Radiology

ORIGINAL RESEARCH • THORACIC IMAGING

Efficacy of Ultrashort Echo Time Pulmonary MRI for Lung Nodule Detection and Lung-RADS Classification

Yoshiharu Ohno, MD, PhD • Daisuke Takenaka, MD • Takeshi Yoshikawa, MD, PhD • Masao Yui, MS • Hisanobu Koyama, MD, PhD • Kaori Yamamoto, RT • Nayu Hamabuchi, MD • Chika Shigemura, MD • Ayumi Watanabe, MD • Takahiro Ueda, MD, PhD • Hirotaka Ikeda, MD, PhD • Hidekazu Hattori, MD, PhD • Kazuhiro Murayama, MD, PhD • Hiroshi Toyama, MD, PhD

From the Department of Radiology (Y.O., N.H., C.S., A.W., T.U., H.I., H.H., H.T.) and Joint Research Laboratory of Advanced Biomedical Imaging (Y.O., K.M.), Fujita Health University School of Medicine, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake 470-1192, Japan; Division of Functional and Diagnostic Imaging Research, Department of Radiology, Kobe University Graduate School of Medicine, Kobe, Japan (Y.O., T.Y.); Department of Diagnostic Radiology, Hyogo Cancer Center, Akashi, Japan (D.T., T.Y.); Canon Medical Systems, Otawara, Japan (M.Y., K.Y.); and Department of Radiology, Osaka Police Hospital, Osaka, Japan (H.K.). Received May 21, 2021; revision requested August 9; revision received September 25; accepted October 1. **Address correspondence to** Y.O. (e-mail: *yohno@fujita-hu.ac.jp*).

Conflicts of interest are listed at the end of this article.

See also the editorial by Wielpütz in this issue.

Radiology 2022; 302:697–706 • https://doi.org/10.1148/radiol.211254 • Content codes: CH MR

Background: Pulmonary MRI with ultrashort echo time (UTE) has been compared with chest CT for nodule detection and classification. However, direct comparisons of these methods' capabilities for Lung CT Screening Reporting and Data System (Lung-RADS) evaluation remain lacking.

Purpose: To compare the capabilities of pulmonary MRI with UTE with those of standard- or low-dose thin-section CT for Lung-RADS classification.

Materials and Methods: In this prospective study, standard- and low-dose chest CT (270 mA and 60 mA, respectively) and MRI with UTE were used to examine consecutive participants enrolled between January 2017 and December 2020 who met American College of Radiology Appropriateness Criteria for lung cancer screening with low-dose CT. Probability of nodule presence was assessed for all methods with a five-point visual scoring system by two board-certified radiologists. All nodules were then evaluated in terms of their Lung-RADS classification using each method. To compare nodule detection capability of the three methods, consensus for performances was rated by using jackknife free-response receiver operating characteristic analysis, and sensitivity was compared by means of the McNemar test. In addition, weighted κ statistics were used to determine the agreement between Lung-RADS classification obtained with each method and the reference standard generated from standard-dose CT evaluated by two radiologists who were not included in the image analysis session.

Results: A total of 205 participants (mean age: 64 years ± 7 [standard deviation], 106 men) with 1073 nodules were enrolled. Figure of merit (FOM) ($P < .001$) had significant differences among three modalities (standard-dose CT: FOM = 0.91, low-dose CT: FOM = 0.89, pulmonary MRI with UTE: FOM = 0.94), with no evidence of false-positive findings in participants with all modalities ($P > .05$). Agreements for Lung-RADS classification between all modalities and the reference standard were almost perfect (standard-dose CT: κ = 0.82, $P < .001$; low-dose CT: κ = 0.82, $P < .001$; pulmonary MRI with UTE: κ = 0.82, $P < .001$).

Conclusion: In a lung cancer screening population, ultrashort echo time pulmonary MRI was comparable to standard- or low-dose CT for Lung CT Screening Reporting and Data System classification.

Online supplemental material is available for this article.

Lung cancer is the leading cause of cancer death for both men and women worldwide. The National Lung Screening Trial found that pulmonary nodules are a common finding at chest radiography and CT not only in routine clinical practice but also in lung cancer screenings with low-dose CT (1). Moreover, several additional trials addressing the same research question regarding the frequency of pulmonary nodules and verifying the findings have been performed worldwide (1–5). The Fleischner Society recommendations and the findings reported by the Lung Cancer Screening Committee for the American College of Radiology Lung CT Screening Reporting and Data System (Lung-RADS) suggest that nodule type and size assessments are effective markers for nodule management to reduce lung cancer–specific mortality (6–13). During the first 5 years of nationwide lung cancer screening, there has been a substantial accumulation of data and experience with many opportunities for continued learning. In 2019, the American College of Radiology released an update of Lung-RADS, version 1.1 (11–13). While many similarities exist between the old and new versions, multiple changes found in version 1.1 were made to adapt to newly acquired evidence in the field and to render the system more suitable for the current lung cancer screening environment.

As a result, thin-section CT examinations with low- and standard-dose protocols are currently widely used for lung nodule detection and assessment even in routine clinical practice. Starting from the early 1990s, however, MRI has been used on a limited basis for different lung, mediastinal, and pleural diseases (14–18). In the 2000s, several techniques based on spin-echo, gradient-echo, and echo-planar imaging sequences with and without

Abbreviations

FOM = figure of merit, JAFROC = jackknife free-response receiver operating characteristic, Lung-RADS = Lung CT Screening Reporting and Data System, UTE = ultrashort echo time

Summary

Pulmonary MRI with ultrashort echo time was comparable to standard- or low-dose CT for Lung CT Screening Reporting and Data System classification in a screening population.

Key Results

- In this prospective study of 205 lung cancer screening participants with 1073 nodules, figure of merit (FOM) for nodule detection ($P < .001$) had significant differences among three modalities (standard-dose CT: FOM = 0.91, low-dose CT: FOM = 0.89, pulmonary MRI with ultrashort echo time [UTE]: FOM = 0.94).
- Standard-dose (87.1%) and low-dose (87.0%) CT had higher accuracy than pulmonary MRI with UTE (83.1%, $P < .001$).
- Lung CT Screening Reporting and Data System classification agreements using all modalities with the reference standard were excellent (standard-dose CT: $\kappa = 0.82$, $P < .001$; low-dose CT: $\kappa = 0.82$, $P < .001$; pulmonary MRI with UTE: $\kappa = 0.82$, $P < .001$).

contrast media have been tested for their nodule detection capability as compared with CT. More recently, suggestions have been made for continuous improvements in image quality and detection capabilities (19–31).

As one of these improvements, during the last several years, pulmonary thin-section MRI with ultrashort echo time (UTE) or zero echo time has been tested for nodule detection or nodule subtype classification and for various radiologic findings evaluations for comparison with standard- or low-dose CT (26–34). This procedure has the potential for better nodule visualization for per-segment and patient-based evaluations comparable to that of standard- or low-dose thin-section CT (27,28). Therefore, the Fleischner Society and other international bodies have endorsed this technique as a promising tool for thoracic MRI in routine clinical practice (26–34). However, to our knowledge, direct comparison among pulmonary MRI with UTE, low-dose CT, and standard-dose CT for lung cancer screening based on Lung-RADS evaluation has not previously been well established in the literature.

