

HHS Public Access

Author manuscript Ann Intern Med. Author manuscript; available in PMC 2016 April 07.

Published in final edited form as:

Ann Intern Med. 2015 April 7; 162(7): 485-491. doi:10.7326/M14-2086.

Performance of Lung-RADS in the National Lung Screening

Trial:

A Retrospective Assessment

Paul F. Pinsky, PhD, David S. Gierada, MD, William Black, MD, Reginald Munden, MD, Hrudaya Nath, MD, Denise Aberle, MD, and Ella Kazerooni, MD

National Cancer Institute, Bethesda, Maryland; Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, Missouri; Geisel School of Medicine at Dartmouth, Hanover, New Hampshire; Houston Methodist Hospital, Houston, Texas; University of Alabama at Birmingham, Birmingham, Alabama; David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California; and University of Michigan, Ann Arbor, Michigan

Abstract

Background—Lung cancer screening with low-dose computed tomography (LDCT) has been recommended, based primarily on the results of the NLST (National Lung Screening Trial). The American College of Radiology recently released Lung-RADS, a classification system for LDCT lung cancer screening.

Objective—To retrospectively apply the Lung-RADS criteria to the NLST.

Design—Secondary analysis of a group from a randomized trial.

Setting—33 U.S. screening centers.

Patients—Participants were randomly assigned to the LDCT group of the NLST, were aged 55 to 74 years, had at least a 30–pack-year history of smoking, and were current smokers or had quit within the past 15 years.

Dr. Black: Department of Radiology, Geisel School of Medicine at Dartmouth, 1 Rope Ferry Road, Hanover, NH 03755. Dr. Munden: 6565 Fannin Street, Houston, TX 77030.

Dr. Aberle: Department of Radiologic Sciences, David Geffen School of Medicine at University of California, Los Angeles, 924 Westwood Boulevard, Los Angeles, CA 90024.

Requests for Single Reprints: Paul F. Pinsky, PhD, National Cancer Institute, 9609 Medical Center Drive, Room 5E108, Bethesda, MD, 20892; pp4f@nih.gov.

Current Author Addresses: Dr. Pinsky: National Cancer Institute, 9609 Medical Center Drive, Room 5E108, Bethesda, MD 20892. Dr. Gierada: Washington University School of Medicine, 510 South Kingshighway Boulevard, St. Louis, MO 63110.

Dr. Nath: University of Alabama at Birmingham School of Medicine, 1720 2nd Avenue South, Birmingham, AL 35294.

Dr. Kazerooni: Department of Radiology, University of Michigan, 1500 East Medical Center Drive, Ann Arbor, MI 48109.

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M14-2108.

Reproducible Research Statement: *Study protocol and statistical code:* Available from Dr. Pinsky (pp4fnih.gov). *Data set:* Available at https://biometry.nci.nih.gov/cdas.

Author Contributions: Conception and design: P.F. Pinsky, W. Black, R. Munden, D. Aberle, E. Kazerooni. Analysis and interpretation of the data: P.F. Pinsky, D.S. Gierada, W. Black, R. Munden, H. Nath, E. Kazerooni. Drafting of the article: P.F. Pinsky, H. Nath.

Critical revision of the article for important intellectual content: P.F. Pinsky, D.S. Gierada, W. Black, R. Munden, H. Nath, D. Aberle, E. Kazerooni.

Final approval of the article: P.F. Pinsky, D.S. Gierada, W. Black, R. Munden, H. Nath, D. Aberle, E. Kazerooni.

Provision of study materials or patients: R. Munden, D. Aberle, E. Kazerooni.

Statistical expertise: P.F. Pinsky. Collection and assembly of data: R. Munden, E. Kazerooni.

Intervention—3 annual LDCT lung cancer screenings.

Measurements—Lung-RADS classifications for LDCT screenings. Lung-RADS categories 1 to 2 constitute negative screening results, and categories 3 to 4 constitute positive results.

Results—Of 26 722 LDCT group participants, 26 455 received a baseline screening; 48 671 screenings were done after baseline. At baseline, the false-positive result rate (1 minus the specificity rate) for Lung-RADS was 12.8% (95% CI, 12.4% to 13.2%) versus 26.6% (CI, 26.1% to 27.1%) for the NLST; after baseline, the false-positive result rate was 5.3% (CI, 5.1% to 5.5%) for Lung-RADS versus 21.8% (CI, 21.4% to 22.2%) for the NLST. Baseline sensitivity was 84.9% (CI, 80.8% to 89.0%) for Lung-RADS versus 93.5% (CI, 90.7% to 96.3%) for the NLST, and sensitivity after baseline was 78.6% (CI, 74.6% to 82.6%) for Lung-RADS versus 93.8% (CI, 91.4% to 96.1%) for the NLST.

Limitation—Lung-RADS criteria were applied retrospectively.

Conclusion—Lung-RADS may substantially reduce the false-positive result rate; however, sensitivity is also decreased. The effect of using Lung-RADS criteria in clinical practice must be carefully studied.

Primary Funding Source—National Institutes of Health.

