A Systematic Review of the Harms of Lung Cancer Screening with Low Dose Computed Tomography

By

Cody Schwartz

A Master's Paper submitted to the faculty of
the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for
the degree of Master of Public Health in
the Public Health Leadership Program.

Chapel Hill

2011

_____
Advisor:

_____
Date

_____
Second Reader:

_____
Date

Abstract:

Background: Lung cancer causes the majority of cancer-related deaths in the United States. Screening for lung cancer with low dose CT (LDCT) has become a prominent topic. This modality has the potential to contribute to a decrease in lung cancer associated mortality. Preliminary data from a large randomized controlled trial have shown a 20% decrease in mortality. Currently, no guidelines endorse LDCT screening for lung cancer. The harms of screening for lung cancer have not been fully characterized in a focused systematic review.

Purpose: To explore and characterize the harms associated with low dose CT screening for lung cancer in high risk populations by evaluating the current evidence. This information will assist policy makers and clinicians as they weigh the risks and benefits of screening high risk populations for lung cancer.

Data Sources: MEDLINE, EMBASE, Cochrane Library, reviews, reference lists, experts

Study Selection: Studies were included that evaluated low dose screening for lung cancer. Those that characterized harms were central to this review.

Data Extraction: Data were abstracted to abstraction forms. Studies were graded according to methods used by the United States Preventive Services Task Force (USPSTF).

Data Synthesis: Studies reported variable rates of abnormal results with positivity of 8 to 45%. A range of 4 to 55% of total invasive procedures and none to 40% of surgeries were performed for benign indications. Incidental findings were reported in 0.77 to 62% of participants. Of the included studies, 30% discussed anxiety, 55% discussed overdiagnosis, 50% reported morbidity and mortality of workup, and 35% reported false negative cases. False positives at baseline ranged from 86 to 98%.

Limitations: This review included observational studies, randomized trials, and systematic reviews of observational studies and short randomized trials. Data from completed large scale randomized trials is currently unavailable. No data from unpublished literature or studies

unavailable in English were included so this review has a potential for publication and language bias.

Conclusions: The harms of low dose CT screening for lung cancer are poorly reported in screening trials. Even though harms are incompletely characterized in the included studies, these harms are likely clinically significant. Clinicians must weigh the risks and benefits of the screening procedure if they are to implement mass screening for lung cancer with LDCT.

Introduction:

Background

To date, lung cancer screening has not proven useful. Systematic reviews have not shown a mortality benefit to screening for lung cancer[1-4]. No current guidelines recommend mass screening for lung cancer, and the American College of Chest Physicians recommends that individuals at high risk should only undergo screening as part of a clinical trial[5,6]. However, preliminary data from the NLST randomized screening trial show a 20% reduction in mortality for those who received a LDCT screening exam compared to the control group[7]. While this data is encouraging in the fight against lung cancer, a good quality screening program requires that the benefits outweigh the harms. No current systematic review has been designed to fully characterize the harms of lung cancer screening with LDCT. In order for clinicians, policy makers, and patients to make informed decisions about screening, a systematic review of screening harms is required. This evidence will help clinicians, policy makers, and patients balance the benefits and risks of screening in a population at risk for lung cancer. Due to the absence of a current systematic review, I will attempt to evaluate the harms of LDCT screening for lung cancer in this review.

Burden of suffering

Lung cancer is the leading cause of overall cancer death in the United States and worldwide[8-10]. According to the American Cancer Society (ACS), lung cancer is expected to cause 157,300 deaths in 2010 in the US, representing approximately 28% of cancer deaths[8]. In

2007, the most recent year for complete data available from the Centers for Disease Control (CDC), lung cancer was responsible for 6.5% of all US deaths[11]. In 2010, the ACS predicted 222,520 new cases of lung cancer will be diagnosed in the United States, 15% of all cancer diagnoses[8]. In perspective, lung cancer causes more deaths per year than stroke (136,000), lower respiratory disease (128,000), and accidents (124,000)[12]. If lung cancer mortality was an individual category, it would be the third leading cause of death in the US, behind only heart disease and all other cancers.

Worldwide, the ACS and the International Agency for Research on Cancer (IARC) estimated about 1.6 million new cases were diagnosed in 2008, representing about 13% of all cancer diagnoses[9,10]. Deaths from lung cancer totaled 1.4 million in 2008, representing 18% of all cancer deaths[9,10]. Lung cancer is the leading cause of cancer death in males, and is second in females behind only breast cancer[9,10].

Lung cancer incidence in males has decreased in the United States over the past decade, but has increased among females from 1991 to 2006[13,14]. During the same period in the US, mortality rates have decreased among males but have increased and remained stable since 2003 for females[13,14]. Even though rates among males have decreased, absolute incidence and mortality remain higher for males compared to females. Tobacco use patterns are likely contributing to the divergent trend observed between sexes. As a result, prevention strategies to decrease lung cancer incidence and mortality among both groups remain a priority.

Prevention Strategies

The primary risk factor for developing lung cancer is tobacco use[15]. In 2000, tobacco use was responsible for 435,000 deaths in the United States[16]. This represents 18.1% of all US deaths in that year. These deaths include not only lung cancer, but other cancers, respiratory illnesses, and cardiovascular diseases associated with tobacco use. According to the CDC and a 2004 report from the Surgeon General, tobacco use is attributable to 80-90% of lung cancer deaths in the US[15,17]. Male smokers are 23 times as likely to develop lung cancer as

nonsmoking males over their lifetime[15]. Female smokers are 13 times as likely to develop lung cancer as nonsmoking females over their lifetime[15]. Former male smokers are 9 times as likely to develop lung cancer as never smokers[15]. Former female smokers are 5 times as likely to develop lung cancer as female never smokers[15]. After 20 years of tobacco cessation, lung cancer risk decreases to about 2 times that of never smokers but never reaches the risk of a never smoker[18]. Lung cancer risk from tobacco exposure is dose dependent, so the more one smokes, the greater the risk of developing cancer[19]. While other risk factors play a role in developing lung cancer, this demonstrates that the main priority for prevention of lung cancer deaths is tobacco cessation and use prevention.

As tobacco use contributes the majority of risk for developing lung cancer[17], prevention of lung cancer is straightforward. Smoking cessation and use prevention is the priority. While the prevalence of smoking in the US has decreased since the 1980s according to data from the National Health Interview Survey, 20.6% of adults, or 46.6 million people, were current smokers in 2009[20]. An estimated 23.5% of men and 17.9% of women were current smokers in 2009 and an additional 21.9% of the population was considered a former smoker[20]. This shows that 42.5% of the US population was either a current or former smoker in 2009, which demonstrates the significant proportion of individuals at increased risk for developing lung cancer compared to never smokers. While prevention of tobacco use is a priority, individuals remain at risk for lung cancer long after they quit smoking as cancer takes many years to develop. Thus, lung cancer screening has been proposed as a tool to reduce the burden of suffering.

In addition to primary prevention, screening for early disease has been proposed to combat lung cancer mortality. While the greatest reduction in the burden of suffering would likely come from a reduction in smoking rates, early detection could further reduce the burden of suffering if screening randomized controlled studies demonstrate a reduction in lung cancer mortality.

History of Screening/Screening Methodology

Screening has the potential to detect early stage cancers which are more amenable to cure than late stage cancers. Most lung cancers present with symptoms and are diagnosed in an advanced stage when prognosis is poor[21]. Only 15% of lung cancers were diagnosed in a local stage in 2006[13]. The ideal screening test would decrease mortality by improving the probability of cure from early detection of early stage cancers. To date, a mortality benefit for lung cancer screening has not been shown. Even if a mortality benefit is eventually shown, screening for early stage cancers can only be effective if the benefits outweigh the harms. Burden of testing, incidental findings, cost, workup morbidity and mortality, anxiety, surgical morbidity, radiation exposure, and screening biases need to be taken into account when weighing the risks and benefits of screening[2,22]. However, because the prognosis of late stage cancers is dismal, even a small benefit in discovering and treating early stage cancers is likely to afford a mortality advantage. As a result of this poor life expectancy, screening has been proposed as a possible mechanism to decrease lung cancer mortality.

Lung cancer screening dates back to the 1960s when researchers studied chest x-ray and sputum cytology as screening modalities. Early trials published in the 1960s and 1970s that evaluated chest x-ray and sputum cytology showed no mortality differences for the intervention group versus controls[23-29]. Observational and experimental studies evaluating chest x-ray published in the 1980s and 1990s also failed to demonstrate a mortality benefit[30-33]. These studies were criticized for lack of adjustment for screening biases, study design problems including lack of power to detect mortality differences, and contamination in control groups[34]. Due to these problems, one arm of the current Prostate Lung Colon Ovarian (PLCO) Cancer screening randomized trial is designed to address mortality differences by comparing annual chest x-ray with usual care for early detection of lung cancer[34]. This study sought to enroll 158,000 individuals aged 55-74 years old with randomization into screening and usual care groups. The lung cancer intervention arm offers smokers annual chest x-ray for 3 years and

non-smokers two annual repeat scans[34]. Final results of this trial are currently unavailable, but are expected soon in 2012.

Two up to date systematic reviews have explored the efficacy of lung cancer screening with chest x-ray for a reduction in mortality. Seven controlled trials comparing chest x-ray with usual care were included in a Cochrane systematic review, six were randomized and one was non-randomized[1]. This review reported a pooled analysis of no mortality benefit and suggested a possible mortality increase in the screening group. A second systematic review also found no mortality benefit from screening with chest x-ray compared to controls[2]. Together these reviews and the chest x-ray screening arm of the PLCO trial will likely address the issue of the efficacy of lung cancer screening with chest x-ray.

Low dose CT screening for lung cancer was proposed in a 1990 study by Naidich et al[35] who compared the feasibility of conventional CT versus low dose CT of the chest. Observational studies evaluating low dose CT screening for lung cancer followed soon thereafter. LDCT screening has been shown to detect early stage lung cancers in uncontrolled observational studies[36-46]. Observational studies and short randomized trials of LDCT have been evaluated in systematic reviews which have demonstrated no mortality benefit to screening[2-4,47]. Due to the screening biases of length time, lead time, and overdiagnosis bias inherent in observational studies, it is difficult to discern whether any observed survival differences are due to actual reductions in mortality from screening or due to bias[22]. Studies without a control group may overestimate survival and only a randomized controlled trial comparing screening with LDCT to a control group can validly evaluate any mortality difference[22]. However, published data from completed randomized controlled trials are not yet available.

The NLST study comparing LDCT screening versus chest x-ray in a population with at least 30 pack-year history smokers will be published soon in 2011[48,49]. The NELSON trial from the Netherlands and Belgium is evaluating LDCT versus usual care in a high risk population and

will be published soon thereafter in 2012[50]. These trials will address the question of whether or not there is a mortality benefit to LDCT screening for lung cancer. However, the goal of this review is to evaluate the harms of LDCT screening. As the characterization of harms likely does not suffer from the same screening biases as mortality evidence, this review will include observational studies, randomized trials, and systematic reviews when available.

Summary

The best method to decrease mortality from lung cancer is to prevent it from occurring in the first place. As mentioned, the majority of risk of developing lung cancer is attributed to smoking. Thus, tobacco cessation programs and use prevention have the greatest likelihood of decreasing lung cancer mortality. But for those who already have a smoking history or are at greater risk for lung cancer, an intervention that decreases lung cancer mortality may benefit the high risk population. Researchers are evaluating LDCT screening as a viable option.

Studies have evaluated individual harms of LDCT screening for lung cancer, but have not fully addressed all harms including, but not limited to, burden of testing, invasive procedures, anxiety, false positives, radiation exposure, and incidental findings. This review will attempt to characterize the harms of screening for lung cancer with LDCT. This will help clinicians, policy makers, and patients weigh the risks and benefits of screening in order to make informed decisions.

This review will attempt to answer the following key questions:

Key Questions

(1) What is the evidence for harms associated with LDCT screening? (burden of testing, incidental findings, false positives, false negatives, testing burden, adverse effects of treatment, radiation exposure)

    (1a) What is the evidence for potential biases in screening for lung cancer? (length time, lead time, overdiagnosis)

(2) Are these harms able to be quantified?

Methods:

As published evidence thus far has shown no mortality benefit, I focused this review on the harms of screening. To address the focused question of identifying and quantifying the harms in lung cancer screening, a brief search of systematic reviews was conducted to determine if a review was indicated. No systematic review has been published that addresses potential harms of lung cancer screening with LDCT.

Data Sources and Searches

I searched the MEDLINE, EMBASE, and CENTRAL online databases from inception to 6/1/2011 to identify relevant articles. I used the MeSH terms of “lung neoplasm,” “mass screening”, "tomography, x-ray computed," and “tomography, spiral computed” plus keyword searches of low dose CT and screening harms. Complete search terms are provided in the appendix. I also hand searched relevant reviews and reference lists of included studies for articles. An attempt to locate gray literature was made in a search of clinicaltrials.gov for trials currently underway and in consultation with experts. An experienced research librarian was consulted in formulating the search strategy.

Study Selection

This review will include original research papers that are relevant to answer the key questions of screening harms. One author reviewed the initial full set of identified titles and abstracts. If it was not clear from the title or abstract if the study fit inclusion criteria, it was included for full text review to maintain a sensitive search. If multiple articles were published on one study, all relevant data was abstracted for our outcomes of interest. Reference lists of included studies were searched for relevant articles.

I included systematic reviews, randomized controlled trials, non-randomized trials, and prospective cohort studies. Cohort studies were included because it is likely that harm data are not subject to lead or length time bias as compared to that of mortality data, however there are other biases that needed to be taken into account. Systematic reviews were included for

identification of primary studies and were not included in the analysis. If systematic reviews quantified harms, I commented on these in the discussion section of this paper. We included any population of current or former smokers who have received a low dose CT scan for the purpose of screening for lung cancer. Studies were eligible for any duration of follow up as potential screening harms can be immediate, such as false positives or anxiety of awaiting results. Studies conducted in any setting that has the potential for lung cancer screening were eligible, including primary care and specialty settings. I required the study to measure at least one of the primary outcomes harm as described in the PICOTS in Table 1. Excluded articles included editorials, non-systematic reviews, case reports, and case series. Studies were excluded if they did not evaluate low dose CT as the intervention screening modality, such as those that evaluated chest x-ray or sputum cytology.

Data Extraction and Quality Assessment

I collected data from the included studies into evidence tables using a standard data collection form. Information collected included study design, population, intervention, comparators, and outcome measures as described in Table 1. Studies reporting incomplete outcomes were reported in the analysis. No attempt was made to contact the primary authors to obtain data. I assessed quality using the USPSTF methodology[51]. One author completed quality assessment for each study.

Data Synthesis and Analysis

I reported data for each of our primary outcomes if available in the included studies. A qualitative synthesis was conducted for each of the primary outcomes of interest due to the heterogeneity of study populations. Heterogeneity was not formally assessed using statistical methods due to the variability of eligibility criteria and differences in study protocols across the included studies. I used the PRISMA reporting guidelines for systematic reviews to preserve completeness[52,53].

PICOTS:

Category	Inclusion criteria
Population	Current and former tobacco smokers, including groups identified in the literature as high risk, low risk, and those whose risk is not defined, no restriction on numbers of pack-years - Former users as defined in literature
Intervention	Screening with low dose CT scan, any interval
Comparators	Usual care, chest x-ray, sputum cytology, different combinations of these
Outcomes	Primary outcomes: - Harms of screening/workup - Testing burden - Invasive procedures (including biopsies, thoracotomies, VATS) - False positives - False negatives - Radiation exposure - Incidental findings - Anxiety - Unnecessary procedures - Repeat imaging – especially associated with radiation exposure, HRCT, PET - Overdiagnosis - Quality of life Secondary outcomes: - Smoking cessation rates of abnormal result
Timing of effect	Harms: any time period
Timing of search	Since MEDLINE inception – 1966 to present
Setting	Primary care, specialty care, inpatient setting, or other relevant lung cancer screening settings
Study Design	SRs, RCTs, non-randomized trials, prospective cohort studies

Table 1. Defining PICOTS framework for systematic review of lung cancer screening harms.

