

Are Lung Imaging Reporting and Data System Categories Clear to Radiologists?

A Survey of the Korean Society of Thoracic Radiology Members on Ten Difficult-to-Classify Scenarios

Dae Hee Han, MD, PhD¹, Jin Mo Goo, MD, PhD^{2,3}, Semin Chong, MD, PhD⁴, Myeong Im Ahn, MD, PhD¹

¹Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea; ²Department of Radiology, Seoul National University College of Medicine, Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul 03080, Korea; ³Cancer Research Institute, Seoul National University, Seoul 03080, Korea; ⁴Department of Radiology, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul 06973, Korea

Objective: To evaluate possible variability in chest radiologists' interpretations of the Lung Imaging Reporting and Data System (Lung-RADS) on difficult-to-classify scenarios.

Materials and Methods: Ten scenarios of difficult-to-classify imaginary lung nodules were prepared as an online survey that targeted Korean Society of Thoracic Radiology members. In each question, a description was provided of the size, consistency, and interval change (new or growing) of a lung nodule observed using annual repeat computed tomography, and the respondent was instructed to choose one answer from five choices: category 2, 3, 4A, or 4B, or "un-categorizable." Consensus answers were established by members of the Korean Imaging Study Group for Lung Cancer.

Results: Of the 420 answers from 42 respondents (excluding multiple submissions), 310 (73.8%) agreed with the consensus answers; eleven (26.2%) respondents agreed with the consensus answers to six or fewer questions. Assigning the imaginary nodules to categories higher than the consensus answer was more frequent (16.0%) than assigning them to lower categories (5.5%), and the agreement rate was below 50% for two scenarios.

Conclusion: When given difficult-to-classify scenarios, chest radiologists showed large variability in their interpretations of the Lung-RADS categories, with high frequencies of disagreement in some specific scenarios.

Keywords: Questionnaires; Pulmonary nodule; Screening; Lung-RADS

INTRODUCTION

Since the National Lung Screening Trial (NLST)

Received June 26, 2016; accepted after revision September 20, 2016.

This study was supported by a grant from the National R&D Program for Cancer Control, Ministry for Health and Welfare, Republic of Korea (1520230).

Corresponding author: Jin Mo Goo, MD, PhD, Department of Radiology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea.

• Tel: (822) 2072-2624 • Fax: (822) 743-6385
• E-mail: jimgoo@plaza.snu.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

demonstrated that low-dose computed tomography (LDCT) screening reduced mortality in a high-risk population (1), many expert groups (the American Association for Thoracic Surgery, the American College of Chest Physicians, the American Society of Clinical Oncology, the Canadian Task Force on Preventive Health Care, the National Comprehensive Cancer Network [NCCN], the US Preventive Services Task Force) have announced recommendations for lung cancer screening (2-6). In the United States, the Centers for Medicare and Medicaid Services determined that the evidence was sufficient to use LDCT to screen for lung cancer (7) and also decided to choose the Lung Imaging Reporting and Data System (Lung-RADS) as the reporting framework.

The Lung-RADS is a risk-stratifying system proposed by the American College of Radiology (ACR) Lung Cancer Screening Committee subgroup for screening chest CT.

According to the ACR itself, the purpose of implementing this clinical practice guideline (CPG) is “to standardize lung cancer screening CT reporting and management recommendations, reduce confusion in lung cancer screening CT interpretations and facilitate outcome monitoring” (8). In order to reduce the false-positive rate and thus improve the cost effectiveness of lung cancer screening, the Lung-RADS criteria increase the nodule size threshold for a positive screen compared with the size that had been used by the NLST and also redefined the assessments of nodule density and growth. Indeed, a retrospective study by Pinsky et al. (9) showed that shifting from the NLST to the Lung-RADS criteria resulted in substantially fewer false-positive test results (Lung-RADS vs. NLST: 12.8% vs. 26.6% at baseline; 5.3% vs. 21.8% after baseline), at the cost of a slight decrease in sensitivity (Lung-RADS vs. NLST: 84.9% vs. 93.5% at baseline; 78.6% vs. 93.8% after baseline). A similar study by McKee et al. (10) suggested that switching from the NCCN’s CPGs in Oncology to Lung-RADS would increase the positive predictive value by a factor of 2.5 with no increase in type two errors.

However, as with any other CPGs, the Lung-RADS should go through a successful dissemination process before being introduced into medical routine, and comprehension by the medical community is a key component in that process (11); two years after its publication, it is unknown to what degree radiologists embrace the Lung-RADS. So far, the Lung-RADS is a brief document dominated by a single table that leaves much room for interpretation; a complete lexicon and atlas of the Lung-RADS have yet to be developed.

In Korea, there is a plan to run a pilot trial before a nationwide lung cancer CT screening. An important point of that trial will be to test whether the screening process can be conducted in a uniform, standardized manner. The Lung-

RADS had been chosen as the platform for reporting the CT results and making recommendations on management plans. In preparation for that pilot trial, the authors of this study and other members of the Korean Imaging Study Group for Lung Cancer (KISLC), a committee of the Korean Society of Thoracic Radiology (KSTR), set out to translate the Lung-RADS into Korean. Before we reached a conclusion, there was a debate on how to interpret and translate some Lung-RADS phrases and findings. In this process, we found that there could be lung-nodule scenarios that were difficult to classify using the Lung-RADS and that this could become a source of non-standardized reporting in lung cancer screening.

To find out more about this issue, we surveyed the members of the KSTR and assessed the uniformity of their responses to these difficult-to-classify scenarios in the Lung-RADS.

The purpose of this study was to evaluate the uniformity of chest radiologists’ decisions in assigning difficult-to-classify nodules into Lung-RADS categories.

MATERIALS AND METHODS

Ten scenarios of difficult-to-classify imaginary lung nodules were prepared by one radiologist; these were derived from Lung-RADS findings that the members of the KISLC had to discuss before reaching a conclusion. Because the categorization of baseline CT findings seemed straightforward, the scenarios consisted of findings on repeat CT only (seven growing nodules and three new nodules). The consistency of the imaginary nodule was solid in one scenario, part-solid in five scenarios, and non-solid (pure ground-glass nodule) in four scenarios. The scenarios were made into an online survey, which was then sent to 240 members of the KSTR via email, along with two documents that explained the Lung-

Table 1. Questions and Consensus Answers for 10 Imaginary-Lung-Nodule Scenarios

Number	Question	Consensus Answer
1	Growing 5 × 4 mm solid nodule at annual repeat CT	4A
2	Growing 5 × 4 mm part-solid nodule with solid-portion of 3 × 2 mm at annual repeat CT	4A
3	Growing 5 × 4 mm part-solid nodule with solid-portion of 4 × 3 mm at annual repeat CT	4B
4	New 5 × 4 mm part-solid nodule with solid-portion of 3 × 2 mm at annual repeat CT	3
5	Growing 12 × 10 mm part-solid nodule with solid-portion of 3 × 2 mm at annual repeat CT	4A
6	Growing 12 × 10 mm part-solid nodule with solid-portion of 7 × 6 mm at annual repeat CT	4B
7	Growing 9 × 8 mm non-solid nodule at annual repeat CT	2
8	New 9 × 8 mm non-solid nodule at annual repeat CT	2
9	Growing 24 × 20 mm non-solid nodule at annual repeat CT	2
10	New 24 × 20 mm non-solid nodule at annual repeat CT	3

RADS in both English (8) and Korean (Supplementary Table 1 in the online-only Data Supplement); we received 44 replies. Two respondents were observed to have submitted multiple replies under the same username within short time intervals (4 and 12 minutes), so only the second entries made by these two respondents were used for analysis. As a result, 420 answers from 42 respondents (response rate 17.5%) were available for analysis.

In the survey, respondents were instructed to create and enter a username and then answer ten multiple-choice questions (Table 1). Each question included a brief description of a lung nodule, including size, consistency (solid, part-solid, or non-solid), and interval change (new or growing); we provided no figures. Respondents were instructed to choose one answer from five choices: category

2, 3, 4A, or 4B or un-categorizable. We received responses electronically over a period of three weeks (May 9 to May 30, 2016) and sent out reminders twice during this period.

