Lung cancer screening: identifying the high risk cohort

Michael W. Marcus, Olaide Y. Raji, John K. Field

Roy Castle Lung Cancer Research Programme, the University of Liverpool Cancer Research Centre, Institute of Translational Medicine, the University of Liverpool, Liverpool, UK

Correspondence to: Prof. John K. Field, Director of Research. Roy Castle Lung Cancer Research Programme, The University of Liverpool Cancer Research Centre, Department of Molecular and Clinical Cancer Medicine, Institute of Translational Medicine, The University of Liverpool, 200 London Road, Liverpool L3 9TA, UK. Email: j.k.field@liv.ac.uk.

Abstract: Low dose computed tomography (LDCT) is a viable screening tool for early lung cancer detection and mortality reduction. In practice, the success of any lung cancer screening programme will depend on successful identification of individuals at high risk in order to maximise the benefit-harm ratio. Risk prediction models incorporating multiple risk factors have been recognised as a method of identifying individuals at high risk of developing lung cancer. Identification of individuals at high risk will facilitate early diagnosis, reduce overall costs and also improve the current poor survival from lung cancer. This review summarises the current methods utilised in identifying high risk cohorts for lung cancer as proposed by the Liverpool Lung Project (LLP) risk model, Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial risk models and the prediction model for lung cancer death using quintiles. In addition, the cost-effectiveness of CT screening and future perspective for selecting high risk individuals is discussed.

Keywords: Lung cancer; screening; low dose computed tomography (LDCT); high-risk individuals; cohort

Introduction

Lung cancer is the leading cause of cancer death among males in both more and less developed countries, and has overtaken breast cancer as the leading cause of cancer death among females in more developed countries (1). The overall 5-year survival rate of lung cancer varies globally but is consistently low due to late stage detection and the paucity of late-stage interventions (2,3). Low dose computed tomography (LDCT) is a viable screening tool for early lung cancer detection and mortality reduction. The USA-based National Lung Screening Trial (NLST) demonstrated a 20% reduction in lung cancer mortality relative to chest X-ray screening (4). In practice, the success of any lung cancer screening programme will depend on successful identification of individuals at high risk (5,6). Risk prediction models incorporating multiple risk factors have been recognised as a method of identifying individuals at high risk of developing lung cancer (7,8). Risk models hold promise for improving patient care by aiding clinicians'

decision making processes regarding choice of interventions and/or treatments. They can also guide selection of individuals at the population level, for screening: this ensures limited resources are focussed on those individuals who are most likely to benefit. The utilisation of a risk strategy ensures minimisation of unnecessary invasive and potentially harmful interventions. Identification of individuals at high risk will facilitate early diagnosis, reduce overall costs and also improve the current poor survival from lung cancer.

In 2012, a systematic review was conducted using MEDLINE, EMBASE and Cochrane library to examine the evidence regarding the benefits and harms of lung cancer screening using LDCT. The study recommended that LDCT for lung cancer screening should be offered to individuals that meet the NLST inclusion criteria; asymptomatic individuals aged 55 to 74 years with a minimum of 30 pack-years of smoking and no more than 15 years since quitting (9). This proposal has been endorsed by a number of prominent societies such as the

American Cancer Society, the American College of Chest Physicians, the American Society of Clinical Oncology, National Comprehensive Cancer Network, the International Association for the Study of Lung Cancer, and the US Preventive Services Task Force (USPSTF) (9-11). The USPSTF recommends annual screening for lung cancer with LDCT in adults aged 55-80 years who have a 30 pack-year smoking history and currently smoke or have quit within 15 years. Screening should be discontinued once a person has not smoked for 15 years or develops health problem that substantially limits life expectancy or the ability or willingness to have curative lung surgery. Henceforth, this recommendation will be referred to as USPSTF criteria in this review.

Although screening of high risk individuals has been recognised as the way forward to reduce the excessive high mortality of lung cancer patients (12), identifying high risk cohort remain an unresolved issue (13,14). In order to maximise the benefit-to-harm ratio of lung cancer screening, it is important to identify which individuals are at sufficiently high risk of the disease, and to target screening to these people. This process involves identification of risk factors, quantitative summary of overall risk, and selection of a suitable cut-off value for CT screening (6). Since lung cancer is mainly attributable to cigarette smoking, and occurs amongst elderly populations (15), the selection criteria for eligible participants in the current two largest randomised controlled trials (NLST and NELSON) were based on smoking history and age (4,9,16). However, the risk of lung cancer is also influenced by other factors such as prior diagnosis of malignant tumour, early onset (<60 years) family history of lung cancer, occupational exposure to asbestos and previous lung diseases such as chronic obstructive pulmonary disease (COPD) (including emphysema and chronic bronchitis), pneumonia, and tuberculosis, cooking fumes, ionising radiation and radon gas (17-20). Therefore, a robust risk assessment which could account for additional risk factors excluded by the NLST and NELSON criteria could improve the selection criteria of lung cancer screening. This review summarises the current methods utilised in identifying high risk cohorts for lung cancer as proposed by the Liverpool Lung Project (LLP) risk model, Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial risk models and the prediction model for lung cancer death using quintiles (*Table 1*). In addition, the cost-effectiveness of CT screening and future perspective for selecting high risk individuals is discussed.

