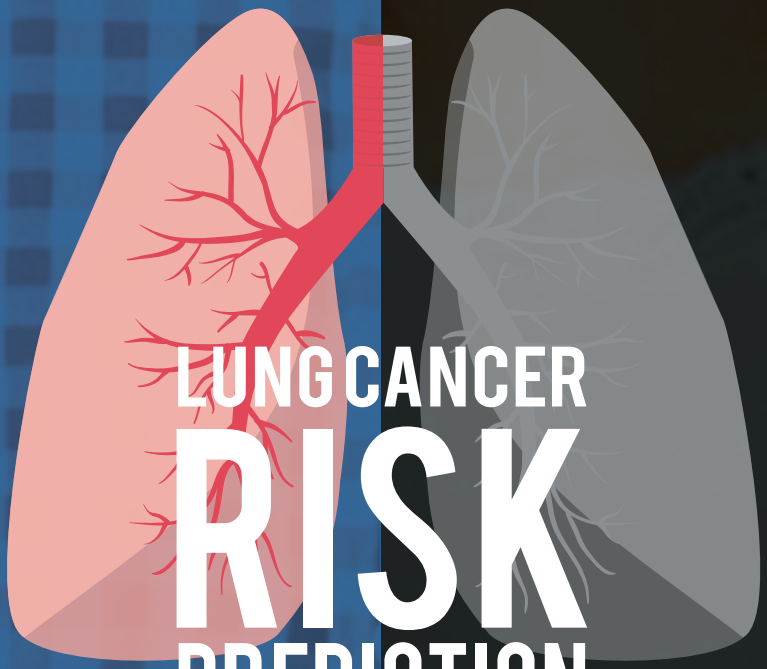
MOH/P/PAK/308-15(TR)

HEALTH TECHNOLOGY ASSESSMENT REPORT



LUNG CANCER RISK PREDICTION MODEL FOR NATIONAL HEALTH RISK ASSESSMENT MODULE

KEMENTERIAN KESIHATAN MALAYSIA

MaHTAS

Malaysia Health Technology Assessment Section

MEDICAL DEVELOPMENT DIVISION
MINISTRY OF HEALTH MALAYSIA

DISCLAIMER

This Health Technology Assessment has been developed from analysis, interpretation and synthesis of scientific research and/or technology assessments conducted by other organizations. It also incorporates, where available, Malaysian data, and information provided by experts to the Ministry of Health Malaysia. While an effort has been made to do so, this document may not fully reflect all scientific research available. Additionally, other relevant scientific findings may have been reported since completion of the review. This report is subjected for external reviewers.

Please contact: htamalaysia@moh.gov.my if you would like further information.

Health Technology Assessment Section (MaHTAS),
Medical Development Division
Ministry of Health Malaysia
Level 4, Block E1, Precinct 1
Government Office Complex
62590 Putrajaya

Tel: 603 88831229

Fax: 603 8883 1230

Available at the following website: http://www.moh.gov.my

ISBN : 978-967-0769-57-8

Authors

Madam Maharita Abd Rahman
Pharmacist
Senior Assistant Director
Health Technology Assessment Section (MaHTAS)
Medical Development Division
Ministry of Health Malaysia

Dr. Izzuna Mudla Mohamed Ghazali
Public Health Physician
Senior Principal Assistant Director
Health Technology Assessment Section (MaHTAS)
Medical Development Division
Ministry of Health Malaysia

Information Specialists

Madam Asmirah Redzuan
Pharmacist
Senior Assistant Director
Health Technology Assessment Section (MaHTAS)
Medical Development Division
Ministry of Health Malaysia

Madam Norharlina Che Zakaria
Nurse/Information Specialist
Health Technology Assessment Section (MaHTAS)
Medical Development Division
Ministry of Health Malaysia

EXPERT COMMITTEE

Dato' Dr. Abdul Razak Muttalif
Senior Consultant Physician (Respiratory)
Institute of Respiratory Medicine
Kuala Lumpur

Dr. Kamaliah Mohamad Noh
Public Health Physician and Deputy Director
Primary Healthcare Section
Family Health Development Division
Ministry of Health Malaysia

Dr. Nik Harlina Roza Binti Nik Kazim
Family Medicine Specialist
Klinik Kesihatan Bachok
Kelantan

Dr. Nurhayati Mohd Marzuki
Consultant Physician (Respiratory)
Institute of Respiratory Medicine
Kuala Lumpur

Dr. Azura Deniel
Oncologist
Radiotherapy & Oncology Department
Hospital Kuala Lumpur

Dr. Nor Saleha binti Ibrahim Tamin
Public Health Physician
Senior Principal Assistant Director
Disease Control Division
Ministry of Health Malaysia

En. Azman Mohammed
Senior Principal Assistant Director
Health Education Division
Ministry of Health Malaysia

Pn. Norhaslida Mohd Nor
Principal Assistant Director
Health Education Division
Ministry of Health

Datin Dr. Rugayah binti Bakri
Public Health Physician and Deputy Director
Health Technology Assessment Section (MaHTAS),
Medical Development Division
Ministry of Health Malaysia

Dr. Junainah binti Sabirin
Public Health Physician and
Head of Health Technology Policy Unit
Health Technology Assessment Section (MaHTAS)
Medical Development Division,
Ministry of Health

EXTERNAL REVIEWERS

Prof. Dato' Dr. Fuad Ismail
Clinical Associate Professor and
Senior Consultant Oncologist
Department of Radiology & Oncology
Hospital Chancellor Tuanku Muhriz (HCTM)

Datu Dr. Andrew Kiyu
Epidemiologist
Sarawak Health Department
Kuching, Sarawak

Dr. Jamalul Azizi Abdul Rahman
Head of Respiratory Medical Service, Ministry of Health Malaysia
Senior Consultant Physician (Respiratory)
Medical Department
Hospital Serdang

ACKNOWLEDGEMENT

The authors this Health Technology Assessment Report would like to express their gratitude and appreciation to the following for their contribution and assistance:

- Health Technology Assessment and Clinical Practice Guidelines Council, Malaysia
- Technical Advisory Committee for Health Technology Assessment, Malaysia
- Madam Sin Lian Thye, and Miss Loong Ah Moi from MaHTAS for their contribution in retrieval of the evidence

DISCLOSURE

The authors of this report have no competing interest in this subject and the preparation of this report is totally funded by the Ministry of Health, Malaysia.

EXECUTIVE SUMMARY

Background

Lung cancer is the most common cancer worldwide, accounting for 1.3 million deaths annually. According to the National Cancer Registry, 1,865 cases of lung cancer were diagnosed and registered in Peninsular Malaysia in 2007. The age standardised rate (ASR) for male was 14.7 per 100,000 and 5.6 per 100,000 for female. The incidence was more than two-fold higher among males compared to females. The incidence increased with age and in 2007 the peak of age-specific incidence rate was among the 70-75 age groups. Most of the lung cancers were detected late where 60% of the cases were detected at stage IV. On the other hand, the percentage of lung cancer detected at stage I and II was only 12%.

The United States, National Cancer Institute reported that the lung cancer five-year survival rate (16.3%) is lower than many other leading cancer sites, such as the colon (62.2%), breast (90.0%) and prostate (99.9%). The five-year survival rate was 52.6 percent for patients with localised disease (within lungs). However, only 15 percent of lung cancer cases were diagnosed at an early stage. Patients who had distant metastases (spread to other organs), the five-year survival rate was only 3.5%.

The National Lung Screening Trial (NLST) suggested that screening programme is appropriate for high risk population. One of the early screening methods is through health risk assessment (HRA) tool. This health risk assessment or also known as health risk appraisal, health & well-being assessment or risk prediction model, is a confidential online questionnaire that asks about risk factors for lung cancer. The HRA incorporates four common key elements; an extended questionnaire, a risk calculation or score, and some form of feedback i.e. face-to-face with a health advisor or an automatic online report.

In Malaysia, currently HRA modules are available for obesity, mental health, diabetes, heart problems, physical activity and smoking habit. Currently there is no risk assessment prediction model for early detection of lung cancer. This review was requested by a Senior Principal Assistant Director, Health Education Division, Ministry of Health (MOH) to review the evidence on lung cancer risk assessment models as a tool in enhancing early detection and diagnosis of lung cancer, towards facilitating implementation of an affordable and effective cancer control in the country.

Technical features

Cancer risk assessment models / health risk assessment tools are statistical models developed for cancer risk prediction and can be divided into two broad categories:

i. To predict the probability of being diagnosed with a particular cancer, and
ii. To predict the likelihood of carrying a gene mutation that predisposes to a particular cancer or set of cancers.

According to Memorial Sloan Kettering Cancer Centre, health risk assessment tools are useful in clinical decision making as they helps clinicians and patients to determine whether the screening will be beneficial.

Policy Questions

i. In the Ministry of Health, should a health risk assessment (HRA) module for lung cancer be introduced as one of the strategies in the prevention of lung cancer under the Malaysian National Cancer Control Programme?
ii. If an HRA module (cancer risk prediction model) for lung cancer is to be introduced, which risk prediction model for lung cancer should be adopted / adapted in Malaysia?

Objectives

i. To assess the effectiveness in terms of predictive accuracy of lung cancer risk assessment/prediction models
ii. To assess the safety, organizational, ethical issues and economic implications related to risk assessment/prediction models for lung cancer

Methods

Studies were identified by searching electronic databases. The following databases were searched through the Ovid interface: MEDLINE(R) In-process and other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to present. Parallel searches were run in PubMed and EMBASE. Appendix 3 shows the detailed search strategies. No limits were applied to the search. The last search was run on 2 July 2015. Additional articles were identified from reviewing the references of retrieved articles.

Results

A total of 2,431 titles were identified through Ovid interface, Pubmed and references of retrieved articles. A total of 55 abstracts were screened using the inclusion and exclusion criteria. After critical appraisal, only six full text articles were included in this review.

Of these, five articles were related to effectiveness (predictive accuracy) of different risk prediction models for lung cancer. The other article was a validation study related to the risk prediction model assessed. The six studies included comprised of three case-control studies, one cohort study, one randomised controlled trial and two non-randomised controlled trials. No evidence on safety and cost-effectiveness / cost-cost utility analysis was retrieved.

Effectiveness (predictive accuracy)

Five risk prediction models namely: Liverpool Lung Project (LLP) risk prediction model, Korean risk prediction model, Bach risk prediction model, Spitz risk prediction model and COSMOS risk prediction model were assessed.

Performance of prediction model is commonly measured by means of calibration and discrimination. A well-fitted model has Expected/Observed (E/O) ratio close to 1, a lower number underestimates the condition's incidence and a higher number overestimates the incidences. The concordance (c)-statistics measure model discrimination performance which is similar to area under the receiver operating characteristic curve (ROC). A c-statistics of 1.0 indicates perfect discrimination and 0.5 equivalents to no discrimination between people who develop the condition and those who do not.

Performance of Risk Prediction Models

Liverpool Lung Project (LLP) Risk Prediction Model (United Kingdom)
LLP risk prediction model is an individualized risk prediction model for lung cancer. The data used were based on data from a case-control study of lung cancer in Liverpool; the Liverpool Lung Project (LLP). The model estimated the absolute risk of lung cancer for a given individual and included variables that are readily available to primary care clinicians to facilitate the referral of high risk individuals. The area under the curve (AUC) was 0.71 and from the 10-fold cross validation of the LLP risk prediction model produced AUC statistics of 0.70 which indicated good discrimination between cases and controls.

Validation study using data from three case-control studies, showed that the LLP risk prediction model had modest discrimination in the European Early Lung Cancer (EUELC) data set (AUC, 0.67 [95% CI 0.64 to 0.69]) and good discrimination in both Harvard (AUC, 0.76 [95% CI 0.75 to 0.78]) and Liverpool Lung Project Population-Based Cohort (LLPC) (AUC, 0.82 [95 % CI 0.80 to 0.85]) data set. The AUC for smoking duration which was the strongest of the risk factors was 0.63, 0.74, and 0.72 in the EUELC, Harvard and LLPC data sets respectively. Besides that, the LLP risk prediction model had moderate overall calibration and improved accuracy at higher values of predicted risks.

Korean Risk Prediction Model (Korea)
Korean Risk Prediction Model was an individualized risk prediction model for lung cancer in Korean men using population-based cohort data. The model was configured to estimate the absolute risk that an individual will have lung cancer in eight years as well as to identify the significant risk factors for lung cancer. C-statistic for Korean risk prediction model showed excellent discrimination of 0.864, 95% CI 0.860 to 0.868. If considering only age and smoking variables, the prediction model also showed excellent discrimination (C-statistic of 0.861 (95% CI: 0.857,0.865). The performance of the risk prediction model also showed excellent discrimination with C-statistic of 0.87 (95% CI: 0.867,0.876) when using external validation dataset.

Spitz Risk Prediction Models (United States)
Spitz risk prediction models are multivariable models that are constructed separately for never smokers, former smokers and current smokers, incorporating into each model variables that exhibit statistically significant

main effects. Each model was well calibrated throughout the entire range of probabilities, as indicated by non-statistically significant Hosmer-Lemeshow goodness-of-fit test statistics (0.777 for never smokers, 0.712 for former smokers and 0.688 for current smokers). Meanwhile by looking at AUC statistics obtained from the validation sets, the AUC was low for never smokers and current smokers compared to former smokers. The results of concordance statistics indicated that the models performed reasonably well to discriminate between cases and controls

Bach Risk Prediction Model (United States)
Bach risk prediction model was developed and validated for individual lung cancer risk that can be applied in both clinical and research settings. The authors examined the predicted 10-year lung cancer risk among subjects enrolled in an ongoing CT screening program. The authors assessed the extent of variation in risk among a cohort of individuals who met typical eligibility criteria for cancer prevention studies to determine whether the risk of lung cancer varies and to ascertain the usefulness of the model as an adjunct to clinical research. The Bach risk prediction model was validated at six study sites; the observed rates of lung cancer matched those that had been predicted by the corresponding model derived from the five included sites. The cross-validated concordance index was 0.72 and the cross-validated calibration plot by risk deciles was consistent with excellent calibration. The risk prediction model only had a cross–validated concordance index of 0.66.

COSMOS (Continuous Observation of Smoking Subjects) Risk Prediction Model (Italy)
COSMOS risk prediction model was based on epidemiologic and clinical risk factors to estimate the probability of individuals in a high-risk population being diagnosed with lung cancer. This model might be useful to stratify individuals and select those at high risk for inclusion in screening programs. Another aim was to develop a second model based on baseline CT findings in a screened population, combined with epidemiologic and clinical risk factors, to stratify individuals according to the probability of being diagnosed with lung cancer at repeat screening scans. The second model was proposed for use in large scale screening programs to select lower risk patients in whom the interval between screening CTs can be lengthened and at the same time to identify those at higher risk of lung cancer in whom surveillance intensity might be increased or who might benefit from prevention intervention studies. At the end, 162 lung cancers were detected in 18,095 person-years of observation from baseline, giving a lung cancer detection rate of 0.90 per 100 years. The detection rates (per 100 years) were slightly higher in men (0.95) compared to women (0.78) and in current smokers (0.92) than former smokers (0.79). However, both differences were not significant. No validation study was conducted for the model.

Safety
There was no retrievable evidence on safety issue of risk prediction model for lung cancer.

Cost Implications
There was no retrieval evidence on risk prediction model or HRA for lung cancer, however, the potential direct cost implicated on the designing, developing, and testing is about RM75,000 to RM100,000.

Organizational
Any risk prediction models or HRA modules require computer literate user / patient and internet access. Statisticians with management capability, computer analysis and risk modelling skills are also required to manage the dataset and undertake statistical analyses. This plan will also involve physicians, nutritionists, health counsellors (psychologists) and physiologists.

Development of risk prediction models require several considerations including research issues, gaps, any priorities, and alternatives needed to advance the field of cancer risk prediction and make specific recommendations for implementations. The model also needs to be continually calibrated and revalidated.

Any uncertainties associated with risk estimates should be addressed and informed particularly when clinical decision has serious consequences especially for those who are at risk. Because of that, the whole plan of the module should include counseling, further diagnosis with physician as well as further management and treatment.

Conclusion
There was fair level of retrievable evidence for risk prediction models for lung cancer. There were five models identified for predicting lung cancer risk. The LLP risk prediction model and Korean risk prediction model were the best models for predicting lung cancer. LLP risk prediction model appeared to have good to excellent discrimination with area under curve (AUC) of 0.71. The LLP risk prediction model also has good ability to distinguish persons who will or will not develop lung cancer by using the predicted 5-year absolute risk. The Korean model is the only model that used Asia population (Korean) and has an excellent discrimination with c-statistic of 0.87.

For other risk prediction models, although they were well calibrated and validated, they appeared to have modest ability to discriminate between subjects who will be having lung cancer and those who will not, in the study population.

There was no retrievable evidence on safety related to risk prediction model or health risk assessment module for the detection of lung cancer in the population. None of the module mentioned any health problem including psychological impact among subjects involved.

There was no retrievable evidence on economic evaluation of risk prediction model or health risk assessment module for lung cancer, or cost implication involved in developing a new health risk assessment retrieved. The cost involved in validating a model by a prospective cohort validation study could be very costly depending on the number of study participants and years of

follow up. However, the potential direct cost implicated to the designing, developing, testing and commissioning of available one risk prediction model of lung cancer is about RM75,000 to RM100,000.

Risk prediction model or health risk assessment module for lung cancer needs continual validation to give meaningful risk estimate and to ensure its capability in the setting it will be used. The complexity to develop and validate the risk prediction model or HRA module is reflected in the necessary local data required. Dedicated research expertise to create a robust risk prediction model with consistent performance is very important.

Recommendation

Health risk assessment (HRA) module / risk prediction model for lung cancer such as Liverpool Lung Project (LLP) risk prediction model and the Korean risk prediction model need to undergo further validation until a well-fitted model with better predictive ability tailored to Malaysian population is established. The model needs continual validation to determine the consistency of its performance. Besides that, the module should only be introduced as part of comprehensive strategies for lung cancer whereby screening, treatment and rehabilitation is available.

