

# AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population: A Randomized Controlled Trial

Ju Gang Nam, MD, PhD • Eui Jin Hwang, MD, PhD • Jayoun Kim, PhD • Nanbee Park, MS • Eun Hee Lee, MD, PhD • Hyun Jin Kim, MD, PhD • Miyeon Nam, MD • Jong Hyuk Lee, MD, PhD • Chang Min Park, MD, PhD • Jin Mo Goo, MD, PhD

From the Department of Radiology (J.G.N., E.J.H., J.H.L., C.M.P., J.M.G.), Artificial Intelligence Collaborative Network (J.G.N.), Medical Research Collaborating Center (J.K., N.P.), and Center for Health Promotion and Optimal Aging (E.H.L., M.N.), Seoul National University Hospital and College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea; Department of Radiology, Ewha Womans University Seoul Hospital, Seoul, Republic of Korea (H.J.K.); Institute of Medical and Biological Engineering, Seoul National University Medical Research Center, Seoul, Republic of Korea (C.M.P.); and Cancer Research Institute, Seoul National University, Seoul, Republic of Korea (J.M.G.). Received July 26, 2022; revision requested September 6; revision received November 18; accepted November 28. Address correspondence to J.M.G. (email: [jmgoo@smu.ac.kr](mailto:jmgoo@smu.ac.kr)).

Supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health and Welfare, Republic of Korea (grant HI19C1129).

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

See also the editorial by Auffermann in this issue.

Radiology 2023; 000:e221894 • <https://doi.org/10.1148/radiol.221894> • Content codes: **CH** **AI**

**Background:** The impact of artificial intelligence (AI)-based computer-aided detection (CAD) software has not been prospectively explored in real-world populations.

**Purpose:** To investigate whether commercial AI-based CAD software could improve the detection rate of actionable lung nodules on chest radiographs in participants undergoing health checkups.

**Materials and Methods:** In this single-center, pragmatic, open-label randomized controlled trial, participants who underwent chest radiography between July 2020 and December 2021 in a health screening center were enrolled and randomized into intervention (AI group) and control (non-AI group) arms. One of three designated radiologists with 13–36 years of experience interpreted each radiograph, referring to the AI-based CAD results for the AI group. The primary outcome was the detection rate, that is, the number of true-positive radiographs divided by the total number of radiographs, of actionable lung nodules confirmed on CT scans obtained within 3 months. Actionable nodules were defined as solid nodules larger than 8 mm or subsolid nodules with a solid portion larger than 6 mm (Lung Imaging Reporting and Data System, or Lung-RADS, category 4). Secondary outcomes included the positive-report rate, sensitivity, false-referral rate, and malignant lung nodule detection rate. Clinical outcomes were compared between the two groups using univariable logistic regression analyses.

**Results:** A total of 10 476 participants (median age, 59 years [IQR, 50–66 years]; 5121 men) were randomized to an AI group ( $n = 5238$ ) or non-AI group ( $n = 5238$ ). The trial met the predefined primary outcome, demonstrating an improved detection rate of actionable nodules in the AI group compared with the non-AI group (0.59% [31 of 5238 participants] vs 0.25% [13 of 5238 participants], respectively; odds ratio, 2.4; 95% CI: 1.3, 4.7;  $P = .008$ ). The detection rate for malignant lung nodules was higher in the AI group compared with the non-AI group (0.15% [eight of 5238 participants] vs 0.0% [0 of 5238 participants], respectively;  $P = .008$ ). The AI and non-AI groups showed similar false-referral rates (45.9% [56 of 122 participants] vs 56.0% [56 of 100 participants], respectively;  $P = .14$ ) and positive-report rates (2.3% [122 of 5238 participants] vs 1.9% [100 of 5238 participants];  $P = .14$ ).

**Conclusion:** In health checkup participants, artificial intelligence-based software improved the detection of actionable lung nodules on chest radiographs.

© RSNA, 2023

Supplemental material is available for this article.

Since the adoption of artificial intelligence (AI), computer-aided detection (CAD) systems have substantially increased their utility in various medical examinations including mammography, brain CT, and chest radiography or CT for diverse indications including lesion detection, differential diagnosis, prioritization of urgent images, or imaging biomarker extraction (1–3). Specifically in chest radiology, detecting lung nodules has been a classic task. Various AI-based CAD systems have been reported to substantially improve radiologists' performance as a second reader (4–8). However, these retrospective publications have several major limitations. First, the performance of AI-based CAD systems

was validated in retrospective data sets, which are often arbitrarily selected to have a disease-enriched and dichotomized distribution. Second, the performance tests were conducted under conditions different from real practice, where readers could be more focused and sensitive, yielding performance bias. Third, proper integration with a picture archiving and communication system (PACS) was lacking, which is necessary to use AI-based CAD in real work processes.

In this context, evidence from prospective trials assessing the impact of AI-based CAD-integrated PACS (hereafter, AI-PACS) in real populations is highly warranted. Health checkup populations could be particularly suitable target

## Abbreviations

AI = artificial intelligence, CAD = computer-aided detection, OR = odds ratio, PACS = picture archiving and communication system

## Summary

In a randomized controlled trial, artificial intelligence–based nodule detection software improved the detection rate of actionable lung nodules on chest radiographs in a health screening population.

## Key Results

- In a pragmatic controlled trial of 10 476 health checkup participants randomized to either an artificial intelligence (AI) or non-AI group, the detection rate of Lung Imaging Reporting and Data System, or Lung-RADS, category 4 nodules on chest radiographs improved with assistance from AI software (odds ratio, 2.4;  $P = .008$ ).
- The AI and non-AI groups showed an equivalent false-referral rate (45.9% vs 56.0%;  $P = .14$ ) and positive-report rate (2.3% vs 1.9%;  $P = .14$ ) of chest radiographs.

populations. Although chest radiography as a screening tool failed to reduce lung cancer mortality in several large, randomized trials (9,10), chest radiography is still frequently used for screening various lung diseases (11–13). In particular, health checkups using chest radiography are commonly performed for the general population in some countries (14–16). Moreover, retrospective studies have proposed the potential of AI-based CAD systems for improving the role of chest radiography in lung cancer screening (14,17–19).

