EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original research article

Scan-based competing death risk model for reevaluating lung cancer computed tomography screening eligibility

Anton Schreuder, Colin Jacobs, Nikolas Lessmann, Mireille JM Broeders, Mario Silva, Ivana Išgum, Pim A de Jong, Michel M van den Heuvel, Nicola Sverzellati, Mathias Prokop, Ugo Pastorino, Cornelia M Schaefer-Prokop, Bram van Ginneken

Please cite this article as: Schreuder A, Jacobs C, Lessmann N, *et al*. Scan-based competing death risk model for reevaluating lung cancer computed tomography screening eligibility. *Eur Respir J* 2021; in press (https://doi.org/10.1183/13993003.01613-2021).

Scan-based competing death risk model for reevaluating lung cancer computed tomography screening eligibility

Author list: Anton Schreuder MD[1], Colin Jacobs PhD[1], Nikolas Lessmann PhD[1], Mireille JM Broeders PhD[2,3], Mario Silva PhD[4,5], Ivana Išgum PhD[6,7], Pim A de Jong PhD[8,9], Michel M van den Heuvel[10], Nicola Sverzellati PhD[7], Mathias Prokop PhD[1], Ugo Pastorino PhD[6], Cornelia M Schaefer-Prokop PhD[1,11], Bram van Ginneken PhD[1,12]

Author affiliations:

[1] Department of Medical Imaging, Radboud University Medical Center, Nijmegen, the Netherlands

[2] Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, the Netherlands

[3] Dutch Expert Centre for Screening, Nijmegen, the Netherlands

[4] Unit of Thoracic Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

[5] Section of Radiology, Unit of Surgical Sciences, Department of Medicine and Surgery (DiMeC), University of Parma, Parma, Italy

[6] Department of Biomedical Engineering and Physics, Amsterdam UMC - location AMC, Amsterdam

[7] Department of Radiology and Nuclear Medicine, Amsterdam UMC - location AMC, Amsterdam

[8] Department of Radiology, University Medical Center Utrecht, the Netherlands

[9] Utrecht University, the Netherlands

[10] Department of Respiratory Diseases, Radboud University Medical Center, Nijmegen, the Netherlands

[11] Department of Radiology, Meander Medisch Centrum, Amersfoort, the Netherlands

[12] Fraunhofer MEVIS, Bremen, Germany

Corresponding author's contact details:

Anton Schreuder

Department of Medical Imaging

Radboud University Medical Center

Geert Grooteplein Zuid 10

6525 GA Nijmegen

The Netherlands

Email: antoniusschreuder@gmail.com

Telephone: +31 24 365 57 31

Author contributions:

Anton Schreuder: Conceived and designed the analysis, performed the analysis, wrote the paper.

Colin Jacobs: Conceived the analysis, supervised the analysis, collected the data, contributed data and analysis tools, critically appraised the paper.

Nikolas Lessmann: Conceived the analysis, collected the data, contributed data and analysis tools, critically appraised the paper.

Mireille JM Broeders: Conceived the analysis, supervised the analysis, critically appraised the paper.

Mario Silva: Conceived the analysis, collected the data, contributed data, critically appraised the paper.

Ivana Išgum: Conceived the analysis, contributed analysis tools, critically appraised the paper.

Pim A de Jong: Conceived the analysis, contributed analysis tools, critically appraised the paper.

Michel M van den Heuvel: Conceived the analysis, critically appraised the paper

Nicola Sverzellati: Conceived the analysis, collected the data, contributed data, critically appraised the paper.

Mathias Prokop: Conceived the analysis, critically appraised the paper.

Ugo Pastorino: Conceived the analysis, collected the data, contributed data, critically appraised the paper.

Cornelia M Schaefer-Prokop: Conceived the analysis, critically appraised the paper.

Bram van Ginneken: Conceived the analysis, supervised the analysis, contributed analysis tools, critically appraised the paper.

Key words: Lung Neoplasms; Mass Screening; Risk; Tomography, X-Ray Computed; Biological Markers

Funding source: No specific funding was obtained for this study.

An online risk calculator for the models described can be accessed online at https://docs.google.com/spreadsheets/d/1lU-UH1mxOmI-O--sNo8IhAgu2WLzRkcOQBCEm4rLdb4/edit?usp=sharing

Scan-based competing death risk model for reevaluating lung cancer computed tomography screening eligibility

Abstract

Purpose

A baseline CT scan for lung cancer (LC) screening may reveal information indicating that certain LC screening participants can be screened less, and instead require dedicated early cardiac and respiratory clinical input. We aimed to develop and validate competing death (CD) risk models using CT information to identify participants with a low LC and a high CD risk.

Methods

Participant demographics and quantitative CT measures of LC, cardiovascular disease, and chronic obstructive pulmonary disease were considered for deriving a logistic regression model for predicting five-year CD risk using a sample from the National Lung Screening Trial (n=15000). Multicentric Italian Lung Detection data was used to perform external validation (n=2287).

Results

Our final CD model outperformed an external pre-scan model (CDRAT) in both the derivation (Area under the curve = 0.744 [95% confidence interval = 0.727 to 0.761] and 0.677 [0.658 to 0.695], respectively) and validation cohorts (0.744 [0.652 to 0.835] and 0.725 [0.633 to 0.816], respectively). By also taking LC incidence risk into consideration, we suggested a risk threshold where a subgroup (6258/23096, 27%) was identified with a number needed to screen to detect one LC of 216 (vs. 23 in the remainder of the cohort) and ratio of 5.41 CDs per LC case (vs. 0.88). The respective values in the validation cohort subgroup (774/2287, 34%) were 129 (vs. 29) and 1.67 (vs. 0.43).

Conclusions

Evaluating both LC and CD risks post-scan may improve the efficiency of LC screening and facilitate the initiation of multidisciplinary trajectories among certain participants.

Abstract word count: 250

Take-Home Message

Lung cancer CT screening participants with a relatively low risk of lung cancer incidence and a high risk of competing death can be identified by applying two respective post-scan risk models, who in turn may benefit from other personalized trajectories.

Abbreviations

CD = competing death
CDRAT = Competing Death Risk Assessment Tool
COPD = chronic obstructive pulmonary disease
CT = low-dose computed tomography
CVD = cardiovascular disease
LC = lung cancer
LCRAT = Lung Cancer Risk Assessment Tool
MILD = Multicentric Italian Lung Detection
NCI = National Cancer Institute
NLST = National Lung Screening Trial
$PLCO_{M2012}$ = Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Model 2012

Introduction

Various randomized controlled trials have demonstrated that lung cancer (LC) screening with low-dose computed tomography (CT) significantly reduces the number of LC deaths compared to chest radiography (1) or no screening (2,3). However, most deaths which occurred in these trials were among LC-free participants (1–3). Even among participants who died of LC, most were reported to have other underlying causes of death. This indicates that preventing a LC death does not always lead to a gain in life years (4,5).

Screening eligibility is primarily based on two demographic predictors of LC incidence: age and smoking history (6,7). In addition, people with contra-indications for curative LC treatments may be excluded. This is to ensure a sufficiently high detection rate of treatable LCs while avoiding potential harms caused by false-positive findings (8). Various LC risk models exist specifically for determining screening participant eligibility among ever-smokers; the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Model 2012 ($PLCO_{M2012}$) and Lung Cancer Risk Assessment Tool (LCRAT) are among the most renown and best performing (9,10).

After the baseline screening round, chest CT biomarkers can be used to improve prediction accuracy. Imaging features are especially beneficial for estimating nodule malignancy risk (11–17). Moreover, quantitative CT measures (QCT) of cardiovascular disease (CVD) and chronic obstructive pulmonary disease (COPD) have also been validated as biomarkers in this setting (17–21). However, determining eligibility for future screening rounds is based on the same criteria as before the first scan (6,7). Whereas post-scan LC incidence risk models have been developed to personalize screening intervals beyond one year (11,12,17,22,23), there are currently no guidelines for reevaluating screening eligibility using CT information.

Regardless of their LC incidence risk, participants with a high risk of competing (i.e., non-LC) death (CD) may require multidisciplinary follow-up to benefit from screening. The ability to identify participants

with a low risk of developing LC and a high risk of encountering a CD (low-LC-high-CD) may enable more personalized recommendations to continue benefiting those in need while reducing potential harms in others. Two primary objectives can be summarized for this study: to develop a risk model combining demographic information and QCTs for predicting CD among LC screening participants, and to demonstrate the potential implications of halting further participation of low-LC-high-CD screenees. A previously developed model which considers the same QCTs from the baseline scan (i.e., of LC, CVD, and COPD) was used to determine LC risk (17). A secondary objective was to investigate whether a CD risk model combining both demographics and QCTs as predictors is superior to only using one or the other.

Methods

The "Scans and data," "Models and variables," and "Data set formation" subsections have been previously and more extensively described in Schreuder et al. (17); the relevant sections are provided in the supplement.

Scans and data

Scans and meta data from the National Lung Screening Trial (NLST) (ClinicalTrials.gov NCT00047385) were used to derive models with permission from the National Cancer Institute (Cancer Data Access System project ID NLST-437) (1). The NLST was the first and largest randomized controlled trial to demonstrate the effectiveness of CT screening in the reduction of LC death. Inclusion criteria were an age between 55 to 74 years, a cigarette smoking history of at least 30 pack years, and a smoking quit time within 15 years (if applicable). 26722 participants were assigned to the CT cohort and 26732 to the radiography cohort. Between 2002 and 2010, both cohorts underwent three annual screening rounds and were subsequently followed-up for five years (median follow-up time = 6.5 years).

The Multicentric Italian Lung Detection (MILD) randomized controlled trial (ClinicalTrials.gov NCT02837809) was used for external validation purposes (2). MILD was the first trial to demonstrate the

continued effectiveness of LC CT screening beyond the fifth year of screening. Inclusion criteria were an age greater or equal to 49 years, at least 20 pack years of smoking history, and a smoking quit time no longer than 10 years (if applicable). 1190 participants were assigned to annual CT screening, 1186 to biennial CT screening, and 1723 to no screening. Between 2005 and 2018, 93.5% of the participants were followed-up for at least nine years; the median time until the last screening round was six years.

Models and variables

Participant demographics and nodule CT features at baseline were directly provided by the NLST and MILD data sets. When multiple nodules were present in one scan, only the features from the nodule with the longest diameter were considered for the models. In the NLST, causes of death were determined by the endpoint verification process (or death certificate if unavailable). Causes of death for MILD were obtained exclusively from death certificates. Only underlying cause of death was considered for this study, either LC death or CD.

QCTs of CVD and COPD were extracted from the baseline CT scans using validated computer algorithms. This included calcium volume and mean calcium density in the coronary arteries (combined), mitral valve, aortic valve, and transthoracic aorta (19), emphysema score ("the percentage of lung voxels below -950HU after resampling the CT images to 3mm slice thickness, normalization, and bullae analysis") (24), and a reference parameter for measuring bronchial wall thickness called Pi10 ("the square root of the airway wall area for a theoretical 10mm lumen perimeter airway derived using the linear regression of the square root of segmented wall areas against the lumen perimeter extracted from the complete segmented airway tree") (20).

Three models for predicting the five-year risk of CD were derived: a model containing only self-reported demographic information (CD_{survey}), a model containing only age, sex, and CT information (CD_{CT}), and a final model considering all variables from the previous two models (CD_{final}). The previously described

LCi_{final} model was used to calculate the five-year LC incidence probability (Equations S1) (17); the model had been calibrated to each cohort. Table 1 lists all variables considered.

Data set formation

Subject exclusion criteria were the lack of a baseline CT scan, a baseline scan with a slice thickness >2.5mm, and missing data on LC incidence, death status, and times of event. The derivation cohort consisted of 15000 unique subjects from the NLST CT cohort which was the maximum allowed data extraction permitted by the National Cancer Institute (NCI) (Figure S1). This included all participants who were diagnosed with LC or died within the trial period (n=2106) and a random sample of those who did not encounter either of the events (n=12894); this latter group was resampled without replacement (n=8096), ultimately forming a derivation cohort of 23096 subjects. This method for weighted analysis was performed to simulate the original NLST cohort, hereby preventing an overestimation of risk.
All 2287 subjects from the MILD CT cohort formed the validation cohort. Multiple imputation was used to create plausible values for other missing data (see supplemental "Methods" section). Missing QCTs were considered an exclusion criterion for the derivation cohort but not the validation cohort, which were given the respective median values from the MILD data set (n=24 for QCTs of CVD and n=132 for Pi10). In addition, a MILD subgroup of NLST-eligible participants was formed, namely those between 55 and 74 years of age, at least 30 pack-years of smoking intensity, and no more than 15 years since smoking cessation if applicable (1).

Statistical analysis

All statistical analyses were performed in R version 3.4.3. The three parsimonious risk prediction models for competing death were derived using logistic regression. Variables were included in the model if the level of significance was <0.20 (backward elimination) (25). R function "mfp" (mfp package) was used to select factorial polynomials for continuous variables (level of significance <0.05) (26). Second-degree

polynomials were considered for longest nodule diameter; only first-degree polynomials were considered for the remaining variables.

Areas under the receiver operating characteristic curve (AUC) with 95% confidence intervals were calculated for our derived models (CD_{survey}, CD_{CT}, and CD_{final}) and the previously described CD Risk Assessment Tool (CDRAT) (10). Internal validation was performed for each model with 1000 bootstraps to assess overfitting in the form of optimism (the difference between the average bootstrap sample AUC and original sample AUC). The values were compared with 500 bootstraps to test whether the discriminative performances were significantly different (p value < 0.05). Calibration is visualized in the form of calibration plots. The models were externally validated on the full MILD cohort.

To demonstrate practical application, the cohorts were stratified into quintiles (five equally sized groups) based on five-year risk of LC incidence (LCi_{final}) and competing death (CD_{final}). The quintiles were used to form five-by-five contingency tables showing the distribution of LC diagnoses and competing deaths within five-years' follow-up. Based on visual findings, we suggest criteria to select low-LC-high-CD participants. The potential consequences of applying these selection criteria are summarized by calculating the number needed to screen to detect one LC (NNS) and the ratio of CDs per LC diagnosis. The same risk probability thresholds were applied to the full and NLST-eligible MILD cohorts for external validation. This analysis was repeated using external pre-scan risk models for LC risk (i.e., $PLCO_{m2012}$ and LCRAT) and CD risk (i.e., CDRAT) (10,27), using risk probability thresholds based on model-dependent quintiles. McNemar's test was used to determine whether the frequencies between the different low-LC-high-CD groups were statistically different (28).

Results

Study participants

Within five-years' follow-up, 756 LC diagnoses (3.3%) and 800 competing deaths (3.5%) were reported in the derivation cohort (n=23096); 22 participants encountered both outcomes. The respective numbers in the validation cohort (n=2287) were 59 (2.6%) and 33 (1.4%), with two participants overlapping. On average, the NNS in the derivation and validation cohorts are 31 (23096/756) and 39 (2287/59), respectively. The respectively ratios of CDs per LC diagnosed within five years' follow-up are 1.06 (800/756) and 0.56 (33/59). Descriptive statistics are summarized in Table 1.

Competing death risk prediction

Three competing death risk models were derived; the variables and coefficients of the CD_{final}, CD_{survey}, and CD_{CT} are reported in Tables 2, S1, and S2, respectively. AUC of CD_{final} was 0.744 (95% confidence interval = 0.727 to 0.761), significantly greater than that of CD_{survey} (0.707, 0.689 to 0.725) and CD_{CT} (0.719, 0.701 to 0.737) ($p<0.001$) (Figure S2). Internal validation revealed an optimism no greater than 0.006 (Table S3). The AUC of the CDRAT model was 0.677 (0.658 to 0.695), significantly inferior to that of CD_{survey} ($p<0.001$).

External validation resulted in AUCs of 0.744 (0.652 to 0.835), 0.721 (0.627 to 0.815), 0.756 (0.667 to 0.844), and 0.725 (0.633 to 0.816) for CD_{final}, CD_{survey}, CD_{CT}, and CDRAT, respectively (no statistically significant differences) (Figure S3). The calibration curves of CD_{final} in the derivation and validation cohort are shown in Figures S4 and S5, respectively. Note that the deviation of the calibration curve from the diagonal in the validation cohort is caused by one outlier case (five-year CD risk probability = 19.7%).

Decision curve analysis of the CD models are available in the supplement (Figures S6, S7) (29).

Stratification by lung cancer risk and competing death risk

Figure 1 is a collection of three-dimensional column charts and two-by-two contingency tables divided into cells based on risk quintiles according to CD_{final} (vertical axis) and LCi_{final} (horizontal axis). The cut-off values are reported in Table S4. Contingency tables for the validation cohort are given in Table S5 and Figure S6.

Based on the 5-year risk estimate, we determined visually that participants with a LCi_{final} risk ≤0.79%, LCi_{final} risk ≤1.38% and CD_{final} risk >2.93%, or LCi_{final} risk ≤2.18% and CD_{final} risk >4.92% had a relatively high ratio (generally >3) of CDs per LC diagnosis. 27% (6258/23096) of the derivation cohort fit these criteria and were stratified into the low-LC-high-CD group (Table 3). 4% (29/756) of the LC cases and 20% (157/800) of the CDs occurred in this group; the NNS was 216 (full derivation cohort average = 31) and 5.41 CDs occurred per LC diagnosis (average = 1.06).

The same threshold risk probabilities were applied for external validation (Table 3). 34% (774/2287) of MILD participants fell into the low-LC-high-CD group, consisting of 10% (6/59) of the LC cases and 30% (10/33) of the CDs. The NNS was 129 (full validation cohort average = 39) and 1.67 CDs occurred per LC diagnosis (average = 0.56). When only considering NLST-eligible MILD participants, the resulting proportions, ratios, and NNS were closer to those from the derivation cohort.

The performance of three external pre-scan models were also assessed using the same methods, first by combining LCRAT with CDRAT, then combining $PLCO_{M2012}$ with CDRAT (Table 4). Note that the risk probability criteria were changed according to each model's quintile cut-off points in the derivation cohort. Compared to the post-scan models, significantly more LC cases ($p<0.001$ for both LCRAT and $PLCO_{M2012}$) and fewer CDs ($p=0.048$ for LCRAT and $p=0.012$ for $PLCO_{M2012}$) were stratified to the low-LC-high-CD groups based on pre-scan model risk. A similar trend was found in the validation cohort, but the differences between the pre- and post-test models were not statistically significant. In the NLST-eligible

validation cohort, the external models resulted in a lower ratio of CDs per LC diagnosis in the low-LC-high-CD group than in the high-LC-low-CD group.

The outcomes of using CD_{CT}, CD_{survey}, or CDRAT for stratifying a low-LC-high-CD group given the same LC risk thresholds (LCi_{final}) are available for comparison in Table S6.

Discussion

Competing death risk prediction

Three five-year CD risk prediction models were derived, including only self-reported patient demographics (CD_{survey}), only QCTs (CD_{CT}), or both (CD_{final}). CD_{CT} and CD_{final} are the first models using the baseline CT scan for this purpose. CD_{survey} and the external model CDRAT were used to quantify the added value of CT information in addition to pre-scan information (Figures S2, S3) (10).

The discriminative performance of CD_{final} was significantly superior to the other models in the derivation cohort (Figure S2). Due to the small differences in AUC and smaller sample size of MILD, no significant differences were determined between the models in the validation cohort (Figure S3). However, the trend remained that CT-based models (CD_{CT} and CD_{final}, 0.756 and 0.744, respectively) had a higher accuracy than pre-scan models (CDRAT and CD_{survey}, 0.725 and 0.721, respectively). This suggests that a model based exclusively on QCTs (CD_{CT}) may be a viable option for automatically calculating risk scores in a high-risk population.

Assessment of model predictors

The most notable observation of the variables included in CD_{final} is the lack of nodule CT features. Despite there being a seemingly high correlation between the risks of LC and CD (Figures 1 and S6), nodule information is not sufficiently distinctive compared to the contribution of other risk factors of non-LC causes of death. In CD_{CT}, only the presence of a solid opacity as the largest nodule was included; this had a negative beta coefficient, suggesting that participants with a nodule were more likely to

encounter a LC death instead of a CD. In turn, QCTs of CVD and COPD significantly improved the accuracy of CD_{final} and CD_{CT}. It had been previously shown that these QCTs consistently contributed to the prediction of CVD death, COPD death, LC death, and LC incidence (17).

Regarding measures of cardiovascular calcifications, we observed that a lower aorta mean calcium density was associated with a greater CD risk. A lower plaque density may indicate instability in an elderly population and therefore be more likely to be the cause of death (17,30,31). This population with overall high CVD risk may represent a niche for specific risk stratification as opposed to the Agatston score used in populations with much broader risk ranges (32). Conversely, a positive association was observed between mitral valve mean calcium density and CD risk (CD_{CT}) which suggests that the pathophysiological mechanism of plaques is location dependent.

No other unexpected relationships between the variables and CD were observed. Note that no tests for multi-collinearity or interactions between variables were performed for this study. Therefore, causative relations between variables and outcomes should not be deduced based solely on these findings.

Stratification by lung cancer risk and competing death risk

The NELSON results suggested that fixed screening intervals greater than two years are contraindicated (33,34). At the same time, the idea that screening intervals can be lengthened among participants determined to have a post-scan low LC risk is gaining momentum (2,11,12,17,22). Schreuder et al. (17) proposed screening intervals extending past two years among participants with a sufficiently low LC risk. However, an approach which focuses solely on LC risk may not be optimal because the potential benefits of early LC detection may be humbled by underlying competing morbidities. Other pre-scan risk models have hereby applied approaches based on the likelihood to survive LC screening, namely by predicting LC death risk (10,35) or life-years gained (4,5).

Based on this idea that low-LC-high-CD participants are less likely to benefit from LC screening, we propose to consider multidisciplinary follow-up among a small group of individuals. This may involve

proactive trajectories according to cardiac and respiratory disease management guidelines for participants who are not yet under the care of a physician. An example of opportunistic risk assessment is given by the Manchester Lung Screening Pilot, which found that 47% of the participants at high risk of CVD (QRISK2 score ≥10%) were not taking lipid-lowering medication which is indicated for primary prevention (36).

Additionally, a longer screening interval of up to five years can be considered based on post-scan risk prediction. In the derivation cohort, one fourth (27%, 6258/23096) of the screenees would have received a counter-indication to continue participating. This would be at the expense of delaying the LC diagnosis in 4% (29/756) of the participants who would develop LC. We considered this a pessimistic estimate because the analysis did not consider participants undergoing additional follow-up CTs in response to suspicious findings in the baseline scan (4/29 participants were diagnosed within one year of the baseline scan); this group also includes participants which were both diagnosed with LC and encountered a CD (1/29). In exchange, 6229 people may have hypothetically avoided four annual screening rounds which would not have resulted in a LC diagnosis. 2.5% (157/6258) of low-LC-high-CD participants encountered competing deaths within five years, equivalent to 20% of all CDs (157/800). We caution that advocating no further screening may negate some of the positive psychological benefits, e.g., interest in lifestyle advice (37).

Using the same risk probability thresholds, a similar trend occurred in the validation cohort. A greater proportion of participants fell within the low-LC-high-CD group (34%, 774/2287) because the median five-year calibrated LC incidence risk was lower than in the derivation cohort (1.28% and 1.73%, respectively); this was despite the median CD risk also being lower (1.40% and 2.36%, respectively). This would have resulted in a greater proportion of both LC diagnoses (10%, 6/59) and CDs (30%, 10/33) within the low-LC-high-CD group.

Compared to pre-scan models, using LCi_{final} and CD_{final} resulted in higher values for both the NNS and CDs per LC diagnosis in the low-LC-high-CD group. The differences were not statistically significant in the validation cohort despite the similar trend. The risk thresholds used for stratification in our study were merely suggestive; we encourage adjusting the thresholds to satisfy each screening program's aims and values. At certain thresholds, a pre-scan CD model may even perform equivalently to a post-scan CD model.

Limitations

The primary limitation is that the external validation cohort had a low number of events of interest (<100). This was reflected in the wide confidence intervals. It was therefore also not considered useful to calibrate or retrain the model to MILD data. Once more LC screening CT data becomes available, more extensive model validation is recommended before considering implementation in practice.

Not having used Cox proportional hazards regression to model CD also introduces possible biases related to censoring (i.e., loss to follow-up) and competing risk from LC-specific mortality. However, logistic regression does not need to assume proportional hazards and is easier to interpret and implement as a risk model, which was sufficient for the purpose of this study (38,39). Though the coefficients of logistic regression may be a bit inflated compared to those from Cox regression, it is the predictive performance rather than the association between the predictors and outcome which is relevant. Also consider that the interpretation is complicated by ignoring the fact that most participants who died of LC were reported to have multiple secondary causes of death. The risk score was arbitrarily restricted to a five-year post-scan period to limit the loss to follow-up while being considerably longer than the standard one-year screening interval.

Another limitation is that nodule diameter was measured manually in the NLST (no volumetric data available), and only the data from the nodule with the longest diameter was considered when multiple nodules were detected. Volume (or mean diameter) obtained by semi-automatic software would offer

better discrimination (40,41). Whereas nodule size is the most predictive variable for malignancy (given a single scan), this risk can be further modified based on other features, mainly nodule type, location, and presence of spiculation (11,13,17). Note that nodule spiculation was not recorded by MILD. Also, QCTs automatically obtained from scans with slice thickness greater than 1 mm (namely from the NLST) may not be reliable (i.e., emphysema score, Pi10, calcium scores). Another possible issue with the measurement of CT calcium volume and density is that the scans were not ECG-gated, but prior studies indicate a strong predictive value (42) and high concordance with gated calcium scores (43).

Future directions

Additional diagnostic interventions (usually in the form of follow-up CTs) are recommended for nodules with an indeterminate malignancy risk (44,45). In current screening practice, this additional work up does not affect the timing of the subsequent annual screening rounds. We hypothesize that it would be of added value to know the outcome of additional diagnostic tests before deciding on post-scan screening eligibility (or screening interval length). Besides nodule growth, the availability of follow-up scans would enable the estimation of CVD and COPD progression rates. An alternative would be to manage nodules independently of other decisions: for example, a low-LC-high-CD participant with an indeterminate nodule would still be followed-up according to nodule guideline recommendations while being ineligible for further screening. Regardless, future studies should simulate decision trees at multiple time points while taking time-varying risk factors into consideration.

Conclusion

We derived five-year CD risk models using either self-reported patient characteristics (CD_{survey}), chest CT image biomarkers (CD_{CT}), or both (CD_{final}). QCTs of CVD and COPD were included in the CT-based models; pulmonary nodule morphological features were not found to be significant predictors. CT information provides an added value to the AUC of at least two percentage points. In a high-risk screening population, there may be little or no added value of patient demographics for predicting CD when QCTs

have been extracted: a CT scan alone may elucidate personalized susceptibility to smoke damage, aging, and other risk factors for CD.

By calculating both post-scan LC incidence risk (LCi_{final}) and CD risk (CD_{final}), a group of participants with a relatively low risk of the former and high risk of the latter can be identified (low-LC-high-CD). This means that the baseline scan can be used to help identify participants who may benefit from multidisciplinary action and can safely be recommended longer screening intervals. Using our suggested criteria for contra-indicating LC screening participation within the next five years may reduce the number of screenees by approximately one-fourth, of whom more than 200 would need to continue participating to detect one LC. Valuable health care resources could simultaneously be redirected towards the prevention of common competing deaths among low-LC-high-CD participants.

Acknowledgements

The authors thank Gabriel Humpire Mamani, Jean-Paul Charbonnier, and Leticia Gallardo-Estrella for their technical support in obtaining quantitative CT measures and Claudio Jacomelli and Frederica Sabia for extracting the requested Multicentric Italian Lung Detection (MILD) trial data. The authors also thank the MILD research teams for access to MILD data, and the NCI for access to the NCI's data collected by the National Lung Screening Trial under project number NLST-437. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*. 2011;365(5):395–409. doi:10.1056/NEJMoa1102873.

2. Pastorino U, Silva M, Sestini S, et al. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. *Annals of Oncology*. 2019;30(7):1162–9. doi:10.1093/annonc/mdz117.

3. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *New England Journal of Medicine*. 2020;382(6):503–13. doi:10.1056/NEJMoa1911793.

4. Cheung LC, Berg CD, Castle PE, Katki HA, Chaturvedi AK. Life-Gained–Based Versus Risk-Based Selection of Smokers for Lung Cancer Screening. *Annals of Internal Medicine*. 2019;171(9):623. doi:10.7326/M19-1263.

5. Caverly TJ, Cao P, Hayward RA, Meza R. Identifying Patients for Whom Lung Cancer Screening Is Preference-Sensitive. *Annals of Internal Medicine*. 2018;169(1):1. doi:10.7326/M17-2561.

6. Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. *The Lancet Oncology*. 2017;18(12):e754–66. doi:10.1016/S1470-2045(17)30861-6.

7. Moyer VA. Screening for Lung Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*. 2014;160(5):330–8. doi:10.7326/M13-2771.

8. Harris RP, Sheridan SL, Lewis CL, et al. The Harms of Screening. *JAMA Internal Medicine*. 2014;174(2):281. doi:10.1001/jamainternmed.2013.12745.

9. Katki HA, Kovalchik SA, Petito LC, et al. Implications of nine risk prediction models for selecting ever-smokers for computed tomography lung cancer screening. *Annals of Internal Medicine*. 2018;169(1):10–9. doi:10.7326/M17-2701.

10. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and Validation of Risk Models to Select Ever-Smokers for CT Lung Cancer Screening. *JAMA*. 2016;315(21):2300. doi:10.1001/jama.2016.6255.

11. Schreuder A, Schaefer-Prokop CM, Scholten ET, Jacobs C, Prokop M, van Ginneken B. Lung cancer risk to personalise annual and biennial follow-up computed tomography screening. *Thorax*. 2018;73(7). doi:10.1136/thoraxjnl-2017-211107.

12. Patz EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: A retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial. *The Lancet Oncology*. 2016;17(5):590–9. doi:10.1016/S1470-2045(15)00621-X.

13. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *New England Journal of Medicine*. 2013;369(10):910–9. doi:10.1056/NEJMoa1214726.

14. Raghu VK, Zhao W, Pu J, et al. Feasibility of lung cancer prediction from low-dose CT scan and smoking factors using causal models. *Thorax*. 2019;74(7):643–9. doi:10.1136/thoraxjnl-2018-212638.

15. Hassannezhad R, Vahed N. Prediction of the Risk of Malignancy Among Detected Lung Nodules in the National Lung Screening Trial. *Journal of the American College of Radiology*. 2018;15(11):1529–35. doi:10.1016/j.jacr.2018.06.009.

16. Silva M, Milanese G, Sestini S, et al. Lung cancer screening by nodule volume in Lung-RADS v1.1: negative baseline CT yields potential for increased screening interval. *European Radiology*. 2020; doi:10.1007/s00330-020-07275-w.

17. Schreuder A, Jacobs C, Lessmann N, et al. Combining pulmonary and cardiac computed tomography biomarkers for disease-specific risk modelling in lung cancer screening. *European Respiratory Journal*. 2021;2003386. doi:10.1183/13993003.03386-2020.

18. Gallardo-Estrella L, Pompe E, De Jong PA, et al. Normalized emphysema scores on low dose CT: Validation as an imaging biomarker for mortality. *PLoS ONE*. 2017;12(12). doi:10.1371/journal.pone.0188902.

19. Lessmann N, van Ginneken B, Zreik M, et al. Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks With Dilated Convolutions. *IEEE Transactions on Medical Imaging*. 2018;37(2):615–25. doi:10.1109/TMI.2017.2769839.

20. Charbonnier J-PP, Pompe E, Moore C, et al. Airway wall thickening on CT: Relation to smoking status and severity of COPD. *Respiratory Medicine*. 2019;146:36–41. doi:10.1016/j.rmed.2018.11.014.

21. Lebrett MB, Balata H, Evison M, et al. Analysis of lung cancer risk model (PLCO M2012 and LLP v2) performance in a community-based lung cancer screening programme. *Thorax*. 2020;thoraxjnl-2020-214626. doi:10.1136/thoraxjnl-2020-214626.

22. Robbins HA, Berg CD, Cheung LC, Chaturvedi AK, Katki HA. Identification of Candidates for Longer Lung Cancer Screening Intervals Following a Negative Low-Dose Computed Tomography Result. *JNCI: Journal of the National Cancer Institute*. 2019;111(9):996–9. doi:10.1093/jnci/djz041.

23. Silva M, Milanese G, Sestini S, et al. Lung cancer screening by nodule volume in Lung-RADS v1.1: negative baseline CT yields potential for increased screening interval. *European radiology*. 2020; doi:10.1007/s00330-020-07275-w.

24. Gallardo-Estrella L, Lynch DA, Prokop M, et al. Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. *European Radiology*. 2016;26(2):478–86. doi:10.1007/s00330-015-3824-y.

25. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. *American journal of epidemiology*. 1989;129(1):125–37. doi:10.1093/oxfordjournals.aje.a115101.

26. Zhang Z. Multivariable fractional polynomial method for regression model. *Annals of Translational Medicine*. 2016;4(9):174–174. doi:10.21037/atm.2016.05.01.

27. Tammemägi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *The New England journal of medicine*. 2013;368(8):728–36. doi:10.1056/NEJMoa1211776.

28. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947;12(2):153–7. doi:10.1007/BF02295996.

29. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*. 2006;26(6):565–74. doi:10.1177/0272989X06295361.

30. Criqui MH, Knox JB, Denenberg JO, et al. Coronary Artery Calcium Volume and Density: Potential Interactions and Overall Predictive Value: The Multi-Ethnic Study of Atherosclerosis. *JACC: Cardiovascular Imaging*. 2017;10(8):845–54. doi:10.1016/j.jcmg.2017.04.018.

31. Criqui MH, Denenberg JO, Ix JH, et al. Calcium Density of Coronary Artery Plaque and Risk of Incident Cardiovascular Events. *JAMA*. 2014;311(3):271. doi:10.1001/jama.2013.282535.

32. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European guidelines on cardiovascular disease prevention in clinical practice. The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts. Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation. *Giornale italiano di cardiologia (2006)*. 2017;18(7):547–612. doi:10.1714/2729.27821.

33. Baldwin DR, Duffy SW, Devaraj A, Field JK. Optimum low dose CT screening interval for lung cancer: The answer from NELSON? Vol. 72, Thorax. BMJ Publishing Group; 2017. p. 6–7. doi:10.1136/thoraxjnl-2016-209011.

34. Yousaf-Khan U, van der Aalst C, de Jong PA, et al. Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval. *Thorax*. 2017;72(1):48–56. doi:10.1136/thoraxjnl-2016-208655.

35. Kovalchik SA, Tammemagi M, Berg CD, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. *New England Journal of Medicine*. 2013;369(3):245–54. doi:10.1056/NEJMoa1301851.

36. Balata H, Blandin Knight S, Barber P, et al. Targeted lung cancer screening selects individuals at high risk of cardiovascular disease. *Lung Cancer*. 2018;124:148–53. doi:10.1016/j.lungcan.2018.08.006.

37. Stevens C, Smith SG, Quaife SL, Vrinten C, Waller J, Beeken RJ. Interest in lifestyle advice at lung cancer screening: Determinants and preferences. *Lung Cancer*. 2019;128:1–5. doi:10.1016/j.lungcan.2018.11.036.

38. Bugnard F, Ducrot C, Calavas D. Advantages and inconveniences of the Cox model compared with the logistic model: application to a study of risk factors of nursing cow infertility. *Veterinary research*. 1994;25(2–3):134–9.

39. Moriguchi S, Hayashi Y, Nose Y, Maehara Y, Korenaga D, Sugimachi K. A comparison of the logistic regression and the cox proportional hazard models in retrospective studies on the prognosis of

patients with castric cancer. *Journal of Surgical Oncology*. 1993;52(1):9–13. doi:10.1002/jso.2930520104.

40. Tammemagi M, Ritchie AJ, Atkar-Khattra S, et al. Predicting Malignancy Risk of Screen-Detected Lung Nodules–Mean Diameter or Volume. *Journal of Thoracic Oncology*. 2019;14(2):203–11. doi:10.1016/j.jtho.2018.10.006.

41. Han D, Heuvelmans MA, Oudkerk M. Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening. *Translational Lung Cancer Research*. 2017;6(1):52–61. doi:10.21037/tlcr.2017.01.05.

42. Fan L, Fan K. Lung cancer screening CT-based coronary artery calcification in predicting cardiovascular events. *Medicine*. 2018;97(20):e10461. doi:10.1097/MD.0000000000010461.

43. Budoff MJ, Nasir K, Kinney GL, et al. Coronary artery and thoracic calcium on noncontrast thoracic CT scans: Comparison of ungated and gated examinations in patients from the COPD Gene cohort. *Journal of Cardiovascular Computed Tomography*. 2011;5(2):113–8. doi:10.1016/j.jcct.2010.11.002.

44. The American College of Radiology. Lung CT Screening Reporting & Data System v1.1 [Internet]. 2019. Available from: https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads

45. Wood DE, Kazerooni EA, Baum SL, et al. Lung Cancer Screening, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network*. 2018;16(4):412–41. doi:10.6004/jnccn.2018.0020.

Figure captions

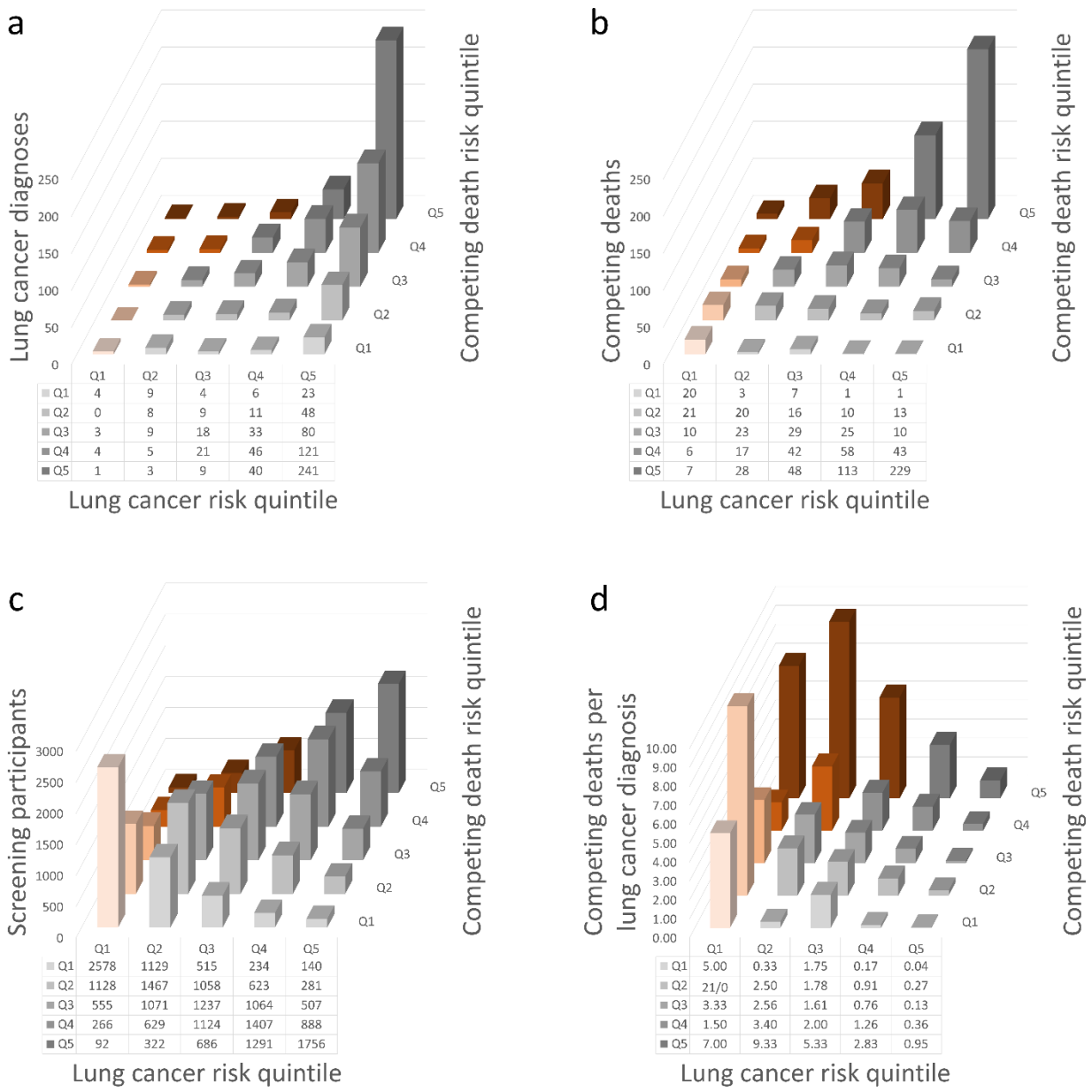


Figure 1: Outcomes per lung cancer and competing death risk quintile (derivation cohort)
Three-dimensional column charts of (a) lung cancer diagnoses, (b) competing deaths, (c) screening participants, and (d) competing deaths per lung cancer diagnosis (y-axis truncated at 10). Each competing death risk quintile is differently shaded, where a darker shade corresponds to a higher risk. The columns are divided into grey and orange columns indicating the suggested separation of screening participants into a group which should continue to be screened (grey) and a group with a relatively low lung cancer risk and high risk of competing death (orange).

Q#, risk quintile where Q1 represents the lowest quintile and Q5 the highest.

