Original Research

Oncology and Translational Medicine

OPEN

Prognostic models for lung cancer in smokers and nonsmokers: an updated systematic review and meta-analysis

Xinyue Pan[a], Boxing Feng[b], Ying Chen[c], Junfeng Wang[d], Xuanqi Pan[e], Taihing Lam[f], Jing Pan[g,*]

Abstract

Background: Lung cancer is the leading cause of cancer-related mortality, and while low-dose computed tomography screening may reduce mortality, emerging prognostic models show superior discriminative efficacy compared to age- and smoking history-based screening. However, further research is needed to assess their reliability in predicting lung cancer risk in high-risk patients.

Methods: This study evaluated the predictive performance and quality of existing lung cancer prognostic models through a systematic review and meta-analysis. A comprehensive search was conducted in PubMed, Cochrane, Web of Science, CNKI, and Wanfang for articles published between January 1, 2000, and February 13, 2025, identifying population-based models incorporating all available modeling data.

Results: Among 72 analyzed studies, models were developed from Asian (28 studies, including 23 Chinese cohorts) and European/American (48 studies) populations, with only 6 focusing on nonsmokers. Twenty-one models included genetic markers, 15 used clinical factors, and 40 integrated epidemiological predictors. Although 37 models underwent external validation, only 4 demonstrated minimal bias and clinical applicability. A meta-analysis of 11 repeatedly validated models revealed calibration and discrimination, though some lacked calibration data.

Conclusions: Few lung cancer prognostic models exist for nonsmokers. Most models exhibit poor predictive performance in external validations, with significant bias and limited application scope. Widespread external validation, standardized model development, and reporting techniques are needed to accurately identify high-risk individuals and ensure applicability across diverse populations.

Keywords: Lung cancer; Prognostic model; Screen; Risk factor

1. Introduction

In 2022, approximately 20 million new cancer cases were diagnosed worldwide, with lung cancer accounting for 12.4% of all cases. Approximately 9.7 million cancer deaths have been reported worldwide, with lung cancer accounting for 18.7%. Notably, lung cancer has emerged as the most prevalent form of cancer among male populations, both in terms of incidence and mortality rates.[1] The mortality rate due to lung cancer is 54.57 per 100,000 in the Chinese population, 75.05 per 100,000 in males and 33.19 per 100,000 in females.[2] The 5-year survival rate for stage I patients after surgery is 77% to 92%, whereas it is only 10% to 36% for stages III and IV patients, indicating that early diagnosis can significantly improve the prognosis and survival of lung cancer patients. Low-dose computed tomography (LDCT) screening significantly reduces lung cancer mortality rates by 20% to 26% in high-risk individuals according to large lung cancer screening trials such as the United States–based National Lung Cancer Trial and Dutch-Belgian Lung Cancer Screening Study (Dutch acronym: NELSON).[3,4] Current LDCT screening trials predominantly focus on age and smoking history as primary criteria for high-risk population selection. However, evidence suggests that these parameters demonstrate lower discriminatory power than comprehensive lung cancer risk prediction models. This limitation persists because established risk factors extend beyond these variables, including male sex, White race, emphysema, occupational asbestos exposure, chronic obstructive pulmonary disease diagnosis, and family history of lung cancer.[5] Risk prognostic models, which can estimate the

X. Pan, B. Feng, and Y. Chen contributed equally to this study.

Supplemental digital content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's Web site (https://journals.lww.com/otm).

[a]School of Public Health, Guangzhou Medical University, Guangzhou, China, [b]School of Nursing, Guangzhou Medical University, Guangzhou, China, [c]Nursing Department, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China, [d]Faculty of Science, Utrecht University, Utrecht, the Netherlands, [e]Faculty of Medical Science, University of Groningen, Groningen, the Netherlands, [f]School of Public Health, University of Hong Kong, China, [g]Science and Education Section, Affiliated Guangzhou Twelfth People's Hospital, Guangzhou Medical University, Guangzhou, China.

