



Published in final edited form as:

*Acad Radiol.* 2022 February ; 29(Suppl 2): S18–S22. doi:10.1016/j.acra.2020.07.040.

## Factors Influencing the False Positive Rate in CT Lung Cancer Screening

Mark M Hammer<sup>1,2</sup>,

Suzanne C Byrne<sup>1,2</sup>,

Chung Yin Kong<sup>2</sup>

<sup>1</sup>Department of Radiology, Brigham and Women's Hospital, Boston, MA

<sup>2</sup>Harvard Medical School, Boston, MA

### Abstract

**Purpose.**—To identify factors influencing the likelihood of a false positive lung cancer screening (LCS) CT, which may lead to increased costs and patient anxiety.

**Materials and Methods.**—In this retrospective study, we examined all LCS CTs performed across our healthcare network from 2014 – 2018, recording Lung-RADS category and diagnosis of lung cancer. A false positive was defined by Lung-RADS 3-4X and no diagnosis of lung cancer within 1 year. Patient demographics and smoking history, presence of emphysema, diagnosis of COPD, radiologist years of experience and annual volume, and screening institution were evaluated in a multivariate logistic regression model for false positive exams.

**Results.**—A total of 5835 LCS CTs were included from 3735 patients. Lung cancer was diagnosed in 142 cases (2%). Of the LCS CTs, 905 (16%) were positive by Lung-RADS, and 766 (13%) represented false positives. Logistic regression analysis showed that screening institution (odds ratios [OR] 0.91 – 2.43), baseline scan (OR 1.43), radiologist experience (OR 0.59), patient age (OR 2.08), diagnosis of COPD (OR 1.34), presence of emphysema (OR 1.32), and income level (OR 0.43) were significant predictors of false positives.

**Conclusions.**—A number of patient-specific and site/radiologist-specific factors influence the false positive rate in CT LCS. In particular, radiologists with less experience had a higher false positive rate. Screening programs may wish to develop quality assurance programs to compare the false positive rates of their radiologists to national benchmarks.

### Keywords

lung cancer screening; Lung-RADS; false positive; radiologist experience

### Introduction

Computed tomography (CT)-based lung cancer screening (LCS) has been shown to reduce overall mortality, most notably demonstrated by the pivotal National Lung Screening Trial

(NLST)<sup>1</sup>. Data from this trial showed CT LCS to be cost effective, with estimates ranging from \$49,000 – \$81,000 per quality-adjusted life year (QALY)<sup>2,3</sup>. However, even this cost may be a barrier to adoption of CT LCS in certain healthcare systems. There are many determinants of the costs of CT LCS, but one important contributor is false positive scans leading to unnecessary interventions. Even more important than the financial costs that false positives produce are negative experiences for patients, who will undergo unnecessary follow-up scans or even invasive procedures. False positives likely contribute to patient anxiety within LCS programs<sup>4</sup> as well as increased exposure to ionizing radiation.

In the NLST, the specificity (i.e. 1 – false positive rate) for the CT screening arm was 73%<sup>5</sup>. In order to reduce the false positive rate of CT LCS, the American College of Radiology developed the Lung-RADS schema, which among other goals, enforces a higher cut-off for a positive scan<sup>6</sup>. A secondary analysis of the NLST data using Lung-RADS showed a higher specificity of 87% (at the cost of reduced sensitivity from 98% to 85%)<sup>7</sup>. Data from other CT LCS programs have shown higher specificity of up to 91%<sup>8</sup>, and the Dutch NELSON LCS trial showed a specificity of 98% using a different nodule follow-up scheme<sup>9</sup>. Given disparate results for the specificity of CT LCS, we undertook an analysis of factors that may influence the false positive rate in LCS.

## Methods

### Patient selection.

This retrospective study was approved by the Institutional Review Board and conducted in compliance with HIPAA guidelines. Informed consent was waived by the Institutional Review Board for this retrospective study. The electronic medical records of a large healthcare network were searched for chest CTs billed as lung cancer screening, from January 2014 through August 31, 2018. LCS CTs were performed at 4 institutions, including two large quaternary academic medical centers (denoted Academic 1 and 2) and two community hospitals with non-academic radiology practices (denoted Community). The interpreting radiologists were distinct among the four sites.

### Data extraction.

Patient and radiology data, including radiology reports, from these patients were downloaded from the electronic medical record. Reports were searched for the presence of a Lung-RADS score, and the score was extracted from the reports using regular expressions in a Perl script (Strawberry Perl, v5.28). This yielded a total of 5,986 studies. Chest CTs with Lung-RADS scores but not billed as lung cancer screening studies were excluded (n=151), leaving 5,835. Patient and radiology data from these studies were then entered into Microsoft Excel (v1806, Microsoft Corp, Redmond, WA).