Results:

A total of 655 titles and abstracts were identified and reviewed using our search strategy. Of these, 215 were included for full text review. Articles were excluded for wrong interventions, poor outcome reporting, duplicate publications, wrong design, poor quality, and inability to obtain an English translation, special populations. Nine articles were reviewed from searching reference lists. On final inclusion, 46 articles from 20 studies were included in this review. Six randomized controlled trials and 14 cohort studies were included. Eight systematic reviews and ten articles from special populations (asbestos, nuclear plants) were identified. Most randomized trials only reported baseline findings. A flow diagram of the results of our search strategy is shown in Figure 1.

Summary of Study Descriptions

Characteristics of included studies are summarized in Table 2. A total of 94,521 low dose CT screening exams were performed for 46,946 participants in the included studies. Median age ranged from 53 to 67 years old. Median smoking history ranged from 30 to 54 pack-years. Low dose CT slice thickness ranged from 0.75 to 11 mm.

The total number of lung cancers diagnosed was 845 with 67% of these diagnosed at baseline screening. Baseline positivity ranged from 5 to 53%. Stage I disease was diagnosed in 38 to 93% of lung cancers with 17 to 85% of all cancers representing adenocarcinomas. Lung cancer outcomes are summarized in Table 3. Study descriptions and lung cancer outcomes are provided in text form in the appendix.

Quality Summary of Included Studies

Quality assessment is summarized in Table 4. One study was rated as good and the remaining demonstrated fair quality. Main issues were poor adherence and limited description of recruiting strategies which has the potential to introduce selection bias. Most studies did not provide details about masking of the intervention to the investigator or participant. Masking would likely be unfeasible in subjects in randomized trials as they likely know if they are

undergoing a CT scan versus a chest x-ray or usual care. Generalizability was difficult to determine due to inadequate description of recruiting strategies inhibiting the ability to determine the source population. A final issue included poor reporting of target outcomes, especially for incident scans. Complete quality assessments for individual studies are reported in the appendix.

Summary of Target Outcomes

Total positivity of low dose CT screening ranged from 8 to 45%. For the studies reporting workup, 5 to 28% of those screened at baseline received some further workup as a result of a screening test. A range of 4 to 55% of total invasive procedures were performed for benign lesions. None to 40% of surgical procedures were performed for benign lesions. In the studies that reported incidental findings, 0.77 to 62% of those screened had abnormal findings other than lung nodules. Studies reported on extra-pulmonary tumors, emphysema, bronchiectasis, lymphadenopathy, aneurysms, coronary calcifications, renal abnormalities, and adrenal abnormalities. 30% of the included studies mentioned anxiety or burden of testing and 55% of included studies mentioned overdiagnosis as potential harms to lung cancer screening. Half of the studies reported observed morbidity and mortality for diagnosis or treatment. 35% of studies reported false negative cases. The false positive rate of baseline LDCT ranged from 86 to 98%. Target outcomes for studies are summarized in Table 5 and are individually addressed in the text below.

Depiscan - Blanchon[54]

Total positivity rate was 45%. I could not determine who received further workup after screening exams. The study reported 12 total invasive procedures with 25% for benign lesions, however it was unclear if participants received more procedures than were reported. 38% of surgeries were performed for benign lesions. Incidental findings were briefly reported and included a 6% rate of severe emphysema, 19% with bronchiectasis, and 5% with mediastinal

abnormalities. No morbidity or mortality was reported in the LDCT screening group. No false negative cases were reported. The false positive rate of LDCT was 95% at baseline.

Colorado - Garg[55]

Total positivity rate was 33%. 28% of LDCT screened participants received further workup. Invasive procedures and surgeries were not reported. Incidental findings were not reported, but one case of metastatic laryngeal cancer was mentioned. Morbidity, mortality, and overdiagnosis were not mentioned as potential harms. No false negative cases were reported. The false positive rate was 90%.

DANTE - Infante[56-58]

Total positivity rate was 28% for baseline and incident screens combined. 18% of those screened received some form of further workup. Of those who screened positive, 57% received a high resolution CT and 16% received a PET scan. Ninety-six invasive procedures were performed, 11% were for benign lesions. 13% of thoracotomies were performed for benign lesions. Five of the 46 total thoracotomies were required for non-lung cancer lesions including mesothelioma, esophageal cancer, complex aspergilloma, empyema, and leiomyoma. Incidental findings were reported in 3% of subjects. These included effusions, pleural lesions, mediastinal masses, hiatal hernias, aortic aneurysms, goiters, renal and adrenal masses, and diaphragmatic paralysis. Two post-operative deaths were reported. Overdiagnosis was mentioned as a potential harm of screening. One false negative case was detected on sputum cytology. The false positive rate was 86% at baseline.

ELCAP - Henschke[36,37]

Total positivity rate was 26% for baseline and incident screens combined. 18% of those at baseline and 3% of those who received a repeat screen received some form of further workup. All participants who received further workup received additional diagnostic imaging. Forty invasive procedures were performed, 13% were for benign lesions. Of note, 3 of these were against study protocol. It was unclear how many surgical procedures were performed but

no thoracotomies were performed for benign lesions. Incidental findings were not reported in the primary articles. Morbidity, mortality, and other potential harms of screening were not reported. Five false negative cases were reported as four lung cancers and one interim cancer were visible on baseline exam. The false positive rate was 88%.

A follow study conducted by Anderson *et al* examined whether negative screening results influenced smoking cessation behaviors on a subpopulation of those in the ELCAP study[59]. The study found that negative scans were not associated with a lower likelihood of prolonged abstinence among baseline smokers or a higher likelihood of relapse among former smokers over 6 years of follow up compared with those with a false positive result. In other words, negative scans were not associated with adverse smoking behavioral changes. However, smoking status was not collected uniformly throughout the study period and could influence measurement of results. In addition, 39% of those who reported quitting during the study period were excluded from analysis because of missing data.

A second follow up study evaluated smoking behaviors among a subpopulation of 313 individuals from the ELCAP sample[60]. Cancer anxiety and perceived health benefits of quitting were found to be predictors of smoking cessation following LDCT exam. However, the result of the screening exam, whether positive or negative, was not statistically significantly associated with smoking cessation.

NY-ELCAP[61]

Total positivity was 10% for baseline and incident screens combined. 8% of those at baseline and 3% of those who received a repeat screen received further workup. I could not determine who received additional diagnostic imaging. A total of158 biopsies were performed but 134 were performed as recommended by the study protocol. Of all biopsies, 21% were performed for benign lesions. Of the recommended biopsies, 7% were performed for benign lesions. The study reported no lobectomies were performed for benign lesions however it is unclear if any thoracotomies or VATS were performed for biopsy or benign lesions. No post-

operative deaths of workup or treatment were observed. No complications were mentioned. Incidental findings were not reported in the primary article, however the study reported four cancers detected on biopsy that were not lung cancer, including 2 cases of metastatic colorectal cancer and 2 cases of lymphoma. The study mentions overdiagnosis as a potential harm to lung cancer screening. The study did not report any false negative cases. The false positive rate of LDCT screening was 89% in this study.

A companion paper analyzed the frequency of mediastinal masses detected in both the ELCAP and NY-ELCAP populations[62]. The authors reported a prevalence of 0.77% of those screened who had any mediastinal mass. These included masses in the thymus, thyroid, esophagus, and trachea. Other incidental findings were noted but not reported in this paper.

Matsumoto - Sone[43,63,64]

Total positivity rate was 4.3% for baseline and incident screens combined. 4.9% of those at baseline, 3.8% of those screened at first repeat, and 3.3% of those screened at second repeat received further workup. I could not determine who received further diagnostic imaging. Seventy-two invasive procedures were reported, all were surgical biopsies. 22% of these surgeries were performed for non-cancerous lesions. Incidental findings were not reported. The study reported 24 false negative cases. No complications were mentioned. The false positive rate was 92%.

A follow article reported 10 year follow up of the study cohort[64]. Main outcomes in this article included survival of patients diagnosed with CT detected lung cancer. However, the study did include an estimated overdiagnosis rate of 13% using a formula which included the volume doubling time of the tumor and the patient's age.

ALCA – Sobue/Kaneko[42,65]

Total positivity rate was 9.5% for baseline and incident screens combined. 12% of those at baseline and 9.8% of those screened at repeat screening received further workup. 97% and 99.7% of positives at baseline and repeat screening, respectively received additional diagnostic

imaging. Seventy-one total invasive procedures were performed, 49% were for benign lesions. Twenty-one total surgeries were performed with a 29% rate for benign disease. Incidental findings were not reported in the primary articles. Two deaths from post-surgical infection were recorded. There were no adverse events of biopsy observed. Anxiety and overdiagnosis were mentioned as potential harms to screening. No false negatives were reported. The false positive rate was 95%.

ITALUNG – Lopes Pegna[66]

Total positivity rate was 30%. 26% of LDCT screened participants received further workup. Fifty-two invasive procedures were performed, 4% were for benign lesions. One of the 18 surgeries performed was for a benign lesion, a rate of 5.5%. Incidental findings were not reported. Morbidity, mortality, and overdiagnosis were not mentioned as potential harms. No false negative cases were reported. The false positive rate was 95%.

COSMOS - Veronesi[67-69]

Total positivity rate was 53% for baseline screening and could not be determined for incident screening. 9.7% of those screened received further workup including diagnostic testing and biopsy. This study reported 104 surgeries with 14% of those for benign lesions. Procedures included 5 VATS wedge resections, 9 open wedge resections, and 1 open lobectomy. This study did not report incidental findings. The study mentions radiation exposure and overdiagnosis as potential harms to LDCT screening for lung cancer. Major post-operative morbidity was reported in 4.6% of cancer resections. No peri-operative morbidity occurred in surgeries for benign disease and no post-operative mortality was observed among surgeries. The study reported 24 false negative cancers that had a prevalent lesion on baseline. The false positive rate was 98% at baseline.

Munster - Diederich[39,40,70]

Total positivity rate was 18% for baseline and repeat screening combined. I could not determine who received further workup after a positive result. The study reported 20% of the 16

invasive procedures performed at baseline were for benign lesions. Three surgical procedures were performed at baseline, but I could not determine the proportion of total surgeries that were performed. This study did not report incidental findings but reported on seven cases of lung metastases from unknown primary malignancies. Overdiagnosis was mentioned as a potential harm to screening for lung cancer. One post-operative death was observed following bilobectomy for cancer treatment. Four false negative cases were reported as abnormal baseline scans but were diagnosed symptomatically within screening intervals. The false positive rate on baseline scan was 96%.

Hitachi - Nawa[41]

Total positivity rate was not reported and I could not determine who received further workup. At baseline, 6.8% received diagnostic CT whereas 2.7% received diagnostic CT at repeat screening. Seventy-one invasive procedures were performed in total, but it was not clear how many were for benign lesions. Seventeen of 57 surgeries, or 30%, were performed for a benign lesion. The authors reported extrapulmonary tumors as incidental findings. These were tumors in the thyroid, parathyroid, mediastinum, and chest wall. No complications of workup were reported. The false positive rate at baseline was 98%.

Milan - Pastorino[71]

Total positivity rate was 15% for baseline and repeat screening combined. I could not determine who received further workup. 6% of those screened at baseline and 3.4% of those screened at repeat screening had further diagnostic imaging performed. Total invasive procedures were not reported. 22% of the 27 total surgeries were performed for benign lesions. Incidental findings were not reported. No post-operative deaths were observed during study follow up. The authors mention anxiety and overdiagnosis as potential harms to lung cancer screening. No false negative exams were reported. The false positive rate was 94% at baseline. The authors reported an average effective radiation dose of 0.7 mSv per screening exam.

Mayo Clinic - Swensen[44,45,72,73]

Total positivity rate was 51% for baseline screening and 74% of those screened had at least one non-calcified nodule (NCN) detected on any screen over the 5 total screening exams. I could not determine who received further diagnostic imaging or invasive diagnostic procedures. At 3 year follow up, 8 of 39 surgeries were performed for benign lesions, a rate of 21%. At 5 year follow up, 15 surgeries were performed for benign lesions out of an unknown total. This study reported that 79% of those screened had either a pulmonary lesion or an incidental finding after 3 years of follow up. The most common extra-pulmonary findings included lymphadenopathy, aortic aneurysms, renal abnormalities, and adrenal abnormalities. In addition, 18 extra-pulmonary tumors were detected including cases of renal cell carcinoma, breast cancer, lymphoma, carcinoid, ovarian cancer, pheochromocytoma, and metastases. The study mentions both anxiety of workup and overdiagnosis as potential harms to LDCT screening for lung cancer. The study reported 5 false negative cases, 1 sputum detected case that was visible on baseline scan and 4 cancers detected on incident scans that were visible on prior scans. The false positive rate was 96% at baseline.

A follow up article by Cox *et al* reported on smoking changes of 97% of the Mayo study participants with smoking data at baseline[74]. 14% of baseline smokers quit at one year and 10% former smokers were smoking at one year. A longer duration of cessation was associated with abstinence at one year for former smokers. A positive screening result was not associated with cessation and a negative screening result was not associated with continued smoking or relapse. The study was not designed to determine if screening was associated with any change as there was no non-screened control group.

A second follow up article by Townsend *et al* evaluated the relationship between smoking cessation and results of screening exam over three annual screening exams[75]. This study included 91% of the original Mayo cohort with 926 current smokers and 594 former smokers. Smoking cessation after baseline exam was associated with older age, worse

baseline lung function, and abnormal CT findings on prior screening exam. For those who received 3 abnormal exams, 42% reported quitting smoking. For those who received 2 abnormal exams, 28% reported quitting. For those with one abnormal exam, 24% quit. For those without an abnormal exam, 20% quit. It appears this association is dose dependent. Longer duration of abstinence was the only predictor of maintaining abstinence among former smokers.

Another follow up study by Crestanello *et al* more closely examined thoracic surgery procedures in the Mayo cohort through 2002[73]. The article reported a total of 60 surgical procedures in 55 patients performed for diagnosis and treatment. Lobectomy was the most common procedure followed by wedge resection. Ten surgeries were performed for benign disease, a rate of 18.1%. One lobectomy and nine limited resections were performed for benign lesions. Complications occurred in 27% of surgeries including air leak, arrhythmia, pneumonia, ileus, stroke, depression, vocal cord paralysis. One death (1.7%) occurred post-operatively from an intracerebral hemorrhage. A total of 79 nodules were resected, 39% of these nodules had benign pathology.

PLuSS - Wilson[76,77]

Total positivity rate was 41% for baseline and incident screens combined. 22% of those screened at baseline had further diagnostic imaging, or 56% of those that screened positive. Total invasive procedures were not reported, but 90 total surgeries were performed. 40% of these were for non-cancerous diagnoses. 82 surgeries were performed for lung cancer suspicion and 28 of these were for benign lung lesions, a rate of 34%. The remaining 8 surgeries were performed for suspicious incidental findings ultimately diagnosed as benign lesions. Incidental findings were not well reported however 9 extra-pulmonary malignancies were discovered including 4 cases of lymphoma and 5 metastatic lesions in the lung. Morbidity and mortality of workup and overdiagnosis were mentioned as harms to screening but were not quantified. The false positive rate at baseline was 96%.

A companion paper explored healthcare usage after LDCT screening in a subpopulation of 400 participants in the PLuSS cohort[78]. There were no observed differences between groups, however data showed an increase in outpatient visits over the first 6 months following the screening exam, regardless of whether the participant received a positive or negative result. Visits declined to pre-screening levels over the next 6 months. These data show that a temporary increase in visits occurred following screening, but this increase was not sustained past 6 months.

A follow article presented results exploring the relationship of emphysema diagnosed on CT with airflow obstruction[76]. This paper found that 24% of those screened had mild to moderate/severe emphysema on LDCT. It is unclear if these results are clinically significant, however emphysema was found to be independently associated with lung cancer.

