**Risk Stratification of Patients Using the Protein-based Lung Cancer Biomarker Panel in China**

**Abstract**

**RATIONALE:** There was no good biomarker panel based model to evaluate the probability of lung cancer. To verify clinical utility of a simple blood test comprising of four protein-based markers for risk stratification of patients presenting symptoms related to lung cancer.

**METHODS:** The study is a multicenter, prospective case-control research (NCT01928836). We tested panel of lung cancer biomarker and collected the diagnosis procedure of a patient with undiagnosed lesion in the lung, including epidemiological information, CT scan result.

**RESULTS:** We first compared the correlation between our biomarker panel based risk model and the pathological biopsy result (training test). And then we test the model among all enrolled patients (validation test). And the aim of the test panel is to screening out the possible lung cancer patients, so we defined a model with higher sensitivity. While for urban models, we suggested to include CT scan result, and also we provided a higher specificity model for the clinicians, which in order to increase the accuracy of the finial pathological diagnosis and decrease the unnecessary biopsy. We found the model included age, sex, smoke history and the biomarker panels (Pro GRP, SCC, Cyfra21-1 and CEA) with or without the sum of lesion diameters could both provide the 95% sensitivity or 95% specificity in the lung cancer risk model. And area under the ROC could be 0.7649- 0.8443 in training cohort and 0.7472- 0.8220 in validation cohort.

**CONCLUSION:** We first time tested the national application of a panel of four biomarkers to assist clinical evaluation of lung cancer risk probability. We found four equations suitable for both rural and urban hospital as well as with or without CT scan. By further health economics analysis, we elucidated that these biomarker panel could increase the efficiency of current lung cancer screening procedure and to decrease the cost of lung cancer in China.

**Background**

The early diagnosis of lung cancer patients is always very important issue. Early in 60 years ago, Dr. Overholt emphased that ‘Every physician who is in active practice shoulders part of the cancer burden.’ 1. Due to the 10-year survival rate of stage I patients could be 88%, while if the patients received surgical resection within 1 month after diagnosis, the survival rate could increase to 92% 2. Current by the National Lung Screening Trail (NLST) (ClinicalTrails.gov number, NCT00047385) study result, low dose computed tomography (LDCT) scan for screening in high-risk population (with a history of cigarette smoking ≥ 30 pack-years and elder between 55 and 74 years old) reduced 20% mortality from lung cancer 3. And also in NELSON trail, by applying CT screening tragedy, more often diagnosed at stage I and less often at stage IIIB-IV 4. While the due to different demographics and health care system, the benefit of LDCT might be limited. Rotterdam Study in Netherlands found there was minority of patients with lung cancer meets the NLST high-risk criteria, which means it should with caution to apply the screening model 5. DANTE trail also found there is limited statistical power to support the efficacy of LDCT screening 6. Besides, from PLuSS cohort, the false positive results by CT scan also lead to unnecessary invasive procedure 7. Currently, in American College of Chest Physican (ACCP) and Chinese Alliance Against Lung Cancer (CAALC) guidelines, the follow-up procedure was mainly depended on the nodule diameters 8,9. Besides, one adjusted model from Mayo Clinic includes Age, Smoke, Cancer, Diameter, Speculation and Position is combined to evaluate the probability 10.

By combining multi-biomarker test in blood improve single assay detection of early cancer 11. Also there were several new blood biomarkers validated to help detect early-stage of lung cancer patients 12,13. And some study tried to combine the biomarker with clinical parameters, which could improve the efficacy of the panel 14. However, plasma DNA level is found limited to improve the accuracy of lung cancer CT screening 15. Previously we did a preliminary single center study on lung cancer biomarker panel study, and found combined ProGRP, CEA, SCC and CYRFRA21-1 could improve CT scan sensitivity for patients with lung cancer 16. In this study we report the development and validation of models and calculators for predicting the probability of lung cancer in population with high risk of lung cancer using data from Lung Cancer Biomarker Panel in China (LCBP) study.

**Methods**

People’s Liberation Army ()

**Overall Study Design**

This study was designed to evaluate a combination of four serum proteins and clinical information to differentiate benign from malignant pulmonary nodules in patients at risk of having lung cancer. For the nodule risk model, we first analyzed a training set of 163 patients with indeterminate nodules using logistic regression analyses that assigned patients to a low-, medium- or high-risk category. These models were then tested in an independent, masked validation set of 126 patients. For the high-risk model, we first analyzed a training set of 240 patients with high risk factor of lung cancer using logistic regression analyses that assigned patients to a low-, medium- or high-risk category. These models were then tested in an independent, masked validation set of 179 patients.

In the model building logistic regression process, we include all the following variables in to the model: age, sex, smoke status, diameter of nodules, Pro GRP, SCC, CYFRA211, CEA.To take all the effect into account, we conduct a full model with all variables included.

**Serum Marker Assays**

We established a lung cancer biomarker panel (LCBP) that measures the serum levels of progastrin-releasing peptide (ProGRP), carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCC) and cytokeratin 19 fragment (CYFRA21-1). The serum samples were tested via a commercial chemiluminescent microparticle immunoassay (CMIA) method with ARCHITECT i2000SR, an automated immunoassay analyzer (Abbott Laboratories, Chicago, Ill). All samples in the training set were run in duplicate. For the validation study, all assays were run on automated analyzers in singlicate per the manufacturer’s instructions.

**Imaging Studies, Nodule Size, and Pathologic Classification**

All patients enrolled in either the training or validation study had a CT study of the thorax. CT studies at five institutions were performed using multiple contiguous 5-mm sequential axial images through the thorax. The size of the nodule was analyzed by Pneuview (Myrian, Pairs). The longest dimension on axial CT images was record. All patients with lung cancer had histologic confirmation and pathologic staging. Patients without lung cancer had a histologic diagnosis, CT resolution of 2-year stability of the lesion, or clinical observation without evidence of lung cancer.

**Statistical Methods**

The association between various characteristics of a pulmonary nodule (serum levels of ProGRP, CEA, SCC, CYFRA21-1 and nodule size) and the presence of lung cancer was assessed within the training set data using logistic regression. The LCBP model was generated using risk model analysiswhich is based on logistic regression model. The logistic model provided an estimated of probability that cancer was present as a function of nodule diameter, ProGRP, CEA, SCC, CYFRA21-1, and other characteristics. Two threshold probabilities were considered for predicting a cancer diagnosis: a calculated probability of 0.68 and higher, and a threshold probability of 0.13. The 0.68 threshold is a standard probability which is the left-top point of ROC curve and will provide most sensitivity and specificity for diagnosis and the 0.13 threshold is intended to reduce the number of false-negative diagnoses, which is the point with sensitivity of 0.95.

An independent validation study was designed to test the accuracy of the logistic regression models developed from the training set. Each model was used to assign each subject within the validation data set a “lung cancer” (high-risk group) or “no lung cancer” (low-risk group) designation. After classification of all patients, the diagnosis was revealed and the sensitivity, specificity, and positive and negative predictive values for each model were determined. This validation study was designed to specifically assess the sensitivity and specificity of the derived classification algorithm for the presence or absence of cancer within a patient group representative of all patients with a pulmonary nodule or focal pulmonary abnormality.

**Results**

**Biomarker level in the total population**

Significant higher levels of ProGRP (P<0.001), SCC (P<0.001), CYFRA21-1 (P<0.001) and CEA (P=0.027) for lung cancer patients were observed in comparison with the clinical inconsiderable lung biopsy population (Figure 2A). Similar results were obtained when we compared the lung cancer patients (ProGRP, P<0.001; SCC, P<0.001; CYFRA21-1, P<0.001; CEA, P=0.304) with patients with normal biopsy result. Besides, significant higher level of ProGRP for small cell lung cancer was observed in both comparison with the adenocarcinoma (P<0.001) and squamous cell carcinoma (P<0.001) (Figure 2B). All P values were based on parametric rank tests (Independent Sample T Test).

Figure 2C presents the serum levels of the 4 biomarkers for lung cancer patients with nodule with diameter less than 20mm and patients with lung nodule or lesion with diameter longer than 20mm. Significant higher level of ProGRP (P=0.035) for patients with larger lesion was observed in comparison with the smaller lesion. Similar results, was obtained when we compared the advanced stage (IIIB-IV) patients (CEA, P<0.01) with patients with limied stage (I-IIIA) (Figure 2D).

**Training Set Results**

A total of 410 achieved serum samples were included in the training dataset. Patients characteristics including age and sex are summarized for the 157 patients with a confirmed diagnosis of lung cancer and the 236 individuals without cancer in Table 1. The logistic regression model used nodule size, ProGRP, CEA, SCC, CYFRA21-1, age, smoke history, malignant tumor history and sex developed to predict diagnosis. If a patient with a calculated cancer probability of 0.65 or higher is considered to have cancer and one with a probability less than 0.21 is considered not to have cancer, the overall sensitivity, specificity are 94.2% and 94.5%.

**Validation Study Results**

A total of 354 patients fulfilled the eligibility criteria and were included in the independent validation study. The diagnosis of primary lung cancer was established in 140 patients and a benign abnormality in 186 (Table 2). LCBP data, nodule size and other clinical data were entered into logistic regression model derived from the training set, and a ‘lung cancer’ (high-risk) or ‘no lung cancer’ (low-risk) diagnosis was assigned for each patients. Assuming a probability threshold of 0.65 for assigning a patient to the high-risk group and assuming a threshold of 0.21 for assigning a patient to the low-risk group, the sensitivity and specificity for all patients in the validation study were 92.9% and 99.5%.

**Discussion**

Many factors will increase the delay of diagnosis of lung cancer, such as failure to recognize documented abnormal imaging results, failure to receive key diagnostic procedures in a short period and errors on understanding of pathological report 18,19. Current early diagnosis procedure for lung cancer is mainly in comprehensive hospitals, which need expensive examination cost and huge human labor resource to collaborate. There are two prospective clinical lung cancer screening trail based on blood biomarker panel, one is Early Cancer detection test – Lung cancer Scotland (ECLS Study) conducted by National Health Service (NHS)[], and one is studied by National Jewish Health hospital[James Jett]. The future 4P medicine model could provide sophisticated model to help patient understand the early disease more effectively. Certain study showed nurse navigator support program will improve patient experience and reduce lung cancer costs after early diagnosis 20. And here we provide an easy model to input different parameters and to stratify patients into difference risk level of lung cancer.

As multi-slice spiral CT and LDCT scan are widely used to screen the subjects at high risk of lung cancer, the detection of solitary pulmonary nodules is significantly increased, ranging from about 8% to 51%. In particular, the solitary pulmonary nodules with a diameter less than 1 cm are detected much more than ever. Only 1.1% to 12.0% of these solitary nodules are carcinous. It is difficult to differentiate malignant tumors from benign ones due to the limitations of current diagnostic methods, especially for the solitary pulmonary nodules with a diameter less than 8 mm. The lung cancer patients detected at early stage will have significant better survival if they are treated by an experienced respiratory physician rather than by a non-respiratory physician. LCBP algorithm can be used to evaluate the risk of lung cancer before operation in the patients with pulmonary nodules. The data of our study in Chinese population have shown that the accuracy of LCBP is nearly 90% in identifying lung cancer. Such a screening tool is helpful for doctors at different levels, especially the doctors working in community and remote areas. Identifying malignant nodules at early stage is favorable for selecting more appropriate and right treatment strategy to achieve optimal treatment efficacy.

Currently, the studies targeting the risk models of pulmonary nodules are generally based on the European and American populations, for example, the diagnostic algorithm for estimating the risk of pulmonary nodules published by the American College of Chest Physicians (ACCP) [Figure 3&4]. However, it is not known if such a model is applicable for Chinese population, or whether the data from a single center are sufficient. So, for the purpose to obtain more scientific and real world data, the first multi-center LCBP study based on Chinese population was initiated in October 2012, which was led by the Department of Respiratory Diseases, Shanghai Zhongshan Hospital, Fudan University. This study was also conducted in other four hospitals, including Peking Union Medical College Hospital, Xinqiao Hospital affiliated to the Third Military Medical University, Henan Provincial Hospital, and Nanjing PLA General Hospital. This study aims to build local dataset, which is then used to validate the accuracy of LCBP. The results of this study will help to establish and standardize the practice processes of early diagnosis of lung cancer in China, and provide sound evidence for clinical management of lung cancer.

**Conclusion**

Tumor biomarker testing is simple to operate at relatively low cost and easy to follow up due to its non-invasive nature. Meanwhile, these biomarkers can reflect the disease state more objectively with shorter turnaround time, which can save the time waiting. As novel tumor biomarkers are emerging, they are providing more and more useful clinical information, e.g., for risk stratification, histological typing, and monitoring treatment efficacy and evaluating prognosis. Our LCBP diagnostic model also incorporates clinical high risk factors and CT imaging information, which can be used as an aide to help clinicians at every level to comprehensively and scientifically evaluate the risk of lung cancer in the patients with pulmonary nodules.

**Reference**

1 Overholt RH, Woods FM. Early diagnosis and treatment of cancer of the lung. The New England journal of medicine 1951;245:555-559.

2 International Early Lung Cancer Action Program I, Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage i lung cancer detected on ct screening. The New England journal of medicine 2006;355:1763-1771.

3 The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395-409

4 Horeweg N, van der Aalst CM, Thunnissen E, Nackaerts K, Weenink C, Groen HJ, Lammers JW, Aerts JG, Scholten ET, van Rosmalen J, Mali W, Oudkerk M, de Koning HJ. Characteristics of lung cancers detected by computer tomography screening in the randomized nelson trial. American journal of respiratory and critical care medicine 2013;187:848-854.

5 Heuvers ME, Stricker BH, Aerts JG. Generalizing lung-cancer screening results. The New England journal of medicine 2012;366:192-193.

6 Infante M, Cavuto S, Lutman FR, Passera E, Chiarenza M, Chiesa G, Brambilla G, Angeli E, Aranzulla G, Chiti A, Scorsetti M, Navarria P, Cavina R, Ciccarelli M, Roncalli M, Destro A, Bottoni E, Voulaz E, Errico V, Ferraroli G, Finocchiaro G, Toschi L, Santoro A, Alloisio M, Group DS. Long-term follow-up results of the dante trial, a randomized study of lung cancer screening with spiral computed tomography. American journal of respiratory and critical care medicine 2015;191:1166-1175.

7 Wilson DO, Weissfeld JL, Fuhrman CR, Fisher SN, Balogh P, Landreneau RJ, Luketich JD, Siegfried JM. The pittsburgh lung screening study (pluss): Outcomes within 3 years of a first computed tomography scan. American journal of respiratory and critical care medicine 2008;178:956-961.

8 Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013 May;143(5 Suppl):e93S-120S.

9 Lung cancer group of the Chinese Medical Association, Chines Alliance Against Lung Cancer Expert Group. Diagnosis and treatment of pulmonary nodules in chinese expert consensus. Chin J Tuberc Respir Dis 2015;38:249-254.

10 Herder GJ, van Tinteren H, Golding RP, Kostense PJ, Comans EF, Smit EF, Hoekstra OS. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. Chest 2005; 128: 2490-2496.

11 Yurkovetsky Z, Skates S, Lomakin A, Nolen B, Pulsipher T, Modugno F, Marks J, Godwin A, Gorelik E, Jacobs I, Menon U, Lu K, Badgwell D, Bast RC, Jr., Lokshin AE. Development of a multimarker assay for early detection of ovarian cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2010;28:2159-2166.

12 Qiu J, Choi G, Li L, Wang H, Pitteri SJ, Pereira-Faca SR, Krasnoselsky AL, Randolph TW, Omenn GS, Edelstein C, Barnett MJ, Thornquist MD, Goodman GE, Brenner DE, Feng Z, Hanash SM. Occurrence of autoantibodies to annexin i, 14-3-3 theta and lamr1 in prediagnostic lung cancer sera. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2008;26:5060-5066.

13 Blasberg JD, Pass HI, Goparaju CM, Flores RM, Lee S, Donington JS. Reduction of elevated plasma osteopontin levels with resection of non-small-cell lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2010;28:936-941.

14 Yee J, Sadar MD, Sin DD, Kuzyk M, Xing L, Kondra J, McWilliams A, Man SF, Lam S. Connective tissue-activating peptide iii: A novel blood biomarker for early lung cancer detection. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2009;27:2787-2792.

15 Sozzi G, Roz L, Conte D, Mariani L, Andriani F, Lo Vullo S, Verri C, Pastorino U. Plasma DNA quantification in lung cancer computed tomography screening: Five-year results of a prospective study. American journal of respiratory and critical care medicine 2009;179:69-74.

16 Yang DW, Zhang Y, Hong QY, Hu J, Li C, Pan BS, Wang Q, Ding FH, Ou JX, Liu FL, Zhang D, Zhou JB, Song YL, Bai CX. Role of a serum-based biomarker panel in the early diagnosis of lung cancer for a cohort of high-risk patients. Cancer 2015; 121 Suppl 17: 3113-3121.

17 Singh H, Hirani K, Kadiyala H, Rudomiotov O, Davis T, Khan MM, Wahls TL. Characteristics and predictors of missed opportunities in lung cancer diagnosis: an electronic health record-based study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2010; 28: 3307-3315.

18 Singh H, Hirani K, Kadiyala H, Rudomiotov O, Davis T, Khan MM, Wahls TL. Characteristics and predictors of missed opportunities in lung cancer diagnosis: An electronic health record-based study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2010;28:3307-3315.

19 Singh H, Sethi S, Raber M, Petersen LA. Errors in cancer diagnosis: Current understanding and future directions. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2007;25:5009-5018.

20 Wagner EH, Ludman EJ, Aiello Bowles EJ, Penfold R, Reid RJ, Rutter CM, Chubak J, McCorkle R. Nurse navigators in early cancer care: A randomized, controlled trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2014;32:12-18.