Table 1: Distribution of variables by five-year competing death event or not

Variables	Derivation cohort (NLST)		Validation cohort (MILD)	
	Competing death (n=800)	No competing death (n=22296)	Competing death (n=33)	No competing death (n=2254)
Patient characteristics				
Age (SD), years	63.6 (5.5)	61.3 (5.0)	62.8 (5.8)	57.5 (5.9)
Female sex (%)	223 (27.9)	9286 (41.6)	8 (24.2)	717 (31.8)
Race or ethnicity				
White (%)	714 (89.3)	19963 (89.5)	33 (100)	2250 (99.8)
Black (%)	51 (6.4)	1000 (4.5)	0	0
Asian (%)	6 (0.8)	542 (2.4)	0	2 (0.1)
Hispanic (%)	7 (0.9)	335 (1.5)	0	1 (0.0)
Mixed or other (%)	22 (2.8)	456 (2.0)	0	1 (0.0)
Educational level (SD), 0-5	3.6 (1.7)	3.7 (1.6)	1.7 (1.5)[b]	1.4 (1.4)[b]
BMI (IQR), kg/m^2	27.1 (24.0 to 30.9)	27.2 (24.4 to 30.5)	26.1 (22.4 to 30.5)	25.7 (23.5 to 28.4)
Current smoker (%)	455 (56.9)	10564 (47.4)	22 (66.7)	1544 (68.5)
Smoking intensity (IQR), pack years	54 (43 to 76)	48 (39 to 66)	45 (34 to 59)	39 (32 to 51)
Smoking duration (SD), years	43.2 (7.4)	39.7 (7.3)	43.1 (7.1)	38.1 (6.7)
Smoking quit time (IQR), years[a]	7 (3 to 12)	8 (4 to 12)	4 (2 to 6)	5 (3 to 8)
Lung cancer in family				
1 (%)	141 (17.6)	4088 (18.3)	3 (9.1)	550 (24.4)
≥2 (%)	30 (3.8)	712 (3.2)	NA	NA
Work asbestos (%)	58 (7.3)	1017 (4.6)	2 (6.1)	195 (8.7)
COPD diagnosis (%)	217 (27.1)	3943 (17.7)	6 (18.2)	266 (11.8)
Asthma diagnosis (%)	102 (12.8)	2158 (9.7)	3 (9.1)	143 (6.3)
Diabetes diagnosis (%)	145 (18.1)	2062 (9.2)	4 (12.1)	132 (5.9)
Heart disease diagnosis (%)	195 (24.4)	2707 (12.1)	4 (12.1)	277 (12.3)
Hypertension diagnosis (%)	363 (45.4)	7665 (34.4)	14 (42.4)	619 (27.5)
Stroke diagnosis (%)	58 (7.3)	565 (2.5)	2 (6.1)	20 (0.9)
Nodule CT features				
Nodule attenuation				
No nodule (%)	600 (75.0)	16296 (73.1)	18 (54.5)	980 (43.5)
Solid (%)	152 (19.0)	4636 (20.8)	11 (33.3)	1004 (44.5)

Partsolid (%)	14 (1.8)	359 (1.6)	0	61 (2.7)
Nonsolid (%)	34 (4.3)	1005 (4.5)	4 (12.1)	209 (9.3)
Longest diameter (IQR), mm[a]	7 (5 to 11)	6 (5 to 9)	4.6 (3.7 to 8.5)	4.9 (3.5 to 7.4)
Perpendicular diameter (IQR), mm[a]	5 (4 to 8)	5 (4 to 7)	3.5 (3.0 to 5.1)	3.9 (2.8 to 5.8)
Nodule in upper lobe (%)[a]	88 (41.7)	2496 (39.1)	7 (21.2)	370 (16.4)
Nodule spiculation (%)[a]	30 (14.2)	789 (12.4)	NA	NA
Nodule count (IQR)[a]	1 (1 to 2)	1 (1 to 2)	1 (1 to 2)	1 (1 to 2)
Quantitative CT measures of CVD				
Coronary calcium volume (IQR), mm^3	188 (36 to 788)	48 (2 to 267)	114 (21 to 466)	23 (0 to 154)
Coronary mean calcium density (IQR), HU	226 (190 to 264)	207 (141 to 251)	286 (227 to 314)	255 (0 to 311)
Transthoracic aorta calcium volume (IQR), mm^3	1134 (314 to 2894)	403 (89 to 1282)	948 (200 to 2992)	200 (45 to 694)
Transthoracic aorta mean calcium density (IQR), HU	326 (277 to 377)	311 (251 to 378)	418 (375 to 451)	434 (363 to 523)
Mitral valve calcium volume (IQR), mm^3	0 (0 to 18)	0 (0 to 0)	0 (0 to 2)	0 (0 to 0)
Mitral valve mean calcium density (IQR), HU	0 (0 to 206)	0 (0 to 0)	0 (0 to 203)	0 (0 to 0)
Aortic valve calcium volume (IQR), mm^3	0 (0 to 20)	0 (0 to 0)	0 (0 to 33)	0 (0 to 0)
Aortic valve mean calcium density (IQR), HU	0 (0 to 181)	0 (0 to 0)	0 (0 to 241)	0 (0 to 0)
Quantitative CT measures of COPD				
Total lung volume (IQR), liters	5.6 (4.7 to 6.7)	5.4 (4.5 to 6.4)	6.2 (5.2 to 6.9)	5.9 (5.1 to 6.8)
Mean lung density (IQR), HU	-834 (-857 to -808)	-839 (-858 to -815)	-845 (-872 to -834)	-846 (-861 to -828)
Emphysema score (IQR)	0.39 (0.08 to 1.77)	0.24 (0.05 to 1.12)	0.08 (0.00 to 1.22)	0.03 (0.00 to 0.17)
Pi10 (IQR)	3.0 (2.5 to 3.6)	2.8 (2.3 to 3.3)	2.5 (2.2 to 3.0)	2.4 (2.2 to 2.6)

[a] Of those applicable; regarding nodule features, applies to only the nodule with the longest diameter.

[b] On a scale of 0 to 4.

Continuous variables with a normal distribution are given in mean and SD; continuous variables with a non-normal distribution are given in median and IQR; categorical variables are given in %. Educational level: a categorical variable applied as a continuous variable, where 0 = did not complete high school, 1 = high school graduate, 2 = post high school training but no college, 3 = some college, 4 = bachelor's degree, and 5 = graduate school or higher. Lung cancer in family: number of first-degree family members diagnosed with lung cancer (a value of "2" was given when two or more family members were diagnosed). COPD diagnosis: includes prior diagnosis of COPD, emphysema, and/or chronic bronchitis. Mean lung density was centered at -1000. Inter-cohort statistics are reported by Schreuder et al. (17).

BMI, body-mass index; COPD, chronic obstructive pulmonary disease; CVD, cardiovascular disease; HU, Hounsfield units; IQR, interquartile range; MILD, Multicentric Italian Lung Detection; NA, not applicable; NLST, National Lung Screening Trial; Pi10, measure of bronchial wall thickness; SD, standard deviation.

Table 2: Final competing death risk model (CD_{final})

Variable	Beta coefficient	Odds ratio	95% CI	P value
Model intercept	-3.73776	0.02	0.01 to 0.10	<0.001
Age, per year	0.02399	1.02	1.01 to 1.04	0.006
Sex, female	-0.54489	0.58	0.49 to 0.69	<0.001
White race, reference	N/A	N/A	N/A	N/A
Black race, yes	0.39969	1.49	1.08 to 2.01	0.011
Asian race, yes	-1.24615	0.29	0.11 to 0.60	0.003
Hispanic race, yes	-0.36998	0.69	0.29 to 1.37	0.341
Mixed or other race, yes	0.34995	1.42	0.88 to 2.16	0.123
Body mass index, per kg/m^2*	3.35128	28.54	5.52 to 143.16	<0.001
Current smoker, yes	0.27488	1.32	1.11 to 1.56	0.002
Smoking duration, per year*	-0.04392	0.96	0.92 to 0.99	0.016
Hypertension diagnosis, yes	0.12417	1.13	0.97 to 1.32	0.118
Diabetes diagnosis, yes	0.45764	1.58	1.29 to 1.93	<0.001
Heart disease diagnosis, yes	0.20551	1.23	1.01 to 1.48	0.034
Stroke diagnosis, yes	0.55247	1.74	1.28 to 2.32	<0.001
Asthma diagnosis, yes	0.33797	1.40	1.11 to 1.75	0.004
COPD diagnosis, yes	0.25723	1.29	1.08 to 1.54	0.004
Emphysema score, per point	0.05331	1.05	1.04 to 1.07	<0.001
Bronchial wall thickness (Pi10), per point	0.16154	1.18	1.06 to 1.30	0.001
Mean lung density, per HU*	-1.83652	0.16	0.08 to 0.33	<0.001
Aorta calcium volume, per mm^3*	0.18081	1.20	1.14 to 1.26	<0.001
Aorta calcium mean density, per HU*	-1.53207	0.22	0.09 to 0.50	0.001
Coronary calcium volume, per mm^{3*}	0.14995	1.16	1.05 to 1.28	0.003
Mitral valve calcium volume, per mm^3*	0.06940	1.07	1.05 to 1.10	<0.001

* Transformed as follows: (body mass index/10)$^{-2}$; (smoking duration/100)$^{-2}$; ((mean lung density + 1000)/100)$^{-1}$; ln((aorta calcium volume + 0.1)/1000); (aorta calcium mean density/1000); (coronary calcium volume/1000); ln((mitral valve calcium volume + 0.1)/100).

To calculate the five-year risk probability of lung cancer incidence, first find the sum of the products of each (transformed) variable and their respective beta coefficient to obtain the linear predictor, then insert the value into the following equation: $\frac{1}{1+e^{-(linear\ predictor)}}$

CI, confidence intervals.

Table 3: Clinical outcomes of risk stratification

Risk models used	Risk group	Participants (%)	5-year LC diagnoses (%)	5-year CDs (%)	Number needed to screen to detect 1 LC	CDs per LC diagnosis
Derivation cohort (NLST)						
LCi_{final} & CD_{final}	High-LC-low-CD	16838 (73)	727 (96)	643 (80)	23	0.88
	Low-LC-high-CD	6258 (27)	29 (4)	157 (20)	216	5.41
LCRAT & CDRAT	High-LC-low-CD	17335 (75)	677 (90)	668 (84)	26	0.99
	Low-LC-high-CD	5761 (25)	79 (10)	132 (16)	73	1.67
$PLCO_{M2012}$ & CDRAT	High-LC-low-CD	17277 (75)	686 (91)	675 (84)	25	0.99
	Low-LC-high-CD	5819 (25)	70 (11)	125 (16)	83	1.79
	Full derivation cohort	*23096 (100)*	*756 (100)*	*800 (100)*	*31*	*1.06*
Validation cohort (MILD)						
LCi_{final} & CD_{final}	High-LC-low-CD	1513 (66)	53 (90)	23 (70)	29	0.43
	Low-LC-high-CD	774 (34)	6 (10)	10 (30)	129	1.67
LCRAT & CDRAT	High-LC-low-CD	1511 (66)	48 (81)	22 (67)	31	0.46
	Low-LC-high-CD	776 (34)	11 (19)	11 (33)	71	1.00
$PLCO_{M2012}$ & CDRAT	High-LC-low-CD	1568 (69)	51 (86)	28 (85)	31	0.55
	Low-LC-high-CD	719 (31)	8 (14)	5 (15)	90	0.63
	Full validation cohort	*2287 (100)*	*59 (100)*	*33 (100)*	*39*	*0.56*
NLST-eligible validation cohort (MILD)						
LCi_{final} & CD_{final}	High-LC-low-CD	980 (80)	44 (94)	18 (72)	22	0.41
	Low-LC-high-CD	245 (20)	3 (6)	7 (28)	82	2.33
LCRAT & CDRAT	High-LC-low-CD	995 (81)	39 (83)	21 (84)	26	0.54
	Low-LC-high-CD	230 (19)	8 (17)	4 (16)	29	0.50
$PLCO_{M2012}$ & CDRAT	High-LC-low-CD	1088 (89)	43 (91)	25 (100)	25	0.58
	Low-LC-high-CD	137 (11)	4 (9)	0 (0)	34	0.00
	NLST-eligible validation cohort	*1225 (100)*	*47 (100)*	*25 (100)*	*26*	*0.53*

CD, competing death; CD_{final}, final competing death model; CDRAT, Competing Death Risk Assessment Tool (10); High-LC-low-CD, risk group considered to have a relatively high LC risk and low risk of competing death (criteria described in the Figure 1 caption); LC, lung cancer; LCi_{final}, final lung cancer incidence model; LCRAT, Lung Cancer Risk Assessment Tool; Low-LC-high-CD, risk group considered to have a relatively low LC risk and high risk of competing death; MILD, Multicentric Italian Lung Detection; NLST, National Lung Screening Trial, $PLCO_{M2012}$, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Model 2012.

Supplement

Contents

Methods 2

Data selection (NLST) 2

Validation data set preparation (MILD) 2

Quantitative CT measures 3

Data set formation 3

Equations S1: Final lung cancer incidence model equation (LCi_{final}) 4

Table S1: Self-reported demographics-based competing mortality risk model (CD_{survey}) 5

Table S2: CT-based competing mortality risk model (CD_{CT}) 6

Table S3: Internal validation competing death risk models 7

Table S4: Contingency table of outcomes per risk quintile (derivation cohort) 8

Table S5: Contingency table of outcomes per risk quintile (validation cohort) 9

Figure S1: Cohort formation flowchart 11

Figure S2: Receiver operating characteristic curves in the derivation cohort 12

Figure S3: Receiver operating characteristic curves in the validation cohort 13

Figure S4: Calibration plot of CD_{final} on the derivation cohort 14

Figure S5: Calibration plot of CD_{final} on the validation cohort 15

Figure S6: Outcomes per lung cancer and competing death risk quintile (validation cohort) 18

References 19

Methods

Data selection (NLST)

We obtained baseline patient characteristics, nodule features, follow-up outcomes, and chest CT images from National Lung Screening Trial (NLST) participants (1). The data was cleaned (missing data was given blank values) and new patient characteristics variables were created, i.e., body mass index (BMI), time since smoking quit, number of first-degree family members diagnosed with lung cancer, diagnosis of any cancer prior to trial, the follow-up time from randomization to event or date last known alive, and disease-specific causes of death (as determined by the endpoint verification process [EVP] and death certificate). Lung cancer mortality, lung cancer incidence, and other causes of death were provided.

In the NLST data set, the variables race, ethnicity and education were separated into the following categories: Race was divided into White, Black or African-American, Asian, American Indian or Alaskan native, Native Hawaiian or other Pacific Islander, and more than one race; ethnicity was binary: Hispanic or Latino, or neither; educational level completed ranged from 8^{th} grade or less, 9^{th} to 11^{th} grade, high school graduate or General Educational Development, post high school training (excluding college), associate degree or some college, bachelor's degree, graduate school, and other. As most other studies and models did not record race and ethnicity as separate variables, we merged ethnicity with race by allocating all participants who had both a Hispanic or Latino ethnicity and White race into a new race category (Hispanic or Latino). Furthermore, due to the low prevalence (<1%) of the race groups "American Indian or Alaskan native" and "Native Hawaiian or other Pacific Islander" and the ambiguity of the "mixed" race group, the three were merged to form the "mixed or other" group. To make the education variable more applicable to other educational systems and due to their low frequencies, "8^{th} grade or less" and "9^{th} to 11^{th} grade" were merged as "did not complete high school," and "graduate school" and "other" were merged to "graduate school or higher." Furthermore, education is a categorical variable but was applied as a continuous variable similar to in the $PLCO_{m2012}$ (2); "11^{th} grade or less" was set zero, plus one per increase in level up to five.