PALCAD - MacRedmond[79,80]

Total positivity rate was 25% for baseline and incident screens combined. I could not determine who received further workup or the total number of invasive procedures performed. Five invasive procedures were performed for benign lesions. One of 4 surgeries was performed for a benign lesion. Incidental findings were well reported. 62% had any incidental finding on baseline screen with 49% having a clinically significant finding requiring further workup. Most common findings included emphysema (29%), coronary calcification (14%), bronchiectasis (10%), and abdominal abnormalities (10%). One complication of workup was reported as a pneumothorax for a FNA biopsy. The false positive rate at baseline was 98%.

LSS – Gohagan[81,82]

Total positivity rate was 23%. 19% and 24% of those screened at baseline and repeat screening received any further workup, respectively. 73% and 40% of positive results at baseline and repeat screening received further diagnostic imaging, respectively. There were 107 invasive procedures performed in total. At baseline, 23 people received an invasive procedure for a benign lesion, 43% of all procedures. Forty-six biopsies and resections were

performed at baseline, 39% of these for a benign diagnosis. Eighteen biopsies and resections were performed after repeat screening and it was unclear how many of these were for benign lesions. Ten participants had complications related to follow up including pneumothorax, infection, stroke, respiratory failure, and pneumonia. The study mentioned anxiety and overdiagnosis as potential harms to lung cancer screening. Incidental findings were not reported in the primary articles, however a follow up letter reported proportions of incidental findings described below. No false negative cases were reported. The false positive rate at baseline was 91%.

Incidental findings of the Lung Screening Study were reported in a letter by Pinsky *et al*[83]. Of those who received a LDCT screening exam, 16% had a nodule or mass, 17% had a granuloma, 29% had scarring or fibrosis, 25% had COPD, 10% had pleural fibrosis, 0.32% had pleural fluid, 1.5% had a bone or soft tissue abnormality, 14% had a cardiac abnormality, and 19% had coronary artery calcifications. It is unclear if these incidental findings were clinically significant.

A follow up study by Croswell *et al* examined the cumulative risk for false positive screening exams[84]. This study defined a false positive result as a positive LDCT screen with negative workup or one year without a lung cancer diagnosis. The primary outcome showed that the cumulative probability of obtaining a false positive screening test was 33% (CI: 31, 35%) after two LDCT screening exams in this population.

Toronto – Menezes[85]

Total positivity rate was 10% for baseline and incident screens combined. It appeared that 100% of those who screened positive had further workup consisting of LDCT, diagnostic CT, and/or biopsy. A total of 78 biopsies were performed with 16 resulting in a non-cancer diagnosis, a rate of 21%. It was unclear how many surgeries were performed and if any surgeries resulted in benign lesions. The authors did not report on incidental findings. The authors mention overdiagnosis as a harm of LDCT screening for lung cancer. There was one

pneumothorax recorded as a complication of a biopsy. The false positive rate at baseline was 91%.

Israel - Shaham[86]

Total positivity rate was 8.2% for baseline and incident screens combined. I could not determine who received further workup. There were 16 reported invasive procedures performed, all were biopsies of lung lesions. Two of the 12 total surgeries were performed for a benign lesion for a benign surgical rate of 17%. Of note, these surgeries were against recommended protocol. Incidental findings were not reported. The study did not report complications of workup or resection. The false positive rate at baseline was 88%.

DLCST – Pedersen[87]

Total positivity rate was 8.7% as only baseline results have been reported. I could not determine who received further medical evaluation after screening. There were 40 total invasive procedures performed with 25% of invasive procedures performed for a benign lesion. Eleven surgeries were performed, 2 of these (18%) were performed for benign lesions. Incidental findings were not recorded. One post-operative death was reported following lobectomy for adenocarcinoma, but no morbidity or mortality was reported for work up. No false negative cases were reported. The false positive rate at baseline was 91%.

A companion article by Ashraf *et al* sought to determine smoking cessation changes in this trial[88]. Primary measures were cessation and relapse rates at one year. No statistically significant differences in cessation or relapse rates for LDCT versus control groups were observed. However, those who received a positive screening result in the CT group had a higher quit rate than those in the CT group who did not have a positive result (17.7% vs 11.4%, p=0.04). Similarly, those who had a positive screening result were less likely to relapse than those without a positive result in the CT group (4.7% vs 10.6%, p<0.01).

Toronto – Roberts[89]

Total positivity rate was 26% for baseline screening. The study states that 100% of those who screened positive had further workup, with 98% receiving further imaging. A total of 26 invasive procedures were performed for workup with 15% of these performed for benign lesions. The authors reported 3 surgeries for workup of which none were performed for benign indications. The majority of cancers were diagnosed with CT-guided biopsy. The authors reported 6 incidental findings consisting of one abnormal pulmonary vein, one mediastinal cyst, one mild lymphadenopathy, one thickened diaphragm, one carcinoid tumor, and one plasmacytoma. The authors mention overdiagnosis and anxiety as a harm of LDCT screening for lung cancer. No complication or mortality was reported as a result of workup. No false negatives were reported. The false positive rate at baseline was 92%.

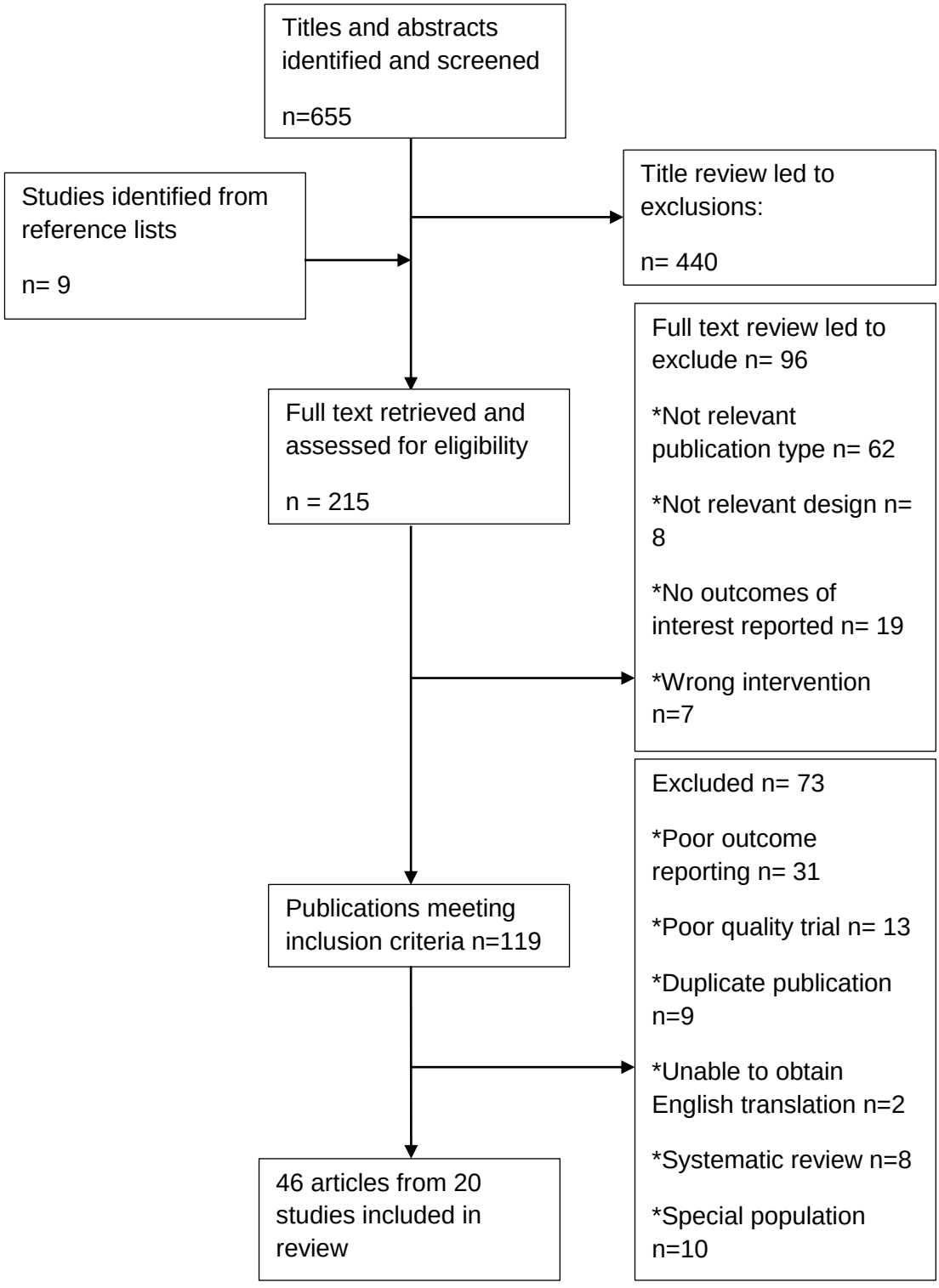

Figure 1. Flow diagram of search strategy.

Author	year started	# screened with LDCT at baseline	total # LDCT screening tests performed	population	females (%)	age range	median age (years)	smoking history (median p-y)	% current/former/never smokers	LDCT slice thickness (mm)
Blanchon	2002	336	336	subjects were under care of GP recruiters, asymptomatic >15 cigs/day for >20 yrs (quit <15 yrs prior)	29	50-75	56	30	64/36	1.5
Diederich	1995	817	2542	asymptomatic volunteers recruited by media, >20 p-y history	28	>40	53	45	-	5
Garg	2001	92	92	veterans, included those with COPD and sputum atypia in previous cohort or >30 p-y. recruited via telephone.	3	50-80	-	-	-	5
Gohagan	2000	1586	2984	volunteers recruited via mass mailing and referral, >30 p-y (quit w/in 10 yrs). From PLCO centers not already in PLCO trial	41	55-74	-	54	58/42	5
Henschke	1992	1000	2184	asymptomatic volunteers recruited from physician offices and hospitals, >10 p-y	46	>60	67	45	-	10

Infante	2001	1276	3612	male volunteers recruited via GPs and mass mailings/media, >20 p-y (quit <10 yrs)	0	60-74	65	47	56/44	5
Sobue/Kaneko	1993	1611	9502	subjects from a dues paying organization that provides lung cancer screening to its members, includes never smokers.	12	40-79	-	-	62/25/14	10
Lopes Pegna		1406	1406	recruited by sending invitation letters to registered patients of GPs, >20 p-y history	35	55-69	61 (mean)	39	65/35	1-3
MacRedmond		449	1371	recruited by media, residents of community served by hospital, >10 p-y smoking at 45 yo	50	>50	56	45	68/32	10
Menezes	2003	3352	6924	volunteers or referred by physician, >10 p-y	54	>50	60	30	NR	1-1.25
Nawa	1998	7956	13524	Hitachi employees, includes never smokers	21	50-69	-	-	62% current/former, 38% never smoker	10
Pastorino	2000	1035	2031	asymptomatic volunteers recruited via media, >20 p-y	29	>50	58	40	86/14	10

Pederse n	2004	2047	2047	volunteers recruited via media, >20 p-y	45	50-70	-	-	76/24	0.75
Roberts	2003	1000	1000	volunteers, >10 p-y history	55	>55	63	34	34/66	1.25
Shaham	1998	842	1784	mixed referral from physician and volunteers from media, >10 p-y	43	>50	56	37	-	3.2 - 11
Sone	1995	5483	13786	volunteers recruited via media, includes never smokers	46	40-74	64 (mean)	-	46% current /former. 54% never smokers	10
Swense n	1999	1520	NR	volunteers recruited via media, >20 p-y	48	>50	59	45	61/39	5
Verones i	2004	5201	10022	asymptomatic volunteers from single center, >20 p-y	34	>50	58 (mean)	44	80/20	2.5
Wilson	2002	3642	7065	recruited volunteers from media, could include symptomatic, >12.5 p-y history	49	50-79	59 (mean)	47	60/40	2.5
NY-ELCAP	2000	6295	12309	recruited via media and physician referral, >10 p-y	51	>60	66	40	33/67	1.25-10

Table 2. Description of included studies.

Author	year started	# screening tests performed	# lung cancers total	description of positive test	# baseline positive (%)	baseline lung cancers detected	prevalence	repeat positive	incident lung cancers detected	% stage I	% adenocarcinoma and BAC
Blanchon	2002	336	8	any NCN without benign features	152 (45%)	8	2.38%	-	-	38 %	63%
Diederich	1995	2552	29	any NCN without benign features	378 (46%)	12	1.35%	89 (13 %)	10	56 %	41%
Garg	2001	92	3	1-6 NCN without benign features	30 (33%)	3	3.26%	-	-	-	-
Gohagan	2000	2984	40	NCN >3 mm or other suspicious abnormalities	325 (21%)	30	1.89%	360 (26 %)	8	48 %	60%
Henschke	1992	2184	36	1-6 NCN without benign features	233 (23%)	27	2.70%	30 (2.5 %)	7	82 %	68%
Infante	2001	3612	63	abnormality of malignancy including NCN >10 mm or smaller with suspicious features. Negative if NCN <5mm	199 (16%)	28	2.19%	152 (4.2 %)	19	65 %	43%

Kaneko/ Sobue	1993	9502	36	those who were requested to undergo thin section CT scan based on suspicion of benign tumor/inflammation/lung cancer	186 (12%)	14	0.87%	721 (9.1%)	22	78%	67%
Lopes Pegna		1406	21	at least 1 solid NCN >=5mm or nonsolid >= 10mm or part solid NCN	426 (30%)	21	1.42%	-	-	48%	48%
MacRedmond		1371	6	Any NCN	105 (23%)	2	0.45%	6 (1.3%)	3	50%	17%
Menezes	2003	6924	65	baseline: >1 NCN >=5mm or non-solid >= 8mm. Repeat: growth or new NCN	600 (18%)	56	1.67%	92	6	65%	68%
Nawa	1998	13524	41	Non-calc SPN >=8 mm without benign features	2099 (26%)	37	0.45%	-	4	85%	85%
Pastorino	2000	2031	22	NCN >= 5 mm	199 (19%)	11	1.06%	99 (10%)	11	77%	77%
Pedersen	2004	2047	17	NCN>=5mm and growing nodules	179 (8.7%)	-	0.83%	-	-	53%	71%
Roberts	2003	1000	20	>=1 solid/part solid NCN >=5 mm or nonsolid >= 8 mm.	256 (26%)	20	2%	-	-	75%	70%
Shaham	1998	1784	14	baseline: 1-6 NCN >=5mm at first. Then adjusted to 1-6 solid/part solid NCN >=5mm or >1 non solid >=8 mm. repeat: new and/or growth	102 (12%)	12	1.43%	45 (5%)	2	86%	57%

Sone	1995	13786	60	positive result included scan read as possible cancer, probable cancer, or small nodule <3 mm	279 (5.1 %)	22	0.40%	309 (3.7 %)	34	88 %	85%
Swensen	1999	NR	68	any NCN	780 (51%)	31	2.04%	773 (NR)	34	57 %	53%
Veronesi	2004	10022	92	any NCN	2754 (53%)	55	1.06%	45 (0.9 %)	36	66 %	68%
Wilson	2002	7065	80	any NCN	1477 (41%)	53	1.46%	42%	24	50 %	NR
NY-ELCAP	2000	12309	124	baseline: at least 1 solid/part solid NCN >= 5 mm or nonsolid NCN >= 8 mm. incident: any new NCN regardless of size or any growth.	906 (14%)	101	1.60%	361 (6.0 %)	20	93 %	60%

Table 3. Results of included studies.