Smoking history entries and billing diagnoses (specifically, lung cancer and chronic obstructive pulmonary disease [COPD]) were extracted from the electronic health records of these patients. LCS CT reports were also searched for the word “emphysema” as well as for the presence of a prior comparison study using a regular expression. The interpreting radiologist was recorded for each LCS CT, and the number of LCS CTs read by each

radiologist was divided by the number of years they read LCS CTs in the database to yield an annual volume. The year of graduation from residency or last fellowship was recorded for each radiologist; radiologist experience was then calculated as the difference between the year of the LCS CT and the graduation year. Patient Zip codes were extracted from the electronic health record, and median incomes of households in those zip codes were retrieved from Income By Zip Code (<http://incomebyzipcode.com>).

### Data analysis.

The resulting data table was imported into JMP Pro (v14.3, SAS Institute, Cary, NC). Lung cancer diagnoses were extracted from billing code records, with the primary endpoint defined as a diagnosis of lung cancer within 1 year of the LCS CT. Positive LCS CTs were defined by Lung-RADS categories 3, 4A, 4B, and 4X, as per the definition in the Lung-RADS document<sup>6</sup>. Thus, false positive scans were positive LCS CTs where lung cancer was not diagnosed within 1 year. A penalized binary logistic regression analysis was performed with the Lasso technique to identify variables associated with false positive scans. The Lasso technique allowed for elimination of correlated variables (whose odds ratios are set to 1). In the model, radiologist experience and patient age were treated as continuous variables.

Differences in categorical variables were analyzed with the chi-squared test except for 2x2 tables which were analyzed with Fisher's exact test (one-tailed). Analysis of per-radiologist false positive rates was performed by weighting each radiologist by the number of studies interpreted.

## Results

### Patient demographics and CT results.

A total of 5,835 LCS CTs were performed within the time period analyzed, representing 3,735 unique patients. Patient demographics are given in Table 1. Just over half of patients (2021 or 54%) were male, with median age of 65. The vast majority (3135 or 85%) were non-Hispanic white. Approximately equal numbers were current and former smokers (1579 or 48% versus 1627 or 50%).

LCS CT data are given in Table 2. The vast majority of LCS CTs were performed at the academic medical centers in our healthcare network (3882 or 67% at Site 1 and 1719 or 29% at site 2, totaling 96%). Thirty percent (1690) of scans represented a baseline CT (i.e. no prior CT for comparison). Fifty percent (2944) of scans reported the presence of emphysema.

Lung-RADS category 1 was reported in 1505 scans (26%), and category 2 was reported in 3417 scans (59%). Positive scans made up 905 or 16%, including 517 category 3 (9%), 233 category 4A (4%), 86 category 4B (1%), and 69 category 4X (1%). Lung cancers were diagnosed within 1 year of the LCS CT in 142 patients (2%), yielding a sensitivity of 97.9% and specificity of 86.5% (false-positive rate of 13.5%). The distribution of lung cancer diagnoses by Lung-RADS category are given in Table 3. The positive predictive value was 15.4%, and the negative predictive value was 99.9%. If a cancer diagnosis within 2 years

was considered positive, then the false positive rate was 729/5607 (13%). If a positive screen was only counted as Lung-RADS 4A-4X, then the false positive rate for cancers diagnosed at 1 year would drop to 263/5693 (4.6%).

There were a total of 75 individual radiologists who interpreted the LCS CTs, with the top 18 radiologists accounting for 80% of the LCS CT volume. The annual volume ranged from 1 to 197 CTs per year; within the group of LCS CTs, the median radiologist volume was 55. Radiologist experience ranged from 0 years (interpreting during fellowship) to 45 years after training; within the group of LCS CTs, the median experience was 16 years. Radiologists with less experience, defined as  $\leq 5$  years, read 27% of cases at Academic 1, 27% at Academic 2, and 25% at Community sites ( $p = \text{NS}$ ). A total of 90 exams (1.5%) were read by an attending with a trainee.

### Factors influencing false positive rate.

There were a total of 766 false positive LCS CTs (13%). The results of a multivariable logistic regression analysis for factors predisposing to false positive scans are shown in Table 4. Variables tested included site, baseline versus follow-up scan, presence of emphysema, radiologist experience and annual volume, patient gender and age, median income by zip code, smoking status, and COPD diagnosis. Baseline scans had a higher rate of false positives (OR 1.43,  $p=0.0002$ ), and patients with COPD and emphysema also had a higher rate of false positives (OR 1.34 and 1.32,  $p=0.001$  for both). Certain sites were associated with higher rates of false positives than others (e.g. Site 1 versus Site 2 OR 2.43,  $p<0.0001$ ). More experienced radiologists had lower rates of false positives (most experienced versus least experienced radiologist OR 0.59,  $p=0.001$ ). Older patients had a higher rate of false positive scans (oldest versus youngest patient OR 2.08,  $p=0.01$ ). Finally, patients living in higher income areas had lower rates of false positive scans (highest income versus lowest income OR 0.43,  $p=0.01$ ). The other variables were not statistically significant.

### Lung-RADS scores by radiologist experience.

LCS CTs were split into two groups by radiologist experience, those read by radiologists with less experience ( $\leq 5$  years) and those read by radiologists with higher experience ( $> 5$  years). The Lung-RADS score distribution and false positive rates are shown in Table 5. Scans read by less experienced radiologists had a 17% positive rate (15% false positive rate) compared to 15% positive rate (13% false positive rate) for more experienced radiologists ( $p < 0.01$  for difference in false positive rate). The distributions of Lung-RADS scores differed between the groups ( $p<0.0001$ ). Regarding the negative Lung-RADS categories, less experienced radiologists had a higher rate of Lung-RADS 2 than more experienced radiologists (65% versus 56%). Less experienced radiologists also had higher rates of positive scans, as noted above, driven by increases in Lung-RADS 4A (5% versus 4%) and 4B (2% versus 1%).

We performed an analysis of per-radiologist false positive rates. Of the 48 radiologists with high experience, the median false positive rate is 13.6% (interquartile range 6% - 18%), with 3 outliers (44% in a radiologist who read 9 cases; 50% in a radiologist who read 2 cases;

and 100% in a radiologist who read 2 cases). Of the 30 radiologists with low experience, the median false positive rate is 15.4% (interquartile range 7% - 20%), with a single outlier (100% in a radiologist who read 1 case). Of note, for 5 early career radiologists, some of the LCS CTs were interpreted when they had low experience, and others when they had high experience, and thus they appear in both groups. For 2 radiologists, year of training completion was not available.

## Discussion

In an analysis of more than 5,800 lung cancer screening CTs across a large healthcare network, we have shown that a number of factors influence false positive rates, including site, baseline scan, emphysema and COPD, radiologist experience, and patient age. In particular, patients undergoing a baseline scan, older patients, and those with COPD or emphysema experience higher rates of false positive scans. Additionally, scans read by less experienced radiologists have a higher rate of false positives.

The fact that baseline scans are associated with a higher rate of false positives is expected, as these patients have no comparison to establish stability of lesions. (Any abnormality that is stable at follow-up in Lung-RADS is downgraded to category 2.) This phenomenon has been shown previously in the NLST and NELSON trials<sup>7,9</sup>. The associations between emphysema or COPD and false positive scans is, perhaps, less intuitive. Previous work has demonstrated an association between COPD and risk of lung cancer within the NLST population<sup>10</sup>, an expected finding given that both are associated with smoking. Several other studies have also shown that COPD and emphysema are associated with an increase in likelihood of a patient having nodules (and therefore a positive screen)<sup>11,12</sup>. The reasons for this are not exactly clear, although our anecdotal clinical experience has been that patients with severe emphysema tend to develop more inflammatory lesions that may simulate spiculated lung cancers. Similarly, the reasons for the association between age or income level and false positives are not entirely clear. Older patients and those with lower income may have worse health status and be overall at increased risk of infectious processes that simulate lung cancer, as with emphysema.

The differences in false positive rates across the sites within our healthcare network were quite striking (ORs ranging from 0.91 to 2.43). The underlying causes of these differences are unclear. It is possible that there are differences in the patient populations across these sites that were not captured within our multivariable model. More likely, these differences reflect different institutional practices regarding the use of the Lung-RADS rubric. Radiologists at some institutions may be more likely to dismiss some lesions, e.g. intrapulmonary lymph nodes or clustered nodules, as probably benign (Lung-RADS 2), while other radiologists may simply use the strict size cut-offs to assign Lung-RADS categories. This phenomenon will likely be mitigated by improvements in Lung-RADS with the new version 1.1, which specifies that intrapulmonary lymph nodes be assigned category 2, but it does not specifically address probable inflammatory lesions in the low risk categories<sup>6</sup>.

We found substantial differences in false positive rates by radiologist experience, with the most experienced radiologist nearly half as likely to report false positive scans as the least experienced radiologist (OR 0.59). The increase in false positives by less experienced radiologists was driven by increases in Lung-RADS 4A and Lung-RADS 4B. Again, this may be related to differences in interpretation of inflammatory lesions – more experienced radiologists may feel more confident in assigning a benign category to such lesions. A similar phenomenon of effect of radiologist experience on false positive rates has been described with breast cancer screening<sup>13,14</sup>. Of note, while one of those studies found independent risks for radiologist volume and experience, we did not find that radiologist volume was predictive if years of experience was included in the model.

The fact that site and radiologist experience have substantial effects on the false positive rate of LCS CT presents an opportunity for lung cancer screening programs. These programs may institute quality assurance and audit programs to evaluate their positive and false positive screening rates. Indeed, this is recommended by the draft American College of Radiology – Society of Thoracic Radiology Practice Parameter for lung cancer screening<sup>15</sup>. By comparing the performance of individual radiologists and the entire program to national metrics, radiologists may be able to adjust their thresholds for positive studies and come more in line with national averages. Additionally, early career radiologists may benefit from lung cancer screening CT training cases, a practice also used in mammography training. These interventions have the potential to reduce false positives and improve the overall cost effectiveness of lung cancer screening across the country.

This study has several limitations. Being a retrospective, multi-institutional study, we are limited by varied institutional practices, particularly for managing positive screening scans. However, that does reflect clinical practice in real world scenarios. Additionally, diagnoses of lung cancer were made by billing codes given the large numbers of patients involved. However, it is unlikely that there would be systematic bias in errors for billing codes with respect to the variables tested, except possibly site (it is possible that the community site would have more patients being diagnosed with cancer outside of our healthcare network). In addition, we included multiple billing codes associated with lung cancer, recorded at any point within one year of the screening CT, to mitigate this issue. We do note that the likelihood of at least a small number of missed lung cancer diagnoses means that our false positive rate more accurately represents an upper bound for the true false positive rate. Another limitation is in calculation of radiologist annual volume, which only measures lung cancer screening CTs read within our network; we did not account for non-lung cancer screening chest CT volume or any work done outside of our network.

In conclusion, a number of factors influence the rate of false positive results in CT LCS, including patient factors such as emphysema/COPD, baseline scans, and radiologist factors including radiologist experience. Policymakers should be aware of these factors when assessing the cost-effectiveness of lung cancer screening in the real world, and radiologists should be aware of these factors when building and running lung cancer screening programs. Hopefully, use of quality assurance and audit programs can help to reduce site-to-site and radiologist-to-radiologist variability and improve the cost-effectiveness of lung cancer screening in the future.

## The authors report no relevant financial disclosures.

C.Y.K. supported by the National Cancer Institute (U01CA199284).

## References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med* 2011;365(5):395–409. [PubMed: 21714641]
2. Black WC, Gareen IF, Soneji SS, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. *N. Engl. J. Med* 2014;371(19):1793–1802. [PubMed: 25372087]
3. Criss SD, Cao P, Bastani M, et al. Cost-Effectiveness Analysis of Lung Cancer Screening in the United States: A Comparative Modeling Study. *Ann. Intern. Med* 2019;
4. Taghizadeh N, Tremblay A, Cressman S, et al. Health-related quality of life and anxiety in the PAN-CAN lung cancer screening cohort. *BMJ Open*. 2019;9(1):e024719.
5. National Lung Screening Trial Research Team, Church TR, Black WC, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N. Engl. J. Med* 2013;368(21):1980–1991. [PubMed: 23697514]
6. American College of Radiology. Lung-RADS<sup>®</sup> Version 1.1. 2019;
7. Pinsky PF, Gierada DS, Black W, et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Ann. Intern. Med* 2015;162(7):485–491. [PubMed: 25664444]
8. Baptiste JV, Jankowich M, Nici LL. Lung Cancer Screening with Low Dose CT: Two Year Experience at Providence Veteran Affairs Medical Center. *American Thoracic Society International Conference Abstracts*. 2017;A5178–A5178.
9. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. *N. Engl. J. Med* 2009;361(23):2221–2229. [PubMed: 19955524]
10. Young RP, Duan F, Chiles C, et al. Airflow Limitation and Histology Shift in the National Lung Screening Trial. The NLST-ACRIN Cohort Substudy. *Am J Respir Crit Care Med*. 2015;192(9):1060–1067. [PubMed: 26199983]
11. Balekian AA, Tanner NT, Fisher JM, Silvestri GA, Gould MK. Factors Associated with a Positive Baseline Screening Exam Result in the National Lung Screening Trial. *Ann Am Thorac Soc*. 2016;13(9):1568–1574. [PubMed: 27387658]
12. Greenberg AK, Lu F, Goldberg JD, et al. CT Scan Screening for Lung Cancer: Risk Factors for Nodules and Malignancy in a High-Risk Urban Cohort. *PLoS One*. 2012;7(7):.
13. Alberdi RZ, Llanes ABF, Ortega RA, et al. Effect of radiologist experience on the risk of false-positive results in breast cancer screening programs. *Eur Radiol*. 2011;21(10):2083–2090. [PubMed: 21643887]
14. Théberge I, Chang S-L, Vandal N, et al. Radiologist interpretive volume and breast cancer screening accuracy in a Canadian organized screening program. *J. Natl. Cancer Inst* 2014;106(3):djt461. [PubMed: 24598715]
15. American College of Radiology. ACR–STR PRACTICE PARAMETER FOR THE PERFORMANCE AND REPORTING OF LUNG CANCER SCREENING THORACIC COMPUTED TOMOGRAPHY. 2019;

**Table 1.**

## Patient demographics

N=3735	
Male	2021 (54%)
Age, median (range), years	65 (39 – 81)
Race	
White	3185 (85%)
Black	206 (6%)
Hispanic	78 (2%)
Asian	65 (2%)
Smoking status	
Current Smoker	1579 (48%)
Former Smoker	1627 (50%)
Never Smoker	65 (2%)
COPD diagnosis	1213 (32%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 2.**

## Screening round characteristics

N=5835	
Site	
Academic 1	3882 (67%)
Academic 2	1719 (29%)
Community	234 (4%)
Baseline scan	
Emphysema on CT	2944 (50%)
Lung-RADS	
0	8 (0.001%)
1	1505 (26%)
2	3417 (59%)
3	517 (9%)
4A	233 (4%)
4B	86 (1%)
4X	69 (1%)
Lung cancer diagnosed within 1 year	
Sensitivity	139/142 (98%)
Specificity	4928/5693 (87%)
Radiologist annual volume, median (range)	55 (1 – 197)
Radiologist experience, median (range), years	16 (0 – 45)

**Table 3.**

## Lung-RADS Score and Diagnoses of Lung Cancer

<b>Lung-RADS</b>	<b>Lung cancers</b>
0	0/8 (0%)
1	1/1505 (0%)
2	2/3417 (0%)
3	15/517 (3%)
4A	38/233 (16%)
4B	35/86 (41%)
4X	51/69 (74%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4.**

Multivariate logistic regression model for false positive rate

Variable	OR (95% CI)	p-value
Site		
Academic 1 vs Academic 2	2.43 (1.94 – 3.04)	<0.0001
Community vs Academic 2	0.91 (0.49 – 1.70)	0.66
Baseline	1.43 (1.19 – 1.72)	0.0002
Radiologist experience		
Per 5 years	0.94 (0.91 – 0.98)	0.001
Range	0.59 (0.43 – 0.81)	
Radiologist annual volume	0.86 (0.61 – 1.21)	0.39
Male Gender	1.10 (0.93 – 1.29)	0.27
Age		
Per 10 years	1.19 (1.04 – 1.36)	0.01
Range	2.08 (1.16 – 3.72)	
Smoking status	1	1
Emphysema	1.32 (1.12 – 1.57)	0.001
COPD	1.34 (1.13 – 1.60)	0.001
Median income by Zip code		
Per \$1,000	1.00 (0.99 – 1.00)	0.01
Range	0.43 (0.22 – 0.84)	

*CI*, confidence interval.

**Table 5.**

Distribution of Lung-RADS scores by radiologist experience

<b>Lung-RADS</b>	<b>Low Experience* (n=1582)</b>	<b>High Experience* (n=4244)</b>
0	1 (0.06%)	7 (0.2%)
1	289 (18%)	1214 (29%)
2	1022 (65%)	2392 (56%)
3	146 (9%)	369 (9%)
4A	72 (5%)	160 (4%)
4B	31 (2%)	55 (1%)
4X	21 (1%)	47 (1%)
Positive	270 (17%)	631 (15%)
False positive	235 (15%)	528 (13%)

\* Low experience is ≤ 5 years; high experience is > 5 years.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript