Article

CT-Detected Subsolid Nodules: A Predictor of Lung Cancer Development at Another Location?

Anton Schreuder [1,*][iD], Mathias Prokop [1], Ernst T. Scholten [1], Onno M. Mets [2], Kaman Chung [1,3], Firdaus A. A. Mohamed Hoesein [4], Colin Jacobs [1] and Cornelia M. Schaefer-Prokop [1,3]

[1] Department of Medical Imaging, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands; mathias.Prokop@radboudumc.nl (M.P.); ethscholten@gmail.com (E.T.S.); kaman.chung10@gmail.com (K.C.); colin.jacobs@radboudumc.nl (C.J.); cornelia.schaeferprokop@gmail.com (C.M.S.-P.)
[2] Department of Radiology and Nuclear Medicine, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands; metsonno@gmail.com
[3] Department of Radiology, Meander Medisch Centrum, 3813 TZ Amersfoort, The Netherlands
[4] Department of Radiology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands; f.a.a.mohamedhoesein@umcutrecht.nl
[*] Correspondence: antoniusschreuder@gmail.com

Simple Summary: The risk assessment of pulmonary nodules on chest computed tomography (CT) is key to the early detection of lung cancer. Whereas a nodule's morphological characteristics are strong predictors of malignancy for that individual nodule, it remains unclear whether the frequency and CT features of pulmonary nodules contribute to the prediction accuracy of future development of LC in another location. This was found by performing risk prediction modelling using age, sex, and CT measures from lung-cancer-free scans as predictors. We found that a greater number of part-solid and ground-glass nodules in the earliest scan was linearly associated with a higher risk of lung cancer developing in another location in the future. There were also indications that CT biomarkers of other pulmonary and heart diseases are risk factors for lung cancer. Our findings endorse the utilization of information from the entire scan to improve the accuracy of lung risk assessment.

Abstract: The purpose of this case–cohort study was to investigate whether the frequency and computed tomography (CT) features of pulmonary nodules posed a risk for the future development of lung cancer (LC) at a different location. Patients scanned between 2004 and 2012 at two Dutch academic hospitals were cross-linked with the Dutch Cancer Registry. All patients who were diagnosed with LC by 2014 and a random selection of LC-free patients were considered. LC patients who were determined to be LC-free at the time of the scan and all LC-free patients with an adequate scan were included. The nodule count and types (solid, part-solid, ground-glass, and perifissural) were recorded per scan. Age, sex, and other CT measures were included to control for confounding factors. The cohort included 163 LC patients and 1178 LC-free patients. Cox regression revealed that the number of ground-glass nodules and part-solid nodules present were positively correlated to future LC risk. The area under the receiver operating curve of parsimonious models with and without nodule type information were 0.827 and 0.802, respectively. The presence of subsolid nodules in a clinical setting may be a risk factor for future LC development in another pulmonary location in a dose-dependent manner. Replication of the results in screening cohorts is required for maximum utility of these findings.

1. Introduction

Lung cancer (LC) is the top cause of cancer deaths worldwide, responsible for 2.1 million deaths in 2018 [1]. A number of 57% of the patients are diagnosed with metastatic disease, for which the 5-year survival rate is approximately 5% [2]. The early diagnosis of

lung cancer almost always involves the early detection and risk assessment of pulmonary nodules on chest computed tomography (CT).

Based on their CT attenuation, lung nodules are categorized as either solid or subsolid (SSN). The great majority of lung nodules are solid with a density in the range of other soft tissue structures, such as muscles. They are very common in a screening population of people that have been smoking for a number of years [3–7]. SSNs are divided into pure ground-glass and part-solid nodules: ground-glass nodules have a vaguely increased CT attenuation, which preserves the visibility of bronchial and vascular margins; part-solid nodules have a ground-glass as well as a solid component.

Prior studies have shown that SSNs are less prevalent but at greater risk of being or becoming malignant compared to solid nodules [8–10]. To date, one study in a LC screening setting reported that the presence of SSNs might be a risk factor for LC development in another location in the lungs [4]. The main goal of our study was to attempt to reproduce these findings in a clinical setting, which may consist of any members from the general population rather than a high-risk population eligible for screening. Namely, we investigated whether the presence and frequency of certain nodule types in a baseline CT scan can predict the development of a LC that has not yet (visibly) originated, and what the added value of this information is on top of other LC risk predictors.

2. Materials and Methods

2.1. Study Population

Anonymized patient information and CT scans from between January 2004 and December 2012 were collected from all patients 40 years or older from the University Medical Center Utrecht (Center A) and Radboud University Medical Center (Center B) in the Netherlands. These data were cross-referenced with the Dutch Cancer Registry to determine the patients who were diagnosed with LC up to December 2014. Informed consent was waived by the institutional review board.

All CT scans from patients diagnosed with LC were retrieved, and the earliest available CT scan from a sample of approximately three times as many LC-free patients was selected at random. Patients and/or scans were subsequently excluded if there were image-retrieval errors, if the image was considered to be of too low quality by a radiologist, if the scan-slice thickness was greater than 3 mm, or if no scan was available at least 2 months prior to the LC diagnosis; the latter criteria was selected arbitrarily to be able to ensure the correct identification of the malignant nodule without excluding too many patients.

2.2. CT Features

All collected scans from Centers A and B were assessed independently by radiologists O.M.M (9 years' experience) and E.T.S. (>20 years' experience), respectively. Their first task was to extract CT features as given in Table 1, which can be divided into two groups: nodule-type variables and control variables, the latter of which was used to adjust for potential confounding. Dedicated in-house software (version 19.9 of CIRRUS Lung Screening, DIAG, Radboudumc, Nijmegen, The Netherlands) was applied to the scans to automatically detect and volumetrically segment pulmonary nodules. The radiologists had to either accept or reject non-calcified nodule candidates provided by the software, and were additionally required to manually scroll through the full scan for potential nodules which had been missed by the software. Only nodules that would trigger a follow-up scan according to nodule management recommendations [11–13] (BTS, LUNG RADS, Fleischner) and perifissural nodules (PFN) were included in the analysis [14]; nodules related to an obvious infection (multifocal, unclearly defined, or confluent SSNs) were excluded [15].