Age (years) and sex (male vs female) were considered for all models. Detailed patient characteristics (which would normally only be obtained via a survey) were race or ethnicity (White, Black or African-American, Asian, Hispanic or Latino, or mixed or other), educational level (range: 0 to 5), BMI (kg/m^2), smoking status (current vs former), smoking intensity (pack years), smoking duration (years), smoking quit time (years, 0 if current smoker), number of first-degree family members diagnosed with lung cancer (range: 0 to 2 [a value of "2" was given when two or more family members were diagnosed due to a very low prevalence of those with three or more]), exposure to asbestos at work for at least one year (yes vs no), and the prior diagnosis of chronic obstructive pulmonary disease (COPD), asthma (child or adult), diabetes, heart disease, hypertension, and stroke (yes vs no). The prior diagnosis of COPD was also considered positive if the subject had a prior diagnosis of chronic bronchitis and/or emphysema. The NLST listed other diseases in their survey, but these were not available in the MILD cohort and were therefore not considered for modelling.

The list of prospectively detected pulmonary nodules as reported by radiologists who participated in the NLST was used. The following nodule features were considered in our study: longest diameter (mm), longest perpendicular diameter (mm), attenuation (solid [soft tissue], nonsolid [ground glass], or partsolid [mixed]), upper lobe location (yes vs no), spiculation (yes vs no), and nodule count. If more than one nodule was recorded, the features of the nodule with the longest diameter was used; subjects who did not have a nodule were given a null value for all nodule features. Note that the NLST only reported non-calcified nodules of at least 4 mm in longest diameter. Nodules reported to have a longest diameter of 20 mm or greater were visually inspected and corrected if necessary.

Validation data set preparation (MILD)

We obtained baseline patient characteristics, nodule features, follow-up outcomes, and chest CT images from Multicentric Italian Lung Detection (MILD) trial participants. The data was cleaned (missing data was given blank values) and the variables were created and transformed to be interchangeable with that of the NLST. Educational level was available in five levels: elementary school graduate, middle school graduate, high school graduate, university attendee, business school graduate, and bachelor's degree. "elementary school graduate" and "middle school graduate" were classified as "did not complete high school," "university attendee" was classified as "post high school training (excluding college)," and "business school graduate" was classified as "associate degree or some

college." The "first-degree family members diagnosed with lung cancer" variable only mentions whether there was at least one or none; the exact number is not given. All other patient characteristics variables selected from the NLST were available.

We utilized the MILD data set of lung nodules retrospectively detected using computer-aided diagnosis, which does not correspond to the data set of nodules retrospectively detected by computer-aided diagnosis (3,4). Only a maximum of five nodules were reported per scan per time point; nodules only counted in the nodule count if larger than or equal to 20 mm^3 in volume. The same nodule features as from the NLST were considered except for spiculation, which was not available and therefore given a null value when applicable. The longest perpendicular diameter was considered the longest diameter in one of the other two planes which was not the plane which the longest diameter was measured.

Quantitative CT measures

The CT images were used to obtain quantitative CT measures (QCT) of CVD – calcium volume and mean density of the coronary arteries, mitral valve, aortic valve, and transthoracic aorta (5,6) – and chronic obstructive pulmonary disease (COPD) – lung volume, mean lung density, normalized emphysema score (the percentage of lung voxels below -950 HU after resampling the CT images to 3mm slice thickness, normalization, and bullae analysis) (7), and bronchial wall thickness (Pi10, the square root of the airway wall area for a theoretical 10mm lumen perimeter airway derived using the linear regression of the square root of segmented wall area against the lumen perimeter)(8). As quality control, cases with extreme outlier values were excluded, i.e., mean lung density > -300 HU, mean lung density < -1000 HU, Pi10 < 0.8, and Pi10 > 6.5.

Data set formation

The primary NLST subject inclusion criterion was the availability of a baseline CT image of slice thickness ≤2.5 mm. Participants with missing data on lung cancer incidence, death status, time of event, nodule features, QCTs of CVD, or QCTs of COPD were excluded from the NLST cohort (Figure 1). All remaining participants who were diagnosed with lung cancer, all who died within the study period, and a random sample of all other participants from the CT screening cohort up to a maximum of 15000 unique subjects were included in the NLST cohort. This was due to the limit of NLST CT images used for one project set by the Cancer Data Access System, project ID "NLST-437. Subsequently, the proportion of non-deceased participants without a lung cancer diagnosed were resampled without replacement to simulate the full NLST cohort. This was calculated by taking the proportion of deceased or participants diagnosed with lung cancer included in our study out of those in the CT arm with a baseline scan, then applying that proportion to the non-deceased participants without a lung cancer diagnosis in the CT arm with a baseline scan to obtain the total number of non-deceased participants without a lung cancer diagnosis who should be included after resampling. This was to maintain the original probabilities of events which occurred in the NLST, in turn preventing the models from overestimating the risk. Resampling was not necessary for the validation cohort as there were no limitations on CT image usage. Furthermore, almost all CT images from MILD were available in 1 mm slice thickness (2271/2287, 99.3%).

Mean lung density was centered to –1000 HU (i.e., 1000 was added to the actual value of the mean lung density) to circumvent modelling issues with negative values.

Multiple imputation using the 'mice' function (R package 'mice', 'cart' method) was performed to impute missing data using classification and regression trees. Of the 15000 NLST subjects included in the NLST data set (before resampling), there was missing data from race (n=13), education (n=21), BMI (n=41), first-degree family history of lung cancer (n=623), exposure to asbestos at work (n=20), and the prior diagnoses of COPD (n=43), asthma (n=15), diabetes (n=14), heart disease (n=33), hypertension (n=16), and stroke (n=14).

2303 subjects were considered for the validation cohort. Of these, 8/2303 (0.3%) were not part of the first screening round and 9/2303 (0.4%) were missing baseline scans. Of the 2287 subjects included, some QCTs of CVD (n=24) and Pi10 values (n=132) could not be extracted from the scans; in contrast to excluding these cases as was done with the NLST cohort, the missing values were replaced with the corresponding median values from the MILD cohort set to avoid excluding cases.

Equations S1: Final lung cancer incidence model equation (LCi_{final})

$$Linear\ predictor$$
$$= (female \times 0.12611) + \left(\frac{BMI}{10} \times (-0.19128)\right) + \left(\left(\frac{smoking\ intensity}{100}\right)^{-2} \times (-0.10613)\right)$$
$$+ \left(\frac{smoking\ duration}{10} \times 0.25895\right) + \left(\frac{smoking\ quit\ time}{10} \times (-0.30764)\right)$$
$$+ (lung\ cancer\ in\ family \times 0.18273) + (\ln(emphysema\ score + 0.1) \times 0.18683)$$
$$+ \left(\left(\frac{mean\ lung\ density + 1000}{100}\right)^{-1} \times (-0.74125)\right) + (Pi10 \times 0.09703)$$
$$+ \left(\ln\left(\frac{aorta\ calcium\ volume + 0.1}{100} \times 0.17668\right)\right)$$
$$+ \left(\left(\frac{aorta\ calcium\ mean\ density + 0.1}{100}\right)^{0.5} \times (-0.57768)\right)$$
$$+ \left(\left(\frac{nodule\ longest\ diameter + 1}{10}\right)^{-2} \times (-0.44997)\right)$$
$$+ \left(\left(\frac{nodule\ longest\ diameter + 1}{10}\right)^{-2} \times ln\left(\frac{nodule\ longest\ diameter + 1}{10}\right) \times (-0.22010)\right)$$
$$+ (\ln(nodule\ perpendicular\ diameter + 1) \times 1.20812) + (nodule\ in\ upper\ lobe \times 0.22769)$$
$$+ (spiculated\ nodule \times 0.64707)$$

$$NLST\ baseline\ hazard\ function\ estimate = H_0(t)_{lung\ cancer\ incidence}$$
$$= 0.0018772 + 0.021448 \times \frac{t + 1}{1000} + (-0.0059062) \times \frac{(t + 1)(\ln(t + 1))}{1000}$$

(where t is the follow-up time in days; for 5 years follow-up, this equals 1826)

$$Cumulative\ survival\ probability = S(t)$$
$$= \exp\left(-exp(linear\ predictor - \ NLST\ mean\ linear\ predictor\ ) \times H_0(t)_{lung\ cancer\ incidence}\right)$$

(where NLST mean linear predictor = 0.10129)

$$Five - year\ lung\ cancer\ incidence\ risk\ calibrated\ to\ the\ National\ Lung\ Screening\ Trial$$
$$= S(t) \times 0.60343 + 0.39822$$

$$Five - year\ lung\ cancer\ incidence\ risk\ calibrated\ to\ the\ Multicentric\ Italian\ Lung\ Detection\ trial$$
$$= S(t) \times 0.85413 + 0.14396$$

Table S1: Self-reported demographics-based competing mortality risk model (CD_{survey})

Variable	Beta coefficient	Odds ratio	95% CI	P value
Model intercept	-7.25928	0.00	0.00 to 0.00	<0.001
Age, per year	0.05829	1.06	1.04 to1.08	<0.001
Sex, female	-0.58787	0.56	0.47 to 0.65	<0.001
White race, reference	N/A	N/A	N/A	N/A
Black race, yes	0.22484	1.25	0.917 to 1.68	0.143
Asian race, yes	-1.41522	0.24	0.10 to 0.50	0.001
Hispanic race, yes	-0.50493	0.60	0.26 to 1.19	0.191
Mixed or other race, yes	0.25358	1.29	0.81 to 1.96	0.261
Body mass index, per kg/m²*	2.30426	10.02	2.12 to 46.00	0.003
Current smoker, yes	0.36904	1.45	1.22 to 1.71	<0.001
Smoking duration, per year*	-0.06592	0.94	0.90 to 0.97	<0.001
Hypertension diagnosis, yes	0.24762	1.28	1.10 to 1.49	0.001
Diabetes diagnosis, yes	0.59631	1.82	1.48 to 2.21	<0.001
Heart disease diagnosis, yes	0.48109	1.62	1.48 to 2.21	<0.001
Stroke diagnosis, yes	0.65990	1.93	1.43 to 2.58	<0.001
Asthma diagnosis, yes	0.32387	1.38	1.10 to 1.72	0.005
COPD diagnosis, yes	0.37986	1.46	1.23 to 1.73	<0.001

* Transformed as follows: $(\text{body mass index}/10)^{-2}$; $(\text{smoking duration}/100)^{-2}$

To calculate the five-year risk probability of lung cancer incidence, first find the sum of the products of each (transformed) variable and their respective beta coefficient to obtain the linear predictor, then insert the value into the following equation: $\frac{1}{1+e^{-(linear\ predictor)}}$

CI, confidence intervals.

Table S2: CT-based competing mortality risk model (CD_{CT})

Variable	Beta coefficient	Odds ratio	95% CI	P value
Model intercept	-3.73187	0.02	0.01 to 0.08	<0.001
Age, per year	0.03097	1.03	1.02 to 1.05	<0.001
Sex, female	-0.47135	0.62	0.53 to 0.74	<0.001
Emphysema score, per point	0.06217	1.06	1.05 to 1.08	<0.001
Bronchial wall thickness (Pi10), per point	0.20689	1.23	1.12 to 1.35	<0.001
Mean lung density, per HU*	-2.02753	0.13	0.07 to 0.26	<0.001
Aorta calcium volume, per mm^3*	0.21075	1.23	1.17 to 1.30	<0.001
Aorta calcium mean density, per HU*	-1.52263	0.22	0.09 to 0.51	0.001
Coronary calcium volume, per mm^{3*}	0.19567	1.22	1.11 to 1.33	<0.001
Mitral valve calcium volume, per mm^3*	0.25931	1.30	1.06 to 1.56	0.009
Mitral valve calcium mean density, per HU*	0.04676	1.05	1.02 to 1.07	<0.001
Solid nodule (largest nodule), present	-0.22020	0.80	0.67 to 0.96	0.0183

* Transformed as follows: $((\text{mean lung density} + 1000)/100)^{-1}$; ln((aorta calcium volume + 0.1)/1000); (aorta calcium mean density/1000); (coronary calcium volume/1000); (mitral valve calcium volume/1000); ln((mitral valve calcium mean density+0.1)/100).

To calculate the five-year risk probability of lung cancer incidence, first find the sum of the products of each (transformed) variable and their respective beta coefficient to obtain the linear predictor, then insert the value into the following equation: $\frac{1}{1+e^{-(linear\ predictor)}}$

CI, confidence intervals.

Table S3: Internal validation competing death risk models

Model	Original sample AUC (%) (95% confidence interval)	Bootstrap sample AUC (%) (95% confidence interval)	Optimism
	5-year	5-year	5-year
CD_{Survey}	0.704 (0.704 to 0.704)	0.710 (0.709 to 0.710)	0.006
CD_{CT}	0.717 (0.717 to 0.717)	0.720 (0.720 to 0.721)	0.003
CD_{final}	0.740 (0.740 to 0.741)	0.746 (0.746 to 0.747)	0.006

One thousand bootstrap replications were performed for each model and time point. The optimism is the difference between the average bootstrap sample AUC and original sample AUC; a greater value indicates more overfitting.

AUC, receiver operating characteristic area under the curve; CD, competing death.

