Protocol and Rationale for the International Lung Screening Trial

Kuan Pin Lim[1,2*], Henry Marshall[3*], Martin Tammemägi[4], Fraser Brims[1,5], Annette McWilliams[2,6], Emily Stone[7,8], Renee Manser[9,10,11], Karen Canfell[12,13,14], Marianne Weber[12,13], Luke Connelly[15,16], Rayleen V. Bowman[3], Ian A. Yang[3], Paul Fogarty[17], John Mayo[18], John Yee[19,20], Renelle Myers[19,21,22], Sukhinder Atkar-Khattra[22], David C. L. Lam[18], Antoni Rosell[23], Christine D. Berg[24], Kwun M. Fong[3], and Stephen Lam[19,22]; on behalf of the ILST (International Lung Screening Trial) Investigators

[1]Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia; [2]Department of Respiratory Medicine, Fiona Stanley Hospital, Murdoch, Western Australia, Australia; [3]University of Queensland Thoracic Research Centre and Department of Thoracic Medicine, The Prince Charles Hospital, Chermside, Queensland, Australia; [4]Department of Health Sciences, Brock University, St. Catharines, Ontario, Canada; [5]Curtin Medical School, Faculty of Health Sciences, Curtin University, Bentley, Western Australia, Australia; [6]Faculty of Health and Medical Sciences, The University of Western Australia, Nedlands, Western Australia, Australia; [7]Department of Thoracic Medicine, St. Vincent's Hospital, Darlinghurst, New South Wales, Australia; [8]Kinghorn Cancer Centre, Darlinghurst, New South Wales, Australia; [9]Department of Respiratory Medicine and [10]Department of Medicine, Royal Melbourne Hospital, University of Melbourne, Parkville, Victoria, Australia; [11]Haematology and Medical Oncology, Peter MacCallum Cancer Centre, Parkville, Victoria, Australia; [12]Cancer Research Division, Cancer Council New South Wales, Sydney, New South Wales, Australia; [13]Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia; [14]Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales, Australia; [15]Centre for the Business and Economics of Health, The University of Queensland, Woolloongabba, Queensland, Australia; [16]Dipartimento di Sociologià e Diritto dell'Economia, Universita di Bologna, Bologna, Italy; [17]Respiratory Department, Epworth Eastern Hospital, Box Hill, Victoria, Australia; [18]Department of Medicine, University of Hong Kong, Hong Kong; [19]Department of Medicine, The University of British Columbia, Vancouver, British Columbia, Canada; [20]Radiology Department and [21]Thoracic Surgery, Vancouver General Hospital, Vancouver, British Columbia, Canada; [22]Department of Integrative Oncology, BC Cancer Research Centre, Vancouver, British Columbia, Canada; [23]Thoracic Institute–German Trias i Pujol University Hospital, Barcelona, Spain; and [24]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland

ORCID IDs: 0000-0001-5362-3517 (K.P.L.); 0000-0002-9626-8014 (H.M.); 0000-0002-6725-7535 (F.B.).

Abstract

Rationale: The NLST (National Lung Screening Trial) reported a 20% reduction in lung cancer mortality with low-dose computed tomography screening; however, important questions on how to optimize screening remain, including which selection criteria are most accurate at detecting lung cancers and what nodule management protocol is most efficient. The $PLCO_{m2012}$ (Prostate, Lung, Colorectal and Ovarian) Cancer Screening Trial 6-year and PanCan (Pan-Canadian Early Detection of Lung Cancer) nodule malignancy risk models are two of the better validated risk prediction models for screenee selection and nodule management, respectively. Combined use of these models for participant selection and nodule management could significantly improve screening efficiency.

Objectives: The ILST (International Lung Screening Trial) is a prospective cohort study with two primary aims: *1*) Compare the accuracy of the $PLCO_{m2012}$ model against U.S. Preventive Services Task Force (USPSTF) criteria for detecting lung cancers and *2*) evaluate nodule management efficiency using the PanCan nodule probability calculator-based protocol versus Lung-RADS.

Methods: ILST will recruit 4,500 participants who meet USPSTF and/or $PLCO_{m2012}$ risk $\geq$1.51%/6-year selection criteria. Participants will undergo baseline and 2-year low-dose computed tomography screening. Baseline nodules are managed according to PanCan probability score. Participants will be followed up for a minimum of 5 years. Primary outcomes for aim 1 are the proportion of individuals selected for screening, proportion of lung cancers detected, and positive predictive values of either selection criteria, and outcomes for aim 2 include comparing distributions of individuals and the proportion of lung cancers in each of three management groups: next surveillance scan, early recall scan, or diagnostic evaluation recommended. Statistical powers to detect differences in the four components of primary study aims were $\geq$82%.

Conclusions: ILST will prospectively evaluate the comparative accuracy and effectiveness of two promising multivariable risk models for screenee selection and nodule management in lung cancer screening.

Clinical trial registered with www.clinicaltrials.gov (NCT02871856).

Keywords: lung cancer; screening; low-dose CT; nodules; protocol

(Received in original form February 2, 2019; accepted in final form January 30, 2020)

*Co–first authors.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Ann Am Thorac Soc Vol 17, No 4, pp 503–512, Apr 2020

DOI: 10.1513/AnnalsATS.201902-102OC
Internet address: www.atsjournals.org

Background

Globally, in 2018, lung cancer caused an estimated 1.76 million deaths (1). This cancer can be detected via systematic screening of high-risk individuals: in 2011, the U.S. NLST (National Lung Screening Trial) reported a 20% reduction in lung cancer mortality with low-dose computed tomography (LDCT) screening compared with screening with chest radiography (2). The NLST eligibility criteria were based on age (55 to 74 yr) and smoking history (⩾30 pack-years, current smokers or former smokers who had quit ⩽15 yr ago).

On the basis of these findings and the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET) modeling, the U.S. Preventive Services Task Force (USPSTF) recommends annual screening in high-risk individuals using NLST criteria, except age was extended to 80 years (3–5). The Canadian Task Force on Preventive Health Care also recommended screening according to NLST criteria for three consecutive years, but the costs of screening are not currently covered (6). Other countries including Australia (7) do not currently endorse screening due to important knowledge gaps including uncertainty in the optimum selection criteria, best recruitment/screening uptake strategies, the action thresholds for early recall CT imaging study, positron emission tomography (PET)/CT or biopsy, false-positive rates and screening-related harms, demonstration of adequate cost-effectiveness, and the translation of U.S. findings to their own healthcare settings (8–10).

The International Association for the Study of Lung Cancer CT screening Task Force has highlighted six areas for further work including identification of high-risk individuals for screening and guidelines for work-up of screen-detected intermediate nodules (10)

The selection of the target population and management of indeterminate nodules are the most fundamental issues to address, given that most screenees do not harbor lung cancer, and most detected nodules are benign. These issues have important downstream effects on costs and cost-benefit, risk of harms, and, ultimately, the feasibility of screening in non-U.S. healthcare settings.

