



Shifts in Pulmonary Nodule Detection After Stopping AI Assistance: A Retrospective Repeated-Measures Study

Jianlin Wu ¹, Junmiao Xiang ², Given Michael Kihaga ³, Yihu Zheng ³, Congcong Pan ⁴, Ningjian Hou ⁵, *

¹ Department of Radiology, Wencheng County People's Hospital, Wenzhou, China

² Department of Gynecology and Obstetrics, Ruian City People's Hospital, Wenzhou, China

³ Department of General Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

⁴ Department of Finance, Wencheng County People's Hospital, Wenzhou, China

⁵ Department of Health Management, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

*Corresponding Author: The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China. Email: 13587688001@163.com

Received: 1 February, 2026; Revised: 13 March, 2026; Accepted: 27 March, 2026

Abstract

Background: Artificial intelligence (AI) systems can improve pulmonary nodule detection, but there is concern that prolonged reliance on AI may alter visual search behavior and affect radiologists' independent interpretive performance when AI support is withdrawn.

Objectives: The objective of this study is to evaluate phase-associated changes in pulmonary nodule detection rate after discontinuation of routine AI assistance.

Patients and Methods: This retrospective study included 980 chest CT examinations that had been originally interpreted before AI implementation by three senior general radiologists (phase I: Baseline clinical reporting phase, during which they had not previously used any chest CT AI assistance). After approximately 26 months of routine AI use for pulmonary nodule detection, the three participating radiologists discontinued use of the chest CT AI system for this study. Each examination was reassigned to its original reporting radiologist according to report signature and independently reread without AI on the basis of the images alone (phase II), and then reread again in the same manner after a 3-month AI-free washout period (phase III). The pulmonary nodule detection rate, defined as the proportion of scans with at least one reported nodule, and the maximum diameter of the largest reported nodule were compared across phases. Because no external lesion-level reference standard was established, the findings reflect changes in reporting rather than sensitivity or specificity.

Results: The pulmonary nodule detection rate decreased from 37.8% (370/980) in phase I to 26.5% (260/980) in phase II and then increased to 43.2% (423/980) in phase III (overall $P < 0.001$). In a generalized estimating equation (GEE) model, using phase I as the reference, the adjusted odds of pulmonary nodule detection were significantly lower in phase II [adjusted odds ratio (aOR) 0.595, 95% confidence interval (CI) 0.530 - 0.667; $P < 0.001$] and significantly higher in phase III (aOR 1.253, 95% CI 1.123 - 1.398; $P < 0.001$). Phase III also showed higher adjusted odds of detection than phase II (aOR 2.106, 95% CI 1.874 - 2.366; $P < 0.001$). The phase-related difference was mainly driven by nodules with a maximum reported diameter of ≤ 5 mm.

Conclusion: Discontinuation of routine AI assistance was associated with a short-term decrease in pulmonary nodule detection rate, particularly for small nodules, followed by recovery after an AI-free washout period. These findings suggest a potential vulnerability window during AI downtime or workflow transitions and highlight the need for resilient clinical workflows and performance monitoring.

Keywords: Artificial Intelligence, Pulmonary Nodules, Computed Tomography, Radiologists, Observer Performance

1. Background

Early detection of pulmonary nodules is crucial for effective lung cancer screening. A meta-analysis, primarily involving studies conducted in Chinese populations, reported an overall detection rate of approximately 27% for pulmonary nodules on chest computed tomography (CT) scans, indicating that more than one in four screened individuals had detectable nodules (1). Studies involving experienced thoracic radiologists have shown detection sensitivities ranging

from approximately 51% to 83%, with substantial interobserver variation in nodule counts and false-positive rates in large-scale screening programs. These findings suggest that pulmonary nodule detection is influenced not only by nodule characteristics but also by radiologists' interpretive skills, experience, and diagnostic performance (2-4). In recent years, artificial intelligence (AI), particularly deep learning-based algorithms, has emerged as a promising tool in medical imaging. Systematic reviews have shown that AI-assisted diagnostic models generally achieve higher sensitivity

and accuracy than manual interpretation alone, significantly reducing missed nodules and interobserver variability (5).

Despite these advantages, the widespread integration of AI into radiologic practice has raised concerns regarding its long-term impact on physicians' diagnostic competence. This issue is particularly relevant as AI technologies are rapidly being adopted in clinical settings. A review published in Artificial Intelligence Review noted that, although AI can enhance clinical decision-making performance, it may also carry a potential risk of clinical skill degradation among physicians (6).