** Corresponding author. Address: Science and Education Section, Affiliated Guangzhou Twelfth People's Hospital, Guangzhou Medical University, No. 1, Tianqiang Road, West Huangpu Avenue, Guangzhou 510000, China. E-mail address: jcdaise@163.com (J. Pan).*

Oncology and Translational Medicine (2025) 11:3;112–117

Received: 13 December 2024; Revised: 16 March 2025; Accepted: 18 March 2025.

http://dx.doi.org/10.1097/ot9.0000000000000091

personalized probability of developing lung cancer, have the potential to assist physicians in efficient and cost-effective screening of high-risk individuals with lung cancer. However, the substantial diversity in predictor types, development and validation of population characteristics, modeling techniques, and other inherent model traits may greatly influence their discriminative capacity and practical utility. In this study, we appraised existing lung cancer prognostic models for their study design, risk of bias, and predictive performance through a systematic literature review and meta-analysis, aiming to provide insights into the strengths and weaknesses of current prognostic models, provide a reliable basis for future clinical lung cancer screening, alleviate the burden on clinicians, and facilitate lung cancer–related care.

2. Methods

2.1. Literature review

Five electronic databases (PubMed, Cochrane, Web of Science, CNKI [China National Knowledge Infrastructure], and Wan Fang data) were searched for relevant reviews and articles published between January 1, 2000, and February 13, 2025, using the following terms in the title and abstract: "lung cancer," "pulmonary neoplasms," "lung neoplasms," and "pulmonary cancer," combined with the terms "screen," "model," "prediction," and "prognostic" according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines. The specific retrieval formula for each database is given in Supplementary Appendix, Table 1, http://links.lww.com/OTM/A16, literature retrieval type. We prioritized a review search and read and collated all the prognostic models for lung cancer risk included in the reviews and found that the latest lung cancer prognostic model studies quoted in the reviews were published on February 1, 2020.[5] The second step of the original literature search was as follows. Lung cancer prognostic model articles published between February 1, 2020, and February 13, 2025, were searched using the same terms in 5 databases. All articles on lung cancer prognostic models were quoted in these relevant reviews, and additional relevant articles were included in our systematic review.

Two reviewers (XP and YC) independently conducted literature review, data collection, and data extraction. In cases of disagreement, a consensus was reached through the input of a third reviewer (BF). Subsequently, one reviewer (XP) synthesized the principal findings of the study, and another reviewer (YC) conducted critical scrutiny and revisions. The inclusion criteria for the systematic review and risk of bias assessment were studies that developed and validated a lung cancer prognostic model. Duplicate studies, studies focusing on other human cancers or nonhuman lung cancers, and models not validated in prospective cohorts were excluded. Furthermore, externally validated models were included in the meta-analysis.

Paper screening followed the following exclusion criteria: (1) duplicate papers, (2) incorrect types of papers (such as case reports), (3) nonhuman subjects (such as animal or cell studies), (4) incomplete papers (with serious missing or unclear data), (5) no lung cancer risk prediction model reported, and (6) the model does not have differentiation or calibration information for internal or external verification.

2.2. Data extraction and analysis

Data were independently extracted by 2 authors (XP and YC). Systematic data were collected for each model. The extracted key information included model name, publication date, country, target population, model construction and validation methods, and model performance metrics. The extracted data were verified by a third researcher (BF). Any discrepancies were discussed individually to reach a consensus. The articles were screened using the Prediction Model Risk of Bias Assessment Tool (PROBAST) to evaluate their risk of bias and applicability.[6] In addition, a meta-analysis was conducted to evaluate the discrimination and calibration of the models in diverse external validation cohorts.

2.3. Model quality assessment

The PROBAST was used to assess the risk of bias and applicability of the included models. The tool has 4 domains: participants, predictors, outcomes, and analysis—with 20 signaling questions. Every question is classified as "low," "high," or "unclear," with the first 3 domains alone evaluating how applicable it is. Two reviewers (XP and YC) independently evaluated each study for risk of bias and applicability, and we evaluated only the model development outcomes based on the study content. Owing to the lack of partial information in the model, the evaluators made a subjective judgment; when the results disagreed, a third evaluator participated in the evaluation of the final results.

In a given field, the assessment results of specific literature are typically "high," particularly when missing data processing, data complexity, and other difficulties are present. This leads to an overall evaluation result of high risk. We made the following request in response to this situation: we contacted the authors and tried to obtain information first. Even if certain models developed earlier lack sufficient information and still present 1 or 2 high-risk issues, we still consider incorporating them into the study when additional critical information becomes available.