Author	design	randomization/allocation concealment
NY-ELCAP	prospective cohort	na
Wilson	prospective cohort	na
Veronesi	prospective cohort	na
Swensen	prospective cohort	na
Sone	prospective cohort	na
Shaham	prospective cohort	na
Roberts	prospective cohort	na
Pedersen	RCT (LDCT vs no screening annual)	2052 in each group, central computer permuted blocks of 10. adequate. Unclear concealment
Pastorino	prospective cohort	na
Nawa	prospective cohort	na

adherence/dropouts	contamination/crossover	selection bias potential(+ to +++ scale)
82% completed first annual repeat	-	++: assessed for differential follow-up, those who attended similar to baseline, but unclear about those who were dropouts
95% at 2 years	-	++: assessed for differential loss to follow-up.
4815 returned for rpt scan (93%)	-	++
98% at 1st repeat, 96% at second repeat, 95% at third, 80% at fourth repeat. Study reported only one person of 1520 who was lost to follow up	-	+: good reporting of adherence, assessed compliance and attempted to maintain adherence
80% adherence at 1st repeat scan, 70% at 2nd repeat	-	++: reported demographics of those who followed up, does not appear to be differential adherence
549/801 (69%) eligible had repeat screen.	-	+++: different recruiting methods, >30% dropout rate
baseline.	-	++: baseline report
2047/2052 (99.8%) received baseline CT	unclear if will assess	+: good descriptions here thus far, only baseline screening though
996 (96%) at 1st repeat scan, 91% overall compliance	-	++: good description of recruiting procedure and adherence
70% received at least 1 annual screen.	-	+++: 30% dropout rate, unclear inclusion criteria

measurement of outcomes	generalizability	study quality
2 rads read independently. Unmasked to participant within the study	fair. Pts paid for workup themselves	fair
1/2 rads read initial scan. Conference reviewed moderate/high suspicion studies. Used compuer algorithm to classify pre-biopsy lung ca risk	fair	fair
used both computer aided detection and manual detection. Radiologist interpreted and lesion size validated by 2 rads at conference.	fair	fair
read by 1 of 4 radiologists.	fair	fair
read by 1 of 4 radiologists, then 2nd reader read suspicious or clinically significant scans. Unmasked to first read.	uncertain: large proportion of females were never smokers, majority of total population were never smokers	fair
changed definition of a positive result during the trial, unclear when. Also changed slice thickness which would affect interpretation of images. One radiologist read all scans	fair	fair
unclear who/how many read scans	fair	fair
read by 2 rads (unknown independence). Discrepancy by consensus.	fair	fair
2 rads independently. Discrepancies resolved by 3rd rad.	fair	fair
2 readers independently read all cases. Discrepancies resolved at weekly conferences.	poor	fair

Author	design	randomization/allocation concealment
Menezes	prospective cohort	-
MacRedmond	prospective cohort (annual screening)	-
Lopes Pegna	RCT (LDCT vs no screening)	randomization by central computer system. Unclear concealment.
Kaneko/Sobue	prospective cohort (6 mo screening interval)	-
Infante	RCT (annual LDCT vs control)	permuted blocks of 4 stratified by center. Unclear of allocation concealment. Comparable baseline groups except for chronic respiratory symptoms in LDCT arm.
Henschke	prospective cohort	-
Gohagan	RCT (annual LDCT vs CXR 2 cycles)	appropriate randomization - blocks of variable size via central system. Comparable baseline groups
Garg	RCT (annual LDCT vs no screening)	randomization procedure NR. Control group appears to have greater p-y history but wide standard deviation so non-significant difference between groups
Diederich	prospective cohort (annual LDCT screen)	-
Blanchon	RCT (annual LDCT vs CXR)	comparable groups at randomization. Unclear of randomization procedure or allocation concealment

adherence/dropouts	contamination/cross over	selection bias potential(+ to +++ scale)
2686/3352 (80%) returned for annual repeat	-	++: 20% of participants who did not return?
413 (92%) completed 2 yr follow up	-	++
1406/1613 (87%) randomized to LDCT underwent baseline study	not assessed	++: non response
1180/1611 (73%) underwent 1st repeat. 445 (27%) at 4 years.	-	+++: for profit organization requires members to pay dues for access to screening, majority subjects dropped out after 3rd screening round (12 months follow up)
2811 total randomized, 2472 (88%) underwent baseline exam. Unclear adherence at followup	6.1% received non-protocol CT. 19% received CXR	++: acceptable rates of adherence and contamination, but unclear adherence at 3 yr followup
841/1000 (84%) had annual repeat scan, 117 were considered dropouts (12%)	-	++: acceptable followup rate, explanation of dropouts, contacted individuals who did not followup
96% and 86% adherence in LDCT group at baseline and 1 yr respectively	random survey: 1.3-2.6% in CXR received CT. 13-20% in LDCT got CXR for any reason	+: appropriate documentation
all 92 randomized to CT group underwent baseline scan.	-	+: all those randomized were adherent
668/817 (82%) returned for repeat scan at 1 yr	-	++: Due to funding, screening discontinued for <55 yo after >=1 normal repeat scan (2 total scans). >55 yo repeat CTs continued. Only n=24 (3%) underwent 5 scans total. N=149 did not respond or did not agree to f/u studies. 206 total nodules could not be followed up bc LTF
621/765 (81%) with baseline data due to withdrawal. Withdrawals more likely to be in CXR group and younger	6 (1%) of those in CXR group given LDCT	+++: concern that those in CXR group who dropped out may be healthier and less likely to have lung cancer

measurement of outcomes	generalizability	study quality
I-ELCAP protocol - Not reported. Thus do not know who read scans	fair	fair
read by 2 rads independently. Discrepancy by consensus.	fair	fair
read by 2 rads independently. Discrepancy by consensus. pulmonologist was not blinded to group.	fair	fair
2 radiologists or chest physicians read each scan. 2nd reader not blinded to first interpretation. 3rd reader determined recommendation for thin section CT. Computer aided diagnosis was introduced in middle of trial.	poor: membership in an organization	fair
2 chest radiologists read each scan independently, consensus for discrepancy.	fair	fair
each scan read by 2 radiologists masked to other's read. Assessed reader agreement	fair	fair
read by radiologist at each site. Results recorded on standardized forms. Unclear of masking or discrepancy in reads.	fair	good
read by 1 chest radiologist, suspicious scans read independently by 2nd radiologist, consensus for discrepancy. No mention of masking	poor (many already identified with sputum atypia in prior study)	fair
scans read by 1 of 2 radiologists, unclear about masking	fair	fair
2 radiologists read each scan independently, discrepancy by consensus	fair - only 41% GPs participated in recruitment, but this would only affect generalizability if GPs served different populations	fair

Table 4. Quality assessment of included studies.

Author	# screening tests performed	total lung cas	% total positive	% of baseline participants with any NCN on any scan (excluding benign) (includes <5 mm)	% of total screened receiving any further workup	% of positives receiving any further workup
NY-ELCAP	12309	124	10%	-	8% at baseline, 3% at repeat	55% at baseline, 54% at repeat
Wilson	7065	80	41%	41%	821 (22%) at baseline.	56%
Veronesi	10022	92	2799 (28%)	53%	504 (9.7%)	18%
Swensen	-	68	-	51%	-	-
Sone	13786	60	588 (4.3%)	-	4.9% at baseline, 3.8% at 1st repeat, 3.3% at 2nd repeat	95% at baseline, 97% at 1st repeat, 95% at 2nd repeat
Shaham	1784	14	8.20%	203 (24%)	-	-
Roberts	1000	20	26%	-	26%	100%
Pedersen	2047	17	179 (8.7%)	29%	-	-
Pastorino	2031	22	15% (of total scans)	-	-	-
Nawa	13524	41	-	-	-	-

#/% receiving diagnostic imaging (DXCT, HRCT, PET)	invasive procedures for diagnostic workup	#, % invasive procedures for benign	total surgeries	#, % surgical (thoracotomy or VATS) for benign of total surgeries	% with any incidental findings
-	158 biopsies	7 (6%) at baseline, 2 (8%) at repeat of 134 recommended. 24 non-recommended all for benign. Total 33 (21%)	-	no lobectomies (unclear about other procedures)	-
821 baseline (22%)	-	-	90	36 (40%)	-
-	104	15 (14%)	104	15 (14%)	-
-	-	15 (CD)	-	15 total. 8 (21% at 3 yr followup)	79% with pulmonary lesion or any IF
-	72	22%	72	16 (22%)	-
-	16	2	12	2 (17%)	-
250 (98%)	26	4 (15%)	3	appears none	6
-	40	10 (25%)	11	2 (18%)	-
61 (5.9%) at baseline, 34 (3.4%) at repeat	-	-	27	6 (22%)	-
6.8% at baseline, 2.7% at repeat	71	-	57	17 (30%)	-

what are incidental findings	anxiety/QOL/burden of testing	overdiagnosis	morbidity/mortality of diagnostic workup
NR. 4 extra-pulm cancers: 2 CRC, 2 lymphoma	-	mentions	no post-op deaths
9 extra-pulm cancer dx: 2 HL, 2 lymphoma, 5 mets to lung	-	mentions	mentioned as harm but not quantified
-	-	Mentions	major post-op morbidity in 4/86 (4.6%). No complications for benign dx. No post – op deaths
lymphadenopathy, aortic aneurysm, renal abnormalities, adrenal abnormalities. 18 extra-pulmonary tumors	mentions	mentions	1 post op lung cancer death. No peri-op mortality for benign
-	-	-	not reported
-	-	-	not reported
1 abnl pulm vein, 1 mediastinal cyst, 1 mild LAD, 1 diaphragm thickening. 1 carcinoid, 1 plasmactyoma.	mentions	mentions	none reported
-	-	-	1 post op death following lobectomy for adenocarcinoma
-	mentions	mentions	no post op deaths
14 extrapulmonary neoplasms: thyroid, parathyroid, mediastinal, chest wall tumors	-	-	-

Author	# screening tests performed	total lung cas	% total postive	% of baseline participants with any NCN on any scan (excluding benign) (includes <5 mm)
Menezes	6924	65	692 (10%)	-
MacRedmond	1371	6	111 (25%)	25%
Lopes Pegna	1406	21	426 (30%)	-
Kaneko/Sobue	9502	36	907 (9.5%) of total screening tests	-
Infante	3612	63	351 (28%) (individuals with positive result)	-
Henschke- ELCAP I	2184	36	263 (26%)	23%
Gohagan	2984	40	685 (23%)	-
Garg	92	3	30 (33%)	33%
Diederich	2552	29	467 (18%)	57%
Blanchon	336	8	45%	152 (45%)

% of total screened receiving any further workup	% of positives receiving any further workup	#/% receiving diagnostic imaging (DXCT, HRCT, PET)	invasive procedures for diagnostic workup
600	100%	12 (2%)	78
-	-	-	-
366 (26%)	366 (86)%	366	52 - 16 biopsies, 18 surgeries, 18 bronch
192 (12%) at baseline, 770 (9.8%) at repeat	192 (103%) at baseline, 770 (107%) at repeat	181 (97%) at baseline, 719 (99.7%) at repeat	22 at baseline, 49 at repeat
226 (18%)	226 (64%)	199 HRCT (57%), 57 PET (16%)	96
180 (18%) at baseline, 28 (3.3%) at repeat	77% at baseline, 93% at repeat	180 at baseline (77%)	31 (13%) at baseline, 9 (30%) at repeat
309 (19%) at baseline, 332 (24%) at repeat	(98%) at baseline, (95%) at repeat	232 (73%) at baseline, 140 (40%) at repeat	75 at baseline, 32 at repeat.
26 (28%)	100%	-	-
not clear	not clear	not clear	-
-	-	-	12

#, % invasive procedures for benign	total surgeries	#, % surgical (thoracotomy or VATS) for benign of total surgeries	% with any incidental findings
16 (21%)	NR	-	-
5 (NR)	4	1 (25%)	276 (62%)
2 (3.8%)	18	1 (5.5%)	-
8 (36%) at baseline, 27 (55%) at repeat	21	6 (29%) (VATS and thoracotomies)	-
6 (11%)	46	6 (13%) of thoracotomies	3%
4 (13%) (3 against recommendation) at baseline. 1 (11%) at repeat.	-	no thoracotomies for benign	-
23 (8%) at baseline, NR at repeat	46 at baseline. 18 at repeat.	18 at baseline. Unclear at repeat.	-
-	-	-	-
4 (20%)	-	3 (unclear)	-
3 (25%)	-	3 (38%) thoracotomies	-

what are incidental findings	anxiety/QOL/burden of testing	overdiagnosis	morbidity/mortality of diagnostic workup
-	-	mentions	1 pneumothorax
emphysema 130 (29%), bronchiectasis 44 (9.8%), CAC 64 (14.3%), IPF 6 (1.3%), inflammation 11 (2.5%), pleural plaques 4 (0.9%), goiter 9 (2%), TAA 1 (0.2%), abdominal findings 46 (10.2%) (includes hepatobiliary, renal, esophagus, fundal mass, endometriosis)	mentions	-	1 pneumothorax from FNA biopsy reported
-	-	-	none mentioned
-	mentions	mentions	2 deaths from post-surgical infection. No adverse events of biopsy.
extrapulmonary findings (effusions, pleural lesions, mediastinal mass, hiatal hernia, aortic aneurysm, goiter, renal mass, adrenal mass, diaphragm paralysis).	-	mentions	2 post op deaths
-	-	-	-
-	mentions	mentions	10 subjects had complications related to f/u. pneumothorax, infection, stroke, respiratory failure, pneumonia
1 mets from laryngeal cancer	-	-	-
7 had lung metastases from unknown primary	-	mentions	1 post op death following bilobectomy
6% severe emphysema, 19% bronchiectasis, 5% mediastinal abnormalities	-	-	none reported in LDCT group

Table 5. Harm outcomes of included studies.

Discussion:

This systematic review attempted to characterize the harms associated with lung cancer screening. Overall, harm outcomes were not uniformly reported across included studies. In addition, harms were difficult to quantify as most studies did not report information or reported incomplete information on harms. However, data from the included studies show that the harms of screening for lung cancer with low dose CT are not trivial. Harms should warrant discussion when deciding whether to initiate mass screening for lung cancer even if a mortality benefit is observed in good quality randomized controlled trials. For a successful screening program, the benefits of undergoing the screening test should outweigh the potential harms.

Healthcare usage

Healthcare usage following screening exam was addressed in the PLuSS study only[78]. Results showed increases in outpatient visits over the first 6 months following screening and then a decline to pre-screening levels by 12 months. This short term increase could be due to distress of testing, participation in a study, and a motivated volunteer population. Measures were self-reported so one must interpret these results with caution as recall and social biases may also play a role in reporting. Further studies should explore the relationship of healthcare usage post screening as this is likely to play a role in the burden of testing and burden on the healthcare system.

Distress/anxiety/quality of life

Psychological factors and quality of life were mentioned in a third of included studies[42,71,72,81,89]. Distress and anxiety were usually briefly mentioned in discussion as a potential psychological harm to screening, especially when a false positive result occurs. None of the included studies quantified distress on a validated scale. In the NELSON trial for which results are not yet published, the psychological effects of screening are specifically addressed in 3 articles[90-92].

In a study conducted by Bunge *et al*, the authors examined distress in 351 subjects before and after LDCT screening in the NELSON randomized trial[91]. Participants were surveyed 1 day before screening and 6 months after using the Impact of Event Scale (IES). Statistically significant results were observed for participant's perceived risk of developing lung cancer before and after screening. 14.5% perceived their risk for lung cancer as high 1 day before screening whereas 10.5% perceived their risk to be high 6 months later ($p<0.01$). Of note, all individuals here received a negative screening result so the results may be explained by reassurance in the screening group. From this study, it is unclear how those with a false positive result would perceive their risk after screening.

In a second study, the authors surveyed 351 participants who were randomized to the LDCT screening arm of the NELSON trial[90]. Health related quality of life (HRQoL) and discomfort were measured 1 day before, within one week after, and 6 months post screening with the 12–item Short Form and the EuroQol questionnaire. This study found that 87 to 99% of participants reported no discomfort related to the LDCT screening procedure and HRQoL did not significantly change over time, but 46% of participants reported at least some discomfort in waiting for results and 51% of participants reported dreading the screening exam results. The authors concluded that LDCT screening demonstrated no adverse effects on quality of life but discomfort was observed in waiting for results. Of note, this study excluded positive screening exams so does not measure the discomfort of false positives.

In a further study, 733 participants in the NELSON trial were surveyed for health related quality of life, anxiety, and lung cancer specific distress before and after screening[92]. Quality of life and anxiety measures did not change before and after screening for those with a negative result. However, IES scores increased after an indeterminate result and decreased after a negative result. This shows that lung cancer distress increases after receiving an indeterminate result and that relief occurs after a negative result. This study excluded positive results and

thus could not determine the effect of a false positive result on quality of life, anxiety, or lung cancer specific distress.