The KISLC has restructured the Lung-RADS (Fig. 1) by rearranging the components of the original ACR document according to nodule type, size, and interval change. Core members of the KISLC established answers to the ten questions (Table 1) using this new table and consensus decision.

RESULTS

Overall, 310 (73.8%) answers agreed with the consensus answers, and 90 (21.4%) did not (Table 2). Respondents did not categorize a nodule a total of 20 times, either by

Solid		
< 6 mm	Baseline	2
	No change	2
	Growing	4A
	New < 4 mm	2
	New (≥ 4 , < 6 mm)	3
≥ 6 , < 8 mm	Baseline	3
	No change	2
	Growing	4A
	New (≥ 6 , < 8 mm)	4A
≥ 8 , < 15 mm	Baseline	4A
	No change	2
	Growing	4B
	New	4B
≥ 15 mm	Baseline	4B
	No change	2
	Growing	4B
	New	4B

A

Part-solid nodule		
< 6 mm	Baseline	2
	No change	2
	Growing (solid < 4 mm)	4A
	Growing (solid ≥ 4 , < 6 mm)	4B
	New	3
≥ 6 mm with solid < 6 mm	Baseline	3
	No change	2
	Growing (solid < 4 mm)	4A
	Growing (solid ≥ 4 , < 6 mm)	4B
	New (solid < 4 mm)	4A
≥ 6 mm with solid ≥ 6 , < 8 mm	New (solid ≥ 4 , 6 mm)	4B
	Baseline	4A
	No change	2
	Growing (solid ≥ 6 , < 8 mm)	4B
≥ 6 mm with solid ≥ 8 mm	New (solid ≥ 6 , < 8 mm)	4B
	Baseline	4B
	No change	2
	Growing	4B
	New (solid ≥ 8 mm)	4B

B

Pure GGN		
< 20 mm	Baseline	2
	No change	2
	Growing	2
	New	2
≥ 20 mm	Baseline	3
	No change	2
	Growing	2
	New	3

C

Special consideration		
Endobronchial nodule		4A
Suspicion of malignancy	Additional features* on category 3 or 4	4X

D

Fig. 1. Lung-RADS categories rearranged by Korean Imaging Study Group for Lung Cancer according to nodule type, size, and interval change. Solid (A), part-solid nodule (B), pure GGN (C), and special consideration (D). *Additional features: spiculation, GGN that doubles in size in 1 year, enlarged lymph nodes etc. GGN = ground-glass nodule, Lung-RADS = Lung Imaging Reporting and Data System

choosing “un-categorizable” ($n = 15$, 3.5%) or by skipping the question ($n = 5$, 1.2%). The numbers of respondents’ answers that agreed with the consensus answers ranged from 3 to 10 (Fig. 2); the most frequent number of agreed-upon answers was 8 (11 respondents), followed by 9 (9 respondents). Eleven (26.2%) respondents submitted 6 or fewer answers that agreed with the consensus answers.

Compared with the consensus answer, respondents more frequently assigned imaginary nodules to higher categories (16.0%) than to lower ones (5.5%) (Fig. 3).

The scenarios that yielded the highest disagreement were questions 4 (new 5 × 4 mm part-solid nodule with a solid portion of 3 × 2 mm at annual repeat CT) and 9 (growing 24 × 20 mm non-solid nodule at annual repeat CT) (Fig. 4). For question 4 (consensus answer, category 3), 19 respondents assigned the scenario to category 3 and 20 respondents to

category 4A. For question 9 (consensus answer, category 2), 19 respondents assigned the scenario to category 2, 13 respondents to category 3, and 6 respondents to un-categorizable. Another frequently disagreed-upon scenario was that in question 3 (consensus answer, 4B; assigned to category 4A ten times).

DISCUSSION

This results of this survey showed that there is considerable variability among radiologists in how they assign difficult-to-classify nodules into Lung-RADS categories.

