scientific reports

Check for updates

OPEN

Predictive performance of risk prediction models for lung cancer incidence in Western and Asian countries: a systematic review and meta-analysis

Yah Ru Juang[1], Lina Ang[1] & Wei Jie Seow[1,2]✉

Numerous prediction models have been developed to identify high-risk individuals for lung cancer screening, with the aim of improving early detection and survival rates. However, no comprehensive review or meta-analysis has assessed the performance of these models across different sociocultural contexts. Therefore, this review systematically examines the performance of lung cancer risk prediction models in Western and Asian populations. PubMed and EMBASE were searched from inception through January 2023. Studies published in English that proposed a validated model on human populations with well-defined predictive performances were included. Two reviewers independently screened the titles and abstracts, and the Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess study quality. A random-effects meta-analysis was performed, and a 95% confidence interval (CI) for model performance was reported. Between-study heterogeneity was adjusted for using the Hartung-Knapp-Sidik-Honkman test. A total of 54 studies were included, with 42 from Western countries and 12 from Asian countries. Most Western studies focused on ever-smokers (19/42; 45.2%) and the general population (17/42; 40.5%), and only two Asian studies developed models exclusively for never-smokers. Across both Western and Asian prediction models, the three most consistently included risk factors were age, sex, and family cancer history. In 45.2% (19/42) of Western and 50.0% (6/12) of Asian studies, models incorporated both traditional risk factors and biomarkers. In addition, 14.8% (8/54) of the studies directly compared biomarker-based models with those incorporating only traditional risk factors, demonstrating improved discrimination. Machine-learning algorithms were applied in eight Western models and two Asian models. External validation of $PLCO_{M2012}$ (AUC = 0.748; 95% CI: 0.719–0.777) outperformed other prediction models, such as Bach (AUC = 0.710; 95% CI: 0.674–0.745) and Spitz models (AUC = 0.698; 95% CI: 0.640–0.755). Despite showing promising results, the majority of Asian risk models in our study lack external validation. Our review also highlights a significant gap in prediction models for never-smokers. Future research should focus on externally validating existing Asian models or incorporating relevant Asian risk factors into widely used Western models ($PLCO_{M2012}$) to better account for unique risk profiles and lung cancer progression patterns in Asian populations.

Keywords Lung cancer, Lung neoplasm, Risk prediction models, Cancer screening

For decades, lung cancer has been the leading cause of cancer-related deaths worldwide[1]. Given that early stage lung cancer is often asymptomatic, most new cases are diagnosed at an advanced-stage[2]. The 5-year relative survival rate for patients with advanced-stage is 6%, compared to 61% for those with localized lung cancer[3]. Therefore, early detection could significantly reduce lung cancer mortality. Extensive randomized trials, including the 2013 National Lung Screening Trial (NLST)[4] and the 2020 NELSON trial[5], have demonstrated the efficacy of low-dose computed tomographic (LDCT) screening as an effective tool for early lung cancer detection. Currently, the U.S. Preventive Services Task Force (USPSTF) recommends screening for high-risk

[1]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore 117549, Singapore. [2]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore 119228, Singapore. ✉email: ephswj@nus.edu.sg

individuals, defined as smokers aged 50 to 80 years with at least 20 pack-years of smoking history, who currently smoke or have quit within the last 15 years[6]. It is important to note that the USPSTF guidelines were established prior to the findings of the NELSON trial and differ from the original NLST-based criteria. Conversely, due to the assumption of low lung cancer risk, light- and never-smokers are not eligible for LDCT screening. However, the rising incidence of lung cancer among never-smokers, particularly among women in Asian countries such as South Korea, Taiwan, and Singapore[7–9], highlights the need to re-evaluate current screening guidelines.

Over the past decades, considerable efforts have been made to develop prediction models that identify high-risk groups for screening and improve screening outcomes. For example, the Bach model[10], developed from the β-Carotene and Retinol Efficacy Trial (CARET), and the Liverpool Lung Project (LLP) model[11], tailored for the UK population, extend beyond traditional risk factors to provide more precise predictions in specific Western contexts. Notably, the $PLCO_{M2012}$model[12], developed from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), has demonstrated superior discriminatory capabilities compared to USPSTF2013 criteria and is recommended for use prior to LDCT screening referral. For instance, using a threshold of 1.70% over 6 years, the $PLCO_{M2012}$ model identified 162 out of 171 lung cancer cases (94.7%; 95% CI: 90.2–97.6), significantly outperforming the USPSTF criteria, which detected only 135 cases (78.9%; 95% CI: 72.1–84.8).

Despite advances in predictive modeling, significant gaps remain in our understanding of how these models perform across diverse sociocultural contexts. For example, existing lung cancer prediction models developed for Western populations may not be directly applicable to Asia contexts due to cultural and socioeconomic differences, including variations in healthcare systems, smoking habits, environmental exposures, and genetic factors[13–15]. These factors can impact the applicability and accuracy of such models in non-Western settings. The increasing lung cancer incidence among never-smokers, particularly women, in Asian regions is likely driven by factors such as family history of lung cancer[16,17], exposure to environmental carcinogens[18–21], dietary habits[22,23], and genetic predispositions like epidermal growth factor receptor (*EGFR*) mutations[24,25], underscoring the need for region-specific risk assessments. Given that lung cancer incidence among never-smokers is considerably higher in Asian countries as compared to Western countries, models like $PLCO_{M2012}$, which heavily weigh smoking history, may not perform as well[26]. As a result, other models have been developed to improve screening selection, particularly for light- and never-smokers, compared to existing guidelines[27,28]. Notably, the TNSF-SQ model has successfully identified 27.03% of high-risk Taiwanese never-smoking females (NSFs) aged 55 to 74 years for LDCT screening, indicating its potential to enhance early lung cancer detection among NSFs[27]. However, given the distinct epidemiological profile of lung cancer in Asian countries compared to Western nations, it is crucial to rigorously compare existing models across different sociocultural contexts to ensure more accurate and culturally sensitive risk predictions that address the diverse needs of populations worldwide.

In practice, the predictive performance of models is typically assessed using the observed/expected (O/E) statistic[29], the Hosmer-Lemeshow (H-L) goodness-of-fit test[30], the C-statistic[31], or the area under the receiving operating characteristic (ROC) curve (AUC)[32]. The O/E statistic measures model calibration by comparing the expected (E) and observed (O) number of events. A model with a good calibration will have a value closer to 1[29]. The H-L goodness-of-fit measures calibration, and small p-values (<0.05) indicate that the model is not a good fit[30]. Model discrimination, which refers to the ability to distinguish between two groups (e.g., presence and absence of a condition), is quantified by metrics such as the C-statistic or AUC. The C-statistic, comparable to AUC in time-fixed binary classification, provides an aggregated measure of the model's performance across all thresholds. Values between 0.9 and 1 are considered excellent, 0.8 to 0.9 good, 0.7 to 0.8 fair or moderate, 0.6 to 0.7 poor, and 0.5 and below as failed[33,34]. In time-to-event analyses, however, the C-statistic incorporates the temporal aspect of events, whereas the AUC does not. Therefore, it is important to carefully differentiate between these metrics when comparing model types.

Previous reviews by Gray et al.[35], Toumazis et al.[36], and Wu et al.[37] have provided valuable insights into lung cancer risk prediction models. Gray et al.[35] evaluated the performance of these models across various populations, primarily focusing on Western populations, with less emphasis on Asian populations. Toumzais et al.[36] reviewed risk-based lung cancer screening strategies, focusing on how integrating risk prediction models can enhance screening programs, while Wu et al.[37] systematically examined the performance and applicability of models for predicting malignancy in pulmonary nodules. Our study builds on these insights by offering a broader and more comprehensive evaluation of risk prediction models, extending beyond nodule-based models to include diverse geographic and sociocultural contexts, specifically comparing Western and Asian countries. Unlike previous reviews, we provided a critical appraisal of the risk of bias (ROB) and performed a meta-analysis of models frequently subjected to external validation, quantitatively synthesizing their performance across diverse settings and populations. The absence of such a comprehensive review and ROB quality assessment leaves uncertainty regarding the applicability of these models across diverse sociocultural contexts. In predictive modeling for lung cancer, some studies focus on mortality-based models, which aim to identify individuals at the highest risk of mortality, potentially maximizing life-saving interventions. However, this review specifically focuses on incidence-based models, which identify individuals at risk of developing lung cancer. These models enable early detection through screening, with the ultimate goal of reducing mortality by diagnosing cases at a more treatable stage. Our review systematically examines the performance of incidence-based models in both Western and Asian populations, stratified by target population, study design, validation type, model development approaches, and model risk factors, as well as meta-analyzes the predictive performance of models validated in external cohorts. This is critical, as models demonstrating good calibration and discrimination across diverse populations and settings have the potential to improve future screening programs.

Methods

This systematic review and meta-analysis was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) 2020 checklist[38].

This review has been registered in the International Prospective Register of Systematic Reviews (PROSPERO) (Registration ID: CRD42022347087). In this paper, we use various specialized terms and acronyms related to our study. Definitions for these terms can be found in the glossary (Table S1).

Eligibility criteria

Studies that meet all the following criteria will be included: (a) full articles published in English; (b) lung cancer incidence risk prediction models developed and/or validated (internally and/or externally) using a human population; (c) reported a well-defined statistical method for model development and their predictive performance (e.g., C-statistic/AUC values with 95% confidence interval (CI)) [or standard error (SE) such that the 95% CI could be estimated]. Studies that meet any of the following criteria will be excluded: (a) non-clinical studies; (b) not related to lung cancer risk prediction models; (c) duplicated studies; (d) incomplete description of the development and predictive performance of prediction model from which estimates could be computed; (e) informal publications (e.g., conference abstracts, summary, editorial, or commentary); (f) not published in peer-reviewed journals. These criteria were outlined in Table S2, which provided a detailed breakdown of the inclusion and exclusion criteria used to assess the relevance of the studies. Discrepancies between reviewers (JYR and LA) regarding study eligibility were resolved through discussion. When consensus could not be reached, a third reviewer (SWJ) was consulted to make the final decision.

Information sources

We searched the following electronic bibliographic databases: Embase and PubMed, which encompass literature from MEDLINE and PubMed Central (PMC), due to their extensive coverage of biomedical, clinical, and scientific research, for relevant articles published from inception to January 2023. No restrictions were applied to the publication date. Backward citation searching was performed on the reference lists of selected full-text articles to identify potentially relevant studies that were not initially captured in the literature search phase.

Search strategy

We conducted a preliminary search across two databases using combinations of relevant keywords such as "lung cancer," "lung neoplasm," and "risk prediction models" to identify all relevant articles published in English. Boolean operators (AND, OR) were used to combine terms and refine the search. The full search strategy with keywords and exact search terms adapted to the requirements of the databases used is presented in Table S3.

Study selection

All studies retrieved from respective databases were imported into EndNote version 20[39] and duplicate records were removed. Two reviewers (JYR and LA) independently screened article titles and abstracts in Rayyan software[40] and those that did not meet the eligibility criteria were discarded. Full texts of potentially relevant articles were subsequently retrieved and evaluated for inclusion. Any discrepancies were resolved through discussion with a third reviewer (SWJ). Backward citation searching was performed to identify additional relevant publications, which were subjected to the same screening and selection process. Two reviewers independently selected articles, extracted data, and assessed the quality of included studies and overall evidence.

Data extraction

Data were independently extracted by two reviewers (JYR and LA), and discrepancies were resolved through discussion with a third reviewer (SWJ). An electronic, standardized data extraction form was developed using Microsoft Excel to extract data from the full-text articles. For each included study, we extracted the following data: study characteristics, i.e., first author, publication year, geographical location, study design, study population; model variables i.e., model type, variable selection approach, internal validation method, risk factors, predictive performance in terms of discrimination (i.e., C-statistic, AUC) and calibration (i.e., O/E statistic, H-L goodness-of-fit test) with 95% CI. In addition, sensitivity and specificity rates of the prediction models at specific risk thresholds were recorded.

Risk of bias and applicability assessment

PROBAST[41] was used to assess study quality, including the ROB introduced and its applicability based on study sample, predictive factors, and the outcome measures. PROBAST provides a checklist of four domains: participants, predictors, outcome, and analysis, with signaling questions which are answered with "yes" (Y), "probably yes" (PY), "no" (N), "probably no" (PN) or "no information" (NI). Specifically, the participant subdomain addresses the adequacy of participant selection and data sources, noting that case-control studies may misrepresent true risk due to sampling biases, and models must be developed in populations similar to the intended use. The predictor selection subdomain evaluates the consistency and blinding of predictor definitions and measurements, emphasizing the need for predictors available at the time of prediction. The outcome subdomain focuses on accurate, unbiased outcome definitions, avoiding incorporation bias, and ensuring consistency across participants and centres. Finally, the analysis subdomain critiques statistical methods, including issues like overfitting, inappropriate dichotomization of continuous predictors, handling of missing data, and reporting of regression coefficients. Studies were rated as "low ROB" when a domain had all signaling questions rated as "yes" or "probably yes." A rating of "no" or "probably no" for one or more questions indicates the potential for bias; however, specific reasons should be provided if the ROB is still considered low. "Unclear ROB" was assigned if relevant information was missing. Scoring was performed independently by two reviewers (JYR and LA), and disagreements in quality scores were resolved via a consensus discussion and re-examination of the full-text articles. A full description of the ROB and applicability signalling questions can be found in Supplementary Table S5.

Statistical analysis

Data extracted from included studies were subjected to qualitative assessment, where we evaluated and summarized key characteristics and findings. In addition, we employed graphical representation, such as tables and figures, to visually present the extracted data. Statistical heterogeneity among studies was evaluated using the I^2 statistic[42]. Considering model heterogeneity, a random-effects meta-analysis was performed only on the same prediction models that had undergone numerous external validations[43]. This approach addresses a significant gap overlooked in previous reviews, which often fail to account for the variability in model performance across different populations and settings. Given that models that have not been externally validated may exhibit biases due to overfitting or limited scope, focusing on models with rigorous external validation minimizes these biases and ensures that the meta-analysis results reflect models that have been tested under real-world conditions. This approach ensures that the conclusions drawn are based on well-tested models, providing a clearer and more dependable assessment of how these prediction models perform in varied sociocultural contexts for lung cancer risk prediction and screening, supporting evidence-based recommendations, and identifying areas for improvement. A random-effects meta-analysis was performed using the DerSimonian and Laird method, which involves extracting effect sizes and their corresponding variances. Each study was assigned a weight based on the inverse of its AUC variance, and weighted AUC values were calculated by multiplying each AUC by its weight before computing the sum of weighted AUC and total weights. The overall combined AUC is obtained by dividing the sum of the weighted AUCs by the sum of weights, accounting for within- and between-study variability using Tau-squared. Specifically, the weight for each study is adjusted to account for both within-study and between-study variability using this formula $w_i* = \frac{1}{v_i + T^2}$ and the random-effects weighted average effect size is calculated based on this formula $\widehat{\theta}_{RE} = \frac{\sum w_i * \widehat{\theta}_i}{\sum w_i *}$. We further adjusted for between-study heterogeneity using the Hartung-Knapp-Sidik-Jonkman (HKSJ) test[44] with the 'metafor' package in R[45]. The analysis involved fitting a random-effects meta-analysis model using the 'rma' function with the Restricted Maximum Likelihood (REML) method to estimate between-study variance. This approach allowed us to refine the confidence intervals, providing more conservative and robust estimates of the effect sizes. Detailed methodology and results of these adjustments are provided in Supplementary Table S4. We employed a mixed-effects model to assess statistically significant differences within subgroups, using AUC as the dependent variable. Results, including estimated coefficients, standard errors, and p-values, were interpreted in accordance with the chosen significance level ($p = 0.05$), with any violations of model assumptions (e.g., linearity and normality of residuals) acknowledged[46]. Publication bias was assessed using Egger's test[47] and illustrated via a funnel plot[48]. The trim-and-fill method[49] was applied to address publication bias, adjusting for funnel plot asymmetry by trimming outliers and filling in potentially missing studies, resulting in more accurate overall effect size estimates. All statistical analyses were conducted using R Version 4.2, and R packages such as meta[50], metafor[45], and dmetar[51] were used.

Results

Study selection

Figure 1 illustrates the study selection process. The preliminary literature search in two electronic databases ($n = 1{,}792$) and backward citation searching ($n = 15$) identified a total of 1,807 potentially relevant papers. Subsequently, 289 duplicate records were removed using the EndNote software, and 1,402 records were excluded after screening the titles and abstracts. The remaining full texts of 116 studies were retrieved and assessed based on the eligibility criteria. A total of 62 studies were excluded, mainly due to wrong outcome measures ($n = 12$), prediction models for lung nodules malignancy ($n = 20$), studies focused on lung cancer screening without proposing a model ($n = 3$), and providing insufficient information regarding the model's development and predictive performance ($n = 27$). A total of 54 studies[12,27,28,52–103] were included for qualitative synthesis, and 15 studies were included for quantitative meta-analysis[12,53,61,64,66,71,76–79,82,91,92,95,98,100].

Study characteristics

Table 1 presents an overview of lung cancer prediction models[10–12,61,76,78,93,100] that have undergone numerous external validations. These models are often employed in numerous studies to validate risk estimates and assess model performance across diverse cohorts and populations. Among these models, age and smoking duration are the most frequently examined risk factors. Most models are developed in cohort studies and primarily designed for ever-smokers. However, the Spitz model[93], developed using data from a case-control study, stands out for its approach of stratifying lung cancer risk estimates based on smoking status. Out of the 8 models, the $PLCO_{M2012}$model[12] which estimates the 6-year lung cancer risk for ever-smokers aged 55–74, has gained global recognition and acceptance in the medical (clinical practice) and research communities. The model incorporates various risk predictors, including age, race/ethnicity, education status, body mass index (BMI), chronic obstructive pulmonary disease (COPD), personal cancer history, family history of lung cancer, smoking status, smoking intensity (measured in pack-years), and time since smoking cessation. The $PLCO_{M2012}$ model demonstrated an AUC of 0.803 in the development dataset (PLCO control arm) and 0.797 in the validation set (PLCO intervention arm). Compared against the NLST eligibility criteria, applying the $PLCO_{M2012}$ model with a 6-year risk threshold of 1.3455% on the intervention arm significantly enhanced sensitivity (0.83 vs. 0.71, $P < 0.0001$) and showed a significantly higher positive predictive value (4.0% vs. 3.4%, $P = 0.01$), without compromising specificity (0.629 vs. 0.627, $P = 0.54$).