Table S4: Contingency table of outcomes per risk quintile (derivation cohort)

			Lung cancer incidence risk (LCi_{final})				
			Q1: ≤0.79%	Q2: 0.79%-1.38%	Q3: 1.38-2.18%	Q4: 2.18-3.85%	Q5: >3.85%
Total participant count (n=23096)	Competing death risk (CD_{final})	Q1: ≤1.11%	2578 (12)	1129 (5)	515 (2)	234 (1)	140 (1)
		Q2: 1.11-1.90%	1128 (5)	1467 (7)	1058 (5)	623 (3)	281 (1)
		Q3: 1.90%-2.93%	555 (3)	1071 (5)	1237 (6)	1064 (5)	507 (2)
		Q4: 2.93-4.92%	266 (1)	629 (3)	1124 (5)	1407 (6)	888 (4)
		Q5: >4.92%	92 (0)	322 (1)	686 (3)	1291 (6)	1756 (8)
5-year lung cancer diagnoses (n=756)		Q1: ≤1.11%	4 (1)	9 (1)	4 (1)	6 (1)	23 (3)
		Q2: 1.11-1.90%	0 (0)	8 (1)	9 (1)	11 (1)	48 (6)
		Q3: 1.90%-2.93%	3 (0)	9 (1)	18 (2)	33 (4)	80 (11)
		Q4: 2.93-4.92%	4 (1)	5 (1)	21 (3)	46 (6)	121 (16)
		Q5: >4.92%	1 (0)	3 (0)	9 (1)	40 (5)	241 (32)
5-year competing deaths (n=800)		Q1: ≤1.11%	20 (3)	3 (0)	7 (1)	1 (0)	1 (0)
		Q2: 1.11-1.90%	21 (3)	20 (3)	16 (2)	10 (1)	13 (2)
		Q3: 1.90%-2.93%	10 (1)	23 (3)	29 (4)	25 (3)	10 (1)
		Q4: 2.93-4.92%	6 (1)	17 (2)	42 (5)	58 (7)	43 (5)
		Q5: >4.92%	7 (1)	28 (4)	48 (6)	113 (14)	229 (29)
Competing deaths per lung cancer diagnosis (mean=1.06)		Q1: ≤1.11%	5.00	0.33	1.75	0.17	0.04
		Q2: 1.11-1.90%	21/0	2.50	1.78	0.91	0.27
		Q3: 1.90%-2.93%	3.33	2.56	1.61	0.76	0.13
		Q4: 2.93-4.92%	1.50	3.40	2.00	1.26	0.36
		Q5: >4.92%	7.00	9.33	5.33	2.83	0.95

The cells are shaded grey according to the proportion of outcomes which occurred in that cell, where a darker shade indicates a greater proportion. Brackets indicate the percentage of the total count where applicable. The double lines indicate the suggested separation of screening participants into a group which should continue to be screened (right of the double line) and a group with a relatively low lung cancer risk and high risk of competing death which is unlikely to benefit from screening (left of the double line).

CD_{final}, final competing death model; LCi_{final}, final lung cancer incidence model (9); n, total count; Q, quintile (e.g., Q1 = 0^{th} to 20^{th} percentile).

Table S5: Contingency table of outcomes per risk quintile (validation cohort)

			Lung cancer incidence risk (LCi_{final})				
			Q1: ≤0.79%	Q2: 0.79-1.38%	Q3: 1.38-2.18%	Q4: 2.18-3.85%	Q5: >3.85%
Total participant count (n=23096)	Competing death risk (CD_{final})	Q1: ≤1.11%	441 (19)	231 (10)	110 (5)	48 (2)	34 (1)
		Q2: 1.11-1.90%	163 (7)	171 (7)	126 (6)	108 (5)	68 (3)
		Q3: 1.90-2.93%	59 (3)	63 (3)	88 (4)	83 (4)	69 (3)
		Q4: 2.93-4.92%	38 (2)	33 (1)	46 (2)	78 (3)	80 (3)
		Q5: >4.92%	4 (0)	17 (1)	19 (1)	42 (2)	68 (3)
5-year lung cancer diagnoses (n=756)		Q1: ≤1.11%	1 (2)	0 (0)	0 (0)	0 (0)	1 (2)
		Q2: 1.11-1.90%	1 (2)	3 (5)	3 (5)	3 (5)	8 (14)
		Q3: 1.90-2.93%	0 (0)	2 (3)	1 (2)	1 (2)	2 (3)
		Q4: 2.93-4.92%	0 (0)	2 (3)	0 (0)	3 (5)	11 (19)
		Q5: >4.92%	0 (0)	0 (0)	2 (3)	4 (7)	11 (19)
5-year competing deaths (n=800)		Q1: ≤1.11%	2 (6)	1 (3)	1 (3)	0 (0)	0 (0)
		Q2: 1.11-1.90%	1 (3)	1 (3)	0 (0)	1 (3)	2 (6)
		Q3: 1.90-2.93%	0 (0)	2 (6)	1 (3)	2 (6)	0 (0)
		Q4: 2.93-4.92%	2 (6)	0 (0)	0 (0)	6 (18)	2 (6)
		Q5: >4.92%	1 (3)	1 (3)	3 (9)	1 (3)	3 (9)
Competing deaths per lung cancer diagnosis (mean=1.06)		Q1: ≤1.11%	2.00	1/0	1/0	0/0	0.00
		Q2: 1.11-1.90%	1.00	0.33	0.00	0.33	0.25
		Q3: 1.90-2.93%	0/0	1.00	1.00	2.00	0.00
		Q4: 2.93-4.92%	2/0	0.00	0/0	2.00	0.18
		Q5: >4.92%	1/0	1/0	1.50	0.25	0.27

The cells are shaded grey according to the proportion of outcomes which occurred in that cell, where a darker shade indicates a greater proportion. Brackets indicate the percentage of the total count where applicable. The double lines indicate the suggested separation of screening participants into a group which should continue to be screened (right of the double line) and a group with a relatively low lung cancer risk and high risk of competing death which is unlikely to benefit from screening (left of the double line).

CD_{final}, final competing death model; LCi_{final}, final lung cancer incidence model (9); n, total count; Q, quintile (e.g., Q1 = 0^{th} to 20^{th} percentile).

Table S6: Clinical outcomes of risk stratification

Risk models used	Risk group	Participants (%)	5-year LC diagnoses (%)	5-year CDs (%)	Number needed to screen to detect 1 LC	CDs per LC diagnosis
Derivation cohort (NLST)						
LCi_{final} & CD_{final}	High-LC-low-CD	16838 (73)	727 (96)	643 (80)	23	0.88
	Low-LC-high-CD	6258 (27)	29 (4)	157 (20)	216	5.41
LCi_{final} & CD_{CD}	High-LC-low-CD	16537 (72)	723 (96)	639 (80)	23	0.88
	Low-LC-high-CD	6559 (28)	33 (4)	161 (20)	199	4.90
LCi_{final} & CD_{survey}	High-LC-low-CD	16530 (72)	725 (96)	643 (80)	23	0.89
	Low-LC-high-CD	6566 (28)	31 (4)	157 (20)	212	5.06
LCi_{final} & CDRAT	High-LC-low-CD	16638 (72)	720 (95)	671 (84)	23	0.93
	Low-LC-high-CD	6458 (28)	36 (5)	129 (16)	179	3.58
	Full derivation cohort	*23096 (100)*	*756 (100)*	*800 (100)*	*31*	*1.06*
Validation cohort (MILD)						
LCi_{final} & CD_{final}	High-LC-low-CD	1513 (66)	53 (90)	23 (70)	29	0.43
	Low-LC-high-CD	774 (34)	6 (10)	10 (30)	129	1.67
LCi_{final} & CD_{CD}	High-LC-low-CD	1538 (67)	54 (92)	23 (70)	28	0.43
	Low-LC-high-CD	749 (33)	5 (8)	10 (30)	150	2.00
LCi_{final} & CD_{survey}	High-LC-low-CD	1414 (62)	52 (88)	21 (64)	27	0.40
	Low-LC-high-CD	873 (38)	7 (12)	12 (36)	125	1.71
LCi_{final} & CDRAT	High-LC-low-CD	1440 (63)	51 (86)	22 (67)	28	0.43
	Low-LC-high-CD	847 (37)	8 (14)	11 (33)	106	1.38
	Full validation cohort	*2287 (100)*	*59 (100)*	*33 (100)*	*39*	*0.56*
NLST-eligible validation cohort (MILD)						
LCi_{final} & CD_{final}	High-LC-low-CD	980 (80)	44 (94)	18 (72)	22	0.41
	Low-LC-high-CD	245 (20)	3 (6)	7 (28)	82	2.33
LCi_{final} & CD_{CD}	High-LC-low-CD	1000 (82)	43 (91)	19 (76)	23	0.44
	Low-LC-high-CD	225 (18)	4 (9)	6 (24)	56	1.50
LCi_{final} & CD_{survey}	High-LC-low-CD	902 (74)	43 (91)	17 (68)	21	0.40
	Low-LC-high-CD	323 (26)	4 (9)	8 (32)	81	2.00
LCi_{final} & CDRAT	High-LC-low-CD	914 (75)	42 (89)	18 (72)	22	0.43
	Low-LC-high-CD	311 (25)	5 (11)	7 (28)	62	1.40
	NLST-eligible validation cohort	*1225 (100)*	*47 (100)*	*25 (100)*	*26*	*0.53*

CD, competing death; CD_{CT}, competing death CT model; CD_{final}, final competing death model; CD_{survey}, competing death survey model; CDRAT, Competing Death Risk Assessment Tool (10); High-LC-low-CD, risk group considered to have a relatively high LC risk and low risk of competing death (criteria described in the Figure 1 caption); LC, lung cancer; LCi_{final}, final lung cancer incidence model; Low-LC-high-CD, risk group considered to have a relatively low LC risk and high risk of competing death; MILD, Multicentric Italian Lung Detection; NLST, National Lung Screening Trial.

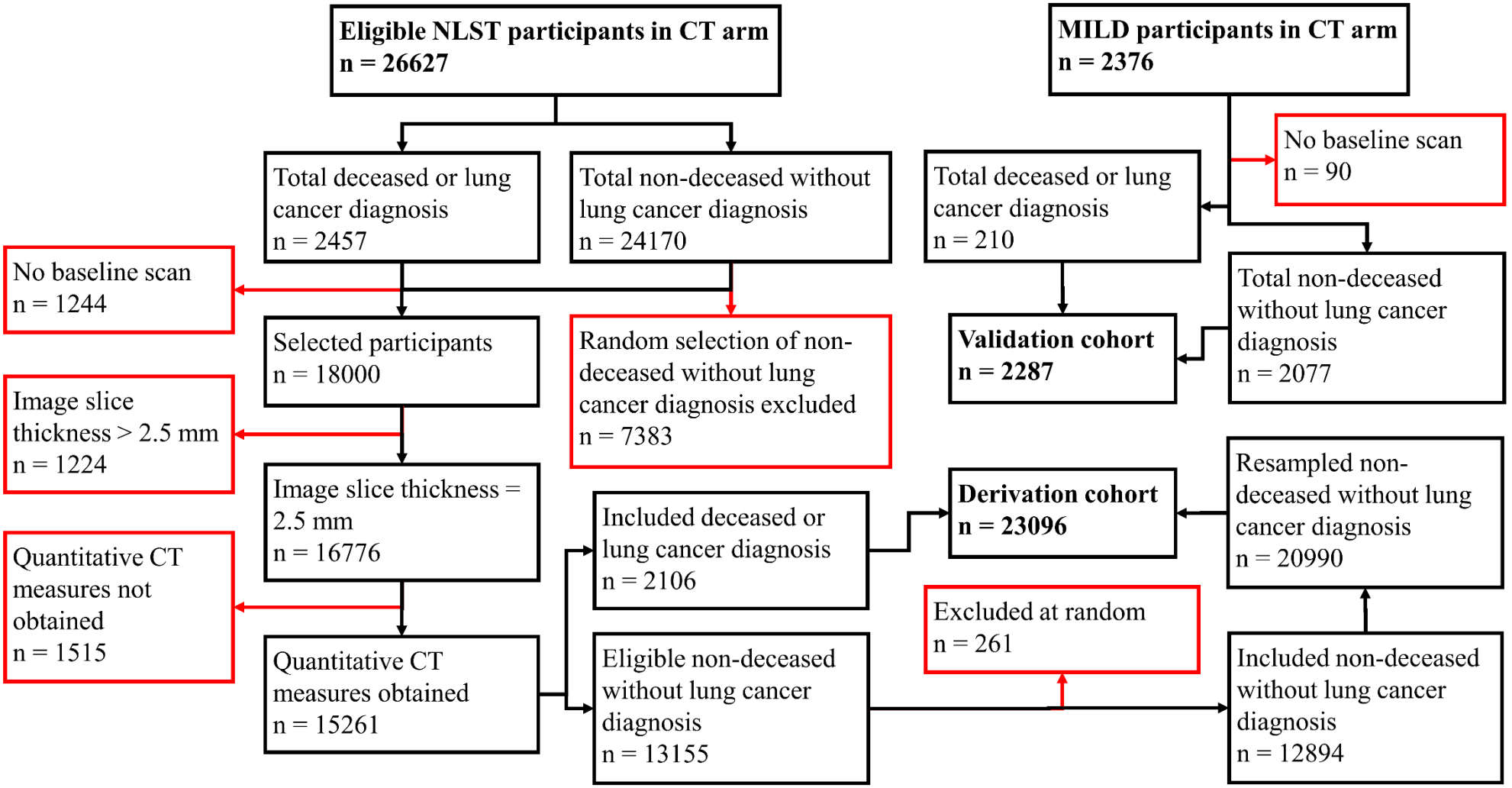


Figure S1: Cohort formation flowchart

For the derivation cohort, permission to use NLST was limited to up to 15000 participants. For this reason, all eligible participants who died or were diagnosed with lung cancer in the trial were included (n=2106) along with a random sample of eligible non-deceased participants without lung cancer (n=12894). To simulate the original CT cohort, the latter group was resampled without replacement to a total of 20990 non-deceased participants without lung cancer. In total, the derivation cohort consisted of 23096 participants. For the validation cohort (MILD), all eligible participants were included. MILD, Multicentric Italian Lung Detection; NLST, National Lung Screening Trial.