The U.S. Preventive Services Task Force recently recommended (grade B) lung cancer screening with low-dose computed tomography (LDCT) for high-risk current and former smokers (1). The primary evidence used by the Task Force was the National Lung Screening Trial (NLST), which reported a 20% reduction in lung cancer–specific death associated with LDCT screening (2). Important considerations for widespread use of LDCT lung cancer screening in clinical practice include the definition of a positive result in computed tomography (CT) screening and the appropriate management of positive screening results.

Much knowledge has accumulated since the NLST was designed in 2002. In this trial, the definition of a positive screening result was a nodule of 4 mm or greater in the longest diameter that had no specific benign calcification patterns. In addition, the NLST achieved its results without a trial-wide specified protocol for diagnostic management for positive screening results. A recent reanalysis of the NLST examined the effect of different cutoffs defining a positive screening result and found that increasing the threshold to 6 or 8 mm would have resulted in substantial decreases in the false-positive result rate with only small corresponding decreases in sensitivity (3). The International Early Lung Cancer Action Program reported similar results for baseline LDCT screenings, showing a substantial reduction in the positivity rate of screening results with increasing size cutoffs and only a few resultant missed cancer cases (4).

Over this period, several professional organizations have promulgated lung cancer screening guidelines, many of which define a positive screening result and include nodule management (5–7). The American College of Radiology recently began efforts to standardize the reporting of LDCT screening results in a manner analogous to the use of the Breast Imaging Reporting and Data System for mammography, based on the best available data. This effort included defining a positive result on lung cancer screening CT in the most effective manner, attempting to reduce the substantial false-positive result rate while having

the least possible effect on test sensitivity, and suggesting management recommendations based on lung cancer risk. Published data from several LDCT screening studies, including the NLST, the International Early Lung Cancer Action Program, and the European NELSON (Nederlands-Leuvens Longkanker Screenings Onderzoek) trial, were used by a consensus panel to help derive positivity criteria for Lung-RADS (4, 8–10), which was officially released in May 2014 (11). Compared with the NLST criteria, Lung-RADS increases the size threshold for a positive baseline screening result from a 4-mm greatest transverse diameter to a 6-mm transverse bidimensional average (and to 20 mm for nonsolid nodules) and requires growth for preexisting nodules.

Although data from the NLST, in part, were used to develop the Lung-RADS criteria, only published summary-level data were considered. These data were sufficient to give an approximate positivity rate for Lung-RADS as applied to the NLST but not to give an exact distribution of Lung-RADS scores. This is especially the case for screenings after baseline, where the individual nodule history over time is critical in defining the Lung-RADS category.

We used participant- and nodule-level data to retrospectively apply the Lung-RADS criteria to the NLST. We evaluate the effect of Lung-RADS on the performance characteristics of LDCT screening, including sensitivity, false-positive result rate, positive predictive value (PPV), and negative predictive value (NPV). In addition, we compare the characteristics of the cancer cases detectable by Lung-RADS with those that it would have missed.

Methods

NLST Design

The NLST randomly assigned participants aged 55 to 74 years to LDCT or chest radiography screening. Eligibility criteria included 30 pack-years of smoking or greater and current smoking status or having quit within the past 15 years (12). Participants were recruited at 33 U.S. centers from 2002 to 2004 and received either LDCT or chest radiography over 3 annual screening rounds (denoted T0, T1, and T2). The NLST was approved by the institutional review board at each screening center, and all participants provided informed consent.

The NLST study protocol defined a noncalcified nodule (NCN) of 4 mm or greater in the longest transverse diameter as a positive screening result. For each NCN that was 4 mm or greater, radiologists used standardized forms to report location, greatest transverse and perpendicular diameters, margins, and attenuation characteristics. At T1 and T2, they reported whether the abnormality was preexisting or new based on examinations of previous images and, if preexisting, whether it had grown and whether a suspicious change in attenuation had occurred since past screenings. Noncalcified nodules that were unchanged from T0 to T2, representing stability for 2 years, could be considered benign and constitute a negative screening result at the radiologist's discretion. Other abnormalities, including adenopathy or effusion, could also trigger positive screening results.

Positive results were tracked for resultant diagnostic procedures and lung cancer diagnoses. In addition, participants were followed with annual surveys to ascertain incident cancer cases. All reported cancer cases were verified with medical records, with stage and histologic characteristics recorded. Deaths were tracked with the annual surveys and supplemented by National Death Index searches.

Lung-RADS

Table 1 describes the primary criteria for defining Lung-RADS categories. Categories 1 (negative) and 2 (benign appearance) correspond to negative screening results, and categories 3 (probably benign) and 4 (suspicious) correspond to positive screening results. Category 4 is further divided into 4A, 4B, and 4X (8). In the context of annual screening, a negative screening result assumes that reevaluation will occur at the next annual screening, whereas a positive screening result means that additional evaluation is recommended before the next annual screening. The distinctions between the positive screening categories are important because Lung-RADS management guidelines differ substantially across categories, ranging from follow-up CT at 6 months for category 3 to positron emission tomography and CT or biopsy for 4B. In addition, category 2 involves tracking small nodules on the next annual screening, and category 1 does not involve tracking nodules.