The characteristics of the 54 included articles can be found in Table 2 (Western countries) and Table 3 (Asian countries). Several papers described more than one prediction model validated in different cohorts while new models were often a modification of previously developed models (e.g., with the addition of one or more risk factors). Among all studies, 77.8% ($n = 42$) of the studies were conducted in Western countries[12,55–57,59–68,70–73,75–85,88–96,99–102]

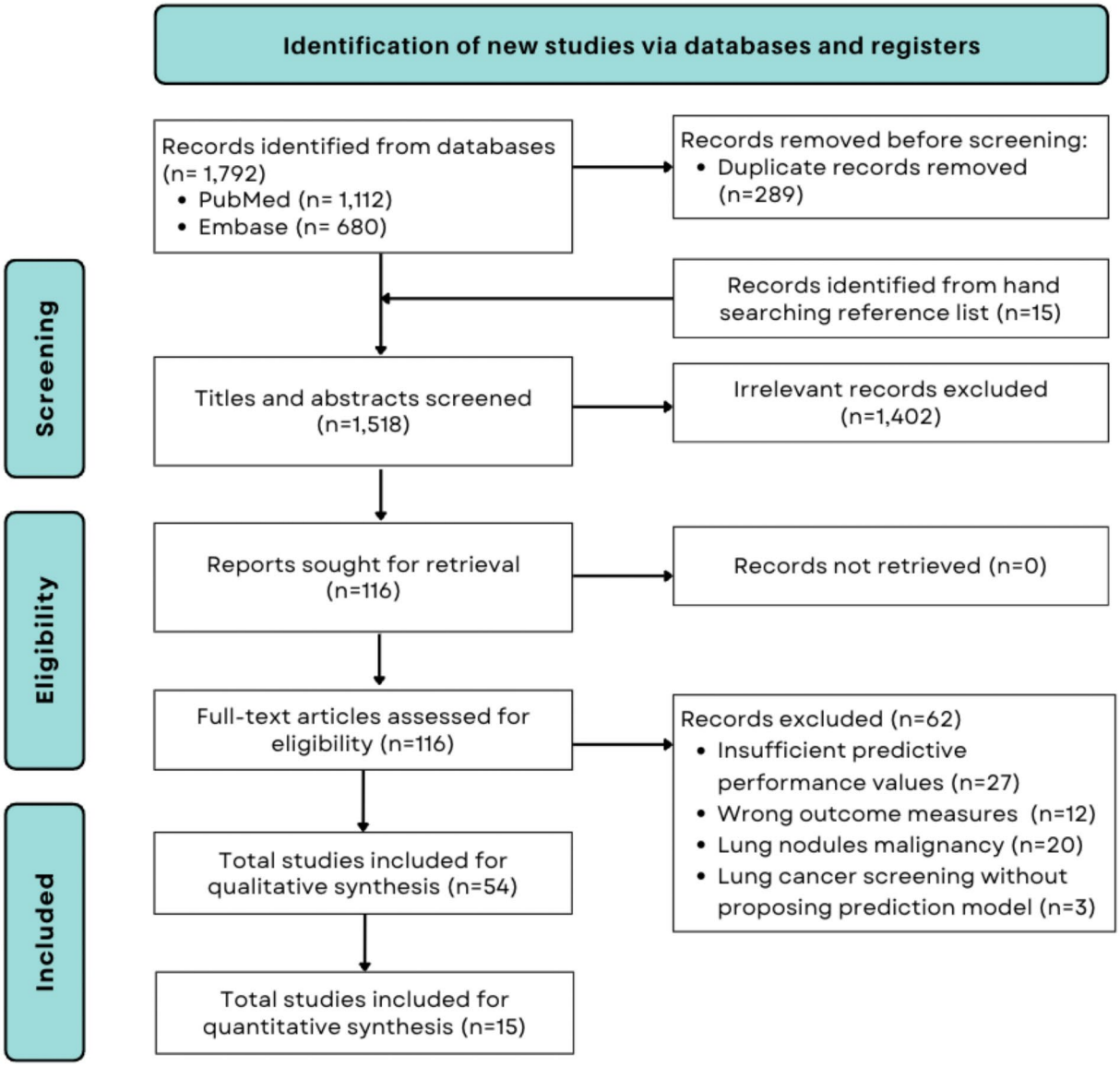

Fig. 1. PRISMA 2020 flow diagram for the selection of studies to be included in the systematic review and meta-analysis.

while 22.2% (n=12) were conducted in Asian countries[27,28,52–54,58,69,74,86,87,97,98]. The included studies were mainly from North America (44%)[12,55–57,59,61–63,65,66,70–72,76,77,79,81,83,88,92,93,95,96,101,102], Europe (26%)[59,67,73,75,78,80,82,84,89–92,94,100], and East Asia (22%)[27,28,52–54,58,69,74,86,87,97,98]. Models developed in Western countries had AUCs ranging from 0.601 to 0.920 while those developed among Asian countries had AUCs ranging from 0.639 to 0.867. Among Asian studies, 58.3% (n=7) of models were developed using case-control studies[27,52,54,69,74,87,98] and 41.7% (n=5) using cohort studies[28,53,58,86,97]. In contrast, Western countries had a higher proportion of models developed using cohort studies (64.3%, n=27) and a lower proportion of case-control studies (28.6%, n=12). Additionally, 7.1% (n=3) of the models were developed using nested-case control studies. Among Western countries, the AUCs of the models developed using cohort studies ranged from 0.668 to 0.920, case-control studies (AUCs=0.601–0.918), and nested-case control studies (AUCs=0.776–0.862). Among Asian countries, the AUCs of the models developed using cohort studies ranged from 0.680 to 0.871, while those developed using case-control studies ranged from 0.635 to 0.759.

Among included studies which reported the methods used for model internal validation, 13 studies employed cross-validation[52,54,55,68–70,79,84,89,93–95,100], six studies employed train-test spilt validation[53,62,72,80,83,98], while 19 studies employed bootstrapping methods[12,27,28,57–60,63–65,70,73,74,78,80,81,85,96,98]. In contrast to models developed in Western populations, those developed within Asian populations lack external validation. The majority of data used to develop Asian models were acquired from single-centre retrospective cohort studies, with only three studies validated using an external dataset[86,97,98]. While current existing Asian risk models have shown promising results,

Model	Author	Country or region	Risk Factors	Number of Lung Cancer Cases	Study Design	Model Type	Length risk (years)	Applicability
Bach[10]	Bach et al. (2003)	US	Age, sex, asbestos exposure, smoking intensity, smoking duration, years since smoking cessation	1,070	Cohort	Cox Proportional Hazards Regression	10	Ever-smokers (Heavy-smokers; 20 + PY)
Hoggart[100]	Hoggart et al. (2012)	UK	Age, smoking intensity, age at smoking initiation, smoking duration	1,250	Cohort	Parametric survival regression (Weibull distribution)	1	Ever-smokers (Light- and Heavy-smokers)
LCRAT[61]	Katki et al. (2016)	US	Age, sex, race, education, BMI, smoking intensity, smoking duration, years since smoking cessation, emphysema, family history of lung cancer	3,109 (PLCO) 2,047 (NLST)	Cohort	Cox Proportional Hazards Regression	1	Ever-smokers (Light- and Heavy-smokers)
LLP[11]	Cassidy et al. (2008)	UK	Smoking duration, history of pneumonia, occupational exposure to asbestos, personal history of cancer, and family history of lung cancer	579	Case-control	Multivariable Logistic Regression	5	General population
LLPi[78]	Marcus et al. (2015)	UK	Age, sex, smoking duration, COPD, Personal history of cancer, Family history of lung cancer	237	Case-control	Cox Proportional Hazards Regression	8.7	General population
Pittsburgh Predictor[76]	Wilson et al. (2015)	US	Smoking duration, smoking status, smoking intensity, age	143 (PLuSS) 1,000 (NLST LDCT) 854 (NLST CXR)	Cohort	Multivariable Logistic Regression	6	Ever-smokers (Heavy-smokers; 25 + PY)
$PLCO_{M2012}$[12]	Tammemägi et al. (2013)	Canada	Age, race, education, BMI, COPD, personal history of cancer, family history of lung cancer, smoking status, smoking intensity, smoking duration, years since smoking cessation	964	Cohort	Multivariable Logistic Regression	6	Ever-smokers (Heavy-smokers; 30 + PY)
Spitz[93]	Spitz et al. (2007)	US	**Never smokers**: Exposure to environmental tobacco smoke, family history of cancer **Former smokers**: Emphysema, dust exposure, family history of cancer, age at smoking cessation **Current smokers**: Emphysema, PY, dust exposure, asbestos exposure, family history of cancer, hay fever	1,851	Case-control	Multivariable Logistic Regression	1	General population

Table 1. Summary of risk factors included in risk prediction models that have undergone numerous validation in external cohorts. *COPD,* Chronic obstructive pulmonary disease; CXR, Chest radiography; *LCRAT,* Lung Cancer Risk Assessment Tool; LDCT, low-dose computed tomography; *LLP,* Liverpool Lung Project Cohort; *LLPi,* Liverpool Lung Project Risk Prediction Model for Lung Cancer Incidence; NLST, National Lung Screening Trial; *PLCO,* Prostate, Lung Colorectal, and Ovarian Screening Trial; PLuSS, Pittsburgh Lung Screening Study; *PY,* Pack-Years; *UK,* United Kingdom; US, United States.

external validation is crucial to ensure their generalizability across populations. Regarding study population, most Western studies were designed for ever-smokers (n = 19)[12,55,56,61,63,64,66,68,70,73,76,77,82,91,95,96,99,100,102] and the general population (n = 17)[57,59,62,71,72,75,78–80,83–85,89,90,92–94] while two studies proposed a model for both ever-smokers and the general population[67,96]. Two studies focused on individuals who underwent lung cancer screening[60,81], and two studies specifically targeted African Americans[88,101]. In contrast to Western studies, most Asian studies (n = 5)[28,54,58,69,91] were designed for the general population, but a smaller proportion of studies specifically targeted ever-smokers (n = 3)[53,87,97]. Among Asian studies, one study proposed a model each for never- and ever-smokers[98], while one study targeted only never-smoking females[27], and two studies were designed for men[52,86]. Among Asian studies targeting never-smokers, the TNSF-SQ model[27] was developed with the aim to effectively identify high-risk Asian never-smoking females (NSFs) who could benefit from LDCT screening. The model demonstrated moderate discriminative power (AUC = 0.77) and identified approximately 27% of high-risk NSFs aged 55 to 74 years, highlighting its potential for early lung cancer detection among NSFs. Another study conducted by Wang et al.[98] externally validated the China NCC-LCm2021 model for lung cancer risk prediction among never-smokers and identified justifiable risk thresholds of > 0.47% and > 0.51% to select never- and ever-smokers respectively for screening. Despite the increasing lung cancer incidence among never-smokers, only a small proportion of studies (15%)[28,52,80,83,93,94,98,100]stratified participants based on smoking status, with only two Asian studies[27,98] focused on developing a model exclusively among never-smokers, highlighting a significant gap in prediction models for this population.

Among the included studies, 29 (23 Western studies and 6 Asian studies) developed models using traditional risk factors[12,52,53,55,58,61,64–67,69,71–73,75–78,80,82,85,86,91–93,95,98,100,101] while 25 (19 Western studies and 6 Asian studies) incorporated both traditional risk factors and biomarkers such as genetic biomarkers (i.e., single nucleotide polymorphisms (SNPs)), protein biomarkers (i.e., carcinoembryonic antigen (CEA)), imaging biomarkers (i.e., CT scan) or serum-based inflammation markers (i.e., high-sensitivity C-reactive protein (hs CRP))[27,28,54,56,57,59,60,62,63,68,70,74,79,81,83,84,87–90,94,96,97,99,102]. Supplementary Figures S1A and S1B presents the