Optimal Selection of High-Risk Populations

Lung cancer screening is most effective when applied to a high-risk population—i.e., maximizing cancer detection and minimizing false positives, cost of work-up, and screen-related adverse events (4, 11–13). NLST reported a number needed to screen to prevent one death from lung cancer of 320 (2).

However, lung cancer risk is heterogenously distributed in populations selected by NLST criteria. For example, Kovalchik *et al.* (11) described striking variations in numbers of lung cancer deaths prevented when NLST participants were stratified to quintiles of risk using an absolute risk prediction model for lung cancer death: 0.2 deaths prevented per 10,000 person-years in the lowest-risk quintile compared with 12.0 deaths prevented per 10,000 person-years in the highest-risk quintile (11). Other work has shown improvements in number-needed-to-screen to prevent one lung cancer death (255 in highest tertile compared with 963 in middle tertile), and in the NLST cost-effectiveness analysis, the cost per quality-adjusted life year varied from $52,000 U.S. dollar (USD) to $169,000 USD between the highest- and lowest-risk quintiles, respectively (12, 14).

Detailed risk assessment using multivariable regression modeling, coupled with risk-based entry criteria, could be advantageous; the model can incorporate other known risk factors aside from age and smoking exposure, better defining risk and avoiding screening and its attendant risks and costs in lower-risk subpopulations. The $PLCO_{m2012}$ (Prostate, Lung, Colorectal and Ovarian) Cancer Screening Trial model was developed using the large prospective PLCO dataset of a general population not limited to high risk of lung cancer (13, 15, 16). In post hoc analyses, $PLCO_{m2012}$ identified the individuals with lung cancer (i.e., the high-risk population) more effectively than either NLST or USPSTF selection criteria, improving sensitivity, positive predictive value, and specificity (12, 13, 17).

The optimal threshold of lung cancer risk for screening selection is unknown. The PanCan (Pan-Canadian Early Detection of Lung Cancer) study used the $PLCO_{m2007}$ model with a ⩾2%/6-year eligibility threshold (18). The cumulative incidence of lung cancer (6.5%) was significantly greater than that observed in the NLST study (18). On the basis of NLST data and validated in the PLCO dataset, the optimal risk threshold for $PLCO_{m2012}$ is 1.51% over 6 years. This is equivalent to the 65th percentile of risk in NLST and the point at which lung cancer mortality in the LDCT arm was consistently lower than that in the chest X-ray arm. This threshold value identified 80% of ever-smokers who developed lung cancer (13).

Optimal Management of Screen-detected Pulmonary Nodules

LDCT screening is highly sensitive for nodule detection; pulmonary nodules' prevalence varies between 22% and 51%, depending on size cutoff for nodule reporting, CT parameters, and study population (2, 18–24). The vast majority of nodules are benign, but may contribute significantly to follow-up costs and may incur unnecessary interventions including surgery.

There is no universally accepted protocol for nodule classification and subsequent management. NLST used a simple axial linear measurement to classify all noncalcified nodules with a maximum diameter of ⩾4 mm as a positive scan. The proportion of positive screening scans was 24.2% in the LDCT arm over all three rounds, and 96.4% of these were not cancer (2). In the NLST, the false-positive rate (1 minus specificity) at baseline scan was 26.6%, at T_1 it was 27.4%, and at T_2 it was 16.1%. Nodule management

A complete list of ILST (International Lung Screening Trial) investigators may be found before the beginning of the REFERENCES.

Supported by the National Health and Medical Research Council, Cancer Australia, Western Australia Cancer and Palliative Care Network, Terry Fox Research Institute, BC Cancer Foundation, VGH-UBC Hospital Foundation, and The Epworth Medical Foundation.

Correspondence and requests for reprints should be addressed to Stephen Lam, M.D., F.R.C.P.C., BC Cancer Research Centre, Rm 10-111, 675 West 10th Avenue, Vancouver, B.C., V5Z 1L3 Canada. E-mail: slam2@bccancer.bc.ca.

guidelines have been previously based on expert opinion and cohorts of clinical patients with high proportions of lung cancer that are unrepresentative of screening cohorts (25–28).

Strategies to reduce the burden of false-positive scans include volumetric measurement, larger axial size threshold, and biennial screening following low-risk (negative) scans. However, none of these strategies have yet been shown to reduce mortality. Recognizing this issue, the American College of Radiology developed the Lung-RADS classification system (29, 30). Retrospective analyses of Lung-RADS have shown reductions in baseline false-positive rates to 10.6–12.8% (31–33). Of note, an updated version of Lung-RADS V1.1 was recently published after the ILST (International Lung Screening Trial) had commenced (29).

An alternative nodule management approach is to estimate cancer risk using regression modeling incorporating nodule and participant variables. The PanCan (or Brock University) nodule malignancy probability calculator (34) was developed from trial data in which individual nodules were longitudinally evaluated. It pertains to nodules detected on baseline scans that accounted for 75% of the lung cancers found in the first 5 years (18). It has superior sensitivity and specificity compared with the Lung-RADS classification in retrospective studies (35–39). For selected nodules, it is recommended by Lung-RADS and the British Thoracic Society nodule guidelines (40, 41), but it has not been prospectively tested.

We seek to prospectively compare the $PLCO_{m2012} \geq 1.51\%$/6-year threshold against the USPSTF selection criteria in terms of proportion of lung cancer detected by either criteria and to prospectively establish the effectiveness of the PanCan nodule malignancy risk model in managing nodules detected at baseline screening LDCT.

Methods

The ILST (clinicaltrials.gov identifier: NCT02871856) is an international, multicenter prospective cohort study with nine recruitment sites in Australia, Canada, Hong Kong, and Spain. ILST will recruit 4,500 ever-smokers between the age of 55 to 80, who will undergo two scheduled LDCTs at baseline and at 2 years. Participants will be followed up for a minimum of 5 years.

The ILST has two primary aims: *1*) compare the predictive accuracies of the $PLCO_{m2012}$ and the USPSTF screening selection criteria and *2*) evaluate a nodule management protocol based on the use of the PanCan nodule probability calculator on baseline screening LDCT.

For primary aim 1, we hypothesize that the predictive accuracy of the $PLCO_{m2012}$ criteria is greater than that of the USPSTF criteria for selecting individuals for screening who are subsequently diagnosed with lung cancer. For the two criteria, comparisons of the following will be made (see Table 1 for calculation guide):

1. The proportion of individuals selected for screening
2. The proportion of lung cancers detected (or, equivalently, McNemar's odds ratio)
3. The positive predictive values

Secondary aims related to primary aim 1 include the following: calculating and comparing the sensitivities in a subset analysis. Calculation of sensitivities requires the "a" in Table 1. Only those who are positive by either criteria get screened. All sites will collect "A," but only selected study sites are following and collecting information on "a" and "A" in Table 1, thus allowing estimation of "a" in a subset analysis.