The Liverpool Lung Project (LLP)

The LLP risk model was developed from a case-control study (21). Using a model-based approach, the LLP estimated the probability that an individual, with a specific combination of risk factors, would develop lung cancer within a 5-year period. In short, data from 579 lung cancer cases and 1,157 age- and sex-matched population-based controls were used. Conditional logistic regression models were used to model significant risk factors. Smoking duration, prior diagnosis of pneumonia, occupational exposure to asbestos, prior diagnosis of malignant tumour and early onset (<60 years) family history of lung cancer were significantly associated with lung cancer. The final multivariable model was combined with age-standardised incident data to estimate the absolute risk of developing lung cancer. In another study, Raji *et al.* evaluated the discrimination of the LLP risk model and demonstrated its predicted benefit for stratifying patients for CT screening by using data from three independent studies from Europe and North America (25). In this study, the LLP risk model performed better than smoking duration or family history alone in stratifying high-risk patients for lung cancer CT screening. The LLP risk model has been used to select high-risk individual in the United Kingdom Lung Screening (UKLS) (26). UKLS is a randomised controlled trial of LDCT for lung cancer screening, following the Wald single-screen design. In short, the UKLS randomised subjects based on their ≥5% risk of developing lung cancer in the next 5 years. Using this selection criterion shows that a screening programme will be more cost-effective if it is limited to the high-risk segment of the population i.e., individuals aged 60-75 years old. Using the LLP risk model with cut-off of ≥5% risk of developing lung cancer in the next five years indicates that excluding 55-59 years old will lead to missing a small number of lung cancer cases.

PLCO Cancer models

Tammemagi *et al.* produced lung cancer risk models using prospective data from 70,962 control subjects in the PLCO Cancer Screening Trial. Models were built for the general population (model 1) and a sub-cohort of ever-smokers (model 2) (22). Both models included age, socioeconomic status (education), body mass index, family history of lung cancer, chronic obstructive pulmonary disease, recent chest X-ray, smoking status (never, former, or current),

Table 1 Current methods utilised in identifying high risk cohort for lung cancer

Study characteristics	LLP risk model (21)	PLCO (22)	PLCO$_{M2012}$ (23)	Kovalchik et al. (24)
Population for modelling	579 lung cancer cases and 1,157 age- and sex-matched population-based controls	70,962 cancer-free population-based individuals. Aged 55-74 at recruitment	36,286 ever smoked population. Aged 55-74 at recruitment	26,604 NLST participants who underwent LDCT vs. 26,554 who underwent chest radiography
Risk factors included in model	Age, sex, smoking duration, prior diagnosis of pneumonia, asbestos exposure, prior diagnosis of malignant tumour, family history of lung cancer	Age, socioeconomic status (education), BMI, family history of lung cancer, COPD, recent chest X-ray and smoking status, pack-year, duration and time since quitting smoking (if applicable)	Age, race/ethnic group, education, BMI, COPD, personal history of cancer, family history of lung cancer and smoking-related factors (status, intensity, duration and quit time)	Age, BMI, family history of lung cancer, pack-years smoked, years since smoking cessation, diagnosis of emphysema
Discriminatory power in modelling population	0.70	0.86 (for all subjects); 0.81 (for ever-smokers)	0.80	–
Discriminatory power in populations for external validation	EUELC =0.67 (25); harvard dataset =0.76 (25); LLPC =0.82 (25)	0.84 (for all subjects); 0.78 (for ever-smokers only)	0.80	–
Strength	It has been used to select high-risk individuals in the UKLS (26)	Use of spline function in modelling; High discriminatory power	Major classical risk factor included; high discriminatory power	LDCT prevented: the greatest number of deaths among high risk and fewer numbers of deaths among low risk
Weakness	Using threshold of ≥5% risk of developing lung cancer in the next 5 years will exclude individuals aged 55-59 years	Healthy volunteer effect may limit external generalisation	Smoker population only. PLCO$_{M2012}$ risk ≥0.0151 is based on quintiles which are population specific. Evaluation of risk threshold was not based on cost-effectiveness	Use of quintiles for the predicted 5 years risk of death from lung cancer may not be generalisable to other cohorts