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Cronin et al. (2006)[82]	Finland	Age, sex, asbestos exposure, smoking intensity, smoking duration, years since smoking cessation	NR	NR	Cohort	Ever-smokers (Light- and Heavy Smokers)	10	333	NA	NA	**E/O ratio (95% CI)** ***ATBC***: 0.89 (0.80 – 0.99)	**C-index (95% CI)** ***ATBC***: 0.69 (0.66 – 0.72)	NR	NR
D'Amelio et al. (2010)[71]	US	Bach, Spitz, and LLP predictors	NR	NR	Case-control	General Population	5	3,197	NA	NA	NR	**AUC (95% CI)** [Bach Model] ***Harvard/MGH***: 0.66 (0.64 – 0.69) [Spitz Model] ***Harvard/MGH***: 0.69 (0.66 – 0.71) [LLP Model] ***Harvard/MGH***: 0.69 (0.67 – 0.71)	[Bach Model] 0.302 (2.5%) 0.155 (5.0%) 0.064 (7.5%) [Spitz Model] 0.266 (2.5%) 0.068 (5.0%) 0.022 (7.5%) [LLP Model] 0.667 (2.5%) 0.455 (5.0%) 0.312 (7.5%)	[Bach Model] 0.112 (2.5%) 0.024 (5.0%) 0.012 (7.5%) [Spitz Model] 0.056 (2.5%) 0.007 (5.0%)] 0.000 (7.5%) [LLP Model] 0.302 (2.5%) 0.155 (5.0%) 0.064 (7.5%)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Katki et al. (2018)[77]	US	Bach, Spitz, LLP, LLPi, $PLCO_{M2012}$, Pittsburgh Predictor, Hoggart, LCRAT model predictors	NR	NR	Population-based	Ever-smokers (Light- and Heavy Smokers)	5	11,590 (NIH-AARP) 2,320 (CPS-II)	NA	NA	**E/O ratio (95% CI)** [Bach Model] ***NIH-AARP***: 1.00 (0.98 – 1.02) ***CPS-II***: 0.97 (0.93 – 1.02) [Spitz Model] ***NIH-AARP***: 2.72 (2.55 – 2.90) ***CPS-II***: 3.69 (3.25 – 4.20) [LLP Model] ***NIH-AARP***: 1.18 (1.15 – 1.22) ***CPS-II***: 1.72 (1.62 – 1.82) [LLPi Model] ***NIH-AARP***: 1.86 (1.82 – 1.90) ***CPS-II***: 2.56 (2.45 – 2.68) [$PLCO_{M2012}$] ***NIH-AARP***: 0.92 (0.90 – 0.95) ***CPS-II***: 0.99 (0.94 – 1.04) [Pittsburgh Predictor] ***NIH-AARP***: 1.01 (0.98 – 1.03) ***CPS-II***: 0.83 (0.78 – 0.87) [Hoggart Model] ***NIH-AARP***: 2.85 (2.67 – 3.04) ***CPS-II***: 2.27 (1.99 – 2.58) [LCRAT] ***NIH-AARP***: 0.97 (0.94 – 0.99) ***CPS-II***: 0.96 (0.91 – 1.02)	**AUC (95% CI)** [Bach Model] ***NIH-AARP***: 0.755 (0.752 – 0.759) ***CPS-II***: 0.750 (0.740 – 0.760) [Spitz Model] ***NIH-AARP***: 0.707 (0.700 – 0.714) ***CPS-II***: 0.624 (0.588 – 0.661) [LLP Model] ***NIH-AARP***: 0.726(0.722 – 0.731) ***CPS-II***: 0.726(0.711 – 0.740) [LLPi Model] ***NIH-AARP***: 0.714 (0.711 – 0.717) ***CPS-II***: 0.721 (0.711 – 0.732) [$PLCO_{M2012}$] ***NIH-AARP***: 0.769 (0.766 – 0.772) ***CPS-II***: 0.754 (0.741 – 0.767) [Pittsburgh Predictor] ***NIH-AARP***: 0.752 (0.748 – 0.757)] ***CPS-II***: 0.747 (0.731 – 0.762) [Hoggart Model] ***NIH-AARP***: 0.714 (0.699 – 0.728) ***CPS-II***: 0.751 (0.722 – 0.781) [LCRAT] ***NIH-AARP***: 0.771 (0.768 – 0.775) ***CPS-II***: 0.770 (0.757 – 0.784)	[Bach Model] 0.68 (2.0%) 0.60 (2.0%) [$PLCO_{M2012}$] 0.59 (2.0%)] 0.58 (2.0%) [LCRAT] 0.62 (2.0%) 0.77 (2.0%)	[Bach Model] 0.71 (2.0%) 0.76 (2.0%) [$PLCO_{M2012}$] 0.78 (2.0%)] 0.79 (2.0%) [LCRAT] 0.77 (2.0%) 0.59 (2.0%)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Li et al. (2015)[91]	Germany	Bach, Spitz, LLP, $PLCO_{M2012}$ predictors	NR	NR	Population-based	Ever-smokers (Heavy-Smokers; 20+pack years)	5	92	NA	NA	**E/O ratio (95% CI)** [Bach Model] ***EPIC-Germany***: 0.88 (0.72–1.08) [Spitz Model] ***EPIC-Germany***: 3.75 (3.06–4.60) [LLP Model] ***EPIC-Germany***: 1.12 (0.92–1.37) [$PLCO_{M2012}$] ***EPIC-Germany***: 1.03 (0.87–1.23)	**C-index (95% CI)** [Bach Model] ***EPIC-Germany***: 0.81 (0.76–0.86) [Spitz Model] ***EPIC-Germany***: 0.78 (0.73–0.83) [LLP Model] ***EPIC-Germany***: 0.79 (0.73–0.83) [$PLCO_{M2012}$] ***EPIC-Germany***: 0.81 (0.76–0.86)	[Bach Model] 0.620 (0.76%) 0.620 (0.90%) 0.554 (1.18%) 0.478 (1.55%) [Spitz Model] 0.533 (3.51%) 0.489 (3.74%) 0.446 (4.18%) 0.370 (4.72%) [LLP Model] 0.587 (0.85%) 0.554 (0.94%) 0.500 (1.17%) 0.435 (1.53%) [$PLCO_{M2012}$] 0.663 (1.01%) 0.641 (1.13%) 0.598 (1.73%) 0.500 (2.11%)	[Bach Model] 0.837 (0.76%) 0.860 (0.90%)] 0.897 (1.18%) 0.931 (1.55%)] [Spitz Model] 0.837 (3.51%) 0.860 (3.74%) 0.897 (4.18%) 0.931 (4.72%) [LLP Model] 0.837 (0.85%) 0.860 (0.94%) 0.897 (1.17%) 0.931 (1.53%) [$PLCO_{M2012}$] 0.838 (1.01%) 0.861 (1.13%) 0.897 (1.73%) 0.932 (2.11%)
Spitz et al. (2007)[93]	US	**Never smokers**: exposure to environmental tobacco smoke, family history of cancer **Former smokers**: emphysema, dust exposure, family history of cancer, age at smoking cessation **Current smokers**: emphysema, pack-years, dust exposure, asbestos exposure, family history of cancer, hay fever	Multivariable Logistic Regression (backward selection)	Three-fold cross-validation	Case-control	General population	1	1,851	**H-L goodness-of-fit** ***MDACC***: Never-smokers: $P = 0.777$ Former smokers: $P = 0.712$ Current smokers: $P = 0.688$	**AUC (95% CI)** ***MDACC***: Never smokers: 0.57 (0.47–0.66) Former smokers: 0.63 (0.58–0.69) Current smokers: 0.58 (0.52–0.64)	NR	NR	NR	NR
Raji et al. (2012)[92]	Europe & North America	Smoking duration, history of pneumonia, occupational exposure to asbestos, personal history of cancer, and family history of lung cancer	Multivariable Logistic Regression (backward stepwise selection)	10-fold cross-validation	Case-control	Individuals aged 40–80 years old	5	585 (EUELC) 1,738 (Harvard/MGH) 420 (LLP)	NA	NA	NR	**AUC (95% CI)** **EUELC**: 0.67 (0.64–0.69) **Harvard/MGH**: 0.76 (0.75–0.78) **LLP**: 0.82 (0.80–0.85)	***EUELC***: 0.552 (2.5%) ***Harvard***: 0.647 (2.5%) ***LLP***: 0.743 (2.5%) ***EUELC***: 0.357 (5.0%) ***Harvard***: 0.494 (5.0%) ***LLP***: 0.574 (5.0%) ***EUELC***: 0.142 (10.0%) ***Harvard***: 0.253 (10.0%) ***LLP***: 0.279 (10.0%)	***EUELC***: 0.698 (2.5%) ***Harvard***: 0.731 (2.5%) ***LLP***: 0.674 (2.5%) ***EUELC***: 0.843 (5.0%) ***Harvard***: 0.871 (5.0%) ***LLP***: 0.811 (5.0%) ***EUELC***: 0.943 (10.0%) ***Harvard***: 0.957 (10.0%) ***LLP***: 0.925 (10.0%)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Marcus et al. (2015)[78]	UK	Age, sex, smoking duration, COPD, Personal history of cancer, Family history of lung cancer	Cox Proportional Hazards Regression (backward selection)	Bootstrap (200 times with repalcement)	Case-control	Individuals aged 40–80 years old	8.7	237	**H-L goodness-of-fit** $P = 0.371$	**C-index (95% CI)** ***LLP***: 0.852 (0.831–0.873)	NR	NR	NR	NR
Etzel et al. (2008)[101]	US	Smoking status, pack- years, age at smoking cessation, years since smoking cessation, self-reported history of COPD or hay fever, exposure to asbestos or wood dusts	Multivariable Logistic Regression (stepwise selection)	NR	Case-control	African Americans	5	491	NR	**AUC (95% CI)** ***MDACC***: 0.75 (0.67–0.82)	NR	**AUC (95% CI)** ***metropolitan Detroit***: 0.63 (0.57–0.69)	NR	NR
Tammemägi et al. (2011)[65]	Canada	**Model 1**: Age, education level, BMI, family history of lung cancer, COPD, recent chest x-ray, smoking status, smoking intensity, smoking duration. **Model 2**: Model 1 + years since smoking cessation	Cox Regression Predictive Model (backward selection)	Bootstrap (200 resampling)	Cohort	**Model 1**: General Population **Model 2**: Ever-smokers (Light- and Heavy Smokers)	9	1,040	**Mean (90th percentile) Absolute error** ***PLCO CTRL***: Model 1: 0.0009 (0.0025) Model 2: 0.0014 (0.0029)	**AUC (95% CI)** ***PLCO CTRL***: Model 1: 0.859 (0.848–0.871) Model 2: 0.809 (0.796–0.822)	**Brier score (95% CI)** Model 1: 0.061 (0.052–0.069)[6] Model 2: 0.061 (0.052–0.069)[6]	**AUC (95% CI)** ***PLCO CTRL***: Model 1: 0.841 (0.813–0.870) Model 2: 0.784 (0.745–0.824)	NR	NR
Tammemägi et al. (2013)[12]	Canada	Age, race, education, BMI, COPD, personal history of cancer, family history of lung cancer, smoking status, smoking intensity, smoking duration, years since smoking cessation	Multivariable Logistic Regression (expert-driven selection)	Bootstrap (200 resampling)	Cohort	Ever-smokers (Heavy-Smokers; 30-pack years)	6	709	**Mean (90th percentile) Absolute error** ***PLCO CTRL***: 0.009 (0.042)	**AUC (95% CI)** ***PLCO CTRL***: 0.803 (0.782–0.813)	NR	**AUC (95% CI)** ***PLCO CXR***: 0.797 (0.782–0.813) ***NLST***: 0.701 (0.689–0.712)	0.83 (1.345%)	0.629 (1.345%)
Kats et al. (2021)[95]	US (Ohio)	$PLCO_{M2012}$ and LCRAT predictors	NR	NR	Cohort	Ever-smokers (Heavy-Smokers; 30-pack years)	6	1,093	NA	NA	**E/O ratio (95% CI)** [$PLCO_{M2012}$] ***I-Ohio***: Assessed graphically [LCRAT] ***I-Ohio***: Assessed graphically	**AUC (95% CI)** [$PLCO_{M2012}$] **I-Ohio**: 0.71 (0.69–0.73) [LCRAT] **I-Ohio**: 0.72 (0.70–0.74)	0.699 (1.75%)	0.583 (1.75%)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Weber et al. (2017)[64]	Australia	Age, ethnicity, education, BMI, personal history of cancer, family history of lung cancer, smoking status, smoking intensity, smoking duration, years since smoking cessation	Cox Proportional Hazards Regression (NR)	Bootstrap (200 resampling)	Cohort	Ever-smokers (Light- and Heavy Smokers)	6	1,035	NA	NA	**Mean (90th percentile) Absolute Risk Difference** ***NSW-Australia***: 0.006 (0.016)	**AUC (95% CI)** ***NSW-Australia***: 0.80 (0.78–0.81)	0.694 (1.51%)	0.720 (1.51%)
Tammemägi et al. (2014)[85]	Canada	Age, race, education, BMI, COPD, personal history of cancer, family history of lung cancer, smoking status, smoking intensity, smoking duration, years since smoking cessation	Multivariable Logistic Regression (expert-driven selection)	Bootstrap (1,000 resampling with replacement)	Cohort	General Population	6	1,040	**Median (90th percentile) Absolute error** ***PLCO CTRL***: 0.0002 (0.0005)	**AUC (95% CI)** ***PLCO CTRL***: 0.859 (0.845 – 0.872)	**Median (90th percentile) Absolute error** ***PLCO CXR***: 0.0003 (0.0014)[8] ***PLCO never-smokers***: 0.0002 (0.0003)[8]	**AUC (95% CI)** ***PLCO CXR***: 0.848 (0.833 – 0.861) ***PLCO never-smokers***: 0.662 (0.607 – 0.709)	NR	NR
Wilson et al. (2015)[76]	US	Smoking duration, smoking status, smoking intensity, age	Multivariable Logistic Regression (NR)	NR	Cohort	Ever-smokers (Heavy-smokers; 25 + pack-years)	6	143 (PLuSS) 1,000 (NLST LDCT) 854 (NLST CXR)	**H-L goodness-of-fit** ***NLST CXR***: $P = 0.19$ ***NLST LDCT***: $P = 0.08$	**AUC (95% CI) [Pittsburgh Predictor]** ***NLST CXR***: 0.688 (0.670–0.705) ***NLST LDCT***: 0.678 (0.662–0.694) **[Bach Model]** ***NLST CXR***: 0.695 (0.678–0.712) ***NLST LDCT***: 0.687 (0.670–0.703) [$PLCO_{M2012}$] ***NLST CXR***: 0.702 (0.685–0.719) ***NLST LDCT***: 0.690 (0.674–0.706)	NR	**AUC (95% CI) [Pittsburgh Predictor]** ***PluSS***: 0.701 **[Bach Model]** ***PluSS***: 0.710 **[$PLCO_{M2012}$]** ***PluSS***: 0.721	NR	NR
Hoggart et al. (2012)[100]	UK	Age, smoking intensity, age at smoking initiation, smoking duration	Parametric survival regression with Weibull distribution (expert-driven selection)	10-fold cross validation	Cohort	Ever-smokers (Light- and Heavy Smokers)	1	820/304 (training set: current-/former-smokers) 92/34 (validation set: current-/former-smokers)	Assessed Graphically	**AUC (95% CI)** [Hoggart Model] ***EPIC***: ***Ever Smokers***: 0.843 (0.810 – 0.875) ***Former Smokers***: 0.830 (0.762 – 0.899) ***Current Smokers***: 0.824 (0.783 – 0.865) [Bach Model] ***EPIC***: 0.775 (0.737–0.813)	NR	NR	NR	NR

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Markaki et al. (2018)[73]	Norway	Sex, age, pack-years, smoking intensity, years since smoking cessation, BMI, daily cough, smoke exposure	Cox Proportional Hazards Regression (backward selection)	Bootstrap (200 resampling)	Population-based	Ever-smokers (Light- and Heavy Smokers)	6	583	Assessed Graphically	**C-index (95% CI)** **HUNT2**: Overall: 0.903 (0.893 – 0.913) Ever-smokers: 0.869 (0.858 – 0.880)	NR	**C-index (95% CI)** ***CONOR***: Overall: 0.892 (0.903–0.979) Ever-smokers: 0·879 (0.866 – 0.891)	0.819 (0.64%)	0.783 (0.64%)
Katki et al. (2016)[61]	US	Age, sex, race, education, BMI, smoking intensity, smoking duration, years since smoking cessation, emphysema, family history of lung cancer	Cox Proportional Hazards Regression (expert-driven selection)	NR	Cohort	Ever-smokers (Heavy-Smokers; 30-pack years)	5	1,505/1,604 (PLCO control/ radiography) 1,083/964 (NLST CT/ radiography)	**E/O ratio (95% CI)** ***PLCO CXR***: 0.94 (0.84 – 1.02)	**AUC (95% CI)** ***PLCO CXR***: 0.80 (0.78–0.81)	**E/O ratio (95% CI)** ***NLST CXR***: 1.06 (0.98 – 1.13)	**AUC (95% CI)** ***NLST CXR***: 0.70 (0.69 – 0.72)	0.987 (1.00%) 0.889 (2.00%)	0.073 (1.00%) 0.332 (2.00%)
Gu et al. (2017)[66]	US	Age, sex, race, education, BMI, smoking intensity, smoking duration, years since smoking cessation, emphysema, family history of lung cancer, time to first cigarette (TTFC)	Cox Proportional Hazards Regression (NR)	NR	Cohort	Ever-smokers (Heavy-Smokers; 30-pack years)	5	773	NR	**AUC (95% CI)** ***NLST-ACRIN***: 0.694 (0.677 – 0.712)	NR	**AUC (95% CI)** ***NLST-ACRIN***: 0.686 (0.668 – 0.704)	0.987 (1.00%) 0.889 (2.00%)	0.093 (1.00%) 0.349 (2.00%)
Hart et al. (2018)[72]	US	Sex, age, BMI, diabetes, smoking status, emphysema, asthma, race, Hispanic ethnicity, hypertension, heart diseases, vigorous exercise habits, history of stroke	Artificial neural network (ANN) (expert-driven selection)	Train-test spilt validation (70/30)	Cohort	General Population	3	454 (training set) 195 (validation set)	NR	**AUC (95% CI)** ***NHIS***: 0.86 (0.84 – 0.89)	NR	NR	Training: 0.798 Validation: 0.753	Training: 0.799 Validation: 0.806
Muller et al. (2017)[80]	UK	**Never-Smokers**: sex, personal history of cancer, family history of lung cancer, emphysema/ chronic bronchitis, hay fever/ allergic rhinitis/ eczema, maximum FEV1% **Former-Smokers**: (never smokers) + age at smoking cessation, smoking intensity **Current-Smokers**: (never smokers) + smoking intensity, difficulty of not smoking for one day	Flexible Parametric Survival Analysis (expert-driven selection)	Bootstrap (1,000 resampling) and Train-test spilt validation (80/20)	Cohort	General UK individuals aged 37 to 73 years old	2	738	Assessed Graphically	**C-index (95% CI)** ***UK Biobank***: Overall: 0.85 (0.82 – 0.87) Never Smokers: 0.74 (0.69 – 0.78) Former Smokers: 0.82 (0.79 – 0.84) Current Smokers: 0.82 (0.79 – 0.84)	NR	NR	0.95 (0.354%)	0.47 (0.354%)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Ward et al. (2019)[55]	US	**Model 1**: $PLCO_{M2012}$ predictors + nicotine dependency (TTFC) **Model 2**: LCRAT predictors + nicotine dependency (TTFC)	**Model 1**: Multivariable Logistic Regression (NR) **Model 2**: Cox Proportional Hazards Regression (NR)	Leave-one-out cross-validation	Cohort	Ever smokers (Heavy-Smokers; 30-pack years)	**Model 1**: 6 **Model 2**: 5	587	**Brier scores** **NLST-ACRIN**: $PLCO_{M2012}$+ TTFC 0.039 LCRAT + TTFC: 0.039	**AUC (SE)** ***NLST-ACRIN***: $PLCO_{M2012}$+ TTFC: 0.70 (0.013) LCRAT + TTFC: 0.69 – 0.72 (0.013 – 0.017)	NR	NR	LCRAT + TTFC: 0.985 (1.0%) 0.876 (2.0%)	LCRAT + TTFC: 0.092 (1.0%) 0.351 (2.0%)
Pan et al. (2023)[67]	UK	Age, gender, BMI, education level, smoking status, smoking initiation age, average number of cigarettes per day, smoking duration, duration since quit smoke, smoking pack-year, history of diabetes, COPD, emphysema, chronic bronchitis, lung cancer family history	Extreme Gradient Boosting (XGBoost algorithm)	5-fold cross-validation	Cohort	General population & Ever-smokers (Light- and Heavy Smokers)	6	1,655 (training set) 564 (validation set 1: UKB Scotland and Wales area) 3,042 (validation set 2: PLCO) 1,132 (validation set 3: NLST)	**E/O ratio (95% CI)** General population: ***UK Biobank***: 1.014 (0.958–1.077) Ever smokers: ***UK Biobank***: 1.045 (0.983–1.117)	**C-index (95% CI)** General population: ***UK Biobank (Training)***: 0.859 (0.846–0.873) Ever-smokers: ***UK Biobank (Training)***: 0.855 (0.839–0.870)	**E/O ratio (95% CI)** General population: **UKB-SW**: 1.037 (0.939–1.159) ***PLCO CTRL***: 0.832 (0.775–0.897) ***PLCO (intervention)***: 0.948 (0.876–1.033) Ever-smokers: ***UKB-SW***: 1.081 (0.969–1.221) ***PLCO CTRL***: 0.813 (0.756–0.879) ***PLCO (intervention)***: 0.914 (0.843–0.997) ***NLST CTRL***: 0.989 (0.920–1.070) ***NLST (intervention)***: 1.255 (1.145–1.389)	**C-index (95% CI)** General population: ***UKB-SW***: 0.855 (0.829–0.880) ***PLCO CTRL***: 0.855 (0.837–0.873) ***PLCO (intervention)***: 0.861 (0.841–0.881) Ever-smokers: ***UKB-SW***: 0.842 (0.814–0.871) ***PLCO CTRL***: 0.791 (0.770–0.812) ***PLCO (intervention)***: 0.792 (0.768–0.815) ***NLST CTRL***: 0.711 (0.688–0.734) ***NLST (intervention)***: 0.718 (0.689–0.747)	**UKB-SW**: 0.71 (0.024%) 0.74 (0.056%) 0.75 (0.095%) 0.78 (0.129%) 0.80 (0.170%) 0.81 (0.213%) 0.80 (0.269%) 0.81 (0.303%)	**UKB-SW** 0.73 (0.024%) 0.73 (0.056%) 0.74 (0.095%) 0.73 (0.129%) 0.73 (0.170%) 0.73 (0.213%) 0.73 (0.269%) 0.73 (0.303%)
Hippisley-Cox et al. (2015)[75]	UK	Age, sex, ethnicity, Townsend deprivation score, BMI, smoking status, alcohol, previous cancer diagnoses, family history of lung cancer, asthma, COPD, personal history of cancer	Cox Proportional Hazards Regression (expert-driven selection)	NR	Cohort	General population	10	32,187	Assessed Graphically	**AUC (95% CI)** ***Qresearch***: Women: 0.905 (0.901 – 0.910) Men: 0.911 (0.908 – 0.914)	NR	NR	Women: 0.673 (1.43%) Men: 0.666 (2.73%)	Women: 0.902 (1.43%) Men: 0.903 (2.73%)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Spitz et al. (2008)[68]	US	**Former Smokers**: emphysema, dust exposure, family history of cancer, age at smoking cessation, hay fever, DNA repair capacity (DRC), bleomycin sensitivity **Current Smokers**: emphysema, pack-years, family history of cancer, hay fever, dust exposure, asbestos exposure, DNA repair capacity, bleomycin sensitivity	Multivariate Logistic Regression (backward selection)	3-fold cross-validation	Case-control	Ever-smokers (Light- and heavy-smokers)	1	725	**H-L goodness of fit** Former Smokers: $P = 0.610$ Current Smokers: $P = 0.433$	AUC (95% CI) ***MDACC (baseline model + DRC + bleomycin)***: Former Smokers: 0.70 (0.66 – 0.74) Current Smokers: 0.73 (0.69 – 0.77)	NR	NR	NR	NR
Yee et al. (2009)[102]	Canada	Age, smoking, FEV1%, CTAP III/NAP-2	Multivariate Logistic Regression (stepwise selection)	NR	Nested case-control	Ever-smokers (Heavy-Smokers; 30-pack years)	Preclinical lung cancer	49	NR	**AUC (95% CI)** Age + smoking status + FEV1%: 0.80 (0.72 – 0.88) Age + smoking status + FEV1% + CTAP III/NAP-2: 0.81 (0.73 – 0.89) Age + smoking status + FEV1% + Haptoglobin: 0.82 (0.74 – 0.90) Age + smoking status + FEV1% + CTAP III/NAP-2 + Haptoglobin + CTAP III/NAP-2*FEV1%: 0.839 (0.765 – 0.913)	NR	NR	NR	NR
El-Zein et al. (2014)[83]	Canada	**Never-Smokers**: CBMN, second-hand smoking, family history of cancer **Former-Smokers**: CBMN, emphysema, dust exposure, family history of cancer, age at smoking cessation **Current-Smokers**: CBMN, emphysema, pack-years, family history of cancer, dust exposure, asbestos exposure	Multivariate Logistic Regression (expert-driven selection)	Train-test spilt validation	Case-control	General Population	1	527 (training set) 239 (validation set)	NA	NA	NR	**AUC (95% CI)** ***MGH***: Overall: 0.918 (0.894 – 0.942) Never-Smokers: 0.918 (0.863 – 0.973) Former-Smokers: 0.910 (0.873 – 0.948) Current-Smokers: 0.925 (0.885 – 0.964)	NR	NR