Table 1. Cohort demographics.

Variable	Type	Categories	Lung Cancer Patients ($n = 163$)	Lung-Cancer-Free Patients ($n = 1178$)
Control patient characteristics				
Age at earliest scan, years	Continuous	N/A	64 (58–69)	61 (52–70)
Sex (male)	Binary	1 = Male; 0 = Female	100 (61)	685 (58)
Control CT features				
Emphysema	Binary	1 = Mild, moderate, or severe; 0 = None	88 (54)	307 (26)
Bronchitis	Binary	1 = Yes; 0 = No	68 (42)	287 (24)
Interstitial lung disease	Binary	1 = Yes; 0 = No	10 (6)	19 (2)
Lymphadenopathy	Binary	1 = Yes; 0 = No	48 (29)	198 (17)
Aortic calcifications	Binary	1 = Mild, moderate, or severe; 0 = None	107 (66)	492 (42)
Coronary calcifications	Binary	1 = Mild, moderate, or severe; 0 = None	115 (71)	582 (49)
Nodule present	Binary	1 = Yes; 0 = No	132 (81)	522 (44)
Mean diameter of largest nodule, mm	Continuous	N/A	9.0 (6.3–14.6)	6.8 (5.0–12.4)
Nodule type variables				
Largest nodule is a SSN	Binary	1 = Yes; 0 = No	36/132 (27)	102/522 (20)
Solid nodule present	Binary	1 = Yes; 0 = No	105 (64)	365 (31)
Solid nodule count	Continuous	N/A	1 (1–2)	1 (1–3)
Part-solid nodule present	Binary	1 = Yes; 0 = No	18 (11)	43 (4)
Part-solid nodule count	Continuous	N/A	1 (1–1)	1 (1–1)
Ground-glass nodule present	Binary	1 = Yes; 0 = No	31 (19)	96 (8)
Ground-glass nodule count	Continuous	N/A	1 (1–1)	1 (1–2)
Perifissural nodule present	Binary	1 = Yes; 0 = No	16 (10)	166 (14)
Perifissural nodule count	Continuous	N/A	1 (1–1)	1 (1–1)

Values given in brackets are percentages for binary variables and twenty-fifth and seventy-fifth percentiles for continuous variables. For continuous variables, only values greater than zero were summarized. N/A: not applicable.

Each nodule was given a unique ID; this enabled the radiologists to retrospectively assess a nodule over time in patients who had been diagnosed with LC. This was relevant for the second task of identifying the malignant nodule. This was done by assessing nodule morphology, growth pattern, and location; the only relevant information provided by the Dutch Cancer Registry was the time of LC diagnosis and the lobe location. A confidence score from 1 to 4 was used to determine the nodule most likely to be the LC: 1 indicated benign morphology, 2 indicated uncertainty due to multiple nodules present, 3 indicated suspicious morphology without a CT scan within 1 month of LC diagnosis, and 4 indicated suspicious morphology with a CT scan within 1 month of LC diagnosis. Only nodules with a score of 3 or 4 were included as cancer cases in this study.

Nodules were classified into one of four mutually exclusive subtypes: solid, part-solid, ground-glass, or PFN. Part-solid and ground-glass nodules fall into the group of SSNs, and the definition of a perifissural nodule—which show morphology indicative of benign intrapulmonary lymph nodes—was that as defined by Hoop et al. [16].

Only the CT features of each patient's earliest scan were considered for the risk prediction analysis. The derivation cohort included all eligible LC-free patients and the eligible LC patients in which it could be said with high confidence that no malignant nodule was retrospectively visible in the earliest scan. Some LC-free patients were diagnosed with other cancer types which may have metastasized to the lungs. LC-free patients who were highly suspect of having pulmonary metastatic disease were excluded.

2.3. Statistical Analysis

All statistical analysis was performed in the statistical program R version 3.6.3 [17]. Weighted Cox proportional hazards regression was performed [18], with LC diagnosis as the event and the time between the earliest scan to LC diagnosis or the end of follow-up as the time to event. Weights were determined based on Barlow's estimate computing the Cox model for case–cohort designs [19]. Since all patients diagnosed with LC were considered, they were given a weight of 1. LC-free cases were given a weight of 37,085/1779 = 20.8, i.e., the total number of LC-free cases in the full cohort divided by the number selected at random. A Kaplan–Meier curve was plotted to visualize the time-dependent relationship between nodule types present and LC incidence risk. The Kaplan–Meier curve was adjusted using rescaled inverse probability weighting based on the following untransformed variables [18]: age, sex, emphysema, bronchitis, interstitial lung disease, lymphadenopathy, aortic calcifications, and coronary calcifications.

Univariable analysis of all independent variables was performed first (Table 1). Transformations of continuous variables with first- or second-degree factorial polynomials were considered [20,21]; transformations were reverted if there were indications for model overfitting during interval validation. Multivariable regression of all variables (i.e., the Full Model) is visualized in the form of a forest plot to indicate which predictors appear to be strongest. To gauge the best predictive performance achievable without nodule-type variables, a parsimonious multivariable model—in which variables which do not contribute significantly to the model are excluded ($p > 0.20$)—containing only control variables was derived (i.e., the Control Model). The backwards elimination procedure was used for variable selection. Interaction terms between the largest nodule's type and its mean diameter were considered, but no interaction was found. Finally, a second parsimonious model including nodule-type variables was derived (i.e., the Nodule Type Model) to assess the added value of nodule-type information compared to the Control Model.

Results are presented in terms of hazard ratios with 95% confidence intervals. The Schoenfeld global test was used to test the proportional hazards assumption [22]. Internal validation was performed for each model by deriving new model coefficients on 1000 weighted bootstrap samples and calculating optimism as the difference between the bootstrap model performance on the bootstrap sample and on the original cohort. With the exception of determining variable inclusion in a parsimonious model, statistical significance was defined as a p value of less than 0.05.

