Recent development in lung risk prediction model and its characteristic

S F Z Bakri¹,², A Hariri² and M Ismail³

¹Faculty of Engineering and Technology, DRB-HICOM University of Automotive Malaysia, Pekan 26607, Pahang, Malaysia
²Faculty of Mechanical and Manufacturing Engineering, University of Tun Hussien Onn Malaysia, Batu Pahat 86400, Johor, Malaysia
³School of Marine and Environmental Sciences, Universiti Malaysia Terengganu, Kuala Nerus 21300, Terengganu, Malaysia.

Abstract. Lung disease is a major public health problem worldwide. A respiratory-related cancer is one of the SEA top ten causes of death listed by WHO. To date, the problem of lung disease in SEA region has received a scant attention in the research literature and the data on the diagnosed lung cancer is still insufficient and not updated. The objective of this review is to provide a clear picture and direction of future innovation and exploration of lung risk prediction in SAE population. A computerized search of the literature resources were conducted through PubMed and Science Direct database to identify a relevant published articles reports from 2000 to 2017. The results reported may provide information on the suitability of lung risk prediction to be used in SEA region as the existing risk prediction model is limited in the quantity. For the past fourteen years, several studies have identified a number of lung cancer risk predictor and most were conducted in a European population with insufficient evaluation in Asian populations. There would, therefore, seem to be a definite need for a development of lung risk prediction model for a Southeast Asian population that had a variety of ethnic and culture.

1. Introduction
Lung disease is a major public health problem worldwide. A respiratory-related cancer is the fourth leading disease in the United States with 243,170 of estimated new cases in 2017 and this disease is one of the top ten causes of death in Southeast Asia (SEA) listed by World Health Organisation (WHO) [1], [2]. There are eleven countries under the SEA: Myanmar, Lao People’s Democratic Republic, Viet Nam, Thailand, Cambodia, Malaysia, Singapore, Indonesia, Brunei, Philippines and the Democratic Republic of Timor-Leste. Majority of the SEA countries is classified as lower middle income (64%) by World Bank while rest is upper middle income and high income with 18% recorded for each level respectively. It has been noted that most low and middle-income countries had accelerated a lung function decrement due to the prevalence of active smoking in the populations and the present of acute lung symptom had also increased due to massive biomass burning in Indonesia and this annual toxic airborne contained-haze had affected neighbour countries of Malaysia, Singapore, Brunei and Thailand [3].

However, to date, the problem of lung disease in SEA region has received a scant attention in the research literature and the data on the diagnosed lung cancer is still insufficient and not updated. In the Western developed country, investigating lung risk is a continuing concern in obtaining data for the development of risk prediction model for a decision making especially in determining study eligibility [4]. These models were capable to predict the probability of developing cancer in individuals over a
defined period and at the same time able to assist clinicians in identifying the high-risk patients. Conversely, the results and variable reported may not necessarily applicable to the SEA country as the lung capacity differs according to the anthropometric measurement of ethnicity. Apart from that, the existence of risk prediction model of lungs only focused on the general application model and not focused exactly on any occupational exposures. Therefore, this review covers relevant studies on the lung risk prediction approach and consideration of related variables in assessing lung conditions. This review is important as the results reported may provide information on the suitability of lung risk prediction to be used in SEA region as the risk prediction model is limited in the quantity. An overall brush up of approaches, development of lung risk model is taking into considerations in this paper. Hence, this effort may provide a clear picture and direction of future innovation and exploration of lung risk prediction in SAE population.

2. Methods
A literature research of related articles on the lung risk prediction model and its developmental approaches are searched in the online electronic databases published in the English language covering the years 2003 to 2017. In obtaining a specific focused on the lung prediction model, the search of journal articles has been made through the journal database by using term “pulmonary AND lung AND risk AND predictor AND model AND risk prediction” within the year 2000 to 2017 listed under category Respiratory System. Eleven relevant research publications were retrieved and selected which specifically focused on the lung health risk and saturated with the keyword ‘lung prediction model’ either in risk development and validation. Each article selected is evaluated in depth of population or ethnicity, study design and approached. The validity and strength of each study were determined based on a qualitative assessment of study objectives, methods and population. The broad study on the statistical data of lung diseases in Southeast Asia, the current trends and challenges are obtained from the computerised search in PubMed and Science Direct journal database by using keywords: lung AND disease AND Southeast Asia AND lung risk. The articles were selected according to the following criteria: original study, cohort and/or case control study design, studies that providing information on the lung cancer risk as well as the incorporated risk factors and area under the curve (AUC).