We hypothesized that pulmonary MRI with UTE may have similar efficacy as standard- or low-dose CT for nodule detection, measurement, and Lung-RADS evaluation. If shown, pulmonary MRI with UTE could function as a screening tool for lung cancer. The purpose of this study was to compare the ability of these three methods for Lung-RADS classification, nodule measurement, and detection in a lung cancer screening population.

Materials and Methods

This multicenter and prospective study was approved by the institutional review boards of Kobe University Hospital and Fujita Health University Hospital. Written informed consent was obtained from all participants. This study was financially and technically supported by Canon Medical Systems. Two of the authors (M.Y. and K.Y.) are employees of Canon Medical Systems but did not have control over any of the data and information submitted for publication or over which data and information were to be included in this study. All data were controlled by D.T., H.K., H.I., and H.H. and analyzed by N.H., C.S., A.W., and T.U. in this study.

Study Participants

Between January 2017 and December 2020, 272 consecutive participants were enrolled. They underwent CT lung cancer screening at our hospitals based on *(a)* American College of Radiology Appropriateness Criteria for lung cancer screening with low-dose CT (35), *(b)* Japanese Society of CT Screening recommendations, *(c)* suspicion of harboring lung nodules because of chest radiographs obtained at nearby hospitals or through Japanese chest radiography lung cancer screenings at one of the outpatient clinics at our hospitals, or *(d)* as part of follow-up CT performed at our outpatient clinics after detection of lung cancer at Japanese chest radiography or low-dose CT lung cancer screening. All participants who had agreed to provide us with their standard- or low-dose CT data after the purpose and protocol of the study had been explained to them and their written informed consent had been obtained underwent standard- or low-dose thin-section CT and pulmonary thin-section MRI with UTE within 1 week (mean, 2 days ± 1 [standard deviation]) of giving their consent. With consideration to this study's inclusion criteria based on the American College of Radiology Appropriateness Criteria for lung cancer screening with low-dose CT (35) and previous articles (27,28,30), the exclusion criteria were as follows: *(a)* adults younger than 50 years or older than 80 years, *(b)* adults of any age with less than 20 packs per year smoking history or who had quit more than 15 years, *(c)* pregnancy or breast feeding, *(d)* contraindication for MRI (pacemaker, ferromagnetic implants, etc), *(e)* no evidence of pulmonary nodules or pulmonary nodules less than 4 mm in diameter or presence of lung mass, and *(f)* complete calcified nodules. Criterion *e* led to the exclusion of 21 participants, criterion *f* to exclusion of 15 participants based on standard-dose thin-section CT findings, criterion *d* to exclusion of 11 participants who had a pacemaker ($n = 7$) and ferromagnetic implants ($n = 4$), criterion *a* to exclusion of eight participants, criterion *b* to exclusion of seven participants, and criterion *c* to exclusion of five smokers who were pregnant or breastfeeding ($n = 3$ and $n = 2$, respectively). None of the patients had been enrolled in any previously presented or published studies, including the data published by us in 2016 and 2017 (26,27) (Fig 1).

Radiologic Examination

All patients underwent thin-section chest CT with tube currents of 60 and 250 mA for low- and standard-dose CT, respectively, on a 320–detector row CT scanner (Aquilion ONE, Canon Medical Systems), an 80–detector row CT scanner (Aquilion Prime; Canon Medical Systems), and a 160–detector row CT scanner (Aquilion Precision. Canon Medical Systems). Moreover, pulmonary thin-section MRI with UTE was performed by means of a respiratory-gated three-dimensional radial UTE pulse sequence (repetition time msec/echo time μsec, 4.0/96–112; flip angle, 5°; voxel size, $1 \times 1 \times 1$ mm^3) in the coronal plane with three 3.0-T MRI systems (Vantage Titan 3T and Vantage

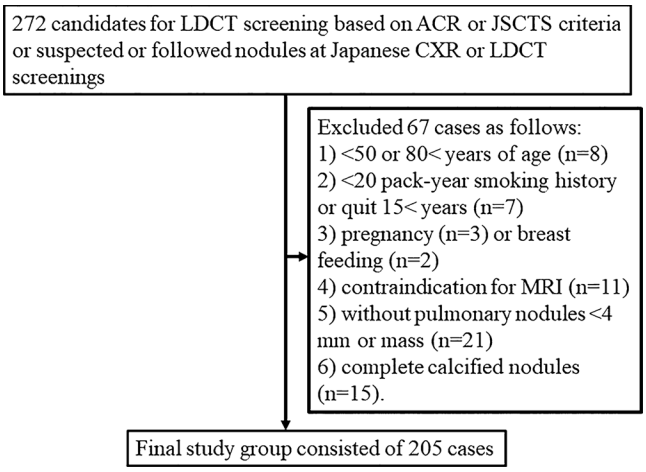

Figure 1: Flowchart of 272 consecutive patients who underwent CT lung cancer screening at our hospitals based on American College of Radiology (ACR) Appropriateness Criteria for lung cancer screening with low-dose CT (LDCT) or Japanese Society of CT Screening (JSCTS) recommendation or who were suspected of having or discovered to have lung nodules following Japanese chest radiography (CXR) or low-dose CT screenings. Sixty-seven patients were excluded due to exclusion criteria. Finally, this study consisted of 205 patients.

Galan 3T/ZGO, Canon Medical Systems) using 16-element phased-array body surface coil and receiver channels combined with parallel imaging capability (Atlas SPEEDER, Canon Medical Systems). Then, all pulmonary MRI scans with UTE were reconstructed with 1-mm section thickness in the axial and coronal planes. Details of MRI technique applied in this study had been reported in the past literature (27,28). Moreover, details of CT and MRI scanning protocols in this study are also stated in Appendix E1 (online).

Reference Standard

To evaluate the efficacy of pulmonary MRI with UTE, thin-section standard-dose CT images for each patient were reviewed by two board-certified radiologists with more than 15 years of experience in thoracic radiology (D.T.) or general radiology (K.M.) on commercially available picture archiving and communication systems (ShadeQuest [Fujifilm Medical Solutions] and RapideyeCore [Canon Medical Systems]) at lung window setting (level, −550 HU; width, 1600 HU) in consensus. These two radiologists were not included in any image analyses in this study. Final evaluation for each patient was determined by consensus of the two reviewers. Location, size, and nodule type of all confirmed nodules were recorded. In addition, nodules of each type were classified into three groups: ground-glass nodules, part-solid nodules, and solid nodules (7–10,26,27). These nodules were also classified into Lung-RADS version 1.1 categories 2, 3, 4A, 4B, and 4X (9,11–13). In addition, the longest diameters of total nodule and of solid components within the nodule were independently measured by the same radiologists, and the two final diameters were used to determine the average values obtained from all investigators' measurements. The results were then used as the reference standard for nodule detection, Lung-RADS classification, and accuracy of nodule evaluation.

