How AI-Augmented Chest Radiography Alters Decisions in Emergency Care

Hyo Kang[1,†], **Hyun Joo Shin**[2,†], **Sukwoong Choi**[3,*], **Eun-Kyung Kim**[2,*], **Namil Kim**[4,*]

1 SNU Business School, Seoul National University; Seoul, 08826, Republic of Korea.
2 Department of Radiology, Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Yongin Severance Hospital, Yonsei University College of Medicine; Yongin-si, Gyeonggi-do, 16995, Republic of Korea.
3 Massry School of Business, University at Albany, State University of New York; Albany, NY, 12222, USA.
* Corresponding author. Email: ekkim@yuhs.ac; namil.kim@yonsei.ac.kr
† These authors contributed equally to this work as co-first authors and are listed in alphabetical order.

Introduction

Recent advancements in AI have revolutionized professional capabilities across fields, with medicine standing out for its demand for precision and speed (*1–3*). In hospitals, AI is shifting from research to routine clinical use, particularly in radiology where it powers over three-quarters of FDA-approved AI medical devices (*4*). These AI tools enhance image quality, detect abnormalities, and segment pathologies (*5*). A pivotal question, however, remains whether and how AI tools reshape physician decision-making and patient management in real-world clinical settings. We tackle this by probing AI-augmented chest radiography in EDs, using a regression discontinuity design (RDD) on comprehensive real-world data from a full ED cohort at a university hospital.

Chest radiographs (CXR), the most common hospital imaging, exemplify AI's rapid integration amid high volumes and radiologist shortages (*2, 6*). In EDs, where timely decisions are life-

critical, delays in expert interpretations heighten risks. Commercial AI solutions have emerged, boosting workflows and lesion detection in early adopters (*4*–*7*). Yet, existing evidence largely focuses on technological performance, such as diagnostic accuracy or results from experimental studies, while rigorous evaluations of AI's effects in real-world clinical use, where clinician behavior, workflows, and constraints interact, remain scarce.

Exploiting a 15% AI-generated abnormality score as a threshold for AI augmentation (i.e., triggering the display of boundary maps and abnormality scores), we applied an RDD to 27,579 visits and chest radiographs from 23,405 patients (March 2020–May 2022). We focus on consolidation and pneumothorax, two common yet clinically distinct lesions that differ in acuity. Our findings reveal context-dependent improvements: For consolidation, AI augmentation significantly increased ED-initiated consultations to relevant specialties without prolonging hospital stays or increasing follow-up imaging. For pneumothorax, a much more acute condition, it boosted overall consultations, extended total stays, and elevated additional CXR and CT scans. These patterns highlight AI-clinician synergies that vary with lesion urgency and severity. For relatively less severe cases such as consolidation, these interactions tend to enhance efficiency, whereas for acute conditions like pneumothorax, they promote caution, potentially reducing risks but increasing diagnostic service utilization. Our findings highlight AI's potential to improve clinical decision making, diagnostic capacity allocation, and patient outcomes across healthcare systems, providing insights for clinicians, hospitals, and policymakers in designing effective AI integration and utilization strategies.

Hospital-wide integration of AI software for CXR

A commercial AI-based lesion detection software (Lunit INSIGHT CXR, versions 2 and 3) was integrated into the imaging workflow for all CXR at a university hospital beginning in March

2020, when the hospital first opened. This 650-bed university hospital includes a designated regional emergency center, and the AI software was applied to every CXR performed. For each CXR, the software detects any covered lesions present and assigns an abnormality score (0–100%) to each identified lesion. In accordance with previous studies and the vendor's guidelines, a lesion was considered present when the AI-generated abnormality score exceeded 15% for any of the lesions it covers (*4*).

The software generates a secondary capture image that highlights the region of interest for each detected lesion using abbreviated labels and displays the corresponding abnormality scores. The annotated image (AI-augmented CXR) is automatically attached to the original radiographs in PACS (Fig. 1). The entire process, from image acquisition to AI result display, typically takes 5-10 seconds, as the image is uploaded to the in-hospital AI server and the results are seamlessly integrated into PACS (*8*). Clinicians could easily access and refer to AI results by scrolling through the PACS, regardless of their location or time of access within the hospital.

In this study, we retrospectively extracted CXR and AI analysis data from the PACS and the AI server. Only the abnormality scores (%) for consolidation and pneumothorax were used, as these two lesions were consistently included in all software versions, are well detected by the AI, and are considered the most common and critical findings in the ED setting for chest diagnosis. We analyze the two lesions separately, as pneumothorax represents a much more acute condition that requires immediate attention and is therefore expected to yield different outcomes.

Measurement of Clinical Impact of AI

We evaluated four outcomes to assess the clinical impact of AI-augmented CXR on ED patient management. The primary outcome was the proportion of consultations to clinically relevant

departments, which captures how AI influences clinician behaviors. For each ED visit, we determined whether a consultation was made to a relevant specialty, regardless of the ordering department. Specifically, for consolidation, appropriate departments included Pulmonology and Allergy, Thoracic Surgery, and Infectious Diseases. For pneumothorax, Thoracic Surgery was considered relevant. Subgroup analysis was also conducted, limited to consultations that were initiated by ED physicians.

Another set of outcomes, designed to capture how AI influences clinical workflow and patient management, were the length of ED stay (in hours) and total hospital stay (in days), defined as the time from initial ED presentation to final discharge—either directly from the ED or after inpatient admission. These measures were intended to assess whether AI influenced the duration of patient care, both within the ED and across the overall hospital course.

The remaining outcomes, to assess diagnostic decision-making, were downstream imaging during the ED stay: (1) follow-up CXR and (2) chest CT scans, regardless of the ordering department. In addition, subgroup analysis was performed by limiting to examinations ordered by relevant departments or ED physicians following the initial CXR. Detailed descriptions of the data, AI system, and regression discontinuity design are provided in the Methods.