In this study, we aimed to investigate the clinical utility of AI-based CAD in health checkup participants through a randomized controlled trial. We integrated AI-based CAD into a commercial PACS and incorporated it into the real clinical work process. The purpose of our randomized controlled trial was to investigate whether commercial AI-based CAD software could improve the detection rate of actionable lung nodules on chest radiographs in health checkup participants.

## Materials and Methods

This single-center, open-label, randomized controlled trial was approved by the institutional review board of Seoul National University Hospital (institutional review board number: D-1908-160-1059) and registered in the Clinical Research Information Service (<https://cris.nih.go.kr>; registration number: KCT0005051). The requirement for written informed consent was waived. We followed the guidelines outlined in the Consolidated Standards of Reporting AI Trials (ie, CONSORT-AI) Extension Checklist (20). The trial was supported by a research grant funded by the Ministry of Health and Welfare, Republic of Korea (grant HI19C1129). None of the study participants have been reported previously.

## Participants and Trial Design

We conducted a pragmatic, randomized controlled trial at a health screening center affiliated with a tertiary referral hospital. All individuals who visited the center and underwent chest radiography for health checkup purposes between June 2020 and December 2021 were enrolled and randomized into either the intervention arm (AI group) or control arm (non-AI group) at

a 1:1 ratio. Individuals aged 18 years or younger were excluded. For those in the AI group, the designated radiologists interpreted chest radiographs aided by AI-based CAD, while for those in the non-AI group, the radiologists interpreted chest radiographs without AI-based CAD results (Fig 1). Most chest radiographs were interpreted by one of three designated board-certified radiologists (E.H.L., H.J.K., and M.N., with 36, 27, and 13 years of experience in chest radiography reading, respectively) at our health screening center. The radiologists, physicians in the health screening center, and outcome assessors were aware of the allocation (single blinded).

Baseline characteristics including age; sex; smoking status; history of lung cancer, other malignancy, lung surgery, and pulmonary tuberculosis; family history of lung cancer; and comorbidities (hypertension, diabetes mellitus, dyslipidemia, and chronic hepatitis) were collected using a self-reported health questionnaire (Fig S1, Appendix S1). For those who underwent chest CT within 3 months after chest radiography, CT images were reviewed, and the presence of actionable lung nodules was determined (Appendix S2) (21–23). For participants who underwent pathologic evaluation, the pathology reports were investigated. All radiologic and clinical information was recorded using an electronic case report (Appendix S1).

## Chest Radiographs

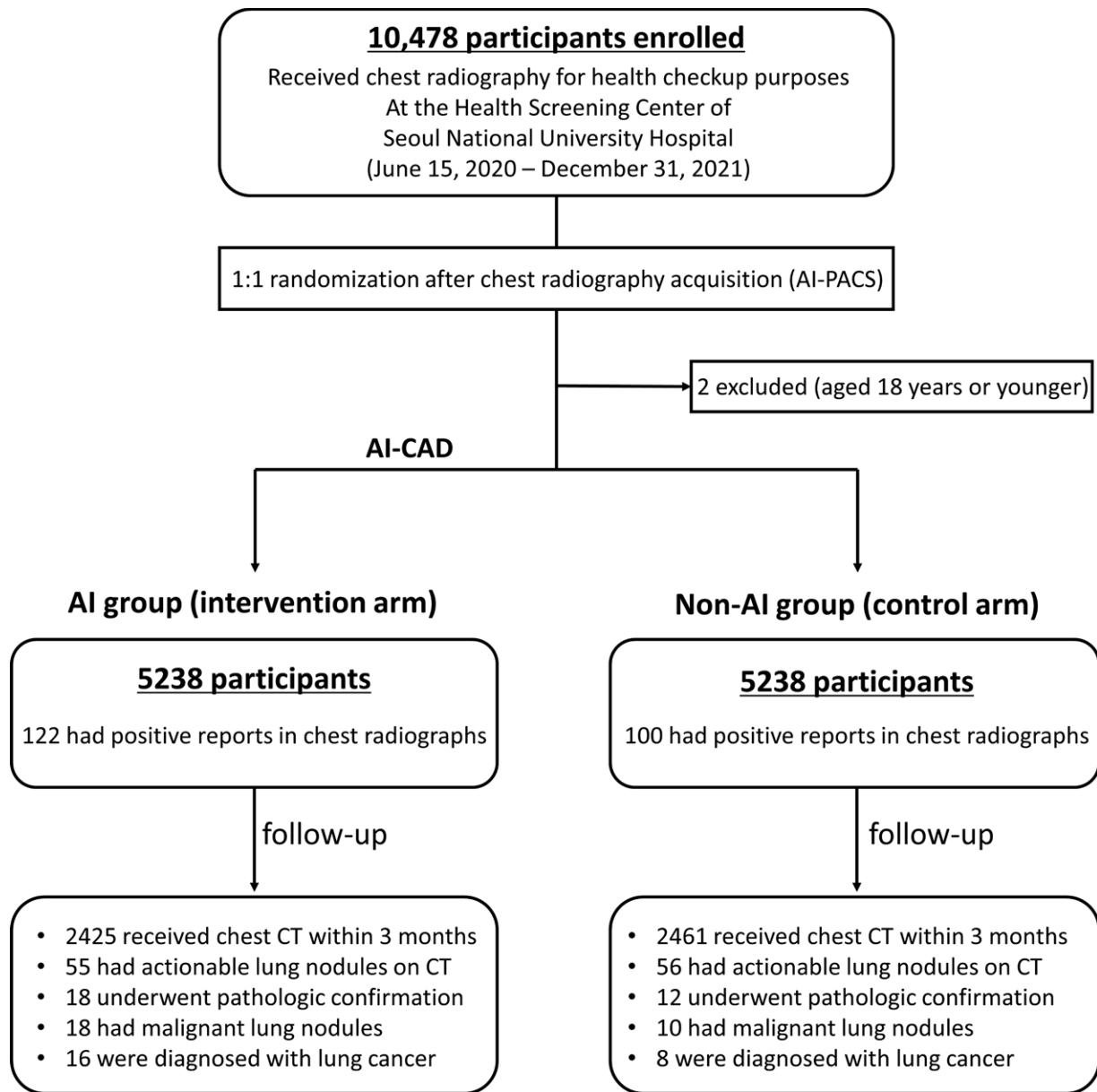
All chest radiographs were obtained in the posteroanterior projection without lateral views using INOVISION-EXII (Dong Kang Medical Systems) (tube voltage, 120 kVp; tube current, 320 mA; variable exposure time, typically 15–18 msec). For some participants who underwent two chest radiographic examinations during the trial, only the latter images were used in the analyses.

