A dynamic nomogram for predicting the risk of asthma: Development and validation in a database study

Lifen Yang\textsuperscript{1,2} | Meihua Li\textsuperscript{1,2} | Qinling Zheng\textsuperscript{1,2} | Chaofeng Ren\textsuperscript{1,2} | Wei Ma\textsuperscript{1,2} | Yanxia Yang\textsuperscript{1,2}

\textsuperscript{1}Department of Respiratory and Critical Care Medicine, The First Hospital of Kunming, Kunming, China
\textsuperscript{2}Kunming Medical University, Kunming, China

Correspondence
Yanxia Yang, Department of Respiratory and Critical Care Medicine, The First Hospital of Kunming, No.504, Qingnian Road, Kunming 650011, China.
Email: yyxdct@163.com

Abstract

\textbf{Background:} Asthma remains a serious health problem with increasing prevalence and incidence. This study was to develop and validate a dynamic nomogram for predicting asthma risk.

\textbf{Methods:} Totally 597 subjects whose age $\geq$ 18 years old with asthma, an accurate age at first cigarette, and clear smoking status were selected from the National Health and Nutrition Examination Survey (NHANES) database (2013–2018). The dataset was randomly split into the training set and the testing set at a ratio of 4:6. Simple and multiple logistic regressions were used for identifying independent predictors. Then the nomogram was developed and internally validated using data from the testing set. The receiver operator characteristic (ROC) curve was used for assessing the performance of the nomogram.

\textbf{Results:} According to the simple and multiple logistic regressions, smoking $\geq$ 40 years, female gender, the age for the first smoking, having close relative with asthma were independently associated with the risk of an asthma attack. The nomogram was thereby developed with the link of https://yanglifen.shinyapps.io/Dynamic_Nomogram_for_Asthma/. The ROC analyses showed an AUC of 0.726 (0.724–0.728) with a sensitivity of 0.887 (0.847–0.928) in the training set, and an AUC of 0.702 (0.700–0.703) with a sensitivity of 0.860 (0.804–0.916) in the testing set, fitting well in calibration curves. Decision curve analysis further confirmed the clinical usefulness of the nomogram.

\textbf{Conclusion:} Our dynamic nomogram could help clinicians to assess the individual probability of asthma attack, which was helpful for improving the treatment and prognosis of asthma.

\textbf{KEYWORDS} asthma, dynamic nomogram, NHANES database, smoking duration
1 | INTRODUCTION

Asthma is a chronic disease principally characterized by episodic wheeze, cough, and breathlessness resulting from airway hyperresponsiveness and inflammation, and it has affected 358 million people in the world. It is estimated that the number of asthmatics would increase by an additional 100 million in 2025. Many environmental influencing factors, including tobacco smoking, air pollution, occupations, and diet, have been identified for the onset and development of asthma. Among them, tobacco smoking has been reported to induce neutrophilic airway inflammation and airway remodeling, thereby accelerating airway obstruction and the decline of lung function, which can contribute to the onset and development of asthma. Many studies have demonstrated the cause-effect relationship between smoking and asthma, suggesting that smoking results in poorer asthma control and more severe asthma exacerbations. Asthmatics were more likely to be regular and heavier smokers than those without asthma. And continuous smokers were more likely to develop asthma, which could be interpreted as the accelerated loss of lung function over time in adult-onset asthma. However, there still exists inconsistency. Eisner et al suggested that patients with asthma had a shorter duration of smoking as compared with smokers in the general population, and no significant difference was found in the duration of smoking.

Currently, a nomogram is a useful tool for estimating the prognosis in oncology and medicine. It can simplify the traditional formulas of the prediction model into an estimated number of the probability of a single endpoint event. Based on the static nomogram, the dynamic nomogram is to generate an online scoring system, where the user-friendly interface is more convenient for clinicians to tailor clinical decisions for an individual patient. To our knowledge, a dynamic nomogram has not yet been established for predicting the probability of asthma development. Herein, this study was to investigate the predictors of the asthma attack, especially the correlation between the duration of smoking and asthma using the National Health and Nutrition Examination Survey (NHANES) database. Further, we aimed to develop and validate a dynamic nomogram for predicting asthma attacks.

