of 70 cancers, 53 occurred in the 1715 Group 1 participants (3.1%) as compared to 17 in the 587 Group 2 participants (2.9%), $p=0.9224$.

Of the 587 NCCN Group 2 participants, 262 had predicted risk above the 1.51% recommended screening threshold (44.6% of the Group 2 sample). Eleven of the 17 lung cancers were in those with risk levels above 1.51%. The sensitivity and specificity of the model for this group at the screening threshold were 64.7% and 56.0%, respectively. PPV was 4.2%. The sensitivity for

this group was much lower than for the complete sample. Almost 40% of the people who developed lung cancer were below the model predicted screening risk threshold.

For NCCN Group 1, 1368 of 1715 participants were above the 1.51% screening risk threshold (79.8% of the Group 1 sample). Forty nine of the 53 lung cancers were in those with risk levels above 1.51%. Sensitivity, specificity, and PPV were 92.5%, 20.6% and 3.6%, respectively. This sub-group had the highest sensitivity at the model predicted screening risk threshold.

Figure 13 – Model Predicted Lung Cancer Risk NCCN Group 1 vs NCCN Group 2



6.6 Model Predictive Performance in NCCN Group 1

In the NCCN Group 1 subgroup (similar criteria as the NLST), the mean predicted risk for participants with lung cancer was 5.249% as compared to 4.144% for those without lung cancer; p=0.0424 (Figure 14).

Model AUC was improved to 0.641 (95% CI 0.576 – 0.706) (Figure 15).

The Hosmer-Lemeshow goodness of fit test indicated poor calibration (p=0.0179), and showed similar large observed to predicted lung cancer risk differences for the 10th decile as did the complete sample (Figure 16). The 10th decile predicted risk of lung cancer was 13.3% as compared to 2.9% observed risk (Figure 17). Similar to the complete sample, the number of predicted lung cancers for the 10th decile was much larger than the number observed, 22.7 vs. 5 lung cancers, respectively (Figure 18).

Figure 14 – Model Predicted NCCN Group 1 Lung Cancer Risk in Screening Program Sample vs Lung Cancers Diagnosed (p=0.0424)

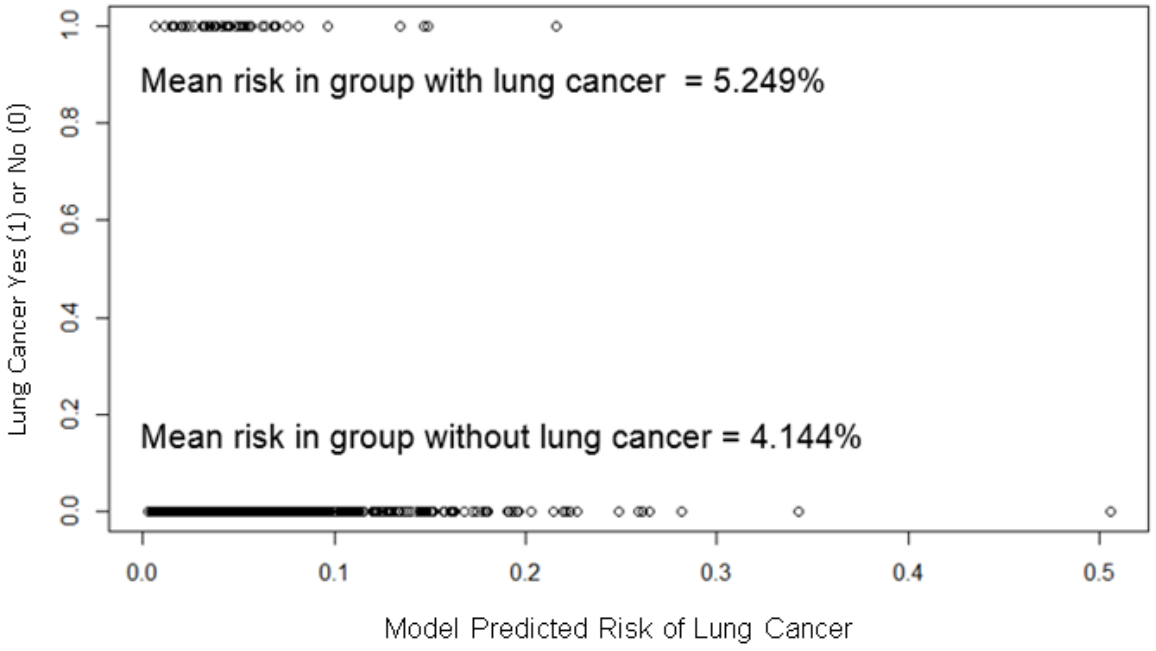


Figure 15 – NCCN Group 1 Area Under the Receiver Operating Curve (AUC) is 0.641 (95% CI 0.576 – 0.706)

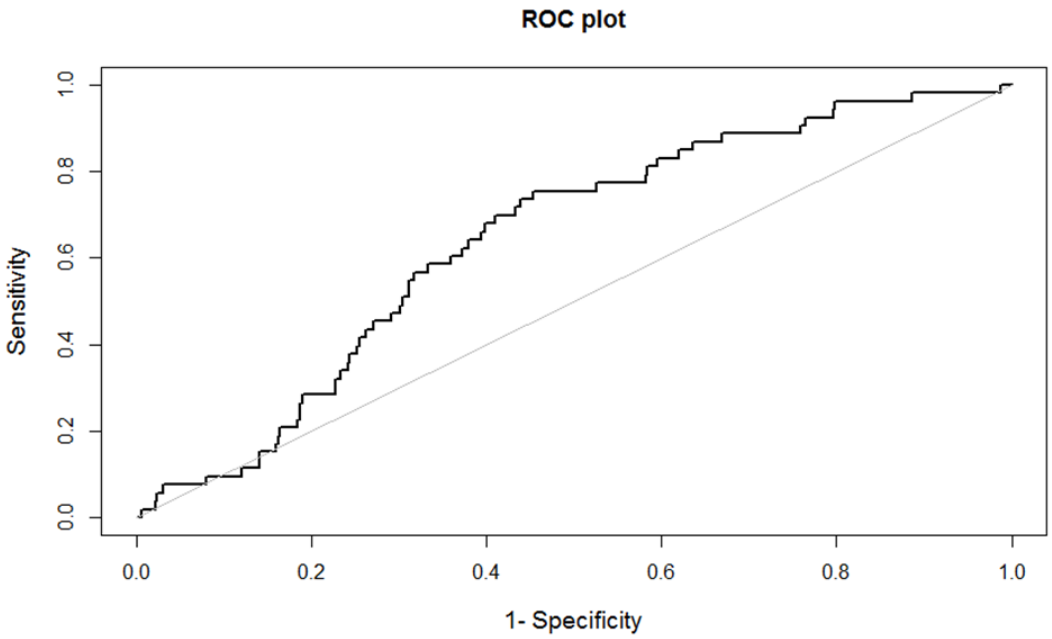


Figure 16 – NCCN Group 1 Hosmer-Lemeshow Test; Observed vs Predicted Risk – 10 Sub-Groups (p=0.0179)

