Cancer Letters 420 (2018) 236–241

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Original Articles

Development and validation of a novel diagnostic nomogram model based on tumor markers for assessing cancer risk of pulmonary lesions: A multicenter study in Chinese population

Qiang Du [a, b, 1], Cunling Yan [c, 1], San-Gang Wu [d, 1], Wei Zhang [e, 1], Chun Huang [b], Yiyong Yao [b], Liyu Wang [f], Qunji Zhang [g], Qinghao Liu [h], Jie Guan [c], Yanfeng Hou [c], Zhiyan Li [c], Andrew Soh [i], Agim Beshiri [i], Qi Wang [a, *], Xun Li [g, *], Yijie Zheng [i, *], Huiling Wang [a, *]

[a] Department of Respiratory Medicine, The Second Affiliated Hospital, Dalian Medical University, Dalian, China
[b] Department of Respiratory Medicine, The North Area of Suzhou Municipal Hospital, Suzhou, China
[c] Department of Clinical Laboratory, Peking University First Hospital, Beijing, China
[d] Department of Radiation Oncology, Xiamen Cancer Hospital, The First Affiliated Hospital of Xiamen University, Xiamen, China
[e] Department of Biostatistics, School of Public Health, Fudan University, Shanghai, China
[f] Department of Oncology, The North Area of Suzhou Municipal Hospital, Suzhou, China
[g] Department of Laboratory Medicine, The First Affiliated Hospital, Medical College of Xiamen University, Xiamen, Fujian 361003, China
[h] Department of Pulmonary Surgery, Peking University First Hospital, Beijing, China
[i] Medical Scientific Liaison Asian Pacific, Abbott Diagnostics Division, Abbott Laboratories, Asian Pacific Group, China

ARTICLE INFO

Article history:
Received 21 October 2017
Received in revised form
8 January 2018
Accepted 30 January 2018

Keywords:
Lung cancer
Biomarker
Diagnosis

ABSTRACT

Purpose: This study aimed to build a valid diagnostic nomogram for assessing the cancer risk of the pulmonary lesions identified by chest CT.

Patients and methods: A total of 691 patients with pulmonary lesions were recruited from three centers in China. The cut-off value for each tumor marker was confirmed by minimum *P* value method with 1000 bootstrap replications. The nomogram was based on the predictive factors identified by univariate and multivariate analysis. The predictive performance of the nomogram was measured by concordance index and calibrated with 1000 bootstrap samples to decrease the overfit bias. We also evaluated the net benefit of the nomogram via decision curve analysis. Finally, the nomogram was validated externally using a separate cohort of 305 patients enrolled from two additional institutions.

Results: The cut-off for CEA, SCC, CYFRA21-1, pro-GRP, and HE4 was 4.8 ng/mL, 1.66 ng/mL, 1.83 ng/mL, 56.55 pg/mL, and 63.24Lpmol/L, respectively. Multivariate logistic regression model (LRM) identified tumor size, CEA, SCC, CYFRA21-1, pro-GRP, and HE4 as independent risk factors for lung cancer. The nomogram based on LRM coefficients showed concordance index of 0.901 (95% CI: 0.842–0.960; $P < 0.001$) for lung cancer in the training set and 0.713 (95% CI: 0.599–0.827; $P < 0.001$) in the validation set. Decision curve analysis reported a net benefit of 87.6% at 80% threshold probability superior to the baseline model.

Conclusion: Our diagnostic nomogram provides a useful tool for assessing the cancer risk of pulmonary lesions identified in CT screening test.

1. Introduction

Lung cancer has been the most common cancer and leading cause of cancer-related deaths worldwide for decades. In 2012, an estimated 1.8 million new lung cancer cases occurred worldwide, accounting for about 13% of total cancer diagnoses [1]. The incidence has kept on growing possibly due to the aging population and

* Corresponding authors.
E-mail addresses: wqdlmu@163.com (Q. Wang), xli2001@xmu.edu.cn (X. Li), yijie.zheng@abbott.com (Y. Zheng), Whl882728@163.com (H. Wang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.canlet.2018.01.079
0304-3835/

environmental pollution [2]. As the stage of lung cancer advances, the prognosis becomes significantly worse, evidenced by the 5-year survival rate of 60–92% for localized cancer, but only 13%–53%% and 0%–10% for advanced and metastatic disease, respectively [3,4]. Thus, development of a reliable tool which can improve lung cancer diagnosis is essential for better patient outcome.

In recent years, the widespread use of computed tomography (CT) has greatly improved the detection of pulmonary nodules [5]. In several large-scale lung cancer screening studies, the detection rate of pulmonary nodules has increased from 8% to 51%. Malignant nodules account for less than 10% of these nodules [6]. Low-dose computed tomography (LDCT) is recommended as an efficient tool for screening lung cancer [7], but it also raises controversy due to higher false positive rate. In a previous report, uncalcified pulmonary nodule was detected in 69% of the people at risk, but less than 4% were finally diagnosed as lung cancer [8]. High false positive rate is associated with adverse effects such as anxiety, additional imaging tests and biopsy procedures, and in rare cases, may result in serious outcomes, including hospitalization and death [7]. When a lung lesion is identified, it's essential to assess the nature of the lesion and the risk of cancer.