Lung-RADS criteria distinguish between baseline (first) and subsequent screenings. For baseline screenings (generally lacking comparison examinations), the criteria are based on nodule size, as measured by average diameter, and nodule attenuation (solid, part-solid, or nonsolid). For subsequent screenings, the criteria also consider the preexistence and growth of the nodule. For baseline screenings, positive screening results for solid and part-solid nodules require a size of 6 mm, and 20 mm is required for nonsolid (that is, ground-glass) nodules. For positivity on subsequent screenings, 4 mm is required for new (solid or part-solid) nodules, and preexisting nodules must show growth, defined as an increase in size of greater than 1.5 mm. New or growing nonsolid nodules still must meet the 20-mm size requirement. For part-solid nodules, the size and/or growth of the solid component is also considered. The overall Lung-RADS score. Category 3 or 4 nodules with additional features (such as spiculation) or imaging findings that increase suspicion for cancer (such as enlarged lymph nodes) can qualify as category 4X.

Applying Lung-RADS to the NLST

The average diameter for NLST nodules was computed as the mean of the longest diameter and the longest perpendicular diameter. The NLST attenuation classifications of soft tissue, ground glass, and mixed were mapped to the Lung-RADS classifications of solid, nonsolid, and part-solid, respectively. The NLST did not report the amount of growth but only whether growth occurred; therefore, report of growth in the NLST was considered nodule growth for Lung-RADS.

For part-solid nodules, the NLST did not report the size of the solid component, which may be required to distinguish among categories 3, 4A, and 4B. Therefore, if NLST data for a part-solid nodule were consistent with 2 or more categories of 3 or higher, a range (such as 3

to 4B) instead of a single category was denoted for our analysis. These category ranges for part-solid nodules were used only if they constituted (at the upper limit of their range) the nodule with the highest degree of suspicion. In addition, the solid component was assumed to be growing if the nodule as a whole was reported as growing or there was a suspicious change in attenuation; growth specifically of the solid component was not recorded in the NLST.

If other suspicious findings, in the absence of any nodules measuring 4 mm or greater, constituted a positive screening result in the NLST, this was classified as category 4X for Lung-RADS.

Quantitative Methods

Lung cancer was deemed to be present at a screening if it was diagnosed within 1 year or before the next screening (whichever came first) or, for positive results, if it was diagnosed after a longer period but with no time gap between diagnostic procedures of more than 1 year. Lung cancer was deemed to be absent if it was not deemed present and there was at least 1 year of follow-up from screening or the participant died within 1 year of screening. Sensitivity, specificity, PPV, and NPV were computed for Lung-RADS with a positive screening result defined as category 3 or higher. Sensitivity was defined as the percentage of screenings with cancer present that were positive, specificity as the percentage of screenings with cancer present that were negative (with the false-positive result rate being 1 minus the specificity rate), PPV as the percentage of positive screening results with cancer present, and NPV as the percentage of negative screening results with cancer absent. These quantities were also computed using the original definition for a positive screening result in the NLST. Participants with indeterminate cancer status (cancer not present but also not deemed to be absent due to follow-up of <1 year) were excluded from the calculations of the previously mentioned statistics.

The statistical significance of differences in sensitivity and specificity between Lung-RADS and the original NLST criteria were determined using the McNemar test to evaluate discordant pairs (positive screening results by only 1 criterion); Wald-type CIs for these differences were computed on the basis of the multinomial distribution (13); CIs for the ratio of PPV and NPV for Lung-RADS versus the NLST criteria were computed using the method of Moskowitz and Pepe (14).

Screen-detected (true-positive) cancer cases in the NLST that were "missed" by Lung-RADS were defined as those with a negative Lung-RADS score (1 or 2) at that screening. The characteristics of missed versus nonmissed true-positive cancer were compared, including stage, histology, and survival. We used the Kaplan–Meier method to estimate lung cancer–specific survival, the log-rank test to assess statistical differences in survival between groups, and the chi-square test to assess group differences in stage and histologic characteristics. All analyses were done using SAS, version 9.2 (SAS Institute).

False-positive screening results in the NLST that would have been "avoided" by Lung-RADS were defined as the (NLST) false-positive results that were negative (score of 1 or 2) on Lung-RADS. In the same way, the number of false-positive results in the NLST with

follow-up invasive procedures that were avoided by Lung-RADS was defined as the number of such screening results that had negative Lung-RADS scores. The number of follow-up chest CT scans after false-positive screening results in the NLST that were avoided by Lung-RADS was computed by summing all such scans for which the corresponding screening results were negative with Lung-RADS.

Because NLST radiologists did not record the amount but only the occurrence of growth and the Lung-RADS criteria defined growth as a size increase greater than 1.5 mm, we did a sensitivity analysis of our assumption that all nodules reported to be growing in the NLST increased in size by at least that amount. We examined the subset of screenings in which the following occurred: The radiologist reported a growing nodule in a given lobe (lobe X), lobe X had a single reported nodule of 4 mm or greater at that screening and the previous screening, and the same radiologist read the current and previous screening. For these screenings, we computed the amount of growth as the difference in the reported average size of the nodule in lobe X at the 2 screenings. These assumptions eliminated interreader variability and attempt to ensure that the same nodule is being measured at both screenings. However, individual nodules were not tracked, so it is possible that the nodule reported at the previous screening was not the growing nodule. Further, individual measurements were not taken with the aim of optimally estimating amount of growth.

