Original Paper

Lung Cancer Risk Prediction Nomogram in Nonsmoking Chinese Women: Retrospective Cross-sectional Cohort Study

Lanwei Guo[1], MD, PhD; Qingcheng Meng[2], MD; Liyang Zheng[1], BSc; Qiong Chen[1], MD; Yin Liu[1], MD; Huifang Xu[1], MD; Ruihua Kang[1], MD; Luyao Zhang[1], MD; Shuzheng Liu[1], MD; Xibin Sun[1], MPH; Shaokai Zhang[1], MD

[1]Department of Cancer Epidemiology and Prevention, Henan Engineering Research Center of Cancer Prevention and Control, Henan International Joint Laboratory of Cancer Prevention, The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, China

[2]Department of Radiology, The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, China

Corresponding Author:
Shaokai Zhang, MD
Department of Cancer Epidemiology and Prevention, Henan Engineering Research Center of Cancer Prevention and Control, Henan International Joint Laboratory of Cancer Prevention
The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital
No.127 Dongming Road
Zhengzhou, 450008
China
Phone: 86 37165587361
Email: shaokaizhang@126.com

Abstract

Background: It is believed that smoking is not the cause of approximately 53% of lung cancers diagnosed in women globally.

Objective: The study aimed to develop and validate a simple and noninvasive model that could assess and stratify lung cancer risk in nonsmoking Chinese women.

Methods: Based on the population-based Cancer Screening Program in Urban China, this retrospective, cross-sectional cohort study was carried out with a vast population base and an immense number of participants. The training set and the validation set were both constructed using a random distribution of the data. Following the identification of associated risk factors by multivariable Cox regression analysis, a predictive nomogram was developed. Discrimination (area under the curve) and calibration were further performed to assess the validation of risk prediction nomogram in the training set, which was then validated in the validation set.

Results: In sum, 151,834 individuals signed up to take part in the survey. Both the training set (n=75,917) and the validation set (n=75,917) were comprised of randomly selected participants. Potential predictors for lung cancer included age, history of chronic respiratory disease, first-degree family history of lung cancer, menopause, and history of benign breast disease. We displayed 1-year, 3-year, and 5-year lung cancer risk–predicting nomograms using these 5 factors. In the training set, the 1-year, 3-year, and 5-year lung cancer risk areas under the curve were 0.762, 0.718, and 0.703, respectively. In the validation set, the model showed a moderate predictive discrimination.

Conclusions: We designed and validated a simple and noninvasive lung cancer risk model for nonsmoking women. This model can be applied to identify and triage people at high risk for developing lung cancers among nonsmoking women.

(JMIR Public Health Surveill 2023;9:e41640) doi: 10.2196/41640

Introduction

China has the most lung cancer death cases around the world in 2020. In 2020, according to estimates provided by the International Agency for Research on Cancer, there were approximately 1.80 million cases of deadly lung cancer globally. China accounted for 39.8% of these cases [1]. In China, the continuous rise in lung cancer deaths during the past 2 decades was attributed to the rising prevalence of lung cancer in women [2]. Additionally, 50% or more of lung cancers in women in Southeast Asia were diagnosed in nonsmokers [3-5]. Most of the Chinese lung cancer cases were reported to be clinically progressed in 2012-2014, with 64.6% of them being stage III-IV lung cancers [6]. The lung cancer survival rate in China, which

was defined as 5 years as standard, grew up to 20% between 2003 and 2015 [7]. The prognosis of lung cancer is strongly associated with the stage in which it was detected; the 5-year survival rate ranges from 0% in cases detected in patients with stage IV cancer to >80% in cases detected in stage I and whose patients underwent surgery [8].

Started in 2002, the National Lung Screening Trial indicated that low-dose computed tomography screening may decrease lung cancer deaths by 20% [9]. However, this project only screened people (41% women) at high risk for lung cancer based on age and smoking history (aged 55-74 years, smoked no fewer than 30 pack-years, and had no more than 15 years of having quit smoking). Women in China have their own characteristics of lung cancer risk factor exposure and incidence patterns, the most critical of which is that although the smoking rate among women is much lower than that of high-income countries such as the United States (2.4% in China and 23.6% in the United States), the lung cancer frequencies are relatively similar (22.8/100,000 in China and 30.8/100,000 in the United States, based on the standardized lung cancer incidence rate of the world population) [10,11]. This finding shows that the existing worldwide guidelines for lung cancer screening focused on smoking as the primary predictor for high-risk individuals, which would be inappropriate for Chinese women, particularly for nonsmoking women. Therefore, determining a way that accurately forecasts the risk of lung cancer in nonsmoking women and directing them toward the more cost-effective low-dose computed tomography screening is a feasible method for achieving efficient early diagnosis and treatment of lung cancer.

Earlier research has developed numerous lung cancer risk predictive models related to specific population demographics [12-41]; however, few of the predictive methods focused on nonsmoking women in mainland China [42]. Consequently, the development of lung cancer risk predicting tools for nonsmoking Chinese women according to consistently established risk factors in earlier studies has become a top goal [43]. Nevertheless, this goal is demanding and difficult. In contrast to the findings of lung cancer caused by tobacco, there are no identified risk variables for the progression of lung cancer in nonsmoking women. Although other risk factors were suggested, their relative importance varies greatly between geographical locations [3,4,44,45]. It was observed that the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) models, which included nearly 2000 Asian nonsmokers and only 7 cases of lung cancer, could be inapplicable to Asian nonsmokers [46]. Among the nonsmokers who participated in the PLCO study (n=65,711), none of them had a 6-year risk that was greater than 0.0151.

On the basis of the Cancer Screening Program in Urban China (CanSPUC), we created such a model [47]. In this paper, we aimed to create and internally validate a lung cancer risk predicting model for nonsmoking Chinese women, with the focus on established risk factors for lung cancer routinely available in general cancer-screening settings.

Methods

Data Source and Subjects

This retrospective, cross-sectional cohort study was carried out inside the scope of CanSPUC, a continuing statewide cancer-screening program for China's urban population. CanSPUC is designed to detect the 5 most common malignancies, including lung cancer, colorectal cancer, upper gastrointestinal cancer, liver cancer, and female breast cancer. The CanSPUC approach was detailed in previous studies [47,48]. All of the qualified subjects were questioned by highly skilled staff to gather information about their exposure to risk variables and to assess their cancer risk using a specific cancer risk score system. The household registration system was used in local communities to identify eligible permanent residents who were aged 40-74 years and asymptomatic for lung cancer with no history of cancer diagnosis. Individuals who were unable to give informed consent, had a medical disability and were unlikely to complete curative lung cancer surgery, had a history of lung cancer, had received treatment for or had evidence of any cancer within the past 5 years (with the exception of nonmelanoma skin cancer and most in situ carcinomas), or had symptoms suggestive of lung cancer (including unexplained weight loss of >7.5 kg within the past 12 months or unexplained hemoptysis) were not eligible to participate. In October 2013, CanSPUC was implemented in Henan province, which encompassed 8 cities with complete cancer registration data (Zhengzhou, Zhumadian, Anyang, Luoyang, Nanyang, Jiaozuo, Puyang, and Xinxiang). We examined the data collected over the first 6 years (from October 2013 to October 2019) in the Henan province for our research. Only nonsmoking women were included in this investigation.