Abbreviations

ASR	:	Age Standardised Rate
AUC	:	Area Under the Curve
BMI	:	Body Mass Index
CARET	:	Carotene and Retinol Efficacy Trial
CDC	:	Centers for Disease Control and Prevention
CI	:	Confidence Interval
COSMOS	:	Continuous Observation of Smoking Subjects
CT	:	Cancer Tomography
E/O	:	Expected/Observed
ETS	:	Environmental Tobacco Smoke
EUELC	:	European Early Lung Cancer
FEF	:	Forced Expiratory Flow
FEV	:	Forced Expiratory Volume
FVC	:	Forced Vital Capacity
HR	:	Hazard Ratio
HRA	:	Health Risk Assessment
LLCC	:	LLP Case-Control
LLP	:	Liverpool Lung Project
LLPC	:	LLP Population-Based Cohort
MOH	:	Ministry of Health
NLST	:	National Lung Screening Trial
OR	:	Odds Ratio
RR	:	Risk Ratio
SEER	:	Surveillance, Epidemiology and End Result Program
US	:	United States
WHO	:	World Health Organization

TABLE OF CONTENTS

LUNG CANCER RISK PREDICTION MODEL FOR NATIONAL HEALTH RISK ASSESSMENT MODULE

1.0 BACKGROUND

Lung cancer is the most common cancer worldwide, accounting for 1.3 million deaths annually. The United State (US) National Institutes of Health estimated that cancers cost the United States an overall $264 billion in 2010. It was estimated that approximately $10.3 billion was spent in the United States on lung cancer treatment alone.[1]

The number of new cases of lung cancer worldwide was estimated to be 1.8 million in 2012 (12.9% of the total 14.1 million new cancer cases, 58% of which occurred in the less developed regions. The disease remains the most common cancer in men worldwide (1.2 million, 16.7% of the total) with the highest estimated age-standardised incidence rates in Central and Eastern Europe (53.5 per 100,000) and Eastern Asia (50.4 per 100,000).[2] The American Cancer Society stated in Cancer Facts and Figure 2014, that lung cancer accounts for more deaths than any other cancer in both men and women and was responsible for nearly one-in-five deaths (1.59 million deaths, 19.4% of the total).[3]

According to the Malaysian National Cancer Registry, 1,865 cases of lung cancer were diagnosed and registered in 2007. The age standardised rate (ASR) for male was 14.7 per 100,000 and 5.6 per 100,000 for female. The incidence was more than two-fold higher among males compared to females. Based on ethnicity, Chinese were found to have higher incidence rate compared to Malay and Indian. The incidence increased with age and in 2007 the peak age-specific incidence rate was among the 70-75 age groups. Most of the lung cancers were detected late where 60% of the cases were detected at stage IV; the percentage of lung cancer detected at stage I and II was only 12%.[4]

The United States National Cancer Institute reported that the lung cancer five-year survival rate (16.3%) was lower than many other leading cancer sites, such as the colon (62.2%), breast (90.0%) and prostate (99.9%). The five-year survival rate was 52.6 percent for patients with localised disease (within the lungs). However, only 15 percent of lung cancer cases were diagnosed at an early stage. Patients who had distant metastases (spread to other organs), the five year-survival rate was only 3.5%.[5]

The National Lung Screening Trial (NLST) suggested that screening programme is appropriate for high risk population. One of the early screening methods is through health risk assessment (HRA) tools.[6] Cancer researchers, clinicians, and the general public are devoting increased attention to the statistical models (HRA) designed to predict the occurrence of cancer.[7] This health risk assessment or also known as health risk appraisal, health & well being assessment or risk prediction model, is a confidential online questionnaire that asks about risk factors for lung cancer.[6] The risk prediction model is one of the most widely used screening tools in the field of health promotion and is often the first step in multi-component health promotions

programme. The questionnaires, allows individuals to evaluate their health risks and quality of life. The HRA incorporates four common key elements; an extended questionnaire, a risk calculation or score, and some form of feedback such as face-to-face with a health advisor or an automatic online report.[7] According to United States Centers for Disease Control and Prevention (CDC), health risk assessment or health risk appraisal is a systematic approach to collect information from individuals that identifies risk factors, provides individualized feedback and links the person with at least one intervention to promote health, sustain function and/or prevent disease. A typical HRA instrument obtains information on demographic characteristics, lifestyle, personal medical history, and family medical history.[8]

In recent years, cancer risk prediction models published in the scientific literature have included refinements of older breast cancers risk models and new models that estimate the risks of melanoma, lung, prostate, colorectal, breast and other cancers. Many of the new models combine clinical and epidemiologic risk factors with new biologic and genetic data to more accurately assess cancer risk. Other than being used to identify individuals at high risk of cancer who may benefit from targeted screening or other interventions, the risk prediction model is also used to develop benefit-risk indices, to estimate the population burden, the cost of cancer and the impact of specific intervention. Risk prediction models are commonly used in clinical decision-making to help physicians and patients determine appropriate screening regimens and or interventions.[7]

In Malaysia, currently the HRA modules are available for obesity, mental health, diabetes, heart problems, physical activity and smoking habit. Currently there is no risk prediction model used nationally for lung cancer. This review was requested by a Senior Principal Assistant Director, Health Education Division, Ministry of Health (MOH) to review the evidence on lung cancer risk prediction model to be included in the HRA module in enhancing early detection and diagnosis of lung cancer, towards facilitating implementation of an affordable and effective cancer control in the country.

2.0 TECHNICAL FEATURES

Cancer risk prediction models are statistical models developed for cancer risk prediction and can be divided into two broad categories:

i. To predict the probability of being diagnosed with a particular cancer, and
ii. To predict the likelihood of carrying a gene mutation that predisposes to a particular cancer or set of cancers.

According to Memorial Sloan Kettering Cancer Centre, health risk assessment tools are useful in clinical decision making as they help clinicians and patients to determine whether the screening will be beneficial.[9]

Risk prediction for lung cancer is the process of identifying characteristics of an individual that are relevant to their risk, and combining those

characteristics into an estimate of probability of developing disease, either over a discrete period of time or over a lifetime. The accuracy of a model for risk prediction depends on the identification of risk factors, but also on how these factors operate in the presence or absence of other factors, how accurately they can be measured and the appropriateness of the statistical model used. [9]

Five lung cancer risk prediction models are included in this report. They are:

i) Liverpool Lung Project (LLP) Risk Prediction Model (United Kingdom)
LLP was developed from the LLP case– control (LLCC) study, provides a single unified model for smokers (current and former) and non-smokers

ii) Korean Risk Prediction Model (Korea)
Involve Korean population which may represent Asian population

iii) Bach Risk Prediction Model (US)
Predict risk of lung cancer in smokers only

iv) Spitz Risk Prediction Model (US)
Require three separate models to predict risk in current, former or non-smokers

v) COSMOS Risk Prediction Model (Italy)

3.0 POLICY QUESTIONS

3.1 In the Ministry of Health, should a health risk assessment (HRA) module for lung cancer be introduced as one of the strategies in the prevention of lung cancer under the Malaysia National Cancer Control Programme?

3.2 If HRA module (cancer risk prediction model) for lung cancer is to be introduced, which risk prediction model for lung cancer should be adopted / adapted in Malaysia?

4.0 OBJECTIVES

4.1 To assess the effectiveness in terms of predictive accuracy of lung cancer risk assessment/prediction models

4.2 To assess the safety, organizational, ethical issues and economic implications related to risk assessment/prediction models for lung cancer

Research Questions

i. What is the predictive accuracies of available risk prediction models for lung cancer in terms of detection rate and stage of disease at diagnosis?

ii. What are the strengths and weaknesses of each available model?

iii. What is the economic impact, ethical, legal and organizational issues of the health risk assessment modules?

5.0 METHODS

5.1. Literature search strategy

Studies were identified by searching electronic databases. The following databases were searched through the Ovid interface: MEDLINE(R) In-process and other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to present. Parallel searches were run in PubMed and EMBASE. Appendix 3 shows the detailed search strategies. No limits were applied to the search. The last search was run on 02 July 2015. Additional studies were identified from reviewing the references of retrieved articles.

5.2. Study Selection

Based on the policy question the following inclusion and exclusion criteria were used:-

5.3. Inclusion criteria

- Population:
  - Adult aged 18 years old and more

- Intervention:
  - Health Risk Assessment Tool / Risk Prediction Model / Health & Well-being Assessment

- Comparators:
  - No HRA module or risk prediction models

- Outcome, one or more of the following outcome measures will be assessed:
  - Effectiveness/benefit of the lung cancer HRA module related to patient outcome as measured by detection rate and stage of diagnosis
  - Performance of available lung cancer risk prediction models in terms of its predictive accuracy - sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), calibration as measured by expected/observed ratio, discrimination as measured by AUC or c-statistic
  - Organizational (operational, training, resources), ethical, legal and economic implication

- Study design:
  - No restriction of study type. HTA reports, systematic review, randomised controlled trial (RCT), non-randomised controlled trials,

diagnostic accuracy studies, cross-sectional, cohort, case-control and economic evaluation studies

- Full text articles published in English.

5.4. Exclusion criteria:-

- Study design: Animal study, narrative review, experimental study
- Non-English full text article
- Prognostic model

Based on the above inclusion and exclusion criteria, study selection was carried out independently by two reviewers. The titles and abstracts of all studies were assessed for the above eligibility criteria. If it was absolutely clear from the title and / or abstract that the study was not relevant, it was excluded. If it was unclear from the title and / or the abstract, the full text article was retrieved. Two reviewers assessed the content of the full text articles. Disagreement was resolved by discussion.

5.5. Quality assessment strategy

The methodological quality of all the relevant full text articles retrieved was assessed using the Critical Appraisal Skills Programme (CASP) tool by two reviewers. For cohort study, the criteria assessed were selection of the cohort, accurate measurement of exposure and outcome, confounding factors, follow-up adequacy and length. The CASP checklist is as in Appendix 4. All full text articles were graded based on guidelines from the U.S./Canadian Preventive Services Task Force (Appendix 1).

5.6. Data extraction strategy

Data were extracted from the included studies by a reviewer using a pre-designed data extraction form (evidence table as shown in Appendix 6) and checked by another reviewer. Disagreements were resolved by discussion. Details on: (1) methods including study design, (2) study population characteristics including gender, age, medical history, cancer history (3) type of intervention, (4) comparators, (5) type of outcome measures including: a) absolute risk ratio, b) quality of life, c) area under the curve or c-statistic, d) validation, e) economic evaluation, and f) organizational issues were extracted. The extracted data were presented and discussed with the expert committee. Other information on author, journal and publication year, and study objectives were also extracted.

5.7. Methods of data synthesis

Data on the safety, efficacy and cost implication of risk prediction model of lung cancer were presented in tabulated format with narrative summaries. Meta-analysis was not conducted for this review.

6.0 SEARCH RESULTS

A total of 2,431 titles were identified through the Ovid interface: MEDLINE (R) In-process and other Non-Indexed Citations and Ovid MEDLINE (R) 1948 to present, Embase 1988 to 2014 Week 46 and PubMed. Ten articles were identified from references of retrieved articles.

After removal of 74 duplicates, two reviewers screened 2,367 titles and 2,312 titles were excluded. A potential of 55 relevant titles and abstracts were screened and the full articles were retrieved. The 55 full text articles were screened using the inclusion and exclusion criteria. Of these, 49 full texts were found to be irrelevant. Only six full text articles were included in the review. The excluded studies were listed in Appendix 3.

Five out of six included articles were related to effectiveness of different risk prediction models of lung cancer; while one article was a validation study related to one of the risk prediction models. The six studies included comprised of three case-control studies, one cohort study, one randomised controlled trial and one non-randomised controlled trial. No evidence on safety and cost-effectiveness / cost-utility analysis was retrieved.

The included studies consisted of five different risk prediction models namely Liverpool Lung Cancer Project (LLP) Risk Prediction Model, Korean Risk Prediction Model, Bach Risk Prediction Model, Spitz Risk Prediction Model and COSMOS Risk Prediction Model – and a validation of LLP Risk Prediction Model.

Forty-nine articles were excluded due to the topics being irrelevant, unrelated health risk assessment modules and risk prediction models and irrelevant study design. There was no HTA report on risk assessment models / risk prediction models / health risk assessment modules for lung cancer identified.

Figure 1 shared the flow of information in this review.

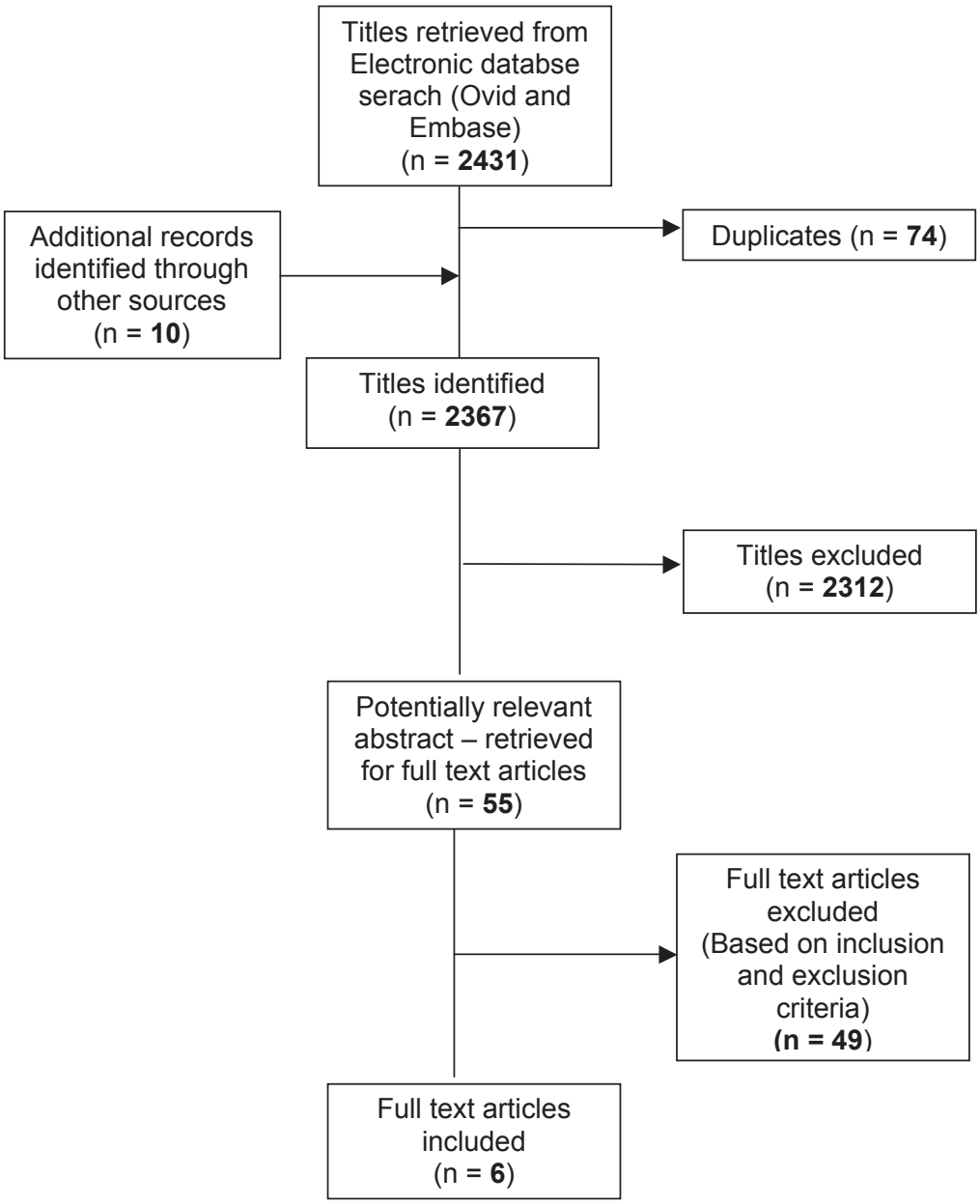


Figure 1: Flow Chart of Study Selection

7.0 RESULTS

7.1 Types of Risk Prediction Models for Lung Cancer

7.1.1 Liverpool Lung Project Risk Prediction Model

Cassidy et al. developed an individual risk prediction model for lung cancer. The data used were based on data from a case-control study of lung cancer in Liverpool; the Liverpool Lung Project (LLP). The aim of the study was to provide a model that would estimate the absolute risk of lung cancer for a given individual and to include in the final model only variables that are readily available to primary care clinicians to facilitate the referral of high risk individuals.[11, level II-2]

Study Population

The LLP case-control study included incident cases of histologically or cytologically confirmed lung cancer between the ages of 20 and 80 years old. The participants were eligible if they were residents within the Liverpool area. Both cases and controls were ineligible for the study if they had a previous cancer within 5-years of interview date. Two population controls per case were selected from registers of general practitioners in Liverpool and matched to lung cancer cases by year of birth (±2 years) and gender. All the participants consisted of 579 incident cases of lung cancer and 1,157 population controls.

Risk Predictors

The risk predictors were collected through standardised lifestyle questionnaires which included:

i. Demographic characteristics and socioeconomics data

ii. Medical history

- Prior non-malignant lung disease (asthma, bronchitis, emphysema, pneumonia or tuberculosis) at any age at least two years before diagnosis of lung cancer (or date of interview for controls)
- Primary set of malignant tumour if they had ever been diagnosed
- Age at time of diagnosis of malignant tumour

iii. Family history of cancer

- First degree relatives (parents, brothers and sisters and biological children)
- Age at diagnosis
- Site of cancer
- Relationship to participants

iv. History of tobacco consumption

- Smoking status
- Inhalation
- Type of cigarette smoked
- Number of cigarettes smoked per day
- Age at start and end

An ever smoker was defined as someone who had smoked at least 100 cigarettes in their lifetime and a current smoker was defined as a participant who reported smoking two years prior to the date of the interview. During the interview, all periods of consumption were defined and counted towards total exposure.

v. Lifetime occupational history

- Asbestos exposure
- Frequency of asbestos exposure:
  - Percentage of working time exposed (categorised as 1 – 5%, 5 – 30% or > 30%)
- Intensity of asbestos exposure:
  - Categorised as low, medium or high

Multivariable Analysis / Discrimination

Overall response rate was 58.3% for cases and 61.5% for controls. Caucasians represented approximately 99% of the cases and the controls.

Final multivariable analysis found that the independent risk factors for lung cancer were family history with lung cancer ($P = 0.01$), individual with history of pneumonia ($P = 0.02$) and other types of cancers ($P = 0.005$), asbestos exposure ($P < 0.001$) and duration of smoking ($P < 0.001$). Those risk factors showed significant increase in risks of lung cancer and were included in the final model. Table 1 showed the details of LLP final multivariable risk model with unadjusted and adjusted odds ratios.

Table 1: LLP Multivariable Risk Model Unadjusted and adjusted Odds Ratios and 95% CIs corresponding to the Model Coefficient[11]

Risk factor/category	Odds ratio[a]	(95% CI)	Odds ratio[b]	(95% CI)	*P*-value	Model coefficient
Smoking duration					<0.001	
Never	1.00	Reference	1.00	Reference		0.000
1–20 years	2.48	(1.47–4.17)	2.16	(1.21–3.85)		0.769
21–40 years	5.81	(3.68–9.18)	4.27	(2.62–6.94)		1.452
41–60 years	19.24	(12.07–30.67)	12.27	(7.41–20.30)		2.507
> 60 years	41.74	(17.86–97.56)	15.25	(5.71–40.65)		2.724
Prior diagnosis of pneumonia					0.002	
No	1.00	Reference	1.00	Reference		0.000
Yes	1.62	(1.21–2.17)	1.83	(1.26–2.64)		0.602
Occupational exposure to asbestos					<0.001	
No	1.00	Reference	1.00	Reference		0.000
Yes	1.94	(1.46–2.59)	1.89	(1.35–2.62)		0.634
Prior diagnosis of malignant tumour					0.005	
No	1.00	Reference	1.00	Reference		0.000
Yes	2.55	(1.76–3.71)	1.96	(1.22–3.14)		0.675
Family history of lung cancer					0.01	
No	1.00	Reference	1.00	Reference		0.000
Early-onset (<60 years)	1.54	(1.03–2.29)	2.02	(1.18–3.45)		0.703
Late-onset (≥60 years)	1.08	(0.80–1.46)	1.18	(0.79–1.76)		0.168

[a]Odds ratios derived from univariate conditional logistic regression. [b]Odds ratios derived from multivariate conditional logistic regression.