From these results, it is likely that distress, quality of life, and anxiety are likely to be important factors that should be addressed before implementing a LDCT screening program. This includes providing timely results and how to handle the psychological effect of an indeterminate result. None of these three studies addressed false positive results, which likely increase anxiety and lung cancer specific distress even further than an indeterminate result.

False positives and false negatives

One third of studies reported any false negative findings. False negatives were not uniformly defined. I included false negatives as they were defined in the primary studies. This definition included cancers detected on incident screen that were visible on prior screening exam, interim cancers that were visible on prior exam, and any cancers missed on CT but diagnosed by other means. Veronesi and Sone reported high numbers of false negative cases, 24 and 18 respectively, whereas other studies reported fewer than 5 cases[43,69]. In a follow up study, Li *et al* describes the errors in interpretation and detection as the primary causes for missing these lung cancers[93].

I was able to determine the false positive rate for all included studies. I defined false positives as screening exams that were interpreted as positive, according to the study definition of a positive test, but did not result in a diagnosis of lung cancer. For the majority of studies, at least 9 out of 10 positive screening tests were false positives at baseline. In addition, Croswell *et al* showed that the probability of having a false positive result after 2 total scans was 33% in the LSS population[84]. This study was conducted in a high risk population with >30 pack-years and >55 years old. However, this high proportion of those with a false positive test after 2 years increases burden on both the individual and the healthcare system as a positive result requires further medical evaluation with uncertain benefit.

These results show the poor predictive value of LDCT in screening for lung cancer. By changing the definition of a positive result, investigators hoped to decrease the false positive rate as was done in the ELCAP study[94]. However, the false positive rates in I-ELCAP and NY-ELCAP were still 90% and 88%, respectively[38,61]. The lowest false positive rate at baseline was 86% in the DANTE trial where the definition of a positive result was any abnormality of malignancy which included NCN ≥10mm or smaller with suspicious features[56]. This definition has a greater cutoff than other definitions as most studies defined a positive result as any NCN ≥5mm. Thus one would expect a smaller degree of false positives with an increased size cutoff. Furthermore, in the DANTE trial there were 13 interim cancers diagnosed because of symptoms out of 63 total, representing a higher proportion of symptom diagnosed cancers than other studies[56]. These data demonstrate that LDCT screening is not without false negative exams and has a high proportion of false positive cases, even when a strict definition of a positive result is used in a high risk population.

Surgeries for benign disease

Most studies reported surgeries resulting in a benign diagnosis, however the proportion of these from total surgeries was variable with a few studies reporting none and others reporting more than 30% with most falling in the 20-30% range. Most studies had a recommended workup protocol after a positive result, but a uniform protocol across the studies did not exist. Furthermore, the treating physician usually had full discretion to follow the recommendation for further medical evaluation. The majority of follow up included further imaging, usually high resolution CT, diagnostic CT, or PET scan. No uniform rule existed for when to biopsy, or whether the biopsy should be non-surgical or surgical. As a result of the variability among workups, one would expect variable rates of surgery for benign disease. The studies that had the least rates of surgeries for benign disease used strict workup algorithms before surgery.

Morbidity and mortality of workup

Complications were incompletely reported as only half of the studies reported observed morbidity and mortality for diagnosis or treatment. Studies reported complications of biopsy including pneumothorax, infection, respiratory failure, and stroke. Nine total deaths related to follow up were reported in total. In a follow up article to the Mayo study, the authors reported complications of workup and treatment[73]. The complication rate in this study was 27%, a value higher than that reported in any other study. The Lung Screening Study reported the next greatest number of complications related to follow up with 10 total complications, representing a much lower percentage than that in the Mayo study[81,82]. From this information, it is likely that complications of workup are underreported in the included studies and further studies should seek to report complications of work up including non-surgical and surgical biopsy. This information is vital to quantifying harms of lung cancer screening workup.

Incidental findings

Seven out of the 20 included studies reported any incidental findings. Clinically relevant findings were reported at a rate of 7 to 49% among the seven studies. Thus, IFs were underreported and variable. The ELCAP study reported a prevalence of 0.77% of mediastinal masses in the study cohort[62] whereas MacRedmond reported that 62% of those screened had an incidental finding on LDCT screening[79,80]. It appeared that only the PALCAD study completely reported incidental findings, including those that did and did not require further workup[79,80]. The Mayo study reported only clinically relevant findings but I could not determine the proportion of those screened with an incidental finding[44,45,72].

Incidental findings were explored as part of the NELSON study[95]. In 1929 participants who underwent LDCT, 129 were deemed to have clinically relevant findings that required further workup. After workup, only 21 had confirmed clinically important findings. Most were cysts and one malignancy was detected. As a result, the authors advised against searching for incidental findings which contradicts the results found in the PALCAD and Mayo studies. Importantly, the

NELSON authors did not report on aneurysms, bronchiectasis, emphysema, or coronary calcifications. From this information, it appears there is selective reporting of incidental findings, with investigators attempting to report all IFs and others reporting only those they deem clinically relevant. Incidental findings may require further workup which adds to the testing burden at unknown benefit to the individual and population. However, incidental findings are an important part of the burden of testing that should be taken into account when deciding to implement a mass screening program. As a result, additional research should be undertaken in future trials as incidental findings in lung cancer screening likely need to be further explored.

Radiation dose

Harmful effects of radiation were addressed in several studies[39,42,43,67,71,72]. The reported effective dose ranged from 0.6 to 1.5 mSv. Diederich estimated that the radiation exposure from screening cohort could induce 3-6 additional lung cancers in 15-20 years[39,40,96]. This dose is also cumulative over time. If high risk smokers underwent annual screening starting at age 50, Brenner estimated that LDCT exams would increase lung cancers by 1.8% (CI: 0.5%, 5.5%) due to radiation induced lung cancers from screening[97]. Thus, the risk of radiation is likely non-negligible and should be considered when attempting to implement a lung cancer screening program.

Overdiagnosis

Overdiagnosis represents a significant potential concern to implementing a mass screening program. Few studies provided estimates on the proportion of overdiagnosed cancers. Lindell *et al* reported that 13 of 48 CT detected cancers may have been overdiagnosed in the Mayo study as all had volume doubling times >400 days[98] (330). Autopsy data from Dammas *et al* and Manser *et al* suggest that overdiagnosis may play a role[99,100]. Dammas found that 25% of nodules detected on CT were not identified at autopsy, suggesting that incidental lung cancers are not identified even at autopsy[99]. Manser found that the majority of incidental lung cancers detected at autopsy were stage I[100]. In a critical appraisal of

overdiagnosis in lung cancer screening, Reich *et al* estimated that 25% of chest x-ray identified cancers were overdiagnosed[101]. As CT detects more cancers, it is likely that a greater proportion of cancers are overdiagnosed with CT. Thus, overdiagnosis of indolent lung cancers that never would be a problem in a person's lifetime are likely to play a significant role in the harms of screening for lung cancer and should be taken into account as a contributing harm.

Limitations of this review

I conducted the research for this review in its entirety which can introduce bias. A good quality systematic review requires at least two investigators and uses independent review with third party or group consensus for discrepancies for abstract and full text inclusion, data abstraction, and quality appraisal. It was unclear if I included all studies published in other countries as I only included English studies. There also exists a potential publication bias as I did not seek to find unpublished data.

Limitations of the evidence

Studies variably reported harm outcomes. It was difficult to accurately determine true rates of procedures for benign disease, morbidity/mortality of workup, false negatives, and incidental findings. Harms are necessary to balance the risks and benefits of a screening test and thus it is difficult to make conclusions based on incomplete evidence.

In addition to variable reporting of desired outcomes, study populations, definition of a positive, and workup algorithms were inconsistent. Study populations were heterogeneous as age and tobacco exposure varied across the studies. For a positive definition, some studies included all NCNs whereas others included only NCNs ≥5mm. Other studies used a vague definition, such as 'a nodule suspicious for lung cancer.' Thus, it is difficult to compare positivity rates across studies without a standardized definition. The workup algorithm for positive results also differed among the included studies. Studies usually included a recommendation for further workup based on the suspiciousness of the test result but participants and providers did not necessarily adhere to follow up recommendations.

Implications for practice

Currently no groups recommend mass screening for lung cancer. However, the NLST reportedly has demonstrated a 20% reduction in mortality for the LDCT screening arm compared to the control arm. Thus, it is likely that this mortality benefit will require policy makers to address lung cancer screening. Clinicians will need to consider both the harms and benefits of screening for lung cancer and balance the tradeoffs so that the benefits outweigh the harms.

Implications for research

Harm data was incompletely reported. Future trials should fully report on harms of screening. Questions that have yet to be worked out include the following:

(a) What is the most effective screening frequency?

(b) What is the appropriate level of risk to start screening based on age and exposure status?

(c) What is the best collimation of the LDCT exam?

(d) What is the most effective protocol for minimizing testing burden and unnecessary procedures?

In addition, a protocol for follow up after a positive test should be standardized so that both clinicians and patients know what to expect. I anticipate final results from randomized trials currently underway will help address these issues.

Conclusion:

The best screening test has the most beneficial tradeoff between harms and benefits. The benefits of screening are important, but in order to make informed decisions clinicians must be aware of the harms as well. Evidence collected in this review shows that harms of lung cancer screening are non-negligible to both the individual and the health care system. Even though there is evidence that harms must be considered, there is insufficient evidence to make a recommendation on what level of benefit is required to outweigh the harms of screening. Clinicians, policy-makers, and patients must weigh both before considering implementing a LDCT lung cancer screening program. I anticipate the final results of the NLST and other randomized trials to help address these concerns.

Appendix A: Introduction

Lifetime lung cancer incidence is 1 in 13 for men and 1 in 16 for women[9]. In 2007, lung cancer incidence was 69.9 cases per 100,000 person-years for males and 51.6 cases per 100,000 person-years for females[13]. The combined incidence for this year was 59.3 cases per 100,000 person-years[13]. Incidence among males has trended downward, from a peak of 102.0 cases per 100,000 person-years in 1984 whereas incidence among females has trended upward from 39.5 cases per 100,000 person-years in 1984[13]. Mortality rates from lung cancer in 2007 were 65.2 deaths per 100,000 person-years in males and 40.0 deaths per 100,000 person-years in females[13]. Combined mortality was 50.7 deaths per 100,000 person-years in 2007[13]. Mortality in males has trended down since a peak in 1987 where the mortality rate was 90.1 deaths per 100,000 person-years whereas mortality in females has increased since the 1970s until about 2000 where it stabilized at 40 deaths per 100,000 person-years[13].

Former smokers have significant smoking histories placing these individuals at increased risk for developing lung cancer. A 55 year old one pack per day smoker with a 25 year history of smoking has a 1% probability of developing cancer over the next 10 years[102]. If this person quits and remains abstinent for the next 10 years, the risk of developing lung cancer decreases to <1%[102]. This risk increases to 14% over the next 10 years for a 65 year old smoker with a 50 year history of smoking 2 packs per day[102]. If this person quits, he or she can decrease their risk to 10% over the next 10 years[102]. This demonstrates the dose dependent relationship between the exposure time (years smoking) and the magnitude of exposure (number cigarettes per day). As a result, it is expected that many new cases of lung cancer will be diagnosed in the coming years.

The 5 year survival of all stages of lung cancer is 16%[8]. The greatest survival is seen in early stages with 53% surviving at 5 five years[8]. However, survival rates can be biased measurements of improvement in prognosis. Instead of serving as a proxy for improving treatments, changes in survival rates often suffer from changing patterns of diagnosis[103]. Even

though trends of 5 year survival may improve from detection of early stage cancers in uncontrolled observational studies, this does not necessarily equate to an improvement in treatment or mortality. Lead time bias, length time bias, and overdiagnosis often plague the measurement of 5 year survival because it is often measured from observational studies[103]. Improvement in survival rates may be because we are diagnosing more cancers unnecessarily rather than improving the prognosis. As a result, I report incidence and mortality rates for lung cancer among males and females rather than 5 year survival rates.

Appendix B: Methods

To determine if a review of harms was indicated, I searched the DARE, CDSR, and NGC databases for reviews of lung cancer screening with low dose CT. Using the search term 'lung cancer screening' in DARE yielded 77 titles (21 DARE, 34 NHS, 22 HTA). Each title and abstract was reviewed for relevance to lung cancer screening with low dose CT. If studies were relevant by abstract, one reviewer read each full text to determine the relevance of the review to LDCT screening harms. Reviews and cost-effectiveness studies have mentioned harms, but none have focused on harms as a primary outcome. Some reviews focus on individual harms, but do not include all potential harms.

Search Strategies:

MEDLINE: search keywords

(lung cancer OR "Lung Neoplasms" [MeSH Terms]) AND (early detection of cancer OR mass screening OR early diagnosis) AND "tomography, x-ray computed"[MeSH Terms] OR "tomography, spiral computed" [MeSH Terms] OR "low dose CT" OR "reduced dose CT" AND (harm OR harms OR risk OR morbidity OR false positive OR false negative OR overdiagnosis OR unnecessary OR quality of life OR incidental finding OR incidentaloma OR excess radiation OR repeat imaging OR anxiety OR stigma) Limits: Humans, English

EMBASE: search keywords

(lung cancer OR Lung Neoplasms) AND (early detection of cancer OR mass screening OR early diagnosis) AND ("tomography, x-ray computed" OR "tomography, spiral computed" OR "low dose CT" OR "reduced dose CT") AND (harm OR harms OR risk OR morbidity OR false positive OR false negative OR overdiagnosis OR unnecessary OR quality of life OR incidental finding OR incidentaloma OR excess radiation OR repeat imaging OR anxiety OR stigma) OR (lung cancer OR Lung Neoplasms) AND (early detection of cancer OR mass screening OR early diagnosis) AND ("tomography, x-ray computed" OR "tomography, spiral computed") AND ("low dose" OR "low dose CT" OR "reduced dose CT")

CENTRAL: search keywords

(lung cancer OR Lung Neoplasms) AND (early detection of cancer OR mass screening OR early diagnosis) AND "tomography, x-ray computed" OR “tomography, spiral computed” OR “low dose CT“ OR “reduced dose CT” AND (harm OR harms OR risk OR morbidity OR false positive OR false negative OR overdiagnosis OR unnecessary OR quality of life OR incidental finding OR incidentaloma OR excess radiation OR repeat imaging OR anxiety OR stigma)

Appendix C: Results

Depiscan – Blanchon[54]

The Depiscan randomized trial compares annual LDCT with chest x-ray. The trial enrolled asymptomatic individuals 50-75 years old who smoked >15 cigarettes per day for >20 years. Enrollees were recruited from the care population of general practitioners involved in the study. A total of 765 individuals were randomized with 385 in the LDCT group and 380 in the CXR group. 336 of those in the LDCT group underwent the baseline LDCT screening exam. 71% were males, median age was 56 years, median smoking history was 30 pack-years, and 64% were current smokers. A positive screening result was defined as any NCN without benign features. Only baseline screening results have been published to date.

A total of 336 LDCT screening tests were performed. Baseline positivity rate was 45%. The study reported 8 lung cancers detected at baseline. Prevalence was 2.38%. 38% of the cancers were stage I at diagnosis and 63% were histologically adenocarcinoma.

The study did not report on the randomization procedure or allocation concealment. Groups were comparable at randomization. 19% of those randomized dropped out before the baseline screening exam. Withdrawals were more likely to be younger and assigned to the CXR group. 1% of those in the CXR group received a LDCT screening exam. There is a concern for selection bias given the non-comparability of those who dropped out compared to those who adhered to the study protocol. Those who dropped out may be healthier and less likely to have lung cancer. Two radiologists read each scan independently with discrepancy resolved by consensus. Overall, quality was rated as fair with fair generalizability.