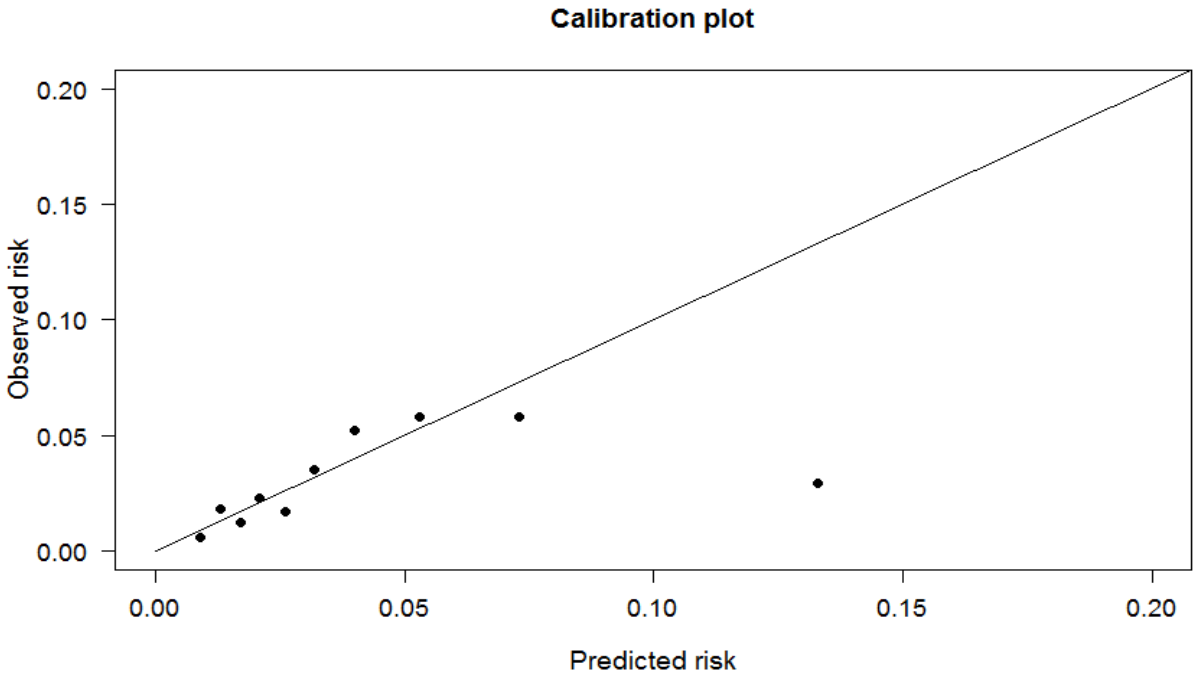


Figure 17 – NCCN Group 1 Predicted and Observed Lung Cancer Risk by Percentile

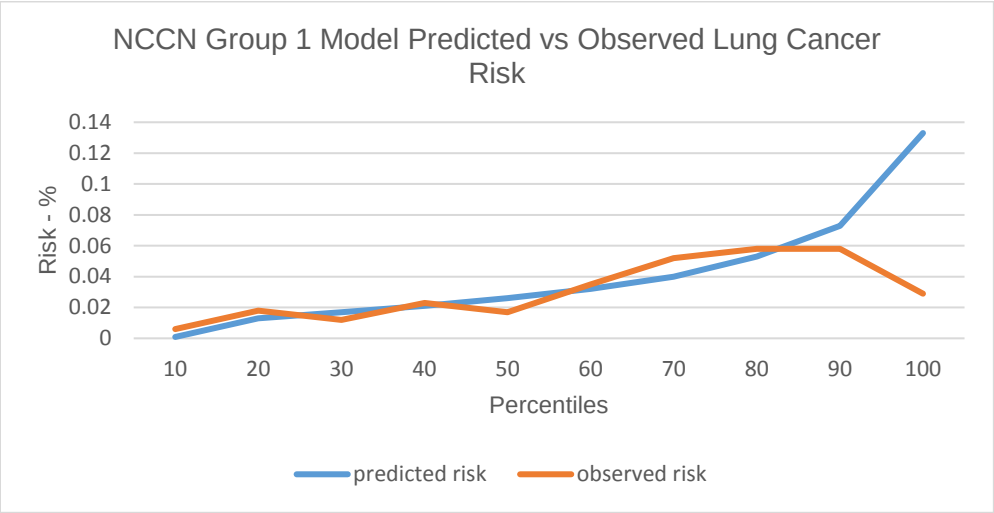
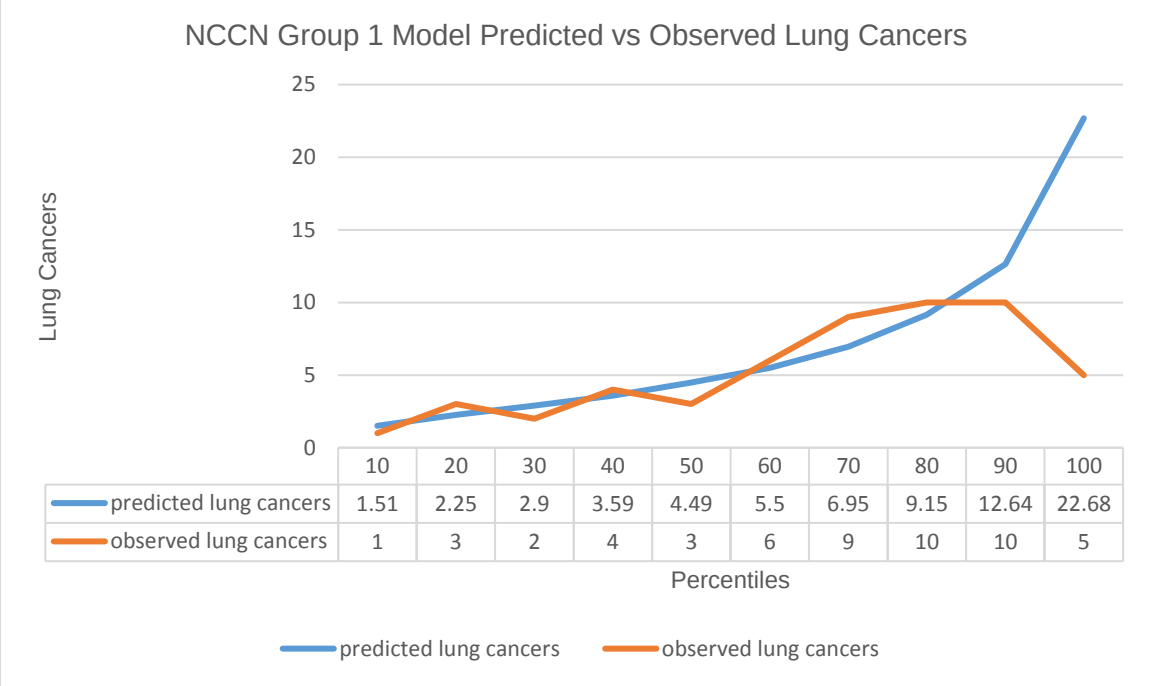


Figure 18 – NCCN Group 1 Number of Lung Cancers Predicted vs Observed by Percentile



6.7 Model Predictive Performance in NCCN Group 2

In the NCCN Group 2 subgroup (younger, lighter smoking history, no limit on time quit and one additional risk factor) the mean predicted risk for participants with lung cancer was 2.394% as compared to 1.826% for those without lung cancer, but the difference was not statistically significant; p=0.2507 (Figure 19).

Model AUC was 0.634 (95% CI 0.522 – 0.746) (Figure 20).

The Hosmer-Lemeshow goodness of fit test indicated poor calibration (p=0.0271) and exhibited more variation between observed and predicted risk across the range of deciles as compared to the complete sample (Figure 21).

Figure 19 – NCCN Group 2 Model Predicted Lung Cancer Risk vs Lung Cancers Diagnosed (p=0.0265)

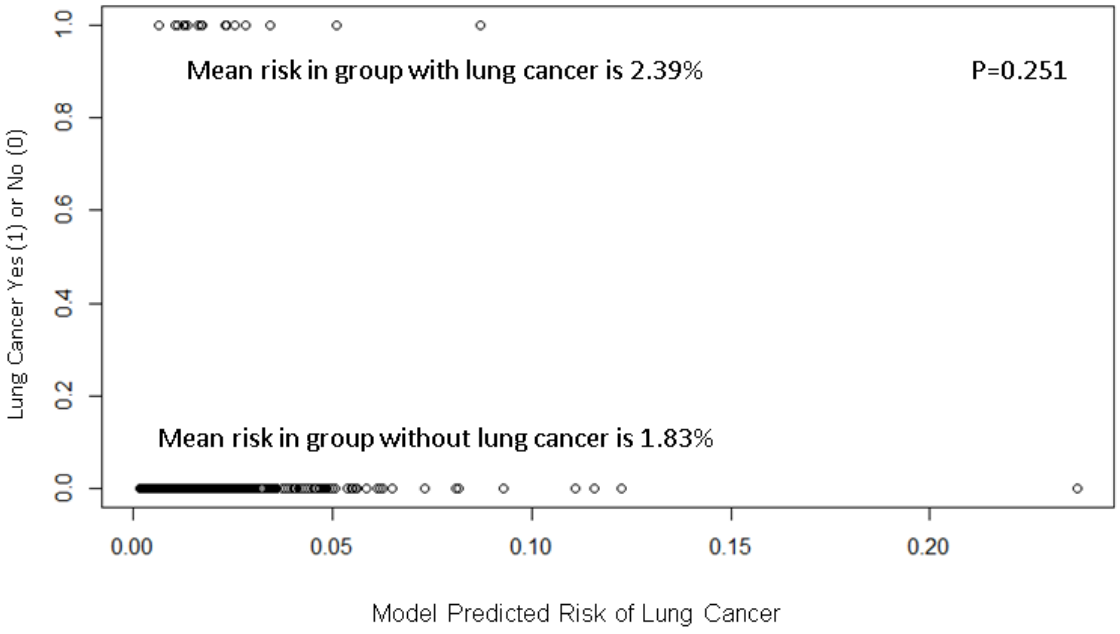


Figure 20 - NCCN Group 2 Area under the Receiver Operating Curve (AUC) is 0.634 (95% CI 0.522 – 0.746)