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Spitz et al. (2013)[88]	US	Age, sex, pack-years, asbestos, emphysema, hay fever, family history of cancer, 6 SNPs (rs950286, rs7124327, rs2736100, rs16969968, rs3087386, rs10519203)	Multivariate Logistic Regression (expert-driven selection)	NR	Case-control	African Americans	5	477 (MDACC) 330 (MDACC-EXHALE)	**H-L goodness of fit** ***MDACC-EXHALE***: $P = 0.576$[38]	**AUC (95% CI)** ***MDACC***: 0.80 (0.77 – 0.84) ***MDACC-EXHALE***: 0.76 (0.73 – 0.79)	NR	**NR**	NR	NR
Marcus et al. (2016)[84]	UK	Age, sex, smoking duration, pneumonia, asbestos exposure, family history of cancer, 3 SNPs (rs1799732 (DRD2), rs5744256 (IL-18) and rs2306022 (ITGA11))	Multivariate Logistic Regression (expert-driven selection)	10-fold cross-validation	Case-control	General Population	5	718	Assessed using AIC and BIC	***AUC (95% CI)*** ***LLP***: 0.81 (0.79 – 0.83)	NR	NR	NR	NR
Sin et al. (2013)[70]	Canada	Age, sex, BMI, personal history of cancer, pneumonia, family history of cancer, smoking duration, cigarettes smoked per day, FEV1%, pro-SFTPB	Multivariate Logistic Regression (expert-driven selection)	Bootstrap (1,000 resampling)	Case-control	Ever-smokers	3	113	**Mean (90th percentile) Absolute error** ***PanCan***: 0.004 (0.010) **Brier Score** ***PanCan***: 0.0438	**AUC (95% CI)** ***PanCan***: 0.741 (0.696 – 0.783)	NR	**AUC (95% CI)** ***CARET***: 0.683 (0.604 – 0.761)	0.804 (3.2%)	0.401 (3.2%)
Guida et al. (2018)[99]	US	Age, smoking history, protein biomarkers (Pro-SFTPB, CA125, CEA, HE4, and CYFRA21 – 1)	Logistic Regression (NR)	NR	Case-control	Ever-smokers (Heavy-Smokers; 30 pack-years)	1	63	NR	**AUC (95% CI)** ***CARET***: 0.80 (0.72 – 0.87)	NR	**AUC (95% CI)** ***EPIC-NSHDS***: 0.83 (0.76 – 0.90)	0.63 (specificity of 0.83: USPSTF)	0.95 (sensitivity of 0.42: USPSTF)
Qian et al. (2016)[89]	Italy	Age, sex, pack-years, SNPs, principal components of subtype-specific pathways (germline) and pathway-smoking interactions	Multinomial Logistic Regression (expert-driven selection)	5-fold cross-validation	Case-control	General Population	5	1,815 (training set) 556 (validation set)	NA	NA	NR	***AUC (95% CI)*** **PLCO**: 0.656 (0.626 – 0.685)	0.66 (NR)	0.63 (NR)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Gould et al. (2021)[79]	US	Smoking duration, age, pack- years, BMI, WBC count, history of COPD, time since quitting smoking, Hispanic ethnicity, HDL-trend and fluctuations, RDW value, ALT value, INR-time, platelets trend, fluctuations and value, cancer history, K + trend and LDL fluctuations	Extreme Gradient Boosting (XGBoost algorithm)	Cross-validation	Nested case-control	General population aged 45–90 years	1	6,505	**Spiegelhalter z test** **KPSC**: $P = 0.075$	**AUC (95% CI)** **[MES model]** ***KPSC***: Ever-smokers: 0.862 (0.845–0.878) NSCLC (0-II): 0.880 (0.858–0.901) NSCLC (III–IV): 0.846 (0.828–0.863) Adenocarcinoma (0–IV): 0.835 (0.814–0.858) Squamous cell carcinoma (0–IV): 0.888 (0.865–0.914)	NR	NR	0.401 (95% specificity)	NR
Hippisley-Cox et al. (2011)[90]	UK	BMI, Townsend score, smoking status, personal history of cancer (female only), COPD, hemoglobin, hemoptysis, appetite loss, weight loss, onset of cough	Cox proportional hazards Regression (expert-driven selection)	NR	Cohort	General Population aged 30 to 84 years	2	3,785 (training set) 2,196 (validation set)	Assessed Graphically	**AUC (95% CI)** ***Qresearch***: Males: 0.92 (0.91 – 0.93) Females: 0.92 (0.91 – 0.93)	NR	NR	0.773 (0.37%) 0.627 (0.68%) 0.362 (2.21%) 0.274 (4.47%)	NR
Hung et al. (2021)[94]	UK	Sex, race, education, BMI, tobacco smoking, COPD history and family history of cancer (stratified by PRS group)	Cox proportional hazards (NR)	10-fold cross-validation	Cohort	General UK individuals aged 40 to 69 years old	5	13,119	**Spiegel halter's z statistic** $P = 0.135$	**AUC (95% CI)** ***UK biobank***: overall (with PRS): 0.832 (0.811–0.853) overall (without PRS): 0.828 (0.807–0.850) Ever-smokers (with PRS): 0.786 (0.762–0.809) Ever-smokers (without PRS): 0.785 (0.762–0.809) Never-smokers (with PRS): 0.687 (0.628–0.746) Never-smokers (without PRS): 0.670 (0.611–0.729)	NR	NR	NR	NR

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Pal Choudhury et al. (2020)[63]	Canada	**Baseline model**: Gender, race, education, BMI, smoking pack-year categories, number of years since stopped smoking cigarettes, number of years smoked, indicator of > 1 pack/ day, presence/absence of emphysema, and lung cancer family history. **Model 1**: Baseline model + CRP **Model 2**: Baseline model + four inflammation markers (CRP, SAA, sTNFRII, CXCL9/ MIG)	Parametric survival regression with Weibull distribution (expert-driven selection)	Bootstrap (1,000 resampling)	Two phase (nested case-control)	Ever-smokers (Light- and Heavy-Smokers)	10	1,253	NR	**AUC (SE)** ***PLCO***: **Model 1**: (≤ 59 years): 0.81 (0.017) (60–64 years): 0.76 (0.016) (65–69 years): 0.808 (0.014) (≥ 70 years): 0.725 (0.019) **Model 2**: (≤ 59 years old): 0.791 (0.02) (60–64 years): 0.759 (0.017) (65–69 years): 0.811 (0.017) (≥ 70 years): 0.738 (0.021)	NR	NR	NR	NR
Tammemägi et al. (2011)[96]	Canada	**Base Model**: Age, sex, education, BMI, family history of lung cancer, smoking pack-years, smoking status (current vs. former), smoking quit-time **Model 2**: Base Model + FEV1% **Model 3**: Base model + FEV1% + sputum DNA image cytometry (SDIC)	Multivariable Logistic Regression (expert-driven selection)	Bootstrap (1,000 resampling)	Cohort	Ever smokers over 40 years old (heavy-smokers; 20 pack-years smoking history)	8	139	**Mean (90th percentile) Absolute error** ***NCI-BCCA***: Base Model: 0.0047 (0.0110) Model 2: 0.0055 (0.0099) Model 3: 0.0058 (0.0089)	**AUC (95% CI)** **NCI-BCCA**: Base Model: 0.718 (0.671 – 0.765) Model 2: 0.767 (0.725 – 0.809) Model 3: 0.773 (0.732 – 0.815)	NR	NR	NR	NR
Wang et al. (2019)[62]	US	Age, gender, smoking, pulmonary diseases, symptoms (i.e., hemoptysis, cough, and chest pain), and abnormal laboratory test results (i.e., C-reactive protein and fibrinogen)	Extreme Gradient Boosting (XGBoost Algorithm)	Train-test spilt validation	Cohort	General Population	1	1,091 (training set) 1,167 (validation set)	Assessed graphically	**AUC (95% CI)** ***Maine***: Overall: 0.881 (0.873 – 0.889) Smoking: 0.865 (0.823–0.907) Age ≥ 65 years: 0.755 (0.738 – 0.772) Age < 45 years: 0.880 (0.776 – 0.984)	NR	NR	NR	NR

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Schreuder et al. (2021)[59]	US & Italy	**Model 1 (Survey)**: Age, gender, race, educational level, BMI, smoking status, intensity, duration, quit time, family history, Work asbestos, COPD, Asthma, Diabetes, heart disease, hypertension and stroke diagnosis **Model 2 (CT)**: Nodule attenuation, longest diameter, perpendicular diameter in upper lobe, spiculation, count **Model 3 (Final Model)**: All variables + Age, mean lung density, emphysema score, bronchial wall thickness and aorta calcium volume	Cox proportional hazards regression (expert-driven selection)	Bootstrap (500 resampling)	Cohort	General Population	5	923 (training set) 108 (validation set)	Assessed Graphically	**AUC (95% CI)** ***NLST***: Model 1 (Survey): 0.706 (0.688–0.724) Model 2 (CT): 0.825 (0.809–0.840) Model 3 (final): 0.840 (0.826–0.855)	Assessed Graphically[64]	**AUC (95% CI)** ***MILD***: Model 1 (Survey): 0.747 (0.680–0.814) Model 2 (CT): 0.801 (0.742–0.860) Model 3 (final): 0.799 (0.739–0.858)	Final Model (5-year): 0.966 (0.1%) 0.966 (0.25%) 0.847 (0.50%) 0.712(0.75%) 0.491(0.90%)	Final model (5-year): 0.103 (0.1%) 0.258 (0.25%) 0.512 (0.50%) 0.765 (0.75%) 0.911(0.90%)
Lu et al. (2020)[57]	US	Age, sex, smoking status (former/current) and chest radiograph images	Convolutional neural network (CNN) (expert-driven selection)	Bootstrap (1,000 resampling)	Cohort	General population	12	962	**E/O ratio (95% CI)** **PLCO**: 1.01 (0.87–1.14)	**AUC (95% CI)** ***PLCO***: 0.755 (0.723–0.786)	NR	**AUC (95% CI)** ***NLST***: 0.659 (0.622–0.695)	0.714 (3.297%)	0.627 (3.297%)
Raghu et al. (2022)[56]	US	Age, sex, smoking status (former/current) and chest radiograph images	Convolutional neural network (CNN) (NR)	NR	Cohort	Ever-smokers (Light- and Heavy-Smokers)	6	361	NA	NA	Assessed Graphically	**AUC (95% CI)** ***MGBH***: Overall: 0.704 (0.67–0.74) Female: 0.704 (0.66–0.75) Male: 0.679 (0.63–0.73) Ethnicity (Black): 0.733 (0.51–0.79)	0.403 (3.297%)	0.825 (3.297%)

Continued

Author	Country or region	Risk Factors	Model Type (Variable Selection Approach)	Internal Validation method	Study Design	Study Population	Length Risk (years)	Number of Lung Cancer Cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Maldonado et al. (2021)[60]	US	LCRAT predictors + screening results	Cox Proportional Hazards Regression (NR)	Bootstrap (10,000 resampling)	Cohort	Individuals with a negative LDCT screen	1	24	NA	NA	**Brier Score (95% CI)** ***LUSI***: 0.004	**AUC (95% CI)** ***LUSI***: 0.73 (0.63–0.82)	0.95 (0.03–0.1%) 0.90 (0.13%) 0.75 (0.17%) 0.65 (0.23%) 0.40 (0.32%) 0.20 (0.48%)	0.10 (0.03%) 0.20 (0.05%) 0.30 (0.07%) 0.40 (0.1%) 0.50 (0.13%) 0.60 (0.17%) 0.70 (0.23%) 0.80 (0.32%) 0.90 (0.48%)
Tammemägi et al. (2019)[81]	Canada	PLCO predictors + screening results	Logistic Regression (expert-driven selection)	Bootstrap (1,000 resampling)	Cohort	Individuals who have undergone lung cancer screening	3	298	**Brier score (95% CI)** ***NLST-LSS***: 0.013 (0.012 – 0.015) **Mean (90th percentile) Absolute error** ***NLST-LSS***: 0.0009 (0.0012)	**AUC (95% CI)** ***NLST-LSS***: 0.769 (0.741 – 0.797)	**Brier score (95% CI)** ***NLST-ACRIN***: 0.012 (0.010 – 0.014) **Mean (90th percentile) Absolute error** ***NLST-ACRIN***: 0.0018 (0.0030)	**AUC (95% CI)** **NLST-ACRIN**: 0.761 (0.716 – 0.799)	0.931 (0.75%) 0.937 (0.75%) 0.856 (1.00%) 0.854 (1.00%) 0.703 (1.50%) 0.656 (1.50%)	0.412 (0.75%) 0.364 (0.75%) 0.532 (1.00%) 0.489 (1.00%) 0.706 (1.50%) 0.673 (1.50%)

Table 2. Summary of the predictive performance of models in western populations. *ACRIN,* American College of Radiology Imaging Network; *ACS CPS-I,* American Cancer Society Cancer Prevention Study I; *ACS CPS-II,* American Cancer Society Cancer Prevention Study II; *AIC,* Akaike information criterion; *ALT,* Alanine transaminase; *ATBC,* Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study; *AUC,* Area under the receiver operator curve; *BCCA,* British Columbia Cancer Agency; *BIC,* Bayesian information criterion; *CA125,* Cancer antigen 125; *CARET,* Carotene and Retinol Efficacy Trial; *CBMN,* Cytokinesis-blocked micronucleus assay; *CEA,* Carcinoembryonic Antigen; *CI,* Confidence interval; *CONOR,* Cohort of Norway; *COPD,* Chronic obstructive pulmonary disease; *CRP,* C-reactive protein; *CTAP-III,* Connective tissue-activating peptide III; *CTRL,* control; *CXCL9/MIG,* Chemokine ligand 9/monokine induced by gamma interferon *CXR,* Chest radiography; *CYFRA 21–1,* Cytokeratin-19 fragment; *EPIC,* European Prospective Investigation into Cancer and Nutrition; *EUELC,* European Early Lung Cancer project; *E/O,* ratio of estimated over observed probabilities; *FEV1,* Forced Expiratory Volume in one second; *Harvard/MGH,* Harvard School of Public Health and Massachusetts General Hospital Case-Control Study; *HDL,* High-density lipoprotein; *HE4,* Human epididymis protein 4; *HUNT2,* Nord-Trøndelag Health Study 2; *INR,* International normalised ratio; *KPSC,* Kaiser Permanente Southern California; *LCRAT,* Lung cancer risk assessment tool; *LDCT,* Low-dose computed tomography; *LDL,* Low-density lipoprotein; *LLP,* Liverpool Lung Project Cohort; *LSS,* Lung Screening Study; *LUSI,* Lung Cancer Screening Intervention Trial; *MDACC,* M.D. Anderson Cancer Center; *MES,* Medial EarlySign; *MGBH,* Mass General Brigham Hospital; *MGH,* Massachusetts General Hospital; *NA,* Not applicable; *NAP-2,* Neutrophil-activating peptide-2; *NCI,* National Cancer Institute; NHIS, National Health Interview Survey; *NIH-AARP,* National Institutes of Health-American Association of Retired Persons Diet and Health Study; *NLST,* National Lung Screening Trial; *NPV,* Negative predictive value; *NR,* Not reported; *NSCLC,* Non-small cell lung cancer; *NSHDS,* Northern Sweden Health and Disease Study; *NSW,* New South Wales; *PANCAN,* Pan-Canadian Early Detection of Lung Cancer; *PPV*, Positive predictive value; *PRS*, Polygenic risk score; *UK*, The United Kingdom; *US*, The United States.

distribution of traditional risk factors incorporated in Western and Asian prediction models. Across both settings, the top three risk factors consistently included are age, sex, and family cancer history. Based on the Venn diagram (Fig. 2), 22 overlapping risk factors were identified. However, there are still notable differences in the risk factors incorporated within each population. For instance, Asian models incorporate unique risk factors such as dietary habits[52] and exposure to radon and carcinogens[52], which are particularly relevant in the Asian context due to cultural and environmental differences that influence cancer risk. Conversely, Western models include specific risk factors such as hay fever[68,71,77,80,88,91,93,101], daily cough[62,72,90], and nicotine dependency[55,66], which reflect different comorbidities profiles, lifestyle exposures that are more prevalent or differently characterized in Western populations. Additionally, clinical risk factors and biomarkers including forced expiratory volume in one second as a percentage of the predicted value (FEV1%), CEA, CRP, and the rs2736100 SNP, were consistently incorporated across both Asian and Western prediction models (Supplementary Tables S4 and S5). Among models that integrated both traditional risk factors and biomarkers; six Western studies[59,63,65,68,94,102] and two Asian studies[54,87] directly compared their biomarker-based models with those incorporating only traditional risk factors, demonstrating better discrimination. For instance, Tammemägi et al. (2011)[65] reported the incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction (AUC = 0.773; 95% CI: 0.732 – 0.815) compared to base PLCO model (AUC = 0.718; 95% CI: 0.671–0.765). Thomas et al. (2020)[87] also reported that the inclusion of either serum or urinary cotinine, in addition to the number of cigarettes per day and years of smoking in the epidemiological model (AUC = 0.68; 95% CI: 0.64–0.71), resulted in a significant increase in the AUC by 0.04 (95% CI: 0.02–0.06, $p = 0.001$). Similarly, Li et al. (2022)[54] observed that compared to the logistic regression model predicted solely with epidemiological risk factors (age, gender, smoking intensity, smoking duration, family history), the addition of SNPs increased the AUC by 0.039 to 0.742 ($p < 0.001$). The comprehensive list of specific biomarkers and SNPs incorporated in both Asian and Western risk prediction models is provided in Supplementary Table S4 and Table S5, respectively.

Among studies that developed a new prediction model ($n = 49$), the majority were developed using statistical approaches such as multivariable logistic regression ($n = 23$)[12,27,52,54,55,69–71,74,76,81,84–86,88–90,92,93,96,99,101,102] and Cox proportional hazards regression ($n = 17$)[28,53,55,5859606164656673,75,78,86,90,94,98]. The remaining studies developed their models using parametric survival regression with Weibull distributions ($n = 2$)[63,100], flexible parametric survival analysis ($n = 1$)[80], and machine-learning approaches such as artificial neural network (ANN) ($n = 1$)[72], convolutional neural network (CNN) ($n = 3$)[56,57,97] and extreme gradient boosting (XGBoost) ($n = 4$)[54,62,67,79]. Among Western countries, the AUCs of the models developed using statistical approaches ranged from 0.601 to 0.920, while those developed using machine-learning approaches showed an AUC ranging from 0.704 to 0.881. Conversely, among Asian countries, the AUCs of the models developed using statistical approaches ranged from 0.639 to 0.867, while those developed using machine-learning approaches ranged from 0.680 to 0.759. Among models developed using machine-learning algorithms ($n = 10$) (8/42; 19.0% in Western studies and 2/12; 16.7% in Asian studies), only one Asian study[54] directly compared and demonstrating significant improvements using XGBoost in their full model (AUC = 0.759; 95% CI: 0.737–0.782) compared to logistic regression model (AUC = 0.742; 95% CI: 0.718–0.765) ($p < 0.001$). Overall, there has been little consistency in how lung cancer models have been reported thus far. While calibration measures including H-L goodness-of-fit and Brier scores were not consistently reported in the included studies, those that were tested exhibited relatively good individual predictions. Conversely, AUC and C-statistic have been the most reported metrics to determine model discrimination.