Image Analysis

To compare the capabilities of thin-section standard- or low-dose CT and pulmonary MRI with UTE for nodule detection and nodule type evaluation, all images were reviewed on the same picture archiving and communication system. Next, images obtained with the three procedures were randomized, and each method was independently evaluated by a board-certified chest radiologist (Y.O.) with 27 years of experience and a board-certified abdominal radiologist with 26 years of experience and more than 10 years of experience with thoracic MRI (T.Y.). Neither of the reviewers had any access to information about the results obtained with other methods. Both reviewers reviewed all thin-section CT images at lung window setting (level, −550 HU; width, 1600 HU) without access to information about the technical parameters. Moreover, neither investigator reviewed CT images with normal section thickness or any other CT images of the participants. All evaluations of CT and MRI scans by the radiologists were performed at different times, days, and reading rooms. The interval between reading CT and MRI scans was more than 1 month.

To compare nodule detection capability of pulmonary MRI with UTE with that of the two types of CT examinations, a five-point visual scoring system was adopted (1, absent; 2, probably absent; 3, equivocal; 4, probably present; 5, present) for a per-nodule basis analysis. The locations of all nodules detected by each investigator were also recorded.

For all nodules, the longest diameters of the total nodule and of solid components within the nodule obtained with each method were measured by each investigator twice, and the final values for each investigator were used to determine the average values for first and second measurements. The final values of the long-axis diameter and of solid components within each nodule obtained with all methods were then used to determine average values for the two investigators.

Lung-RADS classification was independently performed for all detected nodules confirmed with each method by each investigator. In addition, nodules confirmed as reference standard and undetected with any method were established as Lung-RADS category 1.1 and used for assessment. Final Lung-RADS classifications of each nodule for all methods were determined by consensus of the two investigators.

Statistical Analysis

Nodule detection.—To compare nodule detection capabilities of the three methods, each radiologist's assessments of lung nodule detection with the three methods were compared by means of jackknife free-response receiver operating characteristic (JAFROC) analysis (27,35). The same analysis was also used to compare nodule detection performance of the three modalities based on consensus computationally generated by software.

With JAFROC analysis, the figure of merit (FOM), which indicates the probability that a lesion is rated higher than the highest rated nonlesion in a healthy participant, is first computed for each observer, after which a matrix of FOM pseudo values computed with the jackknife statistical technique is analyzed by using analysis of variance. In addition, we assessed the

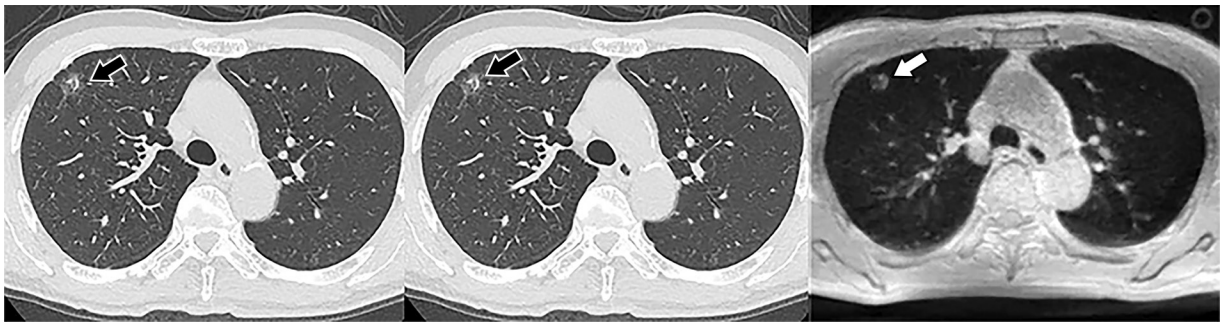

Figure 2: Images in a 63-year-old man with ground-glass nodule (arrow) who was diagnosed with minimally invasive adenocarcinoma. Unenhanced standard-dose (left) and low-dose (middle) CT scans in axial plane clearly show a ground-glass nodule with a long-axis diameter of 13 mm in the right upper lobe. In addition, unenhanced pulmonary MRI scan with ultrashort echo time (repetition time msec/echo time μsec, 4.0/96; flip angle, 5°) (right) demonstrates the presence of a ground-glass nodule with a long-axis diameter of 9 mm. Both reviewers evaluated this nodule as ground-glass nodule and assessed it as Lung CT Screening Reporting and Data System category 2.

differences among the three modalities in terms of FOM values, sensitivity, and false-positive rate by means of one-way analysis of variance. Sensitivity and false-positive rate (per data set) for consensus assessment were also compared among the three methods by using the McNemar test or the signed rank test (27,35). Statistical results for nodule detection capability showed that nodules detected with each modality, assigned a score of at least 3 by both investigators and confirmed by the reference standard, could be assumed to be true-positive findings, while nodules detected with each modality and assigned a score of at least 3 by both investigators but rejected by the reference standard could be assumed to be false-positive findings.

Nodule type evaluation and measurements of longest axis diameters of total nodule and of solid components within nodule.—To determine the interobserver agreement for longest axis diameter measurements of total nodule and of solid components within nodule, correlations of measurements by the two investigators of each longest axis diameter were evaluated by using Pearson correlation analysis. In addition, errors in measurements of each long-axis diameter by the two investigators were assessed with Bland-Altman analysis (36). In addition, the average long-axis diameters of total nodule and solid components in each nodule were also compared among all modalities with Bland-Altman analysis (36).

To compare interobserver agreement with each modality and agreement for nodule type evaluation between each consensus reading result and reference standard, weighted κ statistics were used for evaluations by the two investigators for each method and for consensus reading results for all methods. Agreement was considered poor for κ less than 0.21, fair for κ of 0.21–0.40, moderate for κ of 0.41–0.60, substantial for κ of 0.61–0.80, and excellent for κ of 0.81–1.00 (37). In addition, accuracy of nodule type evaluation was compared between each modality by consensus reading results and the reference standard by means of the McNemar test.

Lung-RADS evaluation capability.—To determine the effect on Lung-RADS evaluation for all three modalities, weighted κ analysis and the χ^2 test were used to determine interobserver agreement between the two investigators for nodule detection using each modality. In addition, agreement on Lung-RADS classification using each modality was determined by the consensus between the two investigators, the reference standard was evaluated using the same statistical analyses, and the intermethod agreement between each of the two modalities was determined by consensus reading for Lung-RADS classification. Interobserver agreements for each modality, agreements between each modality and the reference standard, and intermethod agreements were considered poor for κ less than 0.21, fair for κ of 0.21–0.40, moderate for κ of 0.41–0.60, substantial for κ of 0.61–0.80, and excellent for κ of 0.81–1.00 (37,38).

Except for JAFROC analysis, all statistical analyses were performed with JMP version 14 (SAS Institute Japan). $P < .05$ was considered indicative of a significant difference for all statistical analyses. All statistical analyses were performed by four radiologists who were familiar to statistics (N.H., C.S., A.W., T.U.). For justification of sample size, we performed statistical power analysis and effect size calculation for all statistical analyses by using EZR on R Commander software (version 1.37; Saitama Medical Center, Jichi Medical University).