The authors also calculated the absolute risk for lung cancer within five years period. For example, the absolute risks for a 77 years-old patient with a history of more than 45 years of smoking, with asbestos exposure was 28.68% compared with that of a 77 years-old non-smoker with asbestos exposure which was only 3.17%. The results of other absolute risks were shown in Table 2:

Table 2: Projected 5-Year Absolute Risks and 95% CIs for Combination of Risk Factors[11]

				Prior diagnosis of				
Gender	Age	Smoking duration (Years)	Family history of lung cancer[a]	Malignancy	Pneumonia	Asbestos exposure	Absolute risk (%)	(95% CI)
Female	65	37	Late-Onset	—	Yes	—	2.37	(1.14–4.86)
	68	26	—	—	—	—	1.50	(0.91–2.46)
	69	50	—	—	—	—	4.60	(2.76–7.54)
Male	64	42	Late-Onset	Yes	—	—	9.53	(4.52–18.97)
	66	53	—	—	—	Yes	8.75	(4.89–15.18)
	66	48	—	—	Yes	Yes	14.91	(7.70–26.89)
	67	0	Early-Onset	Yes	—	Yes	3.16	(1.42–6.85)
	73	59	Late-Onset	Yes	—	Yes	27.09	(13.57–46.78)
	77	0	Early-Onset	—	—	Yes	3.17	(1.67–5.95)
	77	45	Early-Onset	—	—	Yes	28.68	(15.07–47.67)

[a]Early onset = <60 years at diagnosis; Late onset = ≥60 years at diagnosis.

The authors applied LLP risk prediction model to the case-control population and found that the area under the curve (AUC) was 0.71. The 10-fold cross validation of the LLP risk prediction model produced AUC statistics of 0.70 which indicated good discrimination between cases and controls.

HEALTH TECHNOLOGY ASSEMENT REPORT

7.1.1.1 LLP Risk Prediction Model Validation Study

Raji YO. et al. conducted an independent validation study for LLP risk prediction model. The study evaluated the discrimination of the LLP risk prediction model and demonstrated its predicted benefit for stratifying patients for Computed Tomography (CT) screening by using data from two case-control studies and a population based prospective cohort from Europe and North America. In addition, the authors also evaluated the potential clinical effect of using the model for making decisions about lung cancer CT screening.[12, level II-2]

Study Population

The authors used data from three studies:

i. EUELC (European Early Lung Cancer) case-control study
- International, collaborative case-control study of early stage lung cancer conducted in eight European countries between 2002 and 2006
- 585 case patients had histologically or cytologically confirmed non-small cell lung cancer with surgically resected primary tumours and were matched with control participants by age, sex and study centre
- 1,238 control participants were recruited from hospitals or from the population registers of general practitioners in the same area as the case patients

ii. Harvard case-control study
- Hospital-based study of cases of non-small cell lung cancer diagnosed at Massachusetts General Hospital, Boston since 1992
- 1,738 patients with histologically confirmed lung cancer
- 1,184 control participants who could be family members and friends of case patients or persons attending the same hospital for other diseases

iii. Liverpool Lung Project population-based prospective cohort (LLPC) study
- Consisted of persons aged 40 to 79 years
- Randomly selected from the Liverpool area (population cohort) or recruited from hospitals to which they came for health episodes other than lung cancer (hospital cohort)
- The cohort comprised of 7,652 participants recruited between 1998 and 2008 and followed annually for lung cancer and mortality outcomes through the Office for National Statistics, the North West Cancer Intelligence Service and hospital case-note review

Risk Predictors

The standardised questionnaire to collect self-reported information included:

i. Demographic and socioeconomic characteristics

ii. Medical history
- History of pneumonia
- History of cancer

HEALTH TECHNOLOGY ASSEMENT REPORT

iii. Family history of cancer
- Age of onset in a first-degree relative (none. Early [<60 years], or late [≥60 years])

iv. History of tobacco consumption
- Age of starting and quitting smoking
- Number of cigarettes smoked per day
- Smoking duration was measured in years (never smoked or smoked for <20, 21 to 40, 41 to 60 or >60 years)

v. Lifetime occupational history
- Asbestos exposure

Statistical Analyses: Methods

The authors conducted several statistical analyses to predict the risk of lung cancers. Those analyses were:

i. LLP risk prediction model developed from LLCC study was used to predict a person's absolute 5-year risk for lung cancer
ii. LLP risk prediction model performance was assessed by measuring discriminative accuracy and by decision curve and relative utility curve analysis
iii. Decision curves plot the predicted net benefit of the risk prediction model versus risk threshold

Statistical Analyses: Results

Based on the response rate, most participants in the EUELC were men, regardless of case-control status. Meanwhile in Harvard study, different sex distributions were observed for case patients and control participants. However, in those three studies the distributions of age, smoking duration, family history, and asbestos exposure had similar patterns particularly for case patients.

In the LLPC, out of 7,652 participants, 420 (approximately 6% of the cohort) developed lung cancer over an average follow-up period of eight years. The lung cancer rates was slightly higher in men than in women, higher in participants with history of pneumonia than in those without, and approximately three times higher in persons with history of cancer than in those without. Lung cancer rates also increased with greater age and longer smoking. The performance of LLP risk prediction model also showed that most absolute risks that was greater than 2.5% (>2.5%) were predicted for patients with cancer, whereas about one half of disease-free patients had absolute risks less than 1%.

The model seems only slightly better compared to duration of smoking or family history of lung cancer as shown in Table 3.

Table 3: Relative Utility of the LLP Risk Prediction Model and Smoking Duration in the EUELC and Havard Data Sets[12]

Absolute Risk Threshold	Risk Group	EUELC (n=1268)		Harvard (n = 2922)		LLPC (n = 7652)	
		Case Patients (n = 585) n (%)	Control Participants (n = 1283) n (%)	Case Patients (n = 1738) n (%)	Control Participants (n = 1184) n (%)	With Disease (n = 420) n (%)	Disease-Free (n = 7232) n (%)
2.50%	<2.5%	262 (44.8)	895 (69.8)	613 (35.3)	865 (73.1)	108 (25.7)	4873 (67.4)
	≥2.5%	323 (55.2)	388 (30.2)	1125 (64.7)	319 (26.9)	312 (74.3)	2359 (32.6)
5.00%	<5.0%	376 (64.3)	1082 (84.3)	880 (50.6)	1031 (87.1)	179 (42.6)	5867 (81.1)
	≥5.0%	209 (35.7)	201 (15.7)	858 (49.4)	153 (12.9)	241 (57.4)	1365 (18.9)
10.00%	<10.0%	502 (85.8)	1210 (94.3)	1299 (74.7)	1133 (95.7)	303 (72.1)	6690 (92.5)
	≥10.0%	83 (14.2)	73 (5.7)	439 (25.3)	51 (4.3)	117 (27.9)	542 (7.5)

*EUELC – European Early Lung Cancer; LLPC – Liverpool Lung Project Prospective Cohort

The LLP risk prediction model had modest discrimination in the EUELC data set (AUC, 0.67 [CI 0.64 to 0.69]) and good discrimination in both Harvard (AUC, 0.76 [95% CI 0.75 to 0.78]) and LLPC (AUC, 0.82 [CI 0.80 to 0.85]) data set. The AUC for smoking duration which was the strongest of the risk factors was 0.63, 0.74, and 0.72 in the EUELC, Harvard and LLPC data sets respectively.

7.1.2 Risk Prediction Model for Lung Cancer in Korean Men

Park S et al. conducted a study to develop an individualized risk prediction model for lung cancer in Korean men using population-based cohort data. The model was configured to estimate the absolute risk that an individual will have lung cancer in eight years as well as to identify the significant risk factors for lung cancer.[13, level II-2]

Study Population

This study used population-based cohort data which involved 1,309,144 men between ages of 30 and 80 years who underwent health examination between 1996 and 1997. The study was restricted to participants who were free of any cancer at baseline and had complete information on the relevant risk factors.

The starting point of the study was the date of health examination, the event date was the date of first diagnosis of lung cancer and the last date of follow-up was December 2007.

Risk Predictors

The risk predictors were collected using standardised lifestyle questionnaires which included:

i. Smoking habits

- Smoking status was classified as never, past and current smoker. Past smoker was defined as a person who 'has quit smoking for at least 1 year' before the time of the health check-up. Meanwhile, for past and current smokers the authors assessed the duration of smoking. The average amount smoked per day was assessed for current smokers. Current smokers were classified into three groups:

  a. Current smoker consuming <0.5 pack per day
  b. Current smoker consuming 0.55-0.99 packs per day
  c. Current smoker consuming >1 pack per day

- Age at smoking initiation was assessed using the information on smoking duration by current smokers only and was divided into five groups based on the school-aging system in Korea and ten-year age intervals. Those groups were:

  a. Less than 16 years (<16) age at smoking initiation
  b. Between 16 and 18 years (16-18) age at smoking initiation
  c. Between 19 and 29 years (19-29) age at smoking initiation
  d. Between 30 and 39 years (30-39) age at smoking initiation
  e. At age of 40 years and more (≥40)

ii. Alcohol drinking

iii. Physical activity

- Physical activity was evaluated based on the intensity (number of exercise sessions per week) and duration (how long per session) of leisure-time physical activity. The physical activity was classified into three groups:

  a. Low (≤4 times per week at <30 minutes per session
  b. Moderate (2-4 times per week at ≥30 minutes per session or ≥5 times per week at <30 minutes per session
  c. High ≥5 times per week at ≥30 minutes per session

iv. Meal preferences (meat versus vegetables)

v. Previous disease history

- All the participants underwent blood and urine laboratory tests to check their health status including blood glucose levels. The health check-up data were obtained from Korea National Health Insurance Cooperation.

vi. History of disease in parents or siblings which included any type of cancers, cardiovascular diseases or diabetes.

vii. Body Mass Index (BMI)

Statistical Analyses: Methods

Several statistical analyses were used to develop the risk prediction models. They were:

i. Cox proportional hazards model to develop a multivariable model for lung cancer risk.
ii. Hierarchical variable selection method
iii. Likelihood ratio tests were used to select the significant variables
iv. Log-log survival plot

Multivariable Analysis / Discrimination

The respondents consisted of 1,309,144 participants with 10,007 newly diagnosed lung cancer cases observed during the 8-year follow-up. The mean age of the cohort was 45 years old; 28.6% were never smokers and 13.9% were current smokers consuming ≥1 pack per day. The majority of the participants were alcohol consumers (84.6%) and had a BMI within normal range (18.5-24.9, 69.0%). Twelve percent of the participants had a family history of any cancer and 6% had fasting glucose levels >126 mg/dL.

Univariable analyses showed that older age, smoking, early age at smoking initiation, high alcohol consumption, and low BMI were significantly associated with a higher lung cancer risk. High glucose levels were also associated with an elevated lung cancer risk. Having a family history of any cancer was not significantly related to lung cancer risk.

After multivariable analyses, the authors found that current smokers with high cigarette consumption showed an approximately 4-fold elevated risk of developing lung cancer and there was significant increasing trend of lung cancer risk by amount smoked (p-value for trend <0.0001). Lean participants with BMI <18.5%, the risk of developing lung cancer increased compared to heavier participants. The heavier participants had approximately 29% decreased risk compared with participants with normal BMI. Other risk predictor that reduced the risk of developing lung cancer was physical activity, 5-13%. High fasting glucose level (≥126mg/dL) was significantly associated with lung cancer. Those participants who initiated smoking at younger age were significantly associated with lung cancer risk. In the multivariable analysis, alcohol consumption was no longer significant when it was included in the model simultaneously with smoking and was excluded from the final model. The details of the multivariable analyses are shown in Table 4:

Table 4: Multivariable Regression Model: Risk Prediction[13]

Risk factor	β	HR* (95% CI)	*p*-value	*p* for trend
Age-Mean$_{age}$, years	0.1668	1.18 (1.18–1.19)	<0.0001	
(Age-Mean$_{age}$)^2, years^2	−0.0020	1. 00 (1.00–1.00)	<0.0001	
Smoke				
Never		1.00 (reference)		<0.0001
Past	0.4180	1.52 (1.41–1.64)	<0.0001	
Current, <0.5 pack/day	0.4444	1.56 (1.42–1.71)	<0.0001	
Current, 0.5–0.99 pack/day	0.9414	2.56 (2.37–2.78)	<0.0001	
Current, ≥1 pack/day	1.3889	4.01 (3.68–4.37)	<0.0001	
Age at smoking initiation				
Age ≥40		1.000 (reference)		<0.0001
30≤ Age <40	0.2194	1.25 (1.16–1.34)	<0.0001	
19≤ Age <30	0.2809	1.32 (1.23–1.43)	<0.0001	
16≤ Age <19	0.5249	1.69 (1.4–2.04)	<0.0001	
Age <16	0.7120	2.04 (1.46–2.84)	<0.0001	
BMI, kg/m^2				
<18.5	0.3306	1.39 (1.28–1.51)	<0.0001	<0.0001
18.5–22.9		1.00 (reference)		
23.0–24.9	−0.2468	0.78 (0.74–0.82)	<0.0001	
≥25.0	−0.3386	0.71 (0.68–0.75)	<0.0001	
Physical activity				
No		1.000 (reference)		<0.0001
Light	−0.0909	0.91 (0.86–0.97)	0.0029	
Moderate	−0.1412	0.87 (0.83–0.91)	<0.0001	
Heavy	−0.0521	0.95 (0.89–1.02)	0.1431	
Fasting glucose level, mg/dL				
<126		1.000 (reference)		
≥126	0.0792	1.08 (1.01–1.16)	0.0201	0.0201

*Hazard ratios were obtained from a Cox proportional hazards model.
doi:10.1371/journal.pone.0054823.t003

Validation

The authors also validated the model. The Korean Risk Prediction Model for lung cancer was validated with an external validation using participants from the Korean National Health Corporation (1998 to 1999). Data of 507,046 male participants were used in the validation analysis. The prediction model is considered good when the discrimination is 0.75.

The model was validated with external validation data set, the C-statistic showed excellent discrimination of 0.864, 95% CI 0.860 to 0.868. If

considering only age and smoking variables the prediction model also showed excellent discrimination (C-statistic 0.861, 95% CI 0.857 to 0.865). Then the model fit was improved by including other covariates such as age at smoking initiation, physical activity, BMI and fasting glucose level. When tested with external validation dataset, the performance of the risk prediction model also showed excellent discrimination with C-statistic of 0.87, 95% CI 0.867 to 0.876.

7.1.3 Risk Model for Prediction of Lung Cancer (Spitz Model)

Spitz MR et al. conducted a study to determine the risk factors of lung cancer and developed risk models for predicting lung cancer. The authors constructed multivariable models separately for never smokers, former smokers and current smokers, incorporating into each model variables that exhibit statistically significant main effects.[9, level II-2]

Study Population

The study population were recruited from an ongoing molecular epidemiologic study of lung cancer from Thoracic Centre at the University of Texas since July 1995. The data was divided into training sets to guide risk prediction model and validation sets to assess the prediction of the risk. The case patients were all newly diagnosed patients presenting with histological confirmed lung cancer and were enrolled before initiation of chemo or radiation therapy. Meanwhile the control subjects were healthy subjects without a prior history of cancer who were recruited from the Kelsey Seybold Clinics. The controls were frequency matched to the case patients by age (±5 years), sex, ethnicity and smoking status. The smoking status was divided into three groups; never smokers group, former smokers group and current smokers group. All the study participants were limited to white non-Hispanic participants within Texas due to inadequate numbers of non-white participants to perform smoking stratum-specific analysis.

Risk Predictors

Data on risk predictors were collected during interviews which included:

i. Demographic characteristics

ii. Smoking history

- Never smokers were defined as individuals who never smoked or had smoked less than 100 cigarettes in a lifetime. Meanwhile the former smokers were defined as individuals who had smoked at least 100 cigarettes in a lifetime but quit smoking more than 12 months before lung cancer diagnosis (case patients) or before the interview (control subjects). Current smokers were those who were currently smoking and recent quitters (less than 12 months) before diagnosis of lung cancer (case patients) or interview (control subjects).

- Other data included under smoking history were smoking duration, number of cigarettes smoked per day, computed pack-years smoked and age at smoking initiation (for all smokers) plus age at smoking cessation and computed years since cessation (for former smokers).

iii. Occupation
- To identify any specific exposure during work such as tobacco smoke, asbestos and other air pollution exposure

iv. Information about specific exposures at work or daily activity
- The authors had concerns on several important exposures especially exposure to second-hand smoke (environmental tobacco smoke or ETS). The exposure was ascertained for never smoker and former smokers and was defined as having been exposed to someone else's cigarette smoke at home or at work on a regular basis.

- Another exposure was asbestos exposure. Any participants who had been employed within a documented asbestos-related occupation or industry were considered as positive for asbestos exposure. Other exposures to chemicals, fumes, dust and wood dust were also documented. The participants were also interviewed as to whether they had ever been diagnosed with emphysema, hay fever or asthma.

v. Medical history
- Past medical history especially respiratory diseases

vi. Family history of cancer in first-degree relatives
- For this data, the authors obtained information for each affected relative regarding year of birth, age at time of interview of the case or control subjects, smoking status, type of cancer, age at diagnosis and year of death.

Multivariable Analysis / Discrimination

There was no statistically significant difference in the distribution of cases and controls by smoking status.

Variables that were statistically significantly associated with lung cancer risk at the 5% level in univariate analysis were included in the multivariable logistic regression analyses for construction of the final risk models.

After multivariable logistic regression analysis, it was found that both exposure to ETS and family history of any cancer were statistically significantly associated with lung cancer in never smokers. Among former smokers, lung cancer was statistically significantly associated with exposure to dust, no prior history of hay fever, personal history of emphysema, family history of any cancer and age at smoking cessation. Among current smokers, lung cancer was shown to have statistically significant association with exposure to dust, no prior history of hay fever, personal history of emphysema, tobacco-related cancers, smoking intensity and exposure to asbestos. The results are shown in Table 5.

Table 5: Multivariable Logistic Model for Lung Cancer by Smoking Status[9]

Risk factor	Regression coefficient	*P*†	OR (95% CI)
Never smoker			
Intercept	−0.8806	<.001	
ETS (yes vs no)	0.5874	.0042	1.80 (1.20 to 2.69)
Family history (≥2 vs <2)‡	0.6954	<.001	2.00 (1.39 to 2.90)
Former smoker			
Intercept	−0.7606	<.001	
Emphysema (yes vs no)	0.9734	<.001	2.65 (1.95 to 3.60)
Dust exposure (yes vs no)	0.4654	<.001	1.59 (1.29 to 1.97)
Family history (≥2 vs <2)‡	0.4636	<.001	1.59 (1.28 to 1.98)
Age stopped smoking§			
<42 y	Referent		
42–53 y	0.2130	.1110	1.24 (0.95 to 1.61)
≥54 y	0.4080	.0018	1.50 (1.16 to 1.94)
	P for trend = .017		
Hay fever (no)	0.3711	.00e55	1.45 (1.12 to 1.88)
Current Smoker			
Intercept	−0.7173	<.001	
Emphysema (yes)	0.7561	<.001	2.13 (1.58 to 2.88)
Pack-years‖			
<28	Referent		
28–41.9	0.2219	.1932	1.25 (0.89 to 1.74)
42–57.4	0.3747	.0241	1.45 (1.05 to 2.01)
≥57.5	0.6151	<.001	1.85 (1.35 to 2.53)
	P for trend<.001		
Dust exposure (yes vs no)	0.3067	.0075	1.36 (1.09 to 1.70)
Asbestos exposure (yes vs no)	0.4109	.0127	1.51 (1.09 to 2.08)
Family history¶			
0	Referent		
≥1	0.3859	.0021	1.47 (1.15 to 1.88)
Hay fever (no)	0.4047	.0054	1.49 (1.13 to 1.99)

* Regression analysis was based on entire dataset (training and validation sets combined). OR = odds ratio; CI = confidence interval; ETS = environmental tobacco smoke.