2. Objectives

The objective of this study is to evaluate phase-associated changes in pulmonary nodule detection rate after discontinuation of routine AI assistance using a three-phase repeated-measures observer design, with a primary analysis that accounted for within-scan correlation.

3. Patients and Methods

3.1. Study Design and Setting

This was a single-center retrospective repeated-measures observer study using archived chest CT examinations. Phase I data were derived from routine baseline clinical reports before AI implementation. Phases II and III consisted of controlled image-only rereads performed without AI assistance, separated by a 3-month AI-free washout period during which the participating radiologists continued routine clinical work.

This study was approved by the Ethics Committee of Wencheng County People's Hospital, Zhejiang province. Given its retrospective design and use of de-identified imaging and diagnostic data, the requirement for written informed consent was waived. This study was reported in accordance with the STROBE statement for observational studies.

3.2. Participants

All chest CT examinations performed between March 1 and June 30, 2023, and independently reported by three senior general radiologists (10 - 15 years of experience in comprehensive radiology practice and high-volume routine reading) were screened. Examinations were derived from routine clinical workflows, including health screening, emergency department, outpatient, and inpatient settings; no age

restrictions were applied to reflect the real-world case mix. Of 1,244 examinations initially identified, 264 were excluded because of incomplete imaging data, missing reports, substantial motion or metal artifacts, or diffuse lung disease precluding reliable nodule assessment. The final dataset comprised 980 examinations, including 48 examinations from minors (< 18 years). A small number of patients underwent repeat examinations (975 unique patients and 980 scans). Each CT examination was treated as an independent observation, while repeated measurements across phases for the same scan were accounted for using the generalized estimating equation (GEE) model. The study flowchart is shown in Figure 1. Before AI implementation (phase I), these radiologists had not used any AI-assisted chest CT tools in routine practice.

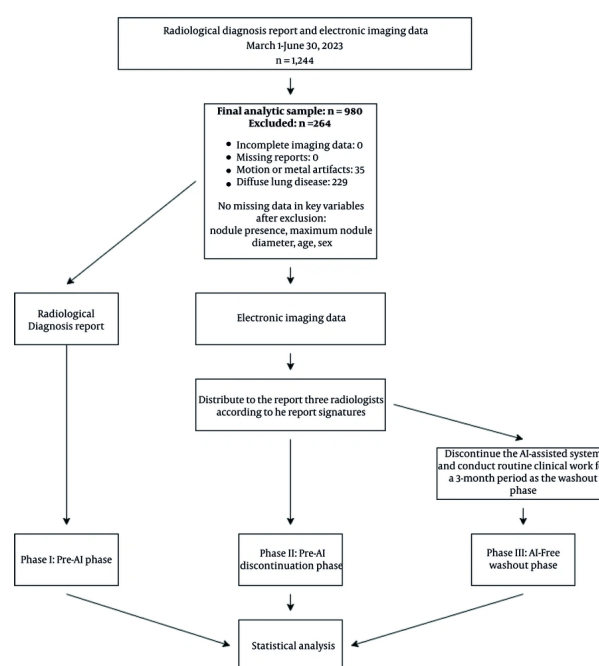


Figure 1. Study flowchart showing case screening, inclusion and exclusion across the three study phases

3.3. CT Acquisition Parameters

All CT images were acquired using a 16-slice multidetector scanner (GE Brightspeed, GE Healthcare, USA) and a 32-slice multidetector scanner (SOMATOM go.All, Siemens Healthineers, Germany) under a standard chest CT protocol. The scan range extended from the thoracic inlet to below the lung bases. Scans were acquired during a single breath-hold at full

inspiration in the craniocaudal direction, without intravenous contrast administration. Images were reconstructed into thin-section lung-window images (slice thickness 1.25/1.5 mm; slice interval 1.25/1.5 mm; window width 1200 HU; window level -500 to -600 HU).

3.4. AI System and Routine Workflow

The AI system used in this study was a commercial deep learning-based pulmonary nodule detection tool (Deepwise Medical, China), which was implemented at our hospital for routine chest CT interpretation on July 1, 2023. AI prompts were displayed concurrently during image review, and radiologists could decide at their discretion whether to accept or disregard the suggestions. For the purposes of this evaluation, the three participating radiologists discontinued routine AI use in their daily work starting on September 1, 2025.