Colorado – Garg[55]

This study is a pilot randomized trial comparing LDCT with no screening in a small sample of individuals. The trial enrolled individuals classified as moderate risk who were 50-80 years old with >30 pack-year history from the VA population in Denver, Colorado and individuals classified as high risk with COPD and known sputum atypia who were previously enrolled in a

cohort study. Out of the 304 eligible participants, 239 agreed to participate and 190 were randomized. Ninety-two were randomized to the LDCT group and 98 to the control group. All of those in the LDCT group underwent baseline exam. 97% of the participants were male. I could not determine overall smoking exposure history for the participants. A positive result was defined as any NCN without benign features. Only baseline screening results have been published.

A total of 92 LDCT screening tests were performed. Baseline positivity rate was 33%. The study reported 3 lung cancers detected. Prevalence was 3.26%. I could not determine the proportion of stage I and adenocarcinomas among the three cancers.

The study did not report on the randomization procedure or allocation concealment. The control group had a higher non-significant pack-year history than the screening group, but confidence intervals were not precise. All those randomized to the LDCT screening group underwent baseline exam. Crossover assessment was not reported. Scans were read by one chest radiologist with suspicious scans read independently by a second radiologist. Consensus was used for discrepancy. Masking of investigators to the study was not mentioned. Quality was rated as fair with poor generalizability given the high risk population of known sputum atypia and small sample size. This is supported by the high prevalence of lung cancer found in the small screened population.

DANTE - Infante[56-58]

The DANTE trial began in 2001 in Italy as a randomized controlled trial comparing LDCT with usual care. Recruitment took place by general practitioners, mass mailings, and media advertisements. All participants received sputum cytology and chest x-ray at baseline. The trial enrolled males 60-75 years old with greater than a 20 pack-year smoking history. Median age was 65 years old, 56% were current smokers, and overall median smoking exposure was 47 pack-years. A total of 2811 eligible participants were randomized with 2472 undergoing the baseline exam, 1276 in the LDCT arm and 1196 in the control arm. A positive result was

defined as an abnormality of malignancy including NCNs >10 mm or smaller with suspicious features. A scan was negative for NCNs <5mm.

A total of 3612 LDCT screening tests were performed. Baseline positivity rate was 16%. The study reported 63 lung cancers detected, 28 at baseline and 19 incident with 13 interim cases. Prevalence was 2.19%. Stage I disease was detected in 65% of those with cancer and 43% of total cancers were adenocarcinomas.

Adequate randomization was conducted using permuted blocks with a central system. Allocation concealment was unclear. Groups were comparable at baseline for age and smoking exposure, but the LDCT group had a higher percentage of co-morbid respiratory conditions. Of those randomized, 88% underwent the baseline exam. The study did not report adherence at 3 year follow up. A survey was conducted for a subset of participants to assess for crossover contamination. The authors reported 6.1% of participants received a non-protocol CT and 19% received a non-protocol CXR, but it was unclear if more control groups received a CT or vice versa so differential contamination was unable to be determined. Scans were read by two chest radiologists independently with discrepancy resolved by consensus. Quality was rated as fair with fair generalizability.

ELCAP - Henchske[36,37]

The Early Lung Cancer Action Project prospective study recruited 1000 asymptomatic volunteers from physician offices and hospitals in New York. Participants were >60 years old with >10 pack-year smoking history. Median age was 67 years old, 54% were male, and median smoking exposure was 45 pack-years. A positive result was defined as 1-6 NCNs without benign features.

A total of 2184 LDCT screening tests were performed. Baseline positivity rate was 23%. The study reported 36 lung cancers, 27 at baseline and 7 at incident screening with 2 interim cases. Prevalence was 2.70%. 82% of the cancers were stage I at diagnosis and 68% were histologically adenocarcinoma.

Of the baseline participants, 84% completed the first annual repeat scan. The study reported loss to follow up for 12% of baseline participants and explained reasons for dropout demonstrating acceptable adherence. Individuals were contacted who did not follow up. Scans were interpreted independently by two radiologists and it was unclear how discrepancies were resolved. Overall, quality was rated as fair with fair generalizability.

NY-ELCAP[61]

The New York Early Lung Cancer Action Project prospective study was a follow up study to the original ELCAP. This study recruited 6295 volunteers from the New York area via media and physician referral. I could not determine if there was overlap between the study cohorts so this study was treated as a separate study. Participants were >60 years old with >10 pack-year smoking history. Median age was 66 years old, 49% were male, 33% were current smokers, and median smoking exposure was 40 pack-years. A positive result at baseline was defined as at least one solid or part-solid NCN ≥5mm or nonsolid NCN ≥8mm. Repeat screening defined a positive screening exam as any new NCN or growth regardless of size.

A total of 12,309 LDCT screening tests were performed. Baseline positivity rate was 14%. The study reported 124 lung cancers, 101 at baseline and 20 at incident screening with 3 interim cases. Prevalence was 1.60%. 93% of the cancers were stage I at diagnosis and 60% were histologically adenocarcinoma.

Of the baseline participants, 82% completed the first annual repeat scan. The study reported the demographics of those who completed baseline and annual follow up exams and it appears that those who followed up were similar with respect to age, gender, and smoking status. Scans were interpreted independently by two radiologists and it was unclear how discrepancies were resolved. Overall, quality was rated as fair with fair generalizability. Of note, the patients paid for workup themselves, which has the potential to affect dropouts and selection bias if those unwilling to pay do not follow up.

Matsumoto - Sone[43,63,64]

The Matsumoto prospective study recruited 5483 volunteers from the general population via media advertising. Participants were between 40 and 74 years old and included never smokers. Mean age was 64 years old, 54% were male, and 46% were current and former smokers. This study defined current and former smokers as those with >1 pack-year exposure history. 93% of females were never smokers. Scans were classified into three categories: possible for cancer, probable for cancer, or a small nodule <3 mm.

A total of 13,786 LDCT screening tests were performed. Baseline positivity rate was 5.1%. The study reported 60 lung cancers, 22 at baseline and 36 at incident screening with four interim cases. Prevalence was 0.40%. 88% of the cancers were stage I at diagnosis and 85% were histologically adenocarcinoma. 52% of the cases were diagnosed in never smokers.

Of the 5483 participants at baseline, 80% completed the 1st annual repeat scan and 70% completed the 2nd repeat scan for a dropout rate of 30%. The study reported the demographics of those who completed the exams for each year and it appears that those who followed up were similar to baseline with respect to age and smoking status. Scans were interpreted by one of four radiologists with a second read conducted for suspicious or clinically significant scans thus the second readers were not masked to the first. Overall, quality was rated as fair with uncertain generalizability given the large proportion of never-smoking females, the overall majority was never smokers, and the majority of cases were diagnosed in never smokers.

ALCA – Sobue/Kaneko[42,65]

Two separate studies were published from the Anti-Lung Cancer Association (ALCA) prospective cohort LDCT screening project which began in 1993. This study enrolled members of the ALCA who were recruited from the general population to join a for profit organization that provided lung cancer screening to dues paying members. Included in the study were those 40-79 years old. 88% were male and 86% had any smoking history (thus 14% were non-smokers). A total of 1611 subjects were screened at baseline with LDCT (71 additional individuals had a

baseline screen but did not fit the age criteria for inclusion in the study). This study also screened its individuals with CXR and sputum cytology in addition to LDCT at 6 month intervals.

A total of 9502 LDCT screening tests were performed. Baseline positivity rate was 12%. The study reported 36 lung cancers, 14 at baseline and 22 at incident screening with no interim cases. Prevalence was 0.87%. 78% of the cancers were stage I at diagnosis and 67% were histologically adenocarcinoma.

All baseline participants received a screening exam and 73% returned for a 6 month follow up screen. Adherence dropped to less than 50% after 3 scans at 12 months and to 27% after 8 scans. It is unclear if those who adhered to the study were different from those who withdrew. This study has the potential for selection bias due to the overall poor adherence and limited generalizability due to its recruiting strategy in a membership organization. Measurement of scans was performed by 2 radiologists or chest physicians with the second reader aware of the first interpretation. A third reader determined the recommendation for undergoing further workup. In the middle of the study, a computer aided detection system was introduced which changed the reading protocol. The study was rated as fair quality with poor generalizability given the membership in the organization.

ITALUNG – Lopes Pegna[66]

The ITALUNG study is a randomized trial comparing LDCT with no screening conducted in Italy. The trial recruited individuals by invitation of patients registered with general practitioners. Eligibility included those who were 55-69 years old with at least a 20 pack-year smoking history. A total of 3206 individuals were enrolled and randomized, 1613 to the LDCT group and 1593 to the control group. 13% of those in the intervention arm did not undergo the baseline LDCT exam. Half of the participants were male, mean age was 61 years old, median smoking exposure was 39 pack-years, and 65% were current smokers. A positive result was defined as at least one solid NCN ≥5 mm or nonsolid NCN ≥10 mm or any part solid NCN. Only baseline screening results have been published.

A total of 1406 LDCT screening tests were performed. Baseline positivity rate was 30%. The study reported that 21 lung cancers were detected. Prevalence was 1.42%. Stage I disease was detected in 48% of individuals with cancer and 48% were adenocarcinomas.

Adequate randomization was performed using a central computer system. Allocation concealment was unclear. 87% of those assigned to the LDCT screening group underwent the baseline exam. Crossover was not assessed. There is a concern for selection bias here if those who did not undergo the baseline exam had different risk factor profiles for lung cancer than those that underwent the exam. Scans were read independently by 2 radiologists with discrepancy resolved for consensus. Quality was fair with fair generalizability.

COSMOS - Veronesi[67-69]

The COSMOS LDCT screening study recruited 5201 asymptomatic volunteers at a single center in Milan, Italy. Participants were >50 years old with >20 pack-year smoking histories. Mean age was 58 years old, 66% were male, and 80% were current smokers. Median exposure was 44 pack-years. A positive result was defined as any NCN detected on a screening exam.

A total of 5201 baseline LDCT screening tests were performed. Baseline positivity rate was 53%. The study reported 92 lung cancers, 55 at baseline and 36 at incident screening with one interim case. Prevalence was 1.06%. 66% of the cancers were stage I at diagnosis and 68% were histologically adenocarcinoma.

Of the 1520 participants at baseline, 93% completed the annual repeat scan for a good level of adherence. Scans were interpreted using both computer aided and manual detection by a single radiologist. Lesion size and characteristics were discussed at conferences by a team of physicians which included radiologists and surgeons. Overall, quality was rated as fair with fair generalizability as we do not know how the subjects were recruited.

Munster - Diederich[39,40,70]

This prospective study recruited 817 asymptomatic volunteers at the University of Munster in Germany using media advertising. Participants were >40 years old with >20 pack-year smoking histories. Median age was 53 years old and 72% were male. Median exposure was 45 pack-years. A positive result was defined as any NCN without benign features detected on a screening exam.

A total of 2552 baseline and repeat LDCT screening tests were performed. Baseline positivity rate was 46%. The study reported 29 lung cancers, 12 at baseline and 10 at incident screening with seven interim cases. Prevalence was 1.35%. 56% of the cancers were stage I at diagnosis and 41% were histologically adenocarcinoma.

Of the 817 participants at baseline, 82% returned for the first annual repeat scan demonstrating acceptable adherence. However, 3% of baseline participants had 5 total screening tests. Selection bias is likely to play an important role in the interpretation of the study results as funding was discontinued for those <55 years old after one normal repeat scan. Those older than 55 years continued repeat scans. This could introduce differential follow up issues as those >55 years old are more likely to have longer exposure to smoking and thus more likely to develop lung cancer. Scans were interpreted by one of two radiologists and team consensus for indeterminate lesions with unclear masking to the study. Overall, quality was rated as fair with fair generalizability however one should use caution due to selection bias.

Hitachi - Nawa[41]

The Hitachi prospective study recruited 7956 Hitachi employees in Japan for low dose CT lung cancer screening. Participants were 50-69 years old, 79% were male, and 62% were current or former smokers with 38% never smokers. A result was positive if a scan had a NCN ≥8mm without benign features.

A total of 13,524 screening tests using LDCT were performed. Baseline positivity rate was 26%. The study reported 41 lung cancers, 37 at baseline and 4 at incident screening with

no interval cases. One individual had 2 primary lung cancers. Prevalence was 0.45%. 85% of the cancers were stage I at diagnosis and 85% were histologically classified as adenocarcinoma. 57% of the cancers were detected in non-smoking individuals.

Of the participants at baseline, 70% returned for at least one follow up exam which is poor adherence compared to other included trials. Inclusion criteria were not reported. With a 30% dropout rate and unclear inclusion criteria, one must consider selection bias in this cohort, especially with 38% never smokers. The scans were read independently by 2 radiologists with discrepancy resolved at weekly conferences so measurement is on par with other studies. Overall, quality was rated as fair with poor generalizability.

Milan - Pastorino[71]

This prospective LDCT screening study recruited 1035 asymptomatic volunteers using media in Lombardy, Italy. Participants were >50 years old with >20 pack-year smoking histories. Median age was 58 years old and 71% were male. 86% and 14% were current and former smokers, respectively, with an overall median 40 pack-year history. A positive result was defined as a NCN ≥5mm detected on a screening exam.

A total of 2031 baseline and repeat LDCT screening tests were performed. Baseline positivity rate was 19%. The study reported 22 lung cancers, 11 at baseline and 11 at incident screening with no interim cases. Prevalence was 1.06%. 77% of the cancers were stage I at diagnosis and 77% were histologically adenocarcinoma.

Of the 1035 participants at baseline, 96% completed the 1st annual repeat scan with overall 91% compliance at follow up. The study reported appropriate descriptions of recruiting procedures and adherence. Scans were independently interpreted by two radiologists with interpretation by a third radiologist for discrepancies. Overall, quality was rated as fair with fair generalizability.

Mayo Clinic - Swensen[44,45,72]

The Mayo LDCT screening study recruited 1520 volunteers via media outlets at the Mayo Clinic in Minnesota. Participants were >50 years old with >20 pack-year smoking histories. Median age was 59 years old, 52% were male, 61% were current smokers, and 39% were former smokers with an overall median 45 pack-year history. A positive result was defined as any NCN detected on a screening exam.

A total of 1520 baseline LDCT screening tests were performed. Baseline positivity rate was 51%. The study reported 68 lung cancers, 31 at baseline and 34 at incident screening with three interim cases. Prevalence was 2.04%. 57% of the cancers were stage I at diagnosis and 53% were histologically adenocarcinoma.

Of the 1520 participants at baseline, 98% completed the 1st annual repeat scan, 96% completed the 2nd repeat scan, 95% completed the third, and 80% completed the fourth for overall good adherence. The study reported only one person of the 1520 was lost to follow up. The investigators attempted to control for selection bias and differential adherence by maintaining appropriate contact with subjects. Scans were interpreted by one of four radiologists with no apparent second interpretation. Overall, quality was rated as fair with fair generalizability.

PLuSS - Wilson[76,77]

The PLuSS prospective study recruited 3642 volunteers via the media at the University of Pittsburgh in Pennsylvania. Participants were 50-79 years old with > 12.5 pack-year smoking history. There was no requirement for these individuals to be asymptomatic. Mean age was 59 years old, 51% were male, median smoking exposure was 47 pack-years, and 60% were current smokers. A positive result was described as any NCN detected on a screening exam.

A total of 7065 screening tests using LDCT were performed. Baseline positivity rate was 41%. The study reported 80 lung cancers, 53 at baseline and 24 at incident screening with

three interim cases. Prevalence was 1.46%. Half of cancers were stage I at diagnosis and an unknown proportion were adenocarcinomas.

Of the baseline participants, 95% received a repeat screening exam for a good level of adherence. The scans were read by 1 of 2 radiologists and moderate and high suspicion studies were discussed in conferences. A computer algorithm was used to classify pre-biopsy risk of lung cancer. I could not determine the validity of this algorithm from included information. Overall, quality was rated as fair with fair generalizability.

PALCAD - MacRedmond[79,80]

The ProActive Lung Cancer Detection (PALCAD) prospective study recruited 449 asymptomatic individuals via the media from the local community in Dublin, Ireland. Participants were >50 years old with > 10 pack-year smoking history and still smoking at 45 years old. Median age was 56 years old, 50% were male, median smoking exposure was 45 pack-years, and 68% were current smokers. A positive result was described as any non-calcified nodule detected on a screening exam.

A total of 1371 screening tests using LDCT were performed. Baseline positivity rate was 23%. The study reported 6 lung cancers, 2 at baseline and 3 at incident screening with one interim case. Prevalence was 0.45%. Half of the 6 cancers were stage I at diagnosis and 1 of the six was an adenocarcinoma.

Of the 449 participants at baseline, 92% completed the 2 year follow up for a good level of adherence. The scans were read independently by 2 radiologists with discrepancy resolved by consensus so measurement is on par with other studies. Overall, quality was rated as fair with fair generalizability.

LSS – Gohagan[81,82]

The Lung Screening Study is a randomized trial comparing LDCT with chest x-ray for lung cancer screening. This trial enrolled volunteers recruited via mass mailing and referral who were 55-74 years old and had >30 pack-year smoking history from six PLCO centers around the

United States. Participants were excluded if they were already enrolled in the PLCO trial. 59% were male with a median smoking history of 54 pack-years. 58% were current smokers. Of the 4828 eligible participants contacted, 3318 were included in the study and randomized. 1660 were assigned to receive LDCT and 1658 were assigned to receive CXR. A positive screen was defined as any NCN >3 mm or other suspicious abnormality.

A total of 2984 LDCT screening tests were performed at baseline and annual repeat screening. Baseline positivity rate was 21%. The study reported 40 total lung cancers detected in the LDCT arm, 30 at baseline, 8 on repeat screen, and two interim cases. Prevalence was 1.89%. 48% of the cancers diagnosed were stage I and 60% were adenocarcinomas.

The study reported an adequate randomization procedure with blocks of variable size randomized by a central system with adequate allocation concealment. Groups were comparable at baseline for demographic information including smoking exposure and age. 96% of those randomized to the LDCT screening group underwent baseline exam and 86% underwent the annual repeat study for an acceptable level of adherence. Crossover was assessed using a survey sent to a random sample of participants. 1.3% to 2.6% in the CXR group received a CT scan during the study period whereas 13-20% of those in the LDCT arm received a CXR for any reason during the study period. Scans were read by a single radiologist at each study site. Results were recorded on a standardized form. I could not determine if masking was implied. Quality was rated as good with fair generalizability.

Toronto – Menezes[85]

The Toronto prospective study recruited 3352 asymptomatic participants who were volunteers who referred by a physician in Toronto, Canada. Participants were >50 years old with >10 pack-year smoking history. Median age was 60 years old, 46% were male, and median smoking exposure was 30 pack-years. A positive result was described as at least 1 solid/part-solid NCN ≥5 mm or non-solid NCN ≥8 mm on baseline and any growth or new NCN on repeat screening.

A total of 6924 screening tests using LDCT were performed. Baseline positivity rate was 18%. The study reported 65 lung cancers, 56 at baseline and 6 at incident screening with 3 interim cases. Prevalence was 1.67%. 65% of cancers were stage I at diagnosis and 68% were histologically adenocarcinoma.

Of the 3352 baseline participants, 80% completed at least one annual screen demonstrating 20% adherence. The study does not report on the characteristics of those who were lost to follow so there is a concern for selection bias. The scans were read according to the I-ELCAP protocol, but the study does not report any further information about who read them. Overall, quality was rated as fair with fair generalizability.

Israel – Shaham[86]

This prospective study recruited 842 current and former smokers by mixed recruiting strategies including physician referral and media recruiting in Israel. 32% were required to pay for the screening exam. Participants were >50 years old with a >10 pack-year smoking history. Median age was 56 years old, 57% were male, and median smoking history was 37 pack-years. The study did not report the proportion of current to former smokers. A positive result initially followed the I-ELCAP protocol with 1-6 NCN ≥5 mm and then was adjusted during the study to 1-6 solid or part solid NCN ≥5 mm or non-solid ≥8 mm on baseline and any new NCN or growth on repeat scan.

A total of 1784 screening tests using LDCT were performed. Baseline positivity rate was 12%. The study reported 14 lung cancers, 12 at baseline and 2 at incident screening with no interim cases. Prevalence was 1.43%. 86% of the cancers were stage I at diagnosis and 57% were classified as adenocarcinomas.

Of the 842 participants at baseline, 801 were eligible for repeat screen and 69% of them had an annual repeat scan representing a 31% dropout rate for a low adherence rate compared to other included studies. Selection bias is likely to play a significant role here as the study used different recruiting strategies and 32% were expected to pay for their exams. It is possible that

those who were required to pay for their exams were more likely to not follow up, however the study did not assess differential loss to follow up. In addition, the study changed the definition of a positive result at some point during follow up, however it was not assessed how this changed the positivity rate or work up of participants. All scans were read by a single radiologist which could induce systematic interpreting error. The study changed LDCT slice thickness during follow up as well so differential detection of nodules could be affected. Overall, quality was rated as fair with uncertain generalizability.

DLCST – Pedersen[87]

The Danish Lung Cancer Screening Trial (DLCST) is a randomized trial comparing LDCT with no screening for lung cancer detection over a period of five years in centers around Denmark. This trial enrolled volunteers recruited via media who were 50-70 years old and had a >20 pack-year smoking history. 55% were male and 76% were current smokers. Of the 5861 participants who volunteered, 4104 were eligible and randomized. 2052 were assigned to receive annual LDCT for five years and 2052 were assigned to receive no screening. A positive screen was defined as any NCN ≥5 mm or any growing nodule. Only baseline results are currently published.

A total of 2047 LDCT screening tests were performed at baseline. Baseline positivity rate was 8.7%. The study reported 17 total lung cancers detected in the LDCT arm. Prevalence was 0.83%. 53% of the cancers diagnosed were stage I and 71% were adenocarcinomas.

The study reported an adequate randomization procedure with blocks randomized by a central system with unknown concealment. Groups were comparable at baseline for demographics. Five of the participants randomized to the LDCT screening group failed to undergo baseline exam for baseline adherence >99%. It is unclear from the article if crossover was assessed. Scans were read by a 2 radiologists with unknown masking, and discrepancy

was resolved by consensus. I could not determine if blinding was performed for investigators or participants. Quality was rated as fair with fair generalizability.

Toronto – Roberts[89]

The I-ELCAP Toronto prospective study reported by Roberts et al recruited 1000 volunteers >55 years old with a >10 pack-year history. Median age was 63 years old, 45% were male, and median smoking exposure was 34-38 pack-years. 34% were current smokers and 66% former smokers. A positive result was described ≥1 solid/part-solid NCN ≥5 mm or non-solid ≥8 mm on baseline and any growth or new NCN on repeat screening. Only baseline results were reported.

A total of 1000 screening tests using LDCT were performed. Baseline positivity rate was 26%. The study reported 20 lung cancers were detected on screening. Prevalence was 2.0%. 75% of cancers were stage I at diagnosis and 70% were histologically adenocarcinoma.

Since the study did not report on repeat screening exams, I could not determine adherence to the study protocol. The scans were interpreted according to the I-ELCAP protocol, but the study does not report any further information regarding who or how many people provided interpretations. Masking was not mentioned. Overall, quality was rated as fair with fair generalizability.

Appendix D: Discussion

Prior Systematic Reviews of LDCT screening

Of the 8 systematic reviews identified in my search strategy, all included a discussion of harms of screening, however limited data on harms prevented pooling of results. Here, I individually discuss the evidence and implications of each for screening harms.

In a short review designed to answer the question of whether screening improves long term survival in asymptomatic lung cancer patients, Hunt *et al* briefly mention false positives and overdiagnosis as harms of screening[104]. This review included observational studies, had a comprehensive search strategy, but did not mention eligibility criteria for articles and did not give a quality assessment of included articles. Thus, it was rated as a poor quality review.

In a systematic review of the prevalence of incidental findings, Jacobs *et al* included populations undergoing coronary artery disease screening and lung cancer screening using CT[105]. The authors grouped incidental findings as lesions representing airway disease, renal disease, liver disease, breast abnormalities, aortic aneurysm, pericardial disease, pleural disease, lymphadenopathy, adrenal mass, gastric tumor, and thyroid disease. Four primary studies were included that screened for lung cancer. This review found that the mean proportion of patients with at least one incidental finding on screening exam was 65.2% (63.5 – 66.9%). This represents about two-thirds of those screened who will have an abnormality on their screening exam not related to the exam's primary purpose. The authors also found that an average of 14.2% (13.2 – 15.2%) of participants in lung cancer screening had further workup for an incidental finding. The authors found that the PALCAD study had the highest rate of clinically significant findings with 26.9% (22.8 – 31.1%) of those screening undergoing further workup for an incidental finding after screening. The review corroborates my finding that variation exists among studies in how they defined an abnormality as clinically relevant. For instance, bronchiectasis and pulmonary fibrosis were included as clinically important in some studies[72,79], but not in the NELSON study[95]. In addition, my conclusion that variation exists in

the reporting of incidental findings in lung cancer screening studies is supported by this review. In addition to incidental findings, this review discussed false positives, anxiety, and costs of workup as harms of screening. This is a fair quality systematic review as the authors included study population characteristics, based their review on a focused key question, had an appropriate search, addressed heterogeneity, but did not report a quality assessment of included studies.

In a systematic review of baseline LDCT screening characteristics, Yau and colleagues sought to assess current knowledge of LDCT screening to determine if LDCT should be introduced for lung cancer screening in a high risk population[47]. The authors reported LDCT test characteristics including a median sensitivity of 81%, median specificity of 81%, positive predictive value of 8%, and negative predictive value of 99%. Yau reported that 80% of all lung cancers detected on screening were stage I NSCLC. The authors discussed false positives, false negatives, anxiety of workup, and costs as harms of screening. Incidental findings, surgery for benign procedures, radiation exposure, and other harms were not mentioned. This review was rated as good as the authors conducted an appropriate search, discussed eligibility criteria, sought unpublished data, and used standard methods to assess quality.

In a systematic review and meta-analysis of baseline findings of randomized trials, Gopal *et al* reported that LDCT screening detects greater numbers of stage I cancers, total cancers, and NSCLCs[106]. The authors pooled data of 14,055 participants in randomized trials of LDCT screening versus a control group (either no screening or CXR). The odds of detecting a Stage I cancer using LDCT was 3.9 (CI: 2.0, 7.4) times as likely as that of the control group and the odds of detecting a NSCLC was 5.5 (CI: 3.1, 9.6) times as likely as that of the control group. Participants in the LDCT arm were 3.1 (2.6-3.7) times as likely to have a false positive screening test compared to controls at baseline. The authors also reported increased rates of unnecessary thoracotomies for benign lesions in the LDCT arm. The event rate of thoracotomy was 3.7 per 1000 screening tests (3.5-3.8) over the study period. Participants in the LDCT arm

were 4 times as likely to have a thoracotomy for a benign lesion as in the control arm. The authors reported that for every 1000 individuals screened, 9 stage I NSCLCs were detected, 235 false positives occurred, and 4 individuals underwent thoracotomies for benign lesions. This fits with my conclusions regarding the high rate of false positives and unnecessary surgeries. The authors also discussed overdiagnosis, further workup, anxiety, and cost as harms of screening. This was a good quality systematic review as the authors included a discussion of the study populations, attempted to find unpublished data, included non-English trials if available, assessed internal validity of included studies, and addressed heterogeneity in control groups.

In a systematic review of observational studies of LDCT screening, Manser *et al* included 8 observational studies as a component of a larger Cochrane Review[107]. This review focused on the benefits of screening, but the authors discussed overdiagnosis, cancers missed by CT, and benign non-surgical and surgical biopsies. Conclusions regarding overdiagnosis and procedures conducted for benign disease were similar to what I found in the included studies in this review. This review was rated as a fair as the authors conducted an appropriate search, but did not mention an attempt to find unpublished data and it was unclear if non-English trials included in search. The authors also did not report an assessment of internal validity of included studies and did not assess heterogeneity.

In a systematic review to evaluate the evidence for lung cancer screening with chest x-ray, sputum, and LDCT, Humphrey *et al* included 6 observational LDCT screening studies[2]. The authors concluded that LDCT diagnosed lung cancer at an earlier stage but they were unable to determine if a mortality benefit exists. The authors discussed false positives, false negatives, anxiety, overdiagnosis, recommendations for further workup, surgery for benign disease, and morbidity and mortality of workup as harms of screening with conclusions that were similar to what I found in my review. This review was rated as good as the authors used an appropriate search, included a discussion of eligibility criteria and study populations, and

performed a quality assessment but did not report the internal validity of included studies. It was also unclear if the authors attempted to find unpublished data or if non-English trials were included.

In a systematic review to evaluate the evidence for clinical effectiveness of LDCT screening, Black *et al* included 12 studies, 2 of which were short randomized trials[3]. The authors reported the proportion of people with an abnormal screening result was 5-51% which is similar to the positivity rate I report in this review. The authors mention radiation dose, anxiety, false positives, quality of life, and adverse events of screening as harms. The authors concluded that screening harms were poorly reported. This review was rated as good due to the discussion of eligibility criteria and study population, the inclusion of non-English articles, and the discussion of the internal validity of the included studies.

In a systematic review to determine whether LDCT reduces mortality, Bepler *et al* included 8 observational studies[4]. The authors included screening related morbidity as a secondary outcome and concluded that screening related data were not well reported. This review was rated as fair due to the appropriate search strategy, discussion of reasons for not pooling data, eligibility criteria, inclusion of non-English articles in search, and the assessment of quality of included studies.

References:

1. Manser RL, Irving LB, Stone C, et al. Screening for lung cancer. *Cochrane Database Syst Rev*. 2004;(1):CD001991.

2. Humphrey LL, Teutsch S, Johnson M. Lung Cancer Screening with Sputum Cytologic Examination, Chest Radiography, and Computed Tomography: An Update for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* 2004;140(9):740-753+I88.

3. Black C, de Verteuil R, Walker S, et al. Population screening for lung cancer using computed tomography, is there evidence of clinical effectiveness? A systematic review of the literature. *Thorax*. 2007;62(2):131-138.

4. Bepler G, Carney DG, Djulbegovic B, Clark RA, Tockman M. A systematic review and lessons learned from early lung cancer detection trials using low-dose computed tomography of the chest. *Cancer Control*. 2003;10(4):306-314.

5. Bach PB, Silvestri GA, Hanger M, Jett JR. Screening for lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest*. 2007;132(3 Suppl):69S-77S.

6. USPSTF. Lung cancer screening: recommendation statement. *Ann. Intern. Med*. 2004;140(9):738-739.

7. Lung cancer trial results show mortality benefit with CT - National Cancer Institute. Available at: http://www.cancer.gov/newscenter/pressreleases/NLSTresultsRelease. Accessed November 26, 2010.

8. American Cancer Society. *Cancer Facts & Figures 2010*. Atlanta: American Cancer Society; 2010.

9. American Cancer Society. *Global Cancer Facts & Figures 2nd Edition*. Atlanta: American Cancer Society; 2011. Available: http://www.cancer.org/Research/CancerFactsFigures/GlobalCancerFactsFigures/global-facts-figures-2nd-ed.

10. Ferlay J, Shin H-R, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*. 2010. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20560135. Accessed March 28, 2011.

11. U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 1999–2007 Incidence and Mortality Web-based Report*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2010. Available at: www.cdc.gov/uscs.

12. Xu JQ, Kochanek KD, Murphy SL, Tejada-Vera B. Deaths: Final data for 2007. National vital statistics reports; vol 58 no 19. Hyattsville, MD: National Center for Health Statistics. 2010.

13. Altekruse SF, Kosary CL, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2007, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2007/.

14. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer*. 2010;116(3):544-573.

15. U.S. Department of Health and Human Services. *The Health Consequences of Smoking: A Report of the Surgeon General.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2004.

16. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA*. 2004;291(10):1238-1245.

17. CDC. Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. *MMWR Morb. Mortal. Wkly. Rep*. 2008;57(45):1226-1228.

18. Burns DM. Primary prevention, smoking, and smoking cessation: implications for future trends in lung cancer prevention. *Cancer*. 2000;89(11 Suppl):2506-2509.