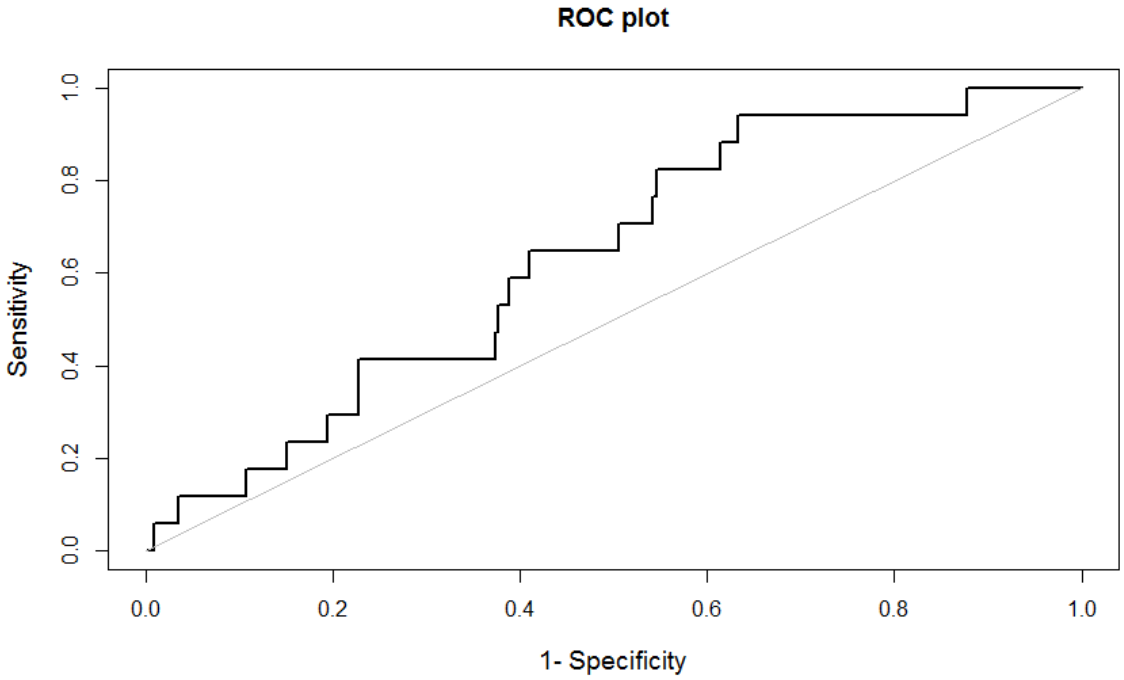
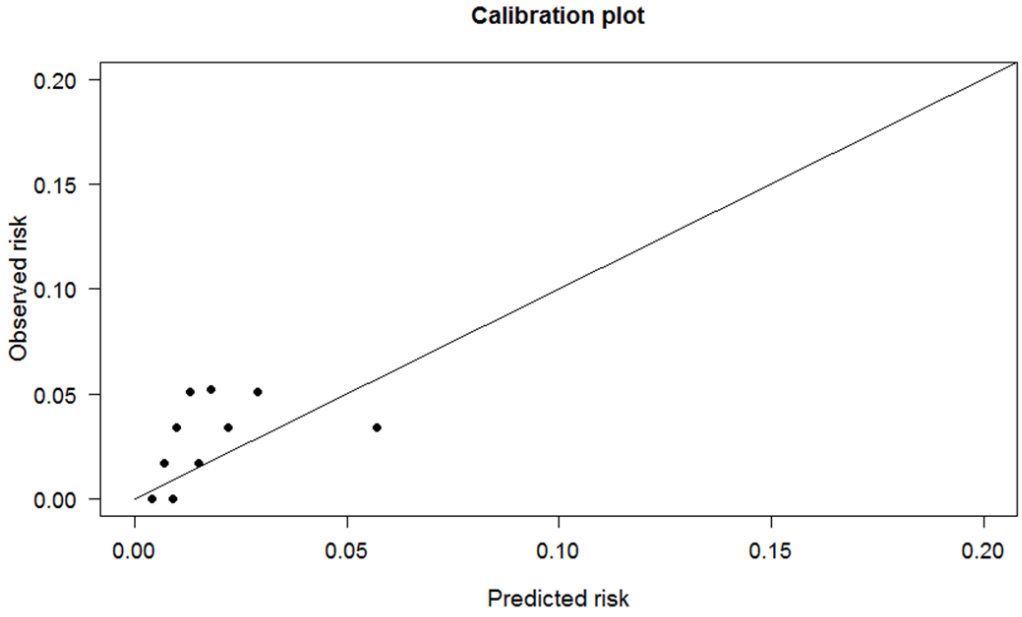


Figure 21 - NCCN Group 2 Hosmer-Lemeshow Test; Observed vs Predicted Risk – 10 Sub-Groups (p=0.0271)



6.8 Sensitivity Analysis

The PLCOm2012noEd model was run for five modified codings of predictor variables and/or modified samples to assess the impact of missing data, uncertain data and differences in predictor variable data collection between the model development and the screening program samples. Table 9 summarizes the results. The “Base case” in the table shows the original analysis with the complete sample.

Table 9 – Sensitivity Analysis Results

Sensitivity Studies							
Case Description	**Sample size**	**Mean risk (No Lung Cancer)**	**Mean risk (Yes lung cancer)**	**Difference in means**	**95% CI**	**P value**	**AUC**
Base case	2302	3.552	4.556	-1.004	-1.888 to -0.121	0.0265	0.63 (0.569-0.69)
Remove adopted, no and limited family history	2160	3.561	4.589	-1.028	-1.922 to -0.133	0.0205	0.631(0.57-0.692)
Only patients self-reported with cancer coded as 1 – personal history of cancer; i.e. patient history of cancer based on chart read only coded as 0 – 427 patients with cancer self-reported (18.5%)	2302	3.367	4.343	-0.976	-1.866 to -0.085	0.03216	0.621(0.56-0.683)
Only patients self-reported with cancer other than only skin cancer coded as 1 for personal history of cancer – 309 pts self-reported and with skin cancers only excluded (13.4%)	2302	3.268	4.119	-0.851	-1.594 to -0.107	0.02554	0.623(0.561-0.684)
With only patient reports coded as positive for COPD	2302	2.847	3.415	-0.568	-1.1788 to 0.0426	0.06779	0.623(0.564-0.681)
71 participant family history of lung cancer changed from code 0 to code 1 – these are participants with family history of cancer – type not specified in medical records	2302	3.641	4.788	-1.147	-2.081 to -0.212	0.01687	0.633 (0.572-0.694)

To investigate the impact of missing family history information, the 142 participants that were adopted or participants that were not asked the question “do you have a family history of lung cancer?” and had no or limited family history on file were deleted from the screening program baseline sample. The mean predicted risk of lung cancer for cases and non-cases remained essentially the same as the baseline, as did the difference in means. The difference in means remained statistically significant; p=0.0205. AUC was essentially unchanged at 0.631 (95% CI 0.57-0.692).

To assess the sensitivity of the analysis to the data sources used to code a person as having a personal history of cancer between the PLCOm2012noEd model development and

validation samples and the screening program (self-report compared to self-report and medical record note), the complete screening program sample was recoded with only participant self-reported cancer coded as a personal history of cancer. This reduced the percentage of participants in the screening program sample with a personal history of cancer from 28.3% to 18.5%. The mean risk for both participants with and without lung cancer decreased slightly (0.213 and 0.185, respectively). The difference in mean risk between the cases and non-cases was still significant at p=0.03216. AUC was reduced slightly to 0.621 (95% CI 0.56-0.683).