Results

Participant and Nodule Characteristics

Representative images for three study participants are shown in Figures 2–4. Eventually, this study group consisted of 205 participants (mean age, 64 years ± 7 [standard deviation]; 106 men) who were evaluated. The presence of a total of 1073 nodules was confirmed by consensus of the same radiologists. On the basis of the standard-dose CT findings, all nodules were divided into solid nodules (n = 759), part-solid nodules (n = 132), and ground-glass nodules (n = 182). Each nodule was also classified in terms of Lung-RADS as category 2 (n = 706), category 3 (n = 219), category 4A (n = 58), category 4B (n = 75), and category 4X (n = 15).

Details of patient characteristics are shown in Table 1.

Nodule Detection Capability Assessment

Results of a comparison of nodule detection efficacy of the three modalities are shown in Table 2. Results of JAFROC analysis for

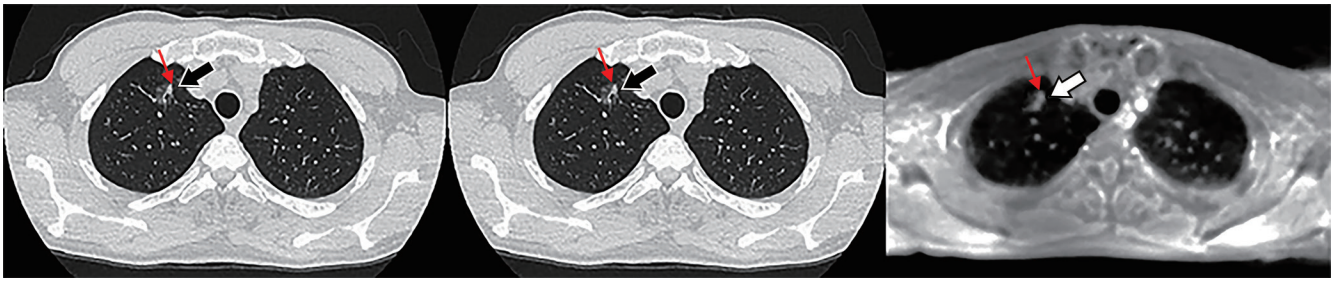

Figure 3: Images in a 65-year-old man with a part-solid nodule who was diagnosed with invasive adenocarcinoma. Unenhanced standard-dose (left) and low-dose (middle) CT scans in axial plane clearly show part-solid nodule with a long-axis diameter of 11 mm (black arrow) and a solid component with a diameter of 3 mm (red arrow) in the right upper lobe. In addition, the unenhanced pulmonary MRI scan with ultrashort echo time (repetition time msec/echo time μsec, 4.0/112; flip angle, 5°) demonstrates the presence of a part-solid nodule with a long-axis diameter of 9 mm (white arrow) and a solid component with a diameter of 5 mm (red arrow). Both reviewers evaluated this nodule as part-solid nodule and assessed it as Lung CT Screening Reporting and Data System category 3.

Figure 4: Images in a 67-year-old man with a solid nodule (arrow) who was diagnosed with adenocarcinoma as well as other subtypes. Unenhanced standard-dose (left) and low-dose (middle) CT scans and pulmonary MRI scan with ultrashort echo time (repetition time msec/echo time μsec, 4.0/96; flip angle, 5°) clearly show a solid nodule with a long-axis diameter of 15 mm in the right lower lobe. Both reviewers evaluated this nodule as solid nodule and assessed it as Lung CT Screening Reporting and Data System category 4B.

each investigator and consensus reading are shown in Figure 5. Investigators 1 and 2 showed differences in FOM ($P < .001$) among standard-dose CT (investigator 1: FOM = 0.91, investigator 2: FOM = 0.91), low-dose CT (investigator 1: FOM = 0.89, investigator 2: FOM = 0.89), and pulmonary MRI with UTE (investigator 1: FOM = 0.94, investigator 2: FOM = 0.94). For consensus evaluation, there were differences in FOM ($P < .001$) among the three modalities (standard-dose CT: FOM = 0.91, low-dose CT: FOM = 0.89, pulmonary MRI with UTE: FOM = 0.94).

Assessment of sensitivity for the two investigators and consensus evaluation showed that sensitivity of pulmonary MRI with UTE (investigator 1: sensitivity = 87.7% [95% CI: 85.6, 89.6]; investigator 2: sensitivity = 88.1% [95% CI: 86.0, 90.0]; consensus evaluation: sensitivity = 87.9% [95% CI: 85.8, 89.8]) was higher than that of standard-dose CT (investigator 1: sensitivity = 87.1% [95% CI: 85.0, 89.1], $P = .03$; investigator 2: sensitivity = 87.1% [95% CI: 85.0, 89.1], $P = .002$; consensus evaluation: sensitivity = 87.1% [95% CI: 85.0, 89.1], $P = .008$) and of low-dose CT (investigator 1: sensitivity = 87.0% [95% CI: 84.8, 88.9], $P = .008$; investigator 2: sensitivity = 87.1% [95% CI: 84.9, 89.0], $P = .001$; consensus evaluation: sensitivity = 87.1% [95% CI: 84.9, 89.0], $P = .004$). We found no evidence of differences in false-positive rate per participant between the two investigators or in consensus evaluation of all modalities ($P > .05$).

Nodule Type Evaluation and Measurement of Longest Axis Diameters of Total Nodules and of Solid Components within Nodule

Results of correlation and the limits of agreement between the two investigators for long-axis diameter measurement for total nodule and solid components within the nodule are shown in Table E3 (online). For each method, correlation coefficients for total nodules ($r = 0.98$, $P < .001$) and solid components within nodules ($r = 0.98$, $P < .001$) between the two investigators were excellent. The mean limits of agreement for the long-axis diameter of total nodules using each modality were determined as follows: standard-dose CT, 0.0 mm ± 2.0; low-dose CT, 0.0 mm ± 2.1; and pulmonary MRI with UTE, 0.0 mm ± 2.2. The mean limits of agreement for the long-axis diameter of solid components within nodules using each modality were determined as follows: standard-dose CT, 0.0 mm ± 1.8; low-dose CT, 0.0 mm ± 1.8; pulmonary MRI with UTE, 0.0 mm ± 2.0. Results of Bland-Altman analysis for long-axis diameter of total nodules and solid components are shown in Figure 6. When the long-axis diameter of total nodules was assessed, the mean limits of agreements were determined as follows: standard-dose CT versus low-dose CT, 0.0 mm ± 3.1; pulmonary MRI with UTE versus others, 0.3 mm ± 2.7. However, when the long-axis diameter of solid components was assessed, the mean limits of agreement between each of two modalities were the same and determined as 0.0 mm ± 2.0.