Synthesis of results

Eligible studies were combined to obtain an overall pooled AUC/C-statistic estimates. Recognizing that it may not be appropriate to combine estimates from all included studies due to heterogeneity in study design and model risk factors, meta-analysis was performed only on the same prediction models that have undergone numerous external validations. Figure 3 shows the forest plot for AUC/C-statistic of the eight models (Table 1), with $PLCO_{M2012}$ showing the best overall performance (AUC = 0.748; $p < 0.001$) when validated in external cohorts. A full description of the heterogeneity and mixed-effects model estimates can be found in Supplementary Table S6.

Risk of bias and applicability

Using PROBAST, Fig. 4 summarises the ROB and applicability for each domain and the overall judgment for each study. Among the 54 included studies, 21 were graded as low ROB, 26 were graded as unclear ROB, and 7 were graded as high ROB. High ROB was mainly observed in the analysis domain, where 10% of the studies did not account for optimism in model performance (e.g., bootstrap and cross-validation) and data complexities (e.g., censoring and competing risks). Unclear ROB was mainly due to incomplete statistical analysis information regarding missing data handling. Among included studies, most were deemed to have a low level of concern ($n = 47$) regarding model applicability, while a handful ($n = 7$) were rated as having an unclear concern. A full description of the ROB and applicability assessment for each study can be found in Supplementary Table S7.

Publication bias

Egger's test (Fig. 5) showed no statistically significant evidence of funnel plot asymmetry ($p > 0.05$) among the studies included in the meta-analysis. The trim-and-fill method was applied to studies with indications of funnel plot asymmetry, despite being statistically insignificant. After adjustment, the LLPi and Pittsburgh models had higher p-values for asymmetry of 0.869, and 0.845, respectively (Supplementary Figure S2), indicating a lower likelihood of publication bias for these models. The adjusted pooled estimates for these models are presented in Supplementary Table S6, and the overall conclusions remain consistent after adjustment.

Discussion

This is the first extensive systematic review evaluating the performance of lung cancer prediction models in both Western and Asian contexts, stratified by target population, study design, validation type, model development approaches, and model risk factors. Our review included 54 studies containing models for lung cancer risk prediction, of which 42 were from Western countries and 12 from Asian countries. In this review, we highlight that the majority of models predominantly examine lung cancer risk in Western populations, with models such as $PLCO_{M2012}$ gaining global recognition and acceptance in clinical practice. The $PLCO_{M2012}$ model was developed as an update to its predecessor ($PLCO_{M2011}$)[65] by addressing limitations and enhancing usability. The original $PLCO_{M2011}$ model, which used more complex modeling techniques such as restricted cubic splines, was considered cumbersome and did not incorporate certain relevant predictors. In contrast, the $PLCO_{M2012}$ model simplified the approach using multivariable fractional polynomials for more straightforward risk calculation, while introducing additional predictors such as race/ethnicity and personal cancer history, and excluding the chest radiography predictor. Furthermore, while the original model based risk estimates on a 9.2-year median follow-up, $PLCO_{M2012}$ truncated follow-up to 6 years for more accurate NLST comparisons. Our meta-analysis revealed that among the eight models that underwent extensive external validations, $PLCO_{M2012}$ demonstrated the best overall predictive discrimination when validated in external cohorts (Fig. 2). This finding is consistent with previous research by Gray et al. (2016)[35] and Tammemägi et al. (2022)[103], who reported that $PLCO_{M2012}$ may be more efficient than the USPSTF2013 criteria for selecting high-risk individuals for lung cancer screening programs. The robustness of $PLCO_{M2012}$ also aligns with findings by Weber et al. (2017)[64], who found that the model outperformed the categorical and dichotomized NLST criteria in identifying high-risk individuals aged 55–74 years. However, $PLCO_{M2012}$[12] is limited in its applicability, as it was developed exclusively in a cohort of ever-smokers in the US and may not be generalizable to never-smokers or Asian populations due to potential differences in baseline lung cancer risk, ethnic, and genetic variations[104–106]. Therefore, to enhance applicability to never-smokers, the $PLCO_{all2014}$ model[85] was developed as an extension of the $PLCO_{M2012}$ model[12], incorporating both never- and ever-smokers from the PLCO control arm and validated in the PLCO intervention arm. This model demonstrated high discrimination (AUC = 0.848) and robust calibration for risk thresholds below 0.10. However, compared to $PLCO_{M2012}$, the $PLCO_{all2014}$ model exhibited an overestimation of risks above 0.15. Furthermore, Cox recalibration analysis revealed that both the $PLCO_{all2014}$ and $PLCO_{M2012}$ models slightly overestimated the intercepts and original model logits (log odds) when assessed using the PLCO intervention arm data, suggesting that further adjustments may be needed to enhance model accuracy and better align with observed data, including their applicability to non-Western populations.

Ultimately, when deploying Western models within Asian populations, it is crucial to consider other well-established and prevalent risk factors (e.g., environmental tobacco smoke, asbestos exposure in certain Asian regions, and genetic factors associated with EGFR mutations)[107] to enhance the applicability and relevance of the model within Asian populations. Currently, it is challenging to directly compare the discriminatory performance between regions, as models are often validated in different populations with varying baseline risks and or follow-up times. To conclusively determine the impact of ethnic and genetic differences on the model's predictive performance in Asian and non-Asian populations, further studies directly evaluating predictive performance in both populations are necessary.

Target population

Given the well-established association between smoking and lung cancer[108,109], the majority of the Western models in our study were developed among ever-smokers. Our review reveals a significant gap in research on risk prediction models for never-smokers, as only a small proportion of studies (15%)[28,52,80,83,93,94,98,100] stratified participants based on smoking status, and only two Asian studies[27,98] focused on developing a model exclusively among never-smokers. However, the rise in lung cancer incidence among never-smokers, particularly in Asian countries[7–9] underscores the critical need to develop a personalized prediction model tailored specifically to this subgroup. Considering that traditional risk factors may not fully explain the risk among never-smokers, there has been a paradigm shift in research focus toward understanding lung cancer etiology and potential risk factors among never-smokers. For instance, studies have shown that environmental factors such as increased exposure to second-hand tobacco smoke and environmental pollutants (i.e., air pollution and carcinogens in the workplace or household) could further increase the risk of developing lung cancer at a younger age[20,110]. Other contributing risk factors, including genetic predisposition and the presence of driver gene mutations, also play a role in the increasing lung cancer incidence among female never-smokers, particularly in Asian populations. A research study conducted in China examined these mutations in never-smoking Chinese females diagnosed with lung adenocarcinoma and the findings revealed a high mutation rate of 76% for the *EGFR* gene[111]. Sun et al.[112] and Ren et al.[113] reported comparable rates of genetic changes, with *EGFR* mutations detected in 79% and 70% of cases, respectively, among never-smoking Chinese women. These findings indicate that oncogenic mutations, particularly involving *EGFR*, predominantly drive lung cancer development in this subgroup. Therefore, understanding the interplay between genetic susceptibility and environmental factors holds promise in informing the development of a refined prediction model tailored to accurately assess lung cancer risk among Asian never-smoking populations.

Study design and validation type

The majority of the models developed among Western and Asian countries were developed using cohort studies (64%; 27/42) and case-control studies (58%; 7/12), respectively. In current practice, cohort[114] and case-control[115]designs are predominantly used for model development using data from healthcare databases or registries. Alternative study designs include nested case-control[116] and case-cohort[117] studies. Theoretically, case-control designs are generally less desirable for developing prediction models for several reasons. Firstly,

Model	Country or region	Risk Factors	Model Type (Variable selection approach)	Internal Validation method	Study Design	Study Population	Length Risk (year)	Number of lung cancer cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Park et al. (2013)[86]	Korea	Age, smoking status, smoking intensity, age at smoking initiation, BMI, physical activity, fasting glucose level	Cox Proportional Hazards Regression (stepwise selection)	NR	Population-based	Korean Men	8	10,007	**H-L goodness of fit**: ***Korea Men***: $P < 0.001$	**C-index (95% CI)** ***Korea Men***: 0.864 (0.860–0.868)	NR	**C-index (95% CI)** ***KNHC***: 0.871 (0.867–0.876)	NR	NR
Lin et al. (2012)[69]	China	Sex, smoking status, history of lung disease, occupational exposure, family history	Multivariable Logistic Regression (stepwise selection)	10-fold cross validation	Case-control	Chinese General Population	Clinical pretest lung cancer risk	633	**H-L goodness of fit** ***Guangdong***: $P = 0.62$	**AUC (95% CI)** ***Guangdong***: 0.75 (0.72–0.78)	NR	NR	NR	NR
Tse et al. (2022)[52]	Hong Kong	Smoking and smoking cessation, education, lung disease history, family history of cancer, residential radon exposure, dietary habits, carcinogens exposure, mask use and dust control	Multivariable Logistic Regression (stepwise selection)	10-fold cross validation	Case-control	Hong Kong Men	1	1,069	**H-L goodness of fit**: ***Hong Kong***: **Never-smokers**: $P = 0.493$ **Former-smokers**: $P = 0.260$ **Current-smokers**: $P = 0.502$	**AUC (95% CI)** ***Hong Kong***: **Overall**: 0.735 (95% CI, 0.714–0.756) **Never-smokers**: 0.710 (0.657–0.762) **Former-smokers**: 0.741 (0.705–0.777) **Current-smokers**: 0.710 (0.668–0.752)	NR	NR	**PPV**: 0.818 (2.83%)	**NPV**: 0.674 (2.83%)
Park et al. (2021)[53]	Korea	Age, sex, pack-years of smoking, years since cessation, physical activity, alcohol consumption, BMI, medical history of COPD, emphysema, pneumoconiosis, and interstitial pulmonary disease	Cox Proportional Hazards Regression (expert-driven selection)	Train-test spilt validation	Population-based	Korean Ever-Smokers (Light- and Heavy-smokers)	6.6	7,767 (training set) 3,368 (validation set)	**E/O ratio (95% CI)** [Korean Model] ***NHIS-Korea***: 0.989 (0.956–1.023) [Bach Model] ***NHIS-Korea***: 2.23 (2.07–2.40) [LCRAT Model] ***NHIS-Korea***: 4.73 (4.50–4.96) [$PLCO_{M2012}$] ***NHIS-Korea***: 1.24 (1.22–1.26) [Pittsburgh Predictor] ***NHIS-Korea***: 1.21 (1.19–1.23) [LLPi] ***NHIS-Korea***: 3.25 (3.20–3.31)	**C-index (95% CI)** *[Korean Model]* ***NHIS-Korea***: 0.816 (0.806–0.826) [Bach Model] ***NHIS-Korea***: 0.661 (0.598–0.665) [LCRAT Model] ***NHIS-Korea***: 0.811 (0.807–0.814) [$PLCO_{M2012}$] ***NHIS-Korea***: 0.772 (0.768–0.777) [Pittsburgh Predictor] ***NHIS-Korea***: 0.781 (0.778–0.784) [LLPi] ***NHIS-Korea***: 0.803 (0.800–0.806)	NR	NR	0.601 (3.7%)	0.830 (3.7%)

Continued

Model	Country or region	Risk Factors	Model Type (Variable selection approach)	Internal Validation method	Study Design	Study Population	Length Risk (year)	Number of lung cancer cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Yeo et al. (2021)[58]	Korea	Age, sex, smoking intensity, BMI, COPD, pulmonary TB and Type 2 diabetes mellitus	Cox Proportional Hazards Regression (expert-driven selection)	Bootstrap (100 resampling)	Cohort	Korean general population	5	16,747 (training set) 7,115 (validation set)	Accessed graphically	**C-index (95% CI)** ***KNHI***: 0.825 (0.810–0.840)	NR	NR	NR	NR
Continued														

Model	Country or region	Risk Factors	Model Type (Variable selection approach)	Internal Validation method	Study Design	Study Population	Length Risk (year)	Number of lung cancer cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Wang et al. (2023)[98]	China	**Never-smokers**: Age, sex, BMI, history of lung cancer in first-degree relatives, history of chronic respiratory diseases **Ever-smokers**: Age, BMI, cigarettes per day, years the person has smoked	Cox Proportional Hazards Regression (expert-driven selection)	Train-test spilt validation and bootstrap	Case-control	Chinese Never- and Ever-Smokers (Light- and Heavy-smokers)	3	925/538 (training set: never-/ever-smokers) 841/421 (validation set A: never-/ever-smokers) 503/127 (validation set B never-/ever-smokers)	NA	NA	**E/O Ratio (95% CI) [China NCC-LC model] Never-smokers**: **NLCS**: 0.958 (0.802–1.124) **Ever-smokers**: **NLCS**: 0.975 (0.815–1.126)	**AUC (95% CI) [China NCC-LC model] Never-smokers**: **NLCS**: 0.698 (0.682–0.714) **WHOLE**: 0.673 (0.650–0.695) **Ever-smokers**: **NLCS**: 0.728 (0.707–0.750) **WHOLE**: 0.752 (0.715–0.789) [Bach model] **NLCS**: 0.662 (0.608–0.615) ***WHOLE***: 0.652 (0.555–0.749) [Spitz model] **NLCS**: 0.723 (0.686–0.760) ***WHOLE***: 0.635 (0.518–0.751) [LLP model] **NLCS**: 0.598 (0.546–0.650) ***WHOLE***: 0.556 (0.500–0.691) [Hoggart model] **NLCS**: 0.544 (0.500–0.591) ***WHOLE***: 0.659 (0.523–0.795) [$PLCO_{M2012}$] **NLCS**: 0.706 (0.660–0.752) ***WHOLE***: 0.668 (0.560–0.775) [LLPi model] **NLCS**: 0.659 (0.613–0.705) ***WHOLE***: 0.567 (0.500–0.712) [Pittsburgh Predictor] **NLCS**: 0.665 (0.615–0.715) ***WHOLE***: 0.630 (0.530–0.730) [LCRAT model] **NLCS**: 0.687 (0.639–0.735) ***WHOLE***: 0.679 (0.585–0.773)	0.24 (0.47%)	0.90 (0.47%)

Continued

Model	Country or region	Risk Factors	Model Type (Variable selection approach)	Internal Validation method	Study Design	Study Population	Length Risk (year)	Number of lung cancer cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Li et al. (2012)[74]	China	Smoking status, 4 SNPs (rs2736100, rs402710, rs4488809 and rs4083914)	Multivariate Logistic Regression (expert-driven selection)	Bootstrap (1,000 resampling)	Case-control	Han Chinese General Population	1	2,283	**H-L goodness of fit** ***Han Chinese Cohort***: $P = 0.154$	**AUC (95% CI)** ***Han Chinese Cohort***: 0.639 (0.621 – 0.652)	NR	NR	NR	NR
Chien et al. (2020)[27]	Taiwan	Age, BMI, COPD, education, family history of lung cancer, 9 SNPs	Multivariate Logistic Regression (expert-driven selection)	Bootstrap (2,000 resampling)	Case-control	Taiwanese Female Never-Smokers (55–70 years old)	6	1,748	NR	**AUC (95% CI)** **TWB2-LCPG**: 0.714 (0.660 – 0.768) ***TWB1-GELAC***: 0.770 (0.749 – 0.791)	NR	NR	0.419 (1.34%) 0.361 (1.51%) 0.242 (2.0%)	0.919 (1.34%) 0.938 (1.51%) 0.975 (2.0%)
Wu et al. (2016)[28]	Taiwan	Age, sex, pack-years, family history of lung cancer, personal cancer history, BMI, lung function test, serum biomarkers (CEA, AFP, and CRP)	Multivariable Cox Regression (NR)	Bootstrap (100 resampling)	Cohort	Taiwanese General Population	5 and 10	1,117	Assessed Graphically	**AUC (95% CI)** ***MJHG***: Overall: 0.851 (0.840 – 0.862) Never-Smokers: 0.806 (0.790 – 0.819) Former Smokers: 0.873 (0.829 – 0.879) Current Smokers: 0.875 (0.864 – 0.887) Light Smokers: 0.847 (0.824 – 0.871) Heavy Smokers: 0.732 (0.708 – 0.752)	NR	NR	NR	NR
Thomas et al. (2020)[87]	China	**Model 1**: Number of cigarettes, years of smoking **Model 2**: Model 1 + serum cotinine **Model 3**: Model 1 + urinary cotinine	Multivariate conditional logistic regression (NR)	NR	Case-control	Ever-smokers (Light- and Heavy-smokers)	NR	452	**Log-likelihood ratio Shanghai cohort**: Model 1 to 4: $P < 0.001$	**AUC (95% CI)** ***Shanghai cohort***: Model 1: 0.68 (0.64–0.71) Model 2: 0.72 (0.69–0.75) Model 3: 0.72 (0.69–0.75)	NR	NR	NR	NR

Continued

Model	Country or region	Risk Factors	Model Type (Variable selection approach)	Internal Validation method	Study Design	Study Population	Length Risk (year)	Number of lung cancer cases	Predictive Performance				Sensitivity (Risk Threshold)	Specificity (Risk Threshold)
									Internal Validation		External Validation			
									Calibration	Discrimination	Calibration	Discrimination		
Li et al. (2022)[54]	China	**Epidemiological Model**: Age, gender, smoking intensity, smoking duration, family history **Full Model**: Epidemiological Model + 61 SNPs	Logistic regression (LR) and extreme gradient boosting (XGBoost) (expert-driven selection)	10-fold cross validation	Case-control	Han Chinese General Population	NR	974	NR	**AUC (95% CI)** ***Han Chinese cohort***: Epidemiological Model (LR): 0.703 (0.681–0.726) Epidemiological Model (XGBoost): 0.744 (0.721–0.766) Full Model (LR): 0.742 (0.718–0.765) Full Model (XGBoost): 0.759 (0.737–0.782)	NR	NR	Full Model (LR): 0.663 Full Model (XGBoost): 0.638	Full Model (LR): 0.699 Full Model (XGBoost): 0.746
Lee et al. (2022)[97]	Korea	Age, sex, smoking status and chest radiograph	Convolutional neural network (CNN) (NR)	NR	Cohort	Korean Ever-Smokers (Heavy-smokers) & who have undergone lung cancer screening	12	6,768 (3-year) 4,874 (5-year)	NA	NA	Assessed graphically	**AUC (95% CI)** ***Seoul-NUH***: 0.68 (0.62–0.73)	0.90 (3.297%)	0.27 (3.297%)

Table 3. Summary of the predictive performance of models in Asian populations. *AFP,* Alpha-fetoprotein; *AUC,* Area under the receiver operator curve; *BMI,* Body mass index; *CEA,* Carcinoembryonic antigen; *CI,* Confidence interval; *COPD,* Chronic obstructive pulmonary disease; *GELAC,* Genetic Epidemiology Study of Lung Adenocarcinoma in Taiwan; *H-L,* Hosmer-Lemeshow; *hsCRP,* high-sensitivity C-reactive protein; *KNHC,* Korean National Health Corporation; *KNHI* Korean National Health Insurance, *LCPG* Taiwan Lung Cancer Pharmacogenomics Study, *LDL-C* Low-density lipoprotein-cholesterol, *NHIS-Korea* The National Health Insurance Corporation, *NLCS,* National Lung Cancer Screening; *NPV,* Negative predictive value; *NA,* Not applicable; *NR,* Not reported; *MJHG,* MJ Health Group; *PPV,* Positive predictive value; *Seoul-NUH,* Seoul National University Hospital; *SNPs,* Single nucleotide polymorphisms;*TB,* Tuberculosis; *TWB,* Taiwan Biobank data; *WHOLE,* Whole Life Cycle of Cancer Screening.

the design may not be well-suited to estimate the absolute risks of outcomes, which is often a key goal of risk prediction modelling[118]. Secondly, the design is highly susceptible to temporal bias affecting the model's ability to predict future events and often leads to miscalibrated predictions[119]. However, in cases where a case-control design is deemed necessary, it is crucial to construct the design with great care. To ensure a fair model evaluation, it is imperative that researchers recalibrate and prospectively validate the model using a cohort design to mitigate potential limitations inherent in case-control designs[120]. A major concern is that prediction models derived from case-control studies lack longitudinal risk or time-to-event data, which hampers the assessment of their robustness and reliability across different timeframes. Conversely, for models developed using prospective cohorts, comparing discriminatory performance can also be challenging due to differences in follow-up durations. Consequently, we recommend that future research utilize time-dependent metrics[121], such as AUCs or hazard ratios, and conduct temporal validation[122] to provide a more nuanced assessment of model performance and robustness across different timeframes.