## CT Protocol

Most CT scans were obtained with low-dose protocol CT without contrast material enhancement. The median volume CT dose index and dose-length product were 0.52 mGy (IQR, 0.43–2.00 mGy) and 20.6 mGy · cm (IQR, 16.5–66.5 mGy · cm), respectively (Appendix S3).

## AI-PACS: AI-based CAD Implementation and Randomization

We used an AI-based CAD–implemented PACS (AI-PACS) devised for this trial. We embedded a commercial AI-based CAD software (Lunit INSIGHT CXR version 2.0.2.0; Lunit) approved by the Ministry of Food and Drug Safety of Korea (4–6) into a commercial PACS (M6; Infinitt Healthcare). When each chest radiograph was acquired, the image was immediately allocated a unique five-digit random number and randomized into the AI or non-AI group. In the AI group, the AI-PACS overlaid the results of AI-based CAD onto the chest radiograph, whereas the AI-based CAD results were blinded in the non-AI group (Fig 2). The AI-based CAD system analyzed frontal chest radiographs to assess the possibility of having major thoracic abnormalities (ie, pulmonary nodule, pneumonia, and pneumothorax) on a percentage scale (0%–100%) and localized them as overlaid heat maps (Fig 3). The radiologists used a structured reporting system



**Figure 1:** Participant flow diagram. AI = artificial intelligence, CAD = computer-aided detection, PACS = picture archiving and communication system.

implemented in AI-PACS (Fig S2) to interpret all chest radiographs from both groups.

**Clinical Outcomes**

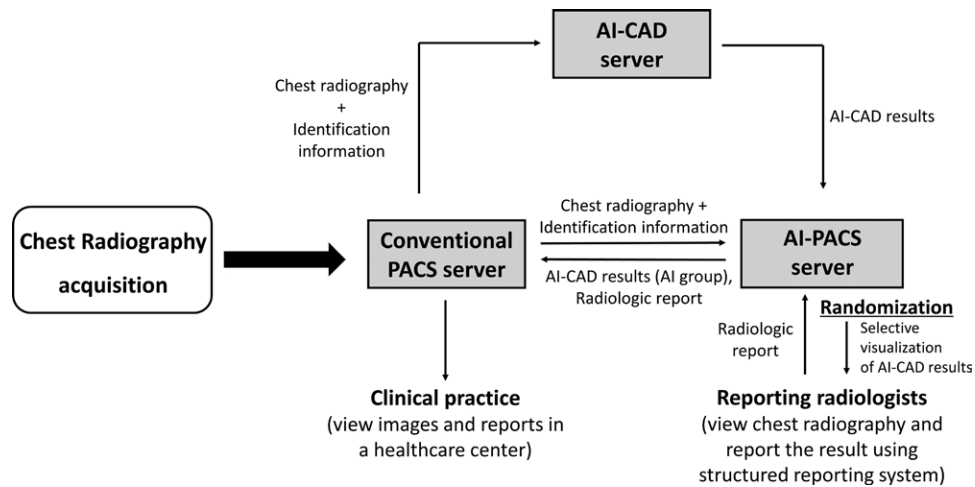
The primary outcome of our trial was the detection rate of actionable lung nodules on chest radiographs. Actionable lung nodules were defined as solid nodules larger than 8 mm or sub-solid nodules with a solid portion larger than 6 mm in average diameter measured on two axes (Lung Imaging Reporting and Data System, or Lung-RADS, category 4), regarding their clinical significance and visibility on chest radiographs (21). The detection rate was defined as the number of true-positive chest radiographs divided by the total number of chest radiographs.

The secondary outcomes included the false-referral rate of chest radiographs, positive-report rate of chest radiographs, rate of performing chest CT, positive rate for actionable lung nodules

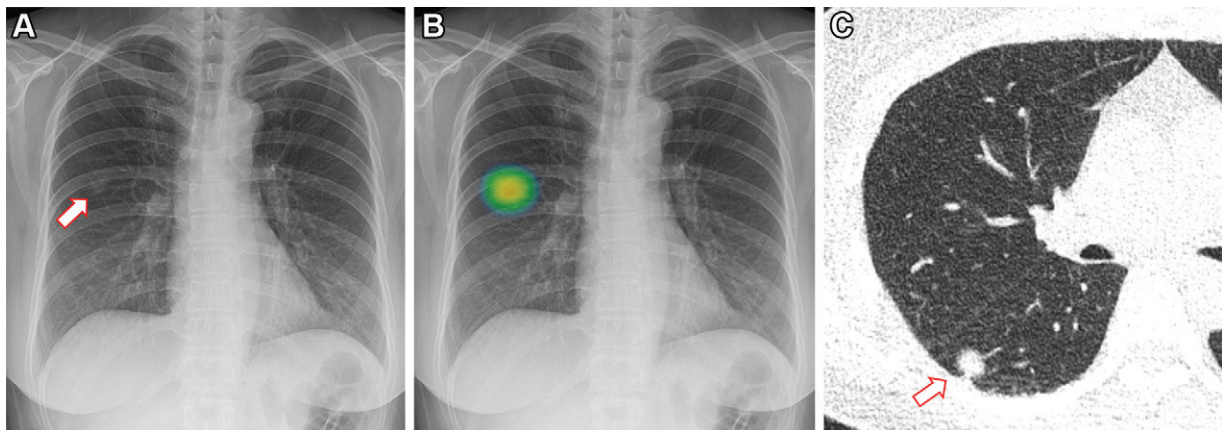
on chest CT scans, rate of performing pathologic evaluations, detection rate of malignant lung nodules on chest radiographs, detection rate of lung cancer on chest radiographs, positive rate of malignant lung nodules, and positive rate of lung cancer. Definitions are listed in Appendix S4.

