

Original Paper

Measuring the Impact of AI on Report-Drafting Efficiency in Chest Computed Tomography Interpretation: Retrospective Analysis

Weiqli Liu^{1,2*}, DrPH; You Wu^{3*}, PhD; Wei Yu⁴, PhD; Mark J Bittle¹, DrPH; Zhuozhao Zheng⁵, PhD; Hadi Kharrazi¹, PhD, MD

¹Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

²Department of Research, Sophmind Technology (Beijing) Co. Ltd, Beijing, China

³Institute for Hospital Management, School of Medicine, Tsinghua University, Beijing, China

⁴Department of Radiology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

⁵Department of Radiology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China

*these authors contributed equally

Corresponding Author:

Zhuozhao Zheng, PhD

Department of Radiology

Beijing Tsinghua Changgung Hospital, School of Clinical Medicine

Tsinghua University

168 Litang Road

Changping District

Beijing, 102218

China

Phone: 86 010 56112345

Email: zzza00509@btch.edu.cn

Abstract

Background: Artificial intelligence (AI), particularly deep learning, has shown promise in enhancing medical image interpretation and improving radiologists' efficiency. In China, growing imaging demand and workforce shortages have placed increasing pressure on radiology services. However, evidence on the operational impact of AI on reporting efficiency remains limited.

Objective: This study aimed to evaluate the effect of an AI system on radiologists' reporting efficiency by examining changes in report-drafting time for lung nodule diagnosis in chest computed tomography (CT) images.

Methods: We analyzed 185,044 chest CT reports from Beijing Anzhen and Tsinghua Changgung Hospitals (2018-2023) using a difference-in-differences design with nonequivalent comparison groups. Report-drafting time before, immediately after, and up to 2 years following AI implementation was compared, adjusting for radiologist gender, seniority, and years of working experience.

Results: The pooled analysis showed a modest overall increase of 0.86 minutes (95% CI 0.14 to 1.57). However, this masked substantial heterogeneity between hospitals due to differing implementation timelines. In the first year after AI deployment, Tsinghua Changgung Hospital experienced a nonsignificant increase of 0.90 minutes (95% CI -0.28 to 2.08). In contrast, at Beijing Anzhen Hospital, the AI-assisted group exhibited an absolute reduction of 0.76 minutes by the first year and a further 1.83-minute reduction by the second year (an approximate 28% time saved vs baseline), while the control group remained stable over time. Using a difference-in-differences framework, this corresponded to a 2.66-minute relative improvement compared with the counterfactual trend ($P < .001$).

Conclusions: AI-assisted lung nodule diagnosis may initially increase report-drafting time due to adaptation and workflow adjustment. Sustained, meaningful efficiency gains were heterogeneous and observed at only 1 of the 2 study sites, indicating that long-term impacts are strongly contingent on site-specific implementation dynamics, learning curves, and local context.

(*J Med Internet Res* 2026;28:e77967) doi: [10.2196/77967](https://doi.org/10.2196/77967)

KEYWORDS

lung nodule; chest computed tomography; radiology; artificial intelligence; AI; efficiency; clinical workflow evaluation

Introduction

Artificial intelligence (AI) can imitate human intelligence, perform tasks, and make progressive enhancements based on amassed data [1]. Particularly, the application of deep learning technology in medical image segmentation and recognition has brought revolutionary changes to the operation of imaging services [2]. The increasing adoption of AI has the potential to alter the working conditions of radiologists [3].

The rapid evolution of radiological modalities has resulted in a substantial increase in the volume of medical images. While this progress has numerous advantages, it has significantly increased radiologists' workloads [4,5]. For example, workplace burnout could weaken radiologists' perceived competence [6], and occupational fatigue may compromise radiologists' diagnostic accuracy, thereby adversely affecting patient care and treatment outcomes [7]. Therefore, it is critical to explore how AI systems can enhance the efficiency of radiologists for more effective applications in clinical practice.

Globally, the scarcity of physicians persists due to the substantial professional requirements and protracted talent development time. This challenge is exacerbated in China, where high patient demand contrasts with inadequate investments in medical education, relatively low remuneration for health care practitioners, and enduring disproportions in medical resource allocation [8]. AI stands as a pivotal technological approach that could potentially alleviate the scarcity of medical resources in China. Furthermore, supportive governmental policies are poised to contribute significantly to the advancement of AI within the Chinese health care sector [9].

The impact of AI systems on the operational efficiency of radiologists remains a topic of uncertainty [10]. Within the medical and scientific communities, there is a prevailing belief that the rapid development of AI in health care will lead to a transformative paradigm shift in medical imaging. Enhancing operational efficiency has been deemed part of the AI's paradigm shift in medical imaging [11]; however, empirical evidence is limited and sometimes conflicting [12,13]. For example, Yi et al [14] unveiled that the implementation of an AI system for breast cancer detection resulted in a substantial reduction in reporting time. Stephan et al [15] reported improved

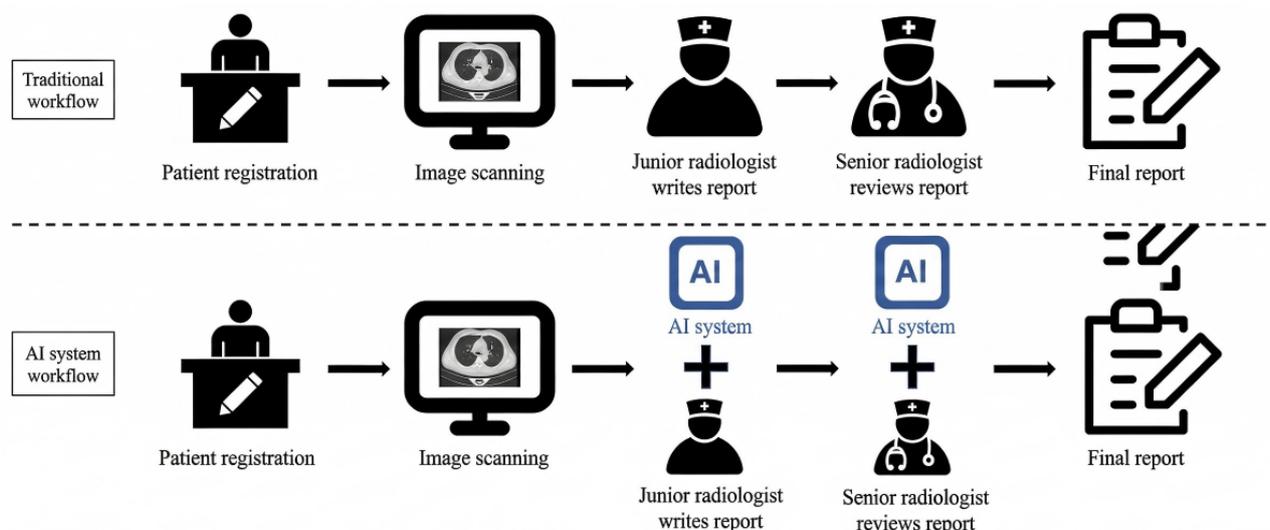
efficiency in generating dental radiology reports, while Zia et al [16] found that implementing an AI system for intracranial hemorrhage detection led to longer turnaround time. Nonetheless, it is essential to acknowledge that these findings may not provide a definitively representative picture, primarily due to the limited sample sizes of such studies [12]. Additionally, the results of these studies often do not provide a comprehensive understanding of the enduring implications of AI system adoption in medical imaging, as the study periods have been relatively brief [12].