For the previously mentioned subset of screenings in which the amount of growth could be estimated, we computed the proportions for which growth exceeded 1.5 mm (denoted as P_{CP} and P_{CA} for screenings with cancer present and cancer absent, respectively). We then used these proportions to compute alternative estimates of true- and false-positive result rates in Lung-RADS after baseline for the sensitivity analysis. Only the screenings after baseline that had positive results in Lung-RADS due solely to a growing nodule (denoted as $N_{CP,G}$ and $N_{CA,G}$ for cancer present and cancer absent, respectively) were affected by the assumptions about the amount of growth; screening results after baseline that were positive in Lung-RADS due to a new nodule or to nonnodule positive findings (denoted as $N_{CP,N}$ and $N_{CA,N}$, respectively) were not affected. Therefore, the alternative estimates of true-positive and false-positive result rates for screening results after baseline under Lung-RADS were computed as $[N_{CP,G} \times P_{CP} + N_{CP,N}] \div N_{CP}$ and $[N_{CA,G} \times P_{CA} + N_{CA,N}] \div N_{CA}$ for true- and false-positive results, respectively, where N_{CP} and N_{CA} were the total number of screenings after baseline with cancer present and cancer absent, respectively.

Role of the Funding Source

The NLST was supported by the National Institutes of Health. The funding source had no role in the study design, analysis or interpretation of data, or writing of the article.

Results

Of 26 722 total LDCT group participants, 26 455 received an NLST screening, with 26 309 receiving their initial screening at T0 and 146 receiving their initial screening at T1 or T2. All initial screenings, regardless of screening round, were denoted as "baseline" screenings. Of the 26 455 screenings, 23 574 received 2 subsequent NLST screenings, 1523 received 1

subsequent screening, and 1358 did not receive a subsequent screening, for a total of 48 671 subsequent screenings.

Table 2 shows the Lung-RADS categories for the baseline screening. Among screening results with cancer present, most were category 4A (26.7%) or 4B (42.5%). Among screening results with cancer absent, most were category 1 (56.2%) or 2 (31.0%). Cancer prevalence generally increased with Lung-RADS category, increasing from 0.1% (category 1) to 34.7% (category 4B).

Table 3 displays Lung-RADS categories for subsequent screenings. For screening results with cancer present, approximately three fourths (72.3%) were either 4A or 4B; the proportion with category 2 increased to 18.2% (from 9.9% at baseline). For screening results without cancer, the proportion with category 2 increased from 31.0% at baseline to 42.5%; the proportion with category 1 (52.2%) was similar to baseline.

Table 4 shows the performance characteristics of Lung-RADS versus the original NLST criteria. At baseline, sensitivity with Lung-RADS (84.9% [95% CI, 80.8% to 89.0%]) was lower than for the NLST criteria (93.5% [CI, 90.7% to 96.3%]) (difference, 8.6% [CI, 5.4% to 11.8%]; P < 0.001). The Lung-RADS false-positive result rate (12.8% [CI, 12.4% to 13.2%]) was also lower than for the NLST (26.6% [CI, 26.1% to 27.1%]) (difference, 13.8% [CI, 13.4% to 14.2%]; P < 0.001). For subsequent screenings, Lung-RADS sensitivity decreased to 78.6% (CI, 74.6% to 82.6%), compared with 93.8% (CI, 91.4% to 96.1%) for the NLST (difference, 15.2% [CI, 11.7% to 18.7%]; P < 0.001). The false-positive result rate at subsequent screenings was 5.3% (CI, 5.1% to 5.5%) for Lung-RADS versus 21.8% (CI, 21.4% to 22.2%) for the NLST (difference, 16.5% [CI, 16.2% to 16.9%]; *P* < 0.001). The PPVs at baseline were 6.9% (CI, 6.1% to 7.7%) for Lung-RADS versus 3.8% (CI, 3.3% to 4.2%) for NLST criteria (P < 0.001); at subsequent screenings, PPVs were 11.0% (CI, 9.9% to 12.2%) for Lung-RADS versus 3.5% (CI, 3.1% to 3.8%) for NLST criteria (P <0.001). The ratio of PPVs (Lung-RADS vs. the NLST) were 1.8 (CI, 1.7 to 1.9) at baseline and 3.2 (CI, 3.0 to 3.4) at subsequent screenings. Negative predictive values were uniformly very high (99.8%) but significantly greater for the NLST than for Lung-RADS at both baseline and subsequent screenings (P < 0.001).

Tables 3 and 4 display aggregate data for the 2 screening rounds after baseline; however, Lung-RADS results for these 2 rounds were generally similar (Appendix Tables 1 to 3, available at www.annals.org).

Of all true-positive cancer cases according to NLST criteria (n = 649), 86 (13%) were missed with Lung-RADS: 25 (9.2%) on baseline screenings and 61 (16.2%) on screenings after baseline (Table 5). Of the 25 cases, 12 had only ground-glass nodules smaller than 20 mm and 13 had solid or part-solid nodules smaller than 6 mm. Of the 61 cases missed on subsequent screenings, 26 had only ground-glass nodules smaller than 20 mm (of which 17 were new and 3 were growing) and 35 had solid (or part-solid) preexisting nongrowing nodules.

The 86 missed cancer cases had a stage distribution (65.1% were stage I) similar to that of the true-positive cancer cases that were not missed (61.1% were stage I) (P = 0.48). For

histology, the proportions with adenocarcinoma and small-cell carcinoma were not statistically different between the groups (60.5% vs. 53.3% for adenocarcinoma among the missed and nonmissed cases, respectively [P = 0.21] and 11.6% and 6.9% for small-cell carcinoma among the missed and nonmissed cases [P = 0.124]). However, squamous cell carcinoma was significantly less frequent among the missed cancer cases (10.5% vs. 22.6%; P = 0.010). Lung cancer–specific survival did not significantly differ between the groups, with 5-year survival of 71.7% for the missed cancer cases versus 64.2% for those not missed (P = 0.22).

Table 5 also shows false-positive screening results and resultant diagnostic procedures that would have been avoided using Lung-RADS. At baseline, 52% of false-positive screening results and a similar percentage of follow-up chest CTs would have been avoided; the percentage of false-positive results with invasive diagnostic procedures avoided was lower (23%). A higher percentage of false-positive results was avoided at subsequent screenings (76%), but the proportions of diagnostic procedures avoided were similar as at baseline.

For false-positive screening results in Lung-RADS after baseline, 56.9% had a new nodule reported, 33.9% had a growing nodule (and no new nodule), and 9.2% had nonnodule-related positive findings. In contrast, among Lung-RADS true-positive screening results after baseline, 59.7% had a growing nodule (without any new nodules). For the sensitivity analysis on nodule growth assumptions, the amount of growth could be estimated for 204 Lung-RADS false-positive screening results and 65 Lung-RADS true-positive screening results, with resulting proportions of growth greater than 1.5 mm of 42.6% and 78.5%, respectively. The alternative Lung-RADS false-positive and true-positive (sensitivity) rates for screenings after baseline were 4.2% and 68.5%, respectively.

Discussion

In this retrospective analysis applying Lung-RADS criteria to the NLST, we found a considerably lower false-positive result rate than in the NLST, especially at screenings after baseline. Sensitivity was lower with Lung-RADS at screenings at and after baseline than with the NLST. The PPV with Lung-RADS was nearly twice as high for baseline and 3-fold higher for screenings after baseline compared with the NLST.

The high false-positive result rate seen in the NLST and LDCT screening studies is generally a major harm and cost driver of LDCT lung cancer screening. Once LDCT screening disseminates and approaches a steady state, most screenings will be done after baseline, at which the reduction in the false-positive result rate with Lung-RADS compared with the NLST was the greatest: approximately 75%. The corresponding reduction (after baseline) in diagnostic procedures was substantially lower (23% for invasive procedures and 46% for chest CTs), reflecting that the reduction in false-positive results was weighted toward lower-risk nodules. Nonetheless, using Lung-RADS still has the potential to substantially reduce the burden of LDCT screening. A critical question, however, is how the corresponding sensitivity reduction might affect the mortality benefit of LDCT screening.

Screen-detected cancer cases in the NLST that would have been missed by Lung-RADS either were smaller than 6 mm or were ground-glass nodules (<20 mm) at baseline and were either nongrowing and/or ground-glass nodules at subsequent screenings. Therefore, Lung-RADS-negative (and NLST screen-detected) cancer cases might be believed to be less aggressive than Lung-RADS-positive cancer cases. However, 5-year lung cancer-specific survival did not significantly differ between the groups, and similar proportions in each group were stage I cancer. The effect of delaying diagnosis of these Lung-RADS missed cancer cases is unknown, but it cannot be assumed that most are indolent and would not affect lung cancer mortality rates. Although we have considered NLST screen-detected cancer cases that are negative on Lung-RADS as contributing to decreased sensitivity of Lung-RADS, it is unknown what proportion would have presented clinically within the next year as true-interval cancer or would have been screen-detected with Lung-RADS on the next round of screening. Still, even delayed screen-detected diagnosis could adversely affect survival. Our sensitivity analysis using an alternate assumption for growing nodules demonstrated that sensitivity after baseline using Lung-RADS could be further reduced, from 78.6% to 68.5%, compared with 93.8% in the NLST. Sensitivity under Lung-RADS will be an important quality indicator of LDCT screening in clinical practice, and it will be critical going forward to monitor it, as well as the false-positive result rate, using population screening registries. As prospective performance characteristics of Lung-RADS become available, it is expected that it will be revised, similar to the process the American College of Radiology has used to revise the Breast Imaging Reporting and Data System classification scheme for breast cancer screening, now in its fifth edition.