There will be comparison of the proportion of lung cancers detected and positive predictive values, when the $PLCO_{m2012}$ threshold for selection is adjusted to include the same number of individuals selected by the USPSTF criteria. (Recent evidence indicates that because populations differ, and smoking patterns have changes, calibrations of risk model selection thresholds require reassessment and readjustment when such models are applied in different populations.)

Primary aim 2 relates to management of screen-detected nodules found on baseline LDCT, and for the purpose of this study, screening results are grouped into three categories: *1*) no or very low risk nodule(s) ("negative") to be followed by the next regular planned surveillance scan (e.g., CAT1 on ILST protocol; see Figure 2), or "positive," to be followed by *2*) early recall scan (CATs 2/3), or *3*) clinical investigation (CATs 4/5). We hypothesize that compared with the Lung-RADS nodule management system, the PanCan nodule malignancy model-based management protocol will have fewer positive results, while detecting an equivalent or higher number of lung cancers in the positive scan results group. For a management protocol to be considered superior, there should be more individuals in group 1, thus reducing systematic costs and risks to individuals, but simultaneously be accompanied by high numbers of lung cancers detected in groups 2 and 3. Because there is no single metric which combinatorially summarizes both distribution of screening results and number of lung cancers detected in the positive results category, we will interpret superiority of a management protocol if both the distribution of screening results (fewer positive scans requiring early recall or investigations) and absolute number and proportion of lung cancers in the positive results category are equivalent or higher.

Ancillary studies will address other screening-related questions of interest including: *1*) associations between outdoor and household air pollution and lung cancer, *2*) the impact of screening on the quality of life and health status of screening participants at a variety of time points, *3*) estimated healthcare economic costs of LDCT screening (in the Australian and Canadian settings), *4*) the utility of spirometric diagnosis of chronic obstructive pulmonary disease as a risk-stratification tool in lung cancer screening, *5*) blood-based biomarkers associated with lung cancer, *6*) the potential role of computer-aided detection (CAD) software to improve radiologist reporting time and quality assurance, *7*) optimal smoking cessation strategies within a lung cancer screening program, *8*) the effectiveness of different recruitment methods in lung cancer screening, and *9*) the implications of incidentally detected abnormalities such as osteoporosis, coronary artery calcification, and interstitial lung abnormalities.

Study Schema and Inclusion–Exclusion Criteria

The study schema up to baseline screening LDCT (including inclusion and exclusion criteria) is summarized in Figure 1.

Recruitment

The recruitment sites and planned number of participants from each site are as follows: The BC Cancer Agency in Vancouver, British Columbia, Canada (2,000), The Prince Charles Hospital in Brisbane, Australia (500), Fiona Stanley Hospital and

Table 1. Analytic schema with cross-stratification of participants by $PLCO_{m2012}$ and USPSTF criteria eligibility for study aim 1

Number of Participants by Screening Eligibility			
	USPSTF −ve	**USPSTF +ve**	**Total**
$PLCO_{m2012}$ −ve	A*	B	A* + B
$PLCO_{m2012}$ +ve	C	D	C + D
Total	A* + C	B + D	T
Lung Cancers by Screening Eligibility			
	USPSTF −ve	**USPSTF +ve**	**Total**
$PLCO_{m2012}$ −ve	a*	B	a* + b
$PLCO_{m2012}$ +ve	C	D	c + d
Total	a* + c	b + d	t

A = Number of individuals who are $PLCO_{m2012}$ (Prostate, Lung, Colorectal and Ovarian) Cancer Screening Trial −ve and U.S. Preventive Services Task Force (USPSTF) −ve*. B = Number of individuals who are $PLCO_{m2012}$ −ve and USPSTF +ve. C = Number of individuals who are $PLCO_{m2012}$ +ve and USPSTF −ve. D = Number of individuals who are $PLCO_{m2012}$ +ve and USPSTF +ve. T = Total number of individuals low-dose CT-screened (B + C + D).
a = Number of lung cancers in $PLCO_{m2012}$ −ve and USPSTF −ve individuals*. b = Number of lung cancers in $PLCO_{m2012}$ −ve and USPSTF +ve individuals. c = Number of lung cancers in $PLCO_{m2012}$ +ve and USPSTF −ve individuals. d = Number of lung cancers in $PLCO_{m2012}$ +ve and USPSTF +ve individuals. t = Total number of lung cancers detected in low-dose CT-screened individuals (b + c + d).
* = Prespecified subgroup analysis on screening ineligible ILST (International Lung Screening Trial) participants from selected study sites. Proportions of individuals selected for screening = (C + D)/T versus (B + D)/T. Proportions of lung cancers detected = (c + d)/(b + c + d) versus (b + d)/(b + c + d). McNemar's odds ratio = c/b. Positive predictive values = (c + d)/(C + D) versus (b + d)/(B + D).

Sir Charles Gairdner Hospitals in Perth, Australia (500 combined), St. Vincent's Hospital in Sydney, Australia (500), The Royal Melbourne Hospital in Melbourne, Australia (500), Epworth Eastern Hospital in Box Hill, Australia (100), Queen Mary Hospital in Hong Kong (400), and German Trias i Pujol University Hospital, Barcelona, Spain (400). ILST will recruit a minimum of 4,500 participants.

Participants will be recruited using a variety of strategies depending on the recruitment site, including via primary care physicians, media advertisements, electoral roll mail-out invitations, and targeted invitations identified from primary care databases.

Potential participants will undergo an initial eligibility assessment including $PLCO_{m2012}$ risk estimation and USPSTF criteria check using a web-based questionnaire. Eligible participants providing informed consent are booked for baseline LDCT scan. Current smokers are offered cessation advice and invited to participate in the national Quitline program using an opt-out approach.

In selected sites, ineligible individuals or eligible individuals who decline to enroll in the study after initial expression of interest are invited to a 5-year health outcomes follow-up study (annual questionnaires and lung cancer registry linkage).

Radiology Protocol

LDCT scans (120 kV, 40–50 mA, pitch 1:0, and gantry rotation time ⩽0.5 s) will be performed on multidetector (⩾16 row) machines with minimum section collimation of ⩽1 mm from lung apices to the adrenals. Low–radiation dose acquisitions (⩽1.5 mSv effective dose) are obtained using reduced mA and a minimum gantry rotation time. The CT dose index volume will be ⩽3.0 mGy (32 cm) for a "standard" person (170 cm, 70 kg, body mass index [BMI] 24). Supine noncontrast images will be acquired in a single inspiratory breath hold with arms overhead. A high–spatial frequency image reconstruction algorithm will be used for lung parenchyma; an intermediate spatial frequency algorithm will be used for mediastinal structures to minimize image noise. Annual calibration using the Radiological Society of North America Quantitative Imaging Biomarker Alliance phantoms will be performed and results subjected to site medical physicist review.