2.4. Meta-analysis investigating model performance

We identified external validation studies for some of the models by reviewing the cited articles of the included studies and conducted a meta-analysis to evaluate the results of applying each developed model to several external validation datasets. Here, n represents the total sample size of the external validation study; "all.events" refers to the actual number of lung cancer occurrences during the entire follow-up period of the cohort; "n.events" refers to the actual number of lung cancer occurrences during the follow-up period; "e.events" refers to the number of lung cancer occurrences predicted by the model during the follow-up period; area under the curve (AUC) is the point estimate of the model's discrimination in external validation; and AUC.95CIl and AUC.95CIu represent the lower and upper bounds of the 95% confidence interval (95% CI) for the model's discrimination in external validation, respectively. When data were missing, the original study authors were contacted via email to obtain complete information. After 1 to 2 rounds of unsuccessful contact, we selected studies with available data for the analysis. The data were analyzed, and a forest plot demonstrating discrimination and calibration abilities was generated. The statistical software used for analysis was R (version 4.3.3) with the "metamisc" and "metafor" packages. Heterogeneity was assessed between studies using the I^2 statistic. Heterogeneity was considered low when $0\% < I^2 \leq 50\%$, moderate when $50\% < I^2 < 75\%$, and high when $I^2 \geq 75\%$. In this study, heterogeneity was primarily addressed through the selection of effect models, subgroup analysis, and sensitivity analysis. Subgroup analyses stratified by geographical region and smoking status were performed when an adequate number of studies were available. Sensitivity analysis was implemented through iterative removal of individual studies to investigate potential sources of heterogeneity and evaluate the robustness of the pooled results. This methodological approach ensured comprehensive exploration of heterogeneity while maintaining analytical rigor. Before data extraction and analysis, the study was registered on PROSPERO (ID: CRD420251010749).

3. Results

3.1. Search and study characteristics

We identified 483 reviews and 6241 original articles after removing duplicates for title and abstract screening; 7 reviews and 81 articles underwent full-text screening for eligibility. We identified 4 reviews and 26 original articles that fulfilled the predefined inclusion criteria. After 2 authors compiled and read the full texts, 46 model articles were found in the 4 reviews. When combined with 26 original articles, this resulted in a total of 72 articles containing 76 lung cancer prognostic models included in this study (Supplementary Appendix, Figure 1, http://links.lww.com/OTM/A16).

Each model exhibited notable distinctions from the other models in terms of study design, objectives, sample size, and various other parameters. Among the 76 lung cancer risk prognostic models identified, 28 were based on data sourced from Asian populations, 23 were predicted from datasets derived from Chinese demographic data, and the remaining 48 were constructed based on European and American populations. Among them, 7 models, the EAGLE (Environment and Genetics in Lung cancer Etiology),[7] basic predictive model,[8] Cancer Screening Program in Urban China nonsmoker (CanSPUC-nonsmoker),[9] CanSPUC-nonsmoker women,[10] metabolic indicator model,[8] Taiwanese NSF Lung Cancer Risk Models using genetic information and simplified questionnaire (TNSF-SQ),[11] and Guo,[12] were constructed for nonsmoking populations only. Forty models exclusively incorporated epidemiological predictors, 15 supplemented them with clinical test variables, and 21 further developed their predictive capacity by incorporating genetic markers. In addition, 37 models (49%) were externally validated. The study characteristics are described in the Supplementary Appendix, Tables 2, 3, and 4, http://links.lww.com/OTM/A16.

3.2. Risk of bias and applicability assessment of the included studies

3.2.1. Models encompassing only epidemiological predictors

Among these 40 models, Bach et al.,[13] Lung Cancer Risk Assessment Tool (LCRAT),[14] QResearch Lung (QLung),[15] and Optimized early Warning model for Lung cancer risk (OWL)[16] had clear statements for 20 questions in 4 domains, and the overall judgment for bias risk was rated as low. Thirteen of these models were rated as high risk in the first question of domain 1 because they were modeled using retrospective data. Four models that did not report the follow-up years of the study were rated as unclear in question 6 of domain 3 and had missing information reported in domain 4; therefore, they were ultimately rated as high risk. In domain 4, the remaining 19 models had 1 to 2 problems, such as improper treatment of continuous variables, selection of predictors based on single-factor analysis, and incomplete model performance evaluation, and we rated them as high risk (Supplementary Appendix, Table 5, http://links.lww.com/OTM/A16).