3.5. Study Procedure and Blinding

In phase I (pre-AI baseline phase), scan-level nodule reporting (≥ 1 nodule: Yes/no) and the maximum diameter of the largest reported nodule were extracted from original routine clinical reports generated before AI implementation. In phase II (post-AI discontinuation phase), after the three participating radiologists discontinued use of the chest CT AI system for this evaluation, each scan was reassigned to its original reporting radiologist on the basis of the report signature. The radiologists independently reread the images without AI assistance and recorded nodule presence and maximum diameter. All available CT slices of each examination were reviewed (whole-scan assessment), rather than selecting a single section per patient. The same whole-scan assessment approach was applied consistently across phases I - III. In phase III (AI-free washout phase), after a 3-month AI-free washout period during routine clinical work, during which chest CT AI assistance remained unavailable to these radiologists and their average chest CT workload exceeded 50 examinations per day, the radiologists repeated the same image-only reread procedure without AI. To mitigate recall bias, the case order was re-randomized before phase III. During rereading, the radiologists were blinded to the original reports and prior phase results and did not have access to additional clinical history or prior imaging. The interval between phase I baseline reporting (March - June 2023) and phase II rereading (after September 1, 2025) exceeded two years, further reducing the likelihood of case recall.

3.6. Outcome Definitions

Pulmonary nodules were defined as focal rounded or irregular pulmonary opacities (solid or subsolid, including pure ground-glass and part-solid components). The primary outcome was scan-level reporting of at least one pulmonary nodule (binary outcome: Event = 1 if ≥ 1 nodule was reported). For scans with multiple reported nodules, the maximum diameter of the largest reported nodule was recorded. For stratified analyses by nodule size, categories (≤ 5 mm, $> 5 - 8$ mm, > 8 mm) were defined separately within each phase based on the maximum reported diameter recorded in that phase. Because no external lesion-level reference standard was established, the term “detection rate” refers to the proportion of scans reported as positive and should not be interpreted as sensitivity or specificity.

3.7. Bias and Sample Size

Potential sources of bias included: (1) Setup differences between Phase I routine reporting and Phase II/III research rereads, such as the absence of clinical context (7); (2) Hawthorne effects and motivational differences during rereading (8); (3) practice effects during the 3-month washout period due to high reading volume; and (4) residual familiarity with cases despite re-randomization. The study size was determined by the number of eligible examinations within the prespecified time window; no a priori power calculation was performed.

3.8. Statistical Analysis

All analyses were performed using SPSS (version 24.0, IBM, Armonk, NY, USA). Categorical variables are summarized as counts (percentages). Continuous variables are summarized as mean $\pm$ standard deviation or median (Q1 - Q3), as appropriate. All tests were two-sided. For post-hoc pairwise phase comparisons, a Bonferroni-adjusted significance threshold of $P < 0.017$ ($0.05/3$) was used.

The primary analysis used a population-averaged logistic GEE model (binomial family, logit link) to evaluate phase-related changes in pulmonary nodule detection (event = 1). The model specified scan_{id} as the clustering variable, an exchangeable working correlation structure, and robust (sandwich) standard errors. Phase (categorical) and radiologist (reader_{id}) were included as covariates. Results are reported as aORs with 95% CIs, where aOR denotes adjusted odds ratio. Pairwise phase comparisons were derived from the GEE model with Bonferroni correction for multiple comparisons.

As supplementary descriptive analyses of repeated measures, overall differences in binary outcomes across phases were assessed using Cochran's Q test; when significant, post-hoc pairwise comparisons were performed using McNemar's test with Bonferroni adjustment. Maximum nodule diameter was compared across phases using the Friedman test; when significant, pairwise comparisons were performed using the Wilcoxon signed-rank test with Bonferroni adjustment. Sensitivity analyses refitted the same GEE model after (1) randomly retaining one scan per patient while keeping all corresponding phase records and (2) excluding minors (<18 years) to assess robustness.

4. Results

4.1. Study Population

The final analysis included 980 chest CT examinations from 975 unique patients, including 5 patients who underwent two examinations. The cohort comprised 494 male and 486 female patients. The mean age was 57.94 ± 19.34 years (range, 3 - 97 years), and the median age was 60 years (Q1 - Q3: 46 - 73 years). Of all examinations, 485 were from patients aged < 60 years and 495 from patients aged ≥ 60 years. Pediatric examinations (< 18 years) accounted for 4.9% (48/980) (Table 1).

Table 1. Study Population Characteristics ^a

Variables	Values
Chest CT examinations included	980
Unique patients	975
Patients with two examinations	5
Male/female patients	494/486
Age, y	57.94 ± 19.34
Median age (Q1-Q3), y	60 (46 - 73)
Age range, y	3 - 97
Patients aged < 60/ ≥ 60 y	485/495
Pediatric examinations (<18 y)	48/980 (4.9%)

Abbreviation: CT, computed tomography.