**Sample Size Estimation**

We conducted a preliminary analysis using a retrospective cohort of 3073 individuals who underwent chest radiography and chest CT in our health screening center (16). In this sample, the estimated detection rates of actionable lung nodules were 1.73% and 1.49% with and without AI-based CAD, respectively. Assuming a two-tailed test with a type I error of .05 and power of 80%, the estimated sample size to determine the difference in detection rates was 83 552. We planned to stop the trial within 18 months (expected sample size, 8000–



**Figure 2:** Work process of chest radiography interpretation during the clinical trial. When a chest radiograph was acquired at the health screening center, the image and identification information were immediately transmitted to the artificial intelligence (AI)-based computer-aided detection (CAD) server and AI picture archiving and communication system (PACS) server. The AI-based CAD server analyzed the image and sent the results to the AI-based CAD-implemented PACS (AI-PACS) server, and the AI-PACS server randomized the participant into either the AI or non-AI group. If the participant was allocated to the AI group, the AI-PACS server provided AI-based CAD results and sent this information back to the conventional PACS server. These sequences took place immediately after the chest radiograph acquisition. The reporting radiologists viewed and reported chest radiographs using a structured reporting system implemented in the AI-PACS, and the final radiologic report was immediately sent to the conventional PACS server. Clinicians at the health screening center viewed the images and radiologic reports using the conventional PACS for clinical practice.



**Figure 3:** Images in a 60-year-old woman who underwent chest radiography for health checkup purposes and was allocated to the artificial intelligence (AI) group. **(A)** Frontal chest radiograph shows a subtle nodular opacity (arrow) in the right middle lung zone. **(B)** The lesion was detected by the AI-based computer-aided detection software, with an abnormality probability of 81.1%. The designated radiologist reported this chest radiograph as positive. **(C)** Axial, noncontrast, low-dose chest CT scan shows a 1.1-cm solid nodule (arrow) in the right lower lobe. The patient underwent percutaneous needle biopsy, and the nodule was confirmed to be adenocarcinoma.

10 000), considering (a) continuous improvements in AI-based CAD performance, (b) policies of the research grant, and (c) expected enhanced effect of AI-based CAD in real clinical practice with low disease prevalence.

### Statistical Analyses

Primary and secondary clinical outcomes were compared between the two groups using the  $\chi^2$  test, Fisher exact test, or Wilcoxon rank sum test, and the impact of the intervention (use of AI-based CAD) on primary and secondary clinical outcomes was evaluated using univariable logistic regression analy-

ses. Subgroup analyses were conducted to evaluate the interactions of confounding factors on the clinical outcomes. Age; sex; smoking status; history of lung cancer, other malignancy, lung surgery, and pulmonary tuberculosis; presence of prior chest radiographs; and the reporting radiologist were included as potential confounders. The subgroup effects of the confounding factors were tested by the significance of the interaction term added to the regression model. The Firth correction was applied when the number of events was relatively too low (24). We also performed multivariable logistic analyses to assess the associations of baseline characteristics of the participants with

**Table 1: Baseline Characteristics of the Participants**

Variable	All Participants ( <i>n</i> = 10 476)	AI Group ( <i>n</i> = 5238)	Non-AI Group ( <i>n</i> = 5238)	<i>P</i> Value*
Age (y)				.88
Mean ± SD	57 ± 13	57 ± 13	57 ± 13	
Median <sup>†</sup>	59 (50–66)	59 (50–66)	60 (50–66)	
Range	19–92	19–92	19–92	
Age group				.68
18–64 years	7267 (69)	3629 (69)	3638 (69)	
65–74 years	2339 (22)	1162 (22)	1177 (22)	
≥75 years	870 (8)	447 (9)	423 (8)	
Sex				.22
F	5355 (51)	2646 (51)	2709 (52)	
M	5121 (49)	2592 (49)	2529 (48)	
Smoking status				.55
Never-smoker	6705 (64)	3326 (64)	3379 (65)	
Former smoker	2633 (25)	1338 (26)	1295 (25)	
Current smoker	1138 (11)	574 (11)	564 (11)	
History of lung cancer	59 (0.6)	31 (0.6)	28 (0.5)	.70
History of other malignancy	1038 (10)	504 (10)	534 (10)	.33
History of lung surgery	73 (0.7)	36 (0.7)	37 (0.7)	.91
Family history of lung cancer	1120 (11)	537 (10)	583 (11)	.15
History of pulmonary tuberculosis	444 (4)	207 (4)	237 (5)	.15
Comorbidities				
Hypertension	2764 (26)	1389 (27)	1375 (26)	.76
Diabetes mellitus	1282 (12)	636 (12)	646 (12)	.77
Dyslipidemia	3072 (29)	1533 (29)	1539 (29)	.90
Chronic hepatitis	419 (4)	222 (4)	197 (4)	.21

Note.—Except where indicated, data are numbers of participants, with percentages in parentheses. AI = artificial intelligence.

\* *P* values were calculated using the Wilcoxon rank sum test (for age) or  $\chi^2$  test (others).

<sup>†</sup> Numbers in parentheses are the IQR.

the performance of chest CT and positive-report rates of chest radiographs, respectively, and the association of nodule characteristics with the sensitivity of chest radiography. All analyses except those for the primary outcome were considered exploratory. Statistical analyses were conducted using SAS 9.4 (SAS Institute) by statisticians with 10 and 15 years of experience (N.P. and J.K., respectively). *P* < .05 was considered to indicate a statistically significant difference.

## Results

### Participant Characteristics

In total, 10 478 participants were enrolled in AI-PACS, and two participants aged 18 years or younger were excluded. In the final analysis, 10 476 participants (median age, 59 years [IQR, 50–66 years]; 5121 men) were allocated into the AI (*n* = 5238) and non-AI (*n* = 5238) groups (Fig 1). In the overall cohort, 11% (1138 of 10 476 participants) and 25% (2633 of 10 476 participants) were current and former smokers, respectively. Of the 10 476 participants, 0.6% (*n* = 59) had a history of lung cancer and 10% (*n* = 1038) had a history of other malignancy (Table 1). Other information is summarized in Table 1, and there was no

evidence of differences in baseline characteristics between the two groups (*P* = .15–.91).