3.2.2. Models encompassing epidemiological predictors and clinical test variables

Among these 15 models, the development groups of The Health Improvement Network (THIN),[17] Liao,[18] Kaeum,[19] Li (LDCT model),[20] and machine learning 2 (ML2)[21] models were derived from retrospective data, and problems such as ignoring competition risks and unknown complex data processing occurred in the modeling process; therefore, they were rated as high risk in domains 1 and 4. The Beane[22] model was rated as unclear in question 6 of domain 3 because of unreported follow-up years and as high risk in question 1 of domain 4 because of its small sample size (only 76 participants). In addition, it has problems similar to those of other models, such as not explicitly mentioning the processing of missing data and ignoring competition risk. The final evaluation results of all 15 models revealed high risk (Supplementary Appendix, Table 6, http://links.lww.com/OTM/A16).

3.2.3. The model encompasses epidemiological predictors, clinical test variables, and genetic markers

Among the 21 models, the Wang[23] and deep Q network[24] models were rated as high risk in domain 1 because of the use of retrospective data. The DNA Damage Binding Protein 2 (DDB2) gene model in the Chinese population,[25] Beane (Biomarker + Clinical),[22] Chinese multigenetic,[26] and Genome-Wide Association Study (GWAS) of the Chinese population models[27] were rated as unclear in domain 3 because of unreported follow-up years. Moreover, except for the deep Q network model, all other models have 1-2 problems in domain 4, the most common being that the method of handling missing data was not explicitly mentioned and that the complicated data were not explained. Therefore, the final evaluation results of these 21 models indicated a higher risk of bias (Supplementary Appendix, Table 7, http://links.lww.com/OTM/A16).

3.3. Meta-analysis of the lung cancer prognostic models

To compare the discrimination and calibration of various prognostic models, we collected external validation studies for all existing models and extracted data related to discrimination and calibration (Supplementary Appendix, Table 8, http://links.lww.com/OTM/A16). The AUCs of the models in external validation were generally suboptimal (Supplementary Appendix, Figure 2, http://links.lww.com/OTM/A16). For the 1-year risk models, the AUC ranged from 0.68 (95% CI, 0.63–0.73; Spitz et al.[28]) to 0.73 (95% CI, 0.45–0.90; Hoggart et al.[29]); for the 5-year risk models, the AUCs were 0.71 (95% CI, 0.69–0.73; Liverpool Lung Project version 3 [LLPv3]), 0.72 (95% CI, 0.69–0.75; Liverpool Lung Project [LLP]), and 0.75 (95% CI, 0.73–0.76; LCRAT); for the 6-year risk models, the AUC ranged from 0.71 (95% CI, 0.69–0.73; Pittsburgh) to 0.74 (95% CI, 0.72–0.75, PLCOm2012, Nord-Trøndelag Health Study [HUNT]); and for the 8/8.7/10-year risk models, the AUCs were 0.75 (95% CI, 0.73–0.76; OWL), 0.75 (95% CI, 0.60–0.85, Liverpool Lung Project Risk Prediction Model for Lung Cancer Incidence [LLPi]), and 0.73 (95% CI, 0.72–0.75, Bach). Among these, the LCRAT, OWL, and LLPi models showed better discrimination than the others. In most development or external validation studies, calibration-related information was scarce (e.g., the actual incidence of lung cancer vs. model-predicted incidence); therefore, calibration data were available for only 6 models. As shown in Supplementary Appendix, Figure 3, http://links.lww.com/OTM/A16, the calibration of the Bach, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial 2012 model (PLCOm2012), LLPv3, OWL, and HUNT models was 0.92 (95% CI, 0.82–1.02), 0.93 (95% CI, 0.84–1.03), 0.98 (95% CI, 0.82–1.17), 1.05 (95% CI, 0.95–1.15), and 1.16 (95% CI, 1.04–1.29), respectively, indicating that these models had relatively accurate predictive capabilities. However, the LCRAT model seemed to overestimate the 5-year lung cancer risk for individuals, with a total O:E ratio of 0.65 (95% CI, 0.58–0.72).

3.4. Heterogeneity analysis results

We performed an overall heterogeneity analysis (Supplementary Appendix, Table 9, http://links.lww.com/OTM/A16) of 11 models in the meta-analysis; the Hoggart et al.[29] model was not analyzed because no subgroup external validation results were reported. The analysis of the remaining 10 models (PLCOm2012,[30] Bach et al.,[13] HUNT,[31] LCRAT,[14] LLP,[32] LLPi,[33] LLPv 3,[34]

OWL,[16] Pittsburgh,[35] and Spitz et al.[28]) showed high interstudy heterogeneity, suggesting that the difference in AUC between studies mainly resulted from real effect differences rather than sampling errors. To explore the source of heterogeneity, we performed subgroup heterogeneity analysis in different regions and smoking status subgroups based on the available outcomes from the literature (Supplementary Appendix, Table 10, http://links.lww.com/OTM/A16). The results indicated that both variables were sources of heterogeneity. The results showed that both variables were sources of heterogeneity. There was no significant difference in the results of the metaregression, which we believe was caused by data imbalance. In addition, the heterogeneity results did not change significantly after sensitivity analysis, indicating that there may be other sources of heterogeneity that were not explained but were limited by the original study data (Supplementary Appendix, Tables 11 and 12, http://links.lww.com/OTM/A16).