3. Application of Lung Prediction Model
In assessing a potential risk associated with lungs, a variety of assessment tools and equations has been generated and applied and this recognised either as diagnosis or prognosis tools. Diagnosis is used to identify and understand the nature of disease and/or disorder among patients while, the prediction of the potential disease and/or disorder in individual is known as prognosis. Globally, an application of prediction model is considered as an effective approach especially for clinical decision making as the number of lung disease cases is increasing. In considering the high mortality due to lung cancer and late diagnosis, a new approach needs to be implemented in identified and monitored the patients’ health conditions [5]. A number of lung cancer risk prediction model have been developed in worldwide starting on the early 2000s with several of the models have considered on the personal factors, smoking history, family history of cancer, exposure and lung conditions [6]. However, out of the listed lung risk prediction model that has been generated and validated internally and externally, only two models of the lung risk predictor are developed based on Asian ethnicity (Korean and Chinese). The details and research aim of each selected lung cancer risk tool are summarised and presented in Table 1. During the past fourteen years, several studies have identified a number of lung cancer risk predictor and most were conducted in a European population with insufficient evaluation in Asian populations. Thus, investigating lung risk prediction is a continuing concern within various ethnicities specifically in the Asian region.

Table 1. Study design review and summary of selected lung cancer risk prediction models.

| Author | Ethnicity / Subject | Model | Study Type | Approach | Area Under the Curve (Validation) | Research Aims |
|--------|---------------------|-------|------------|----------|-----------------------------------|---------------|
|        |                     |       |            |          |                                   |               |

2
| Reference | Study, Cohort | Method | Prediction | Purpose |
|-----------|---------------|--------|------------|---------|
| [16]      | American and Finnish Bach Cohort Study | Hosmer–Lemeshow test and Youden’s Index. | 0.72 (internal) | To study the possibility of variety lung risk cancer among smokers. |
| [15]      | White non-Hispanic Spitz Case-control study | Logistic regression equation: never, former, current smoker & AUC. | 0.57, 0.58 and 0.63 (external: never; current & former smoker) | To construct multivariable models for never, former and current smokers by computing the absolute risk of lung cancer and apply epidemiologic data in lung cancer study. |
| [18]      | Liverpool Lung Project Population-based study | Youden’s Index and AUC. | 0.71 (internal) | To estimate the individual absolute risk of lung cancer to be utilised as prevention and decision making. |
| [19]      | Western Europe and North America African–American Case-control study | Logistic regression and AUC. | 0.63 (external) & 0.75 (internal) | To develop and validate a lung cancer risk prediction model specifically for African-Americans based on epidemiologic case-control data analysis of lung cancer risk factors. |
| [9]       | White non-Hispanic Spitz Model Cohort study. | Hosmer–Lemeshow test and AUC. | 0.7 & 0.73 (internal) | To improve the precision of the models by incorporating DNA repair capacity as a biomarker. |
| [20]      | Northern Italy COSMOS Population study. | Hosmer–Lemeshow test | NR | To develop an epidemiological model and CT-baseline model based on clinical risk factors for a high-risk lung cancer diagnosed population and presence of nodules and emphysema in a screened population. |
| [10]      | NR Pulmonary Function with Lung Response Model Clinical Trial | AUC | 0.773 (internal) | To describe the associations between pulmonary function, sputum DNA image cytometry and lung cancer; and evaluate the possibility of these factors in the improvement of lung cancer risk prediction. |
| [21]      | Western Europe Hoggart Ever smoker cohort. | AUC | 0.843 (external 1yr) & 0.787 (external 5yr) | To build a lung cancer prediction model by applying a healthy cohort of European Prospective Investigation into Cancer and Nutrition (EPIC) with consideration of age and smoking duration. |
| [8]       | Chinese Multifactor Population study. | Logistic regression | 0.639 (external) | To evaluate the discriminatory and |
### 4. Characteristics Approach of Lung Risk Model Development

Eleven selected articles were studied to provide insight view on the development and validation of lung cancer risk prediction model for a better healthcare guideline. Six of the studied articles were categorised as epidemiological models while others are clinical model-type. Even though the risk model has been established for a clinical application in lung cancer disease, however, the models developed were both for the high-risk group and general population based on selected ethnicity and a few of the developed models were taken into considerations of work exposure as a factor and only focused on a limited work environment such as asbestos and dust only. In cancer risk prediction model, two different approaches under mathematical models were preferred: statistical and biomathematical model. A multivariable regression (logistic and linear) of the statistical model is commonly adopted to relate risk factors with cancer incidence, while, biomathematical models help to translate presumed carcinogenicity of biological process into the mathematical term. The probability of individuals to develop cancer (timeframe) based on given risk factor is known as absolute cancer risk [7]. In an existence prediction tools, an absolute cancer risk varies on a model with minimum one-year risk prediction.

The consideration of risk factors is differed according to the demand and intention of lung risk predictor and these include age, gender, BMI, ethnicity, family history of cancer, smoking factors (duration, intensity, cessation age, cessation years, pack-years); dust and/or asbestos exposure. Despite, some lung models were developed to predict the mortality of lung cancer, and while others to estimate the possibility of lung cancers incidence in individuals. Overall, most of these studies highlight the need for personal factors (age), smoking status for former and current together with smoking duration. Unlike of other lung cancer risk model, the rate of sensitivity (0.62 and 0.34) and specificity (0.70 and 0.90) only were reported for Liverpool Lung Project in 2.5% and 6% of prediction rules risk threshold respectively. Although most models rely on epidemiologic and clinical variables, some of the models had added the genetic molecular variables as the improvement of risk assessment [8]–[10]. Several risk assessment had included a traditional risk factor (smoking habit, environmental exposure, personal and/or family history of lung cancer, age, and gender), and in another study, the researcher had found that an inclusion of samples taken from the subject may enhance a predictive ability and stability advantage [8]. In a development of multivariable statistical prediction tools, the identification of suitable predictors is important as this may influence the outcome interprets from this model. The development of risk prediction model involves the four basic steps: a) develop the model, b) validate, c) implement and d) adjust [11].

As the functional assessment of the respiratory systems, results of FEV₁, FVC and FEV₁/FVC are considered by a researcher in early assessment by using spirometry test. However, only FEV₁ is

| Article | Country | Study Design | Model | Method | Validation | Purpose |
|---------|---------|--------------|-------|--------|------------|---------|
| [22]    | Northern Italy | Gender and Pack Years | Case-control study | Lexpit regression model | - | To apply lexpit regression as a novel approach to quantify risk difference associations with population-based case-control data. |
| [12]    | Korean Men Model | Cohort study | Hosmer–Lemeshow test and AUC | 0.875 (external) | | To develop an individualized lung risk prediction model for Korean men by using large population-based cohort data. |

AUC = area under the curve; NR = not reported.
presented as the results of FVC and FEV\textsubscript{i}/FVC ratio are not significantly contributed nor enhance the development of lung risk prognosis tools [10]. Multivariable predictor factors of smoking status and smoking initiation age are found to be the strongest predictor in the model [8], [12]. A simple approach to developing a prediction model is by using an existing epidemiology dataset to predict its performance in estimates individual cancer incidence [11]. The multivariable characteristics applied in the previous study of lung prognostic tools are highlighted in Table 2. Much of the lung risk prediction research has focused on identifying and evaluating the common factors as gender, age, smoking habit (includes pack-years, intensity and duration) as well as environmental exposure. However, the findings of each research may provide differences outcomes due to preference of inclusion and exclusion criteria. It is suggested that the other criteria that can be taken into account in assessing lung cancer risk is the occupational exposure, as nowadays the rate of occupational lung diseases is started to increased due to exposures to hazardous chemicals in occupational settings. Apart of looking on the common environmental exposure factors (dust and asbestos), the development of prediction model can be improvised by considering the specific occupational exposures and/or chemical substances in the workplace as this able to affect the respiratory systems especially due to frequent exposure during time-weighted average of eight hours a day. The history of employment (years of service), exposure time per day (hour), type of chemical and/or metal substance use or handle in work process, the existence of protective measure and application of protective garments/devices are the inclusion criteria for occupational exposures [13], [14].