3. Results

We considered 186 LC patients and 711 LC-free patients from Center A and 292 LC patients and 508 LC-free patients from Center B with adequate CT scans (Figure 1). The malignant nodule was already visible in the earliest available scan in 315 of the 478 LC patients, and 41 of the 1219 LC-patients were suspected of having pulmonary metastasis. In total, 163 (34%) LC patients and 1178 (97%) LC-free patients were included in the derivation cohort. The median time to end of follow-up or LC diagnosis was 1923 days (twenty-fifth and seventy-fifth percentiles = 1148 and 2828 days). The patient and CT demographics are summarized in Table 1.
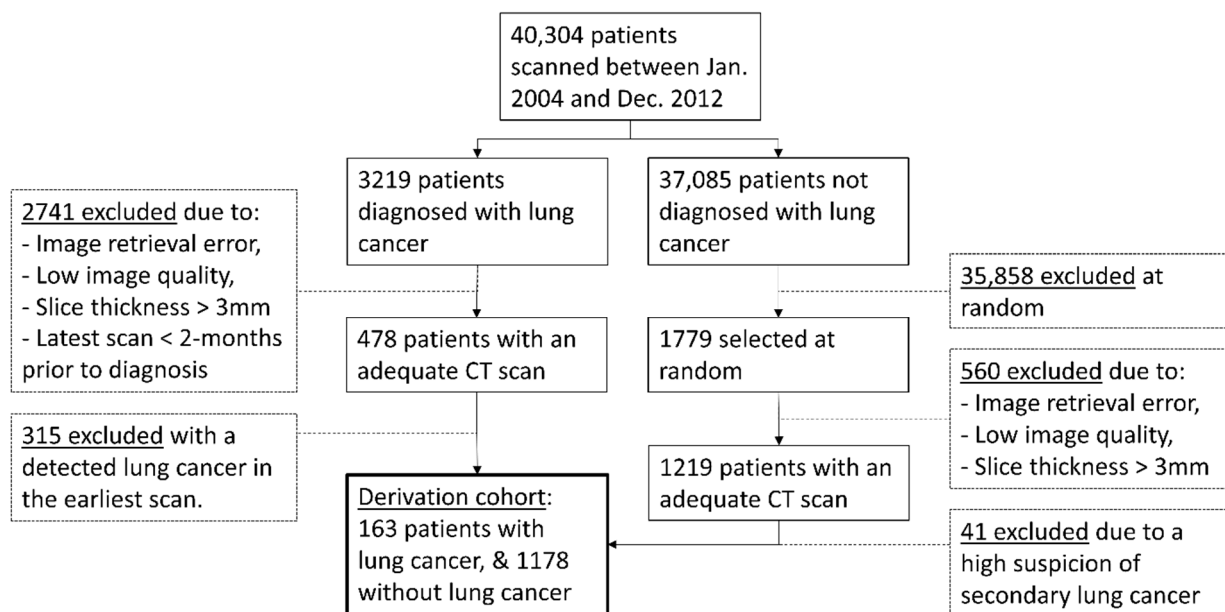
Figure 1. Patient selection flowchart.

3.1. Univariable and Full Multivariable Analysis

The continuous variables, age and solid nodule count, were transformed into second-order polynomials for a better fit of their non-linear relationship with future LC risk. A univariable analysis found all independent variables to be statistically significant predictors of future LC development, with three exceptions: sex, presence of a perifissural nodule, and perifissural nodule count (Table S1).

In the Full Model, the presence of at least one solid nodule was a future LC risk factor, but the risk decreased as the number of solid nodules increased (Figure 2). Additionally, the number of part-solid and ground-glass nodules was linearly correlated to the risk. The presence of at least one perifissural nodule was a protective factor, but was not statistically significant in the Full Model ($p = 0.07$). Compared to univariable analysis, four other variables were no longer considered to be significant predictors, namely "nodule diameter of the largest nodule", "largest nodule is a SSN", "presence of a part-solid nodule", and "presence of a ground glass nodule" (Figure 2).

Independent variable		Hazard ratio (95% confidence interval)		p value
$((\text{Age} - 39)/10)^2$	(n = 1341)	1.62 (1.267 − 2.06)		<0.001 ***
$((\text{Age} - 39)/10)^3$	(n = 1341)	0.88 (0.826 − 0.93)		<0.001 ***
Male sex	0 (n = 556)	reference		
	1 (n = 785)	0.89 (0.630 − 1.25)		0.502
Emphysema	0 (n = 944)	reference		
	1 (n = 395)	2.02 (1.441 − 2.83)		<0.001 ***
Lymphadenopathy	0 (n = 1095)	reference		
	1 (n = 246)	1.78 (1.238 − 2.56)		0.002 **
Bronchitis	0 (n = 986)	reference		
	1 (n = 355)	1.51 (1.076 − 2.12)		0.017 *
ILD	0 (n = 1312)	reference		
	1 (n = 29)	3.63 (1.797 − 7.32)		<0.001 ***
Aorta calcifications	0 (n = 742)	reference		
	1 (n = 599)	1.73 (1.152 − 2.59)		0.008 **
Coronary calcifications	0 (n = 644)	reference		
	1 (n = 697)	1.79 (1.207 − 2.66)		0.004 **
Nodule present	0 (n = 687)	reference		
	1 (n = 654)	4.68 (2.209 − 9.93)		<0.001 ***
Longest diameter (mm)	(n = 1341)	0.99 (0.961 − 1.01)		0.233
Largest nodule is a SSN	0 (n = 1203)	reference		
	1 (n = 138)	1.11 (0.471 − 2.59)		0.818
Solid nodule present	0 (n = 871)	reference		
	1 (n = 470)	0.29 (0.095 − 0.88)		0.028 *
Solid nodule count	(n = 1341)	27.65 (3.765 − 203.03)		0.001 **
$(\text{SNC} + 1) \times \ln(\text{SNC} + 1)$	(n = 1341)	0.23 (0.095 − 0.54)		<0.001 ***
Part-solid nodule present	0 (n = 1280)	reference		
	1 (n = 61)	0.70 (0.234 − 2.12)		0.533
Part-solid nodule count	(n = 1341)	2.02 (1.182 − 3.44)		0.01 *
Ground glass nodule present	0 (n = 1214)	reference		
	1 (n = 127)	0.77 (0.313 − 1.87)		0.559
Ground glass nodule count	(n = 1341)	1.64 (1.132 − 2.38)		0.009 **
Perifissural nodule present	0 (n = 1159)	reference		
	1 (n = 182)	0.35 (0.114 − 1.07)		0.066
Perifissural nodule count	(n = 1341)	0.87 (0.439 − 1.71)		0.681
Weights	(n = 1341)	reference		