This real-world study was conducted in 2 large hospitals in China, using data spanning multiple years and the difference-in-differences (DID) methodology. The study assesses the alterations in report-drafting time for lung nodule diagnosis in computed tomography (CT) scans after AI adoption. Results of this study can inform the radiological community on the prospective advantages and challenges associated with AI integration in medical imaging.

Methods

Study Population

This study was conducted at 2 tertiary health care facilities situated in Beijing, China: Beijing Anzhen Hospital and Tsinghua Changgung Hospital. Figure 1 illustrates the diagnostic process in the radiology departments both before and after the introduction of the AI system. In the traditional workflow, before the AI system's implementation, the typical process involved a junior radiologist writing a diagnostic report after each patient underwent a CT examination. Subsequently, a senior radiologist would review and finalize this report before it was released to clinicians and patients. With the integration of the AI system, both junior and senior radiologists can use the findings of the AI systems as diagnostic aids. The AI system will automatically identify and analyze lung nodules in imported chest CT images. Junior and senior radiologists can reference the diagnostic results provided by the AI system during their diagnostic process and ultimately complete their diagnostic reports. Case assignment follows a randomized queue-based workflow in both hospitals, without systematic allocation of more complex cases to more experienced radiologists.

Figure 1. Traditional vs artificial intelligence (AI)-assisted workflows in the radiology departments of the 2 hospitals in this study.

The lung nodule AI systems implemented at Beijing Anzhen Hospital and Tsinghua Changgung Hospital were developed and integrated by Care.ai [17] (Yitu, integrated into Deepwise's framework in 2021) and Dr.Wise (Deepwise) [18]. Both systems followed a standardized reporting workflow and user interface, and no vendor change occurred during the study period. Both AI systems were approved by the National Medical Products Administration and provided comparable functionalities for lung nodule detection and reporting assistance. These AI systems were introduced at Beijing Anzhen Hospital in January 2019 and at Tsinghua Changgung Hospital in June 2021. The deployment phase, which involved installation and system integration, spanned approximately 1 month. Subsequently, an additional 3-month period was required to attain optimal and seamless system functionality. [Multimedia Appendix 1](#) illustrates the features and capabilities of the AI-assisted diagnostic system as shared with radiologists. This system demonstrates the capability to automatically detect lung nodules; assess their quantity, dimensions, and precise locations; and subsequently generate corresponding diagnostic reports. [Multimedia Appendices 2](#) and [3](#) illustrate the detailed operational workflow of the AI system and show that the AI system use processes in the 2 hospitals are essentially identical.

We analyzed the CT report-drafting time for individual patients from April 2018 to March 2022 at these 2 hospitals. Data for the following periods were excluded from the analysis: January 2019 to March 2019, January 2020 to August 2020, and April 2021 to August 2021, due to the potential impact of the COVID-19 pandemic, AI system integration, and the Spring Festival holiday on patterns of health seeking behavior in China [19]. Due to the considerable volatility in the night shift patterns of radiologists, this study exclusively incorporated day shift reports for analysis. To mitigate potential confounding factors related to workforce turnover, our analysis exclusively examined radiology reports written by the same group of radiologists who had generated reports both before and after the introduction of the AI system.

Data Collection

Using the University of California, San Francisco calculator [20], sample size calculations indicated a necessary sample of 12 to detect an effect size of 1 (difference in minutes between AI-assisted and conventional report-drafting time) with a SD of 0.5, ensuring 80% power at an α level of .05 (2-tailed). With nearly 100,000 reports generated at each hospital during the study period, the sample markedly surpassed requirements.

The data collection procedure encompassed the extraction of all junior radiologist report-drafting time records from the Radiology Information System and Picture Archiving and Communication System of both hospitals throughout the specified sampling period. The collected dataset included various patient-related details, such as age, sex, source of referral, examination specifics, examination time, reporting time by both junior and senior radiologists, and the identities of the reporting radiologists.

The report-drafting time analyzed in this study reflects the first-round interpretation performed by junior radiologists. The primary outcome of interest, report-drafting time, was defined as the duration between loading the CT images and clicking the Submit button on the report form [21]. To ensure data quality, we conducted a detailed assessment of extreme values. Report-drafting time of less than 0.5 minutes or more than 40 minutes—together representing approximately 5% of all observations (short drafting time: 6287/195,020, 3.2%; long drafting time: 3689/195,020, 1.9%)—were excluded, as such values most likely reflected system artifacts or nonclinical interruptions (idle time) rather than true behaviors. To ensure data integrity, a 2-investigator workflow was used. One extracted data from the reporting systems, applied predefined inclusion and exclusion criteria, and calculated report-drafting times. A second investigator independently verified the processed dataset to ensure accuracy, completeness, and consistency.

Statistical Analysis

This study used a DID design, which involved a real-world retrospective assessment of the outcome within an experimental

and a comparison group. The experimental group in our study comprised chest CT images interpreted using the AI-aided diagnostic system following its implementation, while the comparison group included other CT images that were not affected by the introduction of the AI system.

Although the intervention group focused on chest CT images and the control group included other CT images (eg, abdominal, pelvic, and head CT), all examinations followed the same radiology workflow within each institution. This institutional uniformity ensured that reporting procedures were comparable, minimizing systematic bias unrelated to the AI intervention.

Before using the DID methodology, an assessment was conducted to evaluate the parallel trend assumption. This assumption necessitates that the trends exhibited by both the experimental and comparison groups remain consistent before the intervention is introduced [22]. The DID analysis used an ordinary least squares regression model, as shown in equation 1.

$$y_{ij} = \beta + \beta_1 T_{ij} + \beta_2 I_j + \beta_3 (T_{ij} \times I_j) + \gamma^T X_{ij} + \varepsilon_{ij}$$

where the dependent variable (y_{ij}) denotes the report-drafting time for case i interpreted by radiologist j . T_{ij} indicates AI use (1 if the AI-assisted system was used, 0 otherwise), and I_j denotes group assignment (1 for the experimental group and 0 for the comparison group). The interaction term $T_{ij} \times I_j$ captures the DID estimator of interest. X_{ij} represents a vector of covariates, including radiologist sex, professional title rank, and years of working experience. SEs were clustered at the radiologist level to account for within-radiologist correlation. Models included hospital fixed effects to control for time-invariant institutional differences, as well as calendar quarter fixed effects to account for common seasonal and temporal shocks. We further tested for effect modification by radiologist seniority by adding AI-seniority interaction terms to the main model.

In this analysis, a significance level of $P < .05$ was used to determine statistical significance. The statistical analyses were

conducted using Stata (version 15.1; StataCorp LLC) and R (version 4.4.1; R Foundation for Statistical Computing).

Ethical Considerations

This study received ethics approval from the institutional review board (IRB) of Tsinghua Changgung Hospital (23352-4-01) and from the IRB of Beijing Anzhen Hospital (2023078X). Additional IRB approval was also granted by the Johns Hopkins Bloomberg School of Public Health (FWA 00000287). The IRB waived the requirement for informed consent, as this study involved secondary analysis of retrospective, noninterventive data originally collected for operational purposes. According to institutional policy, individual consent is not required for such analyses if the data are fully deidentified and used exclusively for quality improvement or research purposes. No compensation was provided to the participants in this retrospective study.