4. Discussion

4.1. Main findings

This study is the first to systematically review and combine original literature to identify 76 lung cancer risk prognostic models. We applied the PROBAST tool for a comprehensive evaluation of the biases and characteristics of these models, including the population used for model development, modeling methods, types of predictive factors, and model performance. Our findings showed that the earliest models were developed in 2003[13] and originated in regions such as Asia and Europe. Among these, only 23 models were developed using data from Chinese populations, and only 7 were specifically designed for nonsmoking populations. The majority of the models are used to predict lung cancer risk over 1, 3, or 5 years.

Among all the models, age and smoking were the most fundamental predictive factors, followed by sex, race, education level, and family history of lung cancer. In some countries, the incidence of lung cancer among nonsmoking populations was relatively high, and considering only the risks associated with smoking may lead to an increased rate of missed diagnoses and misdiagnoses during screening. Additionally, 40 models were developed based solely on epidemiological factors, whereas 15 models incorporated clinical test variables as predictive factors. With the development of genetic sequencing technologies, 21 models simultaneously considered the influence of genetic factors. Notably, the predictive performance of these models did not vary significantly, with the AUC ranges for both internal and external validation fluctuating between 0.6 and 0.8. Compared with models that include only epidemiological factors, adding variables such as diagnostic tests or genetic factors does not always improve model performance. For example, in the established PLCOm2012 and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) Expand models,[30,36] lung function was added as a predictive factor in the latter; however, the AUC values decreased from 0.857 to 0.77. Additionally, the LLP[32] and Spitz et al.[28] models showed that the addition of genetic factors did not significantly enhance model performance. Therefore, although these models incorporate various predictive factors, they may not fully account for the interactions between these factors, or some factors may be redundant, leading to model overfitting. Future research should involve a deeper analysis of the existing predictive factors. For example, factor interaction effect analysis can be used to explore potential nonlinear relationships and interactions to improve model performance. Some factors may have different importance and predictive power in different populations; therefore, personalized weighting based on these factors and the selection of different predictive factors for populations with distinct characteristics could further enhance model accuracy. Meanwhile, to avoid overly complex and redundant predictive factors in the model, techniques such as least absolute shrinkage and selection operator (LASSO) regression and principal component analysis[37] can be considered to help select more predictive factors and reduce interference from irrelevant factors, thereby improving the stability of the model. Further improvements may be achieved by adding predictors such as clinical variables or biomarkers. In addition, the most commonly used predictors in existing models are traditional epidemiological factors. Future studies should combine the above methods to strengthen the development of this field for the construction of lung cancer risk prognostic models.

In this study, the construction method of the included model mainly relied on traditional statistical approaches, such as logistic regression, with a few models using machine learning algorithms, such as eXtreme Gradient Boosting (XGBoost)[16,38] and random forest (RF).[39] Recent studies (e.g., the Cox proportional hazards [CoxPH] model[40]) have adopted a stacked ensemble approach. Optimizing ensemble strategies is the key to improving model performance, particularly through the use of weighted stacking, which assigns different weights to base learners based on their predictive power, thereby improving the overall prediction accuracy of the model. Moreover, considering regional differences in data and sample size limitations, cross-domain transfer learning[41] can help adjust the model to address sample size deficiencies and enhance its adaptability to new environments. As follow-up data from patients continue to accumulate, dynamic predictions have become increasingly important. Online and incremental learning[42] can help models continuously adapt to new environments and changing data, thereby improving their accuracy and applicability and providing sustained and reliable predictive support for clinical practice, particularly in the context of long-term lung cancer screening, where the models will have greater practical value.