LDCT Reporting

Experienced radiologists (⩾300 CT chest readings in the last 3 yr) use a standardized reporting protocol for all findings. Ancillary analysis using computer-automated detection (Veolity 1.2 system, MeVis Medical Solutions AG, Germany) is being undertaken as a substudy at some sites.

An indeterminate nodule (solid, semisolid, nonsolid) is defined as a nodule ⩾3 mm and <30 mm in average axial diameter. Perifissural nodules and completely calcified nodules are regarded as benign. Nodules detected at baseline scan will be categorized using the PanCan nodule malignancy probability calculator and managed in accordance to the management protocol (Figure 2). In the case of multiple nodules, the nodule with the highest PanCan nodule probability score is used to determine categorization.

Nodule Management

A management algorithm for screen-detected nodules is presented in Figure 2.

Participants with normal baseline LDCT scans (no nodules and no other abnormality suspicious of malignancy) or low-risk nodules with PanCan nodule probability calculator score of <1.5% (CAT1) will have a repeat LDCT in 24 months.

Participants with a PanCan nodule calculator score ⩾1.5% to <6% (CAT2) will have a repeat LDCT annually for up for 2 years (solid nodules) and up to 5 years (subsolid nodules) in accordance with current clinical practice.

Participants with a PanCan nodule calculator score of 6 to <30% (CAT3) will have short-term interval LDCT in 3 months or immediate clinical investigation. CAT3 results are subdivided into CAT3a (PanCan score 6 to <10%) and CAT3b (PanCan score 10 to <30%)—CAT3a by default have an interval LDCT in 3 months, whereas CAT3b may be considered for 3-month interval LDCT or immediate clinical investigation at the discretion of the treating physician. CAT3 participants with a 3-month LDCT that does not show interval growth revert to annual LDCT.

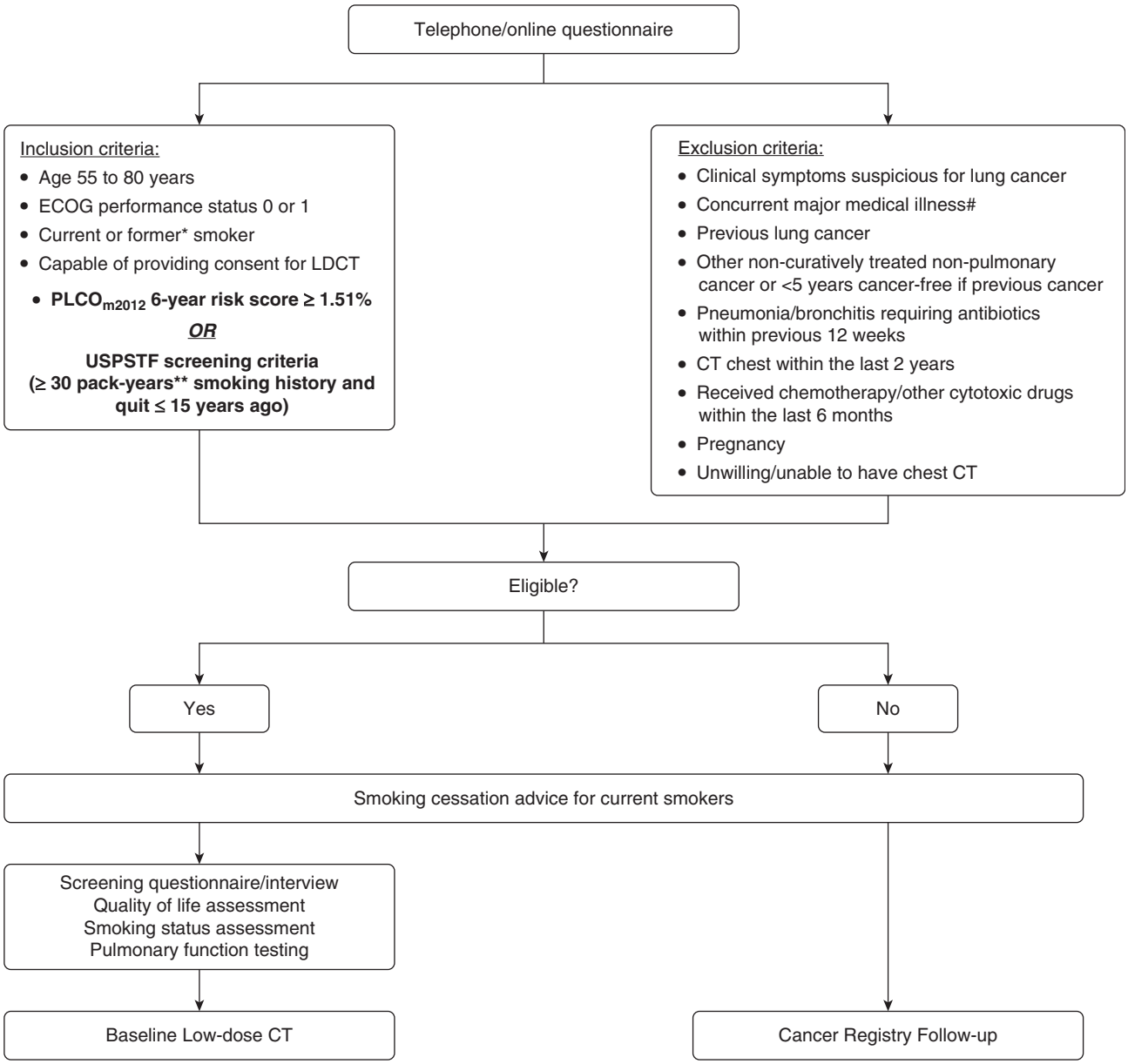

Figure 1. Participant recruitment flow diagram up to baseline screening low-dose computed tomography. * = Former smoker is defined as one who has stopped smoking for ⩾1 year. ** = Pack-year is defined as number of packs of cigarettes smoked per day multiplied by the number of years smoked. (If participant ceased smoking for ⩾6 mo interval, the time will be subtracted from the total duration of smoking in 0.5-year increments.) # = Any medical condition that, in the investigator's opinion, may jeopardize the subject's safety during participation in the study or mean that the subject is unlikely to benefit from screening due to shortened life expectancy; and may include severe cardiac disease (e.g., unstable angina, congestive cardiac failure), acute or chronic respiratory failure, home oxygen therapy for advanced lung disease, bleeding disorders, etc. CT = computed tomography; ECOG performance status = Eastern Cooperative Oncology Group (51); LDCT = low-dose computed tomography; PLCO = Prostate, Lung, Colorectal and Ovarian; USPSTF = U.S. Preventive Services Task Force.

Participants with a PanCan nodule calculator score of ⩾30% (CAT4) or have other findings suspect for lung cancer (CAT5) are considered suspicious for lung cancer and will be reviewed by the site clinician for immediate clinical investigation. CAT5 findings include mass lesion of noninfectious etiology, mediastinal or hilar lymphadenopathy irrespective of nodule size, or endobronchial nodule.