Number of events: 163; Global p value (Log-Rank): 8.6766×10^{-47}
AIC: 2934.71; Concordance index: 0.83

0.1 0.5 1 5 10 50 100
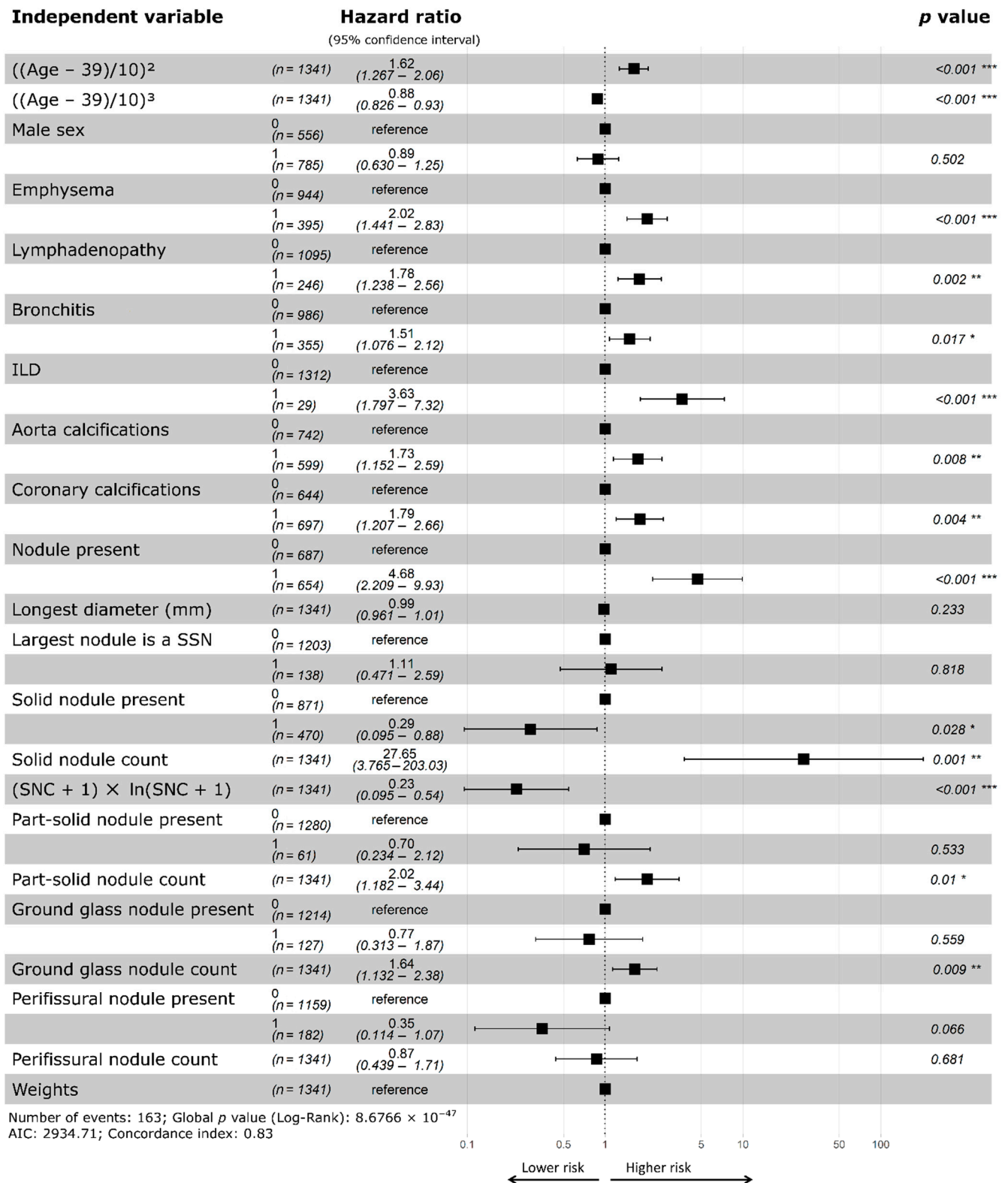
← Lower risk   Higher risk →

Figure 2. Forest plot of multivariable Cox regression. For binary variables, 0 (reference value) indicates absence and 1 indicates presence. For continuous variables (i.e., age, diameter, nodule counts), the hazard ratio refers to each unit increase. A hazard ratio greater than 1 indicates that the presence of the variable increases the risk of lung cancer incidence; conversely, a value less than 1 indicates that the presence of the variable reduces the lung cancer incidence risk. * p value < 0.05; ** p value < 0.01; *** p value < 0.001; ILD, interstitial lung disease; N, number of; PFN, perifissural nodule; SNC, solid nodule count; SSN, subsolid nodule.

3.2. Parsimonious Lung Cancer Incidence Risk Models

Only "nodule diameter of the largest nodule" was excluded from the parsimonious Control Model, leaving in age, sex, emphysema, bronchitis, interstitial lung disease, lymphadenopathy, aortic calcifications, and coronary calcifications (Table S2). Four nodule-type variables were added to form the Nodule Type Model (Table 2): "solid nodule count" (transformed), "part-solid nodule count", "ground glass nodule count", and "presence of a perifissural nodule", the latter indicating a lower future LC risk.

Table 2. Parsimonious lung cancer risk model (Nodule Type Model).