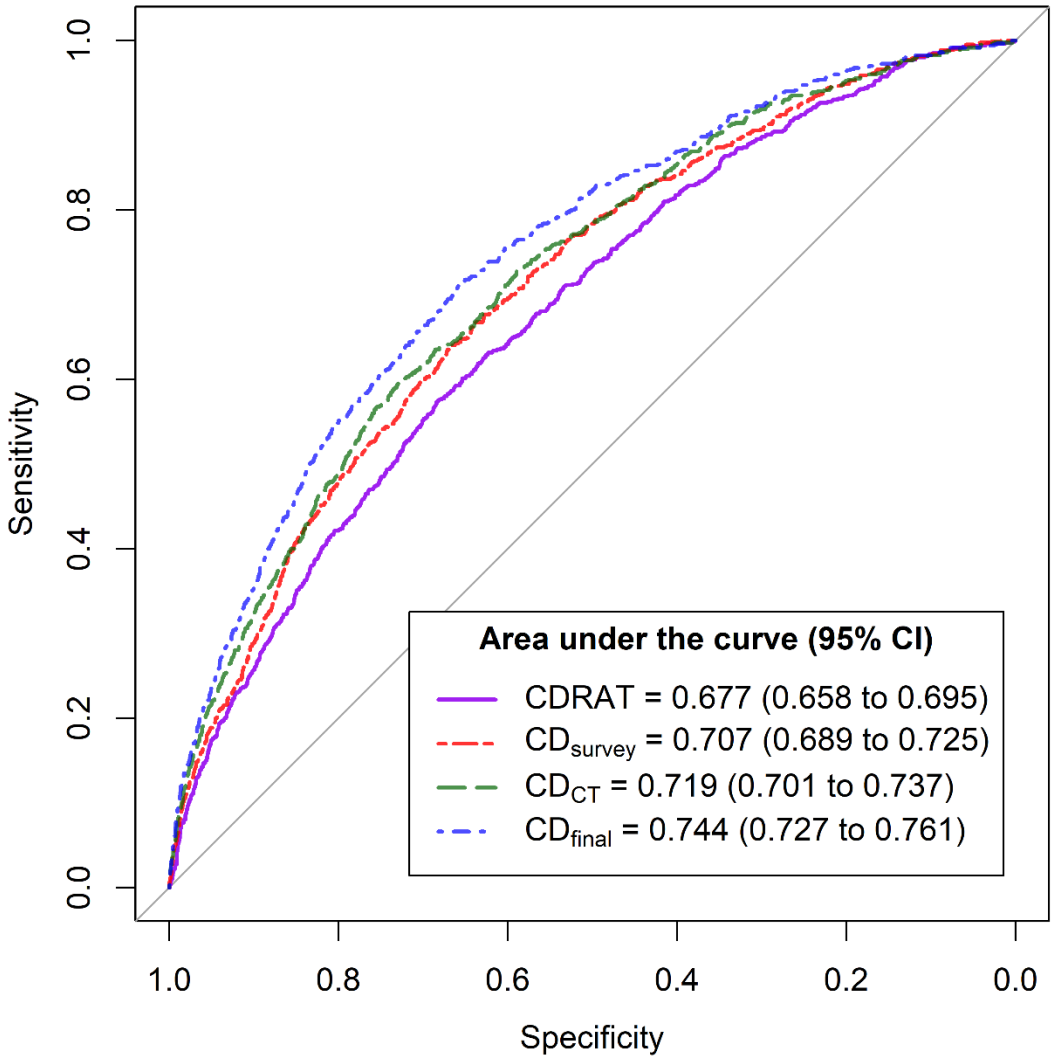


Figure S2: Receiver operating characteristic curves in the derivation cohort

CD_{CT} = CT-based competing death model; CD_{final}, final competing death model; CD_{survey}, self-reported patient characteristics-based competing death model; CDRAT, Competing Death Risk Assessment Tool; CI, confidence interval.

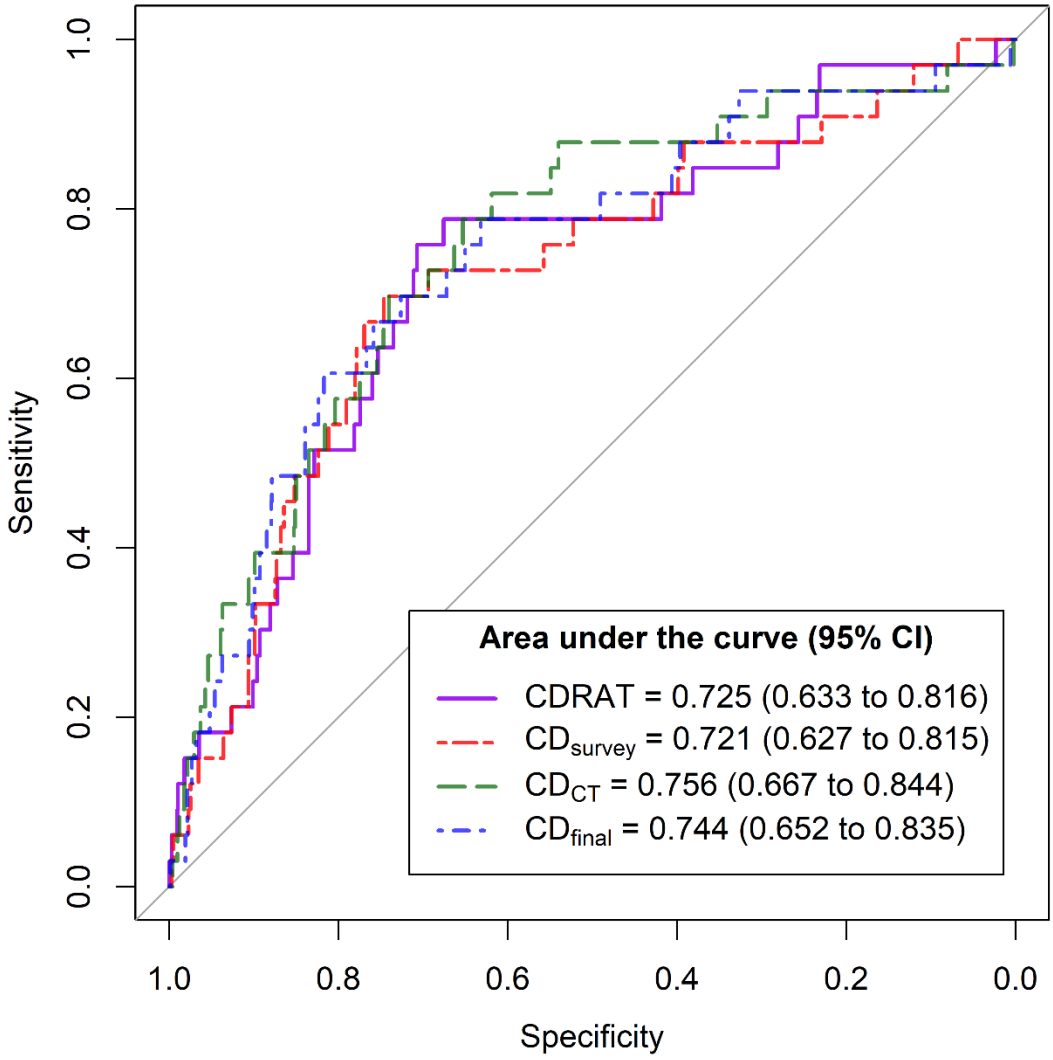


Figure S3: Receiver operating characteristic curves in the validation cohort

CD_{CT} = CT-based competing death model; CD_{final}, final competing death model; CD_{survey}, self-reported patient characteristics-based competing death model; CDRAT, Competing Death Risk Assessment Tool; CI, confidence interval.

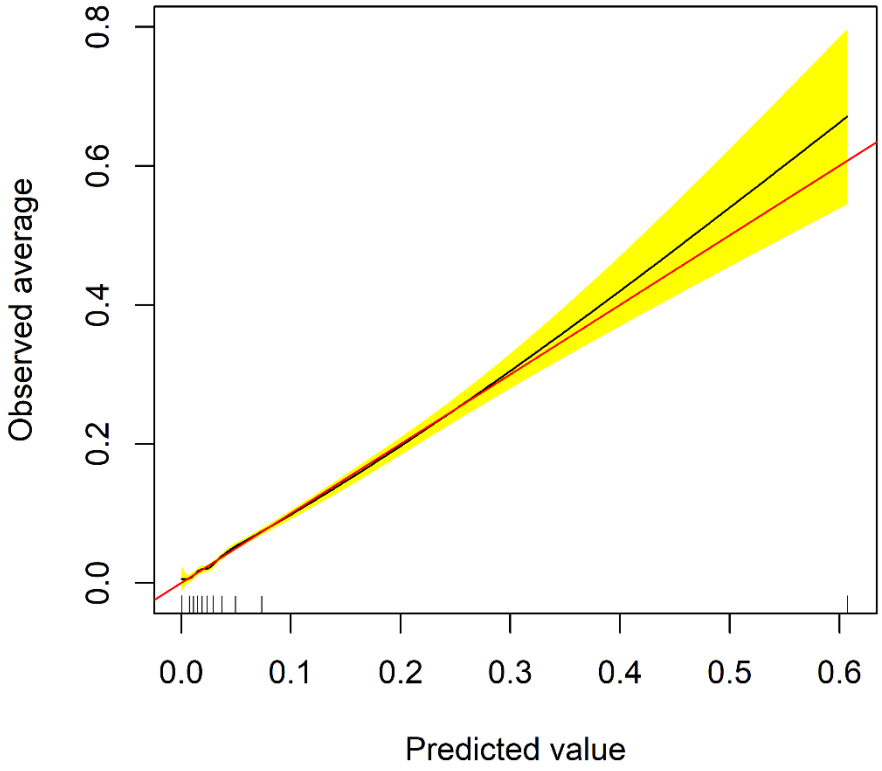


Figure S4: Calibration plot of CD_{final} in the derivation cohort

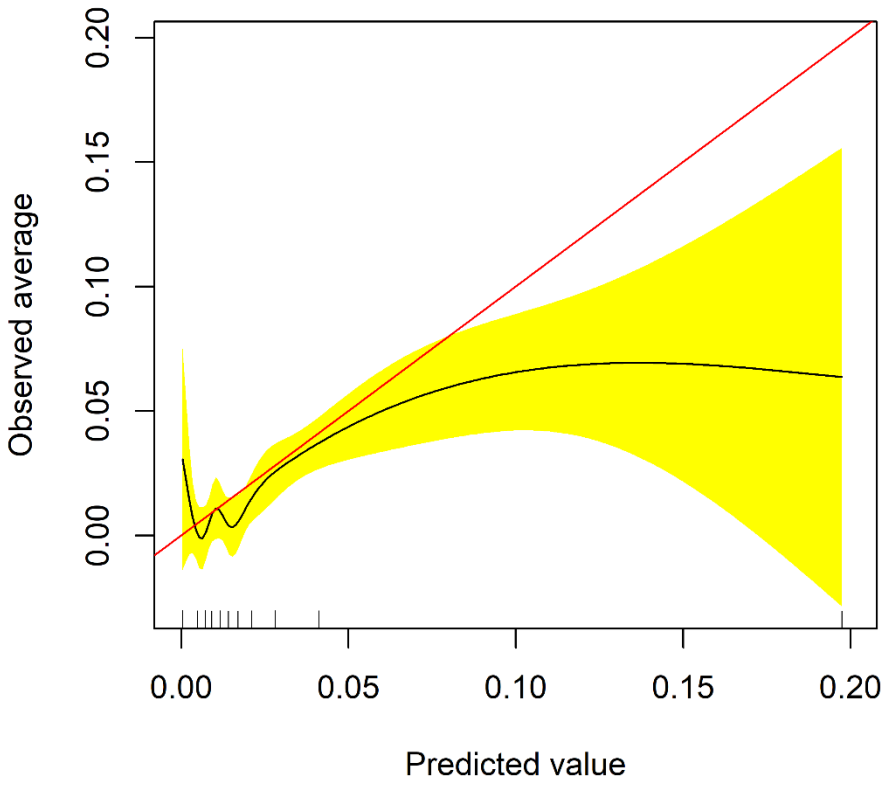


Figure S5: Calibration plot of CD_{final} in the validation cohort

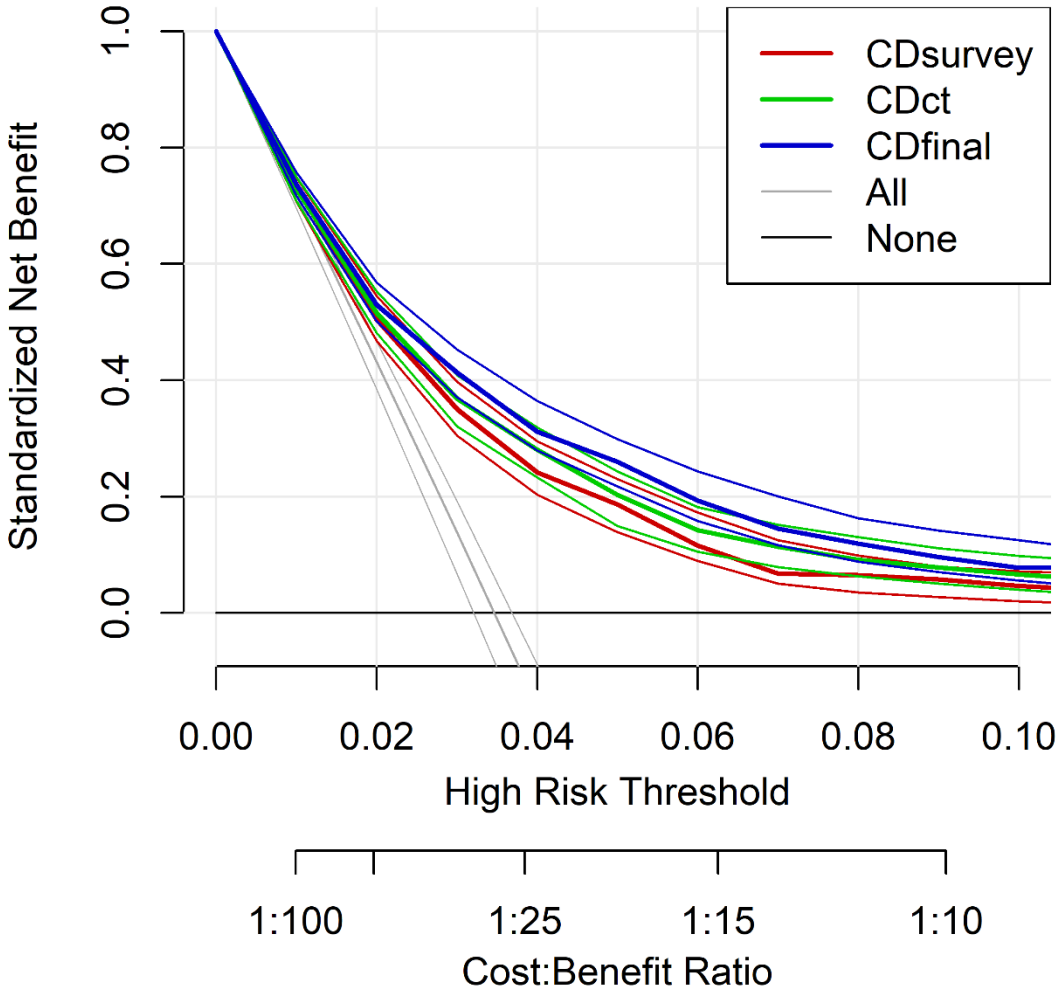


Figure S6: Decision curve analysis of competing death models in the derivation cohort

The thin lines accompanying each thicker line of the same color represent 95% confidence intervals (500 bootstrap samples). The colored lines represent the 5-year competing death models. The gray line ("All") represents the scenario where all participants are predicted to encounter a competing death. The horizontal black line at y-intercept zero represents the scenario where none are predicted to encounter the event. The x-axis represents the subjective preference threshold, where a higher risk threshold indicates a greater weight of false positive test findings per true positive test. The standardized net benefit is calculated as the difference between the true positive rate and the false positive rate (adjusted by the preference threshold and the event prevalence [5-year competing death]). A standardized net benefit greater than zero and greater than that of the gray "All" curve indicates that it would be beneficial to implement a model in practice. For a more detailed guide to the correct use and interpretation of decision curve analysis, please refer to Kerr et al. (11) or Vickers et al. (12).

