Review Article

Review of the future of lung cancer screening

Rowena Yip¹, Artit Jirapatnakul¹, Meng Li¹², David F. Yankelevitz¹, Claudia I. Henschke¹

¹Department of Radiology, Mount Sinai Health System, New York, NY, USA; ²Department of Diagnostic Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Contributions: (I) Conception and design: R Yip, A Jirapatnakul, DF Yankelevitz, CI Henschke; (II) Administrative support: DF Yankelevitz, CI Henschke; (III) Provision of study materials or patients: R Yip, A Jirapatnakul, DF Yankelevitz, CI Henschke; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Claudia I. Henschke, PhD, MD. Department of Radiology, One Gustave L. Levy Place, Box 1234, New York, NY 10029, USA. Email: Claudia.Henschke@mountsinai.org.

Abstract: Lung cancer screening has progressed greatly in the last two decades since the initial publication of ELCAP in 1999. Since then, large randomized trials, the NLST and NELSON studies, have also shown the benefit of lung cancer screening with CT, and it is now widely available. There are many areas for improving the benefit of screening by refining the eligibility criteria, continued optimization of the screening regimen, identifying other related diseases, and improved treatment. As artificial intelligence (AI) techniques continue to improve, an automated report of all diseases can be produced from the low-dose CT scan. New screening tests are being developed for lung cancer including blood-based tests. There is growing recognition that use of risk factor-based criteria which dichotomize age and pack-years of smoking to select eligible individuals for screening may not be the most efficient and may also lead to increased lung cancer disparity. Continued research will focus on optimal approaches for selecting high risk individuals for lung cancer screening and thus hopefully expand eligibility criteria beyond the current criteria which include only 20% of lung cancer patients. With the increase in early stage lungs cancers identified from lung cancer screening, a key focus of ongoing research is the treatment of these cancers.

Keywords: Lung cancer; screening; future

Received: 09 July, 2020. Accepted: 25 June, 2021.
doi: 10.21037/ccts-20-125
View this article at: https://dx.doi.org/10.21037/ccts-20-125

Introduction

In the twenty years since the authors initial Early Lung Cancer Action Project (ELCAP) report in 1999 (1-3), low-dose computed tomography (LDCT) screening has been implemented throughout the world (4-11). At the same time, a remarkable transformation in medicine is starting, due in part to the advancing knowledge in genetics, mathematical advances, and increased computing power.

LDCT radiation doses are currently below the dose of mammographic studies. Ultra LDCT have radiation doses approaching those of chest X-ray (CXR) and are being used for evaluation of chest and heart diseases (12). New image analytics and statistical techniques have been developed, with even more innovations on the horizon. Future image interpretation will increasingly use computer-aided-diagnostics, already started as early as the 1990’s (12). Further developments will continue to improve the assessment of three-dimensional volumetric doubling time (VDT), already introduced in the mid 1990’s by ELCAP. As the experience with LDCT continues to accumulate, the future application of LDCT will be to provide a comprehensive “health check,” perhaps together with expanded routine blood testing.

Possible near-term advances are:
(I) Expansion of the eligibility criteria for screening using risk-based criteria, particularly to women,
never smokers exposed to secondhand tobacco smoke, and to people of younger age.

(II) Integration of artificial intelligence (AI) approaches for LDCT interpretation.

(III) Continued optimization of the regimen of screening.

(IV) Understanding the benefit of identification of other early diseases.

(V) Evaluation of new screening tests will emerge.

(VI) New paradigms of treatment for early lung cancers.

Future development of screening will rely on comprehensive management systems that are integrated into hospital electronic medical record systems and also provide for outreach to potential screening participants. The International Early Lung Cancer Action Program (I-ELCAP) used the ELCAP Management System and provided many new insights for further development of screening regimens. Its success demonstrates the usefulness of comprehensive management systems which provide quality assurance of the interpretation and follow-up management needed in screening programs. The ELCAP Management System started in 1992 and has continued to be updated and used throughout the world. Its database has allowed for further development of the LDCT screening regimen and its efficiency as demonstrated by its contributions (12). The ELCAP Management System has now been translated into an open-source management system called VAPALS-ELCAP Management System for the Veterans Administration Health Care System. These large emerging high-quality databases, developed for comprehensive clinical management of screening, will become invaluable for further development of screening for lung cancer. We present the following article in accordance with the Narrative Review reporting checklist (available at https://dx.doi.org/10.21037/ccts-20-125).

Expansion of the eligibility criteria for screening using risk-based criteria, particularly to women, never smokers exposed to secondhand tobacco smoke, and to people of younger age

The National Lung Screening Trial (NLST) (6) and NELSON trial (11) showed that LDCT screening decreases lung cancer mortality rates for current and former smokers when compared with chest radiographic screening or no screening, respectively. The NLST criteria for screening was current and former smokers with at least 30 pack-years of smoking and for the former smokers, only those who had quit in the last 15 years prior to enrollment.

The authors have previously shown that women are at higher risk to develop lung cancer than equally smoking men (13-16). When the probability of lung cancer among I-ELCAP smokers with sex, age, pack-years of smoking and years since quit was modeled a 55-year-old man with at least 30 pack-years of smoking would have about the same risk as a 47-year-old woman with the same smoking history (17). This result suggested that if the NLST entry criteria used 55 as the age cutoff for men to be eligible for screening, the corresponding age cutoff for women should be lower, at around 47 years old (17). This finding corroborates other studies that women with lung cancer are diagnosed at a younger age than men (18-20). Expanding screening to include women at a lower age would allow identification of those high-risk women who are currently outside of the guidelines. In fact, results of the NELSON trial also support the consideration to extend screening to younger smokers. The NELSON trial (7) included smokers aged 50 and older with a lower smoking history than the NLST, and the trial showed a greater lung cancer mortality reduction when comparing LDCT screening with no screening. Expansion of the NLST criteria is being considered in the United States and elsewhere as it has been estimated that only 20% of the lung cancer patients diagnosed each year in the United States met these criteria.

Recently, lung cancer in never smokers, particularly in women has attracted attention. Never smokers are defined as having smoked less than 100 cigarettes in their lifetime (21). It is estimated that 10-25% of lung cancers occur in never-smokers so that lung cancer in never smokers is the 7th most common malignancy resulting in approximately 300,000 deaths annually (22-24). Lung cancer in never smokers has been considered by some to be a “different lung cancer” compared to smoking associated lung cancer (22,25). Epidemiologic studies showed that lung cancer in never smokers occurred more frequently in women (53%) and at a relatively young age compared to lung cancer in smokers which is traditionally found in elderly men. Among East Asian women, 61–80% of the lung cancer diagnoses were in never smokers (26-28). This high frequency is thought to reflect exposures to secondhand tobacco smoke, indoor air pollution from burning of coal for heating and cooking fumes, and outdoor pollution from ambient fine particles (29,30). The most common cell-type in never smokers was adenocarcinoma. There are also differences in genetics and molecular findings as these cancers are often associated with EGFR mutations and ALK
rarrangements while smoking related lung cancers are associated with KRAS mutations (31-33); these differences can be used to guide specific targeted therapies. In light of the decreasing rates of cigarette smoking, the proportion of lung cancers diagnosed in never smokers is increasing (34).

Expansion of eligibility criteria for individuals who never smoked is becoming an important issue. As eligibility criteria expand, attention to the regimen of screening, and continuous updating of an evidence-based nodule management protocol, is important to limit unnecessary workup and over-treatment. Results from Eastern Asian countries showed that LDCT screening can detect a significant number of lung cancers in never smokers, most of them in early stage; approximately 70% of these lung cancers would have been missed using the NLST selection criteria for screening (35-37). These studies recommended that individuals who have never smoked should be included in eligibility criteria of LDCT screening. Results in the United States also suggest that LDCT screening is beneficial for never smokers exposed to secondhand tobacco smoke for identification of early lung cancer, cardiovascular disease and emphysema (38). Future studies, with focus on young never smoking women, will be needed to address this issue.

Several validated lung cancer risk prediction models are available for ever-smokers. Earlier risk prediction models for lung cancer that considered never smokers, however, were limited as they were either developed in case-control studies which may produce biased samples (39-42) or the predictive accuracy was poor with area under the receiver operating characteristic curves (AUCs) in the region of 0.5 (39,43). In the early 2010s, a few risk prediction models were developed to identify high risk individuals in the general population (regardless of smoking history). However, these models only captured the never smoking status by inverting the smoking duration or intensity to zero, or by including an indicator for ‘never-smokers’ or ‘never-smokers with secondhand tobacco smoke exposure’ in their model (44-47). Furthermore, they did not consider many of the distinct risk factors which are unique to never smokers and thus failed to capture the entire spectrum of exposure among never smokers. Despite the overall good calibration and discrimination demonstrated by the PLCO-M2014 model, when model performance was evaluated separately by never- and ever-smokers, results indicated that the PLCO-M2014 model had much higher discriminatory power in ever-smokers than in never smokers (AUC of 0.85–0.86 for ever-smokers vs. AUC of 0.62 for never-smokers) (44). Concerns about misclassification and inaccuracy in the classification of smoking and exposure status have also been raised (48,49), although others have shown reasonable correlations of self-reported smoking status and biologic confirmation (50).

Since lung cancer is traditionally associated with smoking, the risk in never-smokers remains under-recognized. In recent years there are, however, increasing efforts to understand the etiology of lung cancer in never smokers (21,51-56). Most studies have reported links to family history of cancer, genetic susceptibility, occupational exposure, environmental exposure including secondhand tobacco exposure, air-pollution and cooking oil fumes. Some evidence also suggested the effect of diet and lifestyle on risk of lung cancer while the role of hormones remained unclear. Unfortunately, most of these risk factors have been recently identified and were not collected in historical datasets. These analyses require complex assessment of exposure or additional cost and thus have limited progress in the development and validation of risk models for never smokers. With the advancement in technology, big data analytics and increasing awareness of the risk of lung cancer in never smokers, more researchers are now working to develop personalized risk prediction tools to more accurately stratify lung cancer risk across smoking status by integrating clinical, laboratory and biomarkers/genetic information (57-60). Through better risk assessment, the effectiveness of screening can be increased by identifying those never smokers who have as high a risk of developing lung cancer as some heavy smokers.

In the era of precision medicine, there is growing recognition that use of these risk factor-based criteria which dichotomize age and pack-years of smoking to select eligible individuals for screening may not be the most efficient and may also lead to increase lung cancer disparity (18,61). Furthermore, multiple retrospective analyses and modeling studies have demonstrated that risk prediction models that incorporate additional risk factors help improve risk stratifications. These models demonstrated higher sensitivity and positive predictive value, and increased cost-effectiveness than older, less complex criteria for determining screening eligibility (44,62-64). A few studies are currently underway to prospectively evaluate the use of risk prediction models and to determine the optimal approach for selecting high risk individuals for lung cancer screening (65-68). Despite the superiority over risk factor-based criteria in preventing lung cancer deaths, concerns have been raised that risk prediction models may be suboptimal and increase overdiagnosis. As these lung cancer
risk prediction models do not account for life-expectancy, they are more likely to select older smokers with longer smoking history and more comorbid conditions. These individuals are more likely to die from competing causes and may not live long enough to experience the net benefit of screening. To address these issues, alternative models that select individuals based on life-years gained have also been proposed (69-71).

Integration of AI approaches for LDCT interpretation

AI techniques in the field of medicine are gaining more interest recently, but it should be recognized that initial efforts had started as early as in 1970s (12). Due to enhanced algorithms and computing power, AI has made great strides in many areas—classification of photos, person recognition, self-driving vehicles, natural language processing, and data mining, just to name a few and has generated great enthusiasm for streamlining cancer screening to improve early detection and diagnosis of cancer and personalize treatment and outcome prediction (72-78). The future of AI in lung cancer screening lies in the integration of algorithms that detect and diagnose all diseases visible in a LDCT, not only lung cancer but emphysema, interstitial lung disease, cardiovascular disease, and liver disease to automatically produce a report of everything that is visible on a LDCT chest scan.

AI algorithms for lung cancer screening have focused on the detection and characterization of pulmonary nodules. Early studies used manually engineered features, where an algorithm is designed to compute a specific feature, such as the distribution of density within the nodule, and these studies either targeted the detection of nodules (79,80) or the classification of nodules as benign or malignant (81,82). However, performance of these systems is still not sufficient for completely automated use. Nodule detection results are commonly reported as the sensitivity, the rate of detecting actual nodules, for a given number of false positives, detection of objects that are not nodules. Recent comparative studies of research nodule detection systems report results in the range of 75–85% sensitivity with an average of a single false positive per case (83,84). Likewise, the performance of nodule classification systems still leaves much to be desired, with reported AUC performance of 0.75. Although these results are not sufficiently high for a completely automated system, software systems are now available from many commercial vendors that include tools for nodule detection, measurement, and classification that are intended to help support radiologists’ interpretation.

As a result of recent successes of neural networks and deep learning in the computer vision area, in particular the breakthrough performance deep learning on the ImageNet challenge (85), a competition to classify images into 1,000 different categories, deep learning has been applied to all areas of image interpretation (86). A team of researchers at Google applied deep learning to detecting diabetic retinopathy from retinal fundus photographs with great success—they achieved a performance of over 90% sensitivity and specificity (87). Another recent study developed a deep learning algorithm to predict the risk of lung cancer (88). The authors report an AUC of 0.944 for predicting lung cancer, performing as well as or better than radiologists. However, the system was developed and tested on data from the NLST, which was collected over 15 years ago; CT technology has made great strides during this time, and the high performance on the NLST dataset may not translate to modern images. In addition to lung cancer (89,90), deep learning has also been applied to quantification of aortic calcification (91), emphysema (92), and breast cancer (93,94).

As AI continues to improve, there are two technical obstacles to overcome. Deep learning, more so than previous AI approaches, requires large amounts of data; however, there is a lack of large, public, well-documented databases of medical images. A key reason for the success of computer vision in the ImageNet challenge is its database which allows research groups all over the world to train and test their algorithm on the same database as others in the field, allowing for direct comparisons between algorithms. Unfortunately, while there have been efforts to create such databases for lung screening, such as the public lung database (PLD) (95), NLST (96), ANODE09 (97), LUNA16 (98), and The Cancer Imaging Archive (TCIA) (99) to name a few, these all have few cases—from 50 to a couple thousand. As a point of comparison, the ImageNet database contains over 1 million images. There are many reasons for the lack of such large public databases—patient privacy, funding, and time and difficulty in properly annotating data to name a few.

A second major technical obstacle facing the continued improvement of AI is the lack of consistency between different scanners and protocols in the images that are acquired. This issue has been recognized (100), with studies showing that radiomic features have low reproducibility (101,102), and that CT reconstruction parameters can
affect the performance of radiomics systems (103). There is ongoing work in developing tools and methods for characterizing CT scanner performance (104), so that in the future, radiomic features and quantitative measurements can be consistent across scanners and acquisition protocols.

In conclusion, the field is rapidly progressing towards the goal of automatically producing a report of everything that is visible in the lung, not only related to lung cancer, but also other diseases. The individual algorithms exist today, but much work remains to improve their performance. This will require large, documented databases and high-quality imaging.

Continued optimization of the regimen of screening

Since 1992, when the initial regimen of screening was developed, it was recognized that continual updating was needed to integrate technologic and knowledge advances (12). Growth assessment was an important component to differentiate between lung cancer and benign nodules. This led to recognition of the measurement errors of CT scanners. The need for large data and image repositories was also recognized, particularly in light of the new statistical approaches using AI techniques. Risk models are being developed to provide follow-up recommendations. Lessons learned in the past years since the introduction of LDCT screening studies starting in 1992 (12) are useful in illustrating the importance of accumulated knowledge and data to guide approaches in the future.

When assessing the benefit of screening, it is important to understand implications of the trial designs in assessing the benefit of screening (105-107). This is particularly important when providing information about the benefits and risks of screening to a participant seeking the screening (108,109). The high cure rates of small, early, screening-diagnosed lung cancers are frequently not presented. It needs to be recognized that the lung-cancer mortality reductions of randomized screening trials do not present the benefit of an early diagnosis which is the cure rate of screening-diagnosed lung cancers (108). It needs to be understood that the NLST and NELSON trials were designed to show a mortality reduction of 20% and 25%, respectively to justify screening. These two trials were not designed to provide the cure rate of screen-diagnosed lung cancers. Future development of better tools to explain the benefit of LDCT, or any other screening test for any cancer, should be developed. Increasing patients’ and providers’ awareness of the benefit of screening is key to improving both uptake and adherence to lung cancer screening.

Approaches for comparisons of different regimens of screening are needed to determine the optimal work-up for nodules detected on baseline and annual repeat rounds of screening, by nodule consistency. One such approach was developed to compare the International Early Lung Cancer Action Protocol with those of the American College of Radiology LungRADS and the European Protocol (110). Continuous update of protocols to incorporate advancing technology and knowledge will minimize unnecessary diagnostic work-ups and biopsies/surgeries. Research is being performed to identify new predictors of benign and malignant/aggressive etiology, especially with the help of AI. These include perifissural and costal-pleural nodules (111).

The regimen should strive to maximize the likelihood of early diagnosis of lung cancer while minimizing unnecessary invasive workup. The importance of having a well-defined regimen was demonstrated by the comparison of I-ELCAP results with those of the NLST which did not specify a regimen. I-ELCAP’s higher percentage of Stage I diagnoses and long-term survival rates compared with those of the NLST, after consideration of multiple alternative explanations, was due to the I-ELCAP regimen (112).

Aside from a well-defined regimen, appropriate radiologic interpretation can minimize unnecessary workup and interventional procedures. It has also been shown that high quality LDCT screening can be performed in academic or community settings as long as a quality assurance process is in place (113). Integration of new image analytic tools being developed should further improve the diagnostic interpretation and lead to a reduction of unnecessary further testing and reduce the frequency of surgical resection of benign nodules. It has been shown that, by carefully following a well-defined regimen of screening, the frequency of benign resection can be below 5% (114).

Volumetric measurement is acknowledged to be a better assessment of nodule size. Volume doubling time is a more reliable measure of growth for distinguish between malignant and non-malignant nodules. However, volumetric measurements need to be interpreted cautiously with regard to CT acquisition parameters and CT measurement errors (115). The recent update of Lung-RADS version 1.1 added volume measurements next to the nodule diameter measurements. Use of volume measurement to provide more accurate growth assessment over time is expected to increase as well as further refinement of growth assessment,
including use of phantoms to adjust for measurement variability of the CT will improve measurement accuracy. The ability to reliably assess tumor growth and advancement in prediction tools can help inform follow-up interval/screening interval (116).

Development of blood-based tests will continue to develop and, when proven to be useful, should be integrated into the screening process. New pathologic criteria will continue to emerge together with advances in lung cancer biomarkers as has already been demonstrated by the revised World Health Organization classifications (117-120). In the future, more precise biologic information as to the aggressiveness of the screening-diagnosed lung cancers will be achieved by improved imaging of growth in a timely manner and pathologic biomarkers.

**Understanding the benefit of identification of other early diseases identified**

It has been recognized that LDCT screening provides a comprehensive “health check” of the lungs, heart, and other organs visualized on the LDCT. This vision is gaining increasing recognition throughout the world. An entire session at the 20th World Conference on Lung Cancer (WCLC) in Barcelona, Spain in September 2019 was devoted to these other findings (121). I-ELCAP protocol recommendations for these other findings were developed together with relevant medical specialties, with the recognition that these are findings in asymptomatic screening participants and not patients presenting to physicians with symptoms. Initial focus was on the cardiac findings (122,123) and emphysema (124-126) which, together with lung cancer, are the three big killers of older smokers. This focus has now been expanded to consider the other findings on the LDCT of the chest and the recommendations are provided in the I-ELCAP protocol (127).

Such a comprehensive “health check” optimal LDCT screening requires a carefully-specified, validated regimen which provides for identification and interpretation of critical LDCT findings and the appropriate follow-up recommendations. The recommendations need to be developed together with the relevant medical specialties and may differ in different medical care settings.

**Evaluation of new screening tests will emerge**

As new screening tests emerge, both imaging and blood-based ones, it should be recognized that these new tests can be rapidly evaluated using designs similar to the low-cost, efficient prospective ELCAP cohort design. Two rounds of screening, a baseline round and a single annual repeat round can provide pertinent information on the tumor size at detection, the stage shift (1,2), and after appropriate follow-up, the cure rates (3). This same design can be used to test new blood tests for earlier detection of lung cancer. Both the blood test and a LDCT can be given to individuals at risk of lung cancer and tested within two years.

Hopefully there will be recognition that randomized start-stop trials to test methods of screening for a cancer do not provide the ultimate benefit of the screening test, but rather only provide that a minimum level of benefit is met before providing the screening on a larger population-wide basis. For example the NLST required 3 rounds of screening with 5.5 years of follow-up to show at least a 20% lung cancer-specific mortality reduction by LDCT screening compared to chest radiography; its results took nine years from start to publication (6). The NELSON Trial, started in 2004 to provide 4 rounds of screening with 10 years of follow-up (11), reported that it provided at least a 25% lung cancer-specific mortality reduction compared to no screening.

**New paradigms of treatment for early lung cancers**

As already seen in the past 20 years since screening was introduced, understanding of the pathologic findings of small, early lung cancers has increased. Identification of small, early cancers has stimulated advances in treatment of early lung cancer which in turn has led to updates in the pathologic and staging criteria. The current recommendation of lobar resection of Stage I lung cancers has not changed in more than 50 years.

The authors postulate that, similar to the impact of breast cancer screening which transformed treatment of breast cancer from radical mastectomy to a very nuanced, personalized approach, lung cancer treatment will become more personalized with an emphasis on post-treatment quality of life and preservation of lung tissue as new primary lung cancers may emerge.

LDCT screening has already led to significant changes in the pathologic classification and the 8th Staging Classification (128-130). Screening results stimulated two randomized surgical trials comparing lobectomy with sublobar resection which started in 2007, one in Japan and the other in the United States (131,132). Both anticipate
publishing their final results around 2020 and their interim reports are encouraging as both have shown extremely low rates of surgical deaths (133,134).

New technologies have been introduced such as robotic surgery, navigational bronchoscopy, ablation approaches, and in the future, there will be further innovations. None of these, however, have had critical assessment, and often there is only limited data for small lung cancers. Published cohort studies using the I-ELCAP database (114,135-138) have already provided timely outcome results and Quality of Life measures which will become an increasingly important consideration in treatment determination given the high long-term cure rates of screen-diagnosed lung cancer (139-143).

For timely assessment of new treatments, the Initiative for Early Lung Cancer Research on Treatment (IELCART) (144) was started using the same prospective cohort design used for I-ELCAP and capturing treatment information in the context of clinical care. The ELCAP Management System, used for both management and research purposes allowed for the accumulation of over 82,000 participants with clinical data, imaging and biologic specimens, has been adapted for the multi-institutional, international IELCART database. The system also allows for randomization for future innovative randomized trials. The vision for IELCART is to become as productive as I-ELCAP has been in producing ongoing screening evidence.

Conclusions

Continued advances in CT technology, including reduction in radiation dose as well as new image analytics and statistical techniques being developed will continue to improve LDCT screening, perhaps together with new pre- and post-CT tests, such as those currently under development. Optimizing the screening will reduce the frequency of unnecessary workup and invasive diagnostics, increase the frequency of earlier diagnoses, and provide for an even more comprehensive “health check.” In light of the continued advances in screening for lung cancer and the integration of the advances in screening protocols, a key focus of ongoing and future research should be on the treatment of these cancers.

Acknowledgments

Funding: None.
conflicts of interest to declare.

**Ethical Statement:** the authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

**References**

22. Sun S, Schiller JH, Gazdar AF. Lung cancer in


doi: 10.21037/ccts-20-125