This is the first study to evaluate all lung cancer risk prediction models using the PROBAST (Supplementary Appendix, Table 13, http://links.lww.com/OTM/A16). The overall results indicated that most previous models carry a high risk of bias; only 4 studies were rated as low risk, with the main issue being the lack of a detailed explanation regarding data processing in the reported models. When applying models in clinical practice, it is essential to ensure the integrity and transparency of data. Therefore, in future research, stricter standardized processes should be adopted when collecting data to ensure the quality and consistency of all input data. Additionally, the data processing methods and results should be clearly outlined in the study to reduce model bias. Although some bias issues were considered, to better control bias in the future, model development should enhance the use of randomization or stratified sampling to ensure fair inclusion of various patient groups, thus making the model more widely applicable. Furthermore, sensitivity analysis is recommended to further assess the impact of different biases on model performance.

Model predictive performance is evaluated through discrimination and calibration. Only models that demonstrate good discrimination and calibration through extensive external validation across different populations should be considered to improve clinical use. This study revealed that most studies evaluated discrimination based on the area under the receiver operating characteristic curve, represented by the AUC value. The discriminatory ability of all models varied, with some models reporting predictive performances ranging from poor to good in internal and external validations (AUC values between 0.63 and 0.75). However, good predictive performance does not necessarily imply that the model has good applicability in actual clinical settings. Among the 76 models, only 11 reported the AUC in different external validation cohorts, and the models showed varying results. The heterogeneity test analysis showed that there was high heterogeneity among the studies, and the subgroup analysis showed significant differences in the AUC

values between different regions and smoking status, which may be the source of heterogeneity. However, the metaregression results were different from those of the subgroup analysis; therefore, the interstudy heterogeneity of each model was still too large after the leave-one method, suggesting that other sources of heterogeneity may exist. Regarding calibration, most models were assessed using calibration plots or the Hosmer-Lemeshow goodness-of-fit test, but these methods may not be suitable for all types of models. In future studies, methods such as Platt scaling or isotonic regression could be considered to calibrate the model's probability outputs, as these methods generally improve calibration performance in external validation datasets. We summarized the results of 7 models that reported the predicted and actual incidence numbers in external validation. The calibration assessments for the LCRAT, Bach, PLCOm2012, LLPv3, OWL, and HUNT models ranged from 0.65 to 1.16, demonstrating overall good performance. Calibration assessments for the remaining models were not reported (not reported in the external validation), which was a common issue in previous model development and a key reason for the high risk of bias in the evaluation process, resulting in subsequent studies failing to fully evaluate the model and increasing its difficulty in practical applications. In addition, existing external validations that focus primarily on specific regions or ethnic populations may not be applicable to other regions or populations. To explore and enhance the generalizability of these models, future validation studies should be conducted in more regions and diverse ethnic groups. Furthermore, stratified validation based on factors such as smoking status, age group, and sex should be performed to assess model performance across different subgroups. Sensitivity analysis should also be conducted to evaluate the stability of the model under different parameters or assumptions. Therefore, the best prognostic models should follow a rigorously standardized internal and external validation process with complete information on discrimination and calibration to reduce the risk of bias.

In clinical practice, lung cancer prognostic modeling can help doctors and nurses more easily and accurately screen high-risk individuals for lung cancer, effectively make clinical diagnoses and decisions, and assist in prevention and treatment. Especially in women, misdiagnosis and underdiagnosis can be avoided when nonsmoking factors such as secondhand smoke, environmental exposure, and fumes are considered. However, existing prognostic models for lung cancer risk vary in terms of modeling population, modeling approach, and selection of predictors, resulting in differences in model effects, and no model has been adopted in clinical practice. The predictive performance of some models is relatively good, but the required predictor variables, such as detection test indicators or DNA indicators, may not be applicable to the general population, thus making them difficult to obtain in practical applications, which reduces the utilization of that type of model. Therefore, future research should consider economic issues, such as the cost of application, while considering the predictive performance of the model to maximize the use of the developed model.

4.2. Strengths and limitations

Our study pooled multiple lung cancer risk prognostic models and systematically evaluated them using a biased risk assessment tool and meta-analysis. Specifically, regarding the predictive factors, model structures, and validation methods for lung cancer prognostic models, our study not only provided traditional epidemiological factors but also considered clinical tests and genetic data, offering a broader perspective. Moreover, compared with other studies,[5,43] we not only assessed the predictive ability of the models but also conducted an in-depth analysis of the model quality and bias and provided a reference for the subsequent improvement of the models included in the study. Optimization suggestions for existing models are provided, particularly in the areas of predictive factor selection, model construction methods, and validation strategies. However, one limitation was that, although this study collected 11 datasets of models from different externally validated studies when conducting the meta-analysis, sample heterogeneity in different regions (e.g., sex, age, and genomic differences) may affect the generalizability and applicability of these models. Future research could further strengthen stratified analyses across different regions, ethnicities, or subgroups to ensure the effectiveness of the models across all populations. Another limitation is that, despite our evaluation of existing models, as new data emerge and clinical practices continue to advance, current models may require periodic updates. Moreover, the number of nonsmoking population models included in the study was too small to be further evaluated and discussed, which needs to be improved and perfected in the study of lung cancer risk prognostic models. In the future, more lung cancer risk prognostic models in different regions should be established for nonsmoking populations, especially for female nonsmoking populations.