Compared with the consensus answers, respondents showed a tendency to assign the imaginary nodules to higher categories. The original purpose of implementing the

Table 2. No. of Agreement and Disagreement for Each Question

Questions	1	2	3	4	5	6	7	8	9	10	Sum
Correct answer	33	34	28	19	35	39	31	31	19	41	310
Over by 1 cat	1	4	0	20	6	3	1	9	13	1	58
Over by 2 cat	0	0	0	2	0	0	2	0	2	0	6
Over by 3 cat	0	0	0	0	0	0	1	0	2	0	3
No response	0	0	0	0	0	0	4	1	0	0	5
Uncategorizable	0	2	2	0	1	0	3	1	6	0	15
Under by 1 cat	5	1	10	1	0	0	0	0	0	0	17
Under by 2 cat	3	1	2	0	0	0	0	0	0	0	6
Under by 3 cat	0	0	0	0	0	0	0	0	0	0	0

Cat = category, Over = overestimation of malignancy risk, Under = underestimation of malignancy risk, 0 = agreement with consensus answer

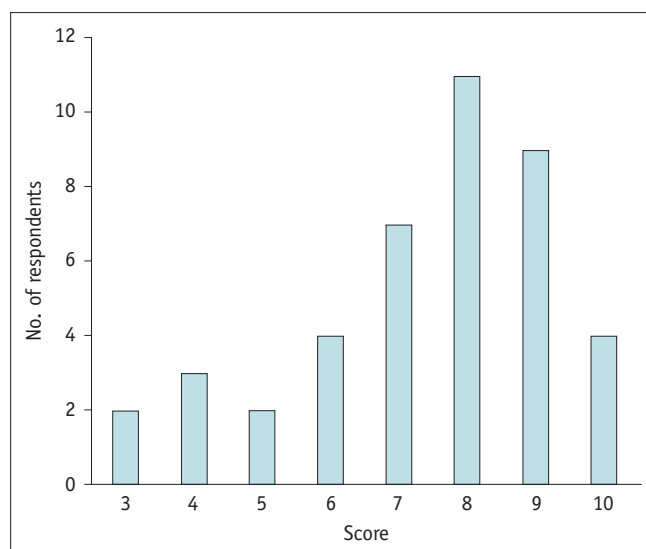


Fig. 2. Respondents' score distributions. Score = number of answers that agreed with consensus answers

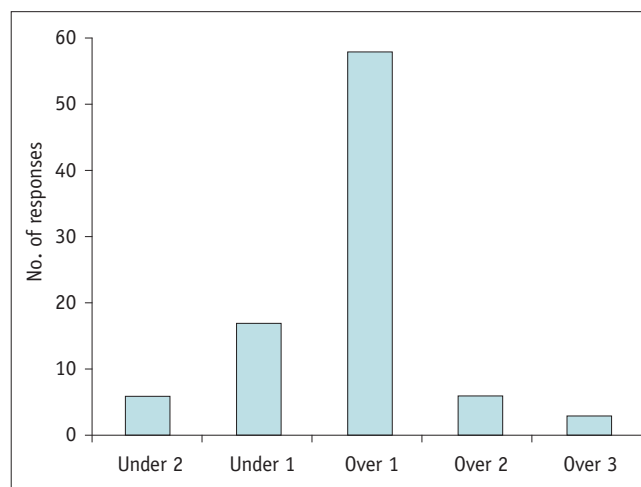


Fig. 3. Number of answers that disagreed with consensus answers. Answers were distributed according to degree of over- or under-estimation of malignancy risk: “under” indicates underestimation of risk by two Lung-RADS categories; “over 1” indicates overestimation by one category. Lung-RADS = Lung Imaging Reporting and Data System