^a Values are presented as median (Q1-Q3) or mean $\pm$ standard deviation (SD) or number as appropriate (Q1-Q3, first to third quartiles).

4.2. Phase-Associated Changes in Scan-Level Pulmonary Nodule Detection

Pulmonary nodule detection rates differed significantly across the three phases. The scan-level detection rate was 37.8% (370/980; 95% CI 34.8 - 40.8) in phase I, 26.5% (260/980; 95% CI 23.9 - 29.4) in phase II, and

43.2% (423/980; 95% CI 40.1 - 46.3) in phase III (overall $P < 0.001$) (Table 2).

The absolute differences were -11.2 percentage points (95% CI, -13.7 to -8.8) for phase II versus phase I, +5.4 percentage points (95% CI, +2.8 to +8.0) for phase III versus phase I, and +16.6 percentage points (95% CI, +14.1 to +19.2) for phase III versus phase II (Table 3).

4.3. Primary GEE Analysis

In the primary population-averaged logistic GEE model adjusted for radiologist and accounting for within-scan correlation, pulmonary nodule detection was significantly lower in phase II than in phase I (aOR 0.595, 95% CI 0.530 - 0.667; $P < 0.001$). In contrast, detection was significantly higher in phase III than in phase I (aOR 1.253, 95% CI 1.123 - 1.398; $P < 0.001$). phase III also showed significantly higher adjusted odds of detection than phase II (aOR 2.106, 95% CI 1.874 - 2.366; $P < 0.001$) (Table 4). Overall, these findings suggest a short-term decline in scan-level pulmonary nodule detection immediately after AI discontinuation, followed by recovery after the AI-free washout period.

4.4. Scan-Level Agreement and Discordance Across Phase Pairs

To further characterize pairwise changes between phases, we summarized scan-level agreement and discordance for each phase pair based on the presence or absence of at least one reported pulmonary nodule on a given examination. Between phase I and phase II, 816/980 examinations were concordant and 164/980 were discordant, including 137 examinations changing from positive to negative (1 $\rightarrow$ 0) and 27 changing from negative to positive (0 $\rightarrow$ 1). Between phase I and phase III, 805/980 examinations were concordant and 175/980 were discordant, including 61 examinations changing from positive to negative and 114 changing from negative to positive. Between phase II and phase III, 791/980 examinations were concordant and 189/980 were discordant, including 13 examinations changing from positive to negative and 176 changing from negative to positive. These agreement and discordance counts were based on scan-level classification rather than lesion-by-lesion matching (Supplementary Table S1 and S2 in Supplementary File).

4.5. Stratified Analyses by Sex, Age, and Nodule Size

A similar phase-associated pattern was observed across sex and age strata. In male patients, pulmonary nodule detection rates were 40.7% (201/494) in phase I, 29.6% (146/494) in phase II, and 45.1% (223/494) in phase

Table 2. Phase-Specific Scan-Level Pulmonary Nodule Detection Rates Stratified by Sex, Age, and Maximum Reported Nodule Diameter ^{a,b}

Characteristic	Phase I: Pre-AI baseline	Phase II: Post-AI discontinuation	Phase III: AI-free washout	Cochran's Q (chi-square)	P-value
Pulmonary nodule detection	370/980 (37.8%) ^A ; [95% CI 34.8 - 40.8]	260/980 (26.5%) ^B ; [95% CI 23.9 - 29.4]	423/980 (43.2%) ^C ; [95% CI 40.1 - 46.3]	157.114	< 0.001
Male	201/494 (40.7%) ^A ; [95% CI 36.4 - 45.1]	146/494 (29.6%) ^B ; [95% CI 25.7 - 33.7]	223/494 (45.1%) ^A ; [95% CI 40.8 - 49.6]	74.905	< 0.001
Female	169/486 (34.8%) ^A ; [95% CI 30.7 - 39.1]	114/486 (23.5%) ^B ; [95% CI 19.9 - 27.4]	200/486 (41.2%) ^C ; [95% CI 36.9 - 45.6]	82.478	< 0.001
Age < 60 (y)	167/485 (34.4%) ^A ; [95% CI 30.3 - 38.8]	109/485 (22.5%) ^B ; [95% CI 19.0 - 26.4]	192/485 (39.6%) ^C ; [95% CI 35.3 - 44.0]	83.038	< 0.001
Age ≥ 60 (y)	203/495 (41.0%) ^A ; [95% CI 36.8 - 45.4]	151/495 (30.5%) ^B ; [95% CI 26.6 - 34.7]	231/495 (46.7%) ^C ; [95% CI 42.3 - 51.1]	74.346	< 0.001
Maximum reported nodule diameter ≤ 5 mm	260/980 (26.5%) ^A ; [95% CI 23.9 - 29.4]	155/980 (15.8%) ^B ; [95% CI 13.7 - 18.2]	306/980 (31.2%) ^C ; [95% CI 28.4 - 34.2]	142.063	< 0.001
Maximum reported nodule diameter > 5 to ≤ 8 mm	73/980 (7.4%); [95% CI 6.0 - 9.3]	69/980 (7.0%); [95% CI 5.6 - 8.8]	77/980 (7.9%); [95% CI 6.3 - 9.7]	2.526	0.283
Maximum reported nodule diameter > 8 mm	37/980 (3.8%); [95% CI 2.8 - 5.2]	36/980 (3.7%); [95% CI 2.7 - 5.0]	40/980 (4.1%); [95% CI 3.0 - 5.5]	1.625	0.444