An important limitation of our analysis is that it was retrospective, applying Lung-RADS criteria to participants previously screened using a different definition for screening positivity. Because there is some variability in manually placing electronic cursors during nodule measurement, a radiologist reporting a 5-mm nodule under criteria in which a nodule of 4 mm or greater constituted a positive screening result might not necessarily report the same 5-mm size when a 6-mm cutoff defined a positive screening result; if there is something concerning about the nodule, the radiologist might be biased toward recording a larger measurement, and vice versa. Therefore, it will be important to assess the performance characteristics of Lung-RADS prospectively in settings where it is being used to actually determine screening outcomes.

Overall, applying Lung-RADS criteria to the NLST substantially reduced the false-positive result rate, with smaller corresponding reductions in sensitivity. These findings suggest good performance characteristics for Lung-RADS; however, the potential effect of reduced sensitivity on the mortality benefit of LDCT screening is unknown. Further validation with prospective data collection will be necessary going forward.

Acknowledgments

Grant Support: The NLST was supported by the following grants and contracts: U01-CA-80098, U01-CA-79778, N01-CN-25522, N01-CN-25511, N01-CN-25512, N01-CN-25513, N01-CN-25514, N01-CN-25515, N01-CN-25516, N01-CN-25518, N01-CN-25524, N01-CN-75022, N01-CN-25476, and N02-CN-63300.

References

- Moyer VA. U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2014; 160:330–338. [PMID: 24378917]. [PubMed: 24378917]
- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011; 365:395–409. [PMID: 21714641]. [PubMed: 21714641]
- Gierada DS, Pinsky P, Nath H, Chiles C, Duan F, Aberle DR. Projected outcomes using different nodule sizes to define a positive CT lung cancer screening examination. J Natl Cancer Inst. 2014; 106 [PMID: 25326638].
- Henschke CI, Yip R, Yankelevitz DF, Smith JP. International Early Lung Cancer Action Program Investigators. Definition of a positive test result in computed tomography screening for lung cancer: a cohort study. Ann Intern Med. 2013; 158:246–252. [PMID: 23420233]. [PubMed: 23420233]
- MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett JR, Naidich DP, et al. Fleischner Society. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society [Editorial]. Radiology. 2005; 237:395–400. [PMID: 16244247]. [PubMed: 16244247]
- National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Fort Washington, PA: National Comprehensive Cancer Network; 2013. Lung Cancer Screening Version 1.2014. Accessed at www.nccn.org/professionals/physician_gls/f_guidelines.asp#detection on 26 January 2015
- Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. Radiology. 2013; 266:304–317. [PMID: 23070270]. [PubMed: 23070270]
- Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al. National Lung Screening Trial Research Team. Results of initial low-dose computed tomographic screening for lung cancer. N Engl J Med. 2013; 368:1980–1991. [PMID: 23697514]. [PubMed: 23697514]
- Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al. National Lung Screening Trial Research Team. Results of the two incidence screenings in the National Lung Screening Trial. N Engl J Med. 2013; 369:920–931. [PMID: 24004119]. [PubMed: 24004119]
- Horeweg N, van der Aalst CM, Vliegenthart R, Zhao Y, Xie X, Scholten ET, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. Eur Respir J. 2013; 42:1659–1667. [PMID: 23845716]. [PubMed: 23845716]
- 11. American College of Radiology. Lung CT Screening Reporting and Data System (Lung-RADS). Accessed at www.acr.org/Quality-Safety/Resources/LungRADS on 11 August 2014.
- Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, et al. National Lung Screening Trial Research Team. The National Lung Screening Trial: overview and study design. Radiology. 2011; 258:243–253. [PMID: 21045183]. [PubMed: 21045183]
- Saeki H, Tango T. Non-inferiority test and confidence interval for the difference in correlated proportions in diagnostic procedures based on multiple raters. Stat Med. 2011; 30:3313–3327. [PMID: 21953516]. [PubMed: 21953516]
- Moskowitz CS, Pepe MS. Comparing the predictive values of diagnostic tests: sample size and analysis for paired study designs. Clin Trials. 2006; 3:272–279. [PMID: 16895044]. [PubMed: 16895044]

Appendix

Page 11

Appendix Table 1

Lung-RADS Classification: First Screening After Baseline*

Lung-RADS Category	Cancer Present, n (%)	Cancer Absent, n (%)	Indeterminate Cancer Status, n (%)	All Classifications, n (%)	With Cancer, $\%^{\dagger}$
1	5 (2.7)	13 085 (52.6)	23 (60.5)	13 113 (52.3)	0.04
2	35 (19.2)	10 450 (42.0)	13 (34.2)	10 498 (41.8)	0.30
3	6 (3.3)	304 (1.2)	0 (0.0)	310 (1.2)	1.90
4A	13 (7.1)	421 (1.7)	1 (2.6)	435 (1.7)	3.00
4A or 4B	17 (9.3)	134 (0.5)	0 (0.0)	151 (0.6)	11.30
4B	99 (54.4)	370 (1.5)	1 (2.6)	470 (1.9)	21.10
4X	7 (3.8)	113 (0.5)	0 (0.0)	120 (0.5)	5.80
All	182 (100)	24 877 (100)	38 (100)	25 097 (100)	0.70

*Percentages may not sum to 100 due to rounding.

 † Excludes participants with indeterminate cancer status.

Appendix Table 2

Lung-RADS Classification: Second Screening After Baseline*

Lung-RADS Category	Cancer Present, n (%)	Cancer Absent, n (%)	Indeterminate Cancer Status, n (%)	All Classifications, n (%)	With Cancer, % [†]
1	8 (3.7)	12 064 (51.7)	16 (45.7)	12 088 (51.3)	0.07
2	38 (17.4)	10 055 (43.1)	15 (42.9)	10 108 (42.9)	0.40
3	4 (1.8)	274 (1.2)	0 (0.0)	278 (1.2)	1.40
4A	26 (11.9)	389 (1.7)	1 (2.9)	416 (1.8)	6.30
4A or 4B	24 (11.0)	105 (0.5)	0 (0.0)	129 (0.5)	18.60
4B	111 (50.7)	313 (1.3)	2 (5.7)	426 (1.8)	26.10
4X	8 (3.7)	120 (0.5)	1 (2.9)	129 (0.5)	6.20
All	219 (100)	23 320 (100)	35 (100)	23 574 (100)	0.90

* Percentages may not sum to 100 due to rounding.

[†]Excludes participants with indeterminate cancer status.

Author Manuscript

Appendix Table 3

Sensitivity, Specificity, PPV, and NPV in the Lung-RADS and Original NLST Readings: First and Second Screenings After Baseline

Variable	First Screenin Baseline in Lun	ig After ig-RADS	First Screening Baseline in N	g After ILST	Second Screeni Baseline in Lun	ng After g-RADS	Second Screeni Baseline in N	ng After LST
	Percentage (95% CI)	N/u	Percentage (95% CI)	N/u	Percentage (95% CI)	N/u	Percentage (95% CI)	N/u
Sensitivity	78.00 (72.00–84.00)	142/182	94.50 (91.20–97.80)	172/182	79.00 (73.60–84.40)	173/219	93.20 (89.80–96.50)	204/219
False-positive result rate *	5.40 (5.10–5.70)	1342/24 877	27.30 (26.80–27.90)	6794/24 877	5.20 (4.90–5.40)	1201/23 320	15.90 (15.50–16.40)	3718/23 320
РРV	9.60 (8.10–11.10)	142/1484	2.50 (2.10–2.80)	172/6966	12.60 (10.80–14.30)	173/1374	5.20 (4.50–5.90)	204/3922
NPV	99.83 (99.77–99.88)	23 535/23 575	99.94 (99.90–99.98)	18 083/18 093	99.79 (99.75–99.83)	22 119/22 165	99.92 (99.89–99.95)	19 602/19 617
NL ST = National Luno Scree	ening Trial: NPV – negati	ve predictive valu	e. PPV = nositive predictiv	ve value				

* 1 minus the specificity rate.

Context

The definitions used to classify low-dose computed tomography findings may markedly influence the benefits and harms of lung cancer screening.

Contribution

This analysis of data from a large screening trial found that using the recently proposed Lung-RADS approach to classifying low-dose computed tomography findings substantially decreased the false-positive result rate but with a concomitant decrease in sensitivity.

Implication

Adopting the Lung-RADS classification system may improve the results of lung cancer screening programs.

Table 1

Summary of Lung-RADS Classification*

Lung-RADS Category	Baseline Screening	Subsequent Screening
1	No nodules; nodules with calcification	No nodules; nodules with calcification
2	Solid/part solid: <6 mm	Solid/part solid: <6 mm
	GGN: <20 mm	GGN: <20 mm or unchanged/slowly growing
	-	Category 3–4 nodules unchanged at 3 mo
3	Solid: 6 to <8 mm	Solid: New 4 to <6 mm
	Part solid: 6 mm with solid component <6 mm	Part solid: New <6 mm
	GGN: 20 mm	GGN: New 20 mm
4A	Solid: 8 to <15 mm	Solid: Growing <8 mm or new 6 and <8 mm
	Part solid: 8 mm with solid component 6 and <8 mm	Part solid: 6 mm with new or growing solid component <4 mm
4B	Solid: 15 mm	Solid: New or growing and 8 mm
	Part solid: Solid component 8 mm	Part solid: 6 mm with new or growing solid component 4 mm
4X	Category 3 or 4 nodules with additional features; imaging findings that increase suspicion of cancer	Category 3 or 4 nodules with additional features; imaging findings that increase suspicion of cancer

GGN = ground-glass nodule.

*Size is the average diameter rounded to the nearest whole number. Growth is a size increase >1.5 mm.

Author Manuscript

Lung-RADS Classification: Baseline Screening*

Lung-RADS Category	Cancer Present, n (%)	Cancer Absent, n (%)	Indeterminate Cancer Status, n (%)	All Classifications, n (%)	With Cancer, $\%^{\dagger}$
1	15 (5.1)	14 660 (56.2)	34 (46.6)	14 709 (55.6)	0.1
2	29 (9.9)	8087 (31.0)	29 (39.7)	8145 (30.8)	0.4
3	21 (7.2)	1672 (6.4)	4 (5.5)	1697 (6.4)	1.2
3 or 4A	0 (0.0)	96 (0.4)	1 (1.4)	97 (0.4)	0.0
3, 4A, or 4B	22 (7.5)	171 (0.7)	0 (0.0)	193 (0.7)	11.4
4A	78 (26.7)	1025 (3.9)	4 (5.5)	1107 (4.2)	7.1
4B	124 (42.5)	233 (0.9)	1 (1.4)	358 (1.4)	34.7
4X	3 (1.0)	146 (0.6)	0 (0.0)	149 (0.6)	2.0
All	292 (100)	26 090 (100)	73 (100)	26 455 (100)	1.1

ò

Ann Intern Med. Author manuscript; available in PMC 2016 April 07.

 $\vec{r}_{\rm Excludes}$ participants with indeterminate cancer status.

Author Manuscript

;	Baseline	
	After	12111
	Screenings	1011100
	sification	TO TRACTICO
,	š	1
į	C)
	ADS	
	Luno-R	

*

Lung-RADS Category	Cancer Present, n (%)	Cancer Absent, n (%)	Indeterminate Cancer Status, n (%)	All Classifications, n (%)	With Cancer, $\%^{\dagger}$
1	13 (3.2)	25 149 (52.2)	39 (53.4)	25 201 (51.8)	0.05
2	73 (18.2)	20 505 (42.5)	28 (38.4)	20 606 (42.3)	0.40
3	10 (2.5)	578 (1.2)	0 (0.0)	588 (1.2)	1.70
4A	39 (9.7)	810 (1.7)	2 (2.7)	851 (1.7)	4.60
4A or 4B	41 (10.2)	239 (0.5)	0 (0.0)	280 (0.6)	14.60
4B	210 (52.4)	683 (1.4)	3 (4.1)	896 (1.8)	23.50
4X	15 (3.7)	233 (0.5)	1 (1.4)	249 (0.5)	6.00
All	401 (100)	48 197 (100)	73 (100)	48 671 (100)	0.80

 $\vec{r} \, \mathrm{Excludes}$ participants with indeterminate cancer status.

Table 4

Sensitivity, Specificity, PPV, and NPV in the Lung-RADS and Original NLST Readings: Baseline and After Baseline*

Variable	Lung	g-RADS at Baseline		N	LST at Ba	seline
	Percentage	(95% CI)	n/N	Percentage	(95% CI)	n/N
Sensitivity	84.90 (80.8	0–89.00)	248/292	93.50 (90.7	0–96.30)	273/292
False-positive result rate	+ 12.80 (12.4	0–13.20)	3343/26 090) 26.60 (26.1	0–27.10)	6939/26 090
PPV	6.90 (6.10)–7.70)	248/3591	3.80 (3.3)	0-4.20)	273/7236
NPV	99.81 (99.7	5–99.86)	22 747/22 79	99.90 (99.8	6–99.94)	19 200/19 219
Lung-RADS After Baseline]	NLST After B	aseline	-	
Percentage (95% CI)	n/N	Percenta	ge (95% CI)	n/N		
78.60 (74.60-82.60)	315/401	93.80 (9	1.40–96.10)	376/401	-	
5.30 (5.10-5.50)	2543/48 197	21.80 (2	1.40–22.20)	10 512/48 197		
11.00 (9.90–12.20)	315/2858	3.50 (3	3.10–3.80)	376/10 888		
99.81 (99.77–99.85)	45 654/45 740	99.93 (9	9.90–99.96)	37 685/37 710		

NLST = National Lung Screening Trial; NPV = negative predictive value; PPV = positive predictive value.

* Totals of 22 screening results at baseline and 28 after baseline with cancer absent were positive in Lung-RADS and had nodule characteristics meeting the positive screening criteria but were nonetheless reported as negative screening results in the NLST. Otherwise, all screening results that were positive according to the Lung-RADS criteria were also positive according to the NLST criteria.

 † 1 minus the specificity rate.

Table 5

NLST True- and False-Positive Screening Results and Diagnostic Procedures Missed or Avoided With Lung-RADS*

Variable	Baseline	After Baseline	All
NLST true-positive † cases of cancer missed with Lung-RADS $^{\ddagger}_{-}$	25 (9.2)	61 (16.2)	86 (13.3)
NLST false-positive results avoided with Lung-RADS			
All§	3618 (52.1)	7997 (76.1)	11 615 (66.6)
With invasive procedures [#]	60 (23.4)	57 (23.3)	117 (23.4)
Chest CTs avoided after false-positive results $^{/\!\!/}$	3557 (50.5)	2150 (45.5)	5707 (48.5)

CT = computed tomography; NLST = National Lung Screening Trial.

*Values are numbers (percentages).

 † Screen-detected.

 \ddagger Denominators for percentages are the total number of cases of NLST true-positive cancer.

 $\ensuremath{\$}^{\ensuremath{\$}}$ Denominators for percentages are the total number of NLST false-positive results.

 $^{/\!/}$ Denominators for percentages are the total number of NLST false-positive results with invasive procedures.

 m Denominators for percentages are the total number of chest CTs after NLST false-positive results.