Methods

Data and Sample

For clinical data analysis, we utilized the electronic medical records, including age, sex, ED visit dates, CXR examination dates, AI triage results, number and types of consultations, hospitalization status, discharge dates, and the total number of CXR and chest CTs obtained during the ED stay.

The inclusion criteria for the analysis sample were as follows: patients aged >18 years who visited the ED between March 2020 and May 2022 and underwent CXR in posteroanterior or anteroposterior projection views, which were analyzed by commercial AI software at the time of image verification. The initial radiograph from each visit was included for the main analysis.

This study was approved by the university hospital's Institutional Review Board with waived informed consent (IRB no. 9-2022-0060) and conducted in accordance with STROBE guidelines and relevant regulations.

Statistical method

Comparisons of demographic data were conducted using Student's t-test. To investigate whether the AI-generated CXR result led to meaningful changes in physician behavior or patient outcomes, we applied an RDD. This quasi-experimental approach leverages the predefined threshold of the AI software: an AI-generated abnormality score of 15% or higher triggers the overlay of lesion-specific annotations (including region-of-interests, annotations, and abnormality scores) on the original radiographs.

Rather than simply comparing outcomes across abnormality score ranges—which could be confounded by the correlation between higher scores and more severe findings—the RDD leverages a predefined threshold (15%) that triggers the AI to display an overlay and focuses on observations within a narrow bandwidth around this cutoff. Under the assumption that patients just above and just below this cutoff are comparable in both observed and unobserved characteristics, any discontinuity in outcomes at the threshold can be interpreted as a causal effect of the AI-augmented radiographs. This quasi-experimental design thus allows for the estimation of causal effects even in the absence of randomized assignments.

For graphical representation, we used a bin width of 1 percentage point and plotted the average value of the outcome variable within each bin. In the case of pneumothorax, due to the smaller number of observations above the threshold, a wider bin width was employed to ensure at least 25 observations per bin above the cutoff.

The main analysis was performed at the patient-visit level. For the formal regression analysis, we employ the linear regression outlined in Equation (1):

$$y_i = \beta_1 + \gamma d_i + \beta_2 d_i s_i + \beta_3 d_i s_i^2 + \beta_4 s_i + \beta_5 s_i^2 + u_i, \quad (1)$$

where i denotes each patient-visit, s_i is the AI-generated abnormality score centered at the 15% cutoff, and d_i is an indicator variable equal to 1 if the score is greater than or equal to 15%. This specification estimates separate quadratic functions on either side of the cutoff. The parameter of interest, γ captures the magnitude of any discontinuous jump at the cutoff, indicating whether the AI output influenced clinical decision-making or patient management.

To account for the bias-variance trade-off inherent in RDD, we used a symmetric bandwidth of [0, 30] around the 15% cutoff for the consolidation analysis. For pneumothorax, where the number of observations above the threshold was much smaller, we extended the right-side bandwidth to [0, 40].

Robustness checks

Further robustness checks substantiated our results. First, splitting the sample into 2020 (Figs. S1–S2, Table S1) and 2021–2022 (Figs. S3–S4, Table S2) showed no significant effects in 2020, consistent with the early phase of AI implementation and disruptions related to COVID-19, whereas effects became stronger and more precise in later years, reflecting physicians' increasing familiarity with AI-augmented radiographs. Second, alternative RDD functional

forms, including cubic polynomials, produced consistent outcomes (Figs. S5-S6, Table S3). Third, excluding multi-radiograph visits maintained similar results (Table S4). Finally, graphical RDD representations across the full abnormality score range confirmed the patterns consistent with those observed in the narrower range (Figs. S7-S8).

Subjects and Main Findings

Subjects

In our final sample, 23,405 patients made a total of 27,579 visits to the ED, each of which included at least one initial CXR ordered by the ED doctors. Our analysis focused exclusively on these ED-initiated initial CXRs to assess how clinical responses varied depending on the AI results.

Among the initial CXR with AI-generated abnormality scores between 0 and 100%, 6,641 (24.08%) for consolidation and 508 (1.84%) for pneumothorax had abnormality scores of 15% or higher. Of the total visits, 14,341 (52%) involved at least one specialty consultation, amounting to 80,200 consultations. Among those with AI-detected consolidation, 3,352 (50.47%) received consultations with one or more of the three relevant departments. Among those with AI-detected pneumothorax, 159 (31.3%) received consultation with the relevant department during the ED stay.

Table 1 presents descriptive statistics at both the patient and patient-visit levels. Note that the comparison of means between the positive (abnormality scores of 15% or higher) and negative groups is provided for descriptive purposes only and does not reflect the validity of the RDD. The core of RDD lies in comparing the trends on either side of the cutoff to identify any discontinuous break, rather than relying on balances in aggregate means.

For consolidation, 1,353 out of 22,291 CXRs (6.1%) had a positive detection and, among them, 478 (35.3%) received consultations with the relevant departments. For pneumothorax, 246 out of 27,317 CXRs (0.9%) had a positive detection and, among them, 29 (11.8%) received consultations with the Thoracic Surgery department. The length of ED stay and total hospital stay were significantly longer in the positive group compared to the negative group for consolidation and pneumothorax (all, $p<0.01$). The number of CXR during each visit was higher in the positive group compared to the negative group, ordered by all departments (all, $p<0.01$). However, for pneumothorax, there was no significant difference in the number of CXR when limited to those ordered by the Thoracic Surgery department and the ED ($p=0.06$). The proportion of chest CTs was significantly higher in the positive group than in the negative group: 25.9% vs. 8.1% for consolidation, and 53.3% vs. 16.4% for pneumothorax (both $p<0.01$). Similar results were observed when the analysis was limited to chest CT ordered by relevant departments. Detailed demographic characteristics are presented in Table 1A.

AI-augmented patient management in consolidation cases

The results of the RDD analysis in consolidation cases are presented in Table 2A and Fig. 2. Regarding the consultation requests to relevant departments, we found significant change in the overall number of consultation around the cutoff value (1) when requests from all departments were included (RD estimate=0.096, SE=0.051, $p=0.059$) and (2) when the analysis was limited to consultations initiated by the ED (RD estimate=0.080, SE=0.043, $p=0.043$).