The chest radiographs were reported to contain nodules in 2% of the 10 476 participants (*n* = 222). Older age (*P* < .001), a history of lung cancer (*P* = .002) or pulmonary tuberculosis (*P* < .001), and the absence of a prior chest radiographic examination (*P* = .001) were associated with positive results, and the positive-report rates differed among the reporting radiologists (*P* < .001) (Table S1). Of the 10 476 participants, 47% (*n* = 4886) underwent chest CT within 3 months after chest radiography and 0.3% (*n* = 30) underwent pathologic evaluation of lung nodules. Participants who were aged 65–74 years (*P* < .001), male (*P* = .004), and former or current smokers (*P* < .001) and those who did not have a prior chest radiograph (*P* = .001) were more likely to undergo chest CT (Table S2).

### Clinical Outcomes

The detection rate for actionable lung nodules on chest radiographs was higher in the AI group than in the non-AI group (0.59% [31 of 5238 participants] vs 0.25% [13 of 5238 participants], respectively; odds ratio [OR], 2.4; 95% CI: 1.3, 4.6; *P* = .008) (Table 2), accomplishing the primary outcome of the trial (Table 1). The positive-report

**Table 2: Summary of Analyses for the Primary and Secondary Outcomes**

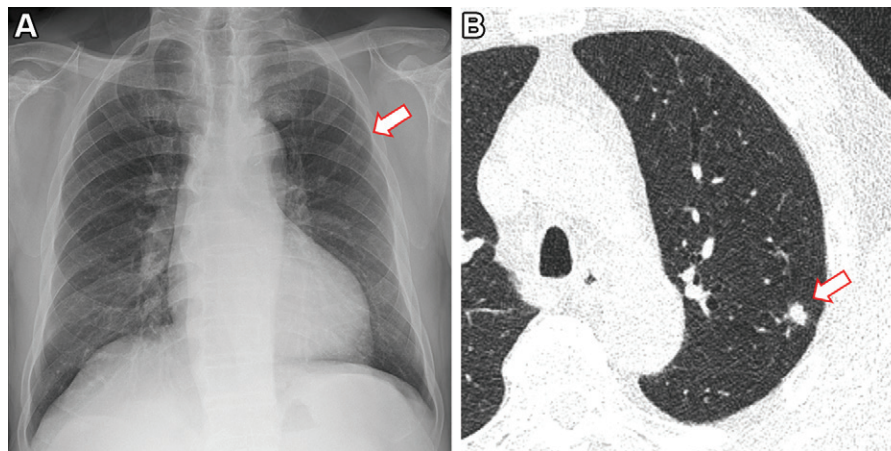
Outcome	All Participants ( <i>n</i> = 10 476)	AI Group ( <i>n</i> = 5238)	Non-AI Group ( <i>n</i> = 5238)	Odds Ratio*	<i>P</i> Value
<b>Primary outcome</b>					
Detection rate of actionable lung nodules on chest radiographs	0.42 (44/10 476)	0.59 (31/5238)	0.25 (13/5238)	2.4 (1.3, 4.6)	.008 <sup>†</sup>
<b>Secondary outcomes</b>					
False-referral rate of chest radiography	50.5 (112/222)	45.9 (56/122)	56.0 (56/100)	0.67 (0.39, 1.1)	.14
Positive-report rate of chest radiography	2.1 (222/10 476)	2.3 (122/5238)	1.9 (100/5238)	1.2 (0.94, 1.6)	.14
Performance of chest CT	46.6 (4886/10 476)	46.3 (2425/5238)	47.0 (2461/5238)	0.97 (0.90, 1.1)	.48
Positive rate of actionable lung nodules	1.1 (111/10 467)	1.1 (55/5238)	1.1 (56/5238)	0.98 (0.68, 1.4)	.92
Performance of pathologic evaluation	0.29 (30/10 476)	0.34 (18/5238)	0.23 (12/5238)	1.5 (0.72, 3.1)	.28
Detection rate of malignant lung nodules on chest radiographs	0.08 (8/10 476)	0.15 (8/5238)	0.0 (0/5238)	17.0 (0.98, 295.1)	.05 <sup>‡</sup>
Detection rate of lung cancer on chest radiographs	0.06 (6/10 476)	0.11 (6/5238)	0.0 (0/5238)	13.0 (0.73, 231.1)	.08 <sup>‡</sup>
Positive rate of malignant lung nodules	0.27 (28/10 476)	0.34 (18/5238)	0.19 (10/5238)	1.8 (0.83, 3.9)	.14
Positive rate of lung cancer	0.23 (24/10 476)	0.31 (16/5238)	0.15 (8/5238)	2.0 (0.86, 4.7)	.10

Note.—Except where indicated, data are percentages, with numbers of participants in parentheses. Actionable lung nodules were defined as solid nodules larger than 8 mm or subsolid nodules with a solid portion larger than 6 mm (Lung Imaging Reporting and Data System, or Lung-RADS, category 4). Odds ratios and *P* values were calculated from univariable logistic regression analyses. AI = artificial intelligence.

\* Numbers in parentheses are 95% CIs.

<sup>†</sup> Statistically significant.

<sup>‡</sup> The Firth correction was applied.



**Figure 4:** Images in a 73-year-old man who underwent chest radiography and low-dose CT for health checkup purposes and was allocated to the non-artificial intelligence (AI) group. **(A)** Frontal chest radiograph shows a small nodular opacity (arrow) in the left upper lung zone, which was missed by the designated reporting radiologist. **(B)** Axial, noncontrast, low-dose chest CT scan shows a 9-mm solid nodule (arrow) in the left upper lobe. The nodule showed low metabolism at PET and decreased in size at follow-up CT. It was confirmed to be an inflammatory nodule.

rate of chest radiography showed no evidence of a difference between the AI group (2.3% [122 of 5238 participants]) and non-AI group (1.9% [100 of 5238 participants]; OR, 1.4; 95% CI: 0.94, 1.6; *P* = .14). Among the 222 participants with positive chest radiographs, the AI and non-AI groups showed a similar false-referral rate (45.9% [56 of 122 participants] vs 56.0% [56 of 100 participants], respectively; OR, 0.67; 95% CI: 0.39, 1.1; *P* = .14). Representative images are presented in Figures 3 and 4.