Table 1: Participant Characteristics

Characteristic	Value
Sex	
No. of men	106/205 (51.7)
No. of women	99/205 (48.3)
Age (y)*	
Men	64 ± 7 (50–80)
Women	65 ± 8 (50–80)
Mean smoking history (pack-years)*	34 ± 11 (20–124)
Nodule type and diameter	
GGN	
No. of nodules	182/1073 (17.0)
Nodule size (mm)*	9.0 ± 4.1 (4–24)
PSN	
No. of nodules	132/1073 (12.3)
Nodule size (mm)*	13.9 ± 7.3 (4–29)
SN	
No. of nodules	759/1073 (70.7)
Nodule size (mm)*	6.4 ± 4.4 (4–29)
Lung-RADS 1.1 category	
2	
No. of nodules	706/1073 (65.8)
Nodule size (mm)*	5.8 ± 2.8 (4–24)
3	
No. of nodules	219/1073 (20.4)
Nodule size (mm)*	7.4 ± 2.2 (4–15)
4A	
No. of nodules	58/1073 (5.4)
Nodule size (mm)*	12.1 ± 2.3 (8–18)
4B	
No. of nodules	75/1073 (7.0)
Nodule size (mm)*	23.7 ± 4.0 (16–29)
4X	
No. of nodules	15/1073 (1.4)
Nodule size (mm)*	13.1 ± 3.1 (7–18)

Note.—Except where indicated, data are numbers of participants or nodules, with percentages in parentheses. GGN = ground-glass nodule, Lung-RADS = Lung CT Screening Reporting and Data System, PSN = part-solid nodule, SN = solid nodule.

* Data are means ± standard deviations, with ranges in parentheses.

Interobserver agreement for nodule type evaluations for each modality is also shown in Table E3 (online). Interobserver agreements for all modalities were almost perfect (standard-dose CT: $\kappa = 1.00$, $P < .001$; low-dose CT: $\kappa = 0.99$, $P < .0001$; pulmonary MRI with UTE: $\kappa = 0.97$, $P < .001$).

Results of agreements between each consensus evaluation and the corresponding reference standard are shown in Table 3. All agreements with reference standards were determined as significant and substantial (standard-dose CT: $\kappa = 0.63$, $P < .001$; low-dose CT: $\kappa = 0.63$, $P < .001$) or moderate (pulmonary MRI with UTE: $\kappa = 0.58$, $P < .001$). Agreements for accuracy of nodule type evaluation for standard-dose CT (87.1% [935 of 1073]) and low-dose CT (87.0% [934 of 1073]) were higher than that of pulmonary MRI with UTE (83.1% [892 of 1073], $P < .001$).

Lung-RADS Evaluation

Interobserver agreements between the two investigators for Lung-RADS classification using each modality are shown in Table E4 (online). Interobserver agreements for all modalities were almost perfect (standard-dose CT: $\kappa = 0.98$, $P < .001$; low-dose CT: $\kappa = 0.98$, $P < .001$; pulmonary MRI with UTE: $\kappa = 0.96$, $P < .001$).

Agreements for Lung-RADS classification of all nodules based on consensus evaluation results for each modality are shown in Table 4. Agreements for Lung-RADS classification using all modalities were almost perfect (standard-dose CT: $\kappa = 0.82$, $P < .001$; low-dose CT: $\kappa = 0.82$, $P < .001$; pulmonary MRI with UTE: $\kappa = 0.82$, $P < .001$). We found no evidence of differences in Lung-RADS classification accuracy among all three modalities (standard-dose CT: 81.9% [879 of 1073], low-dose CT: 81.7% [877 of 1073], pulmonary MRI with UTE: 81.7% [877 of 1073]; $P > .05$). Moreover, fewer Lung-RADS category 4A and more Lung-RADS category 3 nodules were determined by using pulmonary MRI with UTE as compared with standard- and low-dose CT.

Intermethod agreements for Lung-RADS classification between each of two modalities determined by consensus reading are shown in Table E5 (online). All intermethod agreements were almost perfect (standard-dose CT vs low-dose CT: $\kappa = 0.97$, $P < .001$; low-dose CT vs pulmonary MRI with UTE: $\kappa = 0.92$, $P < .001$; pulmonary MRI with UTE vs standard-dose CT: $\kappa = 0.92$, $P < .001$).

Discussion

A comparison of nodule detection capability showed that figures of merit (FOMs) for each investigator and consensus evaluation revealed significant differences among the three modalities. In 205 participants (mean age: 64 years ± 7, 106 men) with 1073 nodules, we found that the FOM for consensus reading ($P < .001$) had significant differences among the three modalities (standard-dose CT: FOM = 0.91, low-dose CT: FOM = 0.89, pulmonary MRI with ultrashort echo time [UTE]: FOM = 0.94) with no evidence of false-positive findings per participant in consensus evaluation of all modalities ($P > .05$). Agreements for Lung CT Screening Reporting and Data System classification between all modalities and the reference standard were almost perfect (standard-dose CT: $\kappa = 0.82$, $P < .001$; low-dose CT: $\kappa = 0.82$, $P < .001$; pulmonary MRI with UTE: $\kappa = 0.82$, $P < .001$). The three modalities produced no significant differences in detection of false-positive nodules. Assessment of quantitatively determined variances in measurements of total nodules and solid components within the nodule and qualitatively assessed interobserver agreements for nodule type evaluation showed that correlations and limits of agreements for the two measurements obtained with each modality were significant, excellent, and small enough for use in clinical settings. In addition, interobserver and intermethod agreements for nodule type evaluation with each method were determined to be significant.

Our data show that pulmonary MRI with UTE has similar performance to that of CT in the detection of pulmonary nodules. This method is complementary to CT because less of the lung structure is depicted (27,28,30). These results are

Table 2: Comparison of Nodule Detection Capability of Standard-Dose CT, Low-Dose CT, and Pulmonary MRI with UTE

Reader and Modality	FOM	*P* Value (vs MRI with UTE)	Sensitivity (%)	*P* Value (vs MRI with UTE)	False-Positive Nodule per Participant	*P* Value
Investigator 1						
Standard-dose CT	0.91 (0.88, 0.93)	<.001	87.1 (85.0, 89.1) [935/1073]	.03	0.28	.86 vs low-dose CT, .52 vs MRI with UTE
Low-dose CT	0.89 (0.86, 0.91)	<.001	87.0 (84.8, 88.9) [933/1073]	.008	0.3	.33 vs MRI with UTE
Pulmonary MRI with UTE	0.94 (0.91, 0.96)	NA	87.7 (85.6, 89.6) [941/1073]	NA	0.17	NA
Investigator 2						
Standard-dose CT	0.91 (0.88, 0.93)	<.001	87.1 (85.0, 89.1) [935/1073]	.002	0.3	.98 vs low-dose CT, .45 vs MRI with UTE
Low-dose CT	0.89 (0.86, 0.91)	<.001	87.1 (84.9, 89.0) [934/1073]	.001	0.3	.45 vs MRI with UTE
Pulmonary MRI with UTE	0.94 (0.91, 0.96)	NA	88.1 (86.0, 90.0) [945/1073]	NA	0.19	NA
Consensus reading						
Standard-dose CT	0.91 (0.88, 0.93)	<.001	87.1 (85.0, 89.1) [935/1073]	.008	0.25	.80 vs low-dose CT, .48 vs MRI with UTE
Low-dose CT	0.89 (0.86, 0.91)	<.001	87.1 (84.9, 89.0) [934/1073]	.004	0.27	.41 vs MRI with UTE
Pulmonary MRI with UTE	0.94 (0.91, 0.96)	NA	87.9 (85.8, 89.8) [943/1073]	NA	0.18	NA

Note.—Numbers in parentheses are 95% CIs, and numbers in brackets are numbers of nodules. FOM = figure of merit, NA = not applicable, UTE = ultrashort echo time.

compatible and reproducible with those of previously published studies (26–31).

However, nodule type evaluation performed using pulmonary MRI with UTE was less accurate than that using standard- and low-dose CT. For clinical practice purposes, nodule size measurements obtained with pulmonary MRI with UTE were almost the same as those obtained with standard- and low-dose CT. Therefore, the significant difference in accuracy may be due to differences in respiratory conditions for CT and MRI. Because volume loss of small ground-glass nodules at end-tidal volume on MRI scans may have increased proton density within nodules as compared with that at the end of full inspiration on each CT scan. Therefore, small ground-glass nodules with a long-axis diameter of less than 6 mm detected on both CT scans were identified as solid nodules or part-solid nodules on MRI scans in our study. Therefore, nodule types may be considered as having the reflection by respiratory condition difference between MRI scans and each CT scan.

In contrast to nodule type evaluation, agreement and accuracy of Lung-RADS classifications for each modality were the same or were

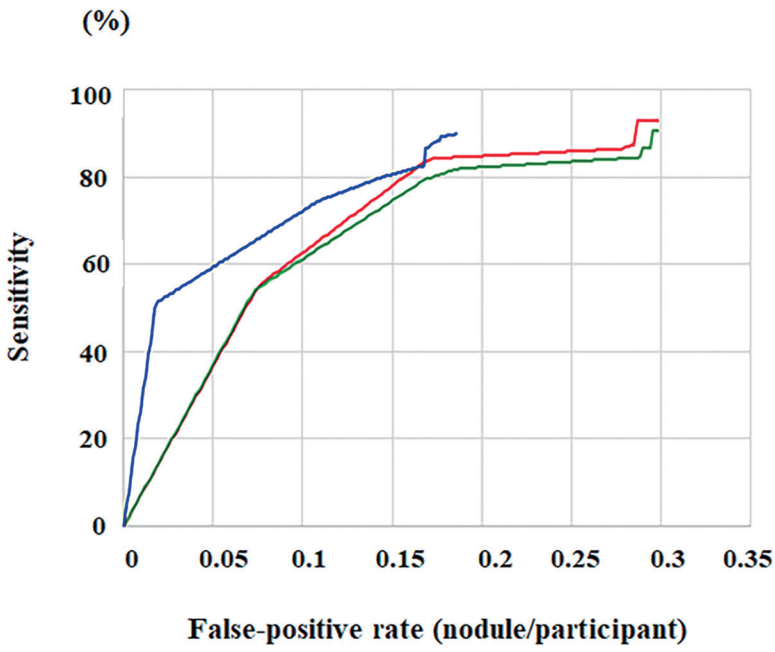

Figure 5: Graph shows jackknife free-response receiver operating characteristic curves for comparison of the three modalities by means of the computationally generated consensus performance from two investigators. There were significant differences in figure of merit (FOM) ($P < .001$) among the three modalities (standard-dose CT [red line]: FOM = 0.91, low-dose CT [green line]: FOM = 0.89, pulmonary MRI with ultrashort echo time [blue line]: FOM = 0.94).

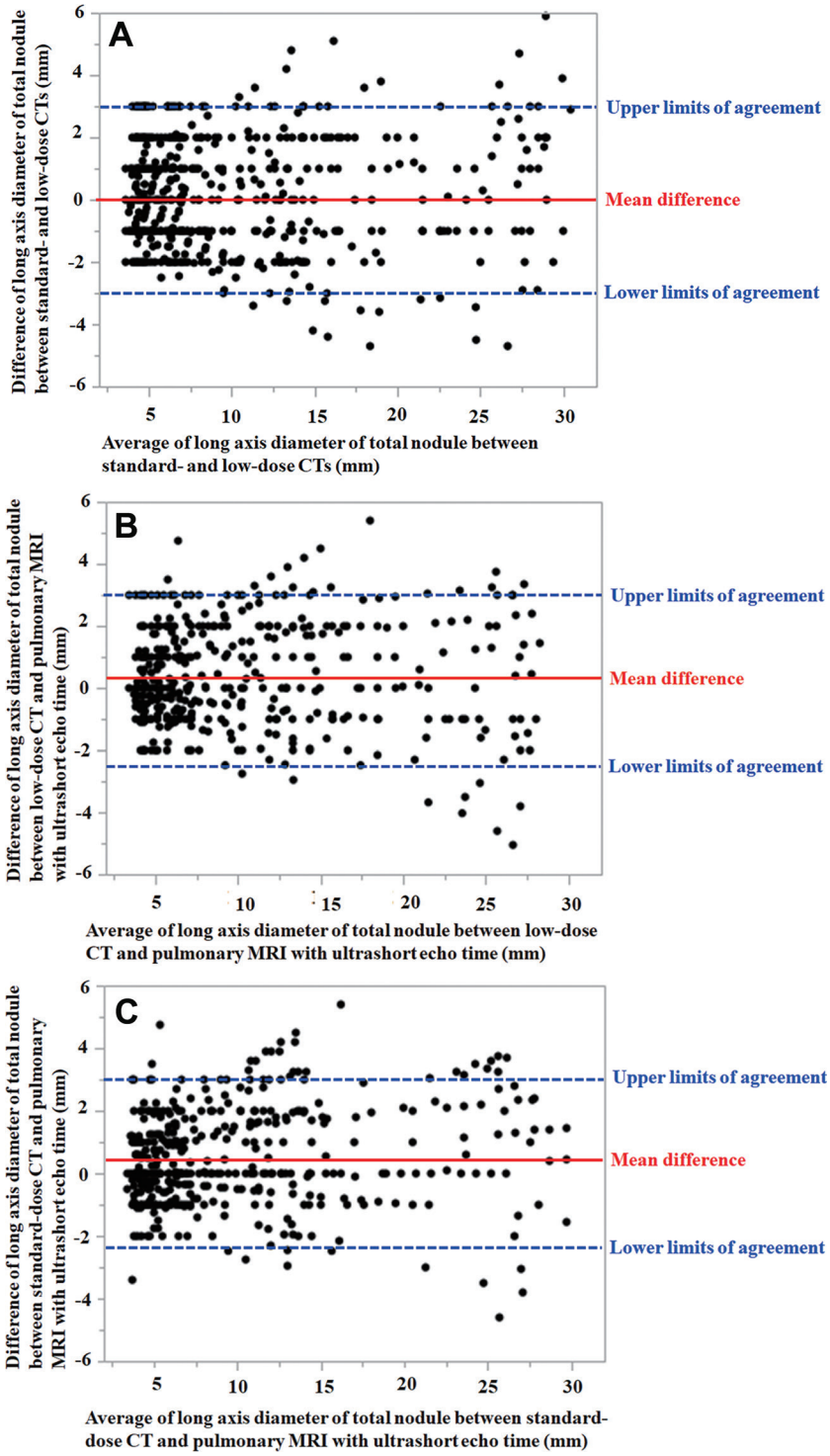

Figure 6: Bland-Altman plots show nodule measurement results for the long-axis nodule diameter. **(A)** Mean difference between standard- and low-dose CT was 0.0 mm. The upper and lower limits of agreement were 3.1 mm and −3.1 mm, respectively. **(B)** Mean difference between low-dose CT and pulmonary MRI with ultrashort echo time was 0.3 mm. The upper and lower limits of agreement were 3.0 mm and −2.4 mm, respectively. **(C)** Mean difference between standard-dose CT and pulmonary MRI with ultrashort echo time was 0.3 mm. The upper and lower limits of agreement were 3.0 mm and −2.4 mm, respectively.

not significantly different among all modalities. Moreover, fewer Lung-RADS category 4A and more Lung-RADS category 3 nodules were determined with pulmonary MRI with UTE than with standard- and low-dose CT. This may result because of differences in the respiratory level for each radiologic examination. In addition, each interobserver agreement for Lung-RADS classification was significant and almost perfect in our study. Therefore, the previously mentioned factors affecting nodule type evaluation can be considered to have only minor effects on Lung-RADS classification in routine clinical practice. In view of the aforementioned results, the potential of pulmonary MRI with UTE for assessing Lung-RADS classification is comparable to that of standard- and low-dose CT.