† *P* value from Wald test.

‡ Number of first-degree relatives with any cancer.

§ Cut points based on the tertile of age at smoking cessation in control subjects in the training set.

‖ Cut points based on the quartile of current smoker pack-years in control subjects in the training set.

¶ Number of first-degree relatives with a smoking-related cancer (i.e., renal cancer and cancers of the lung, upper aerodigestive tract, esophagus, pancreas, bladder, and cervix).

The authors also used the risk prediction model to estimate absolute one year risk for lung cancer of each group. For example, a 75 year-old white man who was a current smoker with a history of 58 pack-year smoking, emphysema and hay fever, two first-degree relatives diagnosed with a smoking-related cancer and prior asbestos exposure has 8.75 higher estimated relative risk of lung cancer compared with a man of similar age but without any of the risk

factors. The estimated one-year absolute risk of lung cancer for this man was calculated as P = 0.0868. The risk was actually more than fifteen times that of the age specific Surveillance, Epidemiology and End Result Program (SEER) incidence rate for lung cancer in white man (0.56%).[9, level II-2]

Validation

The risk prediction model was validated using the same data from an ongoing molecular epidemiologic study of lung cancer from the Thoracic Centre at the University of Texas. The authors validated the risk prediction model with three-phase validation processes. The risk model was well calibrated throughout the entire range of probabilities as indicated by the non-statistically significant Hosmer-Lemeshow goodness-of-fit test statistics. The simplified results were shown in Table 6.[9, level II-2]

Table 6: Model Validation Statistic[9]

Smoking category	*P* from Hosmer–Lemeshow goodness of fit†	AUC† (95% CI)	Concordance statistic‡ (95% CI)
Never smokers	.777	0.57 (0.47 to 0.66)	0.59 (0.51 to 0.67)
Former smokers	.712	0.63 (0.58 to 0.69)	0.63 (0.58 to 0.67)
Current smokers	.688	0.58 (0.52 to 0.64)	0.65 (0.60 to 0.69)

* AUC = area under the curve; CI = confidence interval.

† Derived from validation set.

‡ Derived from threefold cross-validation for combined dataset.

The risk models were well calibrated throughout the entire range of probabilities, as indicated by non-statistically significant Hosmer-Lemeshow goodness-of-fit test statistics (0.777 for never smokers, 0.712 for former smokers and 0.688 for current smokers). The AUC statistics obtained from the validation sets, the AUC was low for never smokers and current smokers compared to former smokers. The results of concordance statistics (Table 6) indicated that the models performed reasonably well to discriminate between cases and controls.[9, level II-2]

7.1.4 Bach Risk Prediction Model

Bach PB et al. conducted a study to develop and validate a model of individual lung cancer risk that can be applied in both clinical and research settings. The authors examined the predicted 10-year lung cancer risk among subjects enrolled in an ongoing CT screening program to determine whether the risk of lung cancer varies and to ascertain the usefulness of the model as an adjunct to clinical research, the authors assessed the extent of variation in risk among a cohort of individuals who meet typical eligibility criteria for cancer prevention studies.[14, level I]

Study Population

For this study, the authors used data from Carotene and Retinol Efficacy Trial (CARET) a multicenter randomised controlled study that evaluated the impact of beta-carotene and vitamin A supplementation on lung cancer incidence and

mortality. CARET consisted of 18,172 subjects who were separated into two populations; 14,254 heavy smokers populations and 4,060 asbestos-exposed men either current smoker or former smokers. Heavy smokers were men and women aged between 50 to 69 years old who had at least 20 pack-years of smoking exposure and were either current smokers or had quit within six years of enrolment. Meanwhile the asbestos group involved men between aged 45 to 69 years old who were either current smokers or former smokers who had quit within 15 years of enrolment. The participants have either radiologic evidence of asbestos exposure or a history of employment in a trade that put them at high risk for asbestos exposure.

Risk Predictors

The model was configured to estimate the absolute risk that an individual will be diagnosed with lung cancer within 10 years. The authors chose the 10-year time horizon because the probable excess time taken for lung cancer to progress from an undetectable size to an untreatable stage; consequently, it may be a useful perspective for patient counseling about screening. The risk predictors assessed were:

i. Age and sex
ii. Prior history of asbestos exposure
iii. Duration of smoking
iv. Average amount smoked per day while smoking
v. Duration of abstinence from smoking for former smoker
- Identifiable from a clinical history
- The participants have established or strongly suspected risk factors for lung cancer

Multivariable Analysis / Discrimination

The authors created two one-year risk models in order to determine the absolute risk of lung cancer for individual within 10-years. One model was used to predict the probability of being diagnosed with lung cancer, which was the main focus of the study. The other model was used to predict the probability of dying without lung cancer diagnosis. Both models were cycled for about ten times to estimate ten-year lung cancer risk.[14, level I]

Cox proportional hazards regression was used to estimate the multivariable relations between the risk factors and the outcomes. In the one-year lung cancer risk model, the authors found that the associations between risk factors and lung cancer occurrence were consistent with other study findings they referred to. The risk factors were duration of smoking, average number of cigarettes smoked per day, duration of abstinence and age. Details of the findings were shown in Figure 2. The study drugs (beta-carotene and retinyl palmitate) increased the risk of lung cancer to a degree consistent with previously published data from CARET (Hazard ratio (HR) = 1.20 (95% CI: 1.06,1.25) P = 0.004). Another independent risk was asbestos exposure which increased lung cancer risk (HR = 1.24 (95% CI: 1.04,1.48) P = 0.02).

Meanwhile, gender did not independently influenced lung cancer risk (HR = 0.94 (95% CI: 0.92,1.08) P = 0.41).

Figure 2: Modelled Multivariable Relations between 1-Year Lung Cancer Risk and Each of 4 Continuous Predictors[14]

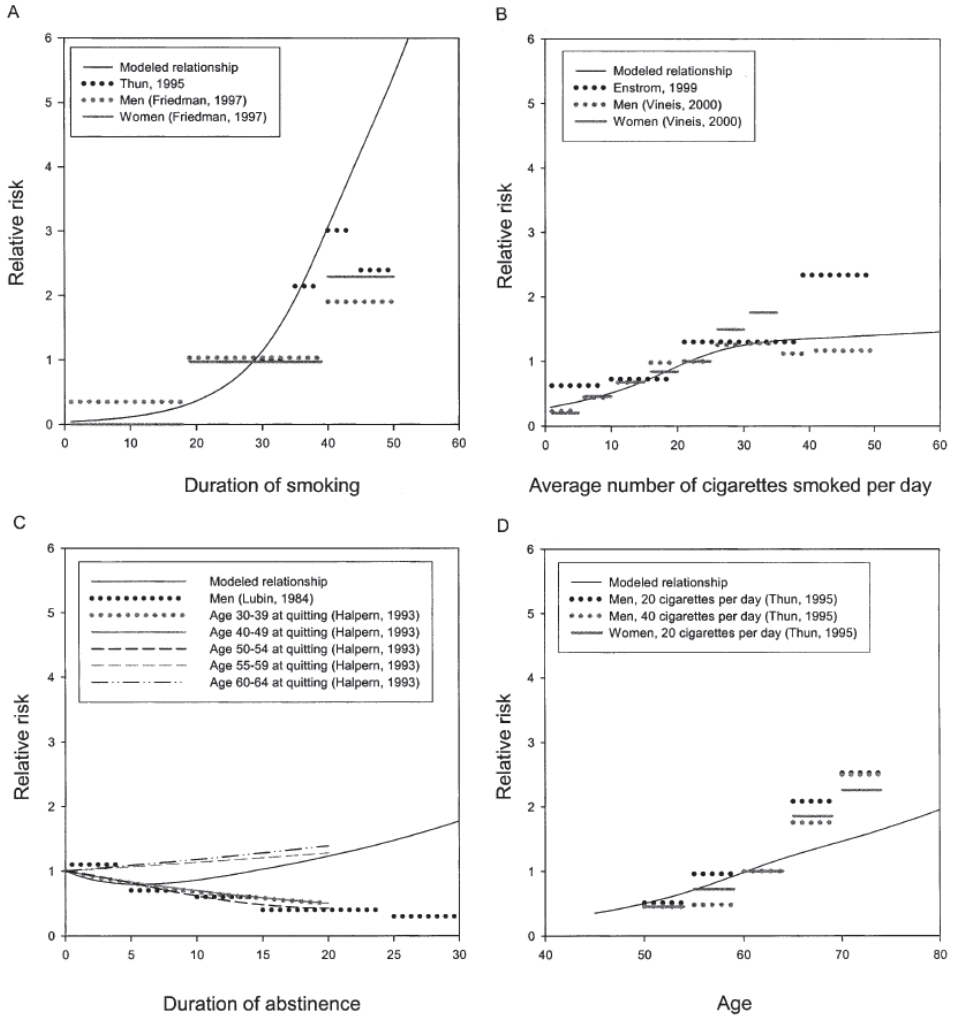


Fig. 1. Modeled multivariable relations between 1-year lung cancer risk and each of the four continuous predictors. Superimposed over each of the modeled relationships are parallel results previously reported by other authors for the same continuous predictors that had been analyzed in similar multivariable contexts in which the other continuous predictors were either controlled for or stratified on. Relative risks of lung cancer are shown for (**A**) duration of smoking *(45,46)*, (**B**) average number of cigarettes smoked per day *(47,48)*, (**C**) duration of abstinence *(33,35)*, and (**D**) age *(46)*. When the previous reports grouped the populations into categories, the expanse of the groupings is indicated by the length of the horizontal bars. When the previous reports included a regression equation, the shape of the reported response has been reproduced.

Validation

The BACH model was validated at all the six study sites; Seattle, Baltimore, New Haven, Portland, San Francisco, and Irvine. The risk prediction model only had a cross–validated concordance index about 0.66.[14, level I]

7.1.5 Italian Risk Prediction Model; COSMOS (Continuous Observation of Smoking Subjects) Trial

Maisonneuve P et al. conducted a study to develop a model based on epidemiologic and clinical risk factors to estimate the probability of individuals in a high-risk population being diagnosed with lung cancer. This model might be useful to stratify individuals and select those at high risk for inclusion in screening programs. Another aim was to develop a second model based on baseline CT findings in a screened population, combined with epidemiologic and clinical risk factors, to stratify individuals according to the probability of being diagnosed with lung cancer at repeat screening scans. The second model was proposed for use in large scale screening programs to select lower risk patients in whom the interval between screening CTs can be lengthened and at the same time to identify those at higher risk of lung cancer in whom surveillance intensity might be increased or who might benefit from prevention intervention studies.[15, level II-1]

Study Population

The data was collected from the ongoing COSMOS single centre non-randomised lung cancer screening trial which was conducted in Northern Italy. A total of 5,203 participants (3,439 men and 1,764 women) were recruited in the COSMOS trial. The participants were considered eligible for the study if they were asymptomatic volunteers aged ≥50 years old, heavy smokers (≥20 pack-years), still smoking or had stopped smoking less than 10 years previously and had not been diagnosed with cancer in the previous five years.

Risk Predictors

i. Smoking history
ii. Lifestyle
iii. Body Mass Index
iv. Fruit and vegetable consumption pattern
v. Alcohol
vi. Asbestos exposure
vii. Lung diseases
viii. Lung spirometry reading

Multivariable Analysis / Discrimination

The multivariable analysis showed that age, smoking duration, number of cigarettes smoked and predicted FEV1 (90% cut-off) were independently associated with lung cancer risk. Table 7 showed the results of univariable and multivariable analysis conducted.

Table 7: Univariable and Multivariable Analysis of Risk Factors[15]

Characteristic (participants)	Lung cancers	Univariate analysis HR[a] (95% CI)	Multivariable analysis HR[a] (95% CI)
Age			
<55 y (*n* = 1,759)	31	1.00	1.00
55–59 y (*n* = 1,725)	49	1.60 (1.02–2.51)	1.41 (0.86–2.31)
60–64 y (*n* = 1,061)	46	2.50 (1.59–3.94)	2.22 (1.27–3.88)
≥65 y (*n* = 658)	36	3.30 (2.04–5.33)	2.16 (1.11–4.21)
Smoking duration			
<35 y (*n* = 865)	11	1.00	1.00
35–40 y (*n* = 1,623)	42	2.01 (1.03–3.90)	1.89 (0.97–3.69)
40–44 y (*n* = 1,510)	54	2.77 (1.45–5.29)	2.02 (1.01–4.01)
45–49 y (*n* = 798)	24	2.40 (1.18–4.91)	1.39 (0.63–3.06)
≥50 y (*n* = 407)	31	6.47 (3.25–12.9)	3.45 (1.51–7.90)
Number of cigarettes smoked			
<20/d (*n* = 911)	20	1.00	1.00
20–24/d (*n* = 2011)	54	1.21 (0.73–2.02)	1.28 (0.77–2.15)
25–29/d (*n* = 557)	16	1.27 (0.66–2.46)	1.35 (0.70–2.61)
30–39/d (*n* = 973)	41	1.91 (1.12–3.26)	2.10 (1.22–3.60)
≥40/d (*n* = 743)	31	1.91 (1.09–3.35)	2.05 (1.16–3.61)
Dyspnea			
No (*n* = 2,956)	82	1.00	1.00
Yes (*n* = 1,686)	64	1.41 (1.02–1.95)	1.20 (0.86–1.68)
Chronic obstructive pulmonary disease			
No (*n* = 4,384)	126	1.00	1.00
Yes (*n* = 819)	36	1.60 (1.10–2.32)	1.17 (0.80–1.73)
Forced expiratory flow$_{25–75}$ ($FEF_{25–75}$% of predicted)			
≥80% (*n* = 1,523)	42	1.00	1.00
50–80% (*n* = 677)	28	1.49 (0.92–2.40)	1.10 (0.65–1.86)
<50% (*n* = 281)	16	2.05 (1.16–3.65)	1.05 (0.54–2.02)
Forced expiratory volume in 1 s (FEV1% of predicted)			
≥90% (*n* = 1,300)	30	1.00	1.00
<90% (*n* = 1,180)	56	2.12 (1.36–3.30)	1.74 (1.03–2.94)

NOTE: Forced expiratory flow$_{25–75}$ and forced expiratory volume in 1 second are missing for 2,722 participants.
[a]HR with 95% CI obtained from univariate and multivariable Cox proportional hazards regression models.

At the end of fourth screening round, 162 lung cancers were detected in 18,095 person-years of observation from baseline, giving a lung cancer detection rate of 0.90 per 100 years. The detection rates (per 100 years) were slightly higher in men (0.95) compared to women (0.78) and in current smokers (0.92) than former smokers (0.79). However, both differences were not significant. Table 8 below showed the lung cancer rates per 100 years and rate ratios (RR) in COSMOS participants according to baseline characteristics.

Table 8: Lung Cancer Rates and RRs in COSMOS Participants According to Baseline Characteristics[15]

Characteristic	Number	Person-years of follow-up	Lung cancers	Rate per 100 years	RR (95% CI)
All participants	5203	18095	162	0.90	
Sex					
Women	1764	6006	47	0.78	1.00
Men	3439	12089	115	0.95	1.22 (0.87-1.71)
Age					
<55 years	1759	6151	31	0.50	1.00
55-59 years	1725	6084	49	0.81	**1.62 (1.03-2.54)**
60-64 years	1061	3669	46	1.25	**2.50 (1.58-3.95)**
≥65 years	658	2191	36	1.64	**3.28 (2.02-5.32)**
Body mass index (kg/m^2)					
Underweight (<18.5)	58	196	1	0.51	0.55 (0.08-4.01)
Normal range (18.5-24.99	2329	8061	74	0.92	1.00
Overweight (25-29.99)	2186	7673	69	0.90	0.98 (0.70-1.36)
Obese (≥30)	606	2113	18	0.85	0.92 (0.55-1.55)
Fruit and vegetable intake					
Less than 3 servings/day	2798	9743	90	0.92	1.00
Three servings/day or more	2279	7938	70	0.88	0.96 (0.70-1.31)
Alcohol consumption					
Never	1463	5010	44	0.88	1.00
Occasionally	2365	8193	70	0.85	0.97 (0.66-1.41)
1-2 glasses/day	1023	3665	34	0.93	1.06 (0.67-1.66)
3-5 glasses/day	156	553	9	1.63	1.85 (0.90-3.81)
More than 5 glasses/day	122	423	3	0.71	0.81 (0.25-2.61)
Smoking status at baseline					
Former smoker	1028	3676	29	0.79	1.00
Current smoker	4175	14419	133	0.92	1.16 (0.78-1.74)
Age started smoking					
<15 years	831	2906	29	1.00	1.00
15-19 years	2913	10194	98	0.96	0.96 (0.63-1.46)
≥20 years	1459	4995	35	0.70	0.70 (0.43-1.15)
Years of cessation					
<5 years	618	2208	17	0.77	1.00
5-10 years	410	1468	12	0.82	1.06 (0.51-2.24)
Smoking duration					
<35 years	865	2970	11	0.37	1.00
35 to 40 years	1623	5713	42	0.74	**2.00 (1.03-3.89)**
40 to 44 years	1510	5357	54	1.01	**2.73 (1.43-5.23)**
45 to 49 years	798	2720	24	0.88	**2.38 (1.16-4.86)**
≥50 years	407	1335	31	2.32	**6.27 (3.14-12.5)**
Cigarette consumption					
<20/day	911	3143	20	0.64	1.00
20-24/day	2011	7001	54	0.77	1.20 (0.72-2.01)
25-29/day	557	1984	16	0.81	1.27 (0.65-2.45)
30-39/day	973	3401	41	1.21	**1.89 (1.11-3.23)**
≥40/day	743	2549	31	1.22	**1.91 (1.08-3.35)**
Pack-years					
20-39 pack-years	1914	6640	35	0.53	1.00
40-59 pack-years	1951	6827	60	0.88	**1.66 (1.09-2.52)**
≥60 pack-years	1338	4628	67	1.45	**2.74 (1.81-4.12)**
Passive smoking					
No	476	1657	14	0.84	1.00
Yes	4620	16070	146	0.91	1.08 (0.62-1.88)
Occupational exposure					
Asbestos	247	837	11	1.31	1.47 (0.79-2.73)
Radon	7	28	0	N/A	N/A
Cadmium	22	77	1	1.30	1.43 (0.20-10.3)
Chromium	69	236	1	0.42	0.46 (0.06-3.31)
Beryllium	9	35	0	N/A	N/A
Aluminium	99	337	3	0.89	0.98 (0.31-3.08)
Silica dust	93	345	5	1.45	1.61 (0.66-3.95)
Sulphuric acid	106	371	3	0.81	0.89 (0.28-2.80)
Ether	66	235	0	N/A	N/A
Carbon	56	194	1	0.52	0.57 (0.08-4.10)
Medical history					
Asbestosis	10	36	1	2.78	3.05 (0.42-22.4)
Asthma	231	792	10	1.26	1.42 (0.74-2.70)
Other allergy	974	3458	22	0.64	0.67 (0.42-1.05)
Pulmonitis	671	2328	17	0.73	0.78 (0.47-1.30)
Tuberculosis	114	375	6	1.60	1.78 (0.78-4.05)
Pleuritis	322	1114	10	0.90	1.01 (0.53-1.93)
Pneumothorax	105	379	7	1.85	2.08 (0.97-4.47)

Chronic obstructive pulmonary disease	819	2755	36	1.31	**1.60 (1.10-2.33)**
Thyroid disease	652	2284	16	0.70	0.80 (0.48-1.35)
Cardiovascular disease	1140	3933	43	1.09	1.27 (0.89-1.80)
Previous cancer	233	809	11	1.36	1.56 (0.84-2.90)
RESPIRATORY SYMPTOMS					
Lung disease limits activity					
No	4663	16285	148	0.91	1.00
Yes	357	1229	9	0.73	0.80 (0.41-1.58)
Wheezing					
No	3795	13313	118	0.89	1.00
Yes	1302	4461	40	0.90	1.01 (0.71-1.45)
Dyspnea					
No	2956	10435	82	0.79	1.00
Yes	1686	5795	64	1.10	**1.39 (1.00-1.93)**
RESPIRATORY FUNCTION*					
Forced vital capacity (FVC% of predicted)					
≥80%	1782	6251	55	0.88	1.00
50-80%	672	2283	29	1.27	1.44 (0.92-2.27)
<50%	27	85	2	2.35	2.67 (0.64-11.1)
Forced expiratory flow$_{25-75}$ (FEF_{25-75}% of predicted)					
≥80%	1523	5269	42	0.80	1.00
50-80%	677	2364	28	1.18	1.48 (0.91-2.39)
<50%	281	986	16	1.62	**2.03 (1.13-3.62)**
Forced expiratory volume in 1 second (FEV1% of predicted)					
≥80%	1810	6306	57	0.90	1.00
50-80%	602	2078	25	1.20	1.33 (0.83-2.14)
<50%	68	231	4	1.73	2.16 (0.78-5.96)
Forced expiratory volume in 1 second (FEV1% of predicted)					
≥90%	1300	4558	30	0.66	1.00
<90%	1180	4056	56	1.38	**2.09 (1.34-3.26)**
FEV1/FVC					
≥70%	2119	7360	70	0.95	1.00
<70%	355	1231	16	1.30	**1.37 (0.79-2.36)**

95% confidence intervals (CI) for rate ratios (RR) calculated assuming events followed a Poisson distribution. *P* for trend: age (<0.0001), alcohol consumption (0.50), age started smoking (0.22), smoking duration (<0.01), cigarette consumption (<0.01), pack-years (<0.01), FVC% (0.05); FEF_{25-75}% (0.01); FEV1% (0.11).
* Missing for 2722 participants

The authors also used Bach model to estimate the lung cancer risk in COSMOS participants. From the Bach model, it was estimated that 21 COSMOS participants would develop symptomatic lung cancer during first year and 55 lung cancers were detected in first screening round. Compared with Bach model, the observed incidence of lung cancer in COSMOS model was higher. However, when the Bach model was recalibrated, the recalibrated Bach model was accurate in predicting the observed incidence (Hosmer-Lemeshow X^2 test = 6.2; P = 0.63).

Validation

No validation study was retrieved for COSMOS model.

7.2 COST IMPLICATION

There was no retrievable evidence related to cost-effectiveness analysis (CEA) or cost-utility analysis (CUA) obtained from the scientific databases for each of the risk prediction models discussed in previous section.

Estimated potential direct cost implicated in designing, developing, testing and commissioning a risk prediction model on lung cancer in a portal is approximately RM75,000 to RM100,000 (Personal communication with Programme Officer, Health Education Division). This amount encompasses the sum of the activities listed below:

Services:

- System development (1 lot) RM50,000
- Installation and configuration (1 lot) RM5,000
- Training and Training of Trainer (1 lot) RM15,000
- Documentation (1 lot) RM1,000

Software:

- Portal Management Application and Content Management System (CMS) (1 lot) RM4,000

The cost does not include the cost incurred to the health care system due to additional reference of subjects with high risk of lung cancer, further screening, diagnosis and treatment.

7.3 OTHER CONSIDERATIONS

7.3.1. Organizational

In Malaysia, there are health risk assessment modules available for health problems such as obesity, mental health, diabetes and heart disease. However, no risk prediction model for lung cancer has been developed yet. The purpose of these models was to assess the risk of developing the disease or to identify early signs of such diseases before it gets worse, and for early management.

The introduction of HRA for lung cancer in Malaysia may give hope to the high risk groups for early detection and management. However, several issues need to be considered if the HRA module is to be implemented such as the next steps after the high risk group is identified. The module has to be supported with proper reference system which includes counseling, further diagnosis by physicians, as well as further management and treatment and reference mechanisms. This plan should involve physicians, radiologists, health counsellors, physiologists, psychologists and family members working together as a team.

Freedman AN et al. in their report stated that cancer risk prediction models can be used to assess the impact of cancer control interventions such as prevention strategies, screening and treatment on population trends in

incidence and mortality. In addition, the model can also be used to project future trends and to help determine optimal cancer control strategies.[17]

Freedman A.N et al also emphasized that the development of cancer risk prediction models require several consideration including research issues, gaps, any priorities, and resource needed to advance the field of cancer risk prediction and make specific recommendations for implementations. The recommendations fell into six broad areas:[17]

i) Revise existing risk assessment models and develop new models to improve predictive power
ii) Encourage the development of new types of risk models
iii) Obtain data to develop more accurate risk models
iv) Support mechanisms and resources to validate risk models
v) Strengthen model development efforts and encourage coordination within large research and clinical centres
vi) Promote effective cancer risk communication and decision-making

The authors stated that, one of the strategies to develop the risk prediction model is to combine case-control data with national registry data to provide detailed information on covariates in relatively short time. However, there are drawbacks of this approach such as potential of recall bias and lack of national registry data for many non-cancer diseases.[17]

7.3.2. Ethical and legal consideration

Freedman AN et al. highlighted in their review the possible use of risk prediction models with high discriminatory power to accurately identify small groups of individuals who will develop a disease so that a population prevention strategy of reducing risk factors prevalence in the whole population could yield maximum benefits. The alternative strategy of targeting high-risk individuals on the basis of a specific risk factor profile could miss a substantial number of individual with disease which might raise an ethical issues and psychological harms for those who are high risk but not included or subjects who were being labelled as 'at risk'.[17] At the same time considering the risk of false positive. The 'at risk' label also has implication for future health care cost. Theoretically increased psychological distress from risk labelling may contribute to other healthcare demands and raising the health care cost.

8.0 DISCUSSION

Five risk prediction models for lung cancer identified. The models were: Liverpool Lung Project (LLP) risk prediction model, Korean Model, Bach Model, Spitz Model and COSMOS Model. Each model have their own risk predictors which varied based on the study population involved. Four models namely LLP, Bach, Spitz and COSMOS model originated in Europe and the United States of America using cohort from other trials and cancer cases. All the models except the Korean model were based on Caucasian population. The Korean risk prediction model was developed based on Korean population.

All the models included smoking history as one of the predictors. Other risk predictors were medical and cancer history of the family as well as the subject and asbestos exposure. However, the Korean model also included body mass index, physical activity and fasting glucose levels as risk predictors. Spitz model was mainly concerned on the smoking factors in developing lung cancer.

Each model did show capabilities in predicting lung cancer. However, the best models among those five were LLP model and Korean model. The LLP model showed a potential in predicting lung cancer in patients with risk predictors within five years. Based on the multivariable analysis, those patients with family history of lung cancer, particularly family members who were under 60 years old had significant increase risks of lung cancer. History of respiratory diseases (e.g. pneumonia) and other cancers (i.e. other than lung cancer) also have increased risk of lung cancer.

Exposure to asbestos and long smoking duration also showed significant increased risk for lung cancer. For the other four models, the same risk predictors also showed significant increased risk of lung cancer except for asbestos exposure; however, COSMOS model found that asbestos exposure was not significantly associated with lung cancer.

The Korean model was the only model identified developed and validated in Asian population. The risk predictors were quite different from the other four models as the model also included BMI, physical activity and fasting blood glucose level as the risk of lung cancer. From the findings, it showed that lean participants with BMI less than 18.5 have increased risk of lung cancer. High fasting glucose level which was 126mg/dL and more was significantly associated with lung cancer. Younger age of smoking initiation was also significantly associated with lung cancer risk. The Korean model also showed that physical activity may decrease the risk of lung cancer by five to thirteen percent (5 to 13%). However, the Korean model is only for men.

Smoking is always the main risk predictors of lung cancer. All the models showed that current smokers or former smokers with long smoking history had increased risks of lung cancer. In the Spitz model, the researcher divided the case and control subjects into three groups of smoking status (never smoker, former smoker and current smokers). Based on the multivariable analysis using the model, never smokers had high risks of having lung cancer if they had both history of exposure to smoke (ETS) and had family history of any cancer. Among former smokers lung cancer was statistically significantly associated with exposure to dust, personal history of emphysema, family history of any cancer and age at smoking cessation. For current smokers, lung cancer was statistically significantly associated with exposure to dust, no prior history of hay fever, personal history of emphysema, family history with tobacco related-cancer and smoking intensity.

Freedman AN et al. emphasised that the most important characteristics of cancer risk prediction models performance are calibration, discrimination and

accuracy. **Calibration** or **reliability** assesses the ability of the model to predict the number of events in subgroups of the population. Calibration is most commonly evaluated by goodness-of-fit or chi-square statistic which compares the observed number of events with the expected numbers of events. Good calibration is important in all models, particularly in those used to estimate population disease burden and to plan population-level interventions. Recalibration of a model can be performed when risk is systematically over estimated or underestimated. **Discrimination** measures a model's ability to distinguish at the individual level between those who will develop disease and those who will not develop disease. Discrimination is commonly quantified by calculating the concordance statistic, which corresponds to the area under a receiver operating characteristics curve. Good discrimination in a model is important for decisions made at the individual level such as clinical decision-making and screening. **Accuracy** scores including positive and negative predictive value can be used to evaluate how well a model categorises specific individuals. It can be especially helpful in evaluating models used for clinical-decision making.[17]

Each model underwent a validation process. Only LLP risk prediction model was validated in a different population while the other models were validated in the same population where the models were developed. Based on those validation processes, LLP risk prediction model showed excellent discrimination compared to other models. The LLP risk prediction model has a good ability to distinguish persons who will not develop lung cancer by using predicted five-year absolute risk. Besides that, the LLP models also seems to be reasonably well-calibrated at high predicted risks and performs better than smoking duration or family history as a tool for deciding which persons to screen for lung cancer. The LLP risk prediction model also combined smoking duration, other important risk factors for lung cancer, and incidence data from cancer registries, thereby combining benefit of each to provide accurate and diverse predicted risks for smokers and non-smokers.

Another risk prediction model that had excellent discrimination was the Korean risk model. Risk prediction models with good discrimination are potentially valuable for identifying high-risk persons in a disease-screening application.[12] Table 9 showed the summary of the lung cancer risk prediction models.

Table 9: Comparison of Lung Cancer RIsk Prediction Models

Study Factor	HRA module				
	LLP Model	**Korean Model**	**Bach Model**	**Spitz Model**	**COSMOS Model**
Method of Estimation	Absolute risk	Absolute risk	Absolute risk	Absolute risk	Absolute risk
Characteristics of Study Population	579 lung cancer cases and 1157 age- and sex-matched population based controls	1,309,144 never, former or current smokers who were free of any cancer at baseline Age 30-80 years	18,172 current or former smokers (including 4,060 asbestos-exposed men) Age 45-69 years	1,851 case patients with history of lung cancer 2,001 control subjects without prior history of lung cancer	Used data from COSMOS trial 4,175 current and heavy smokers and 1,028 former smokers Median age 57 years
Risk Factors Included in Model	Age, medical history, family history of cancer, lifetime occupational history, history of tobacco consumption	Smoking exposure, age at smoking initiation, BMI, physical activity, fasting glucose level	Age, smoking duration, duration of abstinence, asbestos exposure	Smoking history, environmental tobacco smoke (ETS) exposure	Age, smoking intensity, lung cancer and lung diseases history
Strengths	Predict a person's risk of developing lung cancer within 5 years	Follow up patients for 8 years Korean population (Asian)	Calculates a person's risk of developing lung cancer within 10 years	Established levels of risk for smoking categories; high risk and low risk	Identifying lung cancer risks after first screening round
Limitations	Potential of recall and information bias	No women involved	Predictive only for 50-75 year olds who smoked 10-60 cigarettes per day for 25-55 years	Participants limited; white non-Hispanic within Texas	Not mentioned
Proposed Application Model	Identifying high-risk group for screening trials and medical counselling	Identifying high-risk group for screening trials and medical counseling / change lifestyle	Identifying high-risk groups for screening trials and medical counseling	Identifying high-risk groups for screening trials	Identifying high risk groups for further treatment

8.1 Limitations

This review has several limitations. Although there was no restriction in language during the search, only English full text articles were included in the report. Although every effort had been made to retrieve full text articles, there were two abstracts which the authors failed to retrieve full text articles. Most of the articles that meet the inclusion criteria for this review were studies on development of risk prediction for lung cancer study and only one paper was retrieved for validation study of one of the HRA modules. All the models except the Korean models were developed and validated in Caucasian populations.

9.0 CONCLUSION

There was fair level of retrievable evidence for risk prediction model for lung cancer. Five models were identified for predicting lung cancer risk. The LLP risk prediction model and the Korean risk prediction model were the best models for predicting lung cancer. LLP risk prediction model appeared to have excellent discrimination and good ability to distinguish between persons who will or will not develop lung cancer by using the predicted 5-year absolute risk. The Korean model is the only model that used Asian population (Korean) and has an excellent discrimination. For other risk prediction models, although they were well calibrated and validated, they appeared to have only modest ability to discriminate between subjects who will be having lung cancer, than for those who will not in the study population.

There was no retrievable evidence on safety related to risk prediction models or health risk assessment modules for the detection of lung cancer in the population.

There was no retrievable evidence on economic evaluation of risk prediction model or health risk assessment module for lung cancer, or cost implication involved in developing a new health risk assessment retrieved. The cost involved in validating a model by a prospective cohort validation study could be very costly depending on the number of study participants and years of follow up.

Risk prediction model or HRA module for lung cancer needs continual validation to give meaningful risk estimate and to ensure its capability in the setting it will be used. The complexity to develop and validate the HRA module is reflected in the necessary local data required. Dedicated research expertise to create a robust HRA module with consistent performance is very important.

10.0 RECOMMENDATION

Health risk assessment (HRA) module / risk prediction model for lung cancer such as Liverpool Lung Project (LLP) risk prediction model and the Korean risk prediction model need to undergo further validation until a well-fitted model with better predictive ability tailored to Malaysia population is established. The model needs continual validation to determine the consistency of its performance. Besides that, the module should only be introduced as part of comprehensive strategies for lung cancer whereby screening, treatment and rehabilitation is available.

11.0 REFERENCES

1. Trends In Lung Cancer Morbidity and Mortality. American Lung Association. Available at http://www.lung.org/z-testing-2/finding-cures/our-research/trend-reports/lc-trend-report.pdf Access on 24 April 2014
2. Lung Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012; Facts Sheets. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. Available at: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx Access on 12 Jun 2014
3. Cancer Facts & Figure 2014. American Cancer Society Available at: http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-042151.pdf Access on 17 July 2014
4. Malaysia Cancer Statistic – Data and Figure 2007, National Cancer Registry Report. 38-39
5. US National Institutes of Health. National Cancer Institute's: SEER Cancer Statistics Review, 1973-2008
6. Lung Cancer Screening Decision Tool. Available at: http://www.mskcc.org/cancer-care/adult/lung/screening-decision-tool Access on 20 April 2014
7. Health Risk Assessment. Wikipedia https://en.wikipedia.org/wiki/Health_risk_assessment
8. Health Risk Appraisal. Centres for Disease Control and Prevention. Available at: http:///.cdc.gov/nccdphp/dnpao/hwi/programdesign/health_risk_appraisals.htm Access on 20 April 2014
9. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, Shete S & Etzel CJ. A Risk Model for Prediction of Lung Cancer. J Natl Cancer Inst. 2007; 99:715-726
10. Lung Cancer Screening Decision Tool. Memorial Sloan Kettering Cancer Center. Available at: http://www.mskcc.org/cancer-care/adult/lung/screening-decision-tool Accessed on: 24 April 2014
11. Cassidy A, Myles JP, Tongeren MV et al. The LLP risk prediction model: An Individual Risk Prediction Model for Lung Cancer. Br J Cancer. 2008; 98: 270-276
12. 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(4):242-250
13. Park S, Nam BH, Yang HR, et al. Individualized Risk Prediction Model for Lung Cancer in Korean Men. Plos ONE. 2013; 8(2): e54823

14. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, Hsieh LJ & Begg CB. Variations in Lung Cancer Risk Among Smokers. J Natl Cancer Inst. 2003; 95(6);470-478
15. Maisonneuve P, Bagnardi V, Bellomi M et al. Lung Cancer Risk Prediction to Select Smokers for Screening CT – a Model Based on the Italian COSMOS Trial. Cancer Prev Res. 2011; 14:1778-1789
16. Peres J. Lung Cancer Screening Gets Risk-Specific. J Natl Cancer Inst Advance Access. 2012; 1-5
17. Freedman AN, Seminara D, Gail MH et al. Cancer Risk Prediction Models: A Workshop on Development, Evaluation and Application. J Natl Cancer Inst. 2005;97:715-23
18. Cassidy A, Duffy SW, Myles JP et al. Lung Cancer Risk Prediction: A Tool for Early Detection. Int. J. Cancer. 2006; 120:1-6
19. Mills PR, Masloski WS, Bashaw CM et al. Design, Development and Validation of the RedBrick Health Assessment: A Questionnaire-based Study. J R Soc Med Sh Rep. 2011;2:71
20. McCarthy WJ, Meza R, Jeon J et al. Lung cancer in Never Smokers: Epidemiology and Risk Prediction Models. Risk Analysis. 2012; 32(S1):S69-S84
21. Cronin KA, Gail MH, Zou Z et al. Validation of a Model of Lung Cancer Risk Prediction among Smokers. J Natl Cancer Inst. 2006;98(9):637-401

APPENDICIES

Appendix 1

HIERARCHY OF EVIDENCE FOR EFFECTIVENESS STUDIES

DESIGNATION OF LEVELS OF EVIDENCE

I Evidence obtained from at least one properly designed randomized controlled trial.