2 | MATERIAL AND METHODS

2.1 | Study population

In this study, subjects from the NHANES database (2013–2018) were collected. Inclusion criteria: (1) age ≥18 years old; (2) with a confirmed diagnosis of asthma; (3) with an accurate age at first cigarette; (4) with clear smoking status: smokers or non-smokers. Baseline variables including age, gender, race, body mass index (BMI), presence or absence of asthma, age at the first cigarette, the frequency of smoking, age when first had asthma, whether having close relative with asthma, years of smoking, whether having trouble in sleeping, were extracted for analysis (See File S1).

National Health and Nutrition Examination Survey is a program of studies designed to assess the health and nutritional status of adults and children in the United States, which combines interviews and physical examinations (https://www.cdc.gov/nchs/nhanes/index.htm). The sample for the NHANES survey is selected to represent the United States population of all ages. NHANES is an open database, so the Institutional Review Board approval or informed consent by subjects are not required for this study.

2.3 | Statistical analysis

The dataset was randomly split into the training set and the testing set at a ratio of 4:6. Simple and multiple logistic regressions were used for identifying independent predictors of asthma. Then the nomogram was developed and internally validated using data from the testing set. The receiver operator characteristic (ROC) curve was used for assessing the performance of the nomogram, and the ShinyApp (RStudio, PBC) was used for making the online nomogram.

Quantitative variables were expressed as mean ± standard deviation (Mean ± SD) or median and quartile (M [Q1, Q3]), and t test or Wilcoxon rank-sum test was used for comparison between groups. Qualitative variables were expressed as number and percentage (n [%]), and χ² test or Fisher’s exact test was used for comparison. Statistical analysis was performed using SPSS 26.0 software (IBM Corp., Armonk, NY, USA), and p < 0.05 was considered statistically significant.

3 | RESULTS

3.1 | Baseline clinical characteristics

In the present study, 597 eligible subjects including 288 males (48.24%) and 309 females (51.76%) were enrolled and divided into the training set (n = 358) and the testing set (n = 239). The median age at screening was 42.00 (30.00, 57.00) years, and the median age at the first cigarette was 16.00 (14.00, 19.00) years. The 460 subjects (77.05%) reported smoking daily, 116 (19.43%) smoking regularly, and 21 (3.52%) never smoking. The 288 subjects (48.24%) reported asthma in their close relatives, and 283 (47.40%) reported no asthma in relatives. The median age for the first asthma attack was 12.00 (6.00, 30.00) years. The 381 subjects (63.82%) had asthma, and 216 (36.18%) had no asthma or had no recurrence of asthma. The details of the baseline characteristics were shown in Table 1. According to the results, there were no significant differences in all characteristics between the training set and the testing set (all p > 0.05).
3.2 | Simple and multiple logistic regression

The training set was divided into the asthma group (n = 231) and the non-asthma group (n = 127). All baseline characteristics in the training set were included in the univariate analysis, and the results showed that significant differences were found in gender (p = 0.002), age (p = 0.009), the age for the first asthma attack (p < 0.001), whether having close relative with asthma (p = 0.005) and years of smoking (p = 0.023) between the two groups (Table 2).
In this analysis, the significant characteristics in the univariate analysis were included in the multiple logistic regression. After gender and age were adjusted, we found that smoking ≥40 years, female gender, the age for the first smoking, having close relative with asthma were independently associated with the risk of the asthma attack. The risk was 3.448-fold (95% CI: 1.213–16.538) higher in subjects with ≥40 years of smoking compared to non-smokers. And the risk was 1.382-fold (95% CI: 1.461–3.925) higher in subjects having close relative with asthma compared to those without (Figure 1).

In addition, the restricted cubic splines (RCS) were used to flexibly model and visualize the relations between smoking duration and asthma risk, and between age for the first asthma attack and asthma risk. The results showed that there were no nonlinear associations between smoking duration and asthma risk ($p = 0.657$) (Figure 2), nor between age for the first asthma attack and asthma risk ($p = 0.569$) (Figure 3).

### 3.3 Development and validation of the nomogram

According to the results of the multiple logistic regression, predictors including gender, years of smoking, the age for the first asthma attack, and whether having close relative with asthma as well as age were identified and the nomogram was thereby developed with the link of https://yanglifen.shinyapps.io/Dynamic_Nomogram_for_Asthma/ (Figure 4). The results of the ROC analyses showed an AUC