Significant growth of pulmonary nodules on two consecutive scans is considered suspicious for lung cancer, and participants will be evaluated clinically. Significant interval growth is defined as any of: *1*) an increase of >1.5 mm in mean diameter, *2*) volume change of ⩾100% for nodules <5 mm, volume change of ⩾30% for nodules 5 to <10 mm, volume change of ⩾20% for nodules ⩾10 mm, *3*) volume doubling time 30–400 days for nodules <300 mm^3, or *4*) the development of a solid core of ⩾6 mm in a subsolid nodule. Investigations and treatment decisions will

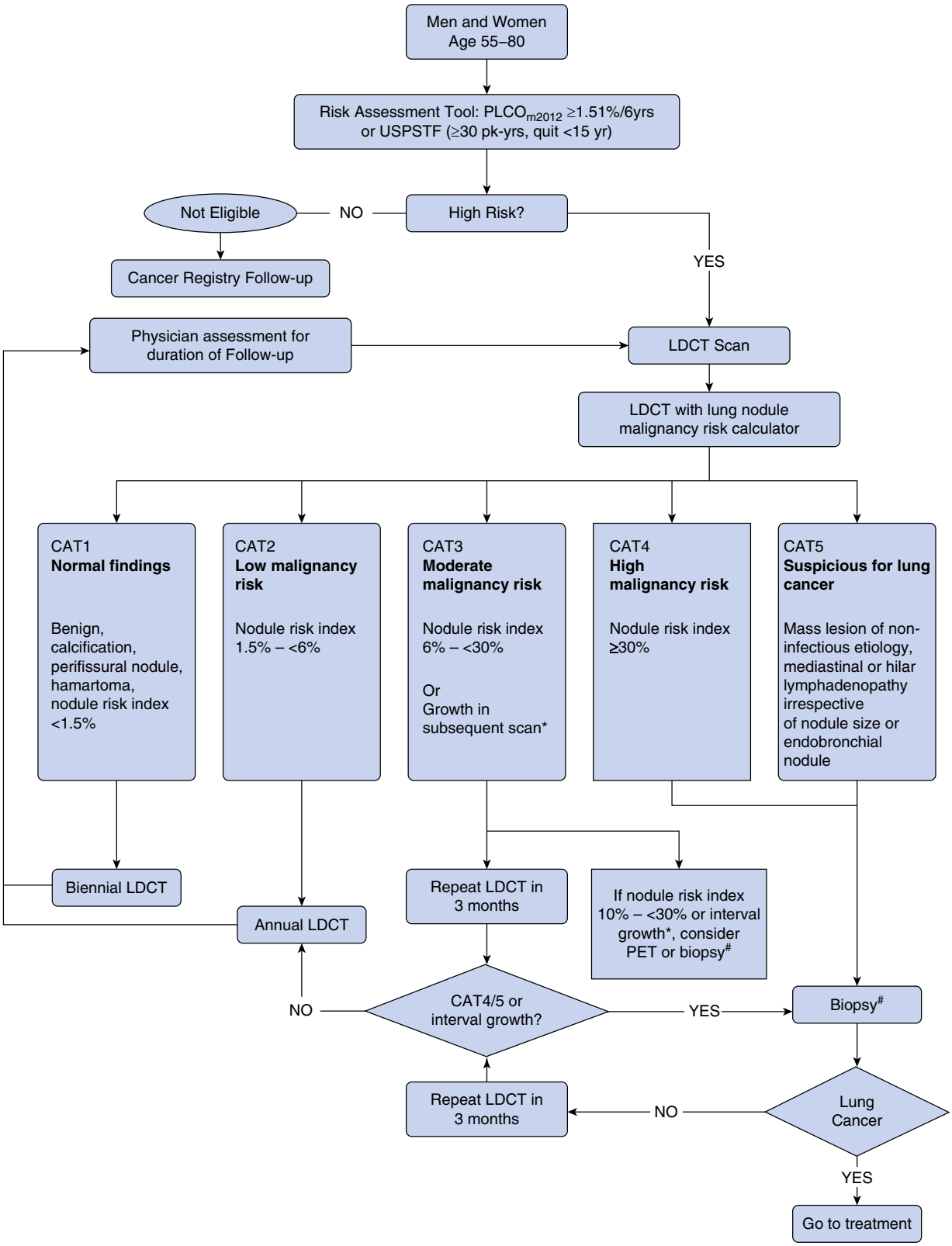

Figure 2. Lung nodule management protocol. *Growth in subsequent scan is defined as: >1.5 mm in mean diameter or solid core of semi-solid nodule ⩾ 6 mm. # = Consider biopsy after appropriate clinical assessment. CAT = computed tomography; LDCT = low-dose computed tomography; PET = positron emission tomography; PLCO = Prostate, Lung, Colorectal and Ovarian; USPSTF = U.S. Preventive Services Task Force.

be led by the clinical team according to local standard of care and may be managed at non-ILST centers.

Any confirmed diagnosis of lung cancer will be treated according to standard of care in the institution as directed by the medical team.

Outcome Evaluation and Follow-Up

All participants will be followed up annually for 5 years or longer (2017 to 2024) to ensure accurate determination of screening outcomes required to evaluate primary study aims 1 and 2.

We recognize that accurate determination of lung cancer outcomes in the noneligible population is required to calculate negative predictive value and specificity and that we have limited resources to achieve this. Collation of accurate data of noneligible individuals will take place by invitation to consent to the contribution to long-term outcome data as well as data linkage to regional and national cancer registries informing of the development of lung cancer or death from lung cancer. While the limitations of registry data are well described, in Australia and Canada, the registration of all cancers (apart from nonmelanoma skin cancers) is required by law, and thus case ascertainment of cancers is considered relatively robust. Both Australian and Canadian cancer registries are governed by national standards of data collection, error checking, deduplication, and linkage to other data sources.

All confirmed lung cancer cases will be evaluated for the following features: histology, TNM stage classification, treatment procedures, length of inpatient stay, and investigation- or treatment-related complications. Annual participant survey will collect smoking status, quality of life and healthcare usage.

In addition to the aforementioned primary outcomes, other outcomes of interest are mortality, rate of detection of other incidental significant disease, types and costs of downstream investigation and treatment related to abnormalities found during screening, logistics/barriers for an early detection program, and quality of life measures and costs.

Computer-aided Nodule Detection

At participating sites, screening scans will undergo semiautomated CAD software analysis (Veolity 1.2 system, MeVis Medical Solutions AG, Germany) as part of a randomized control substudy to evaluate the utility of CAD to improve radiologist reporting time and accuracy. Study LDCTs (baseline or follow-up) will be randomized to either being radiologist-read first then CAD-verified or CAD-verified then radiologist-read for comparisons of time required to report scans and of diagnostic accuracy. Of note, for the purposes of the main ILST study, radiologist review remains the gold standard for LDCT reporting; thus this substudy will not unduly influence the protocoled nodule analysis.