Ethics Approval

The Ethics Committee of the Affiliated Cancer Hospital of Zhengzhou University and Henan Cancer Hospital evaluated and authorized the research (no.2021-KY-0028-001). Our sample was drawn from retrospective encounters documented in the electronic health record; these data were deidentified for both sets of analyses and did not require informed consent.

Outcome, Variables, and Measurements

All new cases of lung cancer were identified by matching with the cancer registry database in Henan province, China (by unique ID number), and histologically confirmed between October 1, 2013, and March 10, 2020. In Henan province, records of lung cancer are first submitted to local cancer registries by the hospitals and medical institutions and then submitted to the Henan Provincial Central Cancer Registry of China by the local cancer registries. The International Classification of Diseases, Tenth Revision was used to classify newly diagnosed lung malignancies by site. Lung cancers were identified by the International Classification of Diseases, Tenth Revision code of C33-C34. To find possible lung cancer risk variables, self-reported information was collected (Textbox 1).

Textbox 1. Self-reported information collected.

---

1. **Demographic factors, such as age, ethnicity, educational status, marital status, height, and body weight**

   - A low-educational level was defined as elementary school or less, a medium educational level as junior or senior high school, and a high-educational level as college or above

   - According to the "Guidelines for the Prevention and Control of Overweight and Obesity in Chinese Adults," BMI was dependent on the individual's height as well as weight and segmented to "<18.5 kg/m^2," "18.5-23.9 kg/m^2," "24.0-27.9 kg/m^2," and "≥28.0 kg/m^2" categories [49]

2. **Dietary habit**

   - Dietary intake of the following food in the past 2 years: vegetables (green-leafy plants and fungi except potatoes, sweet potatoes, and starches) <2.5 kg/week or ≥2.5 kg/week; roughage (all other grains except for white flour and rice) <0.5 kg/week or ≥0.5 kg/week; and fruit <1.25 kg/week or ≥1.25 kg/week. The weight of the food was measured prior to cooking

3. **Living environment, behavior, and habits**

   - Cooking oil fume exposure: exposure is considered as "none or a little" if chimneys, fume extractors, or smokeless pots were used during cooking; otherwise, it was considered as "a lot"

   - Passive smoking: regular living or employment in an enclosed area where people routinely smoke was regarded as "yes"; otherwise, it was regarded as "no"

   - Alcohol consumption: "current" referred to those who had consumed alcohol at least once weekly on average for more than 6 months; "former" referred to those who had ceased drinking; "never" referred to those who had never consumed alcohol

   - Physical activity: swimming; taijiquan, qigong, or walking; long range running; aerobics; sporting events (such as basketball, table tennis, badminton, etc); Yangko dancing or fast walking; and other physical activities (such as mountains climbing, rope jumping, and shuttlecock kicking). Subjects who engage in at least three sessions of practice weekly for a total of ≥90 minutes weekly were classified as engaging in "heavy physical activity"; otherwise, they were classified as engaging in "moderate or no physical activity"

4. **Psychology and emotions, such as a history of serious trauma and more than 6 months of mental depression**

   - Serious trauma was described as a major illness or death of a family member, family conflict and separation, significant loss of property, unexpected job loss, severe unexpected physical injury, violent danger, etc

5. **Comorbidities, such as chronic respiratory disease, tuberculosis, chronic bronchitis, emphysema, asthma bronchiectasis, hypertension, hyperlipidemia, and diabetes**

   - Every self-reported case of comorbidity required an evaluation from a professional medical facility

6. **Family history of lung cancer**

   - Whether first-degree relatives, second-degree relatives, or third-degree relatives had lung cancer or not

7. **Physiology and fertility**

   - Including age of menarche (<12 years or ≥12 years), menopause (yes or no), fertility status (yes or no), lactation status (yes or no), history of benign breast illness (yes or no), and a history of reproductive system surgery (yes or no)

Statistical Analysis

To contrast the profiles of those who have lung cancer and those without cancer, descriptive statistics, presented as percentages for categorical data, were used. Chi-square tests were used to examine the univariate correlation between baseline characteristics and lung cancer progression. For continuous variables, mean (SD) or median (IQR) were used.

In this investigation, the integrated model was applied to generate a nomogram to measure the 1-year, 3-year, and 5-year estimations of the lung cancer risk in the training set, according to the independently prognostic variables using the stepwise multivariable Cox regression (P_{entry}=.15 and P_{stay}=.10). The calibration curve was used to determine the nomogram's validity. By applying 50% and 84% quantiles, the risk predictions were grouped into the low-risk group, medium-risk group, and high-risk group, as suggested previously [50]. As per the risk prediction model, Kaplan-Meier curves were displayed for the low-risk group, medium-risk group, and high-risk group for lung cancer. The log-rank analyses were performed to compare the 3 curves. Receiver operating characteristic curves and the area under the curve (AUC) were used to quantify the prediction performance of 1-year, 3-year, and 5-year lung cancer risk estimations in the training set and validation set. By comparing observed and predicted probabilities, the bootstrap sampling method was used to evaluate the calibration of the current model.

All statistical analysis was carried out via R (version 4.0.3; R Foundation for Statistical Computing) and SAS (version 9.4; SAS Institute) software. The nomogram was drawn using the *rms* package. The receiver operating characteristic curves were drawn by using the *survivalROC* package. Using the *ggplot2*

XSL•FO

RenderX

package, a calibration curve was created. All of the tests were done using 2-tailed hypotheses, and $P<.05$ was determined to be statistically significant.

Results

Characteristics of the Study Population

This research consisted of a total of 151,834 qualified participants with an average age of 55.34 (SD 8.65) years. The subjects were randomly separated into a training set of 75,917 and a validation set of 75,917 (Figure 1). By March 2020, 204 lung cancer cases occurred within 151,834 subjects, resulting in an incident density of 42.24 per 100,000 person-years. Lung cancer cases were more frequent in those who were older ($P<.001$), had a history of respiratory illness ($P=.001$), had a first-degree family history of lung cancer ($P=.02$), and had menopause ($P<.001$). Extra features are shown in Table 1 and Table S1 in Multimedia Appendix 1.