Our study has limitations. First, the participants included in this study were screened based on American College of Radiology Appropriateness Criteria for lung cancer screening with low-dose CT (35) or Japanese chest radiography lung cancer screenings criteria or were undergoing low-dose CT screenings for follow-up of suspected nodules. In addition, the influences of nodule size in the nonselected group of patients or lesion size and type for staging lung cancer were not compared among pulmonary MRI with UTE and standard- and low-dose CT in this study. Second, the number and size of each nodule type differed, so that the distribution of the visual score for each nodule was likely to be severely skewed. In addition, all missed nodules with each modality by the two investigators were ground-glass nodules or solid nodules whose long-axis diameters were determined as less than 4 mm; however, the reference standard was image based and determined with the consensus of reviewers who were not included in the image analysis section of this study and by pathologic results. Therefore, these nodules were one of the biases in this study. Third, a five-point visual assessment for nodule detection was used even though a binary process (yes or no) is used in routine clinical practice. In addition, this study did not evaluate the influence of emphysema, pleural fluid, or respiratory rate on nodule detection. We will therefore use binary process evaluation for nodule detection; evaluate the influence of emphysema, pleural fluid, or respiratory rate; and compare nodule detection capabilities of all three modalities in future studies.

In conclusion, these results suggest that pulmonary MRI with ultrashort echo time is similar to that of standard- or low-dose CT for nodule detection, measurement, and Lung CT Screening Reporting and Data System classification in a lung cancer screening population.

Table 3: Agreements between Consensus Reading Results for Each Modality and Reference Standard

Modality	Nodule Type at Reference Standard (n = 1073)				Weighted κ Value*	P Value	Accuracy (%)†
	Missed Nodules	GGN (n = 182)	PSN (n = 132)	SN (n = 759)			
Standard-dose CT	138	161	132	642	0.63 (0.58, 0.68)	<.001	87.1‡ (935/1073)
Low-dose CT	139	161	132	641	0.63 (0.58, 0.68)	<.001	87.0‡ (934/1073)
Pulmonary MRI with UTE	130	118	165	660	0.58 (0.53, 0.63)	<.001	83.1 (892/1073)

Note.—Except where indicated, data are numbers of nodules. GGN = ground-glass nodule, PSN = part-solid nodule, SN = solid nodule, UTE = ultrashort echo time.

* Numbers in parentheses are 95% CIs.

† Numbers in parentheses are numbers of nodules.

‡ Significantly different from pulmonary MRI with UTE ($P < .001$).

Table 4: Agreement for Lung-RADS Classification of All Nodules Based on Consensus Reading Results for Each Modality

Modality	Lung-RADS Classification						Weighted κ Value*	P Value	Accuracy (%)†
	Missed Nodules	2 (n = 706)	3 (n = 219)	4A (n = 58)	4B (n = 75)	4X (n = 15)			
Standard-dose CT	138	546	232	66	76	15	0.82 (0.80, 0.85)	<.001	81.9 (879/1073)
Low-dose CT	139	540	238	65	76	15	0.82 (0.80, 0.85)	<.001	81.7 (877/1073)
Pulmonary MRI with UTE	130	547	249	57	75	15	0.82 (0.79, 0.84)	<.001	81.7 (877/1073)

Note.—Except where indicated, data are numbers of nodules. Lung-RADS = Lung CT Screening Reporting and Data System, UTE = ultrashort echo time.

* Numbers in parentheses are 95% CIs.

† Numbers in parentheses are numbers of nodules.

Acknowledgments: The authors thank Shinichiro Seki, MD, PhD (Division of Functional and Diagnostic Imaging Research, Department of Radiology, Kobe University School of Medicine); Yuji Kishida, MD, PhD, and Takamichi Murakami, MD, PhD (Department of Radiology, Kobe University Graduate School of Medicine); Yoshimasa Maniwa, MD, PhD (Division of General Thoracic Surgery, Kobe University School of Medicine); Tomoo Ito, MD, PhD (Department of Diagnostic Pathology, Kobe University School of Medicine); Yoshihiro Nishimura, MD, PhD (Division of Respiratory Medicine, Department of Internal Medicine, Kobe University Graduate School of Medicine); Yasushi Hoshikawa, MD, PhD (Department of Thoracic Surgery, Fujita Health University School of Medicine); Masashi Kondo, MD, PhD, and Kazuyoshi Imaizumi, MD, PhD (Department of Respiratory Medicine, Fujita Health University School of Medicine); and Tetsuya Tsukamoto, MD, PhD (Department of Diagnostic Pathology, Fujita Health University School of Medicine) for their valuable contributions to this study.

Author contributions: Guarantor of integrity of entire study, Y.O.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, Y.O., D.T., M.Y., H.K., K.Y., N.H., C.S., A.W., T.U., H.I., H.H., K.M., H.T.; clinical studies, Y.O., D.T., T.Y., H.K., H.I., H.H., K.M., H.T.; statistical analysis, N.H., C.S., A.W., T.U.; and manuscript editing, Y.O., D.T., M.Y., H.K., K.Y., N.H., C.S., A.W., T.U., H.I., H.H., K.M., H.T.

Disclosures of conflicts of interest: **Y.O.** Research grants from Canon Medical Systems, Daiichi-Sankyo, and Fujifilm. **D.T.** No relevant relationships. **T.Y.** Research grants from Canon Medical Systems, Daiichi-Sankyo, and Fujifilm. **M.Y.** No relevant relationships. **H.K.** No relevant relationships. **K.Y.** No relevant relationships. **N.H.** No relevant relationships. **C.S.** No relevant relationships. **A.W.** No relevant relationships. **T.U.** No relevant relationships. **H.I.** No relevant relationships. **H.H.** No relevant relationships. **K.M.** Research grants from Canon Medical Systems, Daiichi-Sankyo, and Fujifilm. **H.T.** Research grants from Canon Medical Systems, Daiichi-Sankyo, and Fujifilm.