In addition, although the Asian risk models in our study have shown promising results, the majority (75.0%) of the Asian studies lack external validation. One potential reason is that many Asian-developed models are relatively new compared to Western models like the Bach model (2003)[10], Spitz model (2007)[93], and LLP model (2008)[11]. Consequently, these newer models may have focused more on internal validation within their own populations or healthcare settings before pursuing external validation. Furthermore, Asian models are often tailored to the specific epidemiological and genetic characteristics of their local populations, resulting in less emphasis on validation in non-Asian populations where risk factors and disease profiles differ. However, this is concerning as internally validated models are highly dependent on the original study sample and may not address selection bias with missing data[123]. Model performance is likely to be overestimated in the absence of additional internal validation methods like cross-validation[124,125] and bootstrapping[126,127]. Conversely, external validation aims to evaluate a model's generalizability in new populations[123] to prevent overestimated performance due to chance or data snooping. Consequently, prediction models typically exhibit worse discrimination, which is closer to their true discriminatory ability when applied in clinical practice[128,129]. Therefore, while internally validated prediction models generally perform better during development phases, current Asian models and future model development studies, should prioritize external validation (gold standard) to better assess the model's generalizability in other populations[128,130].

Model development approaches

Machine learning has emerged as a pivotal tool for healthcare providers and clinicians in the early screening, prediction, and/or prognosis of various diseases[131–133]. While skepticism persists concerning the practical implementation and interpretation of findings derived from machine learning in healthcare settings, the adoption of these methods is rapidly growing. In our review, the predominant approach in model development involved traditional statistical methods, notably regression analyses. Conversely, machine-learning algorithms were employed in eight models originating from Western studies and two models from Asian contexts. In our study, Lu et al. (2020)[57] developed and validated a deep-learning model that uses chest radiographs to identify high-risk PLCO smokers for lung cancer screening. By applying advanced deep learning algorithms to radiographic images and integrating these findings with patient data, the model refines risk prediction and improves screening efficiency by offering an additional layer of risk assessment and supporting more targeted and effective interventions for high-risk smokers. In addition, based on a systematic review and meta-analysis conducted by Kanan et al. (2024)[134], AI-driven models, including convolutional neural networks (CNNs), support vector machines (SVMs), and ensemble methods, have significantly advanced lung cancer diagnosis and prognosis. These models analyze imaging data, pathology slides, and patient demographics, demonstrating promising results in improving diagnostic accuracy by identifying malignancies with high sensitivity and specificity, often outperforming traditional diagnostic methods[134]. However, despite the relatively high predictive discrimination of machine-learning models assessed in our review, the dependency on large amounts of data, sophisticated imaging equipment, research funding, and ethical issues for model development makes it challenging to adopt these models in primary healthcare settings[135]. Moreover, identifying the specific variables utilized in developing the machine learning prediction models is challenging, which could potentially limit the quality and authenticity of these models. Current literature also suggests the importance of prioritizing the development of machine learning-based models using decentralized and non-parametric data[136], which directly process raw data, thereby reducing heterogeneity and data distribution assumptions compared to traditional methods[37].

Risk factors included in the model

The impetus for biomarker research has grown considerably after the NLST trial demonstrated a significant reduction in lung cancer mortality risks through LDCT screening[137]. Subsequently, numerous prediction studies have incorporated genetic and serum-based inflammation markers associated with lung cancer. Among Western studies in our review, Tammemägi et al. (2011)[65] reported the incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction, while Li et al. (2022)[54] found that adding SNPs to the epidemiologic-based logistic regression significantly improved the model's predictive performance in a Chinese population. In our study, it is noteworthy that clinical risk factors, including FEV1% and biomarkers such as CEA, CRP, and the rs2736100 SNP, were consistently incorporated across both Asian and Western prediction models. This consistent inclusion underscores the critical importance and utility of FEV1% in evaluating respiratory health and its strong predictive value for lung cancer across diverse populations[138]. Reduced FEV1% levels have been strongly associated with an increased lung cancer risk, particularly in smokers, as impaired lung function may reflect pathological processes that predispose individuals to malignancy. Conversely, CEA and CRP are widely recognized biomarkers associated with tumor activity and chronic inflammation and have been consistently linked to different cancer types. Elevated levels of CRP[139–141] and CEA[142–144] are frequently

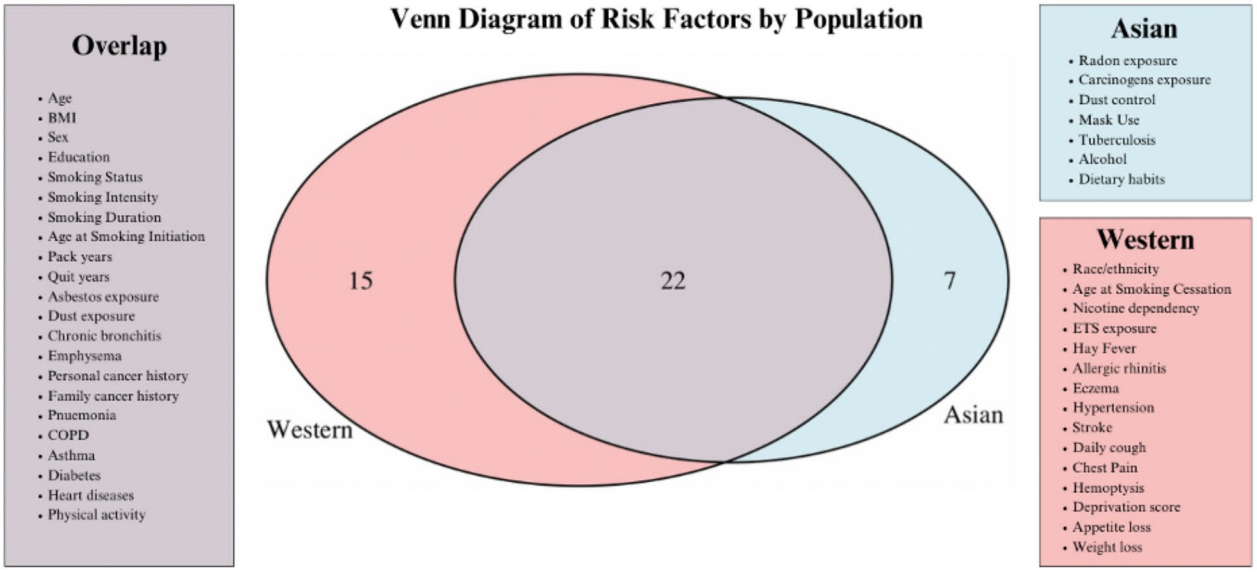

Fig. 2. Venn Diagram illustrating the distribution and overlap of traditional risk factors in Western and Asian risk prediction models.

observed and have been associated with poor prognosis among lung cancer patients, highlighting their potential utility in enhancing early detection and monitoring of the disease. The widespread use of FEV1%, CEA, and CRP highlights their reliability as a standardized measure, making it a key component in risk assessments across different demographic or geographic contexts. Additionally, rs2736100, located in the telomerase reverse transcriptase (TERT) gene, has been extensively studied and is strongly associated with an increased risk of lung cancer due to its role in regulating telomere length[145,146]. Despite its inclusion across diverse prediction models, studies have shown that individuals of Asian descent have a higher prevalence of the C allele of TERT rs2736100 compared to Caucasians[147] and this allele is particularly associated with an increased risk of lung adenocarcinoma in never-smoking Asian women, with effect sizes significantly larger than those reported in European smokers[148]. Consequently, while rs2736100 shows a strong association with lung cancer in certain Asian populations, the significance and reliability of this genetic marker for risk assessments in different geographical regions remain a topic of ongoing debate. Variations in allele frequency, genetic background, and environmental exposures across populations suggest that further research is needed to fully understand the role of rs2736100 in lung cancer risk globally[149]. Furthermore, while studies included in our review consistently showed improved predictive power with the addition of biomarkers, the diversity of included biomarkers may pose a challenge when directly comparing model performance across populations with diverse sociocultural contexts. The relatively smaller sample sizes in Asian studies with biomarkers may lead to underpowered analyses, and caution is warranted when drawing conclusions about the relative performance of biomarker-based models in Asian populations.

Ultimately, the integration of biomarkers into predictive models presents a promising avenue for enhancing the accuracy of lung cancer risk prediction and improving patient outcomes. Future research could refine risk estimates by using a joint-effects approach to model multiple diagnostic criteria markers. This approach aims to identify well-poised biomarkers to predict lung cancer risk within specific populations[137,150]. Nevertheless, it is essential to acknowledge the uncertainty regarding the applicability of employing models that integrate both clinical and biomarker data to the general population. This uncertainty arises from the fact that utilizing clinical and molecular information for risk assessment necessitates physical examination and sample collection, thereby introducing complexity into the decision-making process. Future studies could investigate the clinical utility and cost-effectiveness of biomarker-based models across both Western and Asian settings to validate the use of biomarkers in prediction models to facilitate population-based lung cancer screening.

Patient-centered outcomes and implementation feasibility

When evaluating lung cancer risk prediction models, it is critical to address patient-centered outcomes and the practical feasibility of implementing these models across diverse healthcare settings. For instance, while models such as $PLCO_{M2012}$[12] and LCRAT[61] have shown high predictive accuracy—particularly with $PLCO_{M2012}$outperforming NLST criteria in selecting high-risk individuals for lung cancer screening[64]—their real-world effectiveness relies heavily on patient engagement and adherence to screening protocols[151]. In resource-limited settings, the feasibility of widespread lung cancer screening programs is often hindered by restricted access to LDCT due to its high cost, limited availability, and the requirement for trained personnel to perform and interpret scans[107,152]. Additionally, the lack of comprehensive patient education presents a significant challenge, as individuals may not fully understand the importance of screening, the risks associated with smoking, or the benefits of early detection[153]. Lower health literacy, cultural barriers, and distrust in

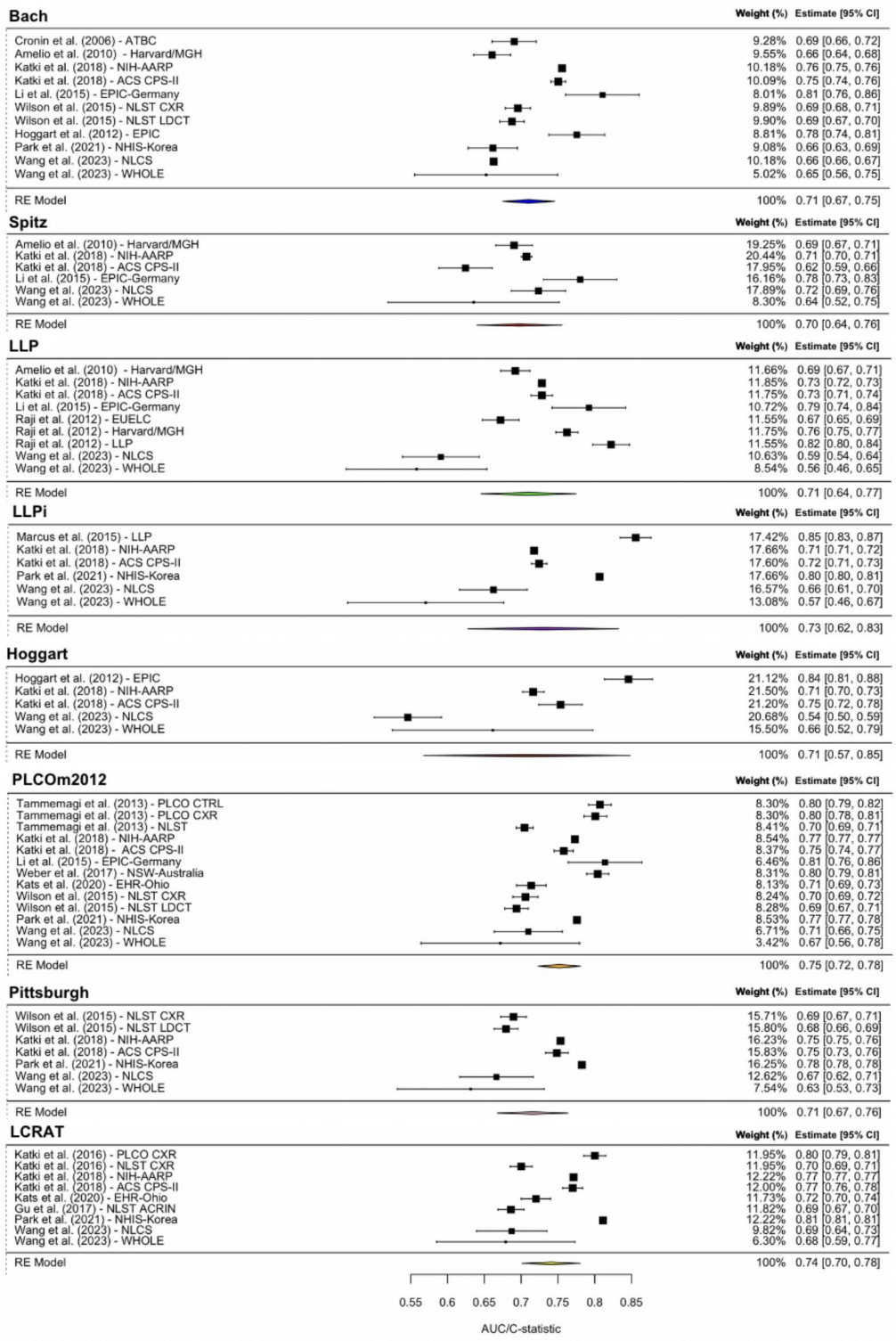

Fig. 3. Forest plots of AUC/C-statistic of eight lung cancer risk prediction models that have undergone numerous validation in external cohorts.

healthcare systems can further impede patient engagement, leading to lower participation in screening and reduced adherence to follow-up care[107,154–156]. These factors ultimately diminish the effectiveness of lung cancer risk prediction models in preventing and detecting the disease early in diverse healthcare settings. Also, most Asian models are developed based on the specific epidemiological and genetic characteristics of their local populations[27,53,86]. While these models are effective within their contexts, their applicability in non-Asian settings may be limited due to differences in disease risk profiles, prevalence of risk factors, and healthcare

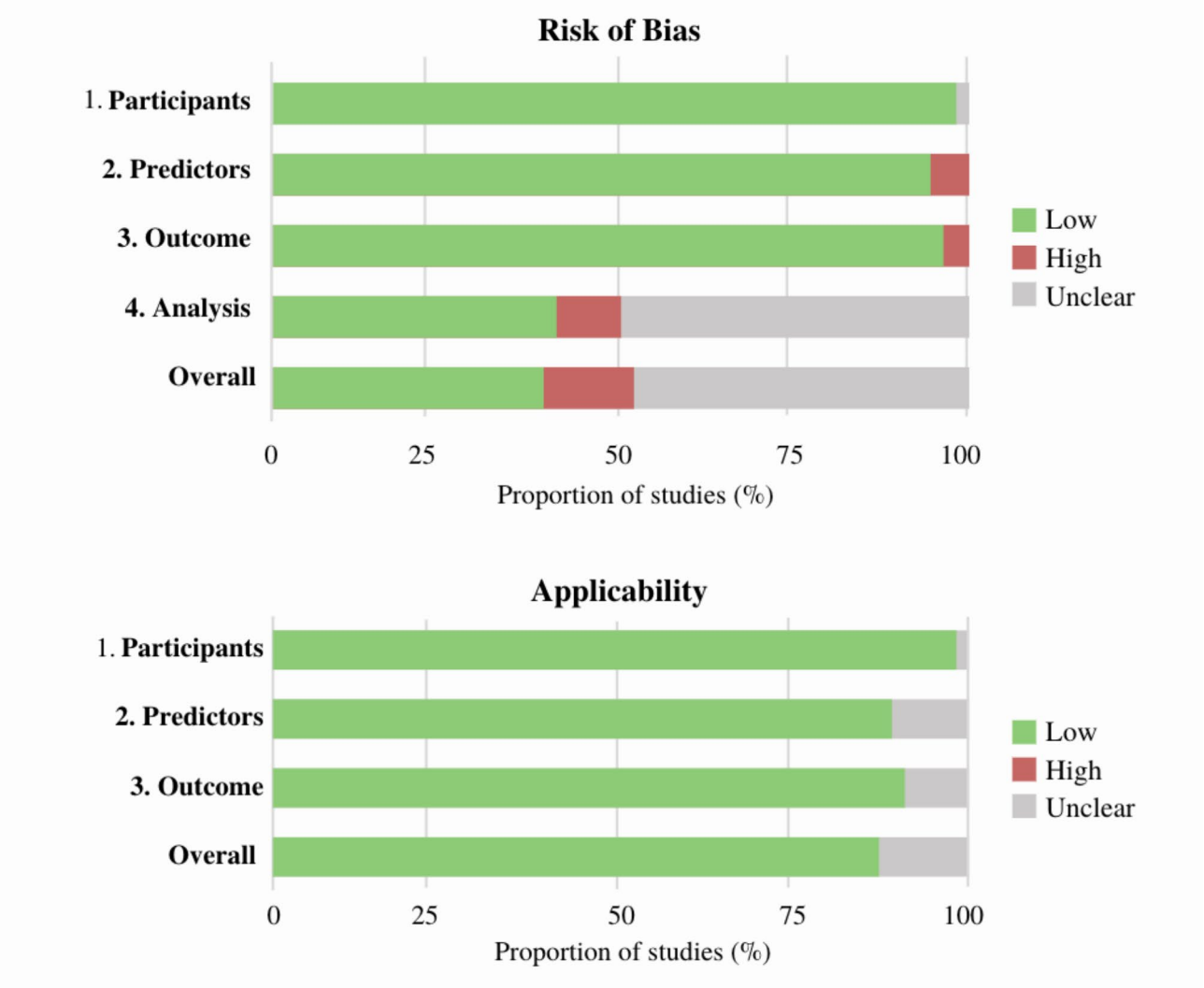

Fig. 4. Bar charts showing the risk of bias and applicability ratings for each domain and the overall judgment, for the 54 included studies. "Overall" indicates the overall risk of bias; "participants" indicates bias introduced by participants or data sources; "predictors" indicates bias introduced by predictors or their assessment; "outcome" indicates bias introduced by outcomes or their assessment; "analysis" indicates bias introduced by the analysis. Values in the bars are proportion (%).

practices, which can affect the model's predictive accuracy and relevance in diverse healthcare settings. By incorporating patient-centered outcomes—such as satisfaction with risk communication and willingness to participate in screening[156]—into the evaluation process, we can more accurately assess the broader clinical applicability and true impact of these models in enhancing lung cancer prevention and early detection efforts globally. Furthermore, when assessing lung cancer prediction models, it is important to distinguish between pre-test risk models and those that incorporate post-screening data. Pre-test models use demographic and clinical data to identify high-risk individuals before screening, while post-screening models incorporate CT findings to enhance predictive accuracy through real-time tumor characteristics, focusing on ongoing risk assessment and changes in risk status over time. Together, these complementary approaches can enhance clinical decision-making for both early detection and longitudinal patient care in lung cancer management.