Abbreviation: AI, artificial intelligence; CI, confidence interval.

^a Values are expressed as n/N (%).

^b Different superscript letters within a row indicate significant pairwise differences based on post-hoc McNemar tests using a Bonferroni-corrected significance threshold of $P < 0.017$; phases sharing the same letter are not significantly different.

Table 3. Absolute Phase Differences in Scan-Level Pulmonary Nodule Detection ^a

Comparison	Absolute difference (percentage points)	95% CI
Phase II vs phase I	-11.2	-13.7 to -8.8
Phase III vs phase I	+5.4	+2.8 to +8.0
Phase III vs phase II	+16.6	+14.1 to +19.2

Abbreviation: CI, confidence interval.

^a Absolute differences in scan-level pulmonary nodule detection are presented as percentage-point differences with 95% confidence intervals (CIs), derived from paired scan-level comparisons of the same 980 examinations.

III. In female patients, the corresponding rates were 34.8% (169/486), 23.5% (114/486), and 41.2% (200/486). Among patients aged <60 years, detection rates were 34.4% (167/485), 22.5% (109/485), and 39.6% (192/485) across phases I, II, and III, respectively. Among those aged ≥ 60 years, the corresponding rates were 41.0% (203/495), 30.5% (151/495), and 46.7% (231/495). In all of these strata, the overall phase-associated difference was significant. Post-hoc comparisons showed that phase II was significantly lower than both phase I and phase III in all sex and age strata. phase I and phase III did not significantly differ in male patients, but differed significantly in female patients and in both age strata (Table 2).

When stratified by maximum reported nodule diameter, the phase effect was mainly driven by small nodules. For scans in which the maximum reported nodule diameter was ≤ 5 mm, the corresponding

detection rates were 26.5% (260/980), 15.8% (155/980), and 31.2% (306/980) in phases I, II, and III, respectively, with a significant overall phase difference. Post-hoc comparisons showed that all three phases differed significantly from one another. No significant phase-associated differences were observed for nodules with a maximum diameter >5 to ≤ 8 mm or >8 mm (Table 2).

Further stratified analysis of scans with a maximum reported nodule diameter of ≤ 5 mm showed the same general pattern across sex and age groups. In men, the rates were 27.1%, 16.4%, and 30.8% across phases I, II, and III; in women, 25.9%, 15.2%, and 31.7%; in patients aged <60 years, 28.0%, 16.1%, and 32.8%; and in those aged ≥ 60 years, 25.1%, 15.6%, and 29.7%, respectively. In each subgroup, phase II was significantly lower than phases I and III (Table 5).

4.6. Changes in Maximum Reported Nodule Diameter

Table 4. GEE-Adjusted Odds Ratios for Pulmonary Nodule Detection by Phase ^a

Comparison	aOR	95% CI	P-value
Phase II vs phase I	0.595	0.530 - 0.667	< 0.001
Phase III vs phase I	1.253	1.123 - 1.398	< 0.001
Phase III vs phase II	2.106	1.874 - 2.366	< 0.001

Abbreviations: aOR, adjusted odds ratio; CI, confidence interval; GEE, generalized estimating equation.

^a Population-averaged logistic GEE model adjusted for radiologist, with scan_id as the clustering variable, an exchangeable working correlation structure, and robust standard errors. Values are presented as aORs with 95% CIs. An aOR >1 indicates higher odds of pulmonary nodule detection relative to the reference phase.