Our study systematically evaluated 76 lung cancer risk prediction models, revealing critical limitations in their methods and providing suggestions for improvement. The models showed a high risk of bias. Only 14.5% of the models underwent external validation with 2 to 15 cohorts, and calibration assessments were neglected. Moreover, the validation cohorts predominantly focused on single or homogeneous ethnic populations, substantially limiting their clinical generalizability. Future research should prioritize refined predictor selection using advanced techniques (e.g., interaction effect analysis) to mitigate redundant variable interference. Standardized modeling and validation protocols should be established by integrating dynamic prediction models to enhance the effectiveness of long-term screening programs. Adaptive analyses across diverse geographical regions and heterogeneous populations should be conducted to improve model robustness and clinical applicability.

Acknowledgments

The authors extend our gratitude to the *Oncology and Translational Medicine* editorial team for their meticulous linguistic refinement and proofreading of our manuscript.

Financial support and sponsorship

This study was funded by the Guangzhou Municipal Science and Technology Bureau (No. 2023A03J0507). The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of interest statement

The authors declare that they have no conflict of interest with regard to the content of this report.

Author contributions

Writing—original draft preparation: XP; methodology: JW; investigation: BF and YC; writing—review and editing: JP, TL, and XP. All authors have read and agreed to the published version of this manuscript.

Data availability statement

All data generated or analyzed during this study are included in the published article (and Appendix Information file).

Ethical approval

Before data extraction and analysis, the study was registered on PROSPERO (ID: CRD420251010749).

References

[1] Bray F, Laversanne M, Sung H, et al. Global Cancer Statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;74(3):229–263.

[2] Qi J, Li M, Wang L, et al. National and subnational trends in cancer burden in China, 2005–20: an analysis of national mortality surveillance data. *Lancet Public Health* 2023;8(12):e943–e955.

[3] National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365(5):395–409.

[4] De Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med* 2020;382(6):503–513.

[5] Toumazis I, Bastani M, Han SS, et al. Risk-based lung cancer screening: a systematic review. *Lung Cancer* 2020;147:154–186.

[6] Chen XP, Zhang Y, Zhauang YY, et al. PROBAST: a tool for assessing risk of bias in the study of diagnostic or prognostic multi-factorial predictive models. *Chin J Evid Based Med* 2020;20(06):737–744.

[7] Kovalchik SA, de Matteis S, Landi MT, et al. A regression model for risk difference estimation in population-based case-control studies clarifies gender differences in lung cancer risk of smokers and never smokers. *BMC Med Res Methodol* 2013;13:143.

[8] Lyu ZY, Li N, Chen SH, et al. Exploratory research on developing lung cancer risk prediction model in female non-smokers. *Chin J Prev Med* 2020;54(11):1261–1267. Chinese.

[9] Guo LW, Lyu ZY, Meng QC, et al. Construction and validation of a lung cancer risk prediction model for non-smokers in China. *Front Oncol* 2022;11:766939.

[10] Guo L, Meng Q, Zheng L, et al. Lung cancer risk prediction nomogram in nonsmoking Chinese women: retrospective cross-sectional cohort study. *JMIR Public Health Surveill* 2023;9:e41640.

[11] Chien LH, Chen CH, Chen TY, et al. Predicting lung cancer occurrence in never-smoking females in Asia: TNSF-SQ, a prediction model. *Cancer Epidemiol Biomarkers Prev* 2020;29(2):452–459.

[12] Guo Y, Zheng Y, Zhong GM, et al. Analysis of the risk prediction model of lung cancer in non-smoking women. *Maternal and Child Health Care in China* 2024;39(06):1081–1084.

[13] Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *J Natl Cancer Inst* 2003;95(6):470–478.

[14] Katki HA, Kovalchik SA, Berg CD, et al. Development and validation of risk models to select ever-smokers for CT lung cancer screening. *JAMA* 2016;315(21):2300–2311.