All data analyzed in this study were fully deidentified before access and analysis, with no personally identifiable information or protected health information included (sample data for review in [Multimedia Appendix 4](#)). Data were stored on secure institutional servers with access restricted to authorized study personnel only. This study does not contain any identifiable images or multimedia materials of individual participants. All figures and tables are based on aggregated, anonymized data.

Results

The study incorporated 185,044 CT images that met the predefined criteria. [Table 1](#) provides an overview of the demographic characteristics of patients from both the experimental and comparison groups, before and after the introduction of the AI system. Beijing Anzhen Hospital accounted for a larger proportion of patients after the AI launch due to its earlier implementation. The sex ratio between the experimental and comparison groups was comparable, but there were slightly more male patients than female patients in both groups. The source of patients varied slightly based on medical needs, with outpatient clinics contributing the largest number of patients.

Table 1. Patient characteristics before and after the artificial intelligence (AI) launch in experimental and control groups.

	Overall (n=185,044), n (%)	Before AI launch (n=93,870)		After AI launch (n=91,174)	
		Control group (n=55,592), n (%)	Experimental group (n=38,278), n (%)	Control group (n=50,751), n (%)	Experimental group (n=40,423), n (%)
Age (years)					
<18	1665 (0.9)	676 (1.2)	253 (0.7)	424 (0.8)	312 (0.8)
18-54	72,493 (39.4)	19,513 (35.4)	16,711 (44.1)	18,703 (37.0)	17,566 (43.6)
>55	109,712 (59.7)	34,939 (63.4)	20,910 (55.2)	31,418 (62.2)	22,445 (55.7)
Sex					
Female	85,216 (46.1)	25,734 (46.3)	17,964 (46.9)	23,846 (47.0)	17,672 (43.7)
Male	99,828 (53.9)	29,858 (53.7)	20,314 (53.1)	26,905 (53.0)	22,751 (56.3)
Hospital					
Beijing Anzhen Hospital	97,366 (52.6)	23,163 (41.7)	7880 (20.6)	36,009 (71.0)	30,314 (75.0)
Tsinghua Changgung Hospital	87,678 (47.4)	32,429 (58.3)	30,398 (79.4)	14,742 (29.0)	10,109 (25.0)
Patient source					
Emergency room	57,687 (31.2)	14,946 (26.9)	8731 (22.8)	18,284 (36.0)	15,726 (38.9)
Inpatient	38,675 (20.9)	17,215 (31.0)	5684 (14.8)	11,084 (21.8)	4692 (11.6)
Outpatient	81,655 (44.1)	23,364 (42.0)	19,686 (51.4)	21,267 (41.9)	17,338 (42.9)
Physical exam	7027 (3.8)	67 (0.1)	4177 (10.9)	116 (0.2)	2667 (6.6)

We assessed the characteristics of the CT images across 2 key dimensions to ensure their comparability over time: case mix and case complexity. First, the anatomical distribution of CT images remained largely consistent, dominated by head and abdominal studies (Figure 2A). Second, the prevalence of primary abnormal findings (eg, “liver cyst” and “cerebral atrophy”), a proxy for interpretation complexity, was also longitudinally stable in high-volume domains such as abdomen and head CTs (Figure 2B). Given that the same 74 radiologists interpreted both intervention and control images across time,

the overall case mix, case complexity, and workload remained stable. However, the job titles and working years varied for the radiologists (Figure 3). Those working at Beijing Anzhen Hospital had a higher job rank and longer working experience. Therefore, these variables were further adjusted in the main analyses. Multimedia Appendix 5 presents the outcomes of the parallel trend assessment, illustrating that the report-drafting times of both experimental and comparison groups displayed a consistent trend before the introduction of the AI system.

Figure 2. Anatomical distribution of scans and prevalence of key abnormal findings. (A) Bar chart showing the anatomical distribution of computed tomography (CT) images for the chest (treatment) and nonchest (control) groups before and after the implementation of artificial intelligence (AI). (B) Stacked area charts depicting the monthly prevalence of primary abnormal findings from 2018 to 2021 for the 3 highest-volume anatomical groups: abdomen and pelvis (left), chest and lung (middle), and head and brain (right).

