Screening for Lung Cancer

Henry M Marshall and Kwun M Fong
Department of Thoracic Medicine, University of Queensland Thoracic Research Centre,
The Prince Charles Hospital, Brisbane, Queensland
Email: Henry_marshall@health.qld.gov.au.

Abstract

Lung cancer is a major global health issue and will remain so for decades to come. Lung cancer screening has the potential to reduce mortality from this disease and represents one of the most exciting developments in recent years. Screening appears to have taken hold in the US, yet has received a more cautious reception in other countries, probably due to a lack of accurate cost-effectiveness data and implementation capacity uncertainties. Refinement of screening using risk prediction to select the highest risk candidates is the next challenge and, coupled with an integrated smoking cessation program, could substantially improve cost-effectiveness. Future research will determine if biomarkers in biological samples will offer a low cost and minimally invasive method of screening and early detection.

Screening with plain chest radiography

Because lung cancer is known to have an asymptomatic, preclinical phase, it was long held that screening, actively searching for early disease in healthy people ‘at risk’, may be effective. From the 1960s onwards, major efforts into screening using plain chest radiography (CXR) were instigated in Europe and the US. Although CXR screening detected more cancers, it had no effect in reducing mortality.13 This apparent paradox is explained by the concept of overdiagnosis, where disease never destined to become clinically-apparent (e.g. a very indolent slow-growing tumour in a person who will die from cardiac disease), is detected by screening, treated and thus apparently ‘cured’.

Overdiagnosis bias is a major concern in screening programs, as it exposes people who would never have developed clinically apparent disease to the risks of unnecessary treatment. It is the subject of much on-going debate in breast and prostate cancer screening.14-17 Although the CXR studies had methodological limitations,15,18 the lack of clear benefit meant CXR screening never became clinical practice. However, there was sufficient doubt that large, well designed randomised Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial was set up in the US in 1993.18 The results of the lung cancer part were published in 2011;19 77,445 participants aged 55-74 years old were randomised to annual CXR screening and 77,456 to usual care (no screening). After 13 years of follow-up, cumulative lung cancer incidence rates were similar (20.1 v 19.2 per 10,000 person years), as were stage and histology. CXR screening made no difference to risk of death from lung cancer, (RR, 0.99; 95% CI, 0.87-1.22), leaving no doubt that CXR screening is ineffective.

However, before the definitive result from the PLCO trial had been reported, interest turned to CT as a potential screening tool, the hypothesis being that CT was too insensitive to detect truly early disease (eg. tumours <1-2cm diameter) but that CT, with its excellent spatial...
resolution, could very easily detect small lesions. To reduce the burden of medical radiation to patients, low-dose protocols were developed, reducing the dose to ~1.5mSv per scan, considerably lower than ~7mSv from conventionally-dosed studies. Thus began the exploration of low-dose CT (LDCT) as a screening tool.

**Screening with Low Dose Computed Tomography (LDCT)**

The first LDCT screening studies were from Japan and the US and were observational in design. They established that LDCT screening was acceptable to the general population, feasible for large numbers of screenees and approximately three times as sensitive as CXR in detecting small tumours. Upwards of 84% of tumours were stage 1 compared to only 16% 'localised' stage in routine clinical practice, thus most patients could be treated surgically. These exciting results led to calls for implementation of screening in the US, however as they were observational studies lacking control groups, no estimate of the effect on lung cancer mortality could be made. Survival was reported as a surrogate endpoint, but this is open to potential bias such as lead time and length bias. Lead time bias gives an apparent increase in survival (the time from diagnosis until death), simply because the date of diagnosis is brought forward, without requiring any effect on the natural history of the disease itself. Length bias describes the preference for screening to detect slower-growing tumours. Early in the piece, it was noted that adenocarcinoma was the predominant histological subtype detected, but that rapidly growing tumours, such as small cell and squamous carcinoma had the potential to grow quickly during the interscan period. These aggressive tumours are thus more likely to be missed on screening and for the patient to present with symptoms or at an advanced stage. More indolent tumours, by definition, have a better prognosis than aggressive tumours, which therefore makes screening appear to be very effective.

To address these methodological limitations, randomised control trials were initiated. The two largest are the National Lung Screening Trial (NLST) in the US and the NELSON study in Holland/Belgium. Both studies are adequately powered to be able to detect a mortality benefit. Several smaller European randomised control trials plan to combine their data in meta analysis. To date, only the NLST has reached its primary endpoint. Their landmark paper in 2011 reported, for the first time, a 20% reduction in lung cancer mortality in the LDCT-screened arm compared to the control (CXR-screened) arm. In response to this result, several US bodies now endorse screening and screening is now reimbursed by several health insurance companies. Certain other countries have followed suit (table 1) yet others, including Australia, remain more cautious. The UK is conducting its own randomised control trial and the British Thoracic Society issued a statement to say that screening in the UK cannot be currently advocated. So why is there such caution?