II-I Evidence obtained from well-designed controlled trials without randomization.

II-2 Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one centre or research group.

II-3 Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.

III Opinions or respected authorities, based on clinical experience; descriptive studies and case reports; or reports of expert committees.

SOURCE: US/CANADIAN PREVENTIVE SERVICES TASK FORCE (Harris 2001)

Appendix 2

PTK-FM-02

HEALTH TECHNOLOGY ASSESSMENT (HTA) PROTOCOL
HEALTH RISK ASSESSMENT MODULE (HRA) FOR LUNG CANCER

1.0 BACKGROUND INFORMATION

Lung cancer is the most common cancer worldwide, accounting for 1.3 million deaths annually. The United States (US) National Institutes of Health estimated that cancers cost the United States an overall $264 billion in 2010. It was estimated that approximately $10.3 billion was spent in the United States on lung cancer treatment alone.

The number of new cases of lung cancer was estimated to be 1.8 million in 2012 (12.9% of the total 14.1 million new cancer cases, in 2012 worldwide), 58% of which occurred in the less developed regions. The disease remains as the most common cancer in men worldwide (1.2 million, 16.7% of the total) with the highest estimated age-standardised incidence rates in Central and Eastern Europe (53.5 per 100,000) and Eastern Asia (50.4 per 100,000). The American Cancer Society stated in Cancer Facts and Figure 2014, that lung cancer accounts for more deaths than any other cancer in both men and women and was responsible for nearly one-in-five deaths (1.59 million deaths, 19.4% of the total).

According to the Malaysia National Cancer Registry, 1,865 cases of lung cancer were diagnosed and registered in 2007. The age standardised rate (ASR) for male was 14.7 per 100,000 and 5.6 per 100,000 for female. The incidence was more than two-fold higher among males compared to females. Based on ethnicity, Chinese were found to have higher incidence rate compared to Malay and Indian. The incidence increased with age and in 2007 the peak age-specific incidence rate was among the 70-75 age groups. Most of the lung cancers were detected late where 60% of the cases were detected at stage IV; the percentage of lung cancer detected at stage I and II was only 12%.

The lung cancer five-year survival rate (16.3%) is lower than many other leading cancer sites, such as the colon (62.2%), breast (90.0%) and prostate (99.9%). The five-year survival rate for lung cancer is 52.6 percent for cases detected when the disease is localized (within the lungs). However, only 15 percent of lung cancer cases are diagnosed at an early stage. For distant cancer metastases (spread to other organs) the five-year survival rate is only 3.5%.

According to the US National Lung Screening Trial (NLST), screening program is suggested for high risk population. One of early screening methods is through health risk assessment (HRA) tools. This health risk

assessment also known as health risk appraisal, health & well-being assessment or risk prediction model is an online questionnaire that asks only about lung cancer risk factors and it is completely confidential. The health risk assessment model is one of the most widely used screening tools in the field of health promotion and is often the first step in multi-component health promotions programme. The questionnaires, allow individuals to evaluate their health risks and quality of life. Commonly the HRA incorporates three key elements: an extended questionnaire, a risk calculation or score, and some form of feedback i.e. face-to-face with a health advisor or an automatic online report. According to Centers for Disease Control and Prevention (CDC), health risk assessment or health risk appraisal is a systematic approach to collect information from individuals that identifies risk factors, provides individualized feedback and links the person with at least one intervention to promote health, sustain function and/or prevent disease. A typical HRA instrument obtains information on demographic characteristics, lifestyle, personal medical history, and family medical history.
In Malaysia, currently there is no risk assessment model for early detection of lung cancer. This review was requested by a Senior Principal Assistant Director, Health Education Division, MOH to review the evidence on lung cancer risk assessment model as a tool in enhancing early detection and diagnosis of lung cancer, towards facilitating implementation of affordable and effective cancer control in the country.

Technical Features
Cancer risk assessment model / health risk assessment tool are statistical models developed for cancer risk prediction and can be divided into two broad categories:

i. To predict the probability of being diagnosed with a particular cancer, and
ii. To predict the likelihood of carrying a gene mutation that predisposes to a particular cancer or set of cancers.

Thus, it is supposed to be useful in clinical decision making. According to Memorial Sloan Kettering Cancer Centre, health risk assessment tools helps clinicians and patients to determine the chance that screening will be beneficial. There are few numbers of health risk prediction tools available. Those are:

i) CLEAR Model
Quantify a smoker's risk of developing lung cancer in the next five, ten or 15 years based on the person's age, sex, smoking history, medical history, and family history of cancer and past exposures to asbestos or wood dust
ii) Liverpool Lung Project Model (LLP)
LLP was developed from the LLP case– control (LLCC) study, provides a single unified model for smokers (current and former) and

non-smokers. LLP model accounts for important lung cancer risk factors besides age, sex, and smoking duration, including history of pneumonia, history of non-lung cancer, asbestos exposure, and family

iii) Bach Model
Predict risk of lung cancer in smokers only

iv) Spitz Model
Require 3 separate models to predict risk in current, former or non-smokers

2.0 POLICY QUESTION

2.1 In the Ministry of Health, should health risk assessment (HRA) module for lung cancer be introduced as one of the strategies in the prevention of lung cancer under the Malaysia National Cancer Control Programme?

2.2 If HRA module (cancer risk prediction model) for lung cancer is to be introduced, which risk prediction model for lung cancer should be adopted / adapted in Malaysia?

3.0 OBJECTIVES

3.1 To assess the effectiveness in term of predictive accuracy of lung cancer risk assessment / prediction models

3.2 To assess the safety, organizational, ethical issues and economic implications related to risk assessment / prediction models for lung cancer

RESEARCH QUESTIONS

1. What are the predictive accuracies of available risk prediction models for lung cancer in terms of detection rate and stage of disease at diagnosis?
2. What are the strengths and weaknesses of each available model?
3. What is the economic impact, ethical, legal and organizational issues of the health risk assessment module?

4.0 METHODS

4.1. Search Strategy

4.1.1 Electronic databases will be searched for published literatures pertaining to risk prediction models or health risk assessment module for lung cancer

The databases are MEDLINE, PubMed, and EBM Reviews-Cochrane Database of Systematic Review, EBM-Reviews-Cochrane Central Register of Controlled Trials, EBM Reviews-Health Technology Assessment, EBM Reviews-Cochrane Methodology Register, EBM Reviews-NHS Economic Evaluation Database, Database of Abstracts of Reviews of Effects (DARE), Horizon Scanning, INAHTA Database, HTA database and FDA database.

4.1.2 Additional literatures will be identified from the bibliographies of the related articles.

4.1.3 General search engine will also be used to get additional web-based information.

4.1.4 There will be no limitation applied in the search such as year and language.

4.1.5 The search strategy will be included in the appendix.

4.2 Inclusion and exclusion criteria

4.2.1 Inclusion criteria

a. Study design	:	HTA reports, systematic review, randomised controlled trial (RCT), diagnostic accuracy studies, cross-sectional, cohort, case-control, and economic evaluation studies.
b. Population	:	Adults at age of ≥ 18 years old
c. Intervention	:	Health Risk Assessment Tool / Risk Prediction Model / Health & Well-being Assessment
d. Comparators	:	No HRA module / risk prediction model
e. Outcome	:	One or more of the following outcome measures will be assessed;

i. Effectiveness/benefit of the lung cancer HRA related to patient outcome as measured by detection rate and stage of diagnosis

ii. Performance of available lung cancer risk prediction models in term of its predictive accuracy – sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), calibration as measured by expected/observed ratio, discrimination as measured by area under the curve (AUC) or c-statistic

iii. Organizational (operational, training, resources>, ethical, legal and economic implication

f. Publication : Full text articles published in English

4.2.2 Exclusion criteria

i. Animal study
ii. Narrative review
iii. Experimental study
iv. Non English full text article
v. Prognostic model

Based on the above inclusion and exclusion criteria, study selection will be carried out independently by two reviewers. Disagreement will be resolved by discussion.

4.3 Data extraction strategy

The following data will be extracted:

4.3.1 Details of methods and study population characteristics.

4.3.2 Detail of intervention and comparators.

4.3.3 Details of individual outcomes for effectiveness, safety and cost associated with Health Risk Assessment Models

Data will be extracted from selected studies by a reviewer using a pre-designed data extraction form and checked by another reviewer. Disagreements will be resolved by discussion.

4.4 Quality assessment strategy

The methodology quality of all retrieved literatures will be assessed using the relevant checklist of Critical Appraisal Skill Programme (CASP).

4.5 Methods of analysis/synthesis

Data on the diagnostic accuracy, effectiveness, safety and cost-effectiveness of risk prediction models of Lung Cancer will be presented in tabulated format with narrative summaries.

5.0 Report writing

Appendix 3

Search strategy:

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citation and Ovid MEDLINE(R) 1946 to Present

1. mass screening/
2. (mass adj 1 screening*).tw.
3. screening* .tw.
4. (risk adj 1 appraisal* health).tw.
5. (indices adj 1 health status).tw.
6. health risk appraisal* .tw.
7. Health Risk Assessment Tool.tw.
8. (assessment* adj 1 (risk or benefit risk or risk benefit or risk-benefit)).tw.
9. risk.mp. and benefits.tw. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
10. (risk adj2 benefits).tw.
11. Risk Prediction Model/
12. Risk Prediction Model.tw.
13. Health & Well-being Assessment/
14. Health & Well-being Assessment.tw.
15. Health Status Indicators/
16 .(index* adj 1 health status).tw.
17. (indicator* adj 1 health status).tw.
18. (risk adj 1 appraisal* health).tw.
19. (indices adj 1 health status).tw.
20. health risk appraisal* .tw.
21. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20
22. lung neoplasms/
23. ((pulmonary or lung) adj1 (cancer* or neoplasm*)).tw.
24. cancer of lung.tw.
25. cancer of the lung.tw.
26. Lung tumour.tw.
27. 22 or 23 or 24 or 25 or 26
28. 21 and 27
29. limit 28 to (English language and humans and “all adult (19 plus years)”)

Embase 1988 to 2014 Week 46

1. mass screening/
2. (mass adj 1 screening*).tw.
3. screening* .tw.
4. (risk adj 1 appraisal* health).tw.
5. (indices adj 1 health status).tw.
6. health risk appraisal* .tw.
7. Health Risk Assessment Tool.tw.
8. (assessment* adj 1 (risk or benefit risk or risk benefit or risk-benefit)).tw.
9. risk.mp. and benefits.tw.
10. (risk adj2 benefits).tw.
11. Risk Prediction Model/
12. Risk Prediction Model.tw.
13. Health & Well-being Assessment/
14. Health & Well-being Assessment.tw.
15. Health Status Indicators/
16 .(index* adj 1 health status).tw.
17. (indicator* adj 1 health status).tw.
18. (risk adj 1 appraisal* health).tw.
19. (indices adj 1 health status).tw.
20. health risk appraisal* .tw.
21. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20
22. lung neoplasms/
23. ((pulmonary or lung) adj1 (cancer* or neoplasm*)).tw.
24. cancer of lung.tw.
25. cancer of the lung.tw.
26. Lung tumour.tw.
27. 22 or 23 or 24 or 25 or 26
28. 21 and 27
29. limit 28 to (English language and humans and “all adult (19 plus years)”)

Appendix 4

RCT

CRITERIA ASSESSED			
Assignment of patients randomised?	Yes ☐	No ☐	Can't tell ☐
Allocation concealment?	Yes ☐	No ☐	Can't tell ☐
Patients, health workers, study personnel blind to treatment?	Yes ☐	No ☐	Can't tell ☐
Intention to treat analysis?	Yes ☐	No ☐	Can't tell ☐
Explanation of loss to follow-up?	Yes ☐	No ☐	Can't tell ☐

COHORT

CRITERIA ASSESSED			
Selection (cohort recruited in an acceptable way?)	Yes ☐	No ☐	Can't tell ☐
Exposure accurately measured?	Yes ☐	No ☐	Can't tell ☐
Outcome accurately measured?	Yes ☐	No ☐	Can't tell ☐
Confounding factors identified and taken account?	Yes ☐	No ☐	Can't tell ☐
Follow-up of subjects complete and long enough?	Yes ☐	No ☐	Can't tell ☐

Appendix 5

LIST OF EXCLUDED STUDIES

1. A screening update for smokers and ex-smokers. Johns Hopkins Medical Letter, Health After 50. 2012;24(11):4.
2. Ahmad U, Detterbeck FC. Current status of lung cancer screening. Semin Thorac Cardiovasc Surg. 2012;24(1):27-36.
3. Albin M. [Authorities' risk assessment was influenced by the asbestos industry. The chemicals legislation proposed by the European Union makes the question of independent expertise a current issue]. Lakartidningen. 2004;101(14):1306-1309.
4. Azari MR, Nasermoaddeli A, Movahadi M, et al. Risk assessment of lung cancer and asbestosis in workers exposed to asbestos fibers in brake shoe factory in Iran. Ind Health. 2010;48(1):38-42.
5. Bach PB. Inconsistencies in findings from the early lung cancer action project studies of lung cancer screening. J Natl Cancer Inst. 2011;103(13):1002-1006.
6. Berman DW. Apples to apples: the origin and magnitude of differences in asbestos cancer risk estimates derived using varying protocols. Risk Anal. 2011;31(8):1308-1326.
7. Berry G. Relative risk and acceleration in lung cancer. Stat Med. 2007;26(18):3511-3517.
8. Bridgman S. Community health risk assessment after a fire with asbestos containing fallout. J Epidemiol Community Health. 2001;55(12):921-927.
9. Burns DM. Primary prevention, smoking, and smoking cessation: implications for future trends in lung cancer prevention. Cancer. 2000;89(11 Suppl):2506-2509.
10. Calabro E, Randi G, La Vecchia C, et al. Lung function predicts lung cancer risk in smokers: a tool for targeting screening programmes. Eur Resp J. 2010;35(1):146-151.
11. Campos-Outcalt D. Lung cancer screening: USPSTF revises its recommendation. J Fam Pract. 2013;62(12):733-740.
12. Chang HY, Chen CR, Wang JD. Risk assessment of lung cancer and mesothelioma in people living near asbestos-related factories in Taiwan. Arch Environ Health. 1999;54(3):194-201.
13. Chen CW. Lingering effect: epidemiological information useful for risk assessment. Regul Toxicol Pharmacol. 2008;52(3):242-247.
14. Chen J. Estimated risks of radon-induced lung cancer for different exposure profiles based on the new EPA model. Health Phys. 2005;88(4):323-333.
15. Chen W, Yang J, Chen J, et al. Exposures to silica mixed dust and cohort mortality study in tin mines: exposure-response analysis and risk assessment of lung cancer. Am J Indust Med. 2006;49(2):67-76.
16. Cox LAT, Jr. Quantifying potential health impacts of cadmium in cigarettes on smoker risk of lung cancer: a portfolio-of-mechanisms approach. Risk Anal. 2006;26(6):1581-1599.
17. Cronin KA, Gail MH, Zou Z, et al. Validation of a model of lung cancer risk prediction among smokers. J Natl Cancer Inst. 2006;98(9):637-640.

18. Crump K. Modeling lung cancer risk from diesel exhaust: suitability of the railroad worker cohort for quantitative risk assessment. Risk Anal. 2001;21(1):19-23.
19. Deslauriers J. Should screening for lung cancer be revisited? J Thorac Cardiovasc Surg. 2001;121(6):1031-1032.
20. Fabianova E, Hettychova L, Koppova K, et al. Health risk assessment for inhalation exposure to arsenic. Cent Eur J Public Health. 2000;8(1):28-32.
21. Finkelstein MM. Silica, silicosis, and lung cancer: a risk assessment. Am J Indust Med. 2000;38(1):8-18.
22. Foy M, Deng L, Spitz M, et al. Chapter 11: Rice-MD Anderson lung cancer model. Risk Anal. 2012;32 Suppl 1:S142-150.
23. Hecht SS. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. Lancet Oncol. 2002;3(8):461-469.
24. Hirayama T. Non-smoking wives of heavy smokers have a higher risk of lung cancer: a study from Japan. 1981. Bull World Health Organ. 2000;78(7):940-942.
25. Kanne JP. Screening for lung cancer: what have we learned? AJR. 2014; Am J Roentgenol. 202(3):530-535.
26. Lee PN. Lung cancer and type of cigarette smoked. Inhal Toxicol. 2001;13(11):951-976.
27. Liu Y, Steenland K, Rong Y, et al. Exposure-response analysis and risk assessment for lung cancer in relationship to silica exposure: a 44-year cohort study of 34,018 workers. Am J Epidemiol. 2013;178(9):1424-1433.
28. Lubin JH. Estimating lung cancer risk with exposure to environmental tobacco smoke. Environ Health Perspect. 1999;107 Suppl 6:879-883.
29. Manser RL, Irving LB, Stone C, et al. Screening for lung cancer. Cochrane Database Syst Rev. 2004 (1):CD001991.
30. Marshall D, Simpson KN, Earle CC, et al. Economic decision analysis model of screening for lung cancer. Eur J Cancer. 2001;37(14):1759-1767.
31. Nerriere E, Zmirou-Navier D, Desqueyroux P, et al. Lung cancer risk assessment in relation with personal exposure to airborne particles in four French metropolitan areas. J Occup Environ Med. 2005;47(12):1211-1217.
32. Nilsson R. Environmental tobacco smoke revisited: the reliability of the data used for risk assessment. Risk Analy. 2001;21(4):737-760.
33. Park RM, Bena JF, Stayner LT, et al. Hexavalent chromium and lung cancer in the chromate industry: a quantitative risk assessment. Risk Analysis. 2004;24(5):1099-1108.
34. Park RM, Bena JF, Stayner LT, et al. Hexavalent chromium and lung cancer in the chromate industry: a quantitative risk assessment. Risk Anal. 2004;24(5):1099-1108.
35. Peres J. Lung cancer screening gets risk-specific. J Natl Cancer Inst. 2013;105(1):1-2.
36. Reynolds P. Epidemiologic evidence for workplace ETS as a risk factor for lung cancer among nonsmokers: specific risk estimates. Environ Health Perspect. 1999;107 Suppl 6:865-872.

37. Russi EW. Lung cancer screening has the potential to save lives, but shall we do it? Swiss Med Wkly. 2011;141:w13185.
38. Sagawa M, Usuda K, Aikawa H, et al. Lung cancer screening and its efficacy. Gen Thorac Cardiovasc Surg. 2009;57(10):519-527.
39. Schwartz AG. Lung cancer: family history matters. Chest. 2006;130(4):936-937.
40. Spitz MR, Etzel CJ, Dong Q, et al. An expanded risk prediction model for lung cancer. Cancer Prev Res. 2008;1(4):250-254.
41. Sturza J. A review and meta-analysis of utility values for lung cancer. Med Decis Making. 2010;30(6):685-693.
42. Swiatkowska B. [Occupational factors influencing lung cancer in women in epidemiological studies]. Medycyna Pracy. 2011;62(6):659-665.
43. Ulm K. Risk assessment between silica dust and lung cancer. Cancer Causes & Control. 2002;13(8):779-780; author reply 781-772.

EVIDENCE TABLE (EFFICACY OF DIFFERENT MODELS)

Evidence Table : LLP RISK PREDICTION MODEL
Question : What is the efficay/effectiveness of the LLP risk prediction model for lung cancer prediction?