LLP, Liverpool Lung Project; EUELC, European Early Lung Cancer; LLPC, Liverpool Lung Project Population cohort; UKLS, United Kingdom Lung Screening; PLCO, Prostate, Lung, Colorectal and Ovarian; BMI, body mass index; COPD, chronic obstructive pulmonary disease; LDCT, low dose computed tomography.

pack-years smoked, and smoking duration. Model 2 also included smoking quit-time (time in years since ever-smokers permanently quit smoking). Logistic regression models were used to model significant risk factors. In addition to population-based design and large sample size, the PLCO models demonstrated high calibration and discrimination. However, these models utilised complicated modelling techniques that makes them difficult to apply in practice. In order to facilitate direct applicability to the NLST data, Tammemägi *et al*. modified and updated the smoker-only version of the PLCO model. The amended model, $PLCO_{M2012}$, included age, race/ethnic group, education, body mass index, COPD, personal history of cancer, family history of lung cancer and smoking status (current *vs*. former), intensity, duration and quit time as predictive variables. It used a simplified evaluation for non-linear effects, applied logistic regression modelling to calculate the probability of developing lung cancer over a period of 6 years (23). In another study, Tammemägi *et al*. evaluated the risk threshold for selecting individuals for screening and compared the efficiency of the threshold with the USPSTF criteria. In addition, they also determine whether never-smokers should be screened and compared lung cancer risks between smokers aged 55-64 and ≥65-80 years old (27). By analysing NLST data using the $PLCO_{M2012}$ model, the 65th percentile in PLCO smokers represents a risk of 0.0151. By using this threshold, mortality rates among NLST participants screened with CT were consistently lower than mortality rates in the chest X-ray arm: 255 people with a $PLCO_{M2012}$ risk ≥0.0151 would need to be screened to prevent one lung cancer deaths. Furthermore, using data collected from smokers in the screened arm of the PLCO trial to compare the efficiency of the $PLCO_{M2012}$ and USPSTF criteria for identifying screenees. Their result showed that 8.8% fewer people had a $PLCO_{M2012}$ risk ≥0.051 than the USPSTF criteria for screening, but 12.4% more lung cancers were identified. Thus, using $PLCO_{M2012}$ improved the sensitivity and specificity of the selection of individuals for lung cancer screening over the UPSTF criteria. However, a major limitation of the $PLCO_{M2012}$ risk ≥0.0151 threshold for selecting individuals for screening is that the evaluation was not based on cost-effectiveness.

Prediction model for lung cancer death using quintiles

Kovalchik *et al*. in an attempt to define high risk populations, investigated whether the benefits and harms of LDCT screening in the NLST vary according to lung cancer risk (24). In their study, they assessed the variation in efficacy, the number of false positive results, and the number of lung cancer deaths prevented among 26,604 participants in the NLST who underwent LDCT *vs*. 26,554 participants who underwent chest radiography, according to the quintile of a 5-year risk of lung cancer death. Lasso regression was used to select predictors of lung cancer death among previously identified risk factors for lung cancer. Selected risk factors for lung cancer death were age, body-mass index, family history of lung cancer, pack-years smoked, years since smoking cessation, and diagnosis of emphysema. Selected risk factors for the model with competitive causes of death were age, sex, race, body-mass index, pack-years smoked, years since smoking cessation, and diagnosis of emphysema. Cox proportional-hazards models of death from lung cancer and competitive causes of death were used to compute the absolute risk of death from lung cancer. The number of lung cancer deaths per 100,000 person-years that were prevented in the CT-screening group *vs*. radiography group increased according to risk quintile and there were significant decreasing trends in the number of participants with false positive results per screening-prevented lung cancer death. Their study concluded that screening with LDCT prevented the greatest number of deaths from lung cancer among participants who were at high risk and prevented very few deaths among those at lowest risk. Although Kovalchik *et al*. reported a new approach to identify high risk subjects based on a patient's risk of lung cancer death, because the primary benefit of LDCT screening is the prevention of lung cancer death, they argue that the prediction models for lung cancer incidence and mortality are likely to have similar discriminatory power. Their argument was further buttressed with the observation of similar trends in the number of CT-prevented lung cancer death across risk quintiles that were defined according to the risk of lung cancer death and the risk of lung cancer incidence (24). A major limitation of using quintiles in risk profiling is that their interpretation depends on their formation. In this study, participants were stratified into five quintiles for the predicted 5-year risk of death from lung cancer (with quintile 1 having the lowest risk and quintile 5 having the highest risk). Although the quintiles contain equal shares of the cohort for the predicted 5-year risk of death from lung cancer, this stratification may not be generalisable to any other cohort.