CDct = CT-based competing death model; CDfinal, final competing death model; CDsurvey, self-reported patient characteristics-based competing death model.

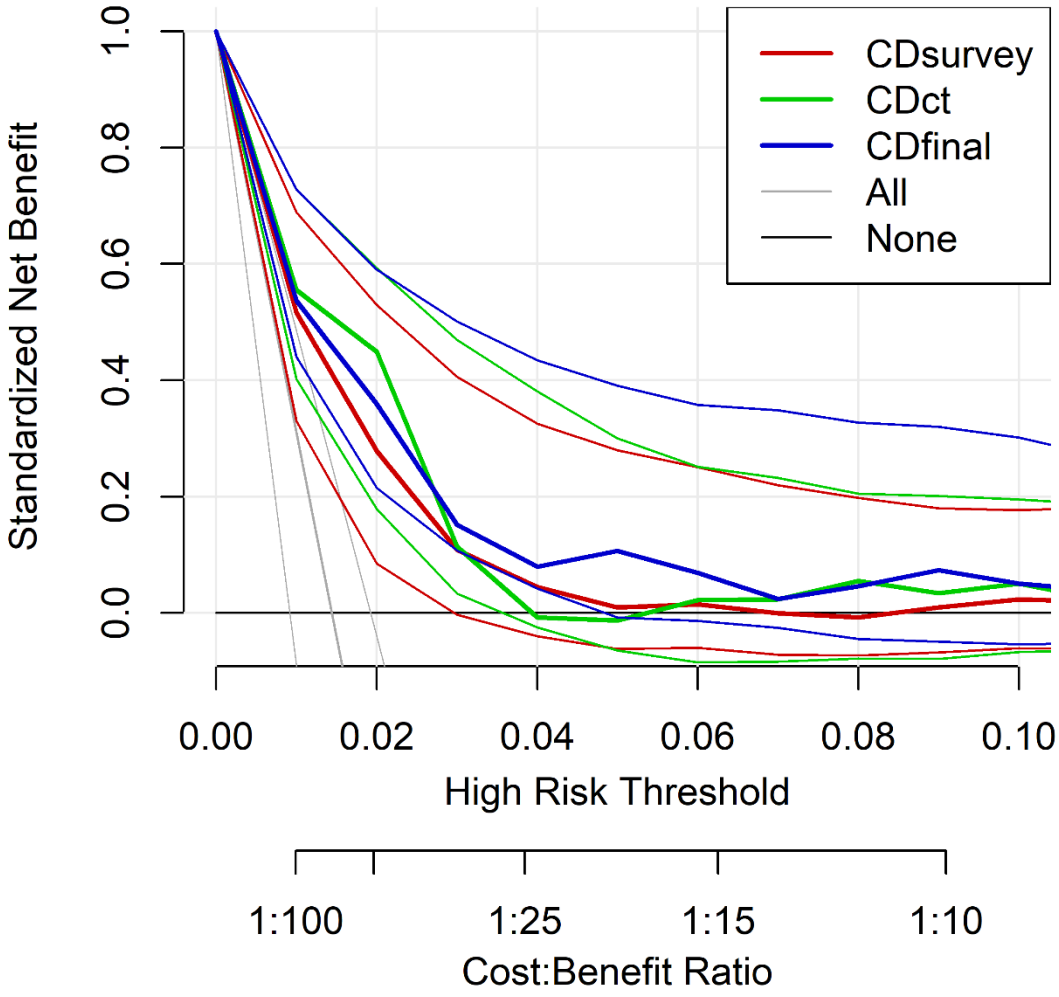
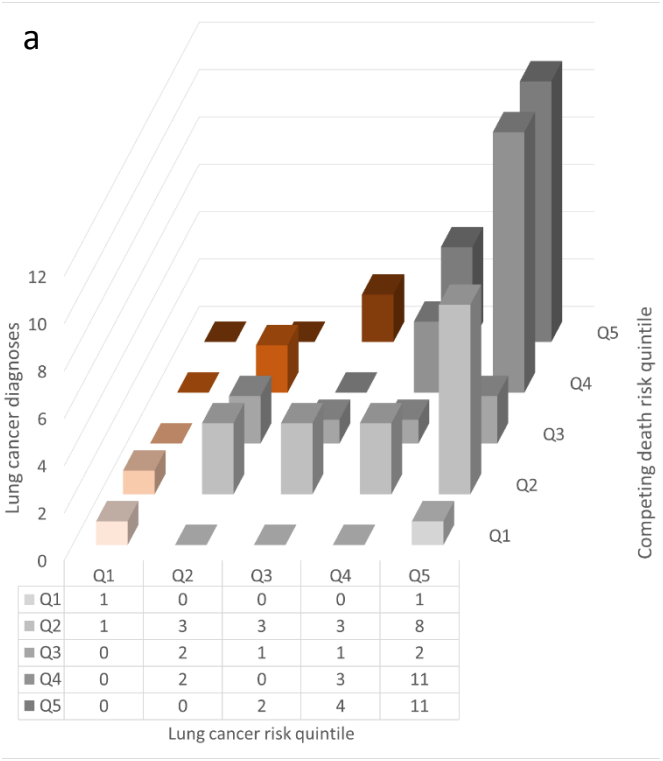
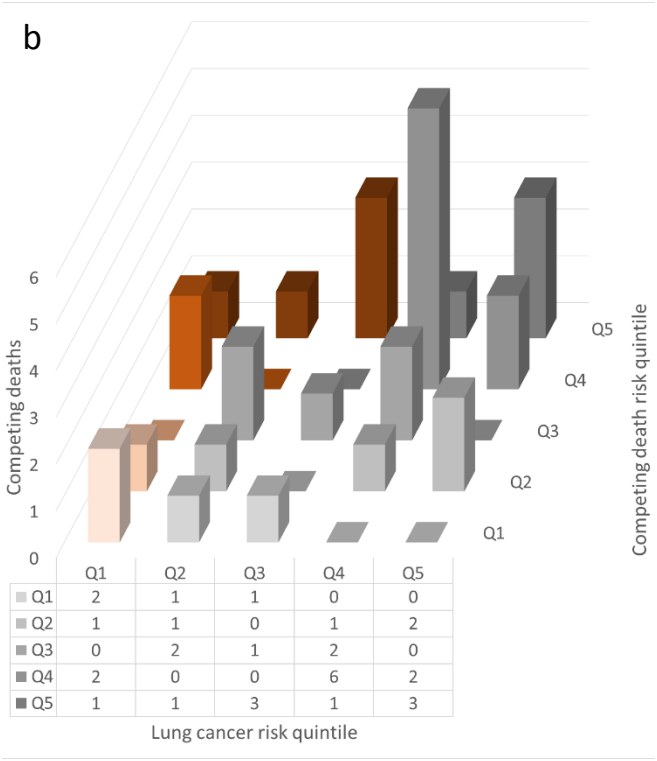


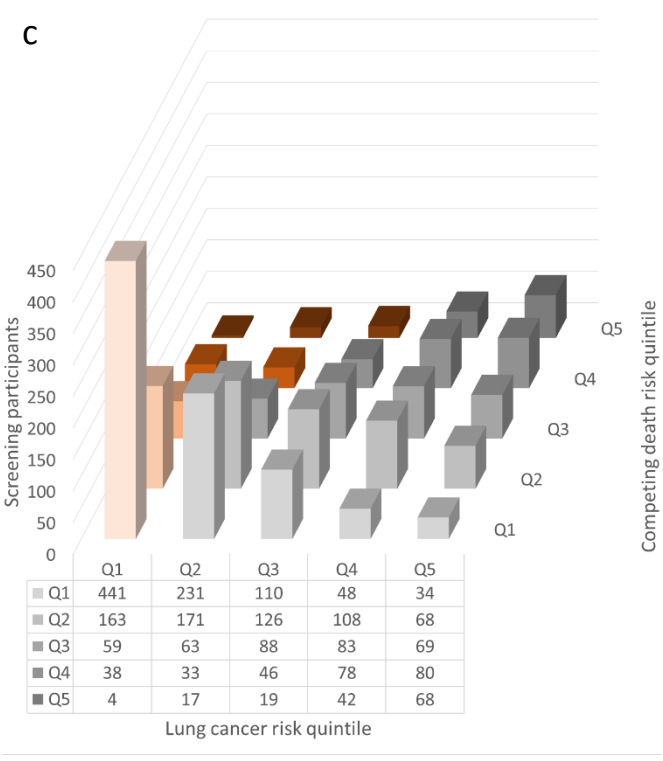
Figure S7: Decision curve analysis of competing death models in the validation cohort

The thin lines accompanying each thicker line of the same color represent 95% confidence intervals (500 bootstrap samples). The colored lines represent the 5-year competing death models. The gray line ("All") represents the scenario where all participants are predicted to encounter a competing death. The horizontal black line at y-intercept zero represents the scenario where none are predicted to encounter the event. The x-axis represents the subjective preference threshold, where a higher risk threshold indicates a greater weight of false positive test findings per true positive test. The standardized net benefit is calculated as the difference between the true positive rate and the false positive rate (adjusted by the preference threshold and the event prevalence [5-year competing death]). A standardized net benefit greater than zero and greater than that of the gray "All" curve indicates that it would be beneficial to implement a model in practice. For a more detailed guide to the correct use and interpretation of decision curve analysis, please refer to Kerr et al. (11) or Vickers et al. (12).

CDct = CT-based competing death model; CDfinal, final competing death model; CDsurvey, self-reported patient characteristics-based competing death model.







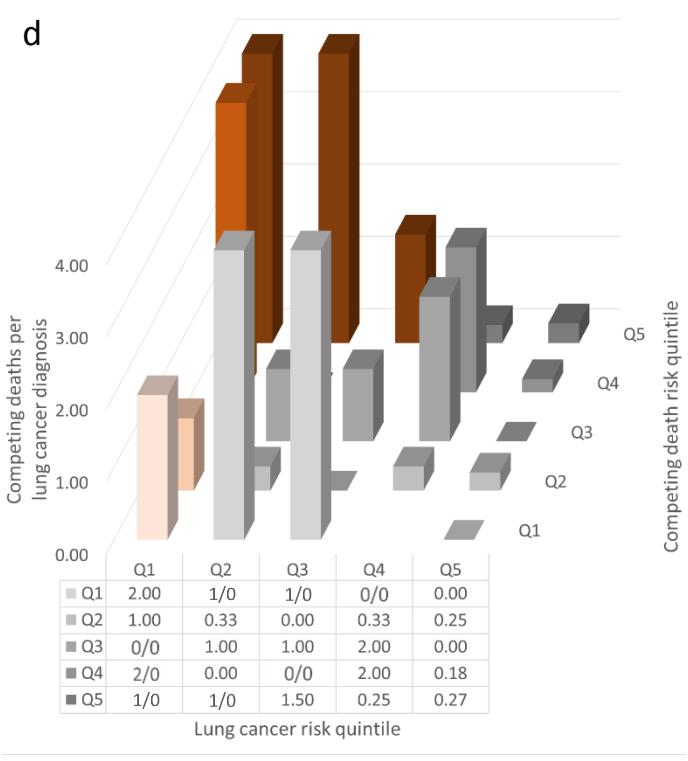


Figure S8: Outcomes per lung cancer and competing death risk quintile (validation cohort)

Three-dimensional column charts of (a) lung cancer diagnoses, (b) competing deaths, (c) screening participants, and (d) competing deaths per lung cancer diagnosis (y-axis truncated at 4). The lung cancer risk quintiles are shaded, where a darker shade corresponds to a higher risk. The columns are divided into grey and orange columns indicating the suggested separation of screening participants into a group which should continue to be screened (grey) and a group with a relatively low lung cancer risk and high risk of competing death (orange).

Q#, risk quintile where Q1 represents the lowest quintile and Q5 the highest.

References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*. 2011;365(5):395–409. doi:10.1056/NEJMoa1102873.

2. Tammemägi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *The New England journal of medicine*. 2013;368(8):728–36. doi:10.1056/NEJMoa1211776.

3. Silva M, Schaefer-Prokop CM, Jacobs C, et al. Detection of Subsolid Nodules in Lung Cancer Screening. *Investigative Radiology*. 2018;53(8):441–9. doi:10.1097/RLI.0000000000000464.

4. Silva M, Prokop M, Jacobs C, et al. Long-Term Active Surveillance of Screening Detected Subsolid Nodules is a Safe Strategy to Reduce Overtreatment. *Journal of Thoracic Oncology*. 2018;13(10):1454–63. doi:10.1016/j.jtho.2018.06.013.

5. Lessmann N, van Ginneken B, Zreik M, et al. Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks With Dilated Convolutions. *IEEE Transactions on Medical Imaging*. 2018;37(2):615–25. doi:10.1109/TMI.2017.2769839.

6. Hoffmann U, Brady TJ, Muller J. Use of New Imaging Techniques to Screen for Coronary Artery Disease. *Circulation*. 2003;108(8). doi:10.1161/01.CIR.0000085363.88377.F2.

7. Gallardo-Estrella L, Lynch DA, Prokop M, et al. Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. *European Radiology*. 2016;26(2):478–86. doi:10.1007/s00330-015-3824-y.

8. Charbonnier J-PP, Pompe E, Moore C, et al. Airway wall thickening on CT: Relation to smoking status and severity of COPD. *Respiratory Medicine*. 2019;146:36–41. doi:10.1016/j.rmed.2018.11.014.

9. Schreuder A, Jacobs C, Lessmann N, et al. Combining pulmonary and cardiac computed tomography biomarkers for disease-specific risk modelling in lung cancer screening. *European Respiratory Journal*. 2021;2003386. doi:10.1183/13993003.03386-2020.

10. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and Validation of Risk Models to Select Ever-Smokers for CT Lung Cancer Screening. *JAMA*. 2016;315(21):2300. doi:10.1001/jama.2016.6255.

11. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *Journal of Clinical Oncology*. 2016;34(21):2534–40. doi:10.1200/JCO.2015.65.5654.

12. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and Prognostic Research*. 2019;3(1):18. doi:10.1186/s41512-019-0064-7.