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Primary target population</th>
<th>Secondary target population</th>
<th>Screening interval and duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Comprehensive Cancer Network</td>
<td>Age 55-74 years</td>
<td>Age ≥50 years</td>
<td>Annual screening Until age 74 years</td>
</tr>
<tr>
<td>American Association for Thoracic Surgery</td>
<td>Age 55-79 years</td>
<td>Treated lung cancer, recurrence-free after four years’ surveillance</td>
<td>Annual screening Until age 79 years</td>
</tr>
<tr>
<td>American College of Chest Physicians and the American Society of Clinical Oncology</td>
<td>Age 55-74 years</td>
<td>Age 50 to 79 years; 20 pack-year smoking history; 5 year cumulative lung cancer risk of ≥ 5%</td>
<td>Annual screening Until age 74 years</td>
</tr>
<tr>
<td>American Cancer Society</td>
<td>Age 55-74 years</td>
<td>Smoking cessation within the past 15 years</td>
<td>none</td>
</tr>
<tr>
<td>French Intergroup for Thoracic Oncology and The French-Speaking Oncology Group</td>
<td>Age 55-74 years</td>
<td>Smoking cessation within the past 15 years</td>
<td>Annual screening Until age 74 years</td>
</tr>
</tbody>
</table>
Screening cost-effectiveness

Although the NLST has answered the fundamental question of whether screening can reduce mortality, several other important questions remain. Because screening is more than simply subjecting people to CT scans, lesion follow-up and downstream evaluation can significantly impact screening effectiveness. In addition, detailed costs of screening are still to be reported. The NLST has yet to publish its cost-effectiveness data, but preliminary reports suggest it will achieve the currently accepted standard of less than $50,000 per incremental cost-effectiveness ratio. Modelled estimates of LDCT screening costs have varied enormously from highly cost-effective through to highly ineffective, making coherent synthesis of their results next to impossible. An Australian study found screening was likely to be expensive and only cost-effective if very high-risk individuals were targeted and screening was either highly effective or very cheap. However, this study took the assumption that screening would be undertaken by case-finding (i.e. screening offered opportunistically to individuals when they sought medical care) rather than centrally-organised mass screening. The healthcare model in the US is substantially different to that in Australia, thus extrapolation of NLST costs to this country requires caution. This issue may be assisted by our Australian study of lung cancer screening, the Queensland Lung Cancer Screening Study. This pilot observational study is modelled on the NLST and should provide data on the cost of an NLST-style screening program locally.

Screening harms

Other factors that will impact on any screening program’s effectiveness and cost-effectiveness include the false-positive and adverse event rates. False-positive scans are a common problem. In the NLST, approximately 25% of scans were found to have nodules, yet >95% of these were proven to be benign (stability or resolution on serial CT follow-up). There is currently no way of conclusively diagnosing lung cancer from CT images; shape, margin and density all have low positive predictive value. Size is the best marker, the risk increasing with larger sized nodules. After follow-up scans have been obtained, growth is also an important clue. Nonetheless, benign lesions can have very similar appearances and thus biopsy is required to establish the diagnosis. Invasive procedures for benign disease are clearly unwanted. They have the potential to cause direct harm and increase the downstream costs of screening. Both the NLST and NELSON studies used protocols requiring biopsy whenever possible, and if the nodule was too small to biopsy (approximately less than 8mm diameter), required serial monitoring for evidence of growth. A recent systematic review of screening benefits and harms found the literature difficult to interrogate in terms of screening harms due to differing reporting methods in many studies, however rates of non-surgical invasive procedures for ultimately benign diagnoses (e.g. needle biopsy and bronchoscopy) were 1.2% in both the NLST and NELSON studies. Rates of surgical procedures (e.g. thoracotomy) for benign lesions were 0.7 and 0.6% respectively. In the NLST, the rate of complications after any invasive procedure was 33 per 10,000 screenees in the LDCT arm, mostly accounted for by post-surgical complications; the rate following bronchoscopy or needle biopsy was only 1.5 per 10,000 screenees.

Overall perspectives

Other questions that need to be addressed for successful implementation of a screening program include how best to recruit high risk individuals. NLST and NELSON found their participants were slightly above average for age and screening criteria, other markers of socioeconomic status. Lower socioeconomic status groups are important to target, as they have higher cancer rates and generally worse outcomes, yet generally are less likely to avail themselves of screening. The remoteness of many Australian communities may compound the problems of access and health care equity.

Despite the potential benefit of screening, it is imperative that we acknowledge that the most important strategy to reduce future lung cancer risk is to help smokers quit. Smoking cessation remains one of the most cost-effective health interventions, and therefore should form an important component of any future screening program. Indeed, the combination of an integrated smoking cessation program within a screening program could improve overall cost-effectiveness.

An area of current interest is the use of predictive risk modelling in the context of screening. The appeal is that such models may better define the ‘high risk’ population and may also help decide the optimal screening interval. The NLST used a simple eligibility strategy based on age and smoking history. Although undoubtedly these are the two most important risk factors, many others are well-known to contribute to risk, for example family history of lung cancer, occupational and environmental exposures and chronic obstructive pulmonary disease diagnosis. Of the published models, the best validated and most comprehensive was developed by Tammemagi et al using the PLCO dataset and subsequently refined using the NLST dataset. In a retrospective comparison to the existing NLST eligibility criteria, they found using the risk model improved screening sensitivity (83% v 71%, P <0.001) and positive predictive value (4.0% v 3.4%, P = 0.01), without loss of specificity (both 63%). Using the risk model, rather than simply age and smoking criteria, to select participants would have missed approximately 40% fewer lung cancers.

Another potential use for risk stratification is in deciding the screening interval. NLST used annual screening and this is advocated by current guidelines. The NELSON study incorporated a two year interval in its study between the second and third scan and an Italian study is randomising participants to annual or biennial screening. Another Italian study found the presence of radiological emphysema on baseline CT scan improved risk stratification and hypothesised that this could be used to help determine the subsequent screening interval.

Theoretically, better selection of high risk screenees and tailoring scan interval to risk will lead to a lower number of false positive scans and a higher yield of lung cancer. In turn, this will lead to a more efficient and cost-effective process. Two current American guidelines recommend the use of risk estimation for potential screenees who fall outside of the NLST eligibility criteria, but who have other known risk factors (table 1). However, it must be stated that currently no risk model has been prospectively tested in a screening context, although the UK Lung Screen is undertaking this task.

The limitations of CT scans have been well documented;
References