Table 5. Stratified Analysis of Scans with Maximum Reported Nodule Diameter ≤ 5 mm ^{a,b}

Subgroup	Phase I: Pre-AI baseline	Phase II: Post-AI discontinuation	Phase III: AI-free washout	Cochran's Q (chi-square)	P-value
Male	134/494 (27.1%) ^A	81/494 (16.4%) ^B	152/494 (30.8%) ^A	67.554	< 0.001
Female	126/486 (25.9%) ^A	74/486 (15.2%) ^B	154/486 (31.7%) ^C	74.909	< 0.001
Age < 60 (y)	136/485 (28.0%) ^A	78/485 (16.1%) ^B	159/485 (32.8%) ^C	80.415	< 0.001
Age ≥ 60 (y)	124/495 (25.1%) ^A	77/495 (15.6%) ^B	147/495 (29.7%) ^A	62.098	< 0.001

Abbreviation: AI, artificial intelligence.

^a Values are expressed as n/N (%).

^b Different superscript letters within a row indicate significant pairwise differences based on post-hoc McNemar tests using a Bonferroni-corrected significance threshold of $P < 0.017$; phases sharing the same letter are not significantly different.

Among scans with reported nodules, the maximum reported nodule diameter differed significantly across phases. The median maximum diameter was 4 mm (Q1-Q3: 3-6) in phase I, 5 mm (4-7) in phase II, and 4 mm (3-6) in phase III. Post-hoc comparisons indicated that phase II differed significantly from both phase I and phase III, whereas phases I and III did not significantly differ. This pattern was also observed in both sex strata and both age groups (Table 6).

4.7. Sensitivity Analyses

Sensitivity analyses yielded results consistent with the primary model. After randomly retaining one scan per patient, the adjusted odds ratios were 0.597 (95% CI 0.532-0.669; $P < 0.001$) for phase II versus phase I, 1.249 (95% CI 1.119-1.393; $P < 0.001$) for phase III versus phase I, and 2.093 (95% CI 1.863-2.351; $P < 0.001$) for phase III versus phase II. After excluding minors, the corresponding estimates were 0.585 (95% CI 0.519-0.658; $P < 0.001$), 1.248 (95% CI 1.114-1.397; $P < 0.001$), and 2.134 (95% CI 1.894-2.405; $P < 0.001$) for the same comparisons, respectively. These findings support the robustness of the main results (Supplementary Table S2).

5. Discussion

This study observed a significant decline-rebound pattern in scan-level pulmonary nodule detection after

discontinuation of routine AI assistance. Among 980 chest CT examinations reread by the same group of radiologists, the proportion of scans with at least one reported pulmonary nodule decreased from 37.8% in phase I to 26.5% in phase II and then increased to 43.2% in phase III. The GEE analysis showed lower adjusted odds of detection in phase II than in phase I, followed by higher odds in phase III than in both earlier phases. Because no external lesion-level reference standard was available, these findings should be interpreted as phase-associated changes in scan-level detection patterns rather than definitive changes in lesion-level sensitivity or specificity.

The phase-related difference was mainly driven by nodules with a maximum reported diameter of ≤ 5 mm. Consistent with this pattern, the median maximum reported nodule diameter increased from 4 mm in phase I to 5 mm in phase II and then returned to 4 mm in phase III. This pattern suggests that the short-term decline after AI withdrawal was concentrated in low-salience targets, whereas larger nodules were less affected. However, because no lesion-level reference standard was available, this finding should be interpreted cautiously: It may reflect reduced reporting of subtle nodules, altered reporting thresholds, or both, rather than directly measured false-negative lesions. Prior computer-aided detection (CAD)/AI studies have

Table 6. Changes in Maximum Reported Nodule Diameter Across Phases ^{a,b}

Characteristic	Phase I: Pre-AI baseline	Phase II: Post-AI discontinuation	Phase III: AI-free washout	Friedman χ^2	P-value
Maximum reported nodule diameter, mm	4 (3 - 6) ^A	5 (4 - 7) ^B	4 (3 - 6) ^A	28.660	< 0.001
Male	5 (4 - 6) ^A	5 (4 - 7) ^B	5 (4 - 6) ^A	14.217	0.001
Female	4 (3 - 6) ^A	5 (4 - 7) ^B	4 (3 - 5) ^A	14.208	0.001
Age < 60 (y)	4 (3 - 5) ^A	5 (4 - 6) ^B	4 (3 - 5) ^A	13.604	0.001
Age ≥ 60 (y)	5 (4 - 7) ^A	5 (4 - 8) ^B	5 (4 - 7) ^A	13.784	0.001

Abbreviation: AI, artificial intelligence.

^a Data are presented as median (Q1-Q3), in millimeters. Q1-Q3, first to third quartiles.