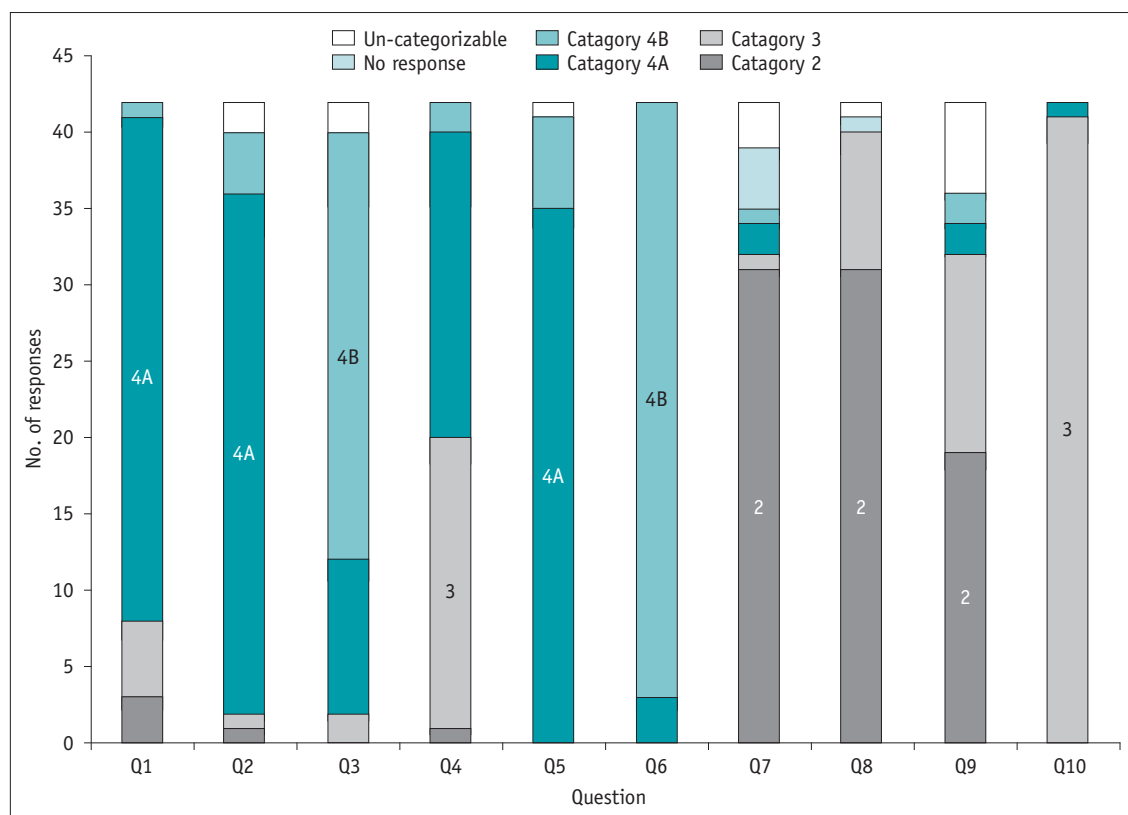


Fig. 4. Number of answers submitted for each question. White letters in graph indicate consensus answers.

Lung-RADS was to increase the cost-effectiveness of lung cancer screening and to reduce the number of unnecessary invasive procedures (8). However, the tendency of Korean radiologists to overestimate the risk of lung cancer, as seen in this survey, raises the concern that the benefits of adopting this new guideline into the pilot trial of lung cancer screening that is planned in Korea may not be as impressive as originally professed by the ACR itself or as subsequently estimated in retrospective studies (9, 10).

In this survey, we attempted to reduce potential sources of variability by providing respondents with the sizes and consistency profiles of the lung lesions so that only the degree of guideline comprehension would be measured. We also provided the respondents with Lung-RADS documents (in both Korean and English) in order to prevent the differences in guideline awareness or in English proficiency from being confounding factors. Nevertheless, the frequency of disagreement among radiologists was high enough to raise concern. The respondents in this survey belonged to an expert group, and still 26.2% of them disagreed with the consensus answers in four or more scenarios. We believe that serious efforts should be made to facilitate awareness of and standardize radiologists' interpretations of the Lung-RADS.

A number of factors may have contributed to the frequent disagreements in category assignment. The current Lung-RADS document lacks detailed explanations of some categories, for example, part-solid nodules of categories 3 and 4 and non-solid nodules; this may account for the low agreement rates for questions 3 and 4, which asked about the categories for growing or new part-solid nodules. Regarding the respondents' categorizations of part-solid nodules as category 4A ("with a new or growing < 4 mm solid component"), there can be a debate whether this describes part-solid nodules ≥ 6 mm or part-solid nodules of any size. Because new part-solid nodules < 6 mm are classified as category 3, in defining category 4A, new part-solid nodules with a solid component < 4 mm need to be interpreted as new part-solid nodules ≥ 6 mm with a solid component < 4 mm, whereas growing part-solid nodules with a solid component < 4 mm need to be interpreted as growing part-solid nodules of any size with a solid component < 4 mm. In addition, because both total size and the solid component should be measured in part-solid nodules, in defining new or growing part-solid nodules, there can be a debate on which size or component should be considered. Although there is no clear guideline on this

