JAMA Network Open

Original Investigation | Imaging

# Evaluation of an Artificial Intelligence Model for Detection of Pneumothorax and Tension Pneumothorax in Chest Radiographs

James M. Hillis, MBBS, DPhil; Bernardo C. Bizzo, MD, PhD; Sarah Mercaldo, PhD; John K. Chin, MD; Isabella Newbury-Chaet, BSc; Subba R. Digumarthy, MBBS; Matthew D. Gilman, MD; Victorine V. Muse, MD; Georgie Bottrell, BSc; Jarrel C.Y. Seah, MBBS; Catherine M. Jones, MBBS; Mannudeep K. Kalra, MBBS, MD; Keith J. Dreyer, DO, PhD

## Abstract

**IMPORTANCE** Early detection of pneumothorax, most often via chest radiography, can help determine need for emergent clinical intervention. The ability to accurately detect and rapidly triage pneumothorax with an artificial intelligence (AI) model could assist with earlier identification and improve care.

**OBJECTIVE** To compare the accuracy of an AI model vs consensus thoracic radiologist interpretations in detecting any pneumothorax (incorporating both nontension and tension pneumothorax) and tension pneumothorax.

**DESIGN, SETTING, AND PARTICIPANTS** This diagnostic study was a retrospective standalone performance assessment using a data set of 1000 chest radiographs captured between June 1, 2015, and May 31, 2021. The radiographs were obtained from patients aged at least 18 years at 4 hospitals in the Mass General Brigham hospital network in the United States. Included radiographs were selected using 2 strategies from all chest radiography performed at the hospitals, including inpatient and outpatient. The first strategy identified consecutive radiographs with pneumothorax through a manual review of radiology reports, and the second strategy identified consecutive radiographs with tension pneumothorax using natural language processing. For both strategies, negative radiographs were selected by taking the next negative radiograph acquired from the same radiography machine as each positive radiograph. The final data set was an amalgamation of these processes. Each radiograph was interpreted independently by up to 3 radiologists to establish consensus ground-truth interpretations. Each radiograph was then interpreted by the AI model for the presence of pneumothorax and tension pneumothorax. This study was conducted between July and October 2021, with the primary analysis performed between October and November 2021.

**MAIN OUTCOMES AND MEASURES** The primary end points were the areas under the receiver operating characteristic curves (AUCs) for the detection of pneumothorax and tension pneumothorax. The secondary end points were the sensitivities and specificities for the detection of pneumothorax and tension pneumothorax.

**RESULTS** The final analysis included radiographs from 985 patients (mean [SD] age, 60.8 [19.0] years; 436 [44.3%] female patients), including 307 patients with nontension pneumothorax, 128 patients with tension pneumothorax, and 550 patients without pneumothorax. The AI model detected any pneumothorax with an AUC of 0.979 (95% CI, 0.970-0.987), sensitivity of 94.3% (95% CI, 92.0%-96.3%), and specificity of 92.0% (95% CI, 89.6%-94.2%) and tension pneumothorax with an AUC of 0.987 (95% CI, 0.980-0.992), sensitivity of 94.5% (95% CI, 90.6%-97.7%), and specificity of 95.3% (95% CI, 93.9%-96.6%).

*(continued)*

## Key Points

**Question** Can a commercial artificial intelligence model accurately detect simple and tension pneumothorax on chest radiographs?

**Findings** This diagnostic study used 1000 chest radiographs from 4 hospitals in the US to compare artificial intelligence model outputs with consensus thoracic radiologist interpretations and found that the model accurately detected pneumothorax (incorporating both simple and tension pneumothorax) and tension pneumothorax. The model's performance exceeded the US Food and Drug Administration benchmarks for computer-assisted triage devices.

**Meaning** These findings suggest that this artificial intelligence model could assist radiologists through its accurate detection of pneumothorax.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

*Abstract (continued)*

**CONCLUSIONS AND RELEVANCE** These findings suggest that the assessed AI model accurately detected pneumothorax and tension pneumothorax in this chest radiograph data set. The model's use in the clinical workflow could lead to earlier identification and improved care for patients with pneumothorax.