However, because chest CT was performed for 47% of the participants regardless of the chest radiography results (mostly for health checkup purposes), considerable actionable lung nodules were found initially at CT. There was no evidence of a difference in CT performance between the AI group (46.3% [2425 of 5238 participants]) and non-AI group (47.0% [2461 of 5238 participants]; OR, 0.97; 95% CI: 0.90, 1.1; *P* = .48). In the overall cohort, 1.1% (111 of 10 476 participants) showed actionable lung nodules at chest CT, and the positive

**Table 3: Diagnostic Performance of Chest Radiography for Detecting Actionable Lung Nodules in the 4886 Participants Who Underwent Chest CT**

Parameter	All Participants with Chest CT ( <i>n</i> = 4886)	AI Group with Chest CT ( <i>n</i> = 2425)	Non-AI Group with Chest CT ( <i>n</i> = 2461)	<i>P</i> Value
Actionable nodules on CT scans	111	55	56	...
Positive report at chest radiography	156	87	69	...
True positive	44	31	13	...
False positive	112	56	56	...
Negative report at chest radiography	4730	2338	2349	...
True negative	4663	2314	2349	...
False negative	67	24	43	...
Diagnostic performance of chest radiography*				
Sensitivity	39.6 (30.5, 48.7)	56.4 (43.3, 69.5)	23.2 (12.2, 34.3)	<.001 <sup>†</sup>
Specificity	97.7 (97.2, 98.1)	97.6 (97.0, 98.3)	97.7 (97.1, 98.3)	.94
Positive predictive value	28.2 (21.1, 35.3)	35.6 (25.6, 45.7)	18.8 (9.6, 28.1)	.02 <sup>†</sup>
Negative predictive value	98.6 (98.3, 98.9)	99.0 (98.6, 99.4)	98.2 (97.7, 98.7)	.03 <sup>†</sup>

Note.—Except where indicated, data are numbers of participants. These results were analyzed among 4886 participants who underwent chest CT for any purpose within 3 months after chest radiography. Actionable lung nodules were defined as solid nodules larger than 8 mm or subsolid nodules with solid a portion larger than 6 mm on CT scans (Lung Imaging Reporting and Data System, or Lung-RADS, category 4). AI = artificial intelligence.

\* Diagnostic performance measures of chest radiography are presented as percentages, with 95% CIs in parentheses. Each measure was compared using the  $\chi^2$  test between the AI and non-AI groups.

<sup>†</sup> Statistically significant.

rate of actionable lung nodules was similar between the AI group (1.1% [55 of 5238 participants]) and the non-AI group (1.1% [56 of 5238 participants]; OR, 0.98; 95% CI: 0.68, 1.4; *P* = .92). When the diagnostic performance of chest radiography was evaluated among the participants who underwent chest CT (*n* = 4886), the AI group showed higher sensitivity (56.4% [31 of 55 participants] vs 23.2% [13 of 56 participants], respectively; *P* < .001), positive predictive value (35.6% [31 of 87 participants] vs 18.8% [13 of 69 participants]; *P* = .02), and negative predictive value (99.0% [2314 of 2338 participants] vs 98.2% [2349 of 2392 participants]; *P* = .03) in detecting actionable lung nodules (Table 3).

In the overall cohort, 0.28% of participants (30 of 10476) underwent pathologic evaluation for lung nodules (22 by means of surgical resection and eight by means of bronchoscopic or percutaneous needle biopsy), where the proportions were similar between the AI group (0.34% [18 of 5238 participants]) and non-AI group (0.23% [12 of 5238 participants]; OR, 1.5; 95% CI: 0.72, 3.1; *P* = .28). Twenty-four participants were diagnosed with malignant lung cancer (20 adenocarcinomas, one squamous cell carcinoma, one adenosquamous carcinoma, and two minimally invasive adenocarcinoma) and four with other malignancies (two metastasis, one atypical carcinoid, and one lymphoma). The detection rates of malignant lung nodules on chest radiographs (0.15% [eight of 5238] in the AI group vs 0% [0 of 5238] in the non-AI group; *P* = .008) and lung cancer on chest radiographs (0.11% [six of 5238] vs 0% [0 of 5238]; *P* = .03) were higher in the AI group; however, there was no evidence of a difference at logistic regression analyses (OR, 17.0 [95% CI: 0.98, 295.1; *P* = .05] and 13.0 [95% CI: 0.73, 231.1; *P* = .08], respectively) (Table 2).

### Subgroup Analyses

The impact of AI-based CAD on the detection of actionable lung nodules on chest radiographs was consistent across the subgroups (Table 4). There was no evidence of significant interactions for age; sex; smoking status; history of lung cancer, other malignancy, lung surgery, and pulmonary tuberculosis; the presence of prior chest radiographs; and the reporting radiologists in detecting actionable lung nodules on chest radiographs (*P* = .05–.98) (Table 4). Specifically, age and sex showed *P* values of .06 and .05, respectively. None of the aforementioned factors showed evidence of significant interactions with AI-based CAD in detecting malignant lung nodules on chest radiographs (*P* > .05 for all) (Table S3) or performing CT (*P* > .05 for all) (Table S4). The sensitivity of chest radiography was not affected by nodule characteristics including size, location, and overlapping structures (Table S5).

### Discussion

Although various artificial intelligence (AI)-based computer-aided detection (CAD) systems have been proposed, few have been prospectively validated. In this study, we conducted a randomized controlled trial at a health screening center to investigate the impact of AI-based CAD on the detection of actionable lung nodules on chest radiographs. We enrolled all 10476 adult participants who underwent at least one chest radiographic examination and randomly allocated them into intervention (AI group, *n* = 5238) and control (non-AI group, *n* = 5238) arms. We demonstrated that the use of AI-based CAD improved the detection rate of actionable lung nodules on chest radiographs (odds ratio, 2.4; 95% CI: 1.3, 4.6; *P* = .008), meeting the primary outcome. There was no evidence of differences in