Variable	Beta Coefficient	Hazard Ratio (95% Confidence Interval)	*p* Value
Patient characteristics			
Age at earliest scan, per year $(((x-39)/10)^2)$	0.50255	1.65 (1.25 to 2.18)	<0.001
Age at earliest scan, per year $(((x-39)/10)^3)$	−0.13941	0.87 (0.82 to 0.93)	<0.001
CT features excluding nodule type			
Emphysema	0.70903	2.03 (1.33 to 3.12)	0.001
Bronchitis	0.39078	1.48 (0.95 to 2.29)	0.08
Interstitial lung disease	1.24998	3.49 (1.36 to 8.97)	0.009
Lymphadenopathy	0.54126	1.72 (1.06 to 2.93)	0.028
Aortic calcifications	0.57696	1.78 (1.08 to 2.93)	0.023
Coronary calcifications	0.51209	1.67 (1.04 to 2.69)	0.035
Nodule present	1.12291	3.07 (1.54 to 6.13)	0.001
Nodule type information			
Solid nodule count, per nodule	1.39874	4.05 (1.30 to 12.63)	0.016
Solid nodule count, per nodule $((x+1) \times \ln(x+1))$	−0.68170	0.51 (0.31 to 0.83)	0.007
Part-solid nodule count, per nodule	0.50634	1.66 (1.11 to 2.47)	0.013
Ground-glass nodule count, per nodule	0.41971	1.52 (1.06 to 2.17)	0.021
Perifissural nodule present	−1.11383	0.33 (0.16 to 0.66)	0.002

A parsimonious model was derived using Cox regression containing only variables with a *p* value < 0.20 using backwards elimination. Some variables were transformed as indicated by the mathematical expression in brackets, in which *x* represents the untransformed variable. N/A: not applicable.

Given the finding that there may be a dose-dependent relationship between the number of SSNs present in a scan, each cohort was divided into four groups: patients with no nodules, patients with at least one nodule (none of which are SSNs), patients with a single SSN, and patients with more than one SSN. The LC-free survival probability over time of the groups was visualized as Kaplan–Meier curves (Figure 3). Among patients with at least 5 years of LC-free follow-up or LC diagnosis within 5 years, LC incidence was 6% (23/379) in those without any nodules at baseline, increasing to 22% (67/300) in those with at least one solid nodule (no subsolid nodules), 22% (18/82) in those with a single SSN, and 40% (14/35) in those with at least two SSNs. The difference in LC-free survival between the group without nodules and the other three groups was statistically significant (log-rank test *p* < 0.001). The difference between having at least two SSNs compared to a single SSN or only solid nodules was also significant (*p* = 0.05 and *p* = 0.03, respectively). The LC probability between the single SSN and only solid nodules groups was equivalent (*p* = 1.00). Weighted 5-year LC-free survival probabilities are given in the Figure 3 legend.
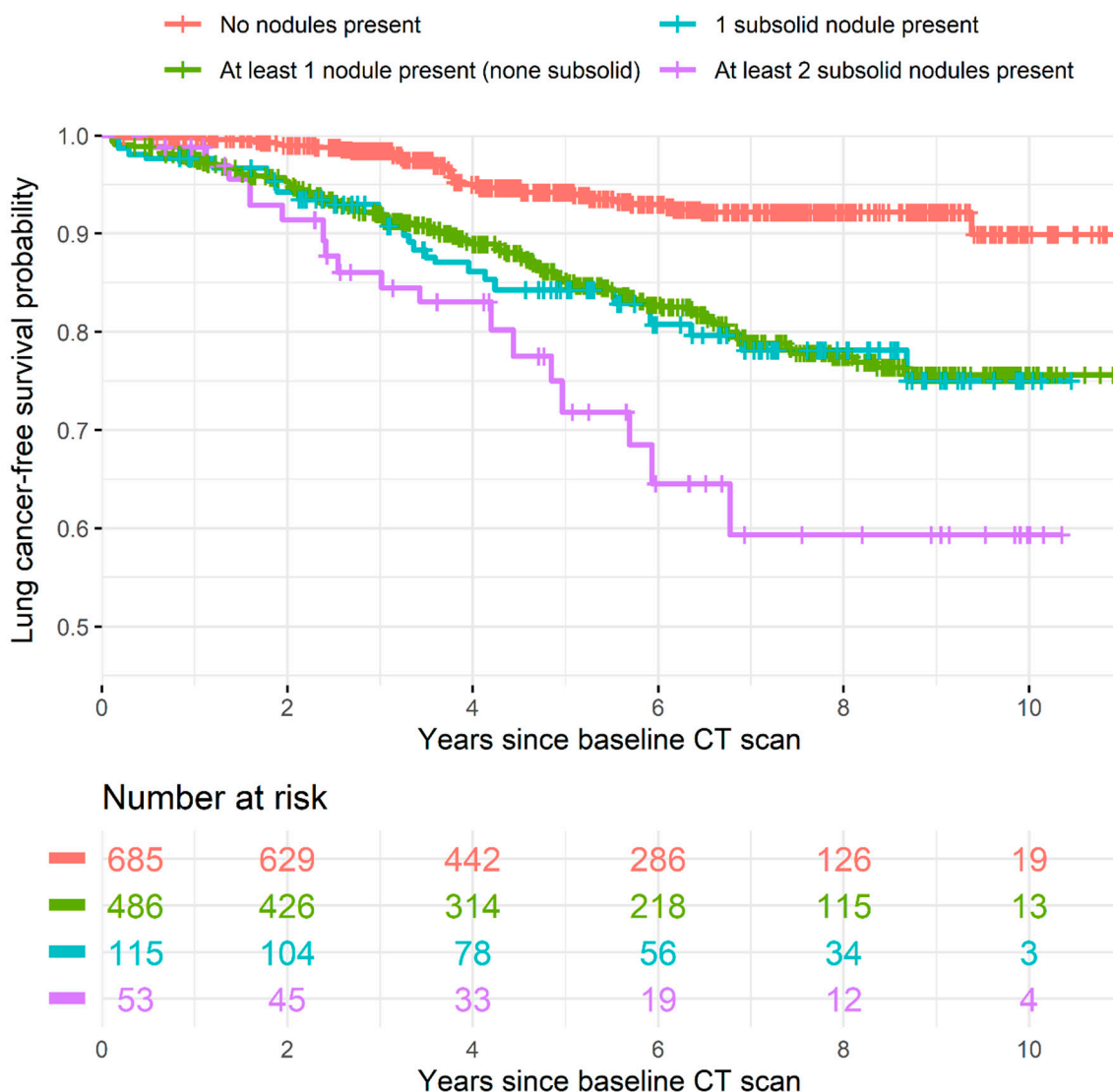
Figure 3. Lung cancer probability by nodule types present at baseline. This figure shows a Kaplan–Meier curve of lung-cancer-free survival, including a "number at risk" table. Vertical dashes on a line indicate the end of patients' follow-up. The 5-year lung-cancer-free weighted survival probabilities were 0.942 (95% confidence intervals = 0.916, 0.969) when no nodules were present at baseline, 0.852 (0.817 to 0.887) when only solid nodules were present, 0.843 (0.775 to 0.917) when one SSN was present, and 0.718 (0.594 to 0.868) when at least two SSNs were present. The same figure with confidence intervals is provided as Figure S1.