Bibliographic Citation	1. Cassidy A, Myles JP, Tongeren MV, Page RD, Liloglou T, Duffy SW & Field JK. The LLP risk prediction model: An Individual Risk Prediction Model For Lung Cancer. British Journal of Cancer. 2008; 98: 270-276
Study Type/Methods	Case-Control Aim: 1. To provide a model that would estimate the absolute risk of lung cancer for a given individual – could be utilised for primary and secondary prevention 2. To include in the final model only variables that are readily available to primary care when patients present (not necessarily with suspected lung cancer), so that it could be applied in primary care setting to facilitate the referral of high-risk individual Both cases and controls were ineligible if they had previous cancer within 5 years of interview date (excluding melanoma) Tools/statistics Involve: i) Multivariable model was built up in 2 phases - All statistically significant covariates (P<0.05) in univariable analyses were included in multivariable model and backward stepwise regression was performed, those factors that losing their significance (P>0.05) in multivariable analyses were dropped - Factors that not significant in the univariable analyses were subsequently fitted to the multivariable model with adjustment for the remaining significant effects, to detect effects, which only seen when the major risk factors are accounted for ii) Pair wise interaction tests were conducted between all the risk factors in the final multivariable model to ensure that they did not modify each other's effects iii) Logistic model was converted to absolute risk iv) AUC analyses The final multivariable model was combined with age-standardised lung cancer incidence data to calculate absolute risk estimate
LE	II-2
Number of Patients & Patient Characteristic	Case-Control Population • 579 lung cancer cases and • 1157 age- and sex-matched Population- based controls Age 20 to 80 years 2 population controls per case were selected from registers of general practitioners in Liverpool and matched to lung cancer cases by year birth (± 2 years) and gender Lung cancer included cancer in any of topographic subcategories of code C34 according to International Classification of Diseases
Intervention	Liverpool Lung Project (LLP) Risk Model Risk Predictors: i. Medical history ii. Family history of cancer iii. Lifetime occupational history iv. History of tobacco consumption – smoking status, inhalation, type of cigarette smoked, number of cigarette smoked per day, age at start and end - Ever smoker = smoked ≥100 cigarettes is lifetime - Current smoker = participant who reported smoking 2 years prior to the date of interview - All period of consumption were defined and counted towards total exposure
Comparison	None
Length of Follow Up (If Applicable)	Recruited between 1998-2005

Outcome Measures/Effect Size	**Results** **Response rates** - Cases 58.3% and controls 61.5%: • Caucasians: 99% in both groups • Men: 61.7% (cases) and 61.6% (controls) • Proportion of ever smokers: 95.3% (cases) and 71% (controls) - Majority of lung cancer cases in the study population presented with non-small cell lung cancer (83.2%) **Univariable Analyses** <u>Epidemiology Risk Factors</u> - Significant differences in Panel of epidemiology risk factors i) History of lung cancer in 1st degree relative P = 0.04 ii) Prior diagnosis of pneumonia P = 0.001 iii) Occupational exposure to asbestos P < 0.0001 iv) Prior diagnosis of malignant tumour P < 0.0001 <u>Risk before and after adjustment:</u> - Significant increase in risk amongst individuals with a prior diagnosis of pneumonia in both before and after adjustment for smoking: i) Before adjustment: OR 1.62, 95% CI 1.21-2.17 ii) After adjustment: OR 1.70, 95% CI 1.21-2.39 - Participants with prior diagnosis of emphysema: • Significant increase in risk before adjustment OR 2.19, 95% CI 1.25-3.84 but • Not after adjustment OR 1.78, 95% CI 0.96-3.30 - No effect was present for prior asthma, bronchitis, and tuberculosis <u>Sex-specific analyses:</u> • Significantly elevated for males who had prior diagnosis of pneumonia OR 1.92, 95% CI 1.25-2.95 Not in females OR 1.30, 95% CI 0.73-2.29 • Women who had prior diagnosis of emphysema exhibited significantly increased lung cancer risk OR 2.72, 95% CI 1.70-3.70 but not in males OR 1.30, 95% CI 0.58-2.94 <u>Cancer History:</u> - Prior cancer : significantly increased lung cancer risk OR 2.18, 95% CI 1.39-3.42 after adjustment for age, sex and smoking • Skin cancer, associated with 2.2 fold increased lung cancer risk 95% CI 1.12-4.26 • Breast cancer, OR 4.81, 95% CI 1.43-16.15 <u>Relatives lung cancer history:</u> - Significant trend of increasing risk with numbers of affected relatives - No significant effect of family history of lung cancer in study population overall or in late-onset cases, regardless the age of affected relatives - However, substantial and statistically significant increase in risk where both the lung cancer case and the affected relative were diagnosed with lung cancer before the age of 60 years OR 4.89, 95% CI 1.47-16.25 - Significant elevated OR were observed in connection with an affected relative diagnosed before age 60 regardless of age-at-onset of the case OR 2.08, 95% CI 1.20-3.59 <u>Smoking</u> - Current smokers OR 13.15, 95% CI 8.43-20.50 were at higher risk than ex-smokers OR 5.72, 95% CI 3.71-8.82 -Fitting total years of smoking duration as a continuous covariate and in 10 and 20 year intervals revealed a steady increase in lung cancer risk - Steady increase in risk with increasing pack-years and average amount smoked - Significant dose-response effect was observed for the daily number of cigarettes (P<0.0001), smoking duration (P<0.0001) and smoking pack years (P<0.0001) - No association was found between smoking pipes or cigars and risk of lung cancer

Outcome Measures/Effect Size	Smoking Exposure -Significant increase in risk for those who reported ever exposure to spousal tobacco smoke OR 1.44, 95% CI 1.04-1.98 • Significant dose-response effect was observed with duration of exposure P = 0.01, with largest increase in risk in the highest exposed group corresponding to more than 50 years exposure OR 2.51, 95% CI 1.33-4.71 - Non-significant excess risk was observed for ever exposure to second hand smoke in the workplace OR 1.29, 95% CI 0.97-1.73 • However, when analysed by duration of exposure, a dose-response effect was not observed P = 0.83. - No evidence on elevated risk of lung cancer due to exposure to secondhand smoke from social sources Asbestos -High lifetime exposure prevalence to asbestos : overall risk of OR 1.88 95% CI 1.36-2.59 which was reduced after adjustment for occupational confounders to 1.51 95% CI 1.02-2.04 **Multivariable Analyses** Final Multivariable Logistic Regression Analyses -Significant increase risks of lung cancer in family history with lung cancer • Particularly high risk in those with relative aged under 60 at diagnosis of lung cancer; P = 0.01 -Significant increase risks of lung cancer in individual history of respiratory disease and other cancer: • Prior diagnosis of pneumonia; P = 0.002 • Prior diagnosis of cancer other than lung cancer; P = 0.005 -Significant increase risk of lung cancer in exposure to asbestos and smoking duration • Occupational exposure to asbestos; P < 0.001 • Duration of smoking; P < 0.001 **Absolute Risk** Absolute Risk of Lung Cancer within 5 year period -Non-smoker : Absolute risk for a man age 77 with family history of lung cancer (relative aged under 60 at diagnosis), a history of asbestos exposure and no other risk factors was 3.17% 95% CI 15.07-47.67 -Smoker: Absolute risk for a man with same risk factor +45 years smoking history, 28.68% 95% CI 15.07-47.67 (smoking history contributes to an approximate 9-fold increase in the 5 year absolute risk of lung cancer) **Area Under the Curve analyses (AUC)** Receiver Operating Characteristic Curve (ROC) - When model applied to case-control population AUC was 0.71 - A 10-fold cross validation of LLP risk prediction model produced and AUC statistic of 0.70 indicating good discrimination between cases and controls **Limitations of LLP Model** i. Absolute risks estimated for each combination of risk factors were based in relative risk derived from case-control study ii. Potential that recall and other information biases could influence the results The results presented in the paper suggest that LLP risk prediction model could predict approximately 2/3 lung cancer within 5 years screening only 30% of population If confirmed in validation studies, the LLP risk prediction model could provide individuals and healthcare professional with easily obtained estimate of lung cancer risk to guide discussions and decision regarding prevention and surveillance
General Comments	

Evidence Table : Individualized Risk Prediction Model for Lung Cancer in Korean Men
Question : What is the efficacy/effectiveness of the Korea risk prediction model for lung cancer prediction?

Bibliographic Citation	2. Park S, Nam BH, Yang HR, Lee JA, Lim H, Han JT, Park IS, Shin HR, & Lee JS. Individualized Risk Prediction Model for Lung Cancer in Korean Men. Plos ONE. 2013; 8(2): e54823
Study Type/Methods	Cohort Obj: To develop an individualized risk prediction model for lung cancer in Korean men using large population-based prospective study Development of the Risk Prediction Model - To identify the significant risk factors for lung cancer in 8 years - To select the best-fit risk prediction model for lung cancer, the author include variables that showed statistical significance at 0.10 level in univariable analysis or that were chosen from stepwise regression model Validation with: - C-statistic - Hosmer-Lemeshow (H-L) - Risk of developing lung cancer for each participant was calculated from prediction model - The average predicted probabilities were compared with actual lung cancer risk estimated by Kaplan-Meier approach • Prediction model is considered good when discrimination is >0.75
LE	II-2
Number of Patients & Patient Characteristic	1,309,144 Study Population and Data Collection - All men between 30 and 80 years who underwent health examination between 1996 and 1997 - Restricted to participants who were free of any cancer at baseline Questionnaire filled
Intervention	Risk Prediction model Risk Predictors: - Smoking exposure - Age at smoking initiation - Body mass index - Physical activity - Fasting glucose levels 5 categories of smoker: i. Never; <0.5 packs/day ii. Current; 0.5-0.99 packs/day Current ≥ 1 pack/day
Comparison	
Length of Follow Up (If Applicable)	8 years
Outcome Measures/Effect Size	**Results** **Response Rate** - 10,007 newly diagnosed lung cancer cases were observed during 8-year follow up - Mean age of the cohort was 45 years - 28.6% never smoke and 13.9% current smokers consuming ≥1 pack per day - Most participants alcohol consumer 84.6% - BMI within normal range 69.0% (18.5-24.9) - 12% had a family history of any cancer - 6%had fasting glucose levels >126mg/dL

Outcome Measures/Effect Size	**Univariable analysis:** • Older age, smoking, early age at smoking initiation, high alcohol consumption and low BMI were significantly associated with higher lung cancer risk • Family history of any cancer was not significantly related to lung cancer risk • High glucose levels were also associated with elevated lung cancer risk **Multivariable analysis:** • Current smokers with high cigarette consumption showed an approximately 4-fold elevated risk of developing lung cancer and there was significant increasing trend of lung cancer risk by amount smoked (p-value for trend <0.0001) • Alcohol consumption was no longer significant when it was included in the model simultaneously with smoking → was excluded from final model • Lean participants (BMI <18.5), 39% increased risk of developing lung cancer • Heavier participants had approximately 29% decreased risk compared with participants with normal BMI • Physical activity decreased lung cancer risk by about 5-13% • High fasting glucose levels (≥126mg/dL) significantly associated with lung cancer • Younger age of smoking initiation significantly associated with lung cancer risk – higher risk of lung cancer **Validation of Risk Prediction Model** i. C-Statistic - The risk prediction model showed excellent discrimination (C-statistic = 0.864, 95% CI 0.860-0.868) - The prediction model with only age and smoking variables also showed excellent discrimination (C-statistic = 0.861, 95% CI = 0.857-0.865 - The model fit was improved by including other covariates (age at smoking initiation, physical activity, BMI and fasting glucose level) - Performance of the developed model was tested on external validation dataset, the discrimination was excellent (C-statistic = 0.871, 95% CI=0.867-0.876) ii. Likelihood ratio - Final model included all significant variables (likelihood ratio test, X^2=442.14, df=11, p<0.0001) iii. Hosmer-Lemeshow - Discrimination of the model was excellent, calibration was rather limited (Hosmer-Lemeshow type X^2 test, p<0.001 **Potential Limitation of the Study** - No assessment of the effects of environmental or occupational risk factors on lung cancer - Information on family history of lung cancer was not available in the data used to develop the risk prediction model - No women included in the study – only men Not able to differentiate effect of smoking intensity among past smokers because no data available
General Comments	

Evidence Table : Risk Model for Prediction of Lung Cancer
Question : What is the efficacy/effectiveness of the risk prediction model for lung cancer prediction?

Bibliographic Citation	3. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, Shete S & Etzel CJ. A Risk Model for Prediction of Lung Cancer. J Natl Cancer Inst. 2007; 99:715-726
Study Type/Methods	Case Control For 3 groups i) Never smokers ii) Former smokers iii) Current smokers Statistical Analysis by - Student's t test - Hosmer-Lemeshow - Discriminate with AUC. AUC of 0.5 indicates chance prediction and statistic of 0.7 indicates good discrimination - CART analysis was used to evaluate higher order interactions in the training sets due to sample size restrictions Data collected: - Demographic characteristics - Smoking history, occupation, information about specific exposures at work or from hobbies, medical history, and family history of cancer - Number of cigarettes smoked per day, smoking duration and age at smoking initiation Risk Model Building - Variables that statistically significantly associated with lung cancer risk at 5% level in univariable analysis in 3 training sets were included in multivariable logistic regression analyses
LE	II-2
Number of Patients & Patient Characteristic	Case-control 1851- Case patients –lung cancer patients recruited from Thoracic Centre at University of Texas since July 1995 2001-Control subjects – healthy without prior history of cancer recruited from Kelsey Seybold clinics Control subjects were frequency matched to case patients by age, sex, ethnicity and smoking status 3 Groups i. Never smokers ii. Former smokers Current smokers
Intervention	Risk Prediction Model
Comparison	None
Length of Follow Up (If Applicable)	
Outcome Measures/Effect Size	**Results** **Response Rate** - No statistically significant differences in the distribution of case patients and control subjects by smoking status, current smokers (39.8%) and fewer former smokers (42.4%) among the case patients (36.9%) than among control subjects, 44.2% - Pack-years smoker – case patients, 51.9% and control subjects, 44.6 for an average of 36.1 yrs (±SD 12.5) in case patients compared with 32.7 yrs (±SD 13.0) for control subjects ($p<0.001$) - Daily smokers - case patients mean cigarettes per day = 28.1 (±SD 13.7), control subjects, 26.4 (±SD 14.4); $P \leq 0.001$

Outcome Measures/Effect Size	**Univariable Analysis** Never Smokers - 330 case patients and 379 control subjects; - Exposure to ETS (OR = 1.77, 95% CI 1.2 to 2.6) or dust (OR = 1.48, 95% CI 1.0 to 2.1); - Family history of any cancer in ≥2 first-degree relatives (OR=1.96, 95% CI1.3 to 2.9); Were all statistically significantly associated with lung cancer risk - Asthma associated with 1.43-fold increase in risk among never smokers but the increasing was not statistically significant (95% CI 0.9 to 2.2) Former Smokers - 784 in case patients, 884 in control subjects risk were statistically significantly elevated with • Exposure to ETS (OR=2.07, 95% CI 1.3 to 3.2), • dust (OR = 1.64, 95% CI1.3 to 2.0), • fumes (OR = 1.32, 95% CI 1.1 to 1.6) and • chemicals (OR = 1.25, 95% CI 1.0 to 1.5), • with history of emphysema (OR = 2.99, 95% CI = 2.2 to 4.0), • with family history in ≥2 relatives of any cancer (OR = 1.84, 95% CI 1.4 to 2.4) or • smoking related cancers (OR = 1.40, 95% CI 1.1 to 1.7) and • with history of hay fever (OR = 0.72, 95% CI 0.6 to 0.9) Current Smokers - 737 case patients, 738 control subjects - The risk factor similar to former smokers with statistically significant associations for • emphysema (OR = 2.69, 95% CI 2.0 to 3.6) • exposure to dust (OR = 1.67, 95% CI 1.4 to 2.1) • fumes (OR = 1.31, 95% CI 1.1 to 1.6) • chemicals (OR = 1.34, 95% CI 1.1 to 1.7) • asbestos (OR = 1.78, 95% CI 1.3 to 2.4) • family history in ≥2 relatives of any cancer (OR = 1.68, 95% CI 1.3 to 2.2) • family history in ≥2 relatives of any smoking-related cancer (OR = 1.58, 95% CI 1.3 to 2.0) • history of hay fever (OR = 0.62, 95% CI 0.5 to 0.8) • Smoking variables (age at smoking cessation for former smokers, and measure of smoking intensity, for both former and current smokers) were statistically significantly associated with lung cancer risk **Multivariable Risk Models (Table 1)** - Exposure to environmental, tobacco and smoke (ETS) and family history of any cancer were statistically significantly associated with lung cancer in never smokers - Among former and current smokers, lung cancer was statistically significantly associated with exposure to dust, no prior history of hay fever (as the risk conferring value of the variable), personal history of emphysema, family history of any cancer (for former smokers) or tobacco-related cancers (for current smokers) and smoking intensity (for current smokers) and age at smoking cessation (for former smokers) - Exposure to asbestos was statistically significantly associated with lung cancer in current smokers but not in former smokers - Smoking status specific risks model stratified by sex: • Former smokers (men) with age at smoking cessation and no prior hay fever were statistically significantly associated with lung cancer risk compared to women • Current smokers (men) with asbestos exposure was statistically significantly associated with lung cancer compared to women **Decision Trees of CART Models** • Among never smokers – exposure to ETS and family history of cancer were strongly associated with lung cancer risk • Among former smokers – history of emphysema is the strongest risk factor (OR = 4.55, 95% CI 3.0 to 6.8), dust exposure without history of emphysema (OR = 2.35, 95% CI 1.7 to 3.3) • Former smokers without emphysema or dust exposure, later age at smoking cessation was associated with 1.88-fold increase in risk (95% CI 1.4 to 2.5) and combination of dust exposure and family history of any cancer was associated with OR 3.41 (95% CI 2.2 to 5.30) • Among current smokers – history emphysema was the strongest risk factor for lung cancer (OR = 4.20, 95% CI = 2.9 to 6.2), whereas smoking intensity (≥37 pack yrs) was strongly associated with lung cancer risk among current smokers without emphysema (OR = 2.55, 95% CI 1.9 to 3.5) • Family-history of smoking-related cancers was associated with lung cancer in subjects with heavier smoking histories (OR = 3.09, 95% CI 2.0 to 4.7) as well as among those with lighter smoking histories and self-reported dust exposure (OR = 5.12, 95% CI 2.6 to 10.3) **Assessment of Model Fit / Model Validation** 3-phase validation process to assess the performance in validation sets of the models developed in the training sets

Outcome Measures/Effect Size	i. Hosmer-Lemeshow Risks model were well calibrated throughout entire range of probabilities as indicated by non-statistically significant Hosmer-Lemeshow goodness of fit tests statistic: • 0.777 for never smokers • 0.712 for former smokers and • 0.688 for current smokers ii. Area Under the Curve (AUC) AUC statistic obtained from validation sets • Low for never smokers (AUC = 0.57, 95% CI 0.47 to 0.66) • Low for current smokers (AUC = 0.58, 95% CI 0.52 to 0.64) • Slightly higher for former smokers (AUC = 0.63, 95% CI 0.58 to 0.69) iii. Threefold Cross-Validation Resulting concordance statistics calculated by threefold cross-validation of the combined dataset: • 0.59 in never smoker • 0.63 in current smoker • 0.65 in former smokers indicated that the models performed reasonably well in discriminating between case patients and control subjects - The authors also evaluated the fit of Bach et al. Model using their own data in ever smokers → AUC was only 0.57 (95% CI 0.56 to 0.59) → indicated that adding clinical and epidemiologic variables improves risk prediction **Estimation of Absolute 1-year Risk for Lung Cancer** - Did on 3 patients i. High risks ii. Moderate risks iii. Low risks **Risk Index / Ordinal Risk Index** Ordinals risk indices was developed from odds ratio derived from multivariable regression analyses for statistically significant risk factors from each model Based on CART Analysis the author established levels of risk for each smoking categories: Low risk group • Former Smokers (lowest risk group) - Validation set = true-negative rates were 66% (95% CI 59% to 72%) - Combined analysis = 66% (95% CI 62% to 70%) • Current Smokers (low risk group) - Validation set = true-negative rates were 65% (95% CI 55% to 75%) - Combined sets = 68% (95% CI 63% to 72%) High risk group • Former Smokers - Validation set = true-positive rates were 73% (95% CI 60% to 84%) - Combine set = 70% (95% CI 65% to 76%) in • Current Smokers - Validation set = true positive rates were 68% (95% CI 52% to 82%) - Combined set = 69% (95% CI 63% to 74%) The above risks scenario were used to classify subjects into 3 risk groups (low, intermediate and high) for each smoking stratum-specific model
General Comments	

Evidence Table : Bach. Model
Question : What is the efficacy/effectiveness of the LLP risk prediction model for lung cancer prediction?