[15] Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2011;61(592):e715–e723.

[16] Pan Z, Zhang R, Shen S, 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* 2023;88:104443.

[17] Iyen-Omofoman B, Tata LJ, Baldwin DR, et al. Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. *Thorax* 2013;68(5):451–459.

[18] Liao Z, Zheng R, Shao G. A lung cancer risk prediction model for nonsmokers: a retrospective analysis of lung nodule cohorts in China. *J Clin Lab Anal* 2022;36(11):e24748.

[19] Choi K, Park JS, Kwon YS, et al. Development of lung cancer risk prediction models based on F-18 FDG PET images. *Ann Nucl Med* 2023;37(10):572–582.

[20] Li J, Li WX, Cheng YL, et al. Construction of lung cancer risk prediction model based on low dose computed tomography screening in Shanghai community population. *J Chin Oncol* 2024;30(08):662–670. Chinese.

[21] Chen A, Wu E, Huang R, et al. Development of lung cancer risk prediction machine learning models for equitable learning health system: retrospective study. *JMIR AI* 2024;3:e56590.

[22] Beane J, Sebastiani P, Whitfield TH, et al. A prediction model for lung cancer diagnosis that integrates genomic and clinical features. *Cancer Prev Res (Phila)* 2008;1(1):56–64.

[23] Wang X, Zhang Y, Hao S, 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* 2019;21(5):e13260.

[24] Chen S, Wu S. Ensemble machine learning models for lung cancer incidence risk prediction in the elderly: a retrospective longitudinal study. *BMC Cancer* 2025;25(1):126.

[25] Lv M, Yang X, Bai Y, et al. Model of lung cancer risk with single nucleotide polymorphism in the DNA damage binding protein 2. *J Environ Occup Med* 2008;25(5):5.

[26] Li H, Yang L, Zhao X, et al. Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model. *BMC Med Genet* 2012;13:118.

[27] Zhu M, Cheng Y, Dai J, et al. Genome-Wide Association Study based risk prediction model in predicting lung cancer risk in Chinese. *Chin J Epidemiol* 2015;36(10):1047–1052. Chinese.

[28] Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst* 2007;99(9):715–726.

[29] Hoggart C, Brennan P, Tjonneland A, et al. A risk model for lung cancer incidence. *Cancer Prev Res (Phila)* 2012;5(6):834–846.

[30] Tammemägi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med* 2013;368(8):728–736.

[31] Markaki M, Tsamardinos I, Langhammer A, et al. A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types: a HUNT study. *EBioMedicine* 2018;31:36–46.

[32] Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer* 2008;98(2):270–276.

[33] Marcus MW, Chen Y, Raji OY, et al. LLPi: Liverpool Lung Project risk prediction model for lung cancer incidence. *Cancer Prev Res (Phila)* 2015;8(6):570–575.

[34] Field JK, Vulkan D, Davies MPA, et al. Liverpool Lung Project lung cancer risk stratification model: calibration and prospective validation. *Thorax* 2021;76(2):161–168.

[35] Wilson DO, Weissfeld J. A simple model for predicting lung cancer occurrence in a lung cancer screening program: the Pittsburgh predictor. *Lung Cancer* 2015;89(1):31–37.

[36] Tammemagi MC, Lam SC, McWilliams AM, Sin DD. Incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction. *Cancer Prev Res (Phila)* 2011;4(4):552–561.

[37] Zhao Q. A review of principal component analysis. *Software Eng* 2016;19(06):1–3.

[38] Levi M, Lazebnik T, Kushnir S, et al. Machine learning computational model to predict lung cancer using electronic medical records. *Cancer Epidemiol* 2024;92:102631.

[39] Dong JY, Zhang L, Zhang J, et al. Construction of a prediction model and validation evaluation of lung cancer influencing factors based on random forest algorithm. *Chin Med J (Engl)* 2023;58(11):1188–1193.

[40] Alonso E, Calle X, Gurrutxaga I, et al. Survival stacking ensemble model for lung cancer risk prediction. *Stud Health Technol Inform* 2024;321:155–159.

[41] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22(10):1345–1359.

[42] Cesa-Bianchi N, Lugosi G. *Prediction, Learning, and Games*. Barcelona: Cambridge University Press; 2006.

[43] Lyu ZY, Tan FW, Lin CQ, et al. The development and validation of risk prediction model for lung cancer: a systematic review. *Chin J Prev Med* 2020;54(4):430–437. Chinese.