**Table 4: Subgroup Analyses for the Primary Outcome**

Parameter	Detection Rate of Actionable Lung Nodules on Chest Radiographs (%)			P Value <sup>†</sup>
	AI Group	Non-AI Group	Odds Ratio*	
Overall	0.59 (31/5238)	0.25 (13/5238)	2.4 (1.3, 4.6)	
Age				.06
18–64 years	0.6 (20/3629)	0.1 (3/3638)	6.7 (2.0, 22.6)	
65–74 years	0.6 (7/1162)	0.6 (7/1177)	1.0 (0.35, 2.9)	
≥75 years	0.9 (4/447)	0.7 (3/423)	1.3 (0.28, 5.7)	
Sex				.05
F	0.5 (13/2646)	0.4 (10/2709)	1.3 (0.58, 3.0)	
M	0.7 (18/2592)	0.1 (3/2529)	5.9 (1.7, 20.0)	
Smoking status				.38
Never-smoker	0.5 (18/3326)	0.3 (10/3379)	1.8 (0.85, 4.0)	
Former smoker	0.7 (9/1338)	0.1 (1/1295)	8.8 (1.1, 69.2)	
Current smoker	0.7 (4/574)	0.4 (2/564)	2.0 (0.36, 10.8)	
History of lung cancer				.90 <sup>‡</sup>
No	0.6 (30/5207)	0.2 (13/5210)	2.3 (1.2, 4.5)	
Yes	3.2 (1/31)	0 (0/28)	2.8 (0.10, 75.8)	
History of other malignancy				.79
No	0.6 (28/4734)	0.3 (12/4704)	2.3 (1.2, 4.6)	
Yes	0.6 (3/504)	0.2 (1/534)	3.2 (0.33, 30.8)	
History of lung surgery				.69 <sup>‡</sup>
No	0.6 (31/5202)	0.2 (13/5201)	2.4 (1.3, 4.6)	
Yes	0 (0/36)	0 (0/37)	...	
History of pulmonary tuberculosis				.98 <sup>‡</sup>
No	0.6 (28/5031)	0.3 (13/5001)	2.2 (1.1, 4.2)	
Yes	1.4 (3/207)	0 (0/237)	8.1 (0.42, 159.3)	
Presence of prior chest radiographs				.80
No	0.8 (13/1682)	0.3 (5/1714)	2.7 (0.95, 7.5)	
Yes	0.5 (18/3556)	0.2 (8/3524)	2.2 (0.97, 5.2)	
Reporting radiologist				.87 <sup>‡</sup>
Radiologist 1	0.5 (18/3387)	0.2 (7/3421)	2.6 (1.1, 6.3)	
Radiologist 2	0.7 (11/1619)	0.4 (6/1580)	1.8 (0.66, 4.9)	
Radiologist 3	1.1 (2/188)	0 (0/189)	5.1 (0.24, 107.4)	
Others	0 (0/44)	0 (0/48)	...	

Note.—Except where indicated, numbers in parentheses are numbers of participants. AI = artificial intelligence.

\* Numbers in parentheses are 95% CIs.

<sup>†</sup> P values were calculated by the significance of the interaction term added to the regression model.

<sup>‡</sup> The Firth correction was applied.

the false-referral rate (45.9% [56 of 122 participants] vs 56.0% [56 of 100 participants];  $P = .14$ ) or positive-report rate (2.3% [122 of 5238 participants] vs 1.9% [100 of 5238 participants];  $P = .14$ ) between the two groups.

A strength of this study is that it was a pioneering randomized controlled trial evaluating the actual effect of AI-based CAD in real clinical practice. Mazzone et al (25) had conducted a similar trial using non-AI CAD on high-risk participants but could not derive meaningful results, possibly due to the limited sample size (1423 enrolled participants with four actionable nodules on CT scans) and a less-efficient CAD system. Our trial was conducted with a pragmatic approach (26,27), including all 10 476 adult participants. This was possible because our institutional review board waived informed consent from

the participants, considering that the AI-based CAD system had been approved by the national Ministry of Food and Drug Safety and robustly validated in various retrospective cohorts (14,16). In addition, we integrated AI-based CAD into a commercial PACS, enabling daily practice using AI-based CAD.

The trial met the primary outcome, although the number of enrolled participants ( $n = 10\,476$ ) was smaller than the estimated sample size ( $n = 84\,000$ ) and the prevalence of actionable lung nodules was smaller than that in a previous retrospective study analyzed using those who received chest CT (16). We believe this reveals the limitations of retrospective studies. Unlike retrospective performance tests, the overall health checkup population contains a very small proportion of positive cases (about 1%), making radiologists less sensitive and less focused



and maximizing the effect of AI-based CAD. The enhanced effect of AI-based CAD in a low disease prevalence environment had been demonstrated in a retrospective simulation test (6). The detection rates of malignant lung nodules (0.15% [eight of 5238 participants] vs 0% [0 of 5238 participants];  $P = .008$ ) and lung cancer (0.11% [six of 5238 participants] vs 0% [0 of 5238 participants];  $P = .03$ ) on chest radiographs were also higher in the AI group; however, these results should be interpreted with caution due to the very low incidence of the disease. Both values showed no evidence of a difference (OR, 17.0 [95% CI: 0.98, 295.1;  $P = .05$ ] and 13.0 [95% CI: 0.73, 231.1;  $P = .08$ ], respectively) in logistic regression analyses after the Firth correction.

As a considerable proportion of participants (47%, 4886 of 10476) underwent chest CT regardless of chest radiography results, we could not evaluate whether AI-based CAD altered patient management and clinical decision-making. Therefore, we evaluated the detection rate on chest radiographs as a primary outcome rather than the diagnosis of lung nodules or lung cancer. That is, our trial purely investigated the impact of AI-based CAD on the diagnostic performance of chest radiography, not its impact on increased diagnoses of lung cancer or participants' prognoses. Nevertheless, the improved detection rate of actionable lung nodules with a similar false-referral rate suggests that using AI-based CAD may improve lung cancer diagnosis without imposing an additional radiation hazard. Meanwhile, the diagnostic performance could be more robustly evaluated owing to the high proportion of CT performance. The AI group exhibited higher sensitivity, positive predictive value, and negative predictive value, while specificity was similar between the two groups.

In the subgroup analyses, no confounding factors induced meaningful interactions with AI-based CAD in detecting actionable lung nodules on chest radiographs. Older age and a history of lung cancer or tuberculosis were associated with positive reports; however, those factors did not affect the impact of AI-based CAD. This result implies that AI-based CAD may work consistently for different populations, even for those with diseased or postoperative lungs. Age and sex showed  $P$  values of .06 and .05 for interaction, suggesting that a younger, male population could have more benefits from AI-based CAD in larger samples. Notably, the reporting radiologists showed varying positive-report rates, but the impact of AI-based CAD was similar among the radiologists ( $P = .87$ ). Therefore, AI-based CAD could be equally helpful for radiologists, regardless of their sensitivity in reporting nodules.