Strengths and limitations

For our review, we strictly adhered to the guidelines for conducting systematic reviews and meta-analyses and used stringent eligibility criteria and quality assessment tools to assess included studies. However, our study must be viewed in consideration of several limitations. Firstly, due to the heterogeneity between studies in terms of study population, model development, and variable selection approaches, we were only able to meta-analyze the predictive performance of the same model across multiple external cohorts. In our analysis, we observed that adjusting for between-study heterogeneity led to slightly wider confidence intervals, reflecting the natural variability among studies and a modest increase in uncertainty in the pooled estimates. However, the overall conclusions remain robust, and this adjustment enhances the reliability of our findings for broader population

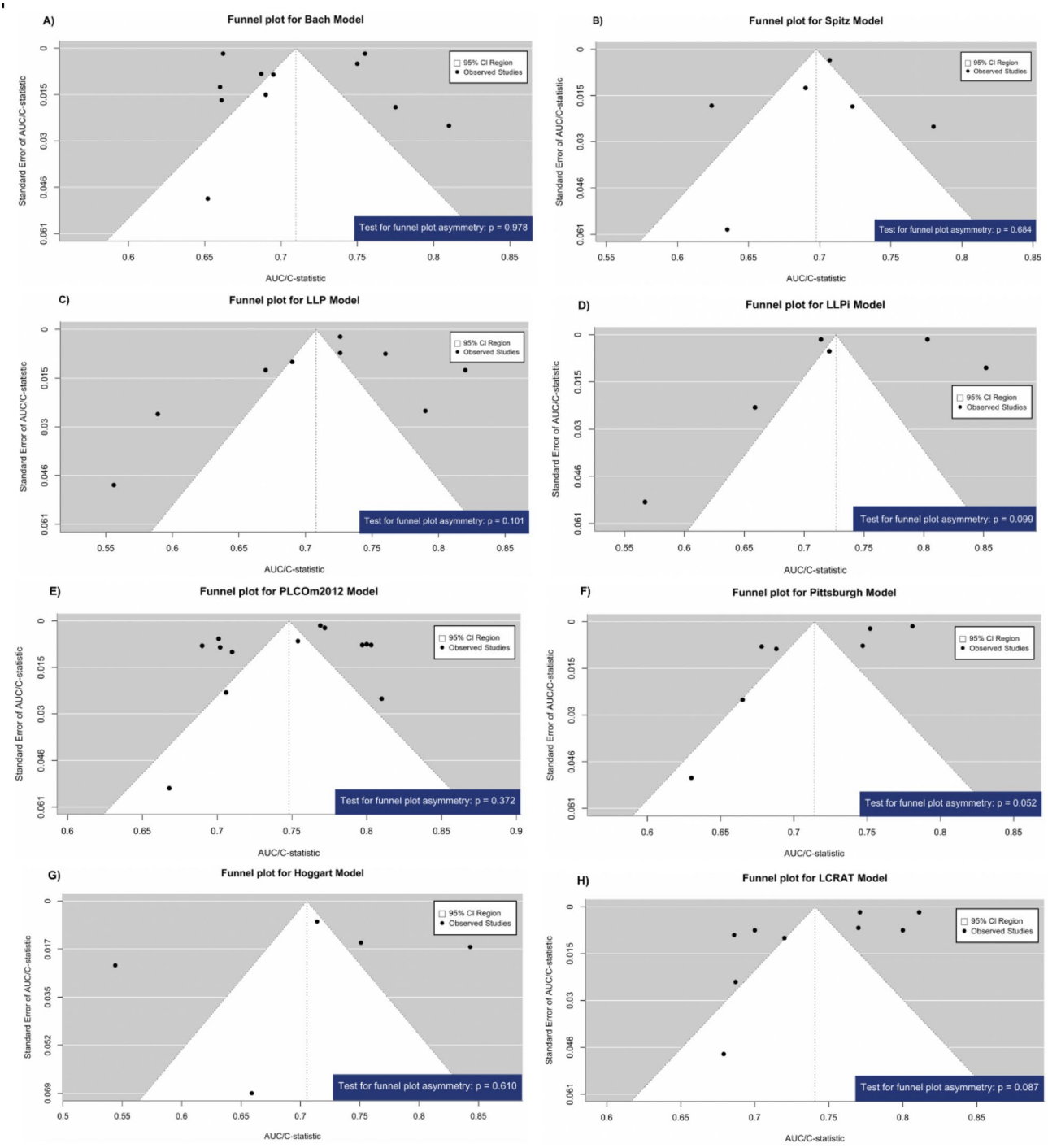

Fig. 5. Funnel plot of studies included in meta-analysis. Each plotted point represents the standard error and AUC/C-statistic estimates of (**A**) Bach, (**B**) Spitz, (**C**) LLP, (**D**) LLPi, (**E**) $PLCO_{M2012}$, (**F**) Pittsburgh, (**G**) Hoggart, and (**H**) LCRAT models. The white triangle represents the region where 95% of the data points would lie in the absence of publication bias. The vertical line represents the average AUC/C-statistic of each model.

applicability. Secondly, as a handful of studies in our analysis showed high ROB due to the absence of optimism consideration in model performance and data complexities, their reported study findings should be interpreted with caution. Thirdly, our review focuses on incidence-based models and does not include mortality-based models, which, while offering valuable insights, fall outside the scope of our current work.

Future work

To inform future guidelines for the selection of high-risk individuals for lung cancer screening across diverse sociocultural contexts, it is crucial that future studies focus on validating multiple prediction models using

the same datasets for direct comparisons[157] and rigorously test these models in large populations with diverse ages and ethnicities[158]. Additionally, it is crucial to carefully assess the implications of employing different study designs, variable selection approaches, and modelling techniques in developing prediction models, as these choices may greatly influence model performance and generalizability across diverse settings. While less frequently reported, performance metrics such as Net Reclassification Improvement (NRI) and Decision Curve Analysis (DCA) provide valuable insights into the clinical utility of prediction models. NRI quantifies the model's improvement in risk classification over existing models[159], while DCA assesses net benefits across threshold probabilities, balancing false positives and negatives to inform clinical decisions[160]. Including these metrics enhances the evaluation's completeness and highlights their practical implications for lung cancer management and outcomes. Future research should also prioritize identifying effective combinations of biomarkers and clinical risk factors, tailored to unique population characteristics, to accurately predict lung cancer risk. For instance, future prediction models could explore integrating results from emerging non-invasive technologies like circulating tumor DNA (ctDNA)[161,162], which was approved in 2015 by the U.S. Food and Drug Administration[163], to detect *EGFR* mutations in the bloodstream, particularly in Asian populations. This enables early identification of precancerous genetic alterations or mutations associated with lung cancer, potentially enhancing the accuracy of risk predictions before clinical symptoms or tumor imaging are evident. Lastly, it is crucial to regularly update and recalibrate prediction models to maintain their accuracy and relevance[164]. Initially developed using specific datasets, models must adapt to evolving demographics, healthcare practices, and sociocultural behaviours such as rising lung cancer rates among never-smokers in Asia and advancements in screening technologies like LDCT. Updating involves re-estimating parameters with recent data to reflect current population characteristics and risk factor distributions[122], while recalibration incorporates new risk factors or biomarkers, enhancing the model's predictive performance and clinical utility[165]. Regular updates ensure models remain accurate and relevant for current risk assessments and targeted interventions.

Conclusion

Despite the significant proportion of lung cancer cases observed among never-smokers, particularly in Asian settings, our review reveals a significant gap in prediction models for this population. Future research should focus on externally validating existing Asian models or adapting widely used Western models, such as $PLCO_{M2012}$, to incorporate pertinent Asian risk factors (e.g., asbestos exposure, and genetic factors associated with EGFR mutations) to optimize risk-based screening for this population. Ultimately, given the unique sociocultural contexts in Western and Asian countries, it is crucial to meticulously develop prediction models that account for the distinct risk profiles and lung cancer progression patterns of the target populations.

Data availability

The datasets generated and analyzed during the current study are derived from published studies and are available in the public domain. Detailed references for each included study are provided in the manuscript.

Received: 18 June 2024; Accepted: 18 December 2024
Published online: 04 March 2025

References

1. Thandra, K. C. et al. Epidemiology of lung cancer. *Contemp. Oncol. (Pozn).* **25**(1), 45–52 (2021).
2. O'Dowd, E., Mackenzie, J. & Balata, H. Lung cancer for the non-respiratory physician. *Clin. Med. (Lond)* **21**(6), e578–e583 (2021).
3. Henschke, C. I. et al. Survival of patients with stage I lung cancer detected on CT screening. *N Engl. J. Med.* **355**(17), 1763–1771 (2006).
4. Aberle, D. R. et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl. J. Med.* **365**(5), 395–409 (2011).
5. de Koning, H. J. et al. Reduced lung-Cancer mortality with volume CT screening in a Randomized Trial. *N Engl. J. Med.* **382**(6), 503–513 (2020).
6. Krist, A. H. et al. Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *Jama* **325**(10), 962–970 (2021).
7. Ko, Y. H. et al. Risk factors for primary lung cancer among never-smoking women in South Korea: a retrospective nationwide population-based cohort study. *Korean J. Intern. Med.* **35**(3), 692–702 (2020).
8. Zhou, F. & Zhou, C. Lung cancer in never smokers-the east Asian experience. *Transl Lung Cancer Res.* **7**(4), 450–463 (2018).
9. Toh, C. K. et al. A decade of never-smokers among Lung Cancer patients-increasing Trend and Improved Survival. *Clin. Lung Cancer* **19**(5), e539–e550 (2018).
10. Bach, P. B. et al. Variations in lung cancer risk among smokers. *J. Natl. Cancer Inst.* **95**(6), 470–478 (2003).
11. Cassidy, A. et al. The LLP risk model: an individual risk prediction model for lung cancer. *Br. J. Cancer* **98**(2), 270–276 (2008).
12. Tammemägi, M. C. et al. Selection criteria for lung-cancer screening. *N Engl. J. Med.* **368**(8), 728–736 (2013).
13. Zhou, W. & Christiani, D. C. East meets West: ethnic differences in epidemiology and clinical behaviors of lung cancer between East asians and caucasians. *Chin. J. Cancer* **30**(5), 287 (2011).
14. Torre, L. A., Siegel, R. L. & Jemal, A. *Lung cancer statistics.* Lung cancer and personalized medicine: current knowledge and therapies, : pp. 1–19. (2016).
15. Zhou, F. & Zhou, C. Lung cancer in never smokers—the east Asian experience. *Translational lung cancer Res.* **7**(4), 450 (2018).
16. Wang, F. et al. Lung cancer risk in non-smoking females with a familial history of cancer: a multi-center prospective cohort study in China. *J. Natl. Cancer Cent.* **1**(3), 108–114 (2021).
17. Lin, H. et al. A family history of cancer and lung cancer risk in never-smokers: a clinic-based case-control study. *Lung Cancer* **89**(2), 94–98 (2015).
18. Myers, R. et al. High-ambient Air Pollution exposure among never smokers versus ever smokers with Lung Cancer. *J. Thorac. Oncol.* **16**(11), 1850–1858 (2021).
19. Kim, A. S. et al. *Exposure to secondhand smoke and risk of Cancer in Never smokers: a Meta-analysis of epidemiologic studies. Int. J. Environ. Res. Public. Health* **15**(9). (2018).
20. Chen, K. C. et al. Indoor air pollution increases the risk of lung cancer. *Int. J. Environ. Res. Public Health* **19**(3), 1164 (2022).

21. Mu, L. et al. Indoor air pollution and risk of lung cancer among Chinese female non-smokers. *Cancer Causes Control* **24**(3), 439–450 (2013).
22. Yin, X. et al. Dietary factors and the risk of lung cancer by epidermal growth factor receptor mutation status and histological subtypes. *Front. Public. Health* **10**, 1079543 (2022).
23. Park, S. Y. et al. Diet quality and risk of lung cancer in the multiethnic cohort study. *Nutrients* **13**(5), 1614 (2021).
24. Sun, Y. et al. Lung adenocarcinoma from east Asian never-smokers is a disease largely defined by targetable oncogenic mutant kinases. *J. Clin. Oncol.* **28**(30), 4616–4620 (2010).
25. Li, C. et al. Spectrum of oncogenic driver mutations in lung adenocarcinomas from east Asian never smokers. *PLoS One* **6**(11), e28204 (2011).
26. Yang, J. J. et al. Lung Cancer Risk Prediction models for Asian ever-smokers. *J. Thorac. Oncol.* **19**(3), 451–464 (2024).
27. Chien, L. H. et al. Predicting Lung Cancer occurrence in Never-Smoking females in Asia: TNSF-SQ, a Prediction Model. *Cancer Epidemiol. Biomarkers Prev.* **29**(2), 452–459 (2020).
28. Wu, X. et al. Personalized Risk Assessment in Never, Light, and heavy smokers in a prospective cohort in Taiwan. *Sci. Rep.* **6**, 36482 (2016).
29. Hilden, J., Habbema, J. D. & Bjerregaard, B. The measurement of performance in probabilistic diagnosis. II. Trustworthiness of the exact values of the diagnostic probabilities. *Methods Inf. Med.* **17**(4), 227–237 (1978).
30. Hosmer, D. W. et al. A comparison of goodness-of-fit tests for the logistic regression model. *Stat. Med.* **16**(9), 965–980 (1997).
31. Pencina, M. J. & D'Agostino, R. B. Evaluating discrimination of risk prediction models: the C statistic. *Jama* **314**(10), 1063–1064 (2015).
32. Nahm, F. S. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J. Anesthesiol* **75**(1), 25–36 (2022).
33. Metz, C. E. Basic principles of ROC analysis. *Semin. Nucl. Med.* **8**(4), 283–298 (1978).
34. Obuchowski, N. A. Receiver operating characteristic curves and their use in radiology. *Radiology* **229**(1), 3–8 (2003).
35. Gray, E. P. et al. Risk prediction models for Lung Cancer: a systematic review. *Clin. Lung Cancer*. **17**(2), 95–106 (2016).
36. Toumazis, I. et al. Risk-based lung cancer screening: a systematic review. *Lung Cancer* **147**, 154–186 (2020).
37. Wu, Z. et al. Lung cancer risk prediction models based on pulmonary nodules: a systematic review. *Thorac. Cancer*. **13**(5), 664–677 (2022).
38. McInnes, M. D. F. et al. Preferred reporting items for a systematic review and Meta-analysis of Diagnostic Test Accuracy studies: the PRISMA-DTA Statement. *Jama* **319**(4), 388–396 (2018).
39. Gotschall, T. EndNote 20 desktop version. *J. Med. Libr. Assoc.* **109**(3), 520–522 (2021).
40. Ouzzani, M. et al. Rayyan-a web and mobile app for systematic reviews. *Syst. Rev.* **5**(1), 210 (2016).
41. Wolff, R. F. et al. PROBAST: A Tool to assess the risk of Bias and Applicability of Prediction Model studies. *Ann. Intern. Med.* **170**(1), 51–58 (2019).
42. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**(11), 1539–1558 (2002).
43. Issanov, A. et al. Risk prediction models for lung cancer in people who have never smoked: a protocol of a systematic review. *Diagn. Progn Res.* **8**(1), 3 (2024).
44. IntHout, J., Ioannidis, J. P. & Borm, G. F. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med. Res. Methodol.* **14**, 25 (2014).
45. W, V. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**(3), 1–48 (2010).
46. Shi, L. et al. A mixed effects model for analyzing area under the curve of longitudinally measured biomarkers with missing data. *Pharm. Stat.* **20**(6), 1249–1264 (2021).
47. Sterne, J. A., Gavaghan, D. & Egger, M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J. Clin. Epidemiol.* **53**(11), 1119–1129 (2000).
48. Biljana, M. et al. Bias in meta-analysis and funnel plot asymmetry. *Stud. Health Technol. Inf.* **68**, 323–328 (1999).
49. Duval, S. & Tweedie, R. A nonparametric trim and fill method of accounting for publication bias in meta-analysis. *J. Am. Stat. Assoc.* **95**(449), 89–98 (2000).
50. Balduzzi, S. & Schwarzer, R. G. G, *How to perform a meta-analysis with R: a practical tutorial. Evid. Based Ment. Health* pp. 153–160. (2019).
51. Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. *dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'.* (2019).
52. Tse, L. A. et al. Risk assessment and prediction for lung cancer among Hong Kong Chinese men. *BMC Cancer* **22**(1), 585 (2022).
53. Park, B. et al. Risk-based prediction model for selecting eligible population for lung cancer screening among ever smokers in Korea. *Translational Lung Cancer Res.* **10**(12), 4390–4402 (2021).
54. Li, Y. et al. *Prediction of lung cancer risk in Chinese population with genetic-environment factor using extreme gradient boosting. Cancer Med.* (2022).
55. Ward, R. C. et al. *Impact of Tobacco Dependence in Risk Prediction Models for Lung Cancer diagnoses and deaths. JNCI Cancer Spectr.*, **3**(2) (2019).
56. Raghu, V. K. et al. Validation of a deep learning–based model to Predict Lung Cancer Risk using chest radiographs and Electronic Medical Record Data. *JAMA Netw. Open.* **5**(12), e2248793–e2248793 (2022).
57. Lu, M. T. et al. Deep learning using chest radiographs to identify high-risk smokers for Lung Cancer Screening computed tomography: development and validation of a prediction model. *Ann. Intern. Med.* **173**(9), 704–713 (2020).
58. Yeo, Y. et al. *Individual 5-Year Lung Cancer Risk Prediction Model in Korea using a Nationwide Representative Database. Cancers (Basel)* 13(14), (2021).
59. Schreuder, A. et al. *Combining pulmonary and cardiac computed tomography biomarkers for disease-specific risk modelling in lung cancer screening. Eur. Respir J.* **58**(3). (2021).
60. Maldonado, S. G. et al. Validation of multivariable lung cancer risk prediction models for the personalized assignment of optimal screening frequency: a retrospective analysis of data from the German lung cancer screening intervention trial (LUSI). *Translational Lung Cancer Res.* **10**(3), 1305–1317 (2021).
61. Katki, H. A. et al. Development and Validation of Risk models to select ever-smokers for CT Lung Cancer Screening. *Jama* **315**(21), 2300–2311 (2016).
62. Wang, X. et al. Prediction of the 1-Year risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the state of Maine. *J. Med. Internet Res.* **21**(5), e13260 (2019).
63. Pal Choudhury, P., Chaturvedi, A. K. & Chatterjee, N. Evaluating discrimination of a Lung Cancer Risk Prediction Model using partial risk-score in a two-phase study. *Cancer Epidemiol. Biomarkers Prev.* **29**(6), 1196–1203 (2020).
64. Weber, M. et al. Identifying high risk individuals for targeted lung cancer screening: independent validation of the PLCO(m2012) risk prediction tool. *Int. J. Cancer* **141**(2), 242–253 (2017).
65. Tammemagi, C. M. et al. Lung cancer risk prediction: prostate, lung, colorectal and ovarian Cancer screening trial models and validation. *J. Natl. Cancer Inst.* **103**(13), 1058–1068 (2011).
66. Gu, F. et al. Potential impact of including time to first cigarette in risk models for selecting ever-smokers for Lung Cancer Screening. *J. Thorac. Oncol.* **12**(11), 1646–1653 (2017).
67. Pan, Z. et al. OWL: an optimized and independently validated machine learning prediction model for lung cancer screening based on the UK Biobank, PLCO, and NLST populations. *EBioMedicine* **88**, 104443 (2023).

68. Spitz, M. R. et al. An expanded risk prediction model for lung cancer. *Cancer Prev. Res. (Phila)*. **1**(4), 250–254 (2008).
69. Lin, H. et al. A clinical model to estimate the pretest probability of lung cancer, based on 1198 pedigrees in China. *J. Thorac. Oncol.* **7**(10), 1534–1540 (2012).
70. Sin, D. D. et al. Pro-surfactant protein B as a biomarker for lung cancer prediction. *J. Clin. Oncol.* **31**(36), 4536–4543 (2013).
71. D'Amelio, A. M. Jr. et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. *Br. J. Cancer* **103**(3), 423–429 (2010).
72. Hart, G. R. et al. *A multi-parameterized artificial neural network for lung cancer risk prediction*. *PLoS One* **13**(10). (2018).
73. Markaki, M. et al. A validated clinical risk prediction model for lung Cancer in smokers of all ages and exposure types: a HUNT study. *EBioMedicine* **31**, 36–46 (2018).
74. Li, H. et al. Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model. *BMC Med. Genet.* **13**, 118 (2012).
75. Hippisley-Cox, J. & Coupland, C. *Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study*. *BMJ Open.* (5). (2015).
76. Wilson, D. O. & Weissfeld, J. A simple model for predicting lung cancer occurrence in a lung cancer screening program: the Pittsburgh Predictor. *Lung Cancer* **89**(1), 31–37 (2015).
77. Katki, H. A. et al. Implications of nine risk prediction models for selecting ever-smokers for computed tomography Lung Cancer Screening. *Ann. Intern. Med.* **169**(1), 10–19 (2018).
78. Marcus, M. W. et al. LLPi: Liverpool Lung Project Risk Prediction Model for Lung Cancer incidence. *Cancer Prev. Res. (Phila)* **8**(6), 570–575 (2015).
79. Gould, M. K. et al. Machine learning for early Lung Cancer Identification using Routine Clinical and Laboratory Data. *Am. J. Respir Crit. Care Med.* **204**(4), 445–453 (2021).
80. Muller, D. C., 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.* **35**(8), 861–869 (2017).
81. Tammemägi, M. C. et al. Development and validation of a multivariable Lung Cancer Risk Prediction Model that includes low-dose computed Tomography Screening results: a secondary analysis of Data from the National Lung Screening Trial. *JAMA Netw. Open.* **2**(3), e190204 (2019).
82. Cronin, K. A. et al. Validation of a model of lung cancer risk prediction among smokers. *J. Natl. Cancer Inst.* **98**(9), 637–640 (2006).
83. El-Zein, R. A. 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.* **23**(11), 2462–2470 (2014).
84. Marcus, M. W. et al. Incorporating epistasis interaction of genetic susceptibility single nucleotide polymorphisms in a lung cancer risk prediction model. *Int. J. Oncol.* **49**(1), 361–370 (2016).
85. Tammemagi, C. M. 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.* **11**(12). (2014).
86. Park, S. et al. Individualized risk prediction model for lung cancer in Korean men. *PLoS One* **8**(2), e54823 (2013).
87. Thomas, C. E. et al. Urinary cotinine is as Good a Biomarker as serum cotinine for cigarette smoking exposure and Lung Cancer Risk Prediction. *Cancer Epidemiol. Biomarkers Prev.* **29**(1), 127–132 (2020).
88. Spitz, M. R. et al. Role of selected genetic variants in lung cancer risk in African americans. *J. Thorac. Oncol.* **8**(4), 391–397 (2013).
89. Qian, D. C. et al. *A novel pathway-based Approach improves Lung Cancer Risk Prediction using germline genetic variations*. *Cancer Epidemiol. Biomarkers Prev.* 25. (2016).
90. Hippisley-Cox, J. & Coupland, C. *Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm*. *Br. J. Gen. Pract.* **61**(592): pp. e715–723 .
91. Li, K. 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)* **8**(9), 777–785 (2015).
92. Raji, O. Y. 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.* **157**(4), 242–250 (2012).
93. Spitz, M. R. et al. A risk model for prediction of lung cancer. *J. Natl. Cancer Inst.* **99**(9), 715–726 (2007).
94. Hung, R. J. et al. Assessing Lung Cancer Absolute Risk Trajectory based on a polygenic risk model. *Cancer Res.* **81**(6), 1607–1615 (2021).
95. Kats, D. J. et al. Assessing different approaches to leveraging historical smoking exposure data to Better Select Lung Cancer Screening candidates: a retrospective validation study. *Nicotine Tob. Res.* **23**(8), 1334–1340 (2021).
96. Tammemagi, M. C. et al. Incremental value of pulmonary function and Sputum DNA Image Cytometry in Lung Cancer Risk Prediction. *Cancer Prev. Res.* **4**(4), 552–561 (2011).
97. Lee, J. H. et al. Deep learning to optimize candidate selection for lung cancer CT screening: advancing the 2021 USPSTF recommendations. *Radiology* **305**(1), 209–218 (2022).
98. Wang, F. et al. Risk-stratified Approach for never-and ever-smokers in Lung Cancer Screening: a prospective cohort study in China. *Am. J. Respir. Crit Care Med.* **207**(1), 77–88 (2023).
99. Guida, F. et al. Assessment of Lung Cancer Risk on the basis of a Biomarker Panel of circulating proteins. *JAMA Oncol.* **4**(10), e182078 (2018).
100. Hoggart, C. et al. A risk model for lung cancer incidence. *Cancer Prev. Res. (Phila)*. **5**(6), 834–846 (2012).
101. Etzel, C. J. et al. Development and validation of a lung cancer risk prediction model for african-americans. *Cancer Prev. Res. (Phila)* **1**(4), 255–265 (2008).
102. Yee, J. et al. Connective tissue-activating peptide III: a novel blood biomarker for early lung cancer detection. *J. Clin. Oncol.* **27**(17), 2787–2792 (2009).
103. Tammemägi, M. C. et al. USPSTF2013 versus PLCOm2012 lung cancer screening eligibility criteria (International Lung Screening Trial): interim analysis of a prospective cohort study. *Lancet Oncol.* **23**(1), 138–148 (2022).
104. Janssen-Heijnen, M. L. G. & Coebergh, J. W. W. The changing epidemiology of lung cancer in Europe. *Lung Cancer* **41**(3), 245–258 (2003).
105. Wong, M. C. S. et al. Incidence and mortality of lung cancer: global trends and association with socioeconomic status. *Sci. Rep.* **7**(1), 14300 (2017).
106. Park, B. et al. Risk-based prediction model for selecting eligible population for lung cancer screening among ever smokers in Korea. *Transl Lung Cancer Res.* **10**(12), 4390–4402 (2021).
107. Lam, D. C. L. et al. Lung cancer screening in Asia: an expert consensus report. *J. Thorac. Oncol.* **18**(10), 1303–1322 (2023).
108. Walser, T. et al. Smoking and lung cancer: the role of inflammation. *Proc. Am. Thorac. Soc.* **5**(8), 811–815 (2008).
109. Furrukh, M. Tobacco Smoking and Lung Cancer: perception-changing facts. *Sultan Qaboos Univ. Med. J.* **13**(3), 345–358 (2013).
110. Seow, W. J. et al. Household air pollution and lung cancer in China: a review of studies in Xuanwei. *Chin. J. cancer* **33**(10), 471 (2014).
111. Zhang, Y. et al. Frequency of driver mutations in lung adenocarcinoma from female never-smokers varies with histologic subtypes and age at diagnosis. *Clin. Cancer Res.* **18**(7), 1947–1953 (2012).
112. Sun, Y. et al. Lung adenocarcinoma from east Asian never-smokers is a disease largely defined by targetable oncogenic mutant kinases. *J. Clin. Oncol.* **28**(30), 4616 (2010).

113. Ren, S. et al. Association of EGFR mutation or ALK rearrangement with expression of DNA repair and synthesis genes in never-smoker women with pulmonary adenocarcinoma. *Cancer* **118**(22), 5588–5594 (2012).
114. Setia, M. S. Methodology Series Module 1: Cohort studies. *Indian J. Dermatol.* **61**(1), 21–25 (2016).
115. Setia, M. S. Methodology Series Module 2: case-control studies. *Indian J. Dermatol.* **61**(2), 146–151 (2016).
116. Ernster, V. L. Nested case-control studies. *Prev. Med.* **23**(5), 587–590 (1994).
117. Barlow, W. E. et al. Analysis of case-cohort designs. *J. Clin. Epidemiol.* **52**(12), 1165–1172 (1999).
118. Steyerberg, E. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating* (Springer Science & Business Media, 2008).
119. Yuan, W. et al. Temporal bias in case-control design: preventing reliable predictions of the future. *Nat. Commun.* **12**(1), 1107 (2021).
120. Janes, H. & Pepe, M. S. Matching in studies of classification accuracy: implications for analysis, efficiency, and assessment of incremental value. *Biometrics* **64**(1), 1–9 (2008).
121. Saha, P. & Heagerty, P. Time-dependent predictive accuracy in the presence of competing risks. *Biometrics* **66**(4), 999–1011 (2010).
122. Moons, K. G. et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* **98**(9), 691–698 (2012).
123. Harrell, J., Frank, E. & Harrell, F. E. *Missing data.* Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis, pp. 45–61. (2015).
124. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **36**(2), 111–133 (1974).
125. Burman, P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika* **76**(3), 503–514 (1989).
126. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap.* (CRC, 1994).
127. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and Their Application.* (Cambridge University Press, 1997).
128. Ramspek, C. L. et al. External validation of prognostic models: what, why, how, when and where? *Clin. Kidney J.* **14**(1), 49–58 (2021).
129. Siontis, G. C. M. et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J. Clin. Epidemiol.* **68**(1), 25–34 (2015).
130. Shipe, M. E. et al. Developing prediction models for clinical use using logistic regression: an overview. *J. Thorac. Dis.* **11**(Suppl 4), S574–s584 (2019).
131. Habehh, H. & Gohel, S. Machine learning in Healthcare. *Curr. Genomics* **22**(4), 291–300 (2021).
132. Woldaregay, A. Z. et al. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. *J. Med. Internet. Res.* **21**(5), e11030 (2019).
133. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**(7639), 115–118 (2017).
134. Kanan, M. et al. *AI-Driven models for Diagnosing and Predicting outcomes in Lung Cancer: a systematic review and Meta-analysis*. *Cancers (Basel)* **16**(3) (2024).
135. Ahmed, M. I. et al. A systematic review of the barriers to the implementation of Artificial Intelligence in Healthcare. *Cureus* **15**(10), e46454 (2023).
136. Warnat-Herresthal, S. et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature* **594**(7862), 265–270 (2021).
137. Atwater, T. & Massion, P. P. *Biomarkers of risk to develop lung cancer in the new screening era*. *Annals Translational Med.* **4**(8). (2016).
138. Fry, J. S., Hamling, J. S. & Lee, P. N. Systematic review with meta-analysis of the epidemiological evidence relating FEV 1 decline to lung cancer risk. *BMC cancer* **12**, 1–15 (2012).
139. Gan, T. et al. Correlation between carcinoembryonic antigen (CEA) expression and EGFR mutations in non-small-cell lung cancer: a meta-analysis. *Clin. Transl Oncol.* **26**(4), 991–1000 (2024).
140. Grunnet, M. & Sorensen, J. Carcinoembryonic antigen (CEA) as tumor marker in lung cancer. *Lung cancer* **76**(2), 138–143 (2012).
141. Yang, Y. et al. Serum carcinoembryonic antigen elevation in benign lung diseases. *Sci. Rep.* **11**(1), 19044 (2021).
142. Nassar, Y. M. et al. *C-Reactive protein as an early predictor of efficacy in Advanced Non-small-cell Lung Cancer patients: a Tumor dynamics-Biomarker modeling Framework*. *Cancers (Basel)* **15**(22). (2023).
143. Chaturvedi, A. K. et al. C-reactive protein and risk of lung cancer. *J. Clin. Oncol.* **28**(16), 2719–2726 (2010).
144. Yin, J. et al. Association between baseline C-reactive protein and the risk of lung cancer: a prospective population-based cohort study. *Cancer Prev. Res.* **15**(11), 747–754 (2022).
145. Seow, W. J. et al. Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from western populations. *Hum. Mol. Genet.* **26**(2), 454–465 (2017).
146. Wang, H. M., Zhang, X. Y. & Jin, B. TERT genetic polymorphism rs2736100 was associated with lung cancer: a meta-analysis based on 14,492 subjects. *Genetic Test. Mol. Biomarkers* **17**(12), 937–941 (2013).
147. Wu, X. et al. Ethnicity-specific association between TERT rs2736100 (A > C) polymorphism and lung cancer risk: a comprehensive meta-analysis. *Sci. Rep.* **13**(1), 13271 (2023).
148. Hsiung, C. A. et al. The 5p15. 33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS Genet.* **6**(8), e1001051 (2010).
149. Zhou, W. & Christiani, D. C. East meets West: ethnic differences in epidemiology and clinical behaviors of lung cancer between East asians and caucasians. *Chin. J. Cancer* **30**(5), 287–292 (2011).
150. Seijo, L. M. et al. Biomarkers in Lung Cancer Screening: achievements, promises, and challenges. *J. Thorac. Oncol.* **14**(3), 343–357 (2019).
151. Kneuertz, P. J. et al. Improving patient engagement, adherence, and satisfaction in lung cancer surgery with implementation of a mobile device platform for patient reported outcomes. *J. Thorac. Disease* **12**(11), 6883 (2020).
152. Jia, Q. et al. *Barriers to low-dose CT Lung Cancer Screening among Middle-aged Chinese*. *Int. J. Environ. Res. Public. Health.* 17(19), (2020).
153. Poon, C. et al. Why is the screening rate in lung cancer still low? A seven-country analysis of the factors affecting adoption. *Front. Public. Health* **11**, 1264342 (2023).
154. Shankar, A. et al. Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward. *Translational lung cancer Res.* **8**(Suppl 1), S106 (2019).
155. Hall, D. L. et al. Lung cancer screening uncertainty among patients undergoing LDCT. *Am. J. Health Behav.* **42**(1), 69–76 (2018).
156. McFadden, K. et al. Lung cancer screening program factors that influence psychosocial outcomes: a systematic review. *Psycho-Oncology* **33**(1), e6252 (2024).
157. Ten Haaf, K. et al. Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study. *PLoS Med.* **14**(4), e1002277 (2017).
158. Steyerberg, E. W. et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* **21**(1), 128–138 (2010).

159. Pencina, M. J. et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat. Med.* **27**(2), 157–172 (2008).
160. Fitzgerald, M., Saville, B. R. & Lewis, R. J. Decision curve analysis. *Jama* **313**(4), 409–410 (2015).
161. Li, C. et al. *Circulating tumor DNA as liquid biopsy in lung cancer: biological characteristics and clinical integration*. *Cancer Lett.* 216365. (2023).
162. Lam, W. J. et al. *Circulating tumour DNA analysis for early detection of lung cancer: a systematic review*. *Annals Translational Med.* **12**(4). (2024).
163. Administration. U.S.F.a.D. *Medical Devices: Cobas EGFR Mutation Test v2—P150047.*. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P150047 (2024).
164. Jenkins, D. A. et al. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? *Diagn. Progn Res.* **5**(1), 1 (2021).
165. Mishra, A. et al. Recalibration methods for Improved Clinical Utility of Risk scores. *Med. Decis. Mak.* **42**(4), 500–512 (2022).

Author contributions

Y.R.J contributed to conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, and writing of the original draft, review, and editing. L.A contributed to data curation, formal analysis, methodology, and writing (review and editing). W.J.S. contributed to conceptualization, formal analysis, methodology, supervision, and writing (review, and editing). All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-83875-6.

Correspondence and requests for materials should be addressed to W.J.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.