Figure 1. Flow chart of participants included in this analysis.
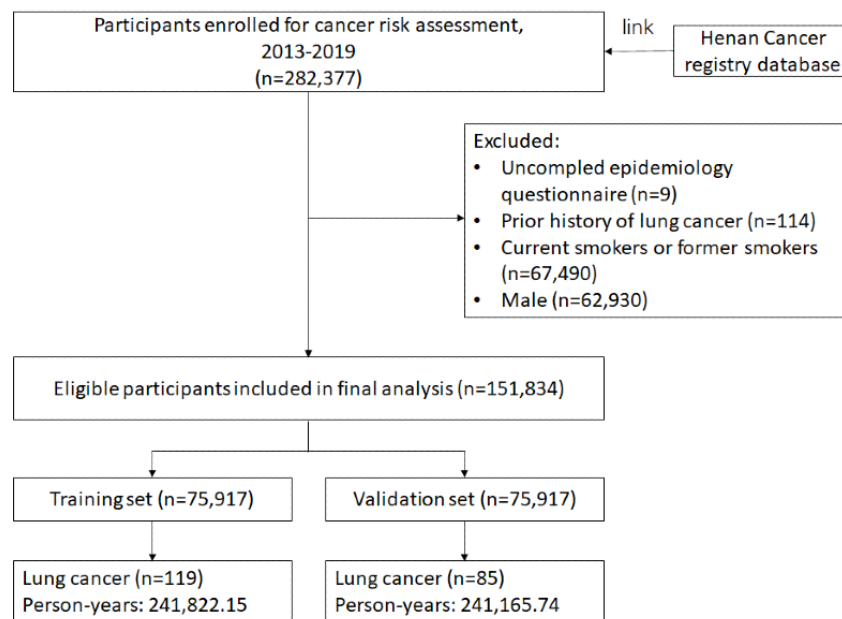
Table 1. Comparison of baseline characteristics between the non–lung cancer and lung cancer groups using chi-square test in the training set.

Variables	Total (N=75,917)[a]	Non–lung cancer[b]	Lung cancer[b]	χ^2 (df)	P value
All participants	75,917 (100)	75,798 (99.84)	119 (0.16)		
Person-years, median (IQR)	2.95 (1.73-4.83)	2.95 (1.73-4.83)	1.56 (0.83-2.38)		
Demographic characteristics					
Age (years), mean (SD)	55.37 (8.65)	55.36 (8.65)	60.37 (7.20)		
Age (years), n (%)				47.96 (6)	<.001 [c]
40-44	9226 (12.15)	9221 (99.95)	5 (0.05)		
45-49	13,558 (17.86)	13,551 (99.95)	7 (0.05)		
50-54	14,389 (18.95)	14,376 (99.91)	13 (0.09)		
55-59	11,857 (15.62)	11,838 (99.84)	19 (0.16)		
60-64	12,927 (17.03)	12,889 (99.71)	38 (0.29)		
65-69	10,181 (13.41)	10,151 (99.71)	30 (0.29)		
70-74	3779 (4.98)	3772 (99.81)	7 (0.19)		
Race, n (%)				0.05 (1)	.83
Han nationality	74,431 (98.04)	74,314 (99.84)	117 (0.16)		
Others	1486 (1.96)	1484 (99.87)	2 (0.13)		
Education[d], n (%)				0.12 (2)	.94
Low	16,139 (21.26)	16,115 (99.85)	24 (0.15)		
Medium	49,922 (65.76)	49,842 (99.84)	80 (0.16)		
High	9856 (12.98)	9841 (99.85)	15 (0.15)		
Marriage, n (%)				1.89 (1)	.17
Unmarried, divorce, or widowed	3193 (4.21)	3191 (99.94)	2 (0.06)		
Married	72,724 (95.79)	72,607 (99.84)	117 (0.16)		
BMI (kg/m^2), n (%)				1.84 (3)	.61
<18.5	1133 (1.49)	1133 (100)	0 (0)		
18.5-24.0	35,445 (46.69)	35,388 (99.84)	57 (0.16)		
24.0-28.0	30,729 (40.48)	30,681 (99.84)	48 (0.16)		
≥28.0	8610 (11.34)	8596 (99.84)	14 (0.16)		
Dietary habit, n (%)					
Vegetables intake (kg/week)				0.01 (1)	.92
≥2.5	39,282 (51.74)	39,221 (99.84)	61 (0.16)		
<2.5	36,635 (48.26)	36,577 (99.84)	58 (0.16)		
Fruit intake (kg/week)				0.71 (1)	.40
≥1.25	43,683 (57.54)	43,610 (99.83)	73 (0.17)		
<1.25	32,234 (42.46)	32,188 (99.86)	46 (0.14)		
Roughage intake (kg/week)				0.64 (1)	.42
≥0.5	51,713 (68.12)	51,636 (99.85)	77 (0.15)		
<0.5	24,204 (31.88)	24,162 (99.83)	42 (0.17)		
Living environment, behavior, and habits, n (%)					
Cooking oil fume exposure				0.05 (1)	.82
None or a little	65,819 (86.70)	65,715 (99.84)	104 (0.16)		

Variables	Total (N=75,917)[a]	Non–lung cancer[b]	Lung cancer[b]	χ^2 (df)	P value
A lot	10,098 (13.3)	10,083 (99.85)	15 (0.15)		
Passive smoking				0.63 (1)	.43
No	49,045 (64.6)	48,964 (99.83)	81 (0.17)		
Yes	26,872 (35.4)	26,834 (99.86)	38 (0.14)		
Alcohol drinking				0.11 (2)	.95
Never	71,567 (94.27)	71,454 (99.84)	113 (0.16)		
Current	3647 (4.8)	3642 (99.86)	5 (0.14)		
Former	703 (0.93)	702 (99.86)	1 (0.14)		
Physical activity				3.19 (1)	.07
Moderate or no	40,014 (52.71)	39,961 (99.87)	53 (0.13)		
Heavy	35,903 (47.29)	35,837 (99.82)	66 (0.18)		
Psychology and emotions, n (%)					
History of a severe trauma				1.22 (1)	.27
No	65,199 (85.88)	65,101 (99.85)	98 (0.15)		
Yes	10,718 (14.12)	10,697 (99.8)	21 (0.2)		
Mental depression for over 6 months				0.00 (1)	.98
No	64,379 (84.8)	64,278 (99.84)	101 (0.16)		
Yes	11,538 (15.2)	11,520 (99.84)	18 (0.16)		
Comorbidities, n (%)					
History of chronic respiratory disease				11.53 (1)	*.001*
No	64,070 (84.39)	63,983 (99.86)	87 (0.14)		
Yes	11,847 (15.61)	11,815 (99.73)	32 (0.27)		
History of tuberculosis				1.24 (1)	.27
No	74,895 (98.65)	74,779 (99.85)	116 (0.15)		
Yes	1022 (1.35)	1019 (99.71)	3 (0.29)		
History of chronic bronchitis				3.44 (*1*)	.06
No	66,728 (87.9)	66,630 (99.85)	98 (0.15)		
Yes	9189 (12.1)	9168 (99.77)	21 (0.23)		
History of emphysema				3.21 (1)	.07
No	75,204 (99.06)	75,088 (99.85)	116 (0.15)		
Yes	713 (0.94)	710 (99.58)	3 (0.42)		
History of asthma bronchiectasis				1.27 (1)	.26
No	73,473 (96.78)	73,360 (99.85)	113 (0.15)		
Yes	2444 (3.22)	2438 (99.75)	6 (0.25)		
History of hypertension				1.66 (1)	.20
No	60,976 (80.32)	60,886 (99.85)	90 (0.15)		
Yes	14,941 (19.68)	14,912 (99.81)	29 (0.19)		
History of hyperlipidemia				1.67 (1)	.20
No	63,309 (83.39)	63,215 (99.85)	94 (0.15)		
Yes	12,608 (16.61)	12,583 (99.8)	25 (0.2)		
History of diabetes				0.00 (1)	.98
No	70,767 (93.22)	70,656 (99.84)	111 (0.16)		

Variables	Total (N=75,917)[a]	Non–lung cancer[b]	Lung cancer[b]	χ^2 (df)	P value
Yes	5150 (6.78)	5142 (99.84)	8 (0.16)		
First-degree family history of lung cancer, n (%)				5.15 (1)	*.02*
No	69,955 (92.15)	69,852 (99.85)	103 (0.15)		
Yes	5962 (7.85)	5946 (99.73)	16 (0.27)		
Physiology and fertility, n (%)					
Age of menarche (years)				0.34 (1)	.56
<12	1910 (2.52)	1908 (99.9)	2 (0.1)		
≥12	74,007 (97.48)	73,890 (99.84)	117 (0.16)		
Menopause				29.26 (1)	*<.001*
No	26,927 (35.47)	26,913 (99.95)	14 (0.05)		
Yes	48,990 (64.53)	48,885 (99.79)	105 (0.21)		
Fertility status				1.67 (1)	.20
No	1047 (1.38)	1047 (100)	0 (0)		
Yes	74,870 (98.62)	74,751 (99.84)	119 (0.16)		
Lactation status				0.06 (1)	.80
No	4233 (5.58)	4227 (99.86)	6 (0.14)		
Yes	71,684 (94.42)	71,571 (99.84)	113 (0.16)		
History of benign breast disease				3.61 (1)	.06
No	53,977 (71.1)	53,883 (99.83)	94 (0.17)		
Yes	21,940 (28.9)	21,915 (99.89)	25 (0.11)		
History of reproductive system surgery				1.75 (1)	.19
No	60,480 (79.67)	60,391 (99.85)	89 (0.15)		
Yes	15,437 (20.33)	15,407 (99.81)	30 (0.19)		

[a]Percentages in this column have denominators of N=75,917.

[b]Percentages in these columns have the n value in the "Total" column in the same row as the denominators.

[c]Italicized values indicate statistical singificance.

[2]Low=primary school or below; medium=junior or senior high school; high=undergraduate degree or above.

Development of the Lung Cancer Risk Assessment Model

Table 2 displays the hazard ratios (HRs) with its 95% CI for every indicator. In the training set, age (≥55 years: HR 1.34, 95% CI 0.38-4.80; ≥60 years: HR 2.33, 95% CI 0.67-8.11; ≥65 years: HR 2.41, 95% CI 0.69-8.49; ≥70 years: HR 1.79, 95% CI 0.43-7.40), history of chronic respiratory disease (HR 1.94, 95% CI 1.24-3.04), first-degree family history of lung cancer (HR 1.60, 95% CI 0.91-2.83), menopause (HR 2.16, 95% CI 0.90-5.19), and history of benign breast disease (HR 0.58, 95% CI 0.36-0.94) were independent risk factors for lung cancer. Consequently, we applied these parameters to construct the model. We drew 1-year, 3-year, and 5-year risk–predicting nomograms for lung cancer (Figure 2A).

Table 2. Multivariable Cox regression prediction model of lung cancer risk in the training set.

Variables	β coefficient	SE	HR[a] (95% CI)	χ^2 (df)	P value
Age (years)					
40-44	N/A[b]	N/A	1.00	N/A	N/A
45-49	–0.19	0.59	0.83 (0.26-2.64)	0.10 (1)	.75
50-54	–0.06	0.62	0.94 (0.28-3.19)	0.01 (1)	.93
55-59	0.30	0.65	1.34 (0.38-4.80)	0.21 (1)	.65
60-64	0.85	0.64	2.33 (0.67-8.11)	1.78 (1)	.18
65-69	0.88	0.64	2.41 (0.69-8.49)	1.89 (1)	.17
70-74	0.58	0.72	1.79 (0.43-7.40)	0.65 (1)	.42
History of chronic respiratory disease					
No	N/A	N/A	1.00	N/A	N/A
Yes	0.66	0.23	1.94 (1.24-3.04)	8.45 (1)	*.004* [c]
First-degree family history of lung cancer					
No	N/A	N/A	1.00	N/A	N/A
Yes	0.47	0.29	1.60 (0.91-2.83)	2.63 (1)	.11
Menopause					
No	N/A	N/A	1.00	N/A	N/A
Yes	0.77	0.45	2.16 (0.90-5.19)	2.95 (1)	.09
History of benign breast disease					
No	N/A	N/A	1.00	N/A	N/A
Yes	–0.55	0.25	0.58 (0.36-0.94)	4.97 (1)	*.03*

[a]HR: hazard ratio.

[b]N/A: not applicable.

[c]Italicized values indicate statistical significance.

XSL•FO
RenderX

Figure 2. (A) Nomogram to calculate the personal 1-year, 3-year, and 5-year risk of lung cancer, and (B) the lung cancer incidence across different cancer risk categories.
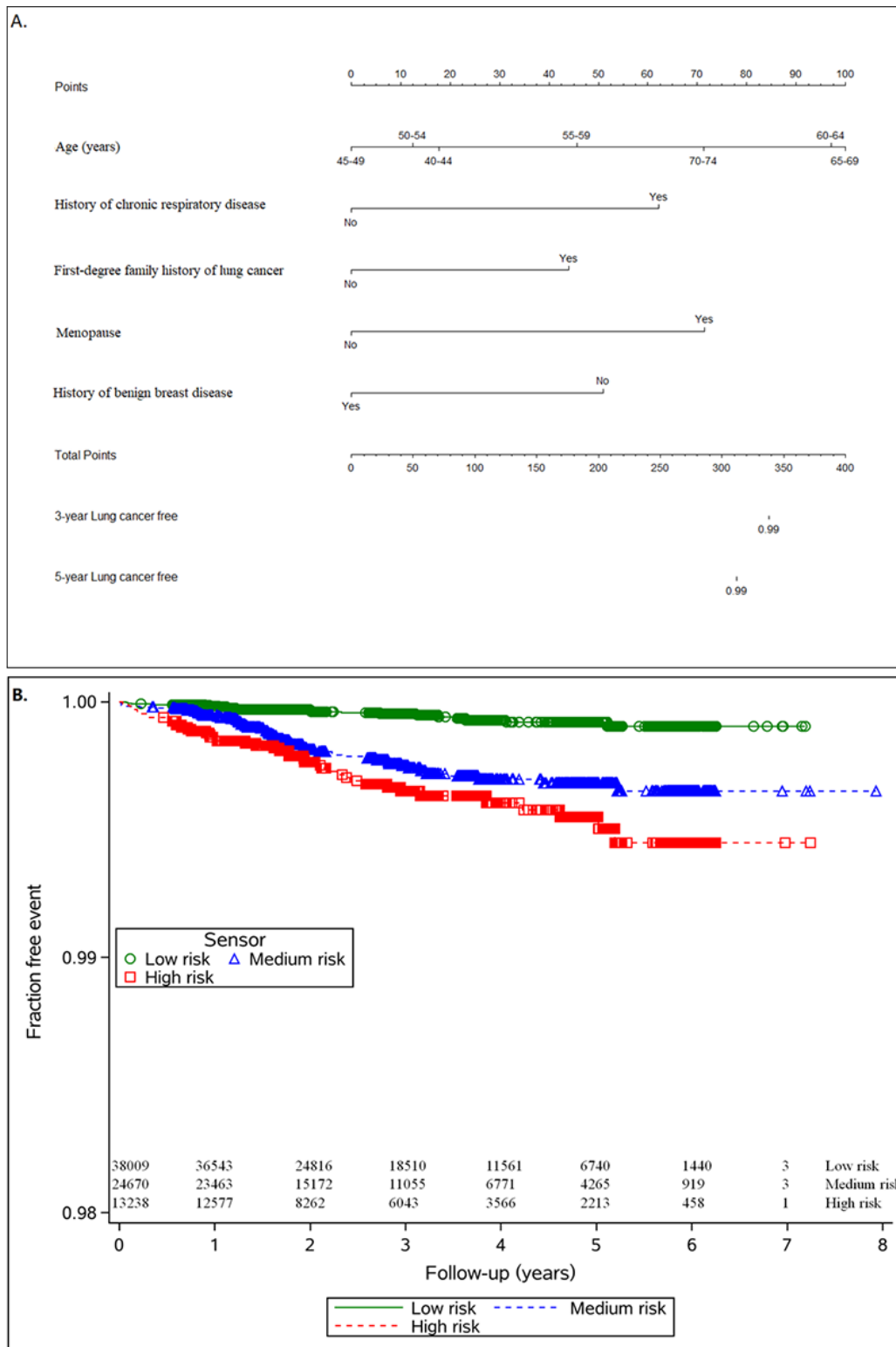
Predictive Performance of the Model

The risk predictions were categorized into low-risk, medium-risk, and high-risk categories, and a log-rank test revealed significant differences across the 3 groups (Figure 2B; $P<.001$).

By using this model, the AUC for 1-year, 3-year, and 5-year lung cancer risk in the training set was 0.762, 0.718, and 0.703, respectively. The model yielded a greater AUC for passive smokers (1-year: 0.787, 3-year: 0.715, and 5-year: 0.745) than for nonpassive smokers (1-year: 0.741, 3-year: 0.721, and 5-year: 0.689; Figure 3). Calibration was acceptable, with very similar observed and predicted hazards (Figure 4).

Figure 3. The receiver operating characteristic curves of prediction models in the training set. (A) Whole population; (B) Nonpassive smokers; (C) Passive smokers. AUC: area under the curve.
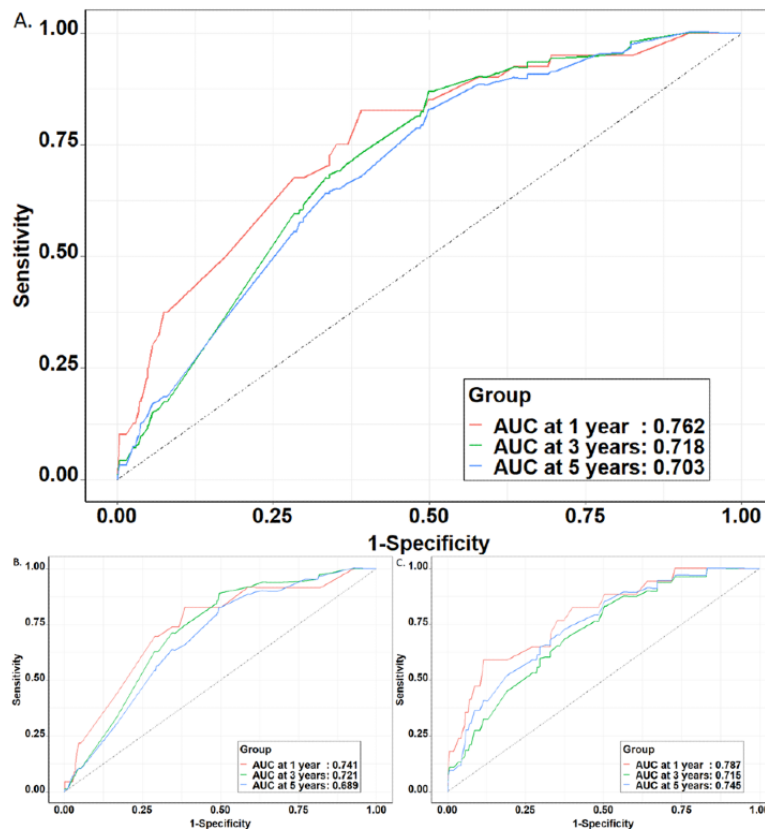
Figure 4. Calibration curves of the nomogram for (A) 1-year, (B) 3-year and (C) 5-year lung cancer–free rates in the training set.
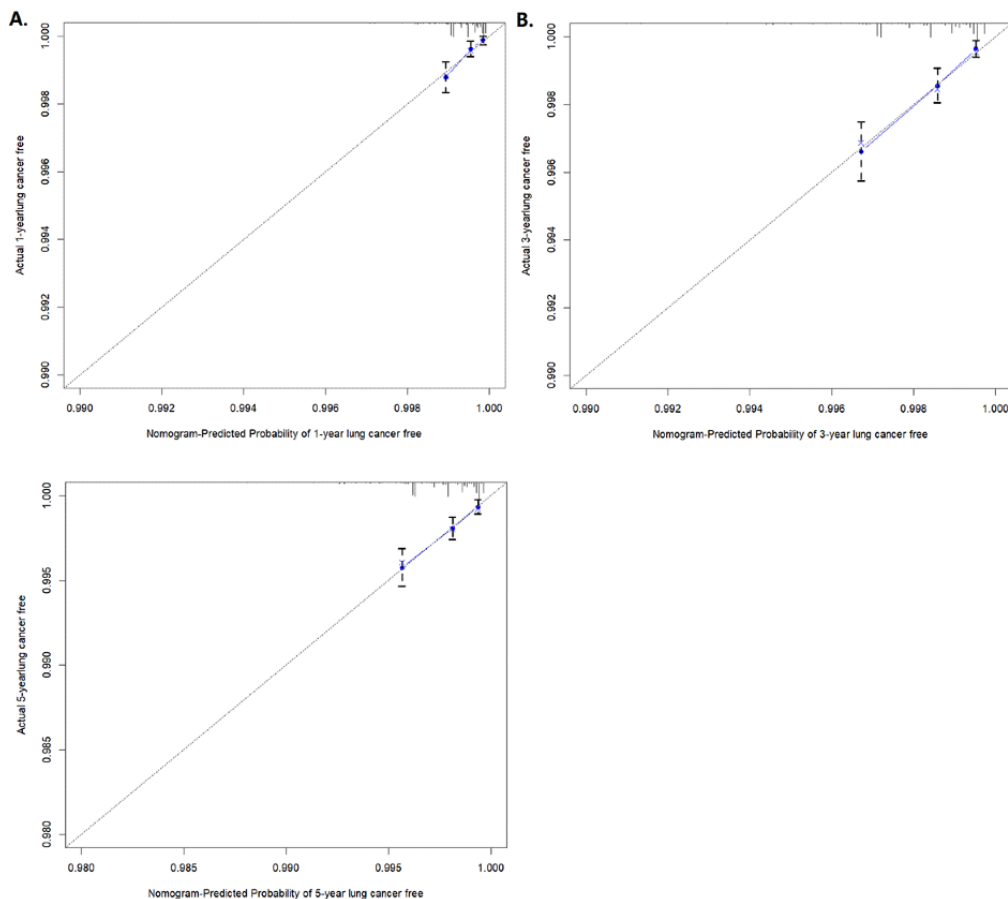
Validation of the Lung Cancer Risk Model

The model demonstrated a moderate predictive discrimination in the validation set, with AUCs of 0.646, 0.658, and 0.650 for 1-year, 3-year, and 5-year lung cancer risks, respectively (Figure S1 in Multimedia Appendix 1), and satisfactory calibration of relative risk (Figure S2 in Multimedia Appendix 1).

Discussion

We constructed and validated a simple risk predictive model internally for lung cancer in nonsmoking women relying on 5 commonly accessible factors such as demographics (age), comorbidities (chronic respiratory disease), first-degree family history of lung cancer, and fertility (menopause and history of benign breast disease). Our results showed that the model has moderate discriminatory accuracy and goodness of fit for both nonpassive smokers and passive smokers.

Multiple lung cancer risk variables were discovered for nonsmoking women, such as passive smoking [51,52], prior lung diseases (tuberculosis, chronic bronchitis, emphysema, and prior lung disorders [chronic obstructive pulmonary disease]) [53], indoor radon [54], cooking oil fume [55], and a family history of lung cancer [56]. The established risk variables for lung cancer, such as age, a family history of lung cancer, and a history of chronic respiratory disease, are similar to the findings. Age is the most important risk variable for lung cancer in nonsmoking women according to our survey, which found that the risk was more than 2.4 times higher in the age group of 65-69 years than 40-44 years.

Menopause was associated with an increased risk of developing lung cancer, with an overall odds ratio of 1.33 (95% CI 0.90-1.96), according to a pooled analysis of nested case-control data [57], which is consistent with our findings. Interestingly, we found that women with a history of benign breast disease were less likely to develop lung cancer, possibly because these women may be more careful about their lifestyle and diet after developing breast disease than those who did not. This finding will need to be validated in future studies.

Besides the accurate indicators, risk predicting models should achieve performance standards for discrimination (the differentiation capacity to distinguish lung cancer cases from control ones) and calibration (defined as the consistency between observed and predicted risk for lung cancer). Since 2010, the substantial growth in the numbers of investigations on lung cancer risk predicting models shows the necessity of using predictive models to drive population triage. Initially, models, such as the Bach model [12], Spitz model [13], Liverpool Lung Project model [14], and PLCO$_{M2012}$ model [58], emphasized the importance of applying the classic epidemiological risk variables, including age, smoking history, personal history of disease, and family history of cancer. To the best of our knowledge, this study is one of the few studies to model the prediction of lung cancer risk among nonsmoking Chinese women. Due to the fact that each model was created in a distinct population with different baseline risks and lengths of follow-up, it is challenging to compare the discriminating performance of risk predictive models. The discriminating ability of every model was quite equal, with C-statistics ranging between 0.72 and 0.86. Compared to prior research, our models showed comparable predictive performance.

In understanding our findings, certain strengths and limitations should be carefully considered. Our research is conducted on a large population-based cancer-screening program in mainland China, which is a strong point. In addition, the variables included in this model could be easily collected and updated without any imaging, sophisticated testing, or calculation. Furthermore, the model will be used as a convenient method to triage high-risk people among nonsmoking women, and it will be involving in public health initiatives, such as recommendations regarding the control of lung cancer in nonsmokers. Nonetheless, the statistics based on self-report may be susceptible to social acceptability bias as well as recall bias. Since data collection and quality control were carried out to a high standard, the vast volume of information can be relied upon. Second, the performance of our risk-predicting model was not validated against an external data set before it was used. The findings of the internal calibration, on the other hand, suggest that this model will function satisfactorily when applied to a variety of populations.

In conclusion, a large-scale lung cancer–screening project in China served as the foundation for the creation and internal calibration of a straightforward risk predictive model for lung cancer in nonsmoking women. The model has moderate discrimination and could be used as a tool for triaging high-risk people to prevent lung cancer in nonsmoking women. To validate the concept in external populations, additional prospective studies are needed.

Data Availability

The data sets for this manuscript are not publicly available because all our data are under regulation of both the National Cancer Center of China and The Affiliated Cancer Hospital of Zhengzhou University and Henan Cancer Hospital. Requests to access the data sets should be directed to SZ.

XSL•FO

RenderX

Authors' Contributions

LG and SZ contributed to the conception and design. LG and L Zheng contributed to statistical analyses. LG, QM, L Zheng, QC, YL, HX, RK, L Zhang, SL, XS, and SZ contributed to data acquisition and data interpretation. LG drafted the article. All authors revised the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary table and figures.
[DOCX File , 169 KB-Multimedia Appendix 1]

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]

2. Zhang SW, Zheng RS, Yang ZX, Zeng HM, Sun KX, Gu XY, et al. Trend analysis on incidence and age at diagnosis for lung cancer in cancer registration areas of China, 2000-2014. Article in Chinese. Zhonghua Yu Fang Yi Xue Za Zhi 2018 Jun 06;52(6):579-585. [doi: 10.3760/cma.j.issn.0253-9624.2018.06.005] [Medline: 29886678]

3. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. Nat Rev Cancer 2007 Oct;7(10):778-790. [doi: 10.1038/nrc2190] [Medline: 17882278]

4. Couraud S, Zalcman G, Milleron B, Morin F, Souquet PJ. Lung cancer in never smokers--a review. Eur J Cancer 2012 Jun;48(9):1299-1311. [doi: 10.1016/j.ejca.2012.03.007] [Medline: 22464348]

5. Scagliotti GV, Longo M, Novello S. Nonsmall cell lung cancer in never smokers. Curr Opin Oncol 2009 Mar;21(2):99-104. [doi: 10.1097/CCO.0b013e328321049e] [Medline: 19532009]

6. Shi JF, Wang L, Wu N, Li JL, Hui ZG, Liu SM, LuCCRES Group. Clinical characteristics and medical service utilization of lung cancer in China, 2005-2014: overall design and results from a multicenter retrospective epidemiologic survey. Lung Cancer 2019 Feb;128:91-100. [doi: 10.1016/j.lungcan.2018.11.031] [Medline: 30642458]

7. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries. Lancet Glob Health 2018 May;6(5):e555-e567 [FREE Full text] [doi: 10.1016/S2214-109X(18)30127-X] [Medline: 29653628]

8. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. Chest 2017 Jan;151(1):193-203. [doi: 10.1016/j.chest.2016.10.010] [Medline: 27780786]

9. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 04;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]

10. Li Q, Hsia J, Yang G. Prevalence of smoking in China in 2010. N Engl J Med 2011 Jun 23;364(25):2469-2470. [doi: 10.1056/NEJMc1102459] [Medline: 21696322]

11. Cepeda-Benito A, Doogan NJ, Redner R, Roberts ME, Kurti AN, Villanti AC, et al. Trend differences in men and women in rural and urban U.S. settings. Prev Med 2018 Dec;117:69-75. [doi: 10.1016/j.ypmed.2018.04.008] [Medline: 29627511]

12. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst 2003 Mar 19;95(6):470-478. [doi: 10.1093/jnci/95.6.470] [Medline: 12644540]

13. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst 2007 May 02;99(9):715-726. [doi: 10.1093/jnci/djk153] [Medline: 17470739]

14. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008 Jan 29;98(2):270-276 [FREE Full text] [doi: 10.1038/sj.bjc.6604158] [Medline: 18087271]

15. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, et al. Development and validation of a lung cancer risk prediction model for African-Americans. Cancer Prev Res (Phila) 2008 Sep;1(4):255-265 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-08-0082] [Medline: 19138969]

16. Spitz MR, Etzel CJ, Dong Q, Amos CI, Wei Q, Wu X, et al. An expanded risk prediction model for lung cancer. Cancer Prev Res (Phila) 2008 Sep;1(4):250-254 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-08-0060] [Medline: 19138968]

17. Young RP, Hopkins RJ, Hay BA, Epton MJ, Mills GD, Black PN, et al. Lung cancer susceptibility model based on age, family history and genetic variants. PLoS One 2009 Apr 23;4(4):e5302 [FREE Full text] [doi: 10.1371/journal.pone.0005302] [Medline: 19390575]

18. D'Amelio AMJ, Cassidy A, Asomaning K, Raji OY, Duffy SW, Field JK, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. Br J Cancer 2010 Jul 27;103(3):423-429 [FREE Full text] [doi: 10.1038/sj.bjc.6605759] [Medline: 20588271]

19. Raji OY, Agbaje OF, Duffy SW, Cassidy A, Field JK. Incorporation of a genetic factor into an epidemiologic model for prediction of individual risk of lung cancer: the Liverpool Lung Project. Cancer Prev Res (Phila) 2010 May;3(5):664-669. [doi: 10.1158/1940-6207.CAPR-09-0141] [Medline: 20424129]

20. Maisonneuve P, Bagnardi V, Bellomi M, Spaggiari L, Pelosi G, Rampinelli C, et al. Lung cancer risk prediction to select smokers for screening CT--a model based on the Italian COSMOS trial. Cancer Prev Res (Phila) 2011 Nov;4(11):1778-1789. [doi: 10.1158/1940-6207.CAPR-11-0026] [Medline: 21813406]

21. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, et al. Lung cancer risk prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial models and validation. J Natl Cancer Inst 2011 Jul 06;103(13):1058-1068 [FREE Full text] [doi: 10.1093/jnci/djr173] [Medline: 21606442]

22. Tammemagi MC, Lam SC, McWilliams AM, Sin DD. Incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction. Cancer Prev Res (Phila) 2011 Apr;4(4):552-561. [doi: 10.1158/1940-6207.CAPR-10-0183] [Medline: 21411501]

23. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Østergaard JN, et al. A risk model for lung cancer incidence. Cancer Prev Res (Phila) 2012 Jun;5(6):834-846 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0237] [Medline: 22496387]

24. Li H, Yang L, Zhao X, Wang J, Qian J, Chen H, et al. Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model. BMC Med Genet 2012 Dec 10;13(1):118 [FREE Full text] [doi: 10.1186/1471-2350-13-118] [Medline: 23228068]

25. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. Ann Intern Med 2012 Aug 21;157(4):242-250 [FREE Full text] [doi: 10.7326/0003-4819-157-4-201208210-00004] [Medline: 22910935]

26. Park S, Nam BH, Yang HR, Lee JA, Lim H, Han JT, et al. Individualized risk prediction model for lung cancer in Korean men. PLoS One 2013 Feb 07;8(2):e54823 [FREE Full text] [doi: 10.1371/journal.pone.0054823] [Medline: 23408946]

27. Spitz MR, Amos CI, Land S, Wu X, Dong Q, Wenzlaff AS, et al. Role of selected genetic variants in lung cancer risk in African Americans. J Thorac Oncol 2013 Apr;8(4):391-397 [FREE Full text] [doi: 10.1097/JTO.0b013e318283da29] [Medline: 23454887]

28. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013 Feb 21;368(8):728-736 [FREE Full text] [doi: 10.1056/NEJMoa1211776] [Medline: 23425165]

29. Veronesi G, Maisonneuve P, Rampinelli C, Bertolotti R, Petrella F, Spaggiari L, et al. Computed tomography screening for lung cancer: results of ten years of annual screening and validation of cosmos prediction model. Lung Cancer 2013 Dec;82(3):426-430. [doi: 10.1016/j.lungcan.2013.08.026] [Medline: 24099665]

30. El-Zein RA, Lopez MS, D'Amelio AM, Liu M, Munden RF, Christiani D, et al. The cytokinesis-blocked micronucleus assay as a strong predictor of lung cancer: extension of a lung cancer risk prediction model. Cancer Epidemiol Biomarkers Prev 2014 Nov;23(11):2462-2470 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-0462] [Medline: 25172871]

31. Li K, Hüsing A, Sookthai D, Bergmann M, Boeing H, Becker N, et al. Selecting high-risk individuals for lung cancer screening: a prospective evaluation of existing risk models and eligibility criteria in the German EPIC cohort. Cancer Prev Res (Phila) 2015 Sep;8(9):777-785. [doi: 10.1158/1940-6207.CAPR-14-0424] [Medline: 26076698]

32. Marcus MW, Chen Y, Raji OY, Duffy SW, Field JK. LLPi: Liverpool Lung Project risk prediction model for lung cancer incidence. Cancer Prev Res (Phila) 2015 Jun;8(6):570-575. [doi: 10.1158/1940-6207.CAPR-14-0438] [Medline: 25873368]

33. Wang X, Ma K, Cui J, Chen X, Jin L, Li W. An individual risk prediction model for lung cancer based on a study in a Chinese population. Tumori 2015 Feb 14;101(1):16-23. [doi: 10.5301/tj.5000205] [Medline: 25702657]

34. Marcus MW, Raji OY, Duffy SW, Young RP, Hopkins RJ, Field JK. Incorporating epistasis interaction of genetic susceptibility single nucleotide polymorphisms in a lung cancer risk prediction model. Int J Oncol 2016 Jul;49(1):361-370 [FREE Full text] [doi: 10.3892/ijo.2016.3499] [Medline: 27121382]

35. Wang X, Ma K, Chi L, Cui J, Jin L, Hu J, et al. Combining telomerase reverse transcriptase genetic variant rs2736100 with epidemiologic factors in the prediction of lung cancer susceptibility. J Cancer 2016;7(7):846-853 [FREE Full text] [doi: 10.7150/jca.13437] [Medline: 27162544]

36. Wu X, Wen CP, Ye Y, Tsai M, Wen C, Roth JA, et al. Personalized risk assessment in never, light, and heavy smokers in a prospective cohort in Taiwan. Sci Rep 2016 Nov 02;6:36482 [FREE Full text] [doi: 10.1038/srep36482] [Medline: 27805040]

37. Muller DC, Johansson M, Brennan P. Lung cancer risk prediction model incorporating lung function: development and validation in the UK Biobank prospective cohort study. J Clin Oncol 2017 Mar 10;35(8):861-869 [FREE Full text] [doi: 10.1200/JCO.2016.69.2467] [Medline: 28095156]

38. Weber M, Yap S, Goldsbury D, Manners D, Tammemagi M, Marshall H, et al. Identifying high risk individuals for targeted lung cancer screening: independent validation of the PLCO risk prediction tool. Int J Cancer 2017 Jul 15;141(2):242-253 [FREE Full text] [doi: 10.1002/ijc.30673] [Medline: 28249359]

39. Charvat H, Sasazuki S, Shimazu T, Budhathoki S, Inoue M, Iwasaki M, JPHC Study Group. Development of a risk prediction model for lung cancer: The Japan Public Health Center-based Prospective Study. Cancer Sci 2018 Mar 21;109(3):854-862 [FREE Full text] [doi: 10.1111/cas.13509] [Medline: 29345859]

40. Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs E, Jemal A, et al. Implications of nine risk prediction models for selecting ever-smokers for computed tomography lung cancer screening. Ann Intern Med 2018 Jul 03;169(1):10-19 [FREE Full text] [doi: 10.7326/M17-2701] [Medline: 29800127]

41. Markaki M, Tsamardinos I, Langhammer A, Lagani V, Hveem K, Røe OD. A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types: a HUNT study. EBioMedicine 2018 May;31:36-46 [FREE Full text] [doi: 10.1016/j.ebiom.2018.03.027] [Medline: 29678673]

42. Lyu ZY, Li N, Chen SH, Wang G, Tan FW, Feng XS, et al. Exploratory research on developing lung cancer risk prediction model in female non-smokers. Article in Chinese. Zhonghua Yu Fang Yi Xue Za Zhi 2020 Nov 06;54(11):1261-1267. [doi: 10.3760/cma.j.cn112150-20200805-01093] [Medline: 33147927]

43. Lam S. Lung cancer screening in never-smokers. J Thorac Oncol 2019 Mar;14(3):336-337 [FREE Full text] [doi: 10.1016/j.jtho.2018.12.019] [Medline: 30782380]

44. Toh C, Gao F, Lim W, Leong S, Fong K, Yap S, et al. Never-smokers with lung cancer: epidemiologic evidence of a distinct disease entity. J Clin Oncol 2006 May 20;24(15):2245-2251. [doi: 10.1200/JCO.2005.04.8033] [Medline: 16710022]

45. Sisti J, Boffetta P. What proportion of lung cancer in never-smokers can be attributed to known risk factors? Int J Cancer 2012 Jul 15;131(2):265-275 [FREE Full text] [doi: 10.1002/ijc.27477] [Medline: 22322343]

46. Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. PLoS Med 2014 Dec;11(12):e1001764 [FREE Full text] [doi: 10.1371/journal.pmed.1001764] [Medline: 25460915]

47. Guo LW, Chen Q, Shen YC, Meng QC, Zheng LY, Wu Y, et al. Evaluation of a low-dose computed tomography lung cancer screening program in Henan, China. JAMA Netw Open 2020 Nov 02;3(11):e2019039 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19039] [Medline: 33141158]

48. Guo L, Zhang S, Liu S, Zheng L, Chen Q, Cao X, et al. Determinants of participation and detection rate of upper gastrointestinal cancer from population-based screening program in China. Cancer Med 2019 Nov;8(16):7098-7107 [FREE Full text] [doi: 10.1002/cam4.2578] [Medline: 31560836]

49. Chen C, Lu FC, Department of Disease Control Ministry of Health, PR China. The guidelines for prevention and control of overweight and obesity in Chinese adults. Biomed Environ Sci 2004;17 Suppl:1-36. [Medline: 15807475]

50. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013 Mar 06;13:33 [FREE Full text] [doi: 10.1186/1471-2288-13-33] [Medline: 23496923]

51. Hackshaw AK, Law MR, Wald NJ. The accumulated evidence on lung cancer and environmental tobacco smoke. BMJ 1997 Oct 18;315(7114):980-988 [FREE Full text] [doi: 10.1136/bmj.315.7114.980] [Medline: 9365295]

52. Stayner L, Bena J, Sasco AJ, Smith R, Steenland K, Kreuzer M, et al. Lung cancer risk and workplace exposure to environmental tobacco smoke. Am J Public Health 2007 Mar;97(3):545-551. [doi: 10.2105/AJPH.2004.061275] [Medline: 17267733]

53. Brenner DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. PLoS One 2011 Mar 31;6(3):e17479 [FREE Full text] [doi: 10.1371/journal.pone.0017479] [Medline: 21483846]

54. Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ 2005 Jan 29;330(7485):223 [FREE Full text] [doi: 10.1136/bmj.38308.477650.63] [Medline: 15613366]

55. Zhao Y, Wang S, Aunan K, Seip HM, Hao J. Air pollution and lung cancer risks in China--a meta-analysis. Sci Total Environ 2006 Aug 01;366(2-3):500-513. [doi: 10.1016/j.scitotenv.2005.10.010] [Medline: 16406110]

56. Lissowska J, Foretova L, Dabek J, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, et al. Family history and lung cancer risk: international multicentre case-control study in Eastern and Central Europe and meta-analyses. Cancer Causes Control 2010 Jul;21(7):1091-1104. [doi: 10.1007/s10552-010-9537-2] [Medline: 20306329]

57. Jin K, Hung RJ, Thomas S, Le Marchand L, Matsuo K, Seow A, et al. Hormonal factors in association with lung cancer among Asian women: a pooled analysis from the International Lung Cancer Consortium. Int J Cancer 2021 May 01;148(9):2241-2254 [FREE Full text] [doi: 10.1002/ijc.33405] [Medline: 33210298]

58. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013 Feb 21;368(8):728-736 [FREE Full text] [doi: 10.1056/NEJMoa1211776] [Medline: 23425165]

Abbreviations

AUC: area under the curve
CanSPUC: Cancer Screening Program in Urban China
HR: hazard ratio
PLCO: Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial