# After Detection: The Improved Accuracy of Lung Cancer Assessment Using Radiologic Computer-aided Diagnosis

Guy J. Amir, MD, MPH, Harold P. Lehmann, MD, PhD

**Rationale and Objectives:** The aim of this study was to evaluate the improved accuracy of radiologic assessment of lung cancer afforded by computer-aided diagnosis (CADx).

**Materials and Methods:** Inclusion/exclusion criteria were formulated, and a systematic inquiry of research databases was conducted. Following title and abstract review, an in-depth review of 149 surviving articles was performed with accepted articles undergoing a Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-based quality review and data abstraction.

**Results:** A total of 14 articles, representing 1868 scans, passed the review. Increases in the receiver operating characteristic (ROC) area under the curve of .8 or higher were seen in all nine studies that reported it, except for one that employed subspecialized radiologists.

**Conclusions:** This systematic review demonstrated improved accuracy of lung cancer assessment using CADx over manual review, in eight high-quality observer-performance studies. The improved accuracy afforded by radiologic lung-CADx suggests the need to explore its use in screening and regular clinical workflow.

**Key Words:** Computer-aided; imaging; medical; lung; cancer.



## INTRODUCTION

Lung cancer is the second leading cause of death in the United States and among the top 10 worldwide. More Americans die each year from lung cancer than from breast, prostate, and colorectal cancers combined. Annually, lung cancer kills more men than prostate cancer and more women than breast cancer (1).

Whereas overall cancer incidence rates are declining, lung cancer incidence rates among women are rising. Between 1960 and 1990, deaths from lung cancer among women increased over 400%. It is the second most common cancer among African American men and kills more African Americans than any other cancer. Five-year survival ranges from 70% for stage I disease to less than 5% for stage IV disease. As of 2014, overall 5-year survival is 17%, with only 15% diagnosed at the localized stage (2).

In this paper, "CADe" is defined as computer-aided detection and "CADx" as computer-aided diagnosis, and unless otherwise specified, will refer to both radiographic and computed tomography (CT) scans of the lungs. These are software outputs that a radiologist supervises/evaluates while viewing the image himself before final assessment. A CADe system detects abnormal nodules, without discriminating malignant from benign. CADx involves further interpretation as to likelihood of cancer. Unsupervised CADx (UCAD) is a standalone reader, a potential technological evolution (cf. Fig 1).

If diagnosed at the early stage (1A, < 3 cm), curative resection of stage I non-small cell lung cancer affords the survival rate of 70–80% (3). Screening current/former heavy smokers aged 55–74 years with low-dose CT by The National Cancer Institute's National Lung Screening Trial (53,454 participants) demonstrated that those who received low-dose CT scans had a 20% lower risk of dying from lung cancer than participants who received chest X-rays (4). Chest radiography (CXR) screening remains controversial (e.g. see Hoop et al. (5)). Because all the above studies excluded CADx, improved outcome is possible. Because of the radiologist workload (6) and false-positive rate after biopsy of up to 50% (7), CADx continues to be investigated.

To our knowledge, there has never been a systematic review focused on CADx (CXR + CT) diagnostic accuracy. The novelty in our paper is in the exhaustive search we performed.

**Acad Radiol 2016;** 23:186–191

From the Division of Health Sciences Informatics, Johns Hopkins University, 2024 East Monument Street, Suite 1-200, Baltimore, MD 21205, USA. Received June 5, 2015; revised October 11, 2015; accepted October 13, 2015. **Address correspondence to:** H.P.L. e-mail: lehmann@jhmi.edu


http://dx.doi.org/10.1016/j.acra.2015.10.014

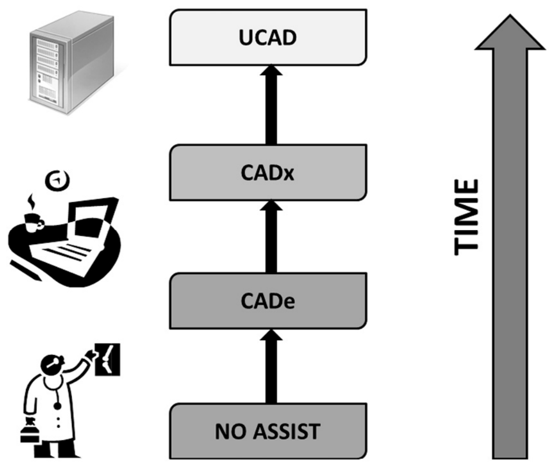


**Figure 1.** Past and possible future evolution of computers in diagnosis. CADe, computer-aided detection; CADx, computer-aided diagnosis; Time, chronological time, assuming continuous technical advancements; UCAD, unsupervised CADx;

The goal of CADe is to differentiate true nodules from normal lung structures with the key outcome being sensitivity and false-positive rate that are reported per scan (patient). The comparative gold standard for CADe is a consensus panel. CADx studies are about correctly classifying/assessing detected nodules as high risk of malignancy (actionable) or low risk of malignancy (no immediate workup indicated, i.e. nonactionable). The comparative gold standard reference for CADx used by the studies is biopsy.

There are different diagnostic CADe/x models described in the literature. Both CADe and CADx employ at a minimum, a two-phase approach (cf. Fig 2). The first phase is segmentation with subtraction/difference image processing using gray-level picture thresholding techniques. The focus of the second phase is feature extraction to reduce false positives and involves predictive modeling techniques such as support vector machines (employing Gaussian, linear, or polynomial kernels), artificial neural networks (ANNs), cluster analysis, Bayesian Wavelet Snake, or other techniques.

CADx involves those systems that use the data acquired from the second phase beyond detection in a third phase to classify nodules as actionable or nonactionable. Rule-based methods, ANN, discriminant analysis, and other classifier techniques are employed in this third phase. CADx represents the entire scheme as depicted on the left in Figure 2.

The goals of this systematic review were to ascertain whether and by how much CADx improves the accuracy of lung cancer assessment over that of radiologists working without the technology. The modalities included are radiography, low-dose computed tomography (LDCT), and high-resolution computed tomography (HRCT).

## MATERIALS AND METHODS

The design of this systematic review was informed by Cochrane guidelines (8). We aimed to include all peer-reviewed journal articles that contained original data on the diagnostic performance of CADx. Databases searched included PUBMED, BioMed Central, the Cochrane Library, CINAHL/CINAHL PLUS, EMBASE, IEEE Xplore, INSPEC, JHSearch, and Web of Science.

Inclusion criteria were comparative evaluations of CADx for lung cancer and the use of CXR or CT. To increase sensitivity of the search, we included "detection" in the search algorithm. Search strategies for each reference database followed the following pattern:

> Computer-aided (Detection OR Diagnosis)
> (Lung Neoplasm) OR (Lung Nodule)
> (Radiography OR CT)

Whenever possible, controlled vocabularies were utilized. When not available, synonyms were tested to increase sensitivity (see online supplement for all search strategies). For example, using MeSH for PubMed, the inclusion criteria became:

> Computer-aided (Detection OR Diagnosis) AND ("Lung Neoplasms"[MeSH] OR Lung Nodule) AND ("Radiography"[MeSH] OR "Tomography, X-Ray Computed"[MeSH]) NOT review[PT]

Additionally, a manual search of the articles' references was conducted to extract further eligible articles. Exclusion criteria were applied after reading titles and abstracts (if needed), and again after surviving articles were read, were any of:

1. Nonoriginal data
2. In title, nothing related to the thoracic region nor computer-aided systems
3. Lack of quantitative data related to CADx, e.g. solely CADe
4. Absence of any quantitative data related to accuracy
5. Absence of reference to either CT or CXR modalities
6. Duplicates

Abstraction forms included a quality scoring form based on the validated Quality Assessment of Diagnostic Accuracy Studies (QUADAS) scale for rating studies of diagnostic accuracy (9) and a content form with design and result tables for statistical analysis. Design data and Accuracy data tables were assembled. Specific factors abstracted from each paper were design data, verification tests, algorithms employed, and statistical data to evaluate improved accuracy. Different accuracy measures were sought, including sensitivity per scan at cutoff, false-positive reading/scan (test positive outputs for normal lung structure or benign lesions), accuracy (TP+TN)/(P + N), where TP = true-positive, TN = true-negative, P = total-positive, N = total-negative, and receiver operating characteristic (ROC)-area index $A_z$ (area under the ROC equal to the probability that a system will rate a randomly chosen positive instance higher than a randomly chosen negative one). A pilot study of 10 articles was used to improve the quality and content article abstraction forms and to ensure inclusion of all relevant variables.