Health Economics

A comparative modeling approach will be used where the costs and cost-effectiveness of lung cancer screening will be evaluated both in Australia and Canada. The health economic analysis for Australia will use a microsimulation model to simulate lung cancer incidence and mortality in the population. The model will consist of core components including lung cancer natural history, diagnosis, treatment, survival, and a smoking history-generator. Some concepts of this model are adapted from CISNET (42, 43). The model will be built, validated, and calibrated using representative datasets and a large, population-based Australian cohort study (the 45 and Up Study) (44). As trial data become available (T_0, T_1, T_2), successive validation exercises will be performed using demographics, participation, cost, and outcome data from trial participants. The modeled health resource utilization and health outcome predictions for the simulated cohort will be compared with the observed data using similar methods as previous work (45). Following completion of this validation step and to perform health economic evaluation of LDCT screening as performed in the trial, projections of 10-year and lifetime outcomes and costs for trial participants will be estimated. Parallel analysis will be done for Canada using the microsimulation platform and/or alternative health economic models (46).

In addition to the above modeling information, healthcare resource utilization rates will be prospectively reported from each of the study centers via an electronic case report form. In Australia, costs will be ascertained from utilization data within the trial, including costs of screening itself, out-of-hospital medical services funded under the Medicare Benefits Schedule and/or the Department of Veteran Affairs, and public hospital utilization. Other Australian datasets may be used to supplement cost data with information regarding emergency department presentations, private hospital admissions, and prescription pharmaceuticals subsidized by the Pharmaceutical Benefits Scheme. In Canada, societal cost data will be ascertained from questionnaires administered to a subgroup of participants and consenting lung cancer patients receiving treatment at the BC Cancer Agency (46). Specifically, costs to screening participants will be evaluated on a per-visit basis with consideration of average travel time, distance, transport modality, employment status, and out-of-pocket expenses related to screening- or treatment-related appointments.

Statistical Considerations

Sample size and power. Our initial study power was based on a sample size of 4,000. Shortly after commencing the study, resources became available to increase the sample to 4,500. Enlarging the sample was deemed to be useful for answering some of the study's ancillary questions and to facilitate analyses stratified by different sites (e.g., Canada vs. Australia). The power calculations presented here are updated to reflect the enlarged sample size. The estimates of proportions in the different power calculations came from PLCO data, our own preliminary data, and the PanCan Study (13, 18). Correlations between $PLCO_{m2012}$ and USPSTF were estimated from PLCO data.

The power for testing if there is a difference in the proportion of lung cancer detected in $PLCO_{m2012}$-selected ([c + d]/T) versus USPSTF-selected ([b + d]/T) individuals (primary aim 1) by McNemar's test is 0.81. Power calculation assumptions: $n = 190$ (lung cancers, including 30 lung cancers in Table 1 [cell a]), and estimated proportion of lung cancers detected is 0.68 for $PLCO_{m2012}$ and 0.59 for USPSTF ($\delta = 0.11$), correlation = 0.55, two-sided alpha error = 0.05.

The power for testing if there is a difference in the PPVs for lung cancer detected in $PLCO_{m2012}$-selected versus USPSTF-selected individuals (primary aim 1) by comparing two proportions with the likelihood-ratio test is 0.82. Power

calculation assumptions: $n = 4{,}000$ individuals will be positive by either criteria (but are not the same group of individuals or lung cancers), PPV = 0.042 for $PLCO_{m2012}$, and PPV = 0.030 for USPSTF ($\delta = 0.012$), two-sided alpha error = 0.05.

The power for testing the difference in positive scan proportions (primary aim 2) by McNemar's test is >0.95. Power calculation assumptions: $n = 4{,}500$ (entire sample), proportion positive is 0.15 for PanCan and 0.19 for Lung-RADS ($\delta = 0.04$), no correlation assumption is made, two-sided alpha error = 0.05.

The power for testing that the proportion of lung cancers detected in positive scans by PanCan is not the same as that detected by Lung-RADS (primary aim 2) by comparison of proportion by the likelihood-ratio test is 0.92. The power calculation includes the following assumptions: $PLCO_{m2012}$ has 675 positive screens (0.15 of 4,500), and USPSTF has 855 (0.19 of 4,500). In the 675 $PLCO_{m2012}$ positive screens, 144 lung cancers (0.90 of 160) will be detected. In the 855 USPSTF positive screens, 125 lung cancers (0.78 of 160) will be detected. The hypothesis tested is 0.146 (125/855) different from 0.213 (144/675).

Power calculations were performed using Stata MP 14.1 software (College Station, Texas).

$PLCO_{m2012}$ risk prediction model performance. The efficiencies of the $PLCO_{m2012}$ risk $\geq$0.0151 and the USPSTF criteria to select high-risk smokers for LDCT screening will be compared by applying these criteria to the prospective data and evaluate study aims 1 and 2. In all four parts of primary study aims 1 and 2, proportions will be compared between the two criteria when they are applied to the same sample. Thus, estimates are not obtained from independent samples, and McNemar's test is most appropriate. Confidence intervals for proportions will be prepared using the exact binomial method (47). As statistical measures of performance fail to estimate clinical benefit of one method over another, decision curve analysis will be performed to compare net benefit differences between the different models and criteria (48).

Missing data. Regarding the primary study hypothesis, determining sensitivity, number screened, and positive predictive value of USPSTF/NLST criteria (the latter is nested in the USPSTF criteria) versus $PLCO_{m2012}$ risk for selection of individuals at high risk for lung cancer screening, all information will be collected by simple direct interview with the prospective study participant. As a consequence we are anticipating zero missing information for these key elements of the study, as those individuals withholding this required information will be excluded from the study.

For secondary study questions which require more specific detailed information we anticipate missing information to be less than 10%, in which case we will consider doing complete cases analysis if the total dataset for analysis is greater than 90%. If the missing data leads to less than 90% of the data being analyzed at an individual level, multiple imputation will be used to handle the missing data (49). Multiple imputation will be implemented using Stata software (50).

Timeline

All recruitment sites have commenced baseline screening in 2017 with the aim of completing baseline screening by end-2019. Completion of 5-year follow-up with final collation of data is expected in 2024.

Ethics and Data Monitoring

This study has been approved by all appropriate local committees of the ILST research sites. All study participants provide written informed consent. An independent monitoring committee will monitor trial processes and protocol deviations.

Discussion

ILST will provide a clearer understanding of the optimum selection criteria for LDCT screening for lung cancer and will evaluate the PanCan nodule malignancy risk protocol using retrospective modeled comparative analysis.

The ILST study is not powered nor designed to detect mortality benefits of LDCT screening. However, from this study other important questions surrounding lung cancer and screening will also be assessed in the form of planned subsidiary studies.

The information derived from the ILST will be important in guiding future international recommendations and healthcare resource allocation. The study is currently in its recruitment phase. Results will be reported in future peer-reviewed publications. ■

ILST (International Lung Screening Trial) Investigator Consortium

Brisbane, Australia: Kwun M. Fong*, Ian Yang, Rayleen Bowman, Richard Slaughter, Katrina Hopcroft, Linda Passmore, Elizabeth McCaul, Susanna Doyle, Henry Marshall, Rachael O'Rourke, Luke Connelly, and Karin Steinke

Melbourne, Australia: Renee Manser, Paul Fogarty, Daniel Steinfort, Louis Irving, Katharine See, Diane Pascoe, Mark McCusker, and Paul Mitchell

Perth, Australia: Annette McWilliams, Fraser Brims, and Kuan Pin Lim

Sydney, Australia: Emily Stone, Matthew Peters, Klaire Garnica, Lisa Tarlinton, Robert Kent, Karen Canfell, Marianne Weber, Nicole Rankin, and Yoon-Jung Kang

St. Catharines, Canada: Martin Tammemägi

Vancouver, Canada: Stephen Lam*, John Yee, Renelle Myers, John Mayo, Ren Yuan, John English, Sonya Cressman, and Sukhinder Atkar-Khattra

Hong Kong: David C. L. Lam

Barcelona, Spain: Antoni Rossell

Baltimore, USA: Christine Berg

*Co–primary investigators.

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors acknowledge Stephan Lam (respirologist, co–principal investigator), John Mayo (radiologist), John Yee (thoracic surgeon), Sukhinder Atkar-Khattra (project manager), Ren Yuan (radiologist), Renelle Myers (respirologist), Anne Dy Buncio (data manager), Nancy Norton (research assistant), Isaac Streit (research assistant), Sim Ladhar (research assistant), Kelly Cho (research assistant), Stephenie Pillainayagam (research assistant), and David Chen (research assistant), BC Cancer Agency, Vancouver, British Columbia, Canada; Martin Tammemägi (clinical epidemiologist), Brock University, St. Catharines, ON, Canada; Kwun Fong (respiratory physician, co–principal investigator), Henry Marshall (respiratory physician), Elizabeth McCaul (research nurse), Linda Passmore (research nurse), Katrina Hopcraft (radiologist), Ian Yang (respiratory physician), Rayleen Bowman (respiratory physician), Richard Slaughter (radiologist), Rachael O'Rourke (radiologist), Anne Williamson (radiologist), Luke Connelly (health economist), and Susanna Doyle (researcher), The Prince Charles Hospital, Brisbane, Australia, and the University of

Queensland; Karin Steinke (radiologist), The Royal Brisbane and Women's Hospital, Brisbane, Australia; Annette McWilliams (respiratory physician), Kuan Pin Lim (respiratory physician, *dual-site*), Lin Mo (respiratory research fellow, *dual-site*), Jacqueline Logan (research nurse, project manager), Stephen Melsom (radiologist), Yi-Jin Kuok (radiologist, *dual-site*), and Bann Saffar (radiologist), Fiona Stanley Hospital, Perth, Australia; Fraser Brims (respiratory physician), Kuan Pin Lim (respiratory physician, *dual-site*), Lin Mo (respiratory research fellow, *dual-site*), Siobhan Dormer (research nurse), Mark Teh (radiologist), Yi-Jin Kuok (radiologist, *dual-site*), and Ramon Sheehan (radiologist), Sir Charles Gairdner Hospital, Perth, Australia; Renee Manser (respiratory physician), Daniel Steinfort (respiratory physician), Louis Irving (respiratory physician), Diane Pascoe (radiologist), Shelley Mesfin (research nurse), Stewart Duncum (clinical research coordinator), Katharine See (respiratory physician), and Mark McCusker (radiologist), The Royal Melbourne Hospital, Melbourne, Australia; Paul Mitchell (medical oncologist), The Austin Hospital, Melbourne, Australia; Paul Fogary (respiratory physician), Belinda Dresser (clinical trial study coordinator), and Hayden Prime (radiologist), Epworth Eastern Hospital, Melbourne, Australia; Francis Thien (respiratory physician), Eastern Health, Melbourne, Australia; Emily Stone (respiratory physician), Kathryn Zeitz (head of diagnostic services), Eugene Hsu (radiologist), Lisa Tarlinton (radiologist), Brad Milner (radiologist), Klaire Garnica (radiographer), Robert Kent (clinical research manager), Trifon Psaroulis (research nurse), and Chris Rofe (data manager), St. Vincent's Hospital, Sydney, Australia; Matthew Peters (respiratory physician), Concord Hospital, Sydney, Australia; Karen Canfell (director of cancer research), Marianne Weber (epidemiologist), and Nicole Rankin (research fellow), Cancer Council, New South Wales, Sydney, Australia; David Lam (respiratory physician), Monica Sze-man Chan (radiologist), Ming-Yen Ng (radiologist), Varut Vardhanabhuti (radiologist), Peony Chong (research nurse), Vanessa Lai (research assistant), and Nerissa Lee (technical officer), Hong Kong; and Antoni Rossell (respiratory physician), Spain.

References

1 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.

2 Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, *et al*.; National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.

3 de Koning HJ, Meza R, Plevritis SK, ten Haaf K, Munshi VN, Jeon J, *et al*. Benefits and harms of CT lung cancer screening strategies. A comparative modeling study for the US Preventive Services Task Force. *Ann Intern Med* 2014;160:311–320.

4 Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, *et al*. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med* 2013;159:411–420.

5 Moyer VA; US Preventive Services Task Force. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:330–338.

6 Canadian Task Force on Preventive Health Care. Recommendations on screening for lung cancer. CMAJ 2016; 188(6) 425–432.

7 Cancer Australia 2015 Standing Committee on Screening. Position statement: lung cancer screening using low-dose computed tomography. *Canberra: Cancer Australia*; 2015 [accessed 2017 June 22]. Available from http://www.cancerscreening.gov.au.

8 Manser R, Lethaby A, Irving LB, Stone C, Byrnes G, Abramson MJ, *et al*. Screening for lung cancer. *Cochrane Database Syst Rev* 2013;6: CD001991. DOI: 10.1002/14651858.CD001991.pub3.

9 Brims F, McWilliams A, Fong K. Lung cancer screening in Australia: progress or procrastination? *Med J Aust* 2016;204:4–5. (Editorial).

10 Field JK, Smith RA, Aberle DR, Oudkerk M, Baldwin DR, Yankelevitz D, *et al*.; IASLC CT Screening Workshop 2011 Participants. (IASLC CT Screening Workshop 2011 Participants). International Association for the Study of Lung Cancer computed tomography screening workshop 2011 report. *J Thorac Oncol* 2012;7:10–19.

11 Kovalchik SA, Tammemägi M, Berg CD, Caporaso NE, Riley TL, Korch M, *et al*. Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med* 2013;369:245–254.