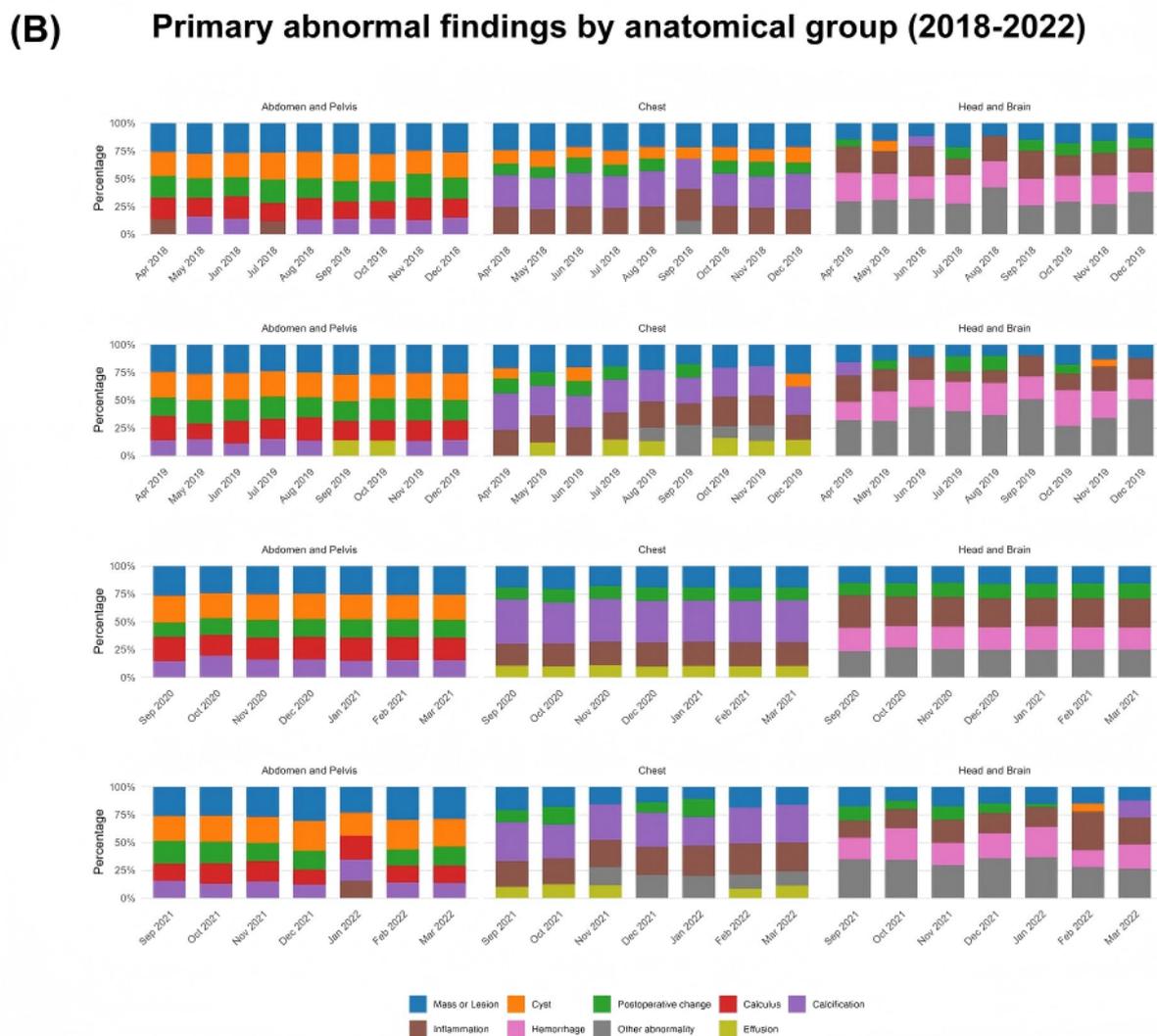
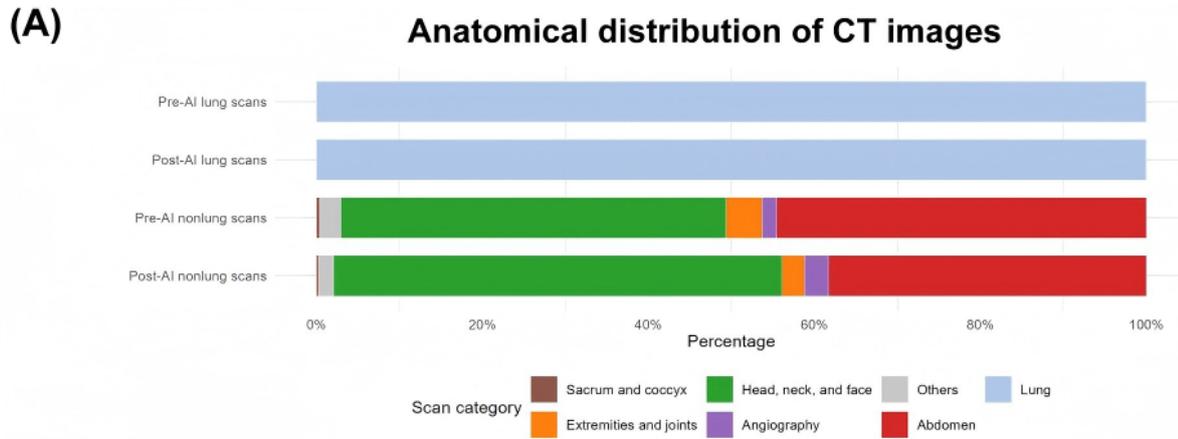
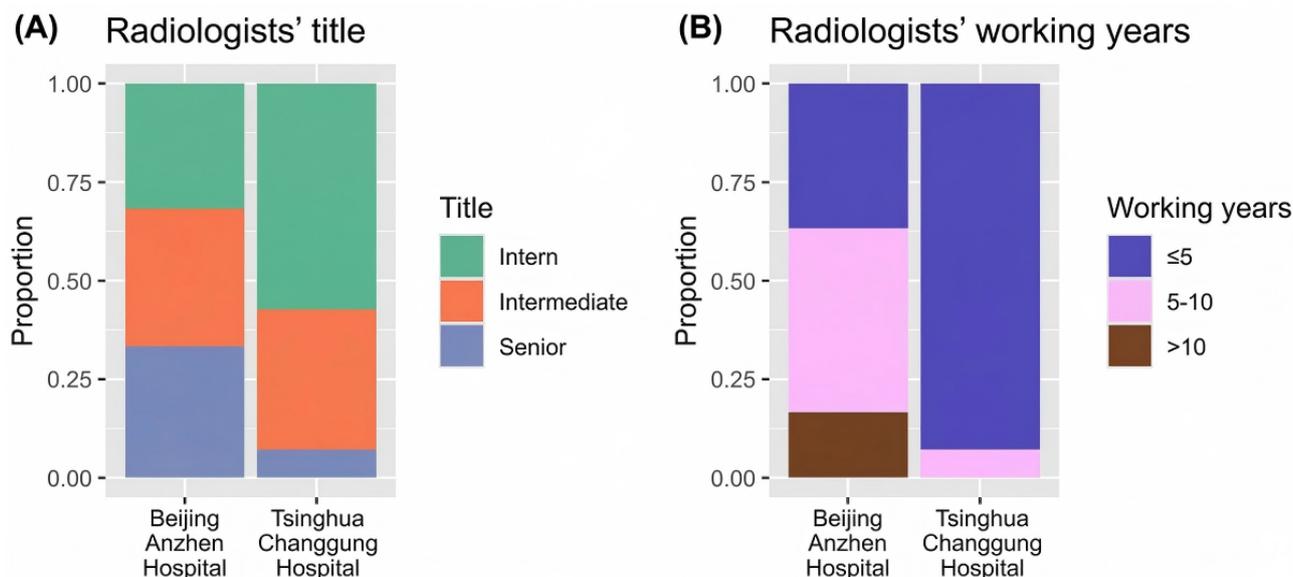


Figure 3. Radiologists' job titles and working years by hospital.



The adjusted DID model revealed a pooled effect size of 0.86 minutes (95% CI 0.14 to 1.57) after AI implementation. To assess the potential influence of anatomical site, we conducted sensitivity analyses by sequentially excluding each major category of nonchest CT images from the control group. As shown in Multimedia Appendix 6, the direction and magnitude of the AI effect remained generally consistent across all exclusion scenarios. However, this aggregated result should be interpreted with caution, as it is averaged over 2 hospitals with

different implementation timelines. When breaking down to the hospital level, Tsinghua Changgung Hospital had a nonsignificant 0.90-minute increase ($P=.12$), while Beijing Anzhen Hospital saw a statistically significant 1.37-minute time saving ($P<.001$; Table 2). No statistically significant interaction was observed between AI exposure and radiologist seniority (all $P>.09$). In sensitivity analyses controlling for patient source, the estimated AI-associated effects remained robust (the overall DID estimate remained 0.86 minutes; $P=.02$).

Table 2. Adjusted report-drafting time differences for chest computed tomography (CT) vs others before and after artificial intelligence (AI) implementation, overall and by hospital.

Intervention and group	Patients, n	Report-drafting time ^a (minutes; 95% CI)	P value
Overall (2 hospitals)	185,044		
Control group^b			
Pre-AI implementation	55,592	7.58 (6.32 to 8.85)	— ^c
Post-AI implementation	50,751	7.51 (6.10 to 8.93)	—
Pre-post difference	—	−0.07 (−0.62 to 0.48)	—
Experimental group^d			
Pre-AI implementation	38,278	6.79 (5.40 to 8.19)	—
Post-AI implementation	40,423	7.58 (6.17 to 8.99)	—
Pre-post difference	—	0.78 (0.26 to 1.30)	—
DID ^e	—	0.86 (0.14 to 1.57)	.02
Tsinghua Changgung			
DID	87,678	0.90 (−0.28 to 2.08)	.12
Beijing Anzhen			
DID	97,366	−1.37 (−2.04 to −0.70)	<.001

^aReport-drafting times are averaged over radiologists' sex, working years, and job title.

^bControl group refers to CT image reports of nonlung sites that were not subject to the launch of the AI system.

^cNot applicable.

^dExperimental group refers to lung CT image reporting, which could be affected by AI assistance.

^eDID: difference-in-differences (comparing changes in report-drafting time for lung vs nonlung CT images before and after AI implementation).

Subgroup analyses were conducted to account for the distinct implementation timelines of the AI systems at the 2 hospitals, as summarized in Figure 4 and Table 3. At Tsinghua Changgung Hospital, both the control group and the experimental group experienced extended report-drafting time, with a 0.90 minute (95% CI -0.28 to 2.08) nonsignificant increase attributable to

AI assistance. However, Beijing Anzhen Hospital demonstrated an absolute 0.76-minute reduction in the AI-assisted group by the first year and a further 1.83-minute reduction by the second year. Using a DID framework, this corresponded to a 2.66-minute relative improvement vs the counterfactual trend ($P < .001$) 2 years into the AI implementation.

Figure 4. Adjusted changes in report-drafting time before and after artificial intelligence (AI) implementation. Gray indicates phases without AI assistance, and blue indicates phases with AI assistance. Tsinghua Changgung Hospital does not have second-year post-AI implementation data because the system was adopted later at that site. Report-drafting times are averaged over radiologist sex, years in practice, and job title. Difference-in-differences estimates compare changes in report drafting time for lung vs nonlung computed tomography reports before and after AI implementation. DID: difference-in-differences.

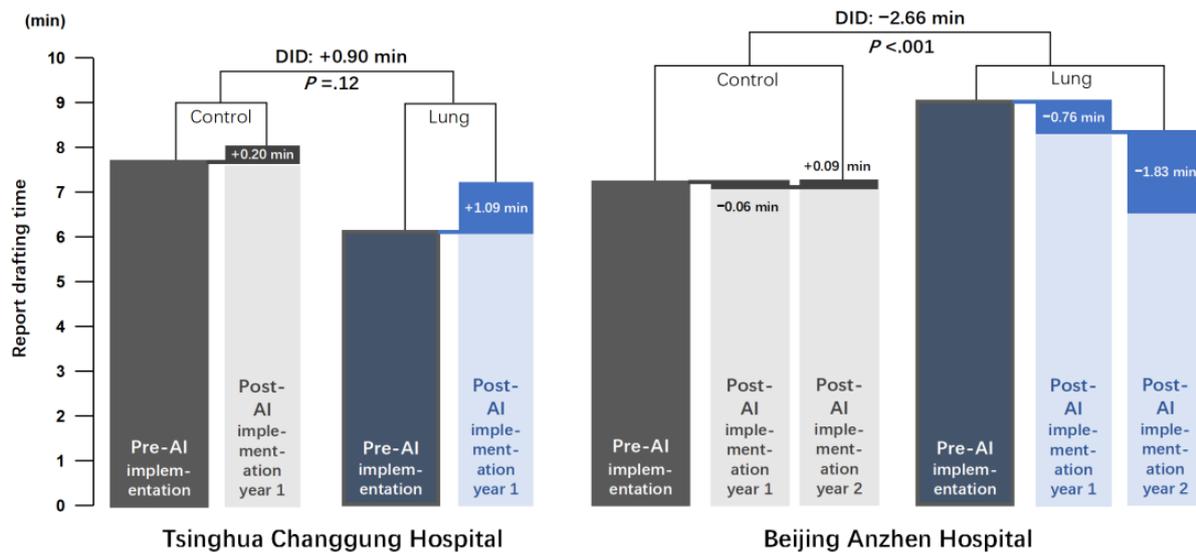


Table 3. Differences in adjusted report-drafting time comparing chest computed tomography (CT) vs others by post-artificial intelligence (AI) adoption duration.

	Adjusted report-drafting time ^a before AI implementation (min; 95% CI)	Adjusted report-drafting time ^a in the first year after AI implementation (min; 95% CI)	Adjusted report-drafting time ^a in the second year after AI implementation (min; 95% CI)
Tsinghua Changgung			
Control group ^b	7.70 (6.61 to 8.79)	7.90 (6.26 to 9.53)	— ^c
Experimental group ^d	6.09 (4.69 to 7.59)	7.18 (5.38 to 8.99)	—
DID ^e	Reference	0.90 (-0.28 to 2.08)	—
Beijing Anzhen			
Control group ^b	7.36 (5.91 to 8.81)	7.30 (5.78 to 8.82)	7.39 (5.22 to 9.56)
Experimental group	9.22 (7.82 to 10.63)	8.46 (6.97 to 9.95)	6.63 (4.75 to 8.51)
DID	Reference	-0.70 (-1.35 to -0.05)	-2.66 (-3.35 to -1.96)

^aReport-drafting times are averaged over radiologists' sex, working years, and job title.

^bControl group refers to CT image reports of nonlung sites that were not subject to the launch of the AI system.

^cNot available.

^dExperimental group refers to lung CT image reporting, which could be affected by AI assistance.

^eDID: difference-in-differences (comparing changes in report-drafting time for lung vs nonlung CT images before and after AI implementation).

Discussion

Adaptation of Diagnostic AI Systems in the Early Stages

Our findings indicate that the implementation of the AI system for lung nodule diagnostics was initially associated with an increase in report-drafting time, particularly in the first year of use. This transitional phase likely reflects a necessary adaptation period, as radiologists familiarized themselves with the system and learned to integrate it into existing workflows.

Technical barriers, such as outdated hospital IT infrastructure and limited interoperability with electronic medical record systems, may also impede seamless AI integration [23]. A potentially more important factor influencing the short-term impact of the AI system on lung nodule diagnosis may be physicians' self-efficacy. In the initial stages of the AI system's implementation, radiologists, especially junior staff, often lacked previous experience working with the AI system and were cautious in interpreting or relying on AI-generated results. This could partially explain the initial reporting delays observed at Tsinghua Changgung Hospital compared with the more experienced radiology team at Beijing Anzhen Hospital. Previous studies have shown that when radiologists lack confidence in the AI system's capabilities or have not yet fully acclimated to its operation, they might spend additional time validating AI outputs, rechecking discordant findings, or temporarily reverting to traditional workflows. This transitional phase can offset the efficiency gains intended by the AI system [24,25]. Limited formal training, reliance on trial-and-error learning, and incomplete understanding of deep learning mechanisms can further undermine early trust and efficiency [26-28].

However, over time, these early barriers diminished with routine use. A previous survey conducted at the same 2 hospitals showed strong user acceptance: radiologists rated the AI system highly in usability (System Usability Scale score: mean 74.3, SD 11.9) and efficiency improvement (mean 4.6, SD 0.6 out of 5) [29]. This growing user confidence likely reduced the need for repeated result validation and enabled more streamlined integration of the system, ultimately contributing to the substantial efficiency gains observed in later stages, such as the significant reduction in report-drafting time during the second year at Beijing Anzhen Hospital.

A related observation at Tsinghua Changgung Hospital further supports the interpretation of early-stage friction. During the first year, the control group did not exhibit a statistically significant change in report-drafting time, whereas the AI group experienced a meaningful increase. We confirmed with the hospital that no workflow reorganization, software upgrades, or staffing changes occurred during the study period. Therefore, the relative inefficiency observed in the AI arm at Tsinghua Changgung Hospital is unlikely to reflect systemic improvements in the control arm; instead, it aligns with our proposed mechanism that AI systems initially introduce cognitive load and workflow disruption. In contrast, Beijing Anzhen Hospital did not demonstrate a comparable degree of early inefficiency among its AI users. We hypothesize that this

might be due to the higher baseline experience and seniority of its radiologists (Figure 3), which may facilitate faster adaptation to AI-assisted workflows.

These findings underscore the importance of considering the user characteristics and testing for potential learning curve when evaluating the impact of AI systems. Early performance metrics may underestimate their long-term value if insufficient time is allowed for user adaptation. If the initial learning curve does exist, structured training and supportive onboarding environments can help accelerate this process and maximize clinical benefits [30].