Bibliographic Citation	4. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, Hsieh LJ & Begg CB. Variations in Lung Cancer Risk Among Smokers. Journal of National Cancer Institute. 2003; 95(6);470-478
Study Type/Methods	RCT Study on development and validation of the risk prediction model To estimate the absolute risk that an individual will be diagnosed with lung cancer within 10 years - Recursively estimated 10 yr lung cancer risk by cycling the two 1-year models for ten times - In each year the risk of lung cancer diagnosis and the risk of death in the absence of lung cancer were estimated Predictors - Age, sex, - Prior history of asbestos exposure, - Smoking durations, - Average amount smoked per day while smoking - Duration of abstinence from smoking for former smokers Derivation of 1-Year Models - Cox proportional hazards regression - Regression analysis – was divided into individual person-time periods Validation of 1-Year Models - Test of correlations with time and examination of residual plots - Concordance index after optimistic bias was reduced through 10-fold cross-validation
LE	1
Number of Patients & Patient Characteristic	CARET Cohort (Data patients from CARET multicenter randomized controlled study) (CARET was large RCT to look for lung cancer prevention) - 2 populations i. 14,254 heavy smokers (men, women aged 50-69yrs) – ≥20 pack-yrs of smoking exposure and were either current smokers or had quit within 6 yrs of enrolment ii. 4,060 asbestos-exposed men (aged 45-69 yrs, either current smokers or former smokers who had quit within 15 yrs of enrolment) who had either radiologic evidence of asbestos exposure or history of employment in trade that put them at high risk for asbestos exposure - Those 18,314 individuals were randomly assigned to receive either placebo or study drug (30mg/day beta carotene and 25000 IU/day retinyl palmitate - Randomization for the pilot study began in June 1985; followed by randomization for full-scale study in Jun 1989; study accrual ended in Sept 1994 - Intervention stopped in January 1996 after preliminary results revealed definitive evidence of no benefit and substantial evidence of possible harm - Study subjects were continually followed annually by mail with additional data collection on reported endpoints - Subjects included in the analyses were 18,172 individuals who had documented history of current or former smoking

	Multicentre (6) - Seattle - Baltimore - New Haven - Portland - San Francisco - Irvine
Intervention	
Comparison	
Length of Follow Up (If Applicable)	10 years (1985 through 1994)
Outcome Measures/Effect Size	**RESULTS** **Response Rate** - 18,172 had documented current or former smoking history (14,245 from heavy smoking cohort and 3,918 from asbestos cohort) - Subjects contributed an average of 13.6 observational intervals (median=13), with mean duration of 265 days per interval (median=200 days), inter-quartile range = 120-365 days) - February 25, 2002 subjects had been followed to an outcome of lung cancer for 168,343 person-yrs and outcome of death for 169,785 person-yrs - 1,070 of the subjects were diagnosed with lung cancer (incidence rate = 636 per 100,000 person-yrs) and 3175 of the subjects died (mortality rate = 1870 per 100,000 person-yrs) - Among the observed lung cancer cases, both distribution of histologic subtypes and survival distribution were consistent with national statistics: - 77% of cases were non-small-cell cancers, - 18% were small-cell cancers and - Median survival after diagnosis was 7.4 months **Lung Cancer Risk Model** - In the 1-year lung cancer risk model, the association between risks factors and lung cancer occurrence were consistent with those in previous reports for both continuous predictors and categorical predictors - The study drugs (beta-carotene and retinyl palmitate) increased risk of lung cancer to degree consistent with previously published data from CARET (HR =1.20, 95% CI = 1.06 to 1.25; P = 0.004) - History of asbestos exposure was associated with independent increase in lung cancer risk (HR = 1.24, 95% CI = 1.04 to 1.48; P = 0.02) - No statistical evidence that sex independently influenced lung cancer risk (HR = 0.94, 95% CI = 0.92 to 1.08; P = 0.41) **Cross Validated Concordance Index** - Cross-validated concordance index was 0.72 and the cross-validated calibration plot by risk deciles was consistent with excellent calibration - The model had a cross-validated concordance index of only 0.66 - This model cannot identify variations in risk between individuals of a particular age which can be quite large, based on individual smoking history - The observed rates of lung cancer across the risk deciles for the held-out site closely matched those that had been predicted by corresponding model derived from the 5 included sites
General Comments	

Evidence Table : COSMOS Trial
Question : What is the efficacy/effectiveness of the COSMOS Risk Prediction Model for lung cancer prediction?

Bibliographic Citation	5. Maisonneuve P, Bagnardi V, Bellomi M et al. Lung Cancer Risk Prediction to Select Smokers for Screening CT – a Model Based on the Italian COSMOS Trial. Cancer Prev Res. 2011; 14:1778-1789
Study Type/Methods	Nonrandomized control trial Aim- 1st model i. To develop a model based on epidemiologic and clinical risk factors to estimate the probability of individuals in a high-risk population being diagnosed with lung cancer → To stratify individuals and select those at higher risk for inclusion in screening programs 2nd model ii. To develop a 2nd model based on baseline CT findings in a screened population, combined with epidemiologic and clinical risks factor → to stratify individuals according to probability of being diagnosed with lung cancer at repeat screening scans Statistical Analysis *Assessment of Lung Cancer Risk in COSMOS Participants* • Lung cancer rates compared between categories using rate ratios and HR - Multivariable Cox Proportional regression modelling - Proportional hazard assumptions - Kaplan-Meier method to represent cumulative incidence of lung cancer - Log-rank test to compare lung cancer incidence between various categories of patients *Prediction of Lung Cancer Risk at 1st Screening Round* - Compared the frequency of lung cancers diagnosed in 1st year in COSMOS trial with frequency predicted by Bach Model - Bach model was evaluated for its ability to distinguish patients with a diagnosis of lung cancer from those without (discrimination) and its agreement with the frequency of lung cancers (calibration) in COSMOS - Hosmer-Lemeshow used to further assessed the calibration - Bach model was recalibrated and the original baseline risk h0 was replaced with baseline risk h0* recalculated from COSMOS dataset *Prediction of Lung Cancer Risk After 1st Screening Round* • Lung cancer cases diagnosed in the 1st screening were excluded • The observation for the model was started at the 2nd screening CT and continued up to date of lung cancer diagnosis or date of latest CT (non-cases) - Multivariable modelling (Multivariable Cox proportional hazard regression analysis) - Linear predictor from Bach model were used to avoid over fitting Net Reclassification Improvement (NRI) Index to evaluate individual risk predictions derived from regression model s for binary outcomes
LE	II-1
Number of Patients & Patient Characteristic	3,439 (66%) men and 1,764 (34%) women Median age 57 years Currents Heavy smokers 4,175 (80%) 1,028 (20%) stop smoked (0-10 yrs) Used data from ongoing COSMOS single-centre nonrandomized lung cancer screening trial, Northern Italy - Asymptomatic volunteers aged 50 yrs or older - Heavy smokers (≥20 pack-yrs) Still smoking or had stopped smoking less than 10 yrs and had not been diagnosed with cancer in previous 5 years
Intervention	
Comparison	
Length of Follow Up (If Applicable)	October 2004 to October 2005

Outcome Measures/Effect Size	**RESULTS** **COSMOS Trial** • 482 pts were recalled for repeat CT (recall rate 9.3%) • 75 (1.4%) for second repeat CT during the 1st year • 160 (3.1%) CT-PET scans were done • Total 525 (10.1%) pts being recalled for CT, CT-PET or both in the 1st year • 1st round Investigations found - 62 pts underwent surgery - 55 lung cancer detected - 7 false positive for benign lesions • 2rd round – 4,822 (92.7%) came back for 2nd scan • 3rd round – 4,582 (88.1%) came back for 3rd scan • 4th round – 4,383 (84.2%) came back for 4th scan (107 underwent surgery for presumed lung cancer, 22 false positive) • Overall 162 lung cancers detected - 116 (71.6%) adenocarcinoma - 19 (11.75) squamous cell carcinoma - 10 (6.2%) small cell carcinoma - 4 (2.5%) carcinoid tumour - 2 (1.2%) bronchoalveolar carcinoma - 1 (0.6%) large cell neuroendocrine tumour - 10 (6.2%) non-small cell types • At diagnosis - 115 (7.1%) stage I - 7 (4.3%) stage II - 25 (15.4%) stage III - 15 (9.3%) stage IV The resectability rate was 89% **Assessment of Lung Cancer Risk in COSMOS Participants** **Univariable Analysis** - 162 lung cancers were detected in 18,095 person years of observation from baseline to end of the 4th screening round – detection rate of 0.90 per 100 years - The detection rate was approximately constant over time and cancer cases usually diagnosed within 6 months of a screening CT - Detection rate was slightly higher in men (0.95) than women (0.78) – differences was not significant - Detection rate was slightly higher in current smokers (0.92) than former smoker (0.79) – differences was not significant - Detection rate was strongly (P<0.0001) dependent on age, increasing from 0.50 in those under 55 years at entry, to1.64 in those more than 65 years at entry - Lung cancer rate did not vary much with age at starting smoking, or years from stopping, but correlated strongly with duration of smoking and cigarette consumption - The rate doubled in those who smoked for 35 to 40 years compared with those who smoked for less than 35 years and was more than 6 times higher (RR=6.27; 95% CI 3.14-12.5) in those who smoked for more than 50 years compared with those who smoked for less than 35 years - Lung cancer rate was doubled in those who smoked more than 40 cigarettes per day than in those who smoked less than 20 per day (RR = 1.91; 95% CI 1.08-3.35) - NO other factors such as demographic data, lifestyle, BMI, fruit and vegetables consumption pattern, alcohol consumption and passive smoking were significantly associated with lung cancer risk - Asbestos exposure was non-significantly associated with the cancer, RR=3.05, 95% CI 0.422-22.4 - Lung cancer rate was high among those who reported history of COPD, RR=1.60; 95% CI=1.10-2.33 - Dyspnea associated with significantly greater risk of lung cancer compared with those who did not report dyspnea (RR = 1.39; 95% CI = 1.00-1.93) - Epidemiologic / clinical risks factor - FEF_{25-75} < 50% of predicted had significantly greater risk of lung cancer (RR = 2.03; 95% CI = 1.13-3.62) than those with $FEF_{25-75} \geq 80\%$ - FVC was unrelated to lung cancer - FEV1 < 90% of predicted had twice the risk of lung cancer compared with those with FEV1 > 90% or more of predicted (RR = 2.09; 95% CI = 1.34-3.26) **Multivariable analysis** - Age, smoking duration, number of cigarettes smoked and predicted FEV1 (90% cut-off) were independently associated with lung cancer risk - The author exclude 55 lung cancers detected during 1st screening round to assessed the extent of lung cancer developed; 4,596 participants involved in 2nd screening round

	- Visual evidence of emphysema on baseline CT doubled the risk (RR = 2.36; 95% CI = 1.59-3.49) of screening-detected lung cancer, - Presence of solid (RR = 2.00; 95% CI = 1.22-3.27), - Presence of partially solid (RR = 3.43; 95% CI = 1.91-6.16) and - Presence of non-solid (RR = 10.1; 95% CI = 5.57-18.5) non-calcified nodule also increased cancer risk - Nodule > 8mm at baseline had a ten-fold risk (RR = 9.89; 95% CI = 5.84-16.8) compared with smaller nodule (<5 mm) *Prediction of lung cancer risk at first screening round* - Age, smoking duration and number of cigarettes smoked per day were the main determinants of lung cancer risk - Duration of quitting and exposure to asbestos were not significantly associated with cancer - Bach model estimated that - 21 COSMOS participants would develop symptomatic lung cancer during 1st year - 55 lung cancers were detected in 1st screening round (standardized incidence ratio = 2.62; 95% CI = 1.97-3.41) • Incidence of lung cancer in COSMOS - Predicted by original Bach model; higher (Hosmer-Lemeshow X^2 test = 70.7; P<0.0001) - Predicted by recalibrated Bach model; (Hosmer-Lemeshow X^2 test = 6.2; P = 0.63) *Prediction of lung cancer risk after first screening round* • Emphysema, nodule type and nodule size were strongly influenced the risk of being diagnosed with lung cancer • Multivariable model - C-index 0.744 after incorporating all the CT variables (Model A) - C-index 0.747; P for difference = 0.87 Discriminatory ability of the model did not improve when FEV1 (<90% vs ≥90% of predicted) was incorporated (Model B) - C-index 0.763; P for difference from previous model = 0.11 discriminatory ability of the model increased after incorporating information about background risk of each participant based on Bach Model (Model C) Development of Prediction Tools • Risk assessment calculator – develop by the author to estimate risk of an individual being diagnosed with lung cancer at screening entry, based on recalibrated Bach model
General Comments	

EVIDENCE TABLE (VALIDATION STUDY)

Evidence Table : LLP RISK PREDICTION MODEL (VALIDATION)
Question : What is the validation of the LLP risk prediction model?

Bibliographic Citation	1. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A & Field JK. 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. Annual International Medicine. 2012;157(4):242-250
Study Type/Methods	Case-Control For Validation study Obj: To evaluate the discrimination if the LLP risk prediction model and demonstrate its predicted benefit for stratifying patients for CT screening by using data from 3 independent studies from Europe and North America Statistical Analysis - Performance of LLP risk prediction model was assessed by i) Measuring discriminative accuracy ii) Decision curve Relative utility curve analysis
LE	II-2
Number of Patients & Patient Characteristic	Participants in i) European Early Lung Cancer (EUELC) case control (LLCC) ii) Harvard case-control studies (LLCC) iii) LLP population-based prospective cohort study (LLPC)
Intervention	LLP risk prediction model
Comparison	
Length of Follow Up (If Applicable)	
Outcome Measures/Effect Size	Distribution of Participant Characteristics EUELC - Mostly men Harvard Study - Different sex distribution Distribution of age, smoking duration, family history and asbestos exposure followed similar patterns in those studies particularly for case pts LLPC - 420 of 7652 participants (6% of the cohort) developed lung cancer over an average follow-up of 8 years (median 7 years) - Lung cancer rates were slightly higher in men than in women - Lung cancer rate higher in participants with history of pneumonia than in those without - Lung cancer rate higher approximately 3 times higher in persons with history of cancer than in those without - Lung cancer rates also increased with greater age and longer smoking duration Performance of the LLP Model • Risk Distribution - Individual absolute risks were lower for control participants than for case patients - Most risk greater than 2.5% were predicted for patients with cancer - One half of disease-free patients had absolute risks less than 1% - Median 5 year predicted values were substantially lower for control participants than for case pts indicating good separation of summary values for pts with and without cancer • Discrimination - LLP model had higher discriminative ability across the data sets than using smoking or family history of lung cancer - LLP model had modest discrimination in EUELC data set (AUC 0.67 [CI 0.64 to 0.69]) - LLP model had good discrimination in both Harvard (AUC 0.76 [95% CI 0.75 to 0.78]) and LLPC (AUC 0.82 [CI 0.80 to 0.85]) data sets - AUC for smoking duration, the strongest of the risk factors was 0.63, 0.74, and 0.72 in EUELC, Harvard and LLPC data sets respectively - LLPC risk model had moderate overall calibration and improved accuracy at higher values of predicted risks

Outcome Measures/Effect Size	Assessment for Potential for Clinical Application - LLP model's sensitivity, specificity and estimated net benefit at threshold of 2.5%, 5.0% and 10% predicted absolute - Positive net indicates that the model had greater net benefit than screen-all strategy - At a threshold of 5 % absolute risk, the model achieved a higher proportion of true-positive classifications than a screen-all strategy (2.3% higher for the LLPC data and 3% higher for EUELC data) at the same proportion of false-positive classifications Refer Appendix 1 - Panels A to C compare the net benefits of using the LLP risk prediction model, the 2 strongest risk factors (smoking duration and family history), or the extreme strategies of screening everyone or no one - LLP risk prediction model has greater net benefit than all alternative strategies at thresholds of absolute risk, ranging from 3% to 15% - LLP risk prediction model performs well relative to the strong predictor if smoking duration which was most often used to stratify high-risk persons for lung cancer CT screening - For LLPC data, receiver-operating characteristic curve showed a moderate increase in discrimination over smoking when using LLP risk prediction model - For relevant risk thresholds at a probability of disease greater than 0.05, the relative utility curve showed moderately higher predicted net benefit, relative to perfect prediction for the LLP risk prediction model than smoking duration - For EUELC and Harvard data, increase in predicted net benefit for LLP risk prediction model versus smoking duration at relevant high-risk threshold was smaller *Discussion* - LLP risk prediction model has a good ability to distinguish persons who will not develop lung cancer by using predicted 5 year absolute risk - LLP model also seems to be reasonably well-calibrated at high predicted risks and performs better than smoking duration or family history as a tool for deciding which persons to screen for lung cancer LLP risk prediction model also unifies smoking duration, other important risk factors for lung cancer, and incidence data from cancer registries, thereby combining benefit of each to provide accurate and diverse predicted risks for smokers and non-smokers
General Comments	

LUNG CANCER RISK PREDICTION MODEL FOR NATIONAL HEALTH RISK ASSESSMENT MODULE

KEMENTERIAN KESIHATAN MALAYSIA

Malaysia Health Technology Assessment Section (MaHTAS)
Medical Development Division, Ministry of Health Malaysia,
Level 4, Bock E1, Complex E, Percint 1,
Federal Gorverment Adminnistrative Centre
65290, Putrajaya, Malaysia

Tel: 03-88831246/47
Fax: 03-8883 1230

ISBN 978-967-0769-57-8