12 Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, *et al*. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. *PLoS Med* 2014;11:e1001764 10.1371/journal.pmed.1001764.25460915

13 Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, *et al*. Selection criteria for lung-cancer screening. *N Engl J Med* 2013;368:728–736.

14 Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, *et al*.; National Lung Screening Trial Research Team. Cost-effectiveness of CT screening in the National Lung Screening Trial. *N Engl J Med* 2014; 371:1793–1802.

15 Tammemägi MC, Lam S. Screening for lung cancer using low dose computed tomography. *BMJ* 2014;348:g2253.

16 Tammemägi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, *et al*. Lung cancer risk prediction: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial models and validation. *J Natl Cancer Inst* 2011;103:1058–1068.

17 Weber M, Yap S, Goldsbury D, Manners D, Tammemägi M, Marshall H, *et al*. Identifying high risk individuals for targeted lung cancer screening: independent validation of the $PLCO_{m2012}$ risk prediction tool. *Int J Cancer* 2017;141:242–253.

18 Tammemägi MC, Schmidt H, Martel S, McWilliams A, Goffin JR, Johnston MR, *et al*.; PanCan Study Team. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study. *Lancet Oncol* 2017;18:1523–1531.

19 Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, *et al*.; National Lung Screening Trial Research Team. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med* 2013;369:920–931.

20 Field JK, Duffy SW, Baldwin DR, Whynes DK, Devaraj A, Brain KE, *et al*. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax* 2016;71:161–170.

21 Horeweg N, van der Aalst CM, Thunnissen E, Nackaerts K, Weenink C, Groen HJ, *et al*. Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *Am J Respir Crit Care Med* 2013;187:848–854.

22 van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, *et al*. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009;361:2221–2229.

23 Pedersen JH, Ashraf H, Dirksen A, Bach K, Hansen H, Toennesen P, *et al*. The Danish randomized lung cancer CT screening trial—overall design and results of the prevalence round. *J Thorac Oncol* 2009;4: 608–614.

24 Walter JE, Heuvelmans MA, Oudkerk M. Small pulmonary nodules in baseline and incidence screening rounds of low-dose CT lung cancer screening. *Transl Lung Cancer Res* 2017;6:42–51.

25 Callister MEJ, Baldwin DR. How should pulmonary nodules be optimally investigated and managed? *Lung Cancer* 2016;91:48–55.

26 MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett JR, Naidich DP, *et al*.; Fleischner Society. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005;237:395–400.

27 Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, *et al*. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 2013;266:304–317.

28 MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, *et al*. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology* 2017;284: 228–243.

29 American College of Radiology. Lung CT screening reporting & data system (Lung-RADS V1.1). Reston, VA: ACR; 2019 [accessed 2019 July 31]. Available from https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads.

30 Kazerooni EA, Armstrong MR, Amorosa JK, Hernandez D, Liebscher LA, Nath H, *et al*. ACR CT accreditation program and the lung cancer screening program designation. *J Am Coll Radiol* 2015;12:38–42.

31 McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR Lung-RADS in a clinical CT lung screening program. *J Am Coll Radiol* 2015;12:273–276.

32 Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, *et al*. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Ann Intern Med* 2015;162:485–491.

33 Mehta HJ, Mohammed TL, Jantz MA. The American College of Radiology lung imaging reporting and data system: potential drawbacks and need for revision. *Chest* 2017;151:539–543.

34 McWilliams A, Tammemägi MC, Mayo JR, Roberts H, Liu G, Soghrati K, *et al*. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910–919.

35 Marshall HM, Zhao H, Bowman RV, Passmore LH, McCaul EM, Yang IA, *et al*. The effect of different radiological models on diagnostic accuracy and lung cancer screening performance. *Thorax* 2017;72:1147–1150.

36 Winkler Wille MM, van Riel SJ, Saghir Z, Dirksen A, Pedersen JH, Jacobs C, *et al*. Predictive accuracy of the PanCan lung cancer risk prediction model—external validation based on CT from the Danish lung cancer screening trial. *Eur Radiol* 2015;25:3093–3099 10.1007/s00330–015–3689–0.

37 Al-Ameri A, Malhotra P, Thygesen H, Plant PK, Vaidyanathan S, Karthik S, *et al*. Risk of malignancy in pulmonary nodules: a validation study of four prediction models. *Lung Cancer* 2015;89:27–30 10.1016/j.lungcan.2015.03.018.

38 van Riel SJ, Ciompi F, Winkler Wille MM, Naqibullah M, Schaefer-Prokop C, van Ginneken B. Lung-RADS versus the McWilliams nodule malignancy score for risk prediction: evaluation of the Danish lung cancer screening trial. *J Thorac Oncol* 2015;10(9)Suppl2:S191

39 van Riel SJ, Ciompi F, Jacobs C, Winkler Wille MM, Scholten ET, Naqibullah M, *et al*. Malignancy risk estimation of screen-detected nodules at baseline CT: comparison of the PanCan model, Lung-RADS and NCCN guidelines. *Eur Radiol* 2017;27:4019–4029 10.1007/s00330–017–4767–2.

40 American College of Radiology. Lung CT screening and reporting system (Lung-RADS). Reston, VA: ACR; 2019 [accessed 2017 June 22]. Available from http://www.acr.org.

41 Callister MEJ, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, *et al*.; British Thoracic Society Pulmonary Nodule Guideline Development Group; British Thoracic Society Standards of Care Committee.British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;70:ii1–ii54 10.1136/thoraxjnl-2015–207168.

42 Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Anal* 2012;32:S51–S68.

43 Meza R, ten Haaf K, Kong CY, Erdogan A, Black WC, Tammemägi MC, *et al*. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* 2014;120:1713–1724.

44 Banks E, Redman S, Jorm L, Armstrong B, Bauman A, Beard J, *et al.*; 45 and Up Study Collaborators. Cohort profile: the 45 and up study. *Int J Epidemiol* 2008;37:941–947.

45 Kitchener HC, Canfell K, Gilham C, Sargent A, Roberts C, Desai M, *et al*. The clinical effectiveness and cost-effectiveness of primary human papillomavirus cervical screening in England: extended follow-up of the ARTISTIC randomised trial cohort through three screening rounds. *Health Technol Assess* 2014;18:1–196.

46 Cressman S, Lam S, Tammemägi MC, Evans WK, Leighl NB, Regier DA, *et al*.; Pan-Canadian Early Detection of Lung Cancer Study Team. Resource utilization and costs during the initial years of lung cancer screening with computed tomography in Canada. *J Thorac Oncol* 2014;9:1449–1458.

47 Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Stat Sci* 2001;16:101–133.

48 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–574.

49 Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med* 1991;10:585–598.

50 Royston P. Multiple imputation of missing values: update of ICE. *Stata J* 2005;5:527–536.

51 Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, *et al*. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982;5:649–655.