^b Different superscript letters within a row indicate significant pairwise differences based on post-hoc Wilcoxon signed-rank tests using a Bonferroni-corrected significance threshold of $P < 0.017$; phases sharing the same letter are not significantly different.

similarly shown that AI support is particularly helpful for subtle and very small nodules (9).

Several factors may explain this short-term decline after AI withdrawal. Prolonged exposure to AI prompts may induce automation bias or imperfect trust calibration, reducing the intensity of independent verification when AI is abruptly removed (10, 11). Artificial intelligence use may also alter visual search behavior, shifting readers toward more cue-driven or confirmatory reading, while the absence of AI may transiently raise the reporting threshold for subtle findings. At the same time, non-AI-related explanations must also be considered. phase I consisted of routine clinical reports, whereas phases II and III were structured rereads performed without clinical context or prior imaging, and these setup differences alone may have affected both perception and reporting behavior (7). Hawthorne effects, motivational differences, and practice effects during the 3-month AI-free washout period may also have contributed (8, 12, 13). Thus, the present findings support an association between AI withdrawal and short-term performance shifts, but they do not establish a single causal mechanism.

The rebound in phase III is also noteworthy. It may reflect a combination of implicit learning, perceptual recalibration, and threshold adjustment, although these interpretations remain speculative (9, 12-14). Clinically, this finding suggests a potential vulnerability window during AI downtime, software migration, hardware replacement, or workflow transitions, when subtle nodules may be less likely to be reported. Although nodules < 6 mm usually have low malignant potential in low-risk adults, stable detection of such findings may still be important in high-risk populations, screening settings, and longitudinal follow-up (15).

Several limitations should be acknowledged. First, phase I data were derived from routine clinical reports,

whereas phases II and III were research rereads performed under controlled conditions without access to the original reports, prior imaging, or additional clinical context; these setup differences may have influenced both perception and reporting thresholds (7). Second, the same case set was re-evaluated in phases II and III; despite the 3-month interval and re-randomization before phase III, residual familiarity, Hawthorne effects, and practice effects cannot be fully excluded (8). Third, this was a single-center study involving only three senior general radiologists, one commercial AI system, and a single local workflow, which may limit generalizability to other institutions, reader populations, AI tools, workflows, or screening settings. In particular, these findings may not be directly generalizable to residents or early-career readers exposed to AI from the beginning of training, whose adaptation to AI withdrawal may differ. Fourth, no external lesion-level reference standard was established; accordingly, the primary endpoint was scan-level reporting of at least one pulmonary nodule rather than lesion-level sensitivity or specificity. Fifth, a small number of patients contributed more than one scan, although sensitivity analyses supported the robustness of the main findings. Finally, the subgroup analyses were exploratory in nature and should therefore be interpreted with caution, even though correction for multiple comparisons was applied.

In conclusion, discontinuation of routine AI assistance was associated with a short-term decrease in scan-level pulmonary nodule detection, particularly for small nodules, followed by recovery after an AI-free washout period. These findings suggest a potential vulnerability window during AI withdrawal or workflow transitions and highlight the need for resilient implementation strategies, reader monitoring, and

further prospective studies incorporating lesion-level reference standards.

Supplementary Material

Supplementary material(s) is available [here](#) [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Footnotes

AI Use Disclosure: The authors declare that no generative AI tools were used in the creation of this article.

Authors' Contribution: Study concept and design: Y. Z. and J. W.; Acquisition of data: J. W.; Analysis and interpretation of data: J. W.; Drafting of the manuscript: J. W.; Critical revision of the manuscript for important intellectual content: Y. Z.; Statistical analysis: J. X., C. P.; Administrative, technical, and material support: N. H.; Study supervision: N. H.; Translation: G. K.

Conflict of Interests Statement: The authors declare no conflicts of interest.

Data Availability: The dataset presented in the study is available on request from the corresponding author during submission or after publication.

Ethical Approval: This study was approved by the Ethics Committee of Wencheng County People's Hospital, Zhejiang province, China [approval No.: Wencheng County People's Hospital Ethics Committee (2026) No. 01].

Funding/Support: The present study received no funding/support.

Informed Consent: Written informed consent was waived due to the retrospective design and the use of de-identified imaging and diagnostic data.