To further assess the sensitivity for the differences in the type of cancer that qualified for a personal history of cancer between the model development and validation sample and the screen program sample, the complete screening program sample with only patient self-reported cancer, excluding non-melanoma skin cancers, were coded as a personal history of cancer. This reduced the percentage of participants with a personal history of cancer further, from 18% to 13.4%. This was still about three times the percentage of personal history of cancer as compared to both the PLCO and NLST samples. Mean risk declined both for participants with and without lung cancer (0.437 and 0.284, respectively) and the difference in the means remained statistically significant at p=0.02554. The AUC was 0.623 (95% CI 0.561- 0.684).

Similarly, to assess the sensitivity of the analysis results to the data sources used to code a person as having COPD between the model development and validation and the screening program sample, only patient self-reported COPD were coded as patient history of COPD for the complete screening program sample. This reduced the percentage of participants with COPD from 73.4% to 19.5%. The 19.5% of participants with COPD with the recoding was still about double and triple the rates in the PLCO and NLST samples, respectively.

Recall this was a biased sample in the screening program as not all participants were asked the question about COPD during program intake. The mean risk in participants without lung cancer was the lowest of all of the samples run at 2.847% (-0.705). Similarly, the mean risk for participants with lung cancer was also reduced to the lowest of the samples; 3.415 (-1.141). Additionally, the difference between the mean risks was about half of what it was for the baseline (0.568 vs. 1.004) and was no longer statistically significant at a p</= 0.05 level (p=0.06779) but still suggestive of a difference between participants with lung cancer and those without lung cancer. AUC was reduced from 0.63 (95% CI 0.569-0.69) to 0.623 (95% CI 0.564-0.681).

To investigate the impact of a family history of an unspecified type of cancer that may have been lung cancer, the family history for those with an unspecified cancer, coded as “0, no family history of lung cancer, in the baseline analysis were recoded as “1”, yes family history of lung cancer. There were 71 participants recoded. There was a slightly increased mean risk for both participants with and without lung cancer, +0.890 and +0.232, respectively to 3.641% and 4.788% mean risk levels. The difference in the mean risk between cases and non-cases was statistically significant, p=0.01687. AUC increased slightly to 0.633 (95% CI 0.572-0.694)

.

Chapter 7: Discussion

7.1 Overall Model Predictive Performance

This study is the first to assess a lung cancer risk prediction model in a clinical lung cancer screening program. Model predictive performance included assessing both discrimination and calibration. The predictive capability of the reparameterized PLCOm2012 model, PLCOm2012noEd, in the screening program sample was not as good as in the model

development sample and not as good as the PLCOm2012 model in the PLCO smoker control and intervention group samples used for model development and validation, respectively. This is expected as models perform best in their development data sets and lose predictive capability in different samples.[26]

The PLCOm2012 model in the development and validations samples had AUC values considered to be excellent and good discrimination, respectively.[26,39] The PLCOm2012noEd model AUC value was considered to be in the excellent discrimination range in the development sample. In the screening program sample, the AUC was in the poor discrimination range. This is not surprising as the model was developed and validated independently in different samples with significant differences in sample characteristics from the screening program sample. A loss in discrimination in a new and different sample for a model is to be expected even in samples with similar characteristics as models perform best in the sample used to develop them; the model coefficients are fit to the random variation specific to that sample.[26] One of the major differences in the samples likely to have impacted discrimination was the difference in follow-up duration. The development and validation samples had six years of follow-up for every participant vs. an average of 1.86 years of follow-up for the screening program participants. It is likely this difference contributed to the lower model AUC scores as some of the non-cases (no lung cancer) in the screening program are likely to become cases (yes lung cancer) with longer follow-up.

The screening program sample included participants with as few as 1 day of follow-up. Deleting screening program participants with fewer than 90 days of follow-up after a scan (127 participants and 2 lung cancers) and re-running the model did not significantly change the results. The PLCOm2012noEd model six year mean risk for participants with lung cancer was 4.51% as compared to 4.56% for the complete sample. The mean risk for those without lung cancer was 3.57% as compared to 3.55% for the complete sample. Area under the curve (AUC)

of the receiver operator characteristics curve (ROC) was 0.622 (95% CI 0.56 - 0.684) as compared to 0.63 (95% CI 0.57 – 0.69) for the complete sample.

There was also a significant difference in the sample size and number of cases (lung cancers) between the model development and the screening program samples. The percentage of cases was actually higher in the screening program as compared to the model development sample, 3.0% vs. 1.8%, respectively (Table 6), however, the number of cases (lung cancers) in the screening program sample was still only approximately 10% of the number in the development and validation samples (70 vs. 709 and 753 cases, respectively). Having fewer cases makes it more difficult for the model to discriminate between a case and a non-case. Increased sample size and/or longer duration follow-up would be required to increase the number of cases in the sample.

Although the Hosmer-Lemeshow goodness-of-fit test indicated poor calibration with p=0.0053, the area of poor calibration was in the 10th and top decile of risk, well above the screening risk threshold. The absolute error between the mean predicted and observed probability of lung cancer in the 3rd and 4th deciles (encompass the 1.51% screening risk threshold) was 0.008 and 0.009, respectively and comparable to the PLCOm2012 model in the PLCO data sample at 0.007 and 0.010. (Figure 7). These indicate good calibration around the screening selection risk threshold.

Sensitivity was also indicative of good model performance. Specificity, and PPV were both low indicating a large number of false positives, i.e. people meeting the screening threshold but not developing lung cancer. In the case of lung cancer screening, high sensitivity is required to reduce lung cancer mortality.[4] CMS and USPSTF guidelines and clinical practice for lung cancer screening include management of false positives to minimize the potential harms of screening.[9,10,51] The sensitivity, specificity and PPV for the PLOCm2012noEd model in the screening program sample at the 1.51% screening risk threshold were 85.7%, 29.7% and 3.7%,

respectively. These compare favorably to the PLCOm2012 model in the PLCO validation sample for sensitivity (80.1%) and are worse for specificity (66.2%) and slightly worse for PPV (4.2%).[30]

The PLCOmnoEd model was fairly good at predicting lung cancers in this independent sample especially given; 1) the differences in the samples; sample size, personal cancer history and COPD prevalence, lack of race diversity; 2) the differences in the study designs; retrospective vs. prospective and the 1.86 years average follow-up vs. 6 years.