Long-Term Benefits of Diagnostic AI Systems

The long-term benefits of AI for diagnostic efficiency are complex and may vary significantly across institutions. While the pooled analysis suggests an overall increase in report-drafting time, this averaged effect masks the true clinical picture, where one hospital showed no immediate efficiency gains, while the other demonstrated a significant reduction in report-drafting time.

In high-volume radiology departments where radiologists face overwhelmingly heavy workloads over prolonged periods [31], information overload can lead to diagnostic errors [32,33]. Under such circumstances, even modest efficiency gains are valuable. In our study, extended use of the AI system for lung nodule analysis was associated with a reduction in average report-drafting time of up to 28% at one site. The AI system's inherent capacity to autonomously identify and measure lesions markedly reduces the burden of manual labor while also expediting the generation of standardized text reports. These collective features of the AI system, in turn, reduce the necessity of repetitive writing tasks [18]. In settings where radiologists typically read more than 50 scans per day, such an improvement amounts to 2 hours of saved time per radiologist daily. This could allow for higher throughput, shorter patient waiting times, and more time allocated to complex cases.

Notably, these efficiency gains emerged alongside previously established diagnostic benefits. Our previous work evaluating the same AI system demonstrated improved lung nodule detection with a modest early rise in false positives, underscoring the need for training and adaptation [34]. That analysis further showed stable diagnostic accuracy after senior review, suggesting no substantial downstream burden attributable to AI-assisted drafting. These findings provide important clinical context for this study, which shifts the focus from diagnostic performance to operational efficiency. While this analysis focuses on first-round report drafting by junior radiologists, the combined evidence suggests that AI adoption does not merely shift workload upstream but could integrate efficiently into routine clinical workflows.

Future Directions

While our study demonstrates the potential of AI to enhance efficiency within high-volume tertiary hospitals, its findings represent a necessary first step rather than a final answer. The successful deployment in these settings provides a proof of concept, but the ultimate goal is to improve health care resource allocation across the entire system. In terms of application

stages, future mixed methods investigations are necessary to identify institutional and user-level factors that may shape AI adoption trajectories. With regard to clinical settings, future research must directly examine AI's applicability in resource-constrained environments, which are often the hardest hit by medical shortages; first, partnering with primary and county-level hospitals as they begin to adopt AI; second, investigating the distinct barriers to adoption and effectiveness in low-resource settings (eg, outdated IT infrastructure, insufficient training, and different disease mixes); and third, examining how AI might be repurposed for tasks such as fast triage or ruling out common conditions in primary care, rather than for handling complex cases. Furthermore, the evaluation and validation of AI systems represent an iterative and dynamic process, necessitating a life cycle-based regulatory framework [35,36].

Strengths and Limitations

Our study has several strengths. First, we conducted a longitudinal analysis of data from 2 tertiary hospitals in China, covering both pre- and post-AI implementation periods in real-world settings. The inclusion of a larger sample size enhances the reliability of our conclusions. Second, we applied an adjusted DID design, a quasi-experimental approach increasingly used in health services research, which strengthened the rigor of examining the causal effect of AI's clinical impact. Finally, we analyzed report-drafting times by the same radiologists who interpreted both chest and nonchest CTs across the entire study period. This ensured that seasonal workload shocks, if any, would be shared across modalities within individuals, thus not differentially impacting the experimental and the control arm.

Nevertheless, our study had several limitations. First, both hospitals were large tertiary urban hospitals with high workloads

and complex cases. As a result, our findings are considered only proof of concept, suggesting AI's potential to alleviate efficiency bottlenecks in saturated diagnostic workflows. Our conclusions may not generalize to primary or rural hospitals with different infrastructural and caseload challenges. Second, although we conducted extensive sensitivity analyses to confirm longitudinal stability in case mix, case complexity, and preintervention trends, the use of nonchest CT examinations as controls during a pandemic-affected period may only serve as imperfect controls. While we adjusted for key physician-level characteristics, the DID model may have omitted unmeasured confounders. However, the exclusion of unstable periods, inclusion of hospital and seasonal fixed effects, and the staggered rollout design lend support to the internal validity of our results. Third, while the different AI vendors deployed at the 2 hospitals had comparable diagnostic functions, variation in their workflows and interfaces may have affected the drafting time. Finally, implementing a triple-difference design was not feasible in this study. We concur that future research incorporating data from nonadopting or primary care sites is a critical next step.

Conclusions

Our findings highlight the heterogeneous impact of diagnostic AI on radiologists' efficiency. While short-term implementation was associated with an increase in report-drafting time in one hospital with a shorter follow-up period, sustained use was associated with significant efficiency gains in another hospital after more than 2 years of adoption. These divergent outcomes underscore that the realization of AI-driven efficiency is not guaranteed but is critically contingent on site-specific implementation dynamics and local workflow integration. Therefore, adequate adaptation and education are necessary for AI systems to meaningfully alleviate radiologists' workloads in the long run.

Acknowledgments

The authors sincerely thank all participating radiologists from Tsinghua Changgung Hospital and Beijing Anzhen Hospital for their invaluable contributions to this study. The authors are also grateful to the IT departments at both institutions for their technical support and assistance with data extraction and system integration. During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to polish the English language and improve readability. After this, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

YW received funding from the Beijing Nova Program (grant 20250484908).

Data Availability

The datasets generated or analyzed during this study are available from the lead and the corresponding author on reasonable request. A sample dataset with the same structure as the original data, but with anonymized, masked, and sequence-perturbed entries for privacy protection, is provided in [Multimedia Appendix 4](#).

Authors' Contributions

Conceptualization: WL, YW, MJB, ZZ, HK

Methodology: WL, YW

Investigation: WL, WY

Formal analysis: WL, YW

Data curation: WL, WY

Visualization: YW
Resources: ZZ
Supervision: HK
Writing—original draft: WL, YW
Writing—review and editing: ZZ, HK
WL and YW contributed equally to this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

User interface of the artificial intelligence–assisted system under study.

[\[PNG File , 684 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Representative photographs of the artificial intelligence–assisted computed tomography reporting workflow.

[\[PNG File , 1035 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Comparison of artificial intelligence system use workflows between the 2 hospitals.

[\[PNG File , 314 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Sample dataset.

[\[XLSX File \(Microsoft Excel File\), 42 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Parallel trend assessment.

[\[PNG File , 97 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

Sensitivity analysis using the leave-one-category-out approach.