| Variables, n (%) | Total (n = 358) | Non-asthma group (n = 127) | Asthma group (n = 231) | Statistics | $p$ |
|------------------|----------------|----------------------------|-----------------------|------------|-----|
| Gender           |                |                            |                       |            |     |
| Male             | 179 (50.00)    | 78 (61.42)                 | 101 (43.72)           | $\chi^2 = 9.567$ | 0.002 |
| Female           | 179 (50.00)    | 49 (38.58)                 | 130 (56.28)           |            |     |
| Age, M (Q1, Q3), years | 43.00 (30.00, 56.00) | 37.00 (28.00, 53.50) | 47.00 (32.00, 58.00) | $Z = -2.614$ | 0.009 |
| Race             |                |                            |                       |            |     |
| Mexican American | 27 (7.54)      | 12 (9.45)                  | 15 (6.49)             | Fisher     | 0.307 |
| Other Hispanic   | 28 (7.82)      | 11 (8.66)                  | 17 (7.36)             |            |     |
| Non-Hispanic white | 180 (50.28)   | 62 (48.82)                 | 118 (51.08)           |            |     |
| Non-Hispanic black | 93 (25.98)   | 30 (23.62)                 | 63 (27.27)            |            |     |
| Non-Hispanic Asian | 11 (3.07)   | 7 (5.51)                   | 4 (1.73)              |            |     |
| Other            | 19 (5.31)      | 5 (3.94)                   | 14 (6.06)             |            |     |
| BMI, kg/m$^2$    |                |                            |                       |            |     |
| Underweight/normal | 104 (29.05)   | 47 (37.01)                 | 57 (24.68)            | $\chi^2 = 5.463$ | 0.019 |
| Overweight       | 254 (70.95)    | 80 (62.99)                 | 174 (75.32)           |            |     |
| Frequency of smoking |            |                            |                       |            |     |
| Daily            | 276 (77.09)    | 92 (72.44)                 | 184 (79.65)           | Fisher     | 0.257 |
| Sometimes        | 69 (19.27)     | 29 (22.83)                 | 40 (17.32)            |            |     |
| Never            | 13 (3.63)      | 6 (4.72)                   | 7 (3.03)              |            |     |
| Age for first asthma attack, M (Q1, Q3), years | 14.00 (6.00, 30.00) | 9.00 (5.00, 16.00) | 18.00 (7.00, 35.00) | $Z = -4.612$ | <0.001 |
| Close relative has asthma |            |                            |                       |            |     |
| Yes              | 168 (46.93)    | 46 (36.22)                 | 122 (52.81)           | $\chi^2 = 10.485$ | 0.005 |
| No               | 173 (48.32)    | 76 (59.84)                 | 97 (41.99)            |            |     |
| Unknown          | 17 (4.75)      | 5 (3.94)                   | 12 (5.19)             |            |     |
| Duration of smoking |            |                            |                       |            |     |
| Not              | 15 (4.19)      | 8 (6.30)                   | 7 (3.03)              | $\chi^2 = 13.092$ | 0.023 |
| 1-10             | 66 (18.44)     | 29 (22.83)                 | 37 (16.02)            |            |     |
| 10+              | 69 (19.27)     | 30 (23.62)                 | 39 (16.88)            |            |     |
| 20+              | 62 (17.32)     | 23 (18.11)                 | 39 (16.88)            |            |     |
| 30+              | 64 (17.88)     | 18 (14.17)                 | 46 (19.91)            |            |     |
| 40+              | 82 (22.91)     | 19 (14.96)                 | 63 (27.27)            |            |     |

Abbreviation: BMI, body mass index.
| Variable                      | OR(95%CI)     | P       |
|-------------------------------|--------------|--------|
| Year of Smoking               |              |        |
| Never                         | REF          |        |
| 1–10                          | 1.274(0.333–5.109) | 0.724 |
| 10+                           | 1.344(0.388–4.827) | 0.641 |
| 20+                           | 1.785(0.534–6.134) | 0.346 |
| 30+                           | 3.018(0.879–10.541) | 0.078 |
| 40+                           | 4.448(1.213–16.538) | 0.023 |
| Age                           | 0.978(0.947–1.009) | 0.164 |
| Gender (Female)               | 1.905(1.175–3.108) | 0.009 |
| Age for First Asthma Attack   | 1.039(1.021–1.060) | <0.001 |
| Close Relative has Asthma     |              |        |
| No                            |              |        |
| Yes                           | 2.382(1.461–3.925) | <0.001 |
| Unknown                       | 2.646(0.862–9.188) | 0.101 |

**FIGURE 1** Multiple logistic regression of the training set

**FIGURE 2** RCS for smoking duration and asthma risk
FIGURE 3  RCS for age for first asthma attack and asthma risk

Dynamic Nomogram for Asthma

FIGURE 4  ROC curves of the training set and the testing set
of 0.726 (0.724–0.728) with a sensitivity of 0.887 (0.847–0.928) in the training set, and an AUC of 0.702 (0.700–0.703) with a sensitivity of 0.860 (0.804–0.916) in the testing set, suggesting the good performance of our nomogram (Figure 5). As shown in Figure 6, the net benefit of the decision curve for the nomogram is higher with a higher threshold, suggesting that our nomogram could improve the clinical outcome. Also, the calibration curve showed that the predicted probability of the nomogram was closely aligned with the actual estimates ($p = 0.226$) (Figure 7), confirming the goodness-of-fit of the model.