Our study had several limitations. First, because not all participants received chest CT, the accurate diagnostic performance of chest radiography was only assessed using a subgroup. Second, the trial was conducted at a single institution and the estimated sample size was not fulfilled (16). Third, we did not assess the standalone performance of AI-based CAD or the potential further impact of AI-based CAD in reporting prioritization or reporting time reduction. Fourth, although frequently performed, the effectiveness of chest radiographic screening in the general population has not been confirmed in large-scale trials. Last, our target population had relatively simple and dichotomized outcome (with or without nodules).

Different effects could be observed in other populations with various disease groups.

In conclusion, in a randomized controlled trial of 10486 health checkup participants, artificial intelligence–based software improved the detection of actionable lung nodules on chest radiographs with a similar false-referral rate.

**Acknowledgments:** We sincerely thank Jeongseon Lee, BS (clinical research coordinator, Seoul National University Hospital), for her assistance in the data acquisition and Hyeong Ju Bak, BS (development manager, Infinitt Healthcare), for developing and implementing the artificial intelligence–integrated picture archiving and communication system.

**Author contributions:** Guarantors of integrity of entire study, J.G.N., J.M.G.; 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, J.G.N., E.J.H., H.J.K., J.M.G.; clinical studies, J.G.N., E.J.H., E.H.L., H.J.K., M.N.; experimental studies, H.J.K., J.M.G.; statistical analysis, J.G.N., J.K., N.P.; and manuscript editing, J.G.N., E.J.H., J.H.L., C.M.P., J.M.G.

**Disclosures of conflicts of interest:** J.G.N. Research grant from VUNO. E.J.H. Research grant from Lunit. J.K. No relevant relationships. N.P. No relevant relationships. E.H.L. No relevant relationships. H.J.K. No relevant relationships. M.N. No relevant relationships. J.H.L. No relevant relationships. C.M.P. Participation on the Big Data Review Board of Seoul National University Hospital; board member of Korean Society of Radiology, Korean Society of Thoracic Radiology, and Korean Society of Artificial Intelligence in Medicine; stock in Promedius; stock options in Lunit and Coreline Soft. J.M.G. Research grants from LG Electronics and Coreline Soft; associate editor for *Radiology*.

## References

- Salim M, Wählin E, Dembrower K, et al. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncol* 2020;6(10):1581–1588.
- Wardlaw JM, Mair G, von Kummer R, et al. Accuracy of Automated Computer-Aided Diagnosis for Stroke Imaging: A Critical Evaluation of Current Evidence. *Stroke* 2022;53(7):2393–2403.
- Nam JG, Kang H-R, Lee SM, et al. Deep Learning Prediction of Survival in Patients with Chronic Obstructive Pulmonary Disease Using Chest Radiographs. *Radiology* 2022;305(1):199–208.
- Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* 2019;290(1):218–228.
- Hwang EJ, Park S, Jin K-N, et al. Development and validation of a deep learning–based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open*. 2019;2(3):e191095-e.
- Nam JG, Kim M, Park J, et al. Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. *Eur Respir J* 2021;57(5):2003061.
- Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 2018;15(11):e1002686.
- Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit Health* 2021;3(8):e496–e506.
- Oken MM, Hocking WG, Kvale PA, et al; PLCO Project Team. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial. *JAMA* 2011;306(17):1865–1873.
- Aberle DR, DeMello S, Berg CD, 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(10):920–931.
- Shankar A, Saini D, Dubey A, et al. Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward. *Transl Lung Cancer Res* 2019;8(Suppl 1):S106–S121.
- Gossner J. Lung cancer screening—don't forget the chest radiograph. *World J Radiol* 2014;6(4):116–118.
- Dominioni L, Rotolo N, Mantovani W, et al. A population-based cohort study of chest x-ray screening in smokers: lung cancer detection findings and follow-up. *BMC Cancer* 2012;12(1):18.

14. Lee JH, Sun HY, Park S, et al. Performance of a deep learning algorithm compared with radiologic interpretation for lung cancer detection on chest radiographs in a health screening population. *Radiology* 2020;297(3):687–696.
15. Kim EY, Kim YJ, Choi W-J, et al. Performance of a deep-learning algorithm for referable thoracic abnormalities on chest radiographs: A multicenter study of a health screening cohort. *PLoS One* 2021;16(2):e0246472. [Published correction appears in *PLoS One* 2021;16(4):e0251045.]
16. Nam JG, Kim HJ, Lee EH, et al. Value of a deep learning-based algorithm for detecting Lung-RADS category 4 nodules on chest radiographs in a health checkup population: estimation of the sample size for a randomized controlled trial. *Eur Radiol* 2022;32(1):213–222.
17. Nam JG, Hwang EJ, Kim DS, et al. Undetected lung cancer at posteroanterior chest radiography: Potential role of a deep learning-based detection algorithm. *Radiol Cardiothorac Imaging* 2020;2(6):e190222.
18. Jang S, Song H, Shin YJ, et al. Deep learning-based automatic detection algorithm for reducing overlooked lung cancers on chest radiographs. *Radiology* 2020;296(3):652–661.
19. Yoo H, Lee SH, Arru CD, et al. AI-based improvement in lung cancer detection on chest radiographs: results of a multi-reader study in NLST dataset. *Eur Radiol* 2021;31(12):9664–9674.
20. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ* 2020;370:m3164.
21. American College of Radiology. Lung CT Screening Reporting & Data System (Lung-RADS) v1.1. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads>. Accessed July 21, 2022.
22. 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.
23. Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: when is it 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):e93S–e120S.
24. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993;80(1):27–38.
25. Mazzone PJ, Obuchowski N, Phillips M, Risius B, Bazerbashi B, Meziane M. Lung cancer screening with computer aided detection chest radiography: design and results of a randomized, controlled trial. *PLoS One* 2013;8(3):e59650.
26. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147.
27. Peterson ED, Harrington RA. Evaluating health technology through pragmatic trials: novel approaches to generate high-quality evidence. *JAMA* 2018;320(2):137–138.