In Fig. 2, each data point represents the average proportion of patient visits involving specialty consultations initiated by all or the ED to at least one of the three relevant departments—within each 1-percentage-point bin of the abnormality score. Separate second-degree polynomial regression curves were fitted for data points above and below the 15% cutoff. A significant

increase in the proportion of consultation requests initiated by ED was observed immediately above the 15% threshold, indicating a discontinuous change in physician behavior in response to AI-augmented findings. AI results prompted more specialty consultations to relevant departments, and because patients near the threshold were comparable in both observable and unobservable characteristics, the increase likely reflects the causal influence of AI-augmented radiographs.

Regarding the length of stay, no significant changes were observed in the length of ED stay (RD estimate=0.013, SE=0.015, p=0.367) or total hospital stay (RD estimate=0.786, SE=1.299, p=0.545) around the cutoff value (Fig. 2B). Regarding imaging examinations (Fig. 2C-D), there were no significant changes in the total number of CXR (RD estimate=0.064, SE=0.622, p=0.918), the proportion of patient-visits that involved chest CT scans (RD estimate=0.049, SE=0.046, p=0.282), or the total number of chest CTs during the visit (RD estimate=0.040, SE=0.053, p=0.448). Similar results were observed when the analysis was limited to examinations ordered by relevant departments and the ED. These results are illustrated by the trend lines shown in Fig. 2B-D.

AI-augmented patient management in pneumothorax cases

Pneumothorax warrants separate investigation because it is an acute condition demanding prompt specialist attention, whereas consolidation typically allows for less urgent management. The results of the RDD analysis in pneumothorax cases are shown in Table 2B and Fig. 3. Regarding the consultation requests to relevant departments, a significant change was observed in the overall number of consultations when requests from all departments were included (RD estimate=0.176, SE=0.083, p=0.035). However, when the analysis was limited to consultations

initiated by the ED, no significant increase was found around the cutoff value (RD estimate=0.096, SE=0.049, p=0.053).

Regarding the length of stay (Fig. 3B), there was no significant change in the length of ED stay (RD estimate=0.023, SE=0.030, p=0.456), whereas total hospital stay showed a significant increase around the cutoff (RD estimate=12.774, SE=4.680, p=0.006).

For imaging examinations (Fig. 3C-D), after adjusting for ED stay duration, significant increases were observed in the total number of CXR (RD estimate=10.121, SE=2.883, p<0.001), the proportion of chest CTs (RD estimate=0.398, SE=0.113, p<0.001), and the total number of chest CTs performed during the visit (RD estimate=0.623, SE=0.197, p=0.002). Similar results were observed when the analysis was limited to imaging studies ordered by relevant departments and the ED.

Discussion

This study demonstrates that AI-augmented chest radiography can meaningfully shape clinical decisions and patient management in high-stakes emergency care, where heavy workloads and time pressures demand efficient, evidence-based decisions. Leveraging a large-scale dataset and sharp regression discontinuity design, we provide causal evidence of AI's influence, revealing effects that vary by lesion types and workflows.

For consolidation, crossing the 15% AI augmentation threshold significantly increased ED-initiated consultations to relevant specialties without prolonging hospital stays or triggering additional imaging. This suggests AI prompts targeted, timely specialist involvement, facilitating coordinated care while avoiding unnecessary additional diagnostic service utilization. In contrast, for pneumothorax which represents a more acute, life-threatening condition, AI augmentation

boosted overall consultations, extended total hospital stays, and elevated follow-up CXR and CT scans. These broader patterns indicate AI escalates vigilance and downstream evaluations in urgent cases. It is also reasonable that the ED-initiated metrics showed limited change, as the increased length of stay occurred mainly during the inpatient phase rather than in the ED, suggesting that patients were promptly transferred from the ED to the appropriate inpatient departments.

These context-dependent effects reflect the interplay of AI and human factors: less severe lesions like consolidation enable efficiency, while acute ones like pneumothorax amplify caution, potentially averting risks but raising costs. Clinical integration of AI is the next frontier (*9*), yet most evidence remains experimental (*5, 10*). Our findings to some extent address widespread concerns about over-testing (*6, 11*) or algorithm aversion (*12*), showing AI as a supportive tool. That is, it enhances the detection of subtle abnormalities and streamlines inter-departmental consultations rather than substitutes for physician expertise. In our setting, where clinicians access both original and augmented images, AI's boundary maps and abnormality scores fostered thorough reviews and better decisions.

Broader implications extend to healthcare policy and practice. Amid lagging AI adoption in medicine (*13*), our empirical evidence from an early-adopter hospital supports integration to alleviate bottlenecks, such as ED overcrowding and delayed care. By prioritizing urgent cases and optimizing diagnostic capacity allocation, thereby reducing inefficiencies without uniformly increasing testing, AI can enhance patient outcomes, lower costs, and promote equitable access in capacity-constrained systems. However, reactions to AI vary by display methods and by behavior or environmental factors (*7, 9, 12*), emphasizing the decisive role of clinicians in workflows.

Limitations include the RDD's local effects around the 15% threshold, single institution focus (one university hospital using one AI model), and potential confounders like comorbidities. Generalizability may differ across settings, populations, or algorithms; future multi-site studies could explore these. Extrapolating to extreme scores (near 0% or 100%) requires caution, as AI's value shines in ambiguous cases.

In conclusion, our work illuminates AI's transformative potential in hospitals, blending technology with human insight to foster timely, evidence-based care. We hope these insights spark discussions among academics, practitioners, and policymakers, advancing AI's role in saving lives and reshaping healthcare delivery.

References

1. E. J. Hwang, J. M. Goo, C. M. Park, AI Applications for Thoracic Imaging: Considerations for Best Practice. *Radiology* 314 (2025).
2. M. Moassefi, S. Faghani, B. Khosravi, P. Rouzrokh, B. J. Erickson, "Artificial Intelligence in Radiology: Overview of Application Types, Design, and Challenges." in *Seminars in Roentgenology* (2023) vol. 58, pp. 170–177.
3. K. G. van Leeuwen, M. de Rooij, S. Schalekamp, B. van Ginneken, M. J. C. M. Rutten, How does artificial intelligence in radiology improve efficiency and health outcomes? *Pediatr. Radiol.* 52, 2087–2093 (2022).
4. H. J. Shin, S. Lee, S. Kim, N. H. Son, E. K. Kim, Hospital-wide survey of clinical experience with artificial intelligence applied to daily chest radiographs. *PLoS One* 18, 1–14 (2023).
5. H. J. Shin, K. Han, L. Ryu, E. K. Kim, The impact of artificial intelligence on the reading times of radiologists for chest radiographs. *npj Digit. Med.* 6 (2023).
6. S. H. Hwang, H. J. Shin, E. K. Kim, E. H. Lee, M. Lee, Clinical outcomes and actual consequence of lung nodules incidentally detected on chest radiographs by artificial intelligence. *Sci. Rep.* 13, 1–10 (2023).
7. S. H. Kwak, E. K. Kim, M. H. Kim, E. H. Lee, H. J. Shin, Incidentally found resectable lung cancer with the usage of artificial intelligence on chest radiographs. *PLoS One* 18, 1–13 (2023).
8. S. Lee, H. J. Shin, S. Kim, E. K. Kim, Successful Implementation of an Artificial Intelligence-Based Computer-Aided Detection System for Chest Radiography in Daily Clinical Practice. *Korean J. Radiol.* 23, 847–852 (2022).

9. M. G. Linguraru, S. Bakas, M. Aboian, P. D. Chang, A. E. Flanders, J. Kalpathy-Cramer, F. C. Kitamura, M. P. Lungren, J. Mongan, L. M. Prevedello, R. M. Summers, C. C. Wu, M. Adewole, C. E. Kahn, Clinical, Cultural, Computational, and Regulatory Considerations to Deploy AI in Radiology: Perspectives of RSNA and MICCAI Experts. *Radiol. Artif. Intell.* 6 (2024).

10. E. J. Hwang, S. Park, K. N. Jin, J. I. Kim, S. Y. Choi, J. H. Lee, J. M. Goo, J. Aum, J. J. Yim, J. G. Cohen, G. R. Ferretti, C. M. Park, Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. *JAMA Netw. open* 2, e191095 (2019).

11. Z. Wang, L. Wei, L. Xue, Overcoming medical overuse with AI assistance: An experimental investigation. *J. Health Econ.* 103, 103043 (2025).

12. C. Longoni, A. Bonezzi, C. K. Morewedge, Resistance to Medical Artificial Intelligence. *J. Consum. Res.* 46, 629–650 (2019).

13. A. Goldfarb, F. Teodoridis, Why is AI adoption in health care lagging? *Brookings Pap. Econ. Act.* 1–10 (2022).

Tables

Table 1. Patient demographics. Panel (A) reports descriptive statistics based on our final sample, regardless of abnormality score (ranging from 0 to 100), at the patient-visit level. Since patients may have multiple Emergency Department visits over time, this table considers only the first visit for each patient within the sample period. Panel (B) further divides the sample into treatment and control groups for each lesion. The treatment group is defined as having a score in the range (15,30] for consolidation and (15,40] for pneumothorax, while the control group includes cases with scores in [0, 15]. Corresponding descriptive statistics are reported for each group. The first number represents the mean, followed by the standard deviation.

A. Patient-level

	Full sample	**Consolidation**			**Pneumothorax**		
Group	All	Negative	Positive	p-value	Negative	Positive	p-value
Observations	(*N*=23,405)	(*N*=18,219)	(*N*=1,068)		(*N*=23,028)	(*N*=176)	
Score range	[0, 100]	[0, 15]	(15,30]		[0, 15]	(15,40]	
Age	55.29±19.97	50.75±18.32	64.51±19.65	$p<0.01$	55.13±19.92	70.18±17.17	$p<0.01$
Sex (M:F)	10,922:12,483 (46.7%)	8,193:10,026 (45.0%)	528:540 (49.4%)	$p<0.01$	10,645:12,383 (46.2%)	120:56 (68.2%)	$p<0.01$
Visits (w/ X-ray)	1.18±0.61	1.15±0.55	1.24±0.83	$p<0.01$	1.18±0.60	1.43±1.51	$p=0.03$

B. Patient-visit level

	Full sample	Consolidation			Pneumothorax		
Group	All	Negative	Positive	p-value	Negative	Positive	p-value
Observations	(*N*=27,579)	(*N*=20,938)	(*N*=1,353)		(*N*=27,071)	(*N*=246)	
Score range	[0, 100]	[0, 15]	(15,30]		[0, 15]	(15,40]	
Consultation requests to relevant dept. (%)							
All	5,782:21,797 (21.0%)	2,430:18,508 (11.6%)	478:875 (35.3%)	$p<0.01$			
	681:26,898 (2.5%)				522:26,549 (1.9%)	29:217 (11.8%)	$p<0.01$
From ED	2,823:24,756 (10.2%)	933:20,005 (4.5%)	189:26,882 (15.6%)	$p<0.01$			
	311:27,268 (1.1%)				2,564:24,507 (0.7%)	9:237 (3.7%)	$p<0.01$
ED length of stay							
Units: days	0.18±0.14	0.17±0.12	0.19±0.14	$p<0.01$	0.18±0.14	0.20±0.12	$p<0.01$
Units: hours	4.24±3.34	3.98±2.84	4.57±3.44	$p<0.01$	4.23±3.34	4.83±2.87	$p<0.01$
Total hospital stay (incl. ED; days)	3.73±8.98	2.54±6.87	6.12±12.82	$p<0.01$	3.64±8.83	9.54±15.48	$p<0.01$
CXR (exam number)							
All	2.68±6.00	1.7±3.42	3.57±6.72	$p<0.01$	2.58±5.82	7.04±9.06	$p<0.01$
From 3 dept. & ED	1.46±3.12	1.14±1.42	1.71±2.99	$p<0.01$			
From 1 dept. & ED	1.13±1.41				1.09±1.24	1.37±2.21	$p=0.06$
Chest CT (%)							
All	4,718:22,861 (17.1%)	1,694:19,244 (8.1%)	351:1,002 (25.9%)	$p<0.01$	4,452:22,619 (16.4%)	131:115 (53.3%)	$p<0.01$
From relevant dept.	3,757:23,822 (13.6%)	1,255:19,683 (6.0%)	281:1,072 (20.8%)	$p<0.01$	3,525:23,546 (13.0%)	109:137 (44.3%)	$p<0.01$
Chest CT (exam number)							
All	0.19±0.45	0.09±0.30	0.28±0.50	$p<0.01$	0.18±0.44	0.63±0.70	$p<0.01$
From relevant dept.	0.14±0.37	0.06±0.24	0.22±0.44	$p<0.01$	0.14±0.37	0.46±0.53	$p<0.01$
Abnormality score(%)							
Overall	24.20±33.60	9.65±19.82	43.76±26.51	$p<0.01$	23.18±32.90	70.11±28.06	$p<0.01$
Consolidation	16.26±28.67	2.24±3.53	22.52±4.03	$p<0.01$			
Pneumothorax	1.69±8.21				0.74±1.63	25.95±6.76	$p<0.01$

Table 2. Effects of AI-augmented CXR on patient management (Linear fitted line). This table reports estimates from a regression discontinuity design as specified in Equation (1). Heteroskedasticity-robust standard errors are shown in parenthesis, with *p*-value in square brackets. (**A**). For the analysis on consolidation, we use a symmetric bandwidth of [0, 30]. (**B**). For pneumothorax, a wider bandwidth of [0, 40] is used to account for the smaller number of observations above the cutoff. In specifications reported in columns 4a and 6b, we include the ED length of stay as a control variable, as longer stays may mechanically increase the likelihood of X-ray examinations or CT scans.

A. Consolidation

	Dependent variables:									
	Consultation requests to relevant depts.		**ED stay**	**Total hospital stay (incl. ED)**	**Examinations**					
					CXR (exam number)		*CT scans (%)*		*CT scans (exam number)*	
Sample	All (1a)	From ED (1b)	(2)	(3)	All (4a)	Relevant depts and ED (4b)	All (5a)	Relevant depts and ED (5b)	All (6a)	Relevant depts and ED (6b)
RD estimate (γ)	0.096* (0.051) [p=0.059]	0.080** (0.043) [p=0.043]	0.013 (0.015) [p=0.367]	0.786 (1.299) [p=0.545]	0.064 (0.622) [p=0.918]	0.578* (0.319) [p=0.070]	0.049 (0.046) [p=0.282]	0.053 (0.043) [p=0.217]	0.040 (0.053) [p=0.448]	0.075 (0.048) [p=0.116]
Control	–	–	–	–	ED Length of Stay		ED Length of Stay		ED Length of Stay	
Adjusted R^2	0.093	0.038	0.013	0.045	0.050	0.019	0.076	0.064	0.076	0.065
Observations	22,291	22,291	22,291	22,291	22,291	22,291	22,291	22,291	22,291	22,291

B. Pneumothorax

	Dependent variables:									
	Consultation requests to relevant depts.		**ED stay**	**Total hospital stay (incl. ED)**	**Examinations**					
					CXR (exam number)		*CT scans (%)*		*CT scans (exam number)*	
Sample	All (1a)	From ED (1b)	(2)	(3)	All (4a)	Relevant dept and ED (4b)	All (5a)	Relevant dept and ED (5b)	All (6a)	Relevant dept and ED (6b)
RD estimate (γ)	0.176** (0.083) [p=0.035]	0.096* (0.049) [p=0.053]	0.023 (0.030) [p=0.456]	12.774*** (4.680) [p=0.006]	10.121*** (2.883) [p<0.001]	2.902*** (1.055) [p=0.006]	0.398*** (0.113) [p<0.001]	0.248** (0.113) [p=0.028]	0.623*** (0.197) [p=0.002]	0.238** (0.114) [p=0.038]
Control	–	–	–	–	ED Length of Stay		ED Length of Stay		ED Length of Stay	
Adjusted R^2	0.011	0.004	0.015	0.036	0.049	0.022	0.113	0.089	0.110	0.088
Observations	27,317	27,317	27,317	27,317	27,317	27,317	27,317	27,317	27,317	27,317

Figures

Fig. 1. Screens displaying the original and AI-augmented radiographs. (A) A boundary map is overlaid on suspicious regions indicating consolidation (Csn), accompanied by its AI-generated abnormality score of 98%. (B) A boundary map is overlaid on suspicious regions indicating pneumothorax (Ptx), also accompanied by its abnormality score of 87%. Lesion-specific boundary map and their corresponding scores are displayed only when the abnormality score is 15% or higher. When the score was below 15%, the score was presented as "low" and there was no boundary map displayed on images. The maximum abnormality score across lesions is shown at the bottom of the image.

A. Consolidation **B. Pneumothorax**

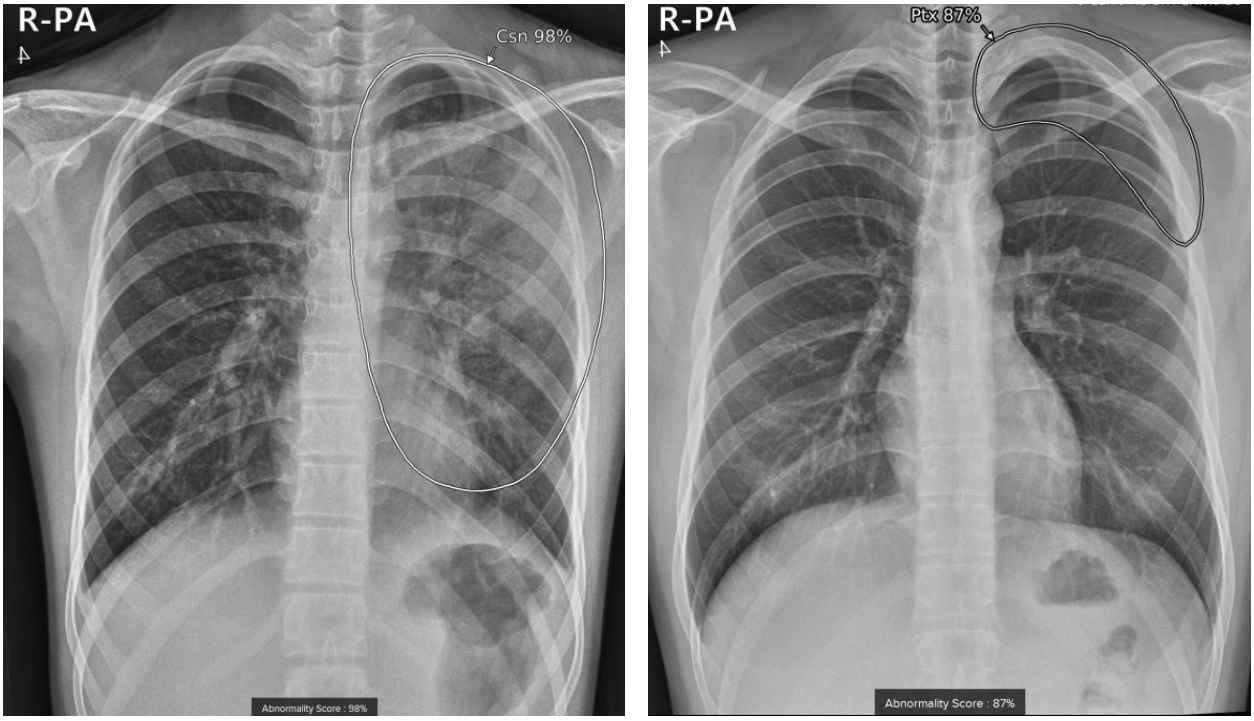
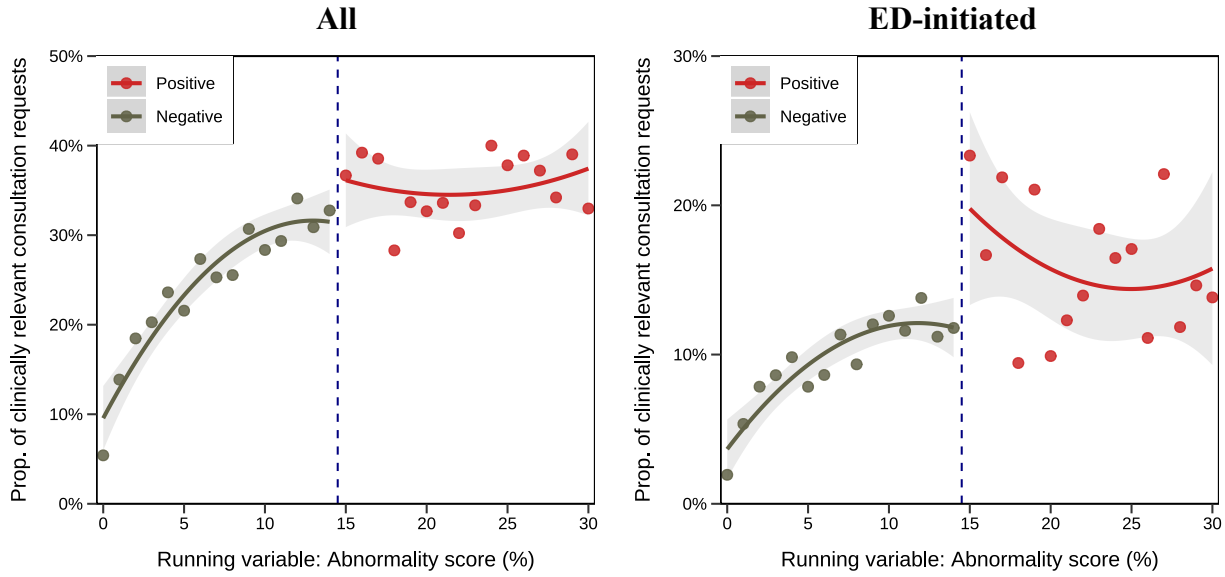
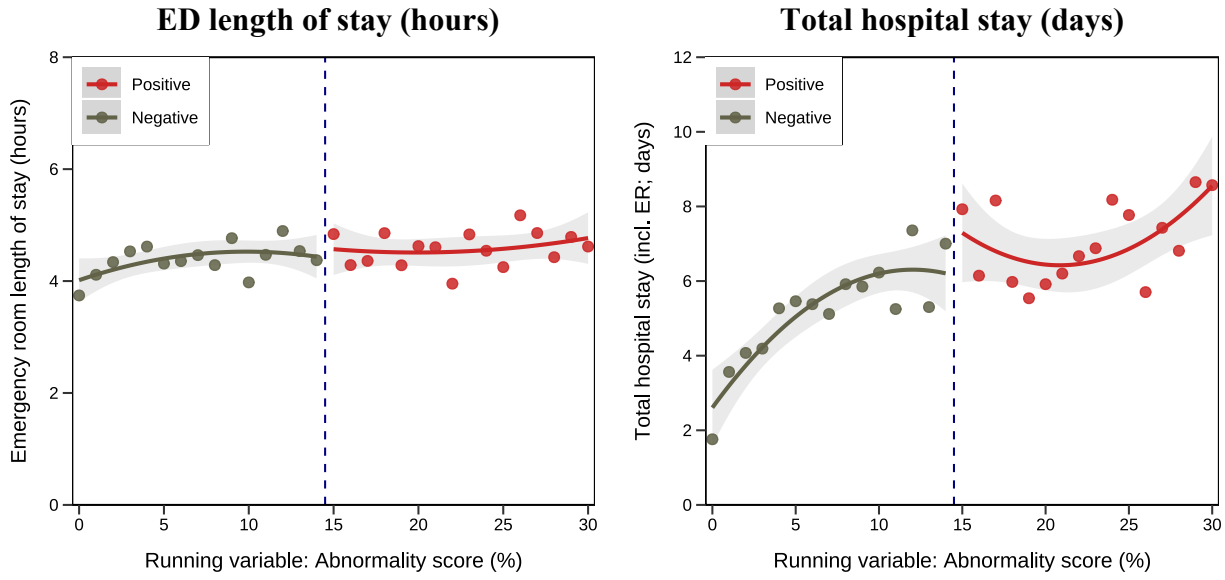


Fig. 2. Effects of AI-augmented CXR on patient management for consolidation. Each point represents a 1% abnormality score bin (cutoff=15%). Outcomes are: (A) proportion of consultations to relevant departments (all departments vs. ED-initiated); (B) length of ED stay and total hospital stay; (C) mean number of CXR during each visit; and (D) proportion of chest CT scan (ordered by all departments vs. ED-initiated) during ED stay. Solid lines indicate second-degree polynomial fits; shaded areas denote standard errors.

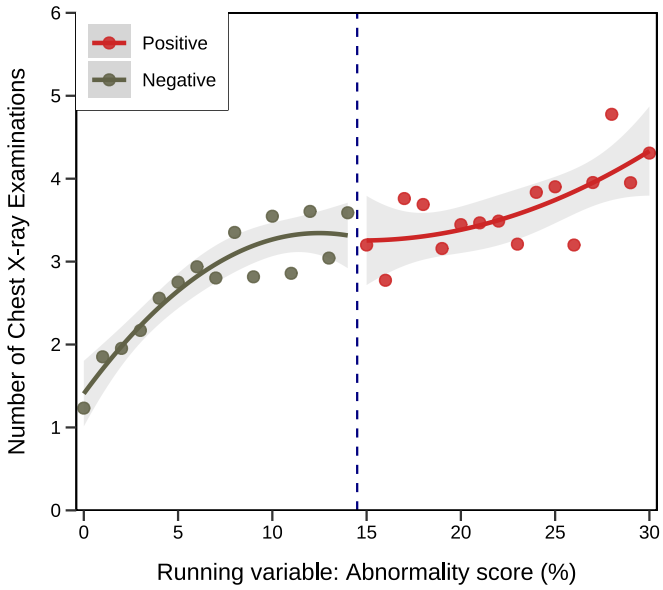
A. Consultation requests to the relevant departments



B. Length of stay



C. Number of CXR examinations



D. Proportion of chest CT scan

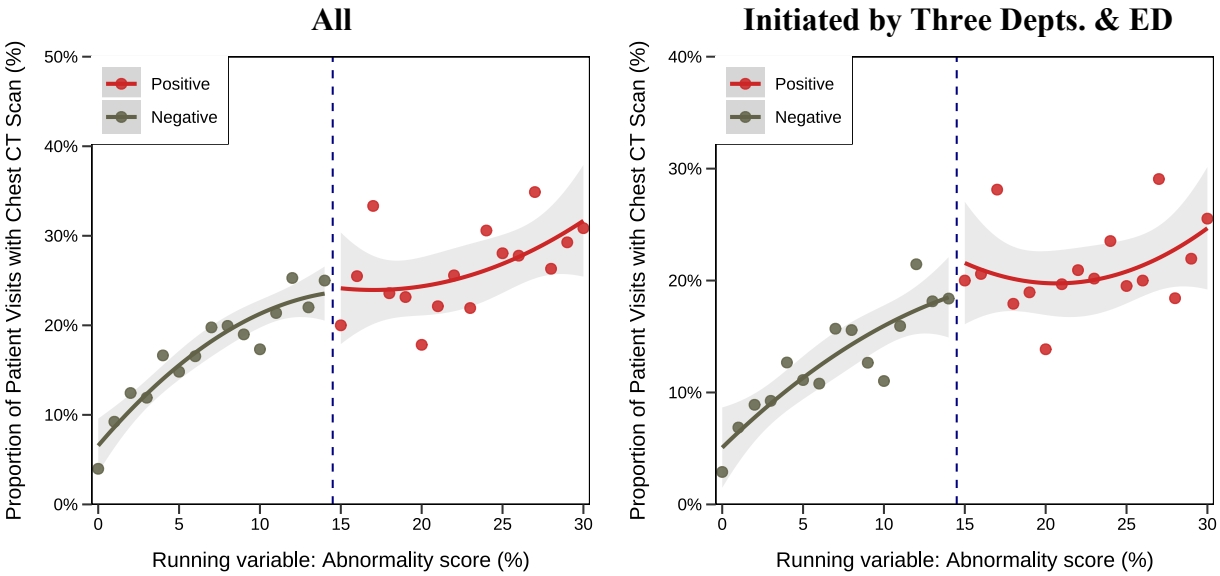
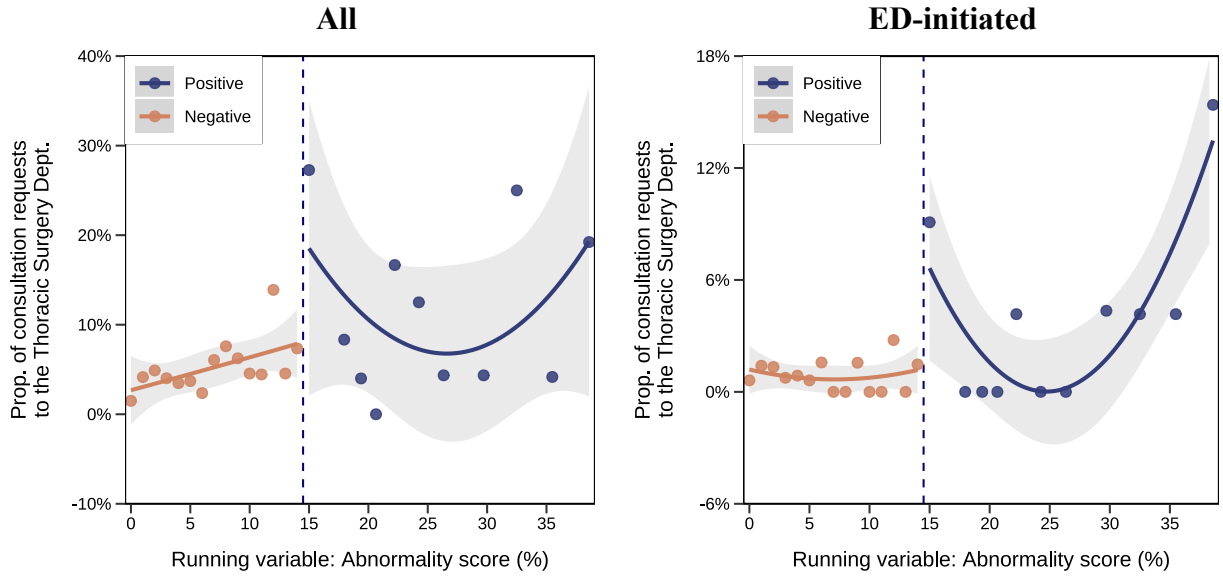
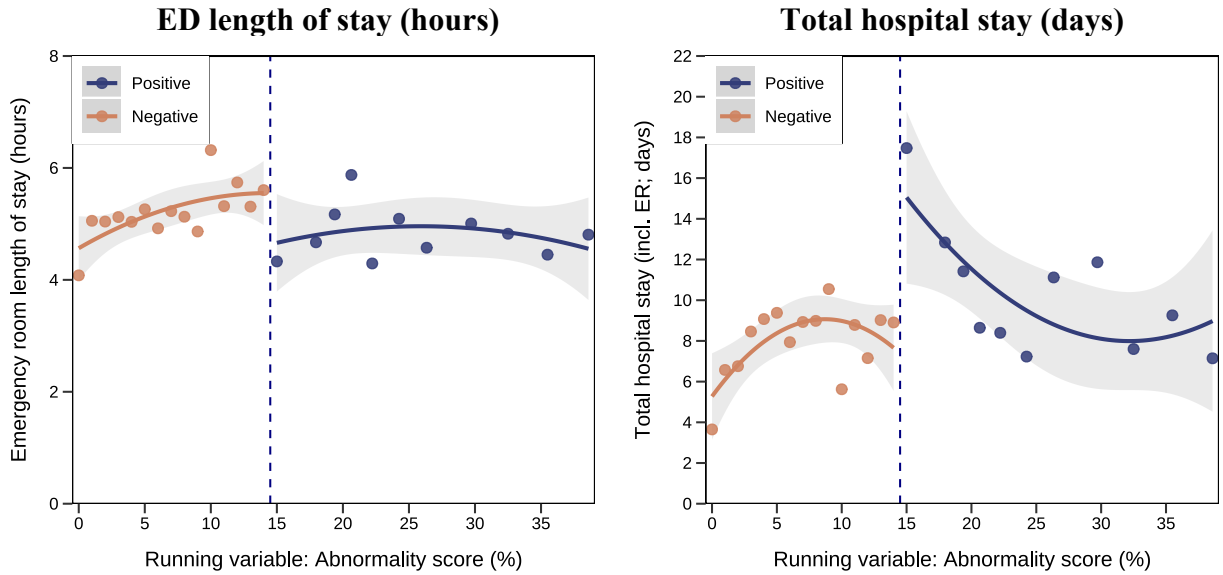


Fig. 3. Effects of AI-augmented CXR on patient management for pneumothorax. Outcomes are: (A) proportion of consultations to the relevant department (all departments vs. ED-initiated); (B) length of ED stay and total hospital stay; (C) mean number of CXR during each visit; and (D) proportion of chest CT scan (ordered by all departments vs. ED-initiated) during ED stay. For scores above 15%, bin width was adjusted to include ≥ 25 observations due to data sparsity. Solid lines indicate second-degree polynomial fits; shaded areas denote standard errors.

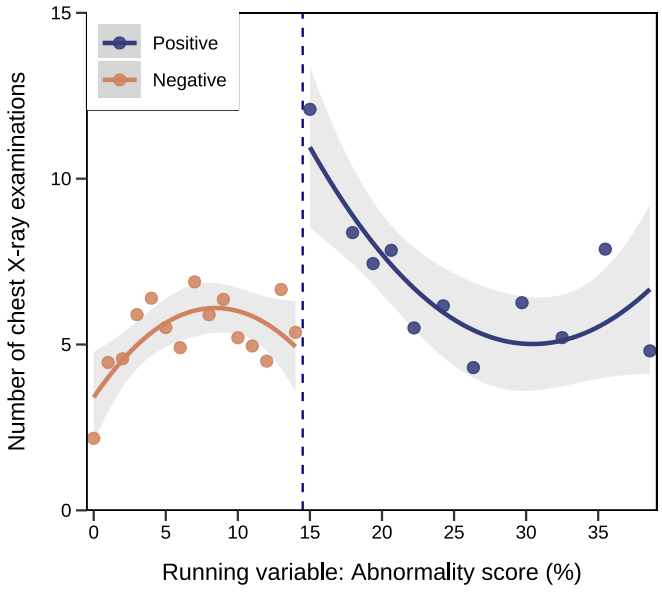
A. Consultation requests to the relevant department



B. Length of stay



C. Number of chest X-ray examinations



D. Proportion of chest CT scan (%)