Internal validation of the Nodule Type Model (Table 2) resulted in an area under the receiver operating characteristic curve (AUC) of 0.827 (95% confidence band = 0.826 to 0.827) on the original sample and 0.838 (0.837 to 0.838) on the bootstrap sample (optimism = 0.011). The Control Model's AUC was 0.802 (0.801 to 0.802) on the original sample and 0.811 (0.810 to 0.812) on the bootstrap sample (optimism = 0.009) (Table S2).

4. Discussion

Our research question was whether the nodule types present in a CT scan had predictive value for the development of LC in another location. To investigate this, we performed weighted time-dependent regression analysis on a clinical case–cohort which only included patients with LC-free baseline CT scans. This was necessary to ensure the correct direction of causality: that the nodule types occurred before LC incidence. Age, sex, and other quantitative CT information were included in the analysis to control for potential confounding.

One of the cohort's inclusion criteria was a high level of certainty that lung cancer was not (yet) detectable in the baseline scan. This can only be done retrospectively with information from follow-up scans; the models and predictors described should be validated and calibrated in a prospective clinical cohort.

Results of the univariable analysis and the Full Model revealed that the presence of at least one solid nodule, part-solid nodule, or ground-glass nodule were significant risk factors for future LC incidence (Table S1, Figure 2). Parsimonious models were derived to simplify the interpretation by only considering the variables with the highest predictive power (Table 2 and Table S2). The Nodule Type Model confirmed that the presence of nodules is a risk factor for future LC development. Though this observation has been previously described [4], a novel finding is that more part-solid and ground-glass nodules increase the risk linearly, whereas more solid nodules (beyond one) reduces the risk. As only 27 patients had more than two SSNs, it remains unclear whether the relationship is indeed linear for higher SSN counts. The presence of at least one perifissural nodule was a protective factor: its beta coefficient was equivalent in magnitude as that for the presence of at least one nodule (−1.11 and 1.12, respectively, Table 2), indicating that patients with nodules which are exclusively perifissural nodules have an equivalent risk to patients without any nodules.

By calculating the difference in AUC performance between the parsimonious models with and without nodule type information (Nodule Type Model and Control Model, respectively), we found that the added value of nodule-type information was 0.025. Note that the AUC is a summary measure that does not reflect clinical utility. To compensate, time-to-event curves were plotted to visually display the trend of higher LC risk among patients with more SSNs (Figure 3).

As would be expected from only including data from LC-free CT scans, the largest nodule diameter was not a predictor of future malignancy in multivariable analysis. In a scenario where LC may already be present, nodule size is otherwise established as the strongest predictor of LC based on a single scan [9,23,24]. No interaction was found between nodule type (i.e., solid vs. SSN) and nodule diameter.

The finding that the presence of an SSN may be a risk factor for future malignancy development in another lung location has thus far been described only once, by Silva et al. [4], based on results from the MILD screening population: of 389 screenees with at least one SSN, 30 LCs were diagnosed, of which 22 were not derived from an SSN. In this context, it is noteworthy that the same author group described an increased malignancy risk for patients with an overall increased lung density, interpreted as an increased inflammatory state [25].

We also found that greater numbers of solid nodules (beyond one) are indicative of lower LC risk. A possible explanation is that some patients naturally develop more solid nodules with no clinical significance. Most benign solid nodules are small (i.e., <6 mm) and represent benign entities [26]; not including small solid nodules in the count may counteract this finding. Note that there is a moderate inter-rater agreement when classifying solid nodules as perifissural nodules [27], and a lack of a standardized definition [14]. Therefore, there may have been an insufficient distinction between small solid nodules and perifissural nodules. We expect that a considerable number of these solid nodules could also be classified as perifissural nodules, of which the LC risk can be considered negligible [14].

The clinical implication of our results is that future LC risk models should consider incorporating SSN count. Such models would likely be most useful in a LC screening setting for the purpose of personalizing scan intervals. As our results were derived from a clinical cohort, subsequent research should first attempt to replicate our findings in a high-risk screening population. Although the predictive value of nodule-type information alone appears to be relatively small, it can be significant when applied to large numbers. To the best of our knowledge, no lung cancer risk model has yet considered the number of SSNs and solid nodules as variables [28]. Schreuder et al. [23] included the "presence of

non-solid nodule" and "presence of part-solid nodule" in their parsimonious model for determining whole scan-based risk, but did not count the SSNs.