Serum biomarkers are easily accessible, which can be analyzed readily as an approach supplementary to the traditional imaging techniques to enhance the overall diagnostic capability. CEA, SCC, CYFRA21-1 and pro-GRP are commonly used biomarkers in lung cancer diagnosis, disease monitoring and prognosis, which is recommended by both the National Academy of Clinical Biochemistry (NACB) and European Group on Tumor Markers (EGTM) [9]. Significantly higher blood levels of pro-GRP for small cell lung cancer patients, CEA for adenocarcinoma lung cancer patients, and both SCC and CYFRA21-1 for squamous lung cancer patients were observed compared with benign diseases [10,11]. HE4 is originally used in ovarian cancer risk stratification. Emerging evidences have been supporting its application in lung cancer, with positive expression in lung cancer cells and with better performance for diagnosis of lung cancer at early stage [12,13]. Particularly, the combination of several biomarkers can improve the detection of lung cancer by overcoming the low sensitivity and specificity of each serum biomarker alone [13,14]. However, tumor biomarkers alone are not enough in confirming the diagnosis given the pathological and histological heterogeneity of lung cancer.

Nomogram is the graphic statistical description and calculates the overall effect of the variables on the probability of specific outcome. A nomogram can provide individualized, evidence-based, and highly accurate estimation of risk by combining multiple independent variables and assigning an appropriate weight to each variable based on its prognostic value. Nomogram is widely used in the evaluation of tumor metastasis, recurrence and prognosis [15,16]. Moreover, nomogram has been claimed as an alternative or even a new standard for the management of many cancers, comparing favorably to the traditional staging systems [15,17,18]. To our knowledge, this is the first attempt to develop a diagnostic nomogram for lung cancer based on tumor markers and the size of lung lesions in a multicenter study in Chinese population.

2. Material and methods

2.1. Study population

The study subjects were recruited from three centers (The Second Affiliated Hospital of Dalian Medical University, The First Affiliated Hospital of Xiamen University, and Peking University First Hospital) in China from October 2015 to May 2017. The diagnosis was confirmed by pathology (surgical resection and/or biopsy) in all patients. The staging of non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) was determined according to the criteria of AJCC Cancer Staging Manual, 8th Edition and the Veterans Administration Lung Cancer Study Group [3,19]. All patients were naïve to antineoplastic therapy, radiotherapy or chemotherapy before surgery or cancer diagnosis. The study protocol was approved by the ethics committees of the above-mentioned hospitals.

2.2. Assay of tumor markers

A commercial chemiluminescent microparticle immunoassay (CMIA) method was used to test serum samples for CEA, SCC, CYFRA21-1, HE4 and pro-GRP on ARCHITECT *i*2000SR, an automated immunoassay analyzer (Abbott Laboratories, Chicago, Ill).

2.3. Image acquisition and analysis

The patients were examined further by high resolution computerized tomography (HRCT) if their lung lesion was less than 30 mm by regular multi-detector CT (MDCT) scan. A standard protocol was used for the scanning with an MDCT scanner (Philips Brilliance CT 64-channel scanner; Philips Healthcare, Best, the Netherlands). The scan was carried out from the thoracic inlet to the middle portion of the kidneys and the parameters were set as follows: slice thickness, 1.00 mm; scan mode, helical; tube voltage, 120 kV; tube current, automatic; and filter, L. No contrast material was used. Crude HRCT data were copied from our hospital's fully digitalized Picture Archiving and Communication System onto dedicated Philips Brilliance CT workstation ver. 4.02, and prepared in blind manner for use in the study using standard software 'Anonymize'. Two senior consultant radiologists reviewed and measured the HRCT scan images of lung lesions in blind manner. No patient data were visible to the readers and they had no access to each other's results. CT findings were interpreted by consensus of readers.

2.4. Statistical analysis

Continuous variables are presented as mean (SD) or median (p25, p75) and categorical variables are expressed as frequency and percentage. T test or Mann-Whitney test were used to compare continuous variables, while categorical variables were compared by using Chi-square test or Fisher exact test. The cut-off value for each marker was confirmed by minimum *P* value method with 1000 bootstrap replications to overcome the variation of a single sample. The cut-off of a single bootstrap sample was estimated by minimum *P* value approach [20] and the mean of 1000 individual cut-off values as the final cut-off.

The factors with $P < 0.1$ in univariate analysis was used as candidate risk factors for stepwise multivariate logistic regression model (LRM) with a backward selection. A nomogram was elaborated upon LRM coefficients by using RMS package of R. The predictive performance of the nomogram was measured by concordance index (C index), calibrated with 1000 bootstrap samples and validated in external sample. Finally, we evaluated whether the model improved the predictive net benefit via decision curve analysis (DCA) [21]. All tests were both sided and 0.05 was set as the *P* value of statistical significance. Statistical analysis was performed using R programming language v.3.3.2.

3. Results

3.1. Pathologic characteristics and biomarkers of patients

Overall, the 386 patients in the training set included 297 lung cancer patients and 89 patients with benign nodules (control