References

- Chen D, Yang L, Zhang W, Shen J, Van Schil PEY, Divisi D, et al. Prevalence and management of pulmonary nodules: a systematic review and meta-analysis. *J Thorac Dis.* 2024;**16**(7):4619-32. [PubMed ID: 39144359]. [PubMed Central ID: PMC11320231]. <https://doi.org/10.21037/jtd-24-874>.
- Armato S3, Roberts RY, Kocherginsky M, Aberle DR, Kazerooni EA, Macmahon H, et al. Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of "truth". *Acad Radiol.* 2009;**16**(1):28-38. [PubMed ID: 19064209]. [PubMed Central ID: PMC2658894]. <https://doi.org/10.1016/j.acra.2008.05.022>.
- Pinsky PF, Gierada DS, Nath PH, Kazerooni E, Amorosa J. National lung screening trial: variability in nodule detection rates in chest CT studies. *Radiology.* 2013;**268**(3):865-73. [PubMed ID: 23592767]. [PubMed Central ID: PMC3750416]. <https://doi.org/10.1148/radiol.13121530>.
- Ming S, Yang W, Cui SJ, Huang S, Gong XY. Consistency of radiologists in identifying pulmonary nodules based on low-dose computed tomography. *J Thorac Dis.* 2019;**11**(7):2973-80. [PubMed ID: 31463127]. [PubMed Central ID: PMC6687997]. <https://doi.org/10.21037/jtd.2019.07.52>.
- Cheo HM, Ong CYG, Ting Y. A Systematic Review of AI Performance in Lung Cancer Detection on CT Thorax. *Healthcare (Basel).* 2025;**13**(13). [PubMed ID: 40648536]. [PubMed Central ID: PMC12250385]. <https://doi.org/10.3390/healthcare13131510>.
- Natali C, Marconi L, Dias Duran LD, Cabitza F. AI-induced Deskilling in Medicine: A Mixed-Method Review and Research Agenda for Healthcare and Beyond. *Artif Intell Rev.* 2025;**58**(11). <https://doi.org/10.1007/s10462-025-11352-1>.
- Hattori S, Yokota H, Takada T, Horikoshi T, Takishima H, Mikami W, et al. Impact of clinical information on CT diagnosis by radiologist and subsequent clinical management by physician in acute abdominal pain. *Eur Radiol.* 2021;**31**(8):5454-63. [PubMed ID: 33515087]. <https://doi.org/10.1007/s00330-021-07700-8>.
- McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol.* 2014;**67**(3):267-77. [PubMed ID: 24275499]. [PubMed Central ID: PMC3969247]. <https://doi.org/10.1016/j.jclinepi.2013.08.015>.
- Sahiner B, Chan HP, Hadjiiski LM, Cascade PN, Kazerooni EA, Chughtai AR, et al. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: analysis of an observer performance study by nodule size. *Acad Radiol.* 2009;**16**(12):1518-30. [PubMed ID: 19896069]. [PubMed Central ID: PMC2810535]. <https://doi.org/10.1016/j.acra.2009.08.006>.
- Dratsch T, Chen X, Rezazade Mehrizi M, Kloeckner R, Mahringer-Kunz A, Pusken M, et al. Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. *Radiology.* 2023;**307**(4). e222176. [PubMed ID: 37129490]. <https://doi.org/10.1148/radiol.222176>.
- Khera R, Simon MA, Ross JS. Automation Bias and Assistive AI: Risk of Harm From AI-Driven Clinical Decision Support. *JAMA.* 2023;**330**(23):2255-7. [PubMed ID: 38112824]. <https://doi.org/10.1001/jama.2023.22557>.
- Waite S, Grigorian A, Alexander RG, Macknik SL, Carrasco M, Heeger DJ, et al. Analysis of Perceptual Expertise in Radiology - Current Knowledge and a New Perspective. *Front Hum Neurosci.* 2019;**13**:213. [PubMed ID: 31293407]. [PubMed Central ID: PMC6603246]. <https://doi.org/10.3389/fnhum.2019.00213>.
- Van De Luecht MR, Reed WM. The cognitive and perceptual processes that affect observer performance in lung cancer detection: a scoping review. *J Med Radiat Sci.* 2021;**68**(2):175-85. [PubMed ID: 33556995]. [PubMed Central ID: PMC8168065]. <https://doi.org/10.1002/jmrs.456>.
- Auffermann WF, Little BP, Tridandapani S. Teaching search patterns to medical trainees in an educational laboratory to improve perception of pulmonary nodules. *J Med Imaging (Bellingham).* 2016;**3**(1):11006. [PubMed ID: 26870749]. [PubMed Central ID: PMC4748144]. <https://doi.org/10.1117/1.JMI.3.1.011006>.
- MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology.* 2017;**284**(1):228-43. [PubMed ID: 28240562]. <https://doi.org/10.1148/radiol.2017161659>.