The evaluation of prediction model is done by applying a dataset into the model and compares the results from the actual patient. This result interpreted the individual’s outcome either they had the possibility of lung cancer incidence based on area under the curve (AUC) at the receiver operating characteristics curve (ROC). Together, AUC, specificity and sensitivity is part of accuracy matrix in the prediction model. Higher event experienced by an individual expressed a high potential for lung cancer incidence to occur. A greater specificity with slightly lower sensitivity implies a higher AUC, however, this conditions had indicated as a poorer choice for clinical use [15].

| Lung Risk Model | Incorporate Risk Factors | Exclusion Criterion | Age Restriction | Absolute Risk Time Frame |
|----------------|--------------------------|---------------------|-----------------|--------------------------|
| Bach           | Age, gender, smoking duration, smoking intensity, cessation years, asbestos exposure. | NR                  | 50 – 75          | 1 – 10 years             |
| LLP            | Age; gender; smoking duration; personal history of cancer; family history of lung cancer; personal history of pneumonia; occupational history; and asbestos exposure. | Malignant tumour (melanoma). | 40 – 84          | 5 years                  |
| Spitz          | Age; gender; smoking habit*; smoking duration; smoking intensity; start smoking age; cessation age; pack years; ETS exposure; family history of smoking-related cancer and lung cancer; environmental work exposure (asbestos, wood dust, chemicals, fumes); personal history of hay fever and/or emphysema. | Chemo and/or radiation therapy; have a history of cancer. | ≥ 20             | ≥ 1 year                  |
| African – American | Age; gender; ethnicity; smoking habit**; cessation age; pack years; cessation years; environmental exposure (asbestos, wood dust, fibres, toluene, xylene, pesticides); personal history of COPD, hay fever and/or pneumonia; family history of lung or other smoking-related cancer. | Had undergone chemotherapy, radiotherapy, and/or recent blood transfusion. | ≥ 20             | 5 years                   |
| Extended Spitz Model | Age; gender; ethnicity; smoking habit; smoking duration; smoking intensity; pack years; start smoking age; cessation age; | Had undergone chemotherapy and restriction | No chemotherapy and restriction | 1 year radiotherapy age. |
In assessing a lung cancer risk, the components of predictors involved is clinical or epidemiological category such as smoking, age, family history, spirometry, gender, tobacco use, COPD, and emphysema [23, 24]). Based on research findings, cigarette smoke is one of the domain risk factor of lung cancer which had a strong relationship between lung cancer developments [25]. However, a smoke pack years provide a misleading findings where the incidence of lung cancer is increased by a factor of 4.5 (correlation with smoking duration) than 1.5 (correlation to daily consumption). Besides pack years, age is also considered as a crucial risk factors and it is recommended to assess an overall performance of the patients [23].

| Study                                      | Components                                                                 | Risk Factors                                                                 | Thresholds                  |
|--------------------------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------------|-----------------------------|
| COSMOS                                     | Age; BMI; CT screening (emphysema); smoking duration; smoking intensity; exposure factors (asbestos); dyspnoea; personal history of COPD; lung function (FEV1 and FEF); non-calcified nodule size; nodule type. | The individual was diagnosed cancers.                                        | ≥ 50 1 year                 |
| Pulmonary Function with Lung Response Model| Age; height; gender; FEV1%; SDCA; BMI; smoking duration; family history of lung cancer; pack years; smoking habit; cessation year; asbestos exposure; personal history of emphysema; | Have a history of cancer, symptomatic congestive heart failure, unstable angina pectoris, cardiac arrhythmia, severe COPD. | ≥ 40 8 years                |
| Hoggart                                    | Age; smoking habit***; start smoking age; smoking duration; smoking intensity; cessation age. | Prevalent cancers; individuals with missing cigarette smoking information; individuals that smoked cigars or pipes after quitting cigarette smoking or smoked pipes or cigars before smoking cigarettes. | ≥ 35 ≥ 1 year              |
| Chinese Multifactorial Genetic Model       | Age; gender; weighted genetic risk score (wGRS); smoking habit; count risk allele (cGRS). | Genetically related to Han Chinese, had a cancer history, had received radiotherapy or chemotherapy. | ± 60 NR                     |
| Gender & Pack Years                        | Age; gender; smoking habit; ETS; pack years; cessation year;                | Incomplete baseline questionnaire.                                           | 35 – 80 3 years             |
| Korean Men Model                           | Age; start smoking age; BMI; smoking habit; alcohol consumption; physical activity; family history of cancer; personal history of disease; fasting glucose level; | Prevalent cancer, incomplete information on the relevant risk factors.        | 30 – 80 8 years             |

NR = not reported; ETS = environmental tobacco smoke @ second-hand smoke; CT = computed tomography; BMI = body mass index; COPD = chronic obstructive pulmonary disease; FEV1 = forced expiratory volume in 1s; FEF25–75 = forced expiratory flow25–75; SDCA = sputum DNA image cytometry.
5. Conclusions
This studies highlights on the arising issues of pulmonary disease and the development of the lung prognosis tools. In this reviews paper, the findings of this research provide a clear direction and important insights into the needs and application of lung risk prediction model for future innovation and exploration, especially in Southeast Asia region as the current risk prediction model is limited. There would, therefore, seem to be a definite need for a development of lung risk prediction model for a Southeast Asian population that had a variety of ethnic and culture. Overall, the development of the risk prediction shall considered on the age, gender, ethnicity and smoking factors by looking on the smoking intensity and pack-years as one of the criteria. As the SEA had multi-cultural ethnicities, which represent a different lung size and capacity, therefore, the concern of lung risk prediction study among Southeast Asian people is suggested by presenting the value of FEV\textsubscript{1} as this is significantly will enhance the prediction tools. The other criteria that should be improvised in assessing the lung cancer risk is the occupational exposure as nowadays the rate of occupational lung diseases is started to increased due to exposures to hazardous chemicals in occupational settings.

Acknowledgment
The authors firmly declare that there are no conflicts of interest related to this work and no funding was received.

References
[1] R. L. Siegel, K. D. Miller, and A. Jemal, ‘Cancer statistics, 2017’, CA. Cancer J. Clin., vol. 67, no. 1, pp. 7–30, Jan. 2017.
[2] WHO, ‘Top 10 Causes of Death 2015’, Global Health Observatory (GHO) Data, 2017. [Online]. Available: http://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/.
[3] A. Hassan, M. T. Latif, C. I. Soo, A. H. Faisal, A. M. Roslina, Y. L. B. Andrea, and T. Hassan, ‘Short communication: Diagnosis of lung cancer increases during the annual southeast Asian haze periods.’, Lung Cancer, vol. 113, pp. 1–3, Nov. 2017.
[4] M. H. Gail, L. A. Brinton, D. P. Byar, D. K. Corle, S. B. Green, C. Schairer, and J. J. Mulvihill, ‘Projecting individualized probabilities of developing breast cancer for white females who are being examined annually.’, J. Natl. Cancer Inst., vol. 81, no. 24, pp. 1879–86, Dec. 1989.
[5] P. Petousis, S. X. Han, D. Aberle, and A. A. T. Bui, ‘Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: A dynamic Bayesian network’, Artif. Intell. Med., vol. 72, pp. 42–55, 2016.
[6] E. P. Gray, M. D. Teare, J. Stevens, and R. Archer, ‘Risk Prediction Models for Lung Cancer: A Systematic Review’, Clin. Lung Cancer, pp. 1–12, 2015.
[7] NIH, ‘Lung Cancer Risk Prediction Models: Absolute Risk Prediction Models’, Epidemiology and Genomics Research Program, 2015. [Online]. Available: https://epi.grants.cancer.gov/cancer_risk_prediction/lung.html. [Accessed: 20-Nov-2017].
[8] H. Li, L. Yang, X. Zhao, J. Wang, J. Qian, H. Chen, W. Fan, H. Liu, L. Jin, W. Wang, and D. Lu, ‘Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model’, BMC Med. Genet., vol. 13, no. 1, p. 118, Dec. 2012.
[9] M. R. Spitz, C. J. Etzel, Q. Dong, C. I. Amos, Q. Wei, X. Wu, and W. K. Hong, ‘An expanded risk prediction model for lung cancer’, Cancer Prev. Res., vol. 1, no. 4, pp. 250–254, 2008.
[10] M. C. Tammemagi, S. C. Lam, A. M. McWilliams, and D. D. Sin, ‘Incremental Value of Pulmonary Function and Sputum DNA Image Cytometry in Lung Cancer Risk Prediction’, Cancer Prev. Res., vol. 4, no. 4, pp. 552–561, Apr. 2011.
[11] G. A. Colditz and E. K. Wei, ‘Risk Prediction Models: Applications in Cancer Prevention’, Curr. Epidemiol. Reports, vol. 2, no. 4, pp. 245–250, Dec. 2015.
[12] S. Park, B.-H. Nam, H.-R. Yang, J. A. Lee, H. Lim, J. T. Han, I. S. Park, H.-R. Shin, and J. S. Lee, ‘Individualized Risk Prediction Model for Lung Cancer in Korean Men’, PLoS One, vol. 8, no. 2, p. e54823, Feb. 2013.
[13] J. M. Antonini, A. B. Lewis, J. R. Roberts, and D. A. Whaley, ‘Pulmonary effects of welding
fumes: Review of worker and experimental animal studies’, *American Journal of Industrial Medicine*, vol. 43, no. 4, pp. 350–360, 2003.

[14] S. A. Meo, M. A. Azeem, and M. M. F. Subhan, ‘Lung Function in Pakistani Welding Workers’, *J. Occup. Environ. Med.*, vol. 45, no. 10, pp. 1068–1073, Oct. 2003.

[15] A. J. Vickers and E. B. Elkin, ‘Decision Curve Analysis: A Novel Method for Evaluating Prediction Models’, *Med. Decis. Mak.*, vol. 26, no. 6, pp. 565–574, Nov. 2006.

[16] P. B. Bach, M. W. Kattan, M. D. Thornquist, M. G. Kris, R. C. Tate, M. J. Barnett, L. J. Hsieh, and C. B. Begg, ‘Variations in lung cancer risk among smokers.’, *J. Natl. Cancer Inst.*, vol. 95, no. 6, pp. 470–8, Mar. 2003.

[17] M. R. Spitz, W. K. Hong, C. I. Amos, X. Wu, M. B. Schabath, Q. Dong, S. Shete, and C. J. Etzel, ‘A Risk Model for Prediction of Lung Cancer’, *JNCI J. Natl. Cancer Inst.*, vol. 99, no. 9, pp. 715–726, May 2007.

[18] A. Cassidy, J. P. Myles, M. van Tongeren, R. D. Page, T. Liloglou, S. W. Duffy, and J. K. Field, ‘The LLP risk model: an individual risk prediction model for lung cancer.’, *Br. J. Cancer*, vol. 98, no. 2, pp. 270–6, 2008.

[19] C. J. Etzel, S. Kachroo, M. Liu, A. D’Amelio, Q. Dong, M. L. Cote, A. S. Wenzlaff, W. K. Hong, A. J. Greisinger, A. G. Schwartz, and M. R. Spitz, ‘Development and Validation of a Lung Cancer Risk Prediction Model for African-Americans’, *Cancer Prev. Res.*, vol. 1, no. 4, pp. 255–265, Mar. 2008.

[20] P. Maisonneuve, V. Bagnardi, M. Bellomi, L. Spaggiari, G. Pelosi, C. Rampinelli, R. Bertolotti, N. Rotmensz, J. K. Field, A. DeCensi, and G. Veronesi, ‘Lung Cancer Risk Prediction to Select Smokers for Screening CT—a Model Based on the Italian COSMOS Trial’, *Cancer Prev. Res.*, vol. 5, no. 6, pp. 834–846, Jun. 2012.

[21] C. Hoggart, P. Brennan, A. Tjonneland, U. Vogel, K. Overvad, J. N. Ostergaard, R. Kaaks, F. Canzian, H. Boeing, A. Steffen, A. Trichopoulos, C. Bamia, D. Trichopoulos, M. Johansson, D. Palli, V. Krogh, R. Tumino, C. Sacerdote, S. Panico, H. Bosuhiuen, H. B. Bueno-de-Mesquita, P. H. M. Peeters, E. Lund, I. T. Gram, T. Braaten, L. Rodriguez, A. Agudo, E. Sanchez-Cantalejo, L. Arriola, M.-D. Chirlaque, A. Barricarte, T. Rasmussen, K.-T. Khaw, N. Wareham, N. E. Allen, E. Riboli, and P. Vineis, ‘A Risk Model for Lung Cancer Incidence’, *Cancer Prev. Res.*, vol. 5, no. 6, pp. 834–846, Jun. 2012.

[22] S. A. Kovalchik, S. De Matteis, M. T. Landi, N. E. Caporaso, R. Varadhan, D. Consonni, A. W. Bergen, H. A. Katki, and S. Wacholder, ‘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.*, vol. 13, no. 1, p. 143, Dec. 2013.

[23] M. Adamek, E. Wachula, S. Szablowska-Siwik, A. Boratyn-Nowicka, and D. Czyzewski, "Risk factors assessment and risk prediction models in lung cancer screening candidates," *Annals of Translational Medicine*, vol. 4, no. 8, pp. 151-151, 2016.

[24] D. Whitney, M. Elashoff, K. Porta-Smith, A. Gower, A. Vachani, J. Ferguson, G. Silvestri, J. Brody, M. Lenburg and A. Spira, "Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy", *BMC Medical Genomics*, vol. 8, no. 1, 2015.

[25] A. J. Alberg, M. V. Brock, J. G. Ford, J. M. Samet, and S. D. Spivack, "Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines", *Chest*, 143, e1S-29S, 2013.