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.
2. Nawa T, Nakagawa T, Mizoue T, et al. A decrease in lung cancer mortality following the introduction of low-dose chest CT screening in Hitachi, Japan. Lung Cancer 2012;78(3):225–228.
3. Saghir Z, Dirksen A, Ashraf H, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. Thorax 2012;67(4):296–301.
4. Infante M, Cavuto S, Lutman FR, et al. Long-Term Follow-up Results of the DANTE Trial, a Randomized Study of Lung Cancer Screening with Spiral Computed Tomography. Am J Respir Crit Care Med 2015;191(10):1166–1175.
5. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N Engl J Med 2020;382(6):503–513.
6. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology 2005;237(2):395–400.
7. Naidich DP, Bankier AA, MacMahon H, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. Radiology 2013;266(1):304–317.
8. Kazerooni EA, Armstrong MR, Amorosa JK, et al. ACR CT accreditation program and the lung cancer screening program designation. J Am Coll Radiol 2015;12(1):38–42.
9. Kazerooni EA, Armstrong MR, Amorosa JK, et al. ACR CT Accreditation Program and the Lung Cancer Screening Program Designation. J Am Coll Radiol 2016;13(2 Suppl):R30–R34.
10. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology 2017;284(1):228–243.
11. American College of Radiology. Lung-Screening Reporting and Data System (LungRADS) Version 1.1. https://www.acr.org/-/media/ACR/Files/

RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf. Published 2019. Accessed January 3, 2020.
12. Dyer SC, Bartholmai BJ, Koo CW. Implications of the updated Lung CT Screening Reporting and Data System (Lung-RADS version 1.1) for lung cancer screening. J Thorac Dis 2020;12(11):6966–6977.
13. Chelala L, Hossain R, Kazerooni EA, Christensen JD, Dyer DS, White CS. Lung-RADS Version 1.1: Challenges and a Look Ahead, From the *AJR* Special Series on Radiology Reporting and Data Systems. AJR Am J Roentgenol 2021;216(6):1411–1422.
14. Ohno Y. New applications of magnetic resonance imaging for thoracic oncology. Semin Respir Crit Care Med 2014;35(1):27–40.
15. Ohno Y, Kauczor HU, Hatabu H, Seo JB, van Beek EJR; International Workshop for Pulmonary Functional Imaging (IWPFI). MRI for solitary pulmonary nodule and mass assessment: Current state of the art. J Magn Reson Imaging 2018;47(6):1437–1458.
16. Ciliberto M, Kishida Y, Seki S, Yoshikawa T, Ohno Y. Update of MR Imaging for Evaluation of Lung Cancer. Radiol Clin North Am 2018;56(3):437–469.
17. Hatabu H, Ohno Y, Gefter WB, et al. Expanding Applications of Pulmonary MRI in the Clinical Evaluation of Lung Disorders: Fleischner Society Position Paper. Radiology 2020;297(2):286–301.
18. Tanaka Y, Ohno Y, Hanamatsu S, et al. State-of-the-art MR Imaging for Thoracic Diseases. Magn Reson Med Sci 2021. doi:10.2463/mrms.rev.2020-0184. Published online April 29, 2021.
19. Vogt FM, Herborn CU, Hunold P, et al. HASTE MRI versus chest radiography in the detection of pulmonary nodules: comparison with MDCT. AJR Am J Roentgenol 2004;183(1):71–78.
20. Koyama H, Ohno Y, Kono A, et al. Quantitative and qualitative assessment of non-contrast-enhanced pulmonary MR imaging for management of pulmonary nodules in 161 subjects. Eur Radiol 2008;18(10):2120–2131.
21. Sieren JC, Ohno Y, Koyama H, Sugimura K, McLennan G. Recent technological and application developments in computed tomography and magnetic resonance imaging for improved pulmonary nodule detection and lung cancer staging. J Magn Reson Imaging 2010;32(6):1353–1369.
22. Regier M, Schwarz D, Henes FO, et al. Diffusion-weighted MR-imaging for the detection of pulmonary nodules at 1.5 Tesla: intraindividual comparison with multidetector computed tomography. J Med Imaging Radiat Oncol 2011;55(3):266–274.
23. Sommer G, Tremper J, Koenigkam-Santos M, et al. Lung nodule detection in a high-risk population: comparison of magnetic resonance imaging and low-dose computed tomography. Eur J Radiol 2014;83(3):600–605.
24. Dewes P, Frellesen C, Al-Butmeh F, et al. Comparative evaluation of non-contrast CAIPIRINHA-VIBE 3T-MRI and multidetector CT for detection of pulmonary nodules: In vivo evaluation of diagnostic accuracy and image quality. Eur J Radiol 2016;85(1):193–198.
25. Meier-Schroers M, Kukuk G, Homsi R, Skowasch D, Schild HH, Thomas D. MRI of the lung using the PROPELLER technique: Artifact reduction, better image quality and improved nodule detection. Eur J Radiol 2016;85(4):707–713.
26. Johnson KM, Fain SB, Schiebler ML, Nagle S. Optimized 3D ultrashort echo time pulmonary MRI. Magn Reson Med 2013;70(5):1241–1250.
27. Ohno Y, Koyama H, Yoshikawa T, et al. Pulmonary high-resolution ultrashort TE MR imaging: Comparison with thin-section standard- and low-dose computed tomography for the assessment of pulmonary parenchyma diseases. J Magn Reson Imaging 2016;43(2):512–532.
28. Ohno Y, Koyama H, Yoshikawa T, et al. Standard-, Reduced-, and No-Dose Thin-Section Radiologic Examinations: Comparison of Capability for Nodule Detection and Nodule Type Assessment in Patients Suspected of Having Pulmonary Nodules. Radiology 2017;284(2):562–573.
29. Jiang W, Ong F, Johnson KM, et al. Motion robust high resolution 3D free-breathing pulmonary MRI using dynamic 3D image self-navigator. Magn Reson Med 2018;79(6):2954–2967.
30. Wielpütz MO, Lee HY, Koyama H, et al. Morphologic Characterization of Pulmonary Nodules With Ultrashort TE MRI at 3T. AJR Am J Roentgenol 2018;210(6):1216–1225.
31. Wielpütz MO, Triphan SMF, Ohno Y, Jobst BJ, Kauczor HU. Outracing Lung Signal Decay - Potential of Ultrashort Echo Time MRI. Rofo 2019;191(5):415–423.
32. Bae K, Jeon KN, Hwang MJ, et al. Comparison of lung imaging using three-dimensional ultrashort echo time and zero echo time sequences: preliminary study. Eur Radiol 2019;29(5):2253–2262.
33. Zhu X, Chan M, Lustig M, Johnson KM, Larson PEZ. Iterative motion-compensation reconstruction ultra-short TE (iMoCo UTE) for high-resolution free-breathing pulmonary MRI. Magn Reson Med 2020;83(4):1208–1221.
34. Zeng F, Nogami M, Ueno YR, et al. Diagnostic performance of zero-TE lung MR imaging in FDG PET/MRI for pulmonary malignancies. Eur Radiol 2020;30(9):4995–5003.
35. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys 2004;31(8):2313–2330.
36. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–310.
37. Svanholm H, Starklint H, Gundersen HJ, Fabricius J, Barlebo H, Olsen S. Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. APMIS 1989;97(8):689–698.
38. Mandrekar JN. Measures of interrater agreement. J Thorac Oncol 2011;6(1):6–7.