issue, changes in any component of part-solid nodules need to be considered. In the description of interval changes in the non-solid nodules, the term “slowly growing” was used without details, leaving room for different interpretations. The Lung-RADS defines growth as an increase in size of > 1.5 mm without mentioning which consistency of nodules this criterion concerns and then recommends categorizing a ground-glass nodule that doubles in size in one year as 4X. Thus, the guideline is not clear regarding which category a non-solid nodule should be assigned to if it has had grown more than 1.5 mm but less than double in size. In our survey, we used the term “growing” rather than “slowly growing,” and this may have affected the low agreement rate for question 9. The low agreement rate for question 3 may be related to incorrect rounding up the measurements by some respondents.

One way to improve radiologist agreement with guidelines is by offering more education opportunities, and another way is to produce and distribute documents or diagrams that elaborate on the meaning of the Lung-RADS. For example, the Lung-RADS categories can be broken down into smaller, easy-to-understand subcategories, which KISLC members have successfully done (Fig. 1). Lastly, there are software tools to assist in standardized Lung-RADS reporting. For example, the KISLC developed Lung Report, which can be run on personal computers and enables categorization as well as structured reporting in accordance with the Lung-RADS. In addition, there is a smartphone application produced by Mevis Medical Solutions AG (Mevis Lung-RADS™, Bremen, Germany). The Mevis Lung-RADS categorization is consistent with our consensus categorizations in all cases except for one scenario of a growing nonsolid nodule of 20 mm or larger. We believe that distributing such software is the most practical way to address the issue of category ambiguity within a relatively short period of time.

This study had the following limitations: 1) the number of survey questions was relatively small, and 2) we only surveyed difficult-to-classify scenarios; thus, the results likely exaggerated any Lung-RADS shortcomings. 3) The respondent pool consisted of only experts, making generalizing our results difficult.

In conclusion, chest radiologists showed large variability in how they categorized difficult-to-classify nodules into Lung-RADS categories, with very high rates of disagreement for some scenarios. Additional efforts are necessary to enhance understanding and reduce variability through

dissemination of more detailed documents and through the use of software applications to assist with category choices.

Supplementary Materials

The online-only Data Supplement is available with this article at <https://doi.org/10.3348/kjr.2017.18.2.402>.

REFERENCES

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409
2. Jaklitsch MT, Jacobson FL, Austin JH, Field JK, Jett JR, Keshavjee S, et al. The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. *J Thorac Cardiovasc Surg* 2012;144:33-38
3. Detterbeck FC, Mazzone PJ, Naidich DP, Bach PB. Screening for lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143(5 Suppl):e78S-e92S
4. Canadian Task Force on Preventive Health Care, Lewin G, Morissette K, Dickinson J, Bell N, Bacchus M, et al. Recommendations on screening for lung cancer. *CMAJ* 2016;188:425-432
5. Wood DE. National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines for Lung Cancer Screening. *Thorac Surg Clin* 2015;25:185-197
6. de Koning HJ, Meza R, Plevritis SK, ten Haaf K, Munshi VN, Jeon J, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med* 2014;160:311-320
7. Decision Memo for Screening for Lung Cancer with Low Dose Computed Tomography (LDCT) (CAG-00439N). Web site. <https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274>. Accessed June 3, 2016
8. Lung CT Screening Reporting and Data System (Lung-RADS™). Web site. <https://www.acr.org/Quality-Safety/Resources/LungRADS>. Accessed June 3, 2016
9. Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Ann Intern Med* 2015;162:485-491
10. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR Lung-RADS in a clinical CT lung screening program. *J Am Coll Radiol* 2015;12:273-276
11. Grimshaw JM, Hutchinson A. Clinical practice guidelines--do they enhance value for money in health care? *Br Med Bull* 1995;51:927-940