[\[PNG File , 112 KB-Multimedia Appendix 6\]](#)

References

1. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. Apr 2017;69S:S36-S40. [doi: [10.1016/j.metabol.2017.01.011](https://doi.org/10.1016/j.metabol.2017.01.011)] [Medline: [28126242](https://pubmed.ncbi.nlm.nih.gov/28126242/)]
2. Alexander A, Jiang A, Ferreira C, Zurkiya D. An intelligent future for medical imaging: a market outlook on artificial intelligence for medical imaging. *J Am Coll Radiol*. Jan 2020;17(1 Pt B):165-170. [FREE Full text] [doi: [10.1016/j.jacr.2019.07.019](https://doi.org/10.1016/j.jacr.2019.07.019)] [Medline: [31918875](https://pubmed.ncbi.nlm.nih.gov/31918875/)]
3. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nat Rev Cancer*. Aug 2018;18(8):500-510. [FREE Full text] [doi: [10.1038/s41568-018-0016-5](https://doi.org/10.1038/s41568-018-0016-5)] [Medline: [29777175](https://pubmed.ncbi.nlm.nih.gov/29777175/)]
4. McDonald RJ, Schwartz KM, Eckel LJ, Diehn FE, Hunt CH, Bartholmai BJ, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad Radiol*. Sep 2015;22(9):1191-1198. [doi: [10.1016/j.acra.2015.05.007](https://doi.org/10.1016/j.acra.2015.05.007)] [Medline: [26210525](https://pubmed.ncbi.nlm.nih.gov/26210525/)]
5. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. *JAMA*. Jun 13, 2012;307(22):2400-2409. [FREE Full text] [doi: [10.1001/jama.2012.5960](https://doi.org/10.1001/jama.2012.5960)] [Medline: [22692172](https://pubmed.ncbi.nlm.nih.gov/22692172/)]
6. Liu Z, Yao Q, Wang P, Shen L, Li H, Zhang J, et al. The price of pressure: nationwide survey on lifestyle disturbances, occupational burnout and compromised perceived-competency among radiology residents in China. *Front Public Health*. Oct 23, 2024;12:1472397. [FREE Full text] [doi: [10.3389/fpubh.2024.1472397](https://doi.org/10.3389/fpubh.2024.1472397)] [Medline: [39507651](https://pubmed.ncbi.nlm.nih.gov/39507651/)]
7. Zhan H, Schartz K, Zygmunt ME, Johnson JO, Krupinski EA. The impact of fatigue on complex CT case interpretation by radiology residents. *Acad Radiol*. Mar 2021;28(3):424-432. [FREE Full text] [doi: [10.1016/j.acra.2020.06.005](https://doi.org/10.1016/j.acra.2020.06.005)] [Medline: [32622748](https://pubmed.ncbi.nlm.nih.gov/32622748/)]

8. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in universal health coverage. *Lancet*. Sep 28, 2019;394(10204):1192-1204. [doi: [10.1016/S0140-6736\(19\)32136-1](https://doi.org/10.1016/S0140-6736(19)32136-1)] [Medline: [31571602](https://pubmed.ncbi.nlm.nih.gov/31571602/)]
9. Kong X, Ai B, Kong Y, Su L, Ning Y, Howard N, et al. Artificial intelligence: a key to relieve China's insufficient and unequally-distributed medical resources. *Am J Transl Res*. May 15, 2019;11(5):2632-2640. [FREE Full text] [Medline: [31217843](https://pubmed.ncbi.nlm.nih.gov/31217843/)]
10. Wenderott K, Krups J, Weigl M, Wooldridge AR. Facilitators and barriers to implementing AI in routine medical imaging: systematic review and qualitative analysis. *J Med Internet Res*. Jul 21, 2025;27:e63649. [FREE Full text] [doi: [10.2196/63649](https://doi.org/10.2196/63649)] [Medline: [40690758](https://pubmed.ncbi.nlm.nih.gov/40690758/)]
11. Bonekamp D, Schlemmer HP. [Artificial intelligence (AI) in radiology?: Do we need as many radiologists in the future?]. *Urologe A*. Apr 2022;61(4):392-399. [doi: [10.1007/s00120-022-01768-w](https://doi.org/10.1007/s00120-022-01768-w)] [Medline: [35277758](https://pubmed.ncbi.nlm.nih.gov/35277758/)]
12. Waller J, O'Connor A, Rafaat E, Amireh A, Dempsey J, Martin C, et al. Applications and challenges of artificial intelligence in diagnostic and interventional radiology. *Pol J Radiol*. Feb 25, 2022;87:e113-e117. [FREE Full text] [doi: [10.5114/pjr.2022.113531](https://doi.org/10.5114/pjr.2022.113531)] [Medline: [35280945](https://pubmed.ncbi.nlm.nih.gov/35280945/)]
13. He C, Liu W, Xu J, Huang Y, Dong Z, Wu Y, et al. Efficiency, accuracy, and health professional's perspectives regarding artificial intelligence in radiology practice: a scoping review. *iRadiology*. Apr 08, 2024;2(2):156-172. [doi: [10.1002/ird3.63](https://doi.org/10.1002/ird3.63)]
14. Yi C, Tang Y, Ouyang R, Zhang Y, Cao Z, Yang Z, et al. The added value of an artificial intelligence system in assisting radiologists on indeterminate BI-RADS 0 mammograms. *Eur Radiol*. Mar 2022;32(3):1528-1537. [doi: [10.1007/s00330-021-08275-0](https://doi.org/10.1007/s00330-021-08275-0)] [Medline: [34528107](https://pubmed.ncbi.nlm.nih.gov/34528107/)]
15. Stephan D, Bertsch A, Burwinkel M, Vinayahalingam S, Al-Nawas B, Kämmerer PW, et al. AI in dental radiology-improving the efficiency of reporting with ChatGPT: comparative study. *J Med Internet Res*. Dec 23, 2024;26:e60684. [FREE Full text] [doi: [10.2196/60684](https://doi.org/10.2196/60684)] [Medline: [39714078](https://pubmed.ncbi.nlm.nih.gov/39714078/)]
16. Zia A, Fletcher C, Bigwood S, Ratnakanthan P, Seah J, Lee R, et al. Retrospective analysis and prospective validation of an AI-based software for intracranial haemorrhage detection at a high-volume trauma centre. *Sci Rep*. Nov 18, 2022;12(1):19885. [FREE Full text] [doi: [10.1038/s41598-022-24504-y](https://doi.org/10.1038/s41598-022-24504-y)] [Medline: [36400834](https://pubmed.ncbi.nlm.nih.gov/36400834/)]
17. Pan F, Li L, Liu B, Ye T, Li L, Liu D, et al. A novel deep learning-based quantification of serial chest computed tomography in Coronavirus disease 2019 (COVID-19). *Sci Rep*. Jan 11, 2021;11(1):417. [FREE Full text] [doi: [10.1038/s41598-020-80261-w](https://doi.org/10.1038/s41598-020-80261-w)] [Medline: [33432072](https://pubmed.ncbi.nlm.nih.gov/33432072/)]
18. Qi LL, Wu BT, Tang W, Zhou LN, Huang Y, Zhao SJ, et al. Long-term follow-up of persistent pulmonary pure ground-glass nodules with deep learning-assisted nodule segmentation. *Eur Radiol*. Feb 2020;30(2):744-755. [doi: [10.1007/s00330-019-06344-z](https://doi.org/10.1007/s00330-019-06344-z)] [Medline: [31485837](https://pubmed.ncbi.nlm.nih.gov/31485837/)]
19. Tan HM, Tan MS, Chang ZY, Tan KT, Ee GL, Ng CC, et al. The impact of COVID-19 pandemic on the health-seeking behaviour of an Asian population with acute respiratory infections in a densely populated community. *BMC Public Health*. Jun 23, 2021;21(1):1196. [FREE Full text] [doi: [10.1186/s12889-021-11200-1](https://doi.org/10.1186/s12889-021-11200-1)] [Medline: [34158019](https://pubmed.ncbi.nlm.nih.gov/34158019/)]
20. Sample Size Calculators for designing clinical research. Sample Size Calculators. URL: <https://sample-size.net/> [accessed 2026-03-03]
21. Ahn JS, Ebrahimian S, McDermott S, Lee S, Naccarato L, Di Capua JF, et al. Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. *JAMA Netw Open*. Aug 01, 2022;5(8):e2229289. [FREE Full text] [doi: [10.1001/jamanetworkopen.2022.29289](https://doi.org/10.1001/jamanetworkopen.2022.29289)] [Medline: [36044215](https://pubmed.ncbi.nlm.nih.gov/36044215/)]
22. Wing C, Simon K, Bello-Gomez RA. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health*. Apr 01, 2018;39(1):453-469. [FREE Full text] [doi: [10.1146/annurev-publhealth-040617-013507](https://doi.org/10.1146/annurev-publhealth-040617-013507)] [Medline: [29328877](https://pubmed.ncbi.nlm.nih.gov/29328877/)]
23. Li J, Zhou L, Zhan Y, Xu H, Zhang C, Shan F, et al. How does the artificial intelligence-based image-assisted technique help physicians in diagnosis of pulmonary adenocarcinoma? A randomized controlled experiment of multicenter physicians in China. *J Am Med Inform Assoc*. Nov 14, 2022;29(12):2041-2049. [FREE Full text] [doi: [10.1093/jamia/ocac179](https://doi.org/10.1093/jamia/ocac179)] [Medline: [36228127](https://pubmed.ncbi.nlm.nih.gov/36228127/)]
24. Brown M, Browning P, Wahi-Anwar MW, Murphy M, Delgado J, Greenspan H, et al. Integration of chest CT CAD into the clinical workflow and impact on radiologist efficiency. *Acad Radiol*. May 2019;26(5):626-631. [doi: [10.1016/j.acra.2018.07.006](https://doi.org/10.1016/j.acra.2018.07.006)] [Medline: [30097402](https://pubmed.ncbi.nlm.nih.gov/30097402/)]
25. Müller FC, Raaschou H, Akhtar N, Brejnebo M, Collatz L, Andersen MB. Impact of concurrent use of artificial intelligence tools on radiologists reading time: a prospective feasibility study. *Acad Radiol*. Jul 2022;29(7):1085-1090. [FREE Full text] [doi: [10.1016/j.acra.2021.10.008](https://doi.org/10.1016/j.acra.2021.10.008)] [Medline: [34801345](https://pubmed.ncbi.nlm.nih.gov/34801345/)]
26. Ooi SK, Makmur A, Soon AY, Fook-Chong S, Liew C, Sia SY, et al. Attitudes toward artificial intelligence in radiology with learner needs assessment within radiology residency programmes: a national multi-programme survey. *Singapore Med J*. Mar 2021;62(3):126-134. [FREE Full text] [doi: [10.11622/smedj.2019141](https://doi.org/10.11622/smedj.2019141)] [Medline: [31680181](https://pubmed.ncbi.nlm.nih.gov/31680181/)]
27. Abuzaid MM, Elshami W, Tekin H, Issa B. Assessment of the willingness of radiologists and radiographers to accept the integration of artificial intelligence into radiology practice. *Acad Radiol*. Jan 2022;29(1):87-94. [doi: [10.1016/j.acra.2020.09.014](https://doi.org/10.1016/j.acra.2020.09.014)] [Medline: [33129659](https://pubmed.ncbi.nlm.nih.gov/33129659/)]