19. Halpern MT, Gillespie BW, Warner KE. Patterns of absolute risk of lung cancer mortality in former smokers. *J. Natl. Cancer Inst*. 1993;85(6):457-464.

20. CDC. Vital signs: current cigarette smoking among adults aged >or=18 years --- United States, 2009. *MMWR Morb. Mortal. Wkly. Rep*. 2010;59(35):1135-1140.

21. Jett JR. Current treatment of unresectable lung cancer. *Mayo Clin. Proc*. 1993;68(6):603-611.

22. Welch HG, Woloshin S, Schwartz LM, et al. Overstating the evidence for lung cancer screening: the International Early Lung Cancer Action Program (I-ELCAP) study. *Arch. Intern. Med*. 2007;167(21):2289-2295.

23. Boucot KR, Weiss W. Is curable lung cancer detected by semiannual screening? *JAMA*. 1973;224(10):1361-1365.

24. An evaluation of radiologic and cytologic screening for the early detection of lung cancer: a cooperative pilot study of the American Cancer Society and the Veterans Administration. *Cancer Res*. 1966;26(10):2083-2121.

25. Nash FA, Morgan JM, Tomkins JG. South London Lung Cancer Study. *Br Med J*. 1968;2(5607):715-721.

26. Brett GZ. The value of lung cancer detection by six-monthly chest radiographs. *Thorax*. 1968;23(4):414-420.

27. Brett GZ. Earlier diagnosis and survival in lung cancer. *Br Med J*. 1969;4(5678):260-262.

28. Dales LG, Friedman GD, Collen MF. Evaluating periodic multiphasic health checkups: a controlled trial. *J Chronic Dis*. 1979;32(5):385-404.

29. Friedman GD, Collen MF, Fireman BH. Multiphasic Health Checkup Evaluation: a 16-year follow-up. *J Chronic Dis*. 1986;39(6):453-463.

30. Kubik A, Parkin DM, Khlat M, et al. Lack of benefit from semi-annual screening for cancer of the lung: follow-up report of a randomized controlled trial on a population of high-risk males in Czechoslovakia. *Int. J. Cancer*. 1990;45(1):26-33.

31. Kubík A, Polák J. Lung cancer detection. Results of a randomized prospective study in Czechoslovakia. *Cancer*. 1986;57(12):2427-2437.

32. Ebeling K, Nischan P. Screening for lung cancer--results from a case-control study. *Int. J. Cancer*. 1987;40(2):141-144.

33. Sobue T, Suzuki T, Naruke T. A case-control study for evaluating lung-cancer screening in Japan. Japanese Lung-Cancer-Screening Research Group. *Int. J. Cancer*. 1992;50(2):230-237.

34. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials*. 2000;21(6 Suppl):273S-309S.

35. Naidich DP, Marshall CH, Gribbin C, Arams RS, McCauley DI. Low-dose CT of the lungs: preliminary observations. *Radiology*. 1990;175(3):729-731.

36. Henschke CI, Naidich DP, Yankelevitz DF, et al. Early lung cancer action project: initial findings on repeat screenings. *Cancer*. 2001;92(1):153-159.

37. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: Overall design and findings from baseline screening. *Lancet*. 1999;354(9173):99-105.

38. Henschke CI, Yankelevitz DF, Libby DM, et al. Survival of patients with stage I lung cancer detected on CT screening. *New Engl. J. Med.* 2006;355(17):1763-1771.

39. Diederich S, Thomas M, Semik M, et al. Screening for early lung cancer with low-dose spiral computed tomography: Results of annual follow-up examinations in asymptomatic smokers. *Eur. Radiol.* 2004;14(4):691-702.

40. Diederich S, Wormanns D, Semik M, et al. Screening for early lung cancer with low-dose spiral CT: Prevalence in 817 asymptomatic smokers. *Radiology*. 2002;222(3):773-781.

41. Nawa T, Nakagawa T, Kusano S, et al. Lung cancer screening using low-dose spiral CT: Results of baseline and 1-year follow-up studies. *Chest*. 2002;122(1):15-20.

42. Sobue T, Moriyama N, Kaneko M, et al. Screening for lung cancer with low-dose helical computed tomography: Anti-Lung Cancer Association project. *J. Clin. Oncol.* 2002;20(4):911-920.

43. Sone S, Li F, Yang Z-G, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br. J. Cancer*. 2001;84(1):25-32.

44. Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am. J. Respir. Crit. Care Med.* 2002;165(4):508-513.

45. Swensen SJ, Jett JR, Hartman TE, et al. Lung cancer screening with CT: Mayo clinic experience. *Radiology*. 2003;226(3):756-761.

46. Tiitola M, Kivisaari L, Huuskonen MS, et al. Computed tomography screening for lung cancer in asbestos-exposed workers. *Lung Cancer*. 2002;35(1):17-22.

47. Yau G, Lock M, Rodrigues G. Systematic review of baseline low-dose CT lung cancer screening. *Lung Cancer*. 2007;58(2):161-170.

48. Aberle DR, Berg CD, Black WC, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011;258(1):243-253.

49. Aberle DR, Adams AM, Berg CD, et al. Baseline characteristics of participants in the randomized national lung screening trial. *J. Natl. Cancer Inst*. 2010;102(23):1771-1779.

50. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int. J. Cancer*. 2007;120(4):868-874.

51. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001;20(3 Suppl):21-35.

52. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann. Intern. Med*. 2009;151(4):W65-94.

53. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med*. 2009;151(4):264-269, W64.

54. Blanchon T, Brechot J-M, Grenier PA, et al. Baseline results of the Depiscan study: A French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR). *Lung Cancer*. 2007;58(1):50-58.

55. Garg K, Keith RL, Byers T, et al. Randomized controlled trial with low-dose spiral CT for lung cancer screening: Feasibility study and preliminary results. *Radiology*. 2002;225(2):506-510.

56. Infante M, Cavuto S, Lutman FR, et al. A randomized study of lung cancer screening with spiral computed tomography: Three-year results from the DANTE trial. *Am. J. Respir. Crit. Care Med.* 2009;180(5):445-453.

57. Infante M, Chiesa G, Solomon D, et al. Surgical procedures in the DANTE trial, a randomized study of lung cancer early detection with spiral computed tomography: Comparative analysis in the screening and control arm. *J. Thorac. Oncol.* 2011;6(2):327-335.

58. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer*. 2008;59(3):355-363.

59. Anderson CM, Yip R, Henschke CI, et al. Smoking cessation and relapse during a lung cancer screening program. *Cancer Epidemiol. Biomarkers Prev*. 2009;18(12):3476-3483.

60. Ostroff JS, Buckshee N, Mancuso CA, Yankelevitz DF, Henschke CI. Smoking cessation following CT screening for early detection of lung cancer. *Prev. Med.* 2001;33(6):613-621.

61. CT Screening for lung cancer: diagnoses resulting from the New York Early Lung Cancer Action Project. *Radiology*. 2007;243(1):239-249.

62. Henschke CI, Lee I-J, Wu N, et al. CT screening for lung cancer: prevalence and incidence of mediastinal masses. *Radiology*. 2006;239(2):586-590.

63. Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet*. 1998;351(9111):1242-1245.

64. Sone S, Nakayama T, Honda T, et al. Long-term follow-up study of a population-based 1996-1998 mass screening programme for lung cancer using mobile low-dose spiral computed tomography. *Lung Cancer*. 2007;58(3):329-341.

65. Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: Screening and detection with low-dose spinal CT versus radiography. *RADIOLOGY*. 1996;201(3):798-802.

66. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer*. 2009;64(1):34-40.

67. Veronesi G, Bellomi M, Veronesi U, et al. Role of Positron Emission Tomography Scanning in the Management of Lung Nodules Detected at Baseline Computed Tomography Screening. *Ann. Thorac. Surg.* 2007;84(3):959-966.

68. Veronesi G, Bellomi M, Mulshine JL, et al. Lung cancer screening with low-dose computed tomography: A non-invasive diagnostic protocol for baseline lung nodules. *Lung Cancer*. 2008;61(3):340-349.

69. Veronesi G, Bellomi M, Scanagatta P, et al. Difficulties encountered managing nodules detected during a computed tomography lung cancer screening program. *J. Thorac. Cardiovasc. Surg.* 2008;136(3):611-617.

70. Diederich S. Screening for early lung cancer with low-dose spiral computed tomography. *Lancet*. 2003;362(9384):588-589.

71. Pastorino U, Bellomi M, Landoni C, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-Year results. *Lancet*. 2003;362(9384):593-597.

72. Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: Five-year prospective experience. *Radiology*. 2005;235(1):259-265.

73. Crestanello JA, Allen MS, Jett JR, et al. Thoracic surgical operations in patients enrolled in a computed tomographic screening trial. *J. Thorac. Cardiovasc. Surg*. 2004;128(2):254-259.

74. Cox LS, Clark MM, Jett JR, et al. Change in Smoking Status after Spiral Chest Computed Tomography Scan Screening. *Cancer*. 2003;98(11):2495-2501.

75. Townsend CO, Clark MM, Jett JR, et al. Relation between smoking cessation and receiving results from three annual spiral chest computed tomography scans for lung carcinoma screening. *Cancer*. 2005;103(10):2154-2162.

76. Wilson DO, Weissfeld JL, Balkan A, et al. Association of radiographic emphysema and airflow obstruction with lung cancer. *Am. J. Respir. Crit. Care Med.* 2008;178(7):738-744.

77. Wilson DO, Weissfeld JL, Fuhrman CR, et al. The Pittsburgh Lung Screening Study (PLuSS): outcomes within 3 years of a first computed tomography scan. *Am. J. Respir. Crit. Care Med*. 2008;178(9):956-961.

78. Byrne MM, Koru-Sengul T, Zhao W, Weissfeld JL, Roberts MS. Healthcare use after screening for lung cancer. *Cancer*. 2010;116(20):4793-4799.

79. MacRedmond R, Logan PM, Lee M, et al. Screening for lung cancer using low dose CT scanning. *Thorax*. 2004;59(3):237-241.

80. MacRedmond R, McVey G, Lee M, et al. Screening for lung cancer using low dose CT scanning: Results of 2 year follow up. *Thorax*. 2006;61(1):54-56.

81. Gohagan J, Marcus P, Fagerstrom R, et al. Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute. *Chest*. 2004;126(1):114-121.

82. Gohagan JK, Marcus PM, Fagerstrom RM, et al. Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. *Lung Cancer*. 2005;47(1):9-15.

83. Pinsky P, Freedman M, Oken M, et al. Prevalence of non-cancer-related abnormalities on low-dose spiral computer tomography versus chest radiograph in a screening population [1]. *Thorax*. 2007;62(2):190.

84. Croswell JM, Baker SG, Marcus PM, Clapp JD, Kramer BS. Cumulative incidence of false-positive test results in lung cancer screening: A randomized trial. *Ann. Intern. Med.* 2010;152(8):505-512.

85. Menezes RJ, Roberts HC, Paul NS, et al. Lung cancer screening using low-dose computed tomography in at-risk individuals: The Toronto experience. *Lung Cancer*. 2010;67(2):177-183.

86. Shaham D, Breuer R, Copel L, et al. Computed tomography screening for lung cancer: Applicability of an international Protocol in a single-Institution environment. *Clin. Lung Cancer*. 2006;7(4):262-267.

87. Pedersen JH, Ashraf H, Dirksen A, et al. The danish randomized lung cancer ct screening trial- overall design and results of the prevalence round. *J. Thorac. Oncol.* 2009;4(5):608-614.

88. Ashraf H, Tønnesen P, Holst Pedersen J, et al. Effect of CT screening on smoking habits at 1-year follow-up in the Danish Lung Cancer Screening Trial (DLCST). *Thorax*. 2009;64(5):388-392.

89. Roberts HC, Patsios D, Paul NS, et al. Lung cancer screening with low-dose computed tomography: Canadian experience. *Can. Assoc. Radiol. J.* 2007;58(4):225-235.

90. van den Bergh KAM, Essink-Bot M-L, Bunge EM, et al. Impact of computed tomography screening for lung cancer on participants in a randomized controlled trial (NELSON trial). *Cancer*. 2008;113(2):396-404.

91. Bunge EM, van den Bergh KAM, Essink-Bot M-L, van Klaveren RJ, de Koning HJ. High affective risk perception is associated with more lung cancer-specific distress in CT screening for lung cancer. *Lung Cancer*. 2008;62(3):385-390.

92. van den Bergh KAM, Essink-Bot ML, Borsboom GJJM, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). *Br. J. Cancer*. 2010;102(1):27-34.

93. Li F, Sone S, Abe H, et al. Lung cancers missed at low-dose helical CT screening in a general population: Comparison of clinical, histopathologic, and imaging findings. *Radiology*. 2002;225(3):673-683.

94. Henschke CI, Yankelevitz DF, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. *Radiology*. 2004;231(1):164-168.

95. van de Wiel JCM, Wang Y, Xu DM, et al. Neglectable benefit of searching for incidental findings in the Dutch-Belgian lung cancer screening trial (NELSON) using low-dose multidetector CT. *Eur. Radiol.* 2007;17(6):1474-1482.

96. Diederich S, Wormanns D, Lenzen H, et al. Screening for asymptomatic early bronchogenic carcinoma with low dose CT of the chest. *Cancer*. 2000;89(11 SUPPL.):2483-2484.

97. Brenner DJ. Radiation Risks Potentially Associated with Low-Dose CT Screening of Adult Smokers for Lung Cancer. *Radiology*. 2004;231(2):440-445.

98. Lindell RM, Hartman TE, Swensen SJ, et al. Five-year lung cancer screening experience: CT appearance, growth rate, location, and histologic features of 61 lung cancers. *Radiology*. 2007;242(2):555-562.

99. Dammas S, Patz EF, Goodman PC. Identification of small lung nodules at autopsy: implications for lung cancer screening and overdiagnosis bias. *Lung Cancer*. 2001;33(1):11-16.

100. Manser RL, Dodd M, Byrnes G, Irving LB, Campbell DA. Incidental lung cancers identified at coronial autopsy: implications for overdiagnosis of lung cancer by screening. *Respir Med*. 2005;99(4):501-507.

101. Reich JM. A critical appraisal of overdiagnosis: Estimates of its magnitude and implications for lung cancer screening. *Thorax*. 2008;63(4):377-383.

102. Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *J. Natl. Cancer Inst*. 2003;95(6):470-478.

103. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? *JAMA*. 2000;283(22):2975-2978.

104. Hunt I, Siva M, Southon R, Treasure T. Does lung cancer screening with low-dose computerised tomography (LDCT) improve disease-free survival? *Interact. Cardiovasc. Thorac. Surg.* 2006;5(5):612-615.

105. Jacobs PCA, Mali WPTM, Grobbee DE, van der Graaf Y. Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. *J Comput Assist Tomogr*. 2008;32(2):214-221.

106. Gopal M, Abdullah SE, Grady JJ, Goodwin JS. Screening for lung cancer with low-dose computed tomography: A systematic review and meta-analysis of the baseline findings of randomized controlled trials. *J. Thorac. Oncol.* 2010;5(8):1233-1239.

107. Manser RL, Irving LB, De Campo MP, et al. Overview of observational studies of low-dose helical computed tomography screening for lung cancer. *Respirology*. 2005;10(1):97-104.

Glossary of acronyms:

NCN: non-calcified nodule
PLuSS: Pittsburgh Lung Screening Study
LDCT: low dose computed tomography
PALCAD: ProActive Lung Cancer Detection study
LSS: Lung Screening Study
IF: incidental finding
DLCST: Danish Lung Cancer Screening Trial
DANTE: RCT from Milan, Italy group
CXR: chest x-ray
ELCAP: Early Lung Cancer Action Project
NY-ELCAP: New York ELCAP
I-ELCAP: International ELCAP
ALCA: Anti-Lung Cancer Association
ITALUNG: RCT from Florence, Italy group
PLCO: Prostate, Lung, Colorectal, and Ovarian Cancer study
NLST: National Lung Screening Trial
NSCLC: non-small cell lung cancer
RCT: randomized controlled trial