From a clinical perspective, use of the PLCOm2012noEd model (or prospectively PLCOm2012) will not replace use of the USPSTF and CMS criteria for lung cancer screening participant selection. The immediate value of the risk model will be in helping both the participant and the ordering physician assess the individualized risk and aid in the shared decision making discussion required by CMS for reimbursement. Knowing that the model has fair predictive capability, especially at the recommended screening threshold criteria will increase confidence in use of the model.

7.2 Model Performance for NCCN Group 1 and NCCN Group 2

The PLCOm2012noEd predictive performance for the sub-group of 1715 participants meeting the criteria for NCCN Group 1, essentially the NLST criteria, had slightly improved discrimination and similar calibration as for the complete sample.

However, the model did not discriminate well between participants with lung cancer and participants without lung cancer in NCCN Group 2.

Group 2 participants had significantly lower model predicted risk of lung cancer as compared to NCCN Group 1 ($p<0.0001$). Additionally, the model predicted mean risk between participants with lung cancer in NCCN Group 2 was not significantly different from those without lung cancer. However, the observed prevalence of lung cancer for the 2 groups was not

distinguishable between the two groups (p=0.9274). This may indicate the lack of a significant predictor variable for this group. PLCOm2012noEd includes all of the additional qualifying risk factors for NCCN Group 2 (COPD, family history of lung cancer, personal history of cancer) except for exposure to occupational carcinogens (radon, asbestos, arsenic, cadmium, and diesel fumes). This exposure predictor variable was not consistently available for the screening program participants, but at least 1.8% in the complete sample are known to have exposure. A subgroup of screening program participants (228) were asked about their occupational carcinogen exposure during program intake. Thirty six of the 228 of participants asked the question (15.8%) indicated an occupational exposure. Thus it is likely the percentage in the complete sample, if all had been asked the question about exposure, would have been larger than 1.8%. The program intake protocol now requires asking all participants about exposure at every annual screening LDCT.

Another risk factor for lung cancer that was not one of the PLCOm2012noEd predictor variables is having more than one relative diagnosed with lung cancer. This may have impacted the NCCN Group 2 predicted risk relative to observed risk. A systematic review of the relationship between family history and lung cancer found an increased risk of lung cancer when multiple family members had lung cancer as compared to only one; RR 1.57% (95% CI 1.34 – 1.84) for one family member as compared to RR 2.52% (95% CI 1.72-3.70) for having two or more family members impacted by lung cancer.[52] Although PLCOm2012noEd included a family history of lung cancer as one of the prediction variables, there was no distinction between having one versus multiple family members diagnosed with lung cancer. In NCCN Group 2, 5.1% of participants (30 of 587) had more than one family member diagnosed with lung cancer while only 2.1% of NCCN Group 1 participants (36 of 1715) had more than one family member diagnosed with lung cancer (p=0.0003).

Future studies evaluating model performance for models including these risk factors may help improve risk prediction for people meeting NCCN Group 2 criteria.

7.3 Sensitivity Analysis

Model discriminatory performance was largely unchanged when participants with missing family history were removed from the sample. Similarly, model discriminatory performance was also largely unchanged with recoding of uncertain family history of lung cancer from no lung cancer to yes lung cancer.

There was also no significant impact on model discriminatory performance due to recoding of personal history of cancer for both the scenario of patient self-report only and the scenario of patient self-report only with exclusion of non-melanoma skin cancer, coded as having personal history of cancer.

The results from these sensitivity analyses are not unexpected based on the lack of statistically significant differences in the screening sample between percentage of participants with and without lung cancer for family history of lung cancer and personal history of cancer (Table 8). However, statistical significance (or lack of) for individual variables is not always predictive of impact on model performance.[26]

In the screening program sample, model discriminatory performance was reduced when only self-reported COPD was coded as having COPD. Mean model predicted risk of lung cancer decreased for both cases and non-cases as did the difference in mean risk between them. The difference in mean risk was no longer statistically significant between the cases and non-cases at p</= 0.05. The percentage of participants in the screening program sample with COPD was the sample characteristic with the largest difference between the screening program sample and the model development and the NLST samples (73.4% vs. 9.1% and 5.1%, respectively). Additionally, it was one of only 2 reported screening program sample characteristics that were

significantly different between participants with lung cancer and participants without lung cancer. Accurately assessing COPD status for lung cancer screening participants will be important for discrimination in using the model for lung cancer screening participant selection.

7.4 Limitations

This study had several limitations. The model development sample had significantly different risk predictor characteristics as compared to the screening program sample. In particular, having COPD was more than 8 times as prevalent in the screening program sample as in the model development sample. The differences in the predictor characteristics were likely due both to differences in the actual sample characteristics as well as to data collection and coding methodology differences between the studies.

The differences in the study designs, including follow-up duration (1.86 years vs. 6 years) have resulted in reduced model performance in the screening program sample. Despite this, model predictive capability was adequate to help individualize risk and to help inform the shared decision making discussion.

The statistical analysis comparing the model calculated probabilities between groups used simple means testing. There are multiple ways of comparing risks and risk differences between groups that were not used for this study

The screening program sample was limited to 2 hospitals in the outskirts of a large metropolitan city in the Northeastern United States. This sample may not be representative of all regions and the results of the study not generalizable across all screening sites.

Implications for Public Health

Lung cancer screening has only recently been recommended by the USPSTF and CMS and thus there is not much experience with screening in clinical lung cancer screening programs. Lung cancer risk prediction models have the potential to provide individualized risk predictions for people helping to stratify the risk beyond the screening eligibility criteria. Screening the highest risk individuals may help to optimize the risk benefit ratio for screening.

The CMS requirement for a shared decision making meeting with the use of one or more decision aids for reimbursement prior to entering a lung cancer screening program provides a window of opportunity to incorporate risk prediction models into the lung cancer screening decision. This study demonstrated that the PLCOm2012noEd model has adequate to good predictive capability, especially around the risk screening threshold, to use in individualizing the risk discussion helping physicians and their patients better balance risk and patient values and preferences. Studies such as this one are important in increasing physician confidence in risk prediction models to encourage incorporation into clinical practice. Decision aid developers will also benefit from including this model into interactive web based aids.