Another finding was that radiologist-rated presence of emphysema, bronchitis, interstitial lung disease, lymphadenopathy, calcifications of the coronary arteries, and calcifications of the transthoracic aorta are non-nodule CT features which contributed to LC prediction accuracy. This supports the notion that the presence of chronic obstructive pulmonary disease and/or cardiovascular disease may indicate susceptibility for LC, likely due to exposure to overlapping risk factors [24,29–33]. Note that while the presence of emphysema, coronary calcifications, and aorta calcifications was originally classified as "none", "mild", "moderate", or "severe", no statistically significant trends were found with increasing severity (results not shown); this was why these variables were converted into binary form for the analysis. While the relationship between these variables with LC incidence is unlikely to be linear [24], it is unclear whether quantitative CT measures would have been better predictors compared to clinician-reported outcome measures.

Limitations

Our study has other limitations that should be taken into consideration. First, no data on smoking status or intensity were available, which would have been a strong demographic predictor of LC and may have been informative as a control variable or to select a higher risk population equivalent to that in a screening setting [34]. Data on patient mortality were also lacking, preventing the investigation of potential differences in LC prognosis; a higher LC risk does not imply a poorer prognosis [35]. Only the size of the largest nodule was considered for the model to avoid overfitting and to simplify model interpretation. The risk models only considered the earliest scan, from which it is not possible to derive information on nodule growth and whether nodules are transient or persistent. Finally, it is possible that the follow-up time was not sufficiently long, and that some LC nodules were erroneously deemed to be benign in our study.

5. Conclusions

In conclusion, our results indicate that the presence of single time-point SSNs on CT in a clinical setting may be a risk factor for future LC development in another lung location in a dose-dependent manner. The risk may increase linearly with the number of SSNs. This finding highlights the potential importance of utilizing information from the entire CT scan to improve the accuracy of LC risk assessment, and work towards personalized medicine.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cancers13112812/s1, Figure S1: Lung cancer probability by nodule type present (with confidence intervals). Table S1: Univariable Cox regression for lung cancer risk prediction. Table S2: Parsimonious lung cancer risk model excluding nodule-type information.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the anonymization of patient data.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Customized data were supplied by the Netherlands Cancer Registry. Instructions for applying for data can be found here: https://iknl.nl/en/ncr/apply-for-data. It is also possible to consult the publicly accessible data available on www.kanker.nl (only in Dutch).

References

1. World Health Organization. The Top 10 Causes of Death 2018. Available online: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed on 6 May 2020).
2. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [CrossRef] [PubMed]
3. Migliore, M.; Fornito, M.; Palazzolo, M.; Criscione, A.; Gangemi, M.; Borrata, F.; Vigneri, P.; Nardini, M.; Dunning, J. Ground glass opacities management in the lung cancer screening era. *Ann. Transl. Med.* **2018**, *6*, 90. [CrossRef] [PubMed]
4. Silva, M.; Prokop, M.; Jacobs, C.; Capretti, G.; Sverzellati, N.; Ciompi, F.; van Ginneken, B.; Schaefer-Prokop, C.M.; Galeone, C.; Marchianò, A.; et al. Long-Term Active Surveillance of Screening Detected Subsolid Nodules is a Safe Strategy to Reduce Overtreatment. *J. Thorac. Oncol.* **2018**, *13*, 1454–1463. [CrossRef]
5. Walter, J.E.; Heuvelmans, M.A.; Yousaf-Khan, U.; Dorrius, M.D.; Thunnissen, E.; Schermann, A.; Groen, H.J.; Van Der Aalst, C.M.; Nackaerts, K.; Vliegenthart, R.; et al. New Subsolid Pulmonary Nodules in Lung Cancer Screening: The NELSON Trial. *J. Thorac. Oncol.* **2018**, *13*, 1410–1414. [CrossRef]
6. Henschke, C.I.; Yip, R.; Smith, J.P.; Wolf, A.S.; Flores, R.M.; Liang, M.; Salvatore, M.M.; Liu, Y.; Xu, D.M.; Yankelevitz, D.F.; et al. CT Screening for Lung Cancer: Part-Solid Nodules in Baseline and Annual Repeat Rounds. *Am. J. Roentgenol.* **2016**, *207*, 1176–1184. [CrossRef] [PubMed]
7. Yankelevitz, D.F.; Yip, R.; Smith, J.P.; Liang, M.; Liu, Y.; Xu, D.M.; Salvatore, M.M.; Wolf, A.S.; Flores, R.M.; Henschke, C.I.; et al. CT Screening for Lung Cancer: Nonsolid Nodules in Baseline and Annual Repeat Rounds. *Radiology* **2015**, *277*, 555–564. [CrossRef] [PubMed]
8. Hammer, M.M.; Palazzo, L.L.; Kong, C.Y.; Hunsaker, A.R. Cancer Risk in Subsolid Nodules in the National Lung Screening Trial. *Radiology* **2019**, *293*, 441–448. [CrossRef] [PubMed]
9. McWilliams, A.; Tammemagi, M.C.; Mayo, J.R.; Roberts, H.; Liu, G.; Soghrati, K.; Yasufuku, K.; Martel, S.; Laberge, F.; Gingras, M.; et al. Probability of Cancer in Pulmonary Nodules Detected on First Screening CT. *N. Engl. J. Med.* **2013**, *369*, 910–919. [CrossRef]
10. Henschke, C.I.; Yankelevitz, D.F.; Mirtcheva, R.; McGuinness, G.; McCauley, D.; Miettinen, O.S. CT Screening for Lung Cancer. *Am. J. Roentgenol.* **2002**, *178*, 1053–1057. [CrossRef]
11. Callister, M.E.J.; Baldwin, D.R.; Akram, A.R.; Barnard, S.; Cane, P.; Draffan, J.; Franks, K.; Gleeson, F.; Graham, R.; Malhotra, P.; et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules: Accredited by NICE. *Thorax* **2015**, *70*, ii1–ii54. [CrossRef]
12. The American College of Radiology. Lung CT Screening Reporting & Data System v1.0 2014. Available online: https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads (accessed on 29 April 2021).
13. MacMahon, H.; Naidich, D.P.; Goo, J.M.; Lee, K.S.; Leung, A.N.C.; Mayo, J.R.; Mehta, A.C.; Ohno, Y.; Powell, C.A.; Prokop, M.; et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* **2017**, *284*, 228–243. [CrossRef]
14. Schreuder, A.; Jacobs, C.; Scholten, E.T.; Van Ginneken, B.; Schaefer-Prokop, C.M.; Prokop, M. Typical CT Features of Intrapulmonary Lymph Nodes: A Review. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e190159. [CrossRef] [PubMed]
15. Chung, K.; Ciompi, F.; Scholten, E.T.; Goo, J.M.; Prokop, M.; Jacobs, C.; Van Ginneken, B.; Schaefer-Prokop, C.M. Visual discrimination of screen-detected persistent from transient subsolid nodules: An observer study. *PLoS ONE* **2018**, *13*, e0191874. [CrossRef]
16. De Hoop, B.; Van Ginneken, B.; Gietema, H.; Prokop, M. Pulmonary Perifissural Nodules on CT Scans: Rapid Growth Is Not a Predictor of Malignancy. *Radiology* **2012**, *265*, 611–616. [CrossRef]
17. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
18. Therneau, T.M. A Package for Survival Analysis in R. 2020. Available online: https://CRAN.R-project.org/package=survival (accessed on 20 April 2021).
19. Therneau, T.M.; Li, H. Computing the Cox model for case cohort designs. *Lifetime Data Anal.* **1999**, *5*, 99–112. [CrossRef]
20. Zhang, Z. Multivariable fractional polynomial method for regression model. *Ann. Transl. Med.* **2016**, *4*, 174. [CrossRef]
21. Royston, P.; Altman, D.G. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1994**, *43*, 429. [CrossRef]
22. Schoenfeld, D. Chi-Squared Goodness-of-Fit Tests for the Proportional Hazards Regression Model. *Biometrika* **1980**, *67*, 145–153. [CrossRef]
23. Schreuder, A.; Schaefer-Prokop, C.M.; Scholten, E.T.; Jacobs, C.; Prokop, M.; Van Ginneken, B. Lung cancer risk to personalise annual and biennial follow-up computed tomography screening. *Thorax* **2018**, *73*, 626–633. [CrossRef] [PubMed]