4 | DISCUSSION

In recent decades, asthma remains as a serious health problem with increasing prevalence and incidence. Identifying high-risk population who may suffer from asthma attacks is necessary to improve the treatment and outcome of the disease. In this population-based study, we developed and validated an online nomogram for predicting the individual probability of asthma attack (link: https://yangifenshinyapps.io/DynamNomogram_for_Asthma/). The ROC analysis for the training set showed an AUC of 0.726. After internal validation, the AUC of the testing set was 0.702, which reveals the good predictive performance of our model.

Smoking ≥40 years, female gender, having close relative with asthma, and age for the first asthma attack were identified as independent predictors for asthma. It is well-recognized that tobacco smoking can lead to inflammatory changes in the airway tract and result in the development of persistent wheezing and worsening of respiratory symptoms.\textsuperscript{17,18} Long-term smoking can accelerate the loss of lung function and finally cause the onset of asthma.\textsuperscript{12} Our findings suggested a 3.448-fold higher risk of asthma in subjects
with ≥40 years of smoking compared to non-smokers. A population-based Korean study reported similar findings that the years of smoking is likely to play a role in asthma manifestation. The long duration of smoking may lead to the vulnerability of developing lung to harmful stimuli, which accelerates the asthma onset. However, Eisner et al demonstrated that no significant difference was found in the duration of smoking. It was probably due to differences in the sample size and the smoking years of enrolled subjects between the two studies. Also, our results showed that females exhibited a 0.905-fold higher risk of asthma compared to males, which may indicate gender difference in the asthma attack. Gender disparity has been reported to exist in asthma development. Studies have found that women are more susceptible than men to respiratory diseases including asthma. According to our results, subjects who have close relative with asthma showed a 1.382-fold higher risk of asthma attack than those not, which could be hypothesized that family history may play a role in the asthma attack. A longitudinal study in Sweden noted the family history of asthma as an important risk factor for asthma development. Young et al also demonstrated that the level of airway responsiveness was increased when there was a family history of asthma.

Although previous studies have developed several prediction models for the diagnosis of asthma, we established an online nomogram which is more convenient for clinicians to evaluate the individual probability of asthma attack. Clinicians could access the website directly through their computer or phone, and input the information of patients to predict the risk of asthma attack, which would greatly facilitate the clinical use. And it enables to predict the patient's risk of asthma by evaluating simple factors such as demographic characteristics and medical history, without the need for complex examinations.

However, several limitations should also be considered. Firstly, the sample size in this study was relatively small, which requires a larger sample size for verification. Secondly, all data were extracted from NHANES, which is consisted of an in-home interview and a health examination in the Mobile Examination Center (MEC). It may impose some limitations on the accuracy of our data, thus affecting the objectivity of the results. Thirdly, the established nomogram was internally validated, which requires different populations for external validation.

In our study, smoking ≥40 years, female gender, having close relative with asthma, and age for the first asthma attack were all correlated with asthma. We developed and internally validated a dynamic nomogram to help clinicians to estimate the individual probability of asthma, which was useful for improving the treatment and prognosis of asthma.

ACKNOWLEDGMENTS
None.

CONFLICT OF INTEREST
There is no conflict of interest in this work.

AUTHOR CONTRIBUTIONS
Lifen Yang and Yanxia Yang designed the study. Lifen Yang wrote the manuscript. Meihua Li, Qinling Zheng, Chaofeng Ren, and Wei Ma collected, analyzed, and interpreted the data. Yanxia Yang critically reviewed, edited, and approved the manuscript. All authors read and approved the final manuscript.

CONSENT FOR PUBLICATION
All the contributing authors agreed for the publication in this journal.

DATA AVAILABILITY STATEMENT
All the data released to this work are available at the corresponding author.

FIGURE 7 Calibration curve for the nomogram