Studies such as this should be done at other screening sites with sufficient sample size, or once sufficient sample size is attained. Additionally, a prospective, multi-center study, to assess model predictive capability in clinical screening programs would provide both additional evidence for use and increase adoption in clinical practice. Assessment of models with alternate/additional risk predictor variables should be developed and assessed for the NCCN Group 2 population. Ideal would be for CMS to cover NCCN Group 2, with evidence development, to collect the data required to assess risk and develop accurate risk prediction models.

Other studies, such as the Fiscella et al. NHANES evaluation of predicted risk in different race/ethnic groups,[31] should be undertaken to identify the population at equivalent or higher risk that is not covered by USPSF/CMS guidelines. Recent studies in people diagnosed with lung cancer have shown half to two thirds do not meet the guidelines.[33,34] As smoking rates continue to decline, fewer people at risk will meet the guidelines. Having validated risk prediction models will help refine the screening criteria with the potential for improved risk benefit ratios and more lives saved.

Chapter 8: Conclusions

Lung cancer risk prediction model, PLCOm2012noEd, predictive performance in a clinical lung cancer screening program was adequate to help patients and their physicians assess individual risk of lung cancer relative to the recommended model risk screening threshold (1.51%) and to supplement USPSTF and CMS screening program entry criteria for shared decision making.

An accurate assessment of COPD status is important for model predictive performance.

Model risk predictive capability for NCCN Group 2 did not match actual screening program lung cancer results. This group was at equal risk to those meeting USPSTF/CMS criteria. More work is needed to understand the additional risk factors needed to increase model performance for this group.

References:

1. National Cancer Institute. Cancer prevalence and cost of care projects. Available at: http://costprojections.cancer.gov/graph.php. Accessed on November 29, 2015.

2. Bradley CJ, Yabroff KR, Dahman B, Feuer EJ, Mariotto A, Brown ML. Productivity Costs of Cancer Mortality in the United States: 2000–2020. *JNCI Journal of the National Cancer Institute*. 2008;100(24):1763-1770. doi:10.1093/jnci/djn384. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2720777/

3. American Cancer Society. Cancer Facts & Figures 2016. Atlanta: American Cancer Society; 2016. http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/index

4. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening*. N Engl J Med* 2011;365(5):395-409. PM:21714641

5. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and study design. Radiology. 2011;258(1):243-253. http://radiology.rsna.org/lookup/suppl/doi:10.1148/radiol.10091808/-/DC1.

6. National Comprehensive Cancer Network. NCCN guidelines for lung cancer screening (V.1.2016) http://www.nccn.org/professionals/physician_gls/pdf/lung_screening.pdf Accessed March 31, 2016.

7. McKee AB, McKee BJ, Wald C, et al. Rescue Lung, Rescue Life: Translating the National Lung Screening Trial Results into Clinical Practice. *Oncology Issues*. 2014;29(2):20-29.

8. McKee BJ, McKee AB, Flacke S, Lamb CR, Hesketh PJ, Wald C. Initial experience with a free, high-volume, low-dose CT lung cancer screening program. *Journal of the American College of Radiology: JACR.* 2013;10(8):586-592.

9. Moyer VA. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of internal medicine*. 2014;160(5):330-338.

10. Decision Memo for Screening for Lung Cancer with Low Dose Computed Tomography (LDCT) (CAG-00439N). http://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274&NcaName=Screening+for+Lung+Cancer+with+Low+Dose+Computed+Tomography+%28LDCT%29&MEDCACId=68&IsPopup=y&bc=AAAAAAAAAgAAAA%3d%3d Published February 2015. Accessed December 8, 2015.

11. U.S.Preventive Services Task Force. Screening for Lung Cancer. Topic Page. December 2013. http://www.uspreventiveservicestaskforce.org/uspstf/uspslung.htm.

12. Tammemagi CM, Pinsky PF, Caporaso PA, et al. Lung cancer risk prediction: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial models and validation. *J Natl Cancer Inst.* 2011;103(13):1058-1068. Doi: 10.1093/jnci/djr173.

13. Tammemagi MC, Katkiha HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med*. 2013;368(8):728-236 DOI: 10.1056/NEJMoa1211776.

14. Jimbo M, Rana GK, Hawley S, et al. What is lacking in current decision aids on cancer screening? *CA Cancer J Clin*. 2013;63:193-214. Doi:10.3322/caac.2180.

15. Edwards A, Elwyn G. The potential benefits of decision aids in clinical medicine. *JAMA.* 1999;282:779-780.

16. Barry MJ. Health decision aids to facilitate shared decision making in an office practice. *Ann Intern Med.* 2002;136:127-135.

17. Legare F, Stacey D, Cossi MJ, et al. Interventions for improving the adoption of shared decision making by health care professionals. *Cochrane Data Base Syst Rev*. 2014;9. Doi: 10.1002/14651858.CD006732.pub3.

18. Lewis JA, Petty WJ, Tooze JA, et al. Low-Dose CT lung cancer screening practices and attitudes among primary care providers at an academic medical center. *Cancer Epidemiol Biomarkers Prev.* 2015;24(4):664-670. Doi: 10.1158/1055-9965.EPI-14-1241.

19. Iaccarino JM, Clark J, Bolton R, et al. A national survey of pulmonologists' views on Low-Dose CT screening for lung cancer*. Annals ATS*. First published online 14 Sep 2015 as DOI: 10.1513/AnnalsATS.201507-467OC.

20. Stacey D, Legare F, Col NF, et al. Decision aids for people facing health treatments or screening decisions. *Cochrane Data Base Syst Rev*. 2014;1. Doi: 10.10.1002/14651858.CD001431.pub4.

21. Kim SP, Knight SJ, Tomori C, et al. Health literacy and shared decision making for prostate cancer patients with low socioeconomic status*. Cancer Invest*. 2001;19(7):684-691. Doi:10.1081/CNV-100106143.

22. Volk RJ, Jiobaja-Weiss ML, Hawley ST, et al. Entertainment education for prostate cancer screening: A randomized trial among primary care patients with low health literacy. *Patient Educ Couns*. 2008;7(33):482-489. Doi:10.1016/j.pec.2008.07.033.

23. Schroy PC, Mylvaganam S, Davidson P. Provider perspectives on the utility of a colorectal cancer screening decision aid for facilitating shared decision making. *Health Expectations.* 2011;17:27-35. Doi: 10.111/j.1369-7625.2011.0073.x.

24. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Schunemann HJ. GRADE: what is "quality of evidence" and why is it important to clinicians? *BMJ*.2008;336:995-998.

25. Usher-Smith J, Emery J, Hamilton W, Griffin SJ, Walter FM. Risk prediction tools for cancer in primary care. *British Journal of Cancer.* 2015;113:1645–1650. doi:10.1038/bjc.2015.409

26. Tammemagi MC. Application of risk prediction models to lung cancer screening. *J Thorac Imaging*. 2015;30(2):88-100. DOI: 10.1097/RTI.0000000000000142.

27. Li K, Husing A, Sookthai D, Bergmann M, and et al. Selecting high-risk individuals for lung cancer screening: A prospective evaluation of existing risk models and eligibility in the German EPIC cohort. *Cancer Prev Res*. 2015;8(9):777-785. Doi:101158/1940-6207.CAPR-14-0424.

28. Brock University. Lung cancer risk calculators. https://www.brocku.ca/lung-cancer-risk-calculator. Accessed March 31, 2016.

29. LauYK, Caverly TJ, Cherng ST, et al. Development and validation of a personalized, web-based decision aid for lung cancer screening using mixed methods: A study protocol. *JMIR Res Protoc*. 2014;3(4):e78. Doi:10.2196/resprot.4039.

30. Tammemagi MC, Church TR, Hocking WG, et al. Evaluation of the lung cancer risks at which to screen ever-and never-smokers: Screening rules applied to the PLCO and NLST cohorts. *PLoS Medicine* 2014;11(12):e1001764. doi:10.1371/journalpmed1001764.

31. Fiscella K, Winters P, Farah S, Sanders M, Mohile SG. Do lung cancer eligibility criteria align with risk among blacks and Hispanics? *PLos ONE*. 2015;10(11):e0143789. Doi:10.1371/journal.pone.0143789.

32. Pinsky PF, Kramer BS. Lung cancer risk and demographic characteristics of current 20-29 pack-year smokers: Implications for screening. *JNCI J Natl Cancer Inst*. 2015;107 (11): djv226. doi: 10.1093/jnci/djv226

33. Farjah F, Wood DE, Zadworny ME, Rusch VW, Rizk NP. Resected lung cancer patients who would and would not have met screening criteria. *Ann Thorac Surg.* 2016;101(1):274-279. Doi: 10.1016/j.athoracsur.2015.06.010.

34. Yang P, Wang Y, Wampfler JA, et al. Trends in subpopulations at high risk for lung cancer. *J Thorac Oncol.* 2015;11(2):194-202. http://dx.doi.org/10.1016/j.jtho.2015.10.016.

35. National Cancer Institute, Division of Cancer Prevention. Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO). http://prevention.cancer.gov/major-programs/prostate-lung-colorectal/background-information. Accessed on March 31, 2016.

36. Hocking WG, Hu P, Oken MM, et al. Lung cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial. *J Natl Cancer Inst.* 2010;102(10):722-731. Doi:10.1093/jnci/djq126.

37. Smoking Pack Years website. http://smokingpackyears.com/. Accessed on April 2, 2016.

38. Gail MH, Pfeiiffer RM. On criteria for evaluating models of absolute risk. *Biostatistics*. 2005;6(2):227-239. DOI: 10.1093/BIOSTATISTICS/KXI005.

39. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models. A framework for traditional and novel measures. *Epidemiology*. 2010;21(1). DOI: 10.1097/EDE.0b013e3181c0fb2.

40. Lloyd-Jones DM. Cardiovascular risk prediction. Basic concepts, current status, and future directions. Circulation. 2010;121:1768-1777. doi: 10.1161/CIRCULATIONAHA.109.849166.

41. Cook NR. Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry*. 2007;54(1):17-23. DOI: 10.1373/clinchem.2007.096529.

42. Hosmer DW, Hosmer T, Cessie LE, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med*. 1997;16:965-980. DOI: 10.1002/(SICI)1097-0258(19970515)16:9<965::AID-SIM509>3.0.CO;2-O.

43. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

44. National Cancer Institute, National Cancer Access System. NLST data sets. Data collection forms. https://biometry.nci.nih.gov/cdas/datasets/nlst/forms/. Accessed on March 31, 2016.

45. National Cancer Institute, Division of Cancer Prevention. Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO). Baseline questionnaire for female participants (BQF3). October 1999. https://biometry.nci.nih.gov/app_media/cdas-bioshare/documents/studies/plco/forms/BQF3.pdf Accessed on March 31, 2016.

46. NIH. National Heart Lung and Blood Institute. What is COPD? http://www.nhlbi.nih.gov/health/health-topics/topics/copd/. Accessed on April 1, 2016.

47. Aberle DR, Adams AM, Berg CD, et al. Baseline characteristics of participants in the randomized National Lung Screening Trial. *J Natl Cancer Inst*. 2010;102:1771-1779. DOI: 10.1093/jnci/djq434. http://www.nhlbi.nih.gov/health/health-topics/topics/copd/

48. National Cancer Institute, Division of Cancer Prevention. Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO). Early questionnaires. https://biometry.nci.nih.gov/cdas/learn/plco/early-qx/. Accessed on March 31, 2016.

49. National Cancer Institute, Division of Cancer Prevention. Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO).Follow-up questionnaires. https://biometry.nci.nih.gov/cdas/learn/plco/follow-up-qx/. Accessed on March 31, 2016.

50. Bastarrika G, Wisnivesky JP, Pueyo JC, et al. Low-dose volumetric computed tomography for quantification of emphysema in asymptomatic smokers participating in an early lung cancer detection trial. Thorac Imaging. 2009;24(3):206-211.

51. McKee BJ, McKee AB, Borondy Kitts A, Regis SM, Wald C. Low-dose computed tomography screening for lung cancer in a clinical setting: Essential elements of a screening program. *Thorac Imaging*. 2015;30(2):115-129. Doi:10.1097/RTI.0000000000000139.

52. Matakidou A, Eisen T, Houlstn RS. Systematic review of the relationship between family history and lung cancer risk. *British Journal of Cancer*. 2005;93:825-833. Doi:10.1038/sj.bjc.6602769.