28. Bergquist M, Rolandsson B, Gryska E, Laesser M, Hoefling N, Heckemann R, et al. Trust and stakeholder perspectives on the implementation of AI tools in clinical radiology. *Eur Radiol*. Jan 2024;34(1):338-347. [FREE Full text] [doi: [10.1007/s00330-023-09967-5](https://doi.org/10.1007/s00330-023-09967-5)] [Medline: [37505245](https://pubmed.ncbi.nlm.nih.gov/37505245/)]
29. Liu W, Wu Y, Zheng Z, Yu W, Bittle MJ, Kharrazi H. Evaluating artificial intelligence's role in lung nodule diagnostics: a survey of radiologists in two pilot tertiary hospitals in China. *J Clin Imaging Sci*. Aug 23, 2024;14:31. [doi: [10.25259/JCIS_72_2024](https://doi.org/10.25259/JCIS_72_2024)] [Medline: [39246733](https://pubmed.ncbi.nlm.nih.gov/39246733/)]
30. Sun L, Yin C, Xu Q, Zhao W. Artificial intelligence for healthcare and medical education: a systematic review. *Am J Transl Res*. Jul 15, 2023;15(7):4820-4828. [FREE Full text] [Medline: [37560249](https://pubmed.ncbi.nlm.nih.gov/37560249/)]
31. Gampala S, Vankeshwaram V, Gadula SS. Is artificial intelligence the new friend for radiologists? A review article. *Cureus*. Oct 24, 2020;12(10):e11137. [FREE Full text] [doi: [10.7759/cureus.11137](https://doi.org/10.7759/cureus.11137)] [Medline: [33240726](https://pubmed.ncbi.nlm.nih.gov/33240726/)]
32. Paul S, Nazareth DL. Input information complexity, perceived time pressure, and information processing in GSS-based work groups: an experimental investigation using a decision schema to alleviate information overload conditions. *Decis Support Syst*. Apr 2010;49(1):31-40. [FREE Full text] [doi: [10.1016/j.dss.2009.12.007](https://doi.org/10.1016/j.dss.2009.12.007)]
33. Vossenhric J, Brantner P, Cyriac J, Boll DT, Merkle EM, Heye T. Quantifying radiology resident fatigue: analysis of preliminary reports. *Radiology*. Mar 2021;298(3):632-639. [doi: [10.1148/radiol.2021203486](https://doi.org/10.1148/radiol.2021203486)] [Medline: [33497316](https://pubmed.ncbi.nlm.nih.gov/33497316/)]
34. Liu W, Wu Y, Zheng Z, Bittle M, Yu W, Kharrazi H. Enhancing diagnostic accuracy of lung nodules in chest computed tomography using artificial intelligence: retrospective analysis. *J Med Internet Res*. Jan 27, 2025;27:e64649. [FREE Full text] [doi: [10.2196/64649](https://doi.org/10.2196/64649)] [Medline: [39869890](https://pubmed.ncbi.nlm.nih.gov/39869890/)]
35. Liu Y, Chen PH, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. *JAMA*. Nov 12, 2019;322(18):1806-1816. [doi: [10.1001/jama.2019.16489](https://doi.org/10.1001/jama.2019.16489)] [Medline: [31714992](https://pubmed.ncbi.nlm.nih.gov/31714992/)]
36. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med*. Jul 15, 2021;385(3):283-286. [FREE Full text] [doi: [10.1056/NEJMc2104626](https://doi.org/10.1056/NEJMc2104626)] [Medline: [34260843](https://pubmed.ncbi.nlm.nih.gov/34260843/)]

Abbreviations

- AI:** artificial intelligence
CT: computed tomography
DID: difference-in-differences
IRB: institutional review board

Edited by A Coristine; submitted 23.May.2025; peer-reviewed by J Jiang, J Zhang; comments to author 17.Jun.2025; revised version received 13.Feb.2026; accepted 14.Feb.2026; published 27.Mar.2026

Please cite as:

Liu W, Wu Y, Yu W, Bittle MJ, Zheng Z, Kharrazi H

Measuring the Impact of AI on Report-Drafting Efficiency in Chest Computed Tomography Interpretation: Retrospective Analysis
J Med Internet Res 2026;28:e77967

URL: <https://www.jmir.org/2026/1/e77967>

doi: [10.2196/77967](https://doi.org/10.2196/77967)

PMID:

©Weiqi Liu, You Wu, Wei Yu, Mark J Bittle, Zhuozhao Zheng, Hadi Kharrazi. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 27.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.