Cost-effectiveness of LDCT screening

The NLST trial has shown that screening with LDCT compared with chest radiology reduced lung cancer mortality. The potential effectiveness of lung cancer screening using LDCT might become a major economic driver for implementing lung cancer screening in national screening programmes. However, the relationship between the costs and benefits of lung cancer screening remains a controversial topic, especially in Europe, particularly in countries such as the UK where the health care system is government-funded. It would be politically problematic to offer publicly-funded medical interventions solely to heavy smokers, when non-/light smokers (although in a smaller proportion) may be at equally high risk due to other environmental and genetic factors and their interactions (6). Black *et al*. have examined the cost-effectiveness of screening with LDCT in the NLST (28). In their study, they estimated mean life-years, quality-adjusted life-years (QALYs), cost per person, and incremental cost-effectiveness ratios (ICERs) for three alternative strategies: screening with LDCT, screening with radiography and no screening. Screening compared with no screening cost an additional $1,631 per person (95% CI: 1,557-1,709) and provided an additional 0.0316 life-years per person (95% CI: 0.0154-0.0478) and 0.021 QALYs per person (95% CI: 0.0088-0.0314). The corresponding ICERS were $52,000 per life year gained (95% CI: 34,000-106,000) and $81,000 per QALY gained (95% CI: 52,000-186,000). In addition, they observed ICERs varied widely in subgroup and sensitivity analyses. Although Black *et al*. estimated that screening for lung cancer with LDCT would cost $81,000 per QALY gained, they suggested that modest changes in their assumptions would greatly alter this price projection. Limitations of the price projection in the study described above depend on (I) the variables that were considered in their sensitivity analyses and the method of implementing screening; (II) 150 NLST participants were excluded from their analysis which may have resulted in a small bias against screening with LDCT.

In addition, they assumed that screening with LDCT did not affect smoking status after the time of entry into the NLST and thus reclassified current smokers as former smokers and thus underestimated the cost-effectiveness of screening with LDCT. Furthermore, the result of their study will be difficult to implement in external data because they did not consider the effect of factors such as stringent selection criteria and high quality of care provided at NLST screening centres (28).

Future perspective

The question: 'who should be screened?' will continue to generate meaningful debate within the lung cancer research community. The current recommendation by USPSTF, the LLP risk model used to select individuals with at least a 5% risk of developing lung cancer in a 5-year period in UKLS, the $PLCO_{M2012}$ risk ≥0.0151 and the recently proposed risk model based on the use of quintile of the risk of lung cancer death at five years does not sufficiently answer this question. Advancement in high throughput methodologies and their application in molecular and genetic epidemiological studies have expanded the potential for biomarker-based risk prediction (8). Genome-wide association studies have identified inherited susceptibility patterns for lung cancer at different loci (29,30) and several methylation (31-33) and microRNA biomarkers (34-38) associated with lung cancer have been identified. Currently, most biomarkers are used mainly for diagnosis, but their value in risk prediction has not been widely explored (6). Cost-effective robust risk models incorporating biomarkers that will account for addition risk information not considered in the NLST/NELSON criteria could improve the selection process for lung cancer screening.

Conclusions

The high mortality associated with lung cancer is mainly due to the late presentation of the disease. Screening is an effective preventive strategy which aims to facilitate early detection and treatment in order to improve the high mortality rate. The selection criteria for screening in eligible participants in the current two largest randomised controlled trials (NLST and NELSON) and also in the recommendation of USPSTF were based on smoking history and age. The success of lung cancer screening will be dependent on successfully identifying a sufficiently high proportion of early-stage cases from the population. To achieve this goal, a cost-effective robust risk prediction algorithm incorporating elements of the current methods utilised in identifying high risk cohort for lung cancer as proposed by the LLP risk model, $PLCO_{M2012}$ risk model, the prediction model for lung cancer death using quintiles and models incorporating biomarkers is required.

Acknowledgements

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.

2. van Rens MT, Zanen P, Brutel de La Rivière A, et al. Survival in synchronous vs. single lung cancer: upstaging better reflects prognosis. Chest 2000;118:952-8.

3. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. CA Cancer J Clin 2010;60:277-300.

4. 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;365:395-409.

5. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. Lung Cancer 2009;64:34-40.

6. Field JK, Chen Y, Marcus MW, et al. The contribution of risk prediction models to early detection of lung cancer. J Surg Oncol 2013;108:304-11.

7. Cassidy A, Duffy SW, Myles JP, et al. Lung cancer risk prediction: a tool for early detection. Int J Cancer 2007;120:1-6.

8. Field JK. Lung cancer risk models come of age. Cancer Prev Res (Phila) 2008;1:226-8.

9. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of CT screening for lung cancer: a systematic review. JAMA 2012;307:2418-29.

10. Field JK, Smith RA, Aberle DR, et al. International Association for the Study of Lung Cancer Computed Tomography Screening Workshop 2011 report. J Thorac Oncol 2012;7:10-9.

11. Humphrey LL, Deffebach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. Ann Intern Med 2013;159:411-20.

12. Jett JR, Midthun DE. Screening for lung cancer: for patients at increased risk for lung cancer, it works. Ann Intern Med 2011;155:540-2.

13. Young RP, Hopkins RJ, Midthun DE. Computed tomographic screening for lung cancer. JAMA 2012;308:1320; author reply 1320-1.

14. Pinsky PF, Berg CD. Applying the National Lung Screening Trial eligibility criteria to the US population: what percent of the population and of incident lung cancers would be covered? J Med Screen 2012;19:154-6.

15. Centers for Disease Control and Prevention (CDC). Annual smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 1997-2001. MMWR Morb Mortal Wkly Rep 2005;54:625-8.

16. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). Int J Cancer 2007;120:868-74.

17. Alberg AJ, Brock MV, Samet JM. Epidemiology of lung cancer: looking to the future. J Clin Oncol 2005;23:3175-85.

18. Alberg AJ, Ford JG, Samet JM. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 2007;132:29S-55S.

19. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med 2011;32:605-44.

20. de Groot P, Munden RF. Lung cancer epidemiology, risk factors, and prevention. Radiol Clin North Am 2012;50:863-76.

21. 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;98:270-6.

22. Tammemagi CM, Pinsky PF, Caporaso NE, et al. Lung cancer risk prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial models and validation. J Natl Cancer Inst 2011;103:1058-68.

23. Tammemägi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013;368:728-36.

24. Kovalchik SA, Tammemagi M, Berg CD, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. N Engl J Med 2013;369:245-54.

25. Raji OY, Duffy SW, Agbaje OF, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. Ann Intern Med 2012;157:242-50.

26. McRonald FE, Yadegarfar G, Baldwin DR, et al. The UK Lung Screen (UKLS): demographic profile of first 88,897 approaches provides recommendations for population screening. Cancer Prev Res (Phila) 2014;7:362-71.

27. Tammemägi MC, Church TR, Hocking WG, et al. Evaluation of the lung cancer risks at which to screen ever-

S162

Marcus et al. Identifying high risk cohort in lung cancer screening

and never-smokers: screening rules applied to the PLCO and NLST cohorts. PLoS Med 2014;11:e1001764.

28. Black WC, Gareen IF, Soneji SS, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. N Engl J Med 2014;371:1793-802.

29. Easton DF, Eeles RA. Genome-wide association studies in cancer. Hum Mol Genet 2008;17:R109-15.

30. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.

31. Das PM, Singal R. DNA methylation and cancer. J Clin Oncol 2004;22:4632-42.

32. Baylin SB. DNA methylation and gene silencing in cancer. Nat Clin Pract Oncol 2005;2 Suppl 1:S4-11.

33. Nikolaidis G, Raji OY, Markopoulou S, et al. DNA methylation biomarkers offer improved diagnostic

efficiency in lung cancer. Cancer Res 2012;72:5692-701.

34. Lin PY, Yu SL, Yang PC. MicroRNA in lung cancer. Br J Cancer 2010;103:1144-8.

35. Shen J, Todd NW, Zhang H, et al. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. Lab Invest 2011;91:579-87.

36. Zheng D, Haddadin S, Wang Y, et al. Plasma microRNAs as novel biomarkers for early detection of lung cancer. Int J Clin Exp Pathol 2011;4:575-86.

37. Hennessey PT, Sanford T, Choudhary A, et al. Serum microRNA biomarkers for detection of non-small cell lung cancer. PLoS One 2012;7:e32307.

38. Bediaga NG, Davies MP, Acha-Sagredo A, et al. A microRNA-based prediction algorithm for diagnosis of non-small lung cell carcinoma in minimal biopsy material. Br J Cancer 2013;109:2404-11.