24. Schreuder, A.; Jacobs, C.; Lessmann, N.; Broeders, M.J.; Silva, M.; Išgum, I.; de Jong, P.A.; Sverzellati, N.; Prokop, M.; Pastorino, U.; et al. Combining pulmonary and cardiac computed tomography biomarkers for disease-specific risk modelling in lung cancer screening. *Eur. Respir. J.* **2021**, *2021*, 2003386. [CrossRef] [PubMed]

25. Sverzellati, N.; Randi, G.; Spagnolo, P.; Marchianò, A.; Silva, M.; Kuhnigk, J.-M.; Vecchia, C.L.; Zompatori, M.; Pastorino, U. Increased mean lung density: Another independent predictor of lung cancer? *Eur. J. Radiol.* **2013**, *82*, 1325–1331. [CrossRef]

26. Honma, K.; Nelson, G.; Murray, J. Intrapulmonary lymph nodes in South African miners—An autopsy survey. *Am. J. Ind. Med.* **2007**, *50*, 261–264. [CrossRef] [PubMed]

27. Schreuder, A.; Van Ginneken, B.; Scholten, E.T.; Jacobs, C.; Prokop, M.; Sverzellati, N.; Desai, S.R.; Devaraj, A.; Schaefer-Prokop, C.M. Classification of CT Pulmonary Opacities as Perifissural Nodules: Reader Variability. *Radiology* **2018**, *288*, 867–875. [CrossRef]

28. Toumazis, I.; Bastani, M.; Han, S.S.; Plevritis, S.K. Risk-Based lung cancer screening: A systematic review. *Lung Cancer* **2020**, *147*, 154–186. [CrossRef] [PubMed]

29. De Verdier, M.G. The Big Three Concept: A Way to Tackle the Health Care Crisis? *Proc. Am. Thorac. Soc.* **2008**, *5*, 800–805. [CrossRef] [PubMed]

30. Zulueta, J.J.; Wisnivesky, J.P.; Henschke, C.I.; Yip, R.; Farooqi, A.O.; McCauley, D.I.; Chen, M.; Libby, D.M.; Smith, J.P.; Pasmantier, M.W.; et al. Emphysema Scores Predict Death from COPD and Lung Cancer. *Chest* **2012**, *141*, 1216–1223. [CrossRef]

31. Fan, L.; Fan, K. Lung cancer screening CT-based coronary artery calcification in predicting cardiovascular events. *Medicine* **2018**, *97*, e10461. [CrossRef] [PubMed]

32. Robbins, H.A.; Berg, C.D.; Cheung, L.C.; Chaturvedi, A.K.; Katki, H.A. Identification of Candidates for Longer Lung Cancer Screening Intervals Following a Negative Low-Dose Computed Tomography Result. *J. Natl. Cancer Inst.* **2019**, *111*, 996–999. [CrossRef]

33. Lebrett, M.B.; Balata, H.; Evison, M.; Colligan, D.; Duerden, R.; Elton, P.; Greaves, M.; Howells, J.; Irion, K.; Karunaratne, D.; et al. Analysis of lung cancer risk model (PLCOM2012 and LLPv2) performance in a community-based lung cancer screening programme. *Thorax* **2020**, *75*, 661–668. [CrossRef] [PubMed]

34. Snowsill, T.; Yang, H.; Griffin, E.; Long, L.; Varley-Campbell, J.; Coelho, H.; Robinson, S.; Hyde, C. Low-dose computed tomography for lung cancer screening in high-risk populations: A systematic review and economic evaluation. *Heal. Technol. Assess.* **2018**, *22*, 1–276. [CrossRef]

35. Kim, T.J.; Goo, J.M.; Lee, K.W.; Park, C.M.; Lee, H.J. Clinical, pathological and thin-section CT features of persistent multiple ground-glass opacity nodules: Comparison with solitary ground-glass opacity nodule. *Lung Cancer* **2009**, *64*, 171–178. [CrossRef] [PubMed]