*JAMA Network Open.* 2022;5(12):e2247172. doi:10.1001/jamanetworkopen.2022.47172

## Introduction

The diagnosis of a pneumothorax, especially tension pneumothorax, on chest radiograph can lead to critical interventions for patient care.[1] The automated detection of pneumothorax and tension pneumothorax through artificial intelligence (AI) has been hypothesized to improve patient care in multiple ways.[2,3] First, it could assist in triaging chest radiographs for sooner interpretation by a radiologist based on the suspected presence of a pneumothorax. Second, it could provide a second "set of eyes" to support identification of a pneumothorax.

Several AI models have been developed to assess for the presence of pneumothorax.[4-9] Many models have also received US Food and Drug Administration (FDA) clearance for use in pneumothorax identification.[10-17] These cleared models are all computer-assisted triage devices, which are intended to aid in prioritization and triage of time-sensitive findings.[18] The FDA regulations permit these devices to provide binary outputs about the presence or absence of findings; they cannot provide localization or segmentation outputs. The sensitivities for pneumothorax identification by these FDA-cleared devices range between 84.3% and 94.65%, with specificities between 87.95% and 95.1%.[10-17] At the time of its clearance, the model assessed in this study is the only device that separately identifies tension pneumothorax.

This retrospective study aimed to assess a commercial AI model that detects both pneumothorax and tension pneumothorax. This model used a deep-convolutional neural network and was trained on more than 750 000 chest radiographs.[19] This study examined the model performance across a range of technical and demographic subgroups to assess the generalizability of the model. It also examined the model performance in the presence or absence of ancillary findings to determine how they impact model accuracy.

## Methods

### Study Design

This retrospective diagnostic study using a standalone model performance design was conducted using radiographs from 4 hospitals within the Mass General Brigham network. It was approved by the Mass General Brigham institutional review board with waiver of informed consent per the Common Rule. It was conducted in accordance with relevant guidelines and regulations, including the Health Insurance Portability and Accountability Act. This report followed the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline.

### Cohort Selection

The cohort was selected based on 2 strategies that used the radiology reports within the MGB radiology archive: consecutive radiographs with pneumothorax through manual review and consecutive radiographs with tension pneumothorax through a natural language processing search engine. Each strategy involved taking the next negative radiograph acquired on the same radiography machine after each positive radiograph to avoid temporal and technical bias. The cohort considered all chest radiographs performed at a hospital, including inpatient and outpatient. There were no limitations on the original chest radiography clinical indication. The chest radiographs were

obtained from patients aged at least 18 years. The chest radiographs were each taken from unique patients; only the first chest radiograph from a given patient was included.

For the consecutive radiographs with pneumothorax, the consecutive radiographs were identified in a prospective order between June 1, 2019, and May 31, 2021. There were 85 report-positive radiographs and 85 report-negative radiographs from each of the 4 hospitals. A radiologist reviewed consecutive chest radiography reports at each of the hospitals until these numbers were achieved.

For the consecutive radiographs with tension pneumothorax, the report-positive radiographs were identified between June 1, 2015, and May 31, 2021, using a natural language processing search for *tension pneumothorax* among chest radiography reports from the same 4 hospitals using a commercial radiology report search engine (Nuance mPower Clinical Analytics). A radiologist then confirmed these radiographs were positive through a chest radiology report review. There were 160 report-positive radiographs selected across all 4 hospitals in a consecutive, retrospective manner from the most recent. An equal number of report-negative radiographs were then selected by taking the next report-negative radiograph after each report-positive radiograph on each radiography machine.

All radiographs were deidentified and underwent an image quality review by an American Board of Radiology–certified radiologist. The image quality review identified the patient positioning (erect or supine) and projections present (anteroposterior [AP], posteroanterior [PA], and/or lateral). Radiographs were excluded if they did not include a chest radiograph or did not include a frontal projection (AP or PA). The review was performed using the FDA-cleared eUnity image visualization software (version 6 or higher) and an internal web-based annotation system.

## Ground Truth Interpretations

Ground truth interpretations were performed by 3 American Board of Radiology–certified radiologists with fellowship training in thoracic radiology. They provided their interpretations independently, without access to the original radiology reports and in different worklist orders. They used the same image visualization software and annotation systems as were used in the image quality review.

The radiologists answered 1 or 3 multiple choice questions about pneumothorax: whether a pneumothorax was present or absent, and, if present, whether the size of pneumothorax was less than 2 cm or 2 cm or greater, and whether features of tension pneumothorax were present or absent. They also indicated the presence of 10 ancillary findings: pleural effusion (including hemothorax), rib fracture, pneumomediastinum, pneumoperitoneum, subcutaneous emphysema, focal or diffuse pulmonary anomalies (including nodules, masses, emphysema, airspace, or interstitial processes), skin fold, intercostal drain, evidence of thoracic surgery (including thoracotomy wires or sutures), and other lines, tubes, or devices (including pacemaker, endotracheal tube, and central line).

For determining consensus for the 3 pneumothorax findings, a 2 + 1 strategy was used: the first 2 radiologists interpreted every radiograph and a third radiologist then interpreted radiographs with discrepant interpretations. Any persistent discrepancies were resolved at a meeting of all 3 radiologists (which could occur for the size and features of tension pneumothorax questions after the 3 interpretations if only 2 radiologists interpreted the radiograph as having a pneumothorax present). After review by the first 2 radiologists, there were 80 radiographs that required adjudication of pneumothorax, 34 radiographs that required adjudication of size, and 47 radiographs that required adjudication of features of tension. After review by the third radiologist, there were 10 radiographs that required adjudication of size and 2 radiographs that required adjudication of features of tension. The ancillary findings were considered present if any radiologist interpreted them as being present.

## Model Inference

The evaluated AI model was version 2.0.0 of the Annalise Enterprise CXR Triage Pneumothorax device. This device is a variant of the Annalise Enterprise (CXR module) device, which is commercially

available in some non-US markets and whose development has been previously described.[19] In brief, it consists of 3 convolutional neural networks designed for clinical decision support: an image projection classification model (attributes model), a clinical finding classification model (classification model), and a clinical finding segmentation model (segmentation model). It can identify more than 100 different radiological findings and was trained on more than 750 000 chest radiographs, which were each labeled by 3 radiologists.

The Annalise Enterprise CXR Triage Pneumothorax device only provides binary classification outputs about the identification of pneumothorax and tension pneumothorax, which is consistent with FDA regulations for computer-assisted triage devices. The model was installed at Mass General Brigham for use in this study and received only the Digital Imaging and Communications in Medicine (DICOM)–formatted chest radiographs. It outputted a classification score between 0 and 1 for each of pneumothorax and tension pneumothorax. A binary output for each of pneumothorax and tension pneumothorax could be derived using predefined operating points, which corresponded to the operating points used in the commercially available model.

### Statistical Analysis

The statistical analysis was performed in R (version 4.0.2) on the full analysis set. The predefined primary end points were the areas under the receiver operating characteristic curves (AUCs) for the detection of pneumothorax (incorporating nontension and tension pneumothorax) and tension pneumothorax. AUCs were calculated using the consensus annotations and the classification scores from the AI model. The predefined secondary end points were the sensitivities and specificities for the detection of pneumothorax and tension pneumothorax. Sensitivities and specificities were calculated by comparing the binary model output with the consensus annotations (ie, by calculating the number of true-positive, false-negative, true-negative, and false-positive radiographs).

These analyses for any pneumothorax and tension pneumothorax were repeated as exploratory analyses for the subgroups of sex, age, manufacturer, patient positioning, chest radiograph projections, and ancillary finding presence. The analyses for pneumothorax were additionally repeated for the subgroups of features of tension pneumothorax and pneumothorax size. The patient's sex and age radiography machine manufacturer were derived from clinical databases or DICOM fields for each radiograph. Any missing data were treated as unknown, and no data were imputed. The manufacturers in the DICOM field may have represented manufacturers other than the radiography machine (eg, the cassette). The patient positioning, projections, and ancillary findings were obtained as part of the cohort selection or ground truth interpretations as described.

All 95% CIs were calculated using bootstrapped intervals with 2000 resamples. The predefined passing criterion for the primary end points was AUC greater than 0.95 based on recognized FDA performance benchmarks.[18] While not predefined passing criteria, this study refers to additional benchmarks of sensitivity greater than 80% and specificity greater than 80%. The sample sizes for each of pneumothorax and tension pneumothorax were calculated using the model results in Seah et al[19] and to ensure the lower bound of the 95% CI for AUC was greater than 0.95. This study was conducted between July and October 2021, with the primary analysis performed between October and November 2021.

## Results

### Participants

An initial cohort of 1000 chest radiographs were selected for this project (eFigure in Supplement 1). Three radiographs were excluded because they were not a chest radiograph during an image quality review. Twelve radiographs were unsuccessful at model inference. The remaining radiographs from 985 patients (mean [SD] age, 60.8 [19.0] years; 436 [44.3%] female patients) were used for analysis. They included 435 radiographs (44.2%) determined to be positive for pneumothorax and 550 radiographs (55.8%) determined to be negative for pneumothorax. There were 128 radiographs that

were positive for tension pneumothorax (29.4% of radiographs with pneumothorax; 13.0% of all radiographs). The demographic and technical details are provided in **Table 1**.

## Pneumothorax Detection

The AI model identified pneumothorax (incorporating nontension and tension pneumothorax) with an AUC of 0.979 (95% CI, 0.970-0.987) (**Figure**, A), sensitivity of 94.3% (95% CI, 92.0%-96.3%), and specificity of 92.0% (95% CI, 89.6%-94.2%). When considering the subgroup analyses for sex, age, manufacturer, patient positioning, and projections, all subgroups achieved an AUC of at least 0.95, sensitivity of at least 80%, and specificity of at least 80%, except the Kodak and unknown manufacturer subgroups (**Table 2**). These 2 subgroups had small cohort sizes and underperformed due to false-negative radiographs: 2 false-negative radiographs of 7 total positive radiographs for the Kodak manufacturer subgroup and 1 false-negative radiograph of 2 total positive radiographs for the unknown manufacturer subgroup. The model achieved higher sensitivity for detecting pneumothorax among tension (compared with nontension) pneumothorax and for pneumothorax 2 cm or greater (compared with <2 cm). It detected a pneumothorax in all radiographs that had a tension pneumothorax.

Table 1. Demographic and Technical Breakdown of Included Patients With Chest Radiographs

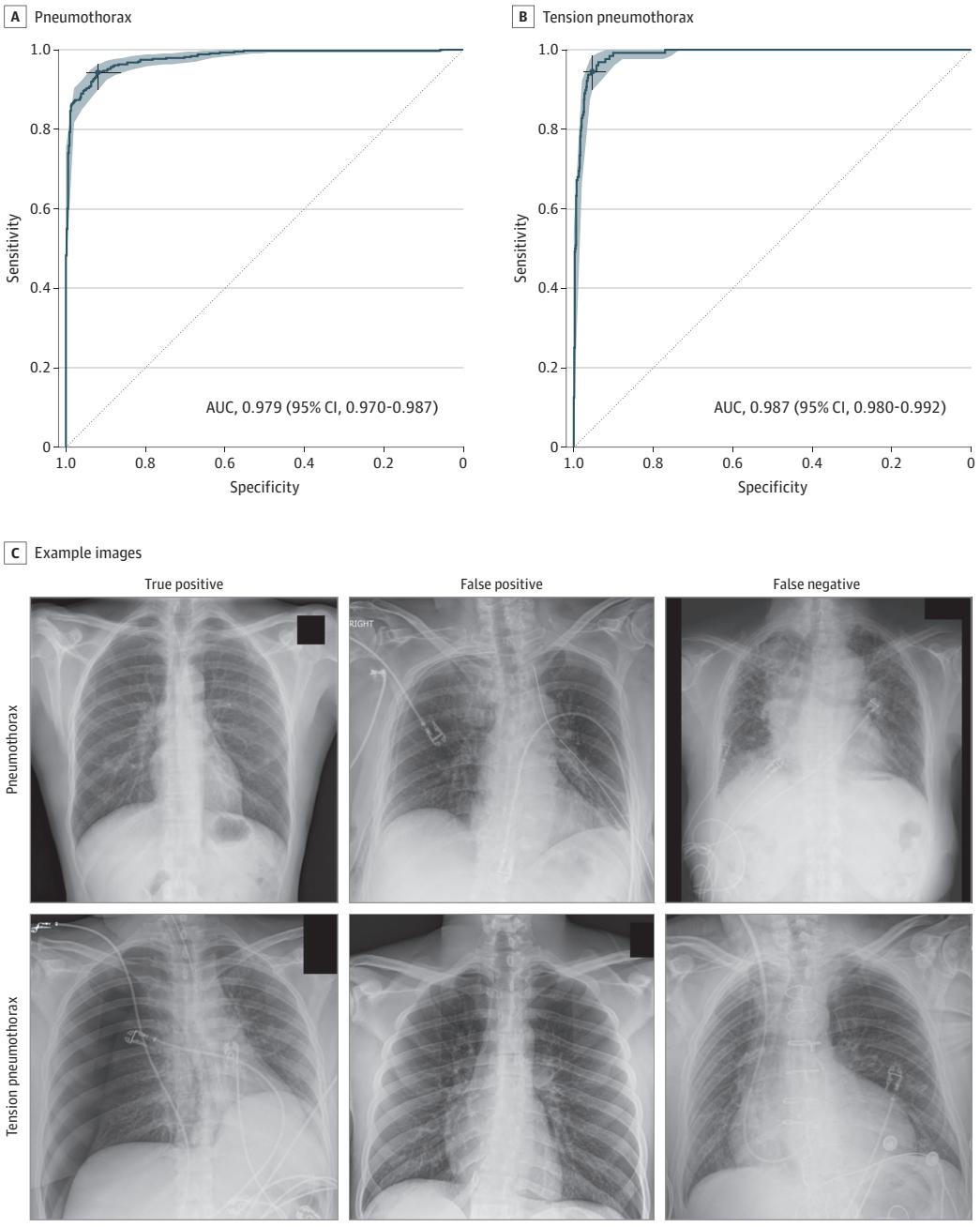
| Characteristic | Patients, No. (%) | | | |
|---|---|---|---|---|
| | Tension pneumothorax (n = 128) | Nontension pneumothorax (n = 307) | Any pneumothorax (n = 435) | No pneumothorax (n = 550) |
| Sex | | | | |
| Female | 56 (43.8) | 125 (40.7) | 181 (41.6) | 255 (46.4) |
| Male | 72 (56.2) | 182 (59.3) | 254 (58.4) | 295 (53.6) |
| Age, y | | | | |
| ≤65 | 80 (62.5) | 154 (50.2) | 234 (53.8) | 289 (52.5) |
| >65 | 48 (37.5) | 153 (49.8) | 201 (46.2) | 261 (47.5) |
| Mean (SD) | 57.5 (18.9) | 60.2 (19.4) | 59.4 (19.3) | 61.9 (18.7) |
| Radiography machine manufacturer | | | | |
| Agfa | 70 (54.7) | 91 (29.6) | 161 (37.0) | 194 (35.3) |
| Carestream | 4 (3.1) | 37 (12.1) | 41 (9.4) | 47 (8.5) |
| Fujifilm | 7 (5.5) | 24 (7.8) | 31 (7.1) | 74 (13.5) |
| GE Healthcare | 0 | 3 (1.0) | 3 (0.7) | 7 (1.3) |
| Kodak | 0 | 7 (2.3) | 7 (1.6) | 14 (2.5) |
| Konica Minolta | 24 (18.8) | 38 (12.4) | 62 (14.3) | 78 (14.2) |
| McKesson | 7 (5.5) | 30 (9.8) | 37 (8.5) | 13 (2.4) |
| Philips | 5 (3.9) | 43 (14.0) | 48 (11.0) | 52 (9.5) |
| Siemens | 5 (3.9) | 15 (4.9) | 20 (4.6) | 32 (5.8) |
| Varian | 6 (4.7) | 17 (5.5) | 23 (5.3) | 26 (4.7) |
| Unknown | 0 | 2 (0.7) | 2 (0.5) | 13 (2.4) |
| Patient positioning | | | | |
| Erect | 69 (53.9) | 221 (72.0) | 290 (66.7) | 370 (67.3) |
| Supine | 59 (46.1) | 86 (28.0) | 145 (33.3) | 180 (32.7) |
| Projections (can include multiple) | | | | |
| AP | 111 (86.7) | 202 (65.8) | 313 (72.0) | 386 (70.2) |
| PA | 17 (13.3) | 105 (34.2) | 122 (28.0) | 164 (29.8) |
| Lateral | 20 (15.6) | 104 (33.9) | 124 (28.5) | 177 (32.2) |
| Pneumothorax size, cm | | | | |
| <2 | 0 | 164 (53.4) | 164 (37.7) | NA |
| ≥2 | 128 (100) | 143 (46.6) | 271 (62.3) | NA |

Abbreviations: AP, anteroposterior; NA, not applicable; PA, posteroanterior.

## Tension Pneumothorax Detection

The AI model identified tension pneumothorax with an AUC of 0.987 (95% CI, 0.980-0.992) (Figure, B), sensitivity of 94.5% (95% CI, 90.6%-97.7%), and specificity of 95.3% (95% CI, 93.9%-96.6%). When considering the subgroup analyses for sex, age, manufacturer, patient positioning, and projections, all subgroups achieved an AUC of at least 0.95, sensitivity of at least 80%, and specificity of at least 80% (**Table 3**).

**Figure. Areas Under the Receiver Operating Characteristic Curve (AUC) for Pneumothorax and Tension Pneumothorax Detection and Example Images**




The shaded region reflects the bootstrapped 95% CI. The selected point on each graph reflects the model operating point. All 3 radiographs with pneumothorax were correctly interpreted by the artificial intelligence model to be negative for tension pneumothorax and all 3 radiographs with tension pneumothorax were correctly interpreted by the AI model to be positive for pneumothorax. Solid boxes cover text annotations on images.

### Detection of Pneumothorax and Tension Pneumothorax When Ancillary Findings Present

The AI model accuracy was assessed in the presence or absence of 10 ancillary findings. It mostly achieved an AUC of 0.95, sensitivity of 80%, and specificity of 80% for the detection of both pneumothorax and tension pneumothorax in the presence or absence of each finding (**Table 4**). The exceptions included performing with an AUC less than 0.95 and specificity less than 80% for the detection of pneumothorax in the presence of findings commonly associated with pneumothorax, including rib fracture, pneumomediastinum, subcutaneous emphysema, and intercostal drain. The model also had an AUC less than 0.95 for the detection of pneumothorax when there was evidence of thoracic surgery. In these situations, the model still performed with an AUC greater than 0.90 (Table 4).

## Discussion

This retrospective diagnostic study assessed the performance of an AI model in detecting pneumothorax and tension pneumothorax via chest radiography. The model overall achieved an AUC

Table 2. Performance for Detecting Pneumothorax Across Demographic and Technical Subgroups

| | Pneumothorax, No. | | | % (95% CI) | |
|---|---|---|---|---|---|
| Characteristic | Positive | Negative | AUC (95% CI) | Sensitivity | Specificity |
| Overall | 435 | 550 | 0.979 (0.970-0.987) | 94.3 (92.0-96.3) | 92.0 (89.6-94.2) |
| Sex | | | | | |
| Female | 181 | 255 | 0.978 (0.963-0.989) | 92.8 (89.0-96.1) | 92.2 (88.6-95.3) |
| Male | 254 | 295 | 0.981 (0.971-0.989) | 95.3 (92.5-97.6) | 91.9 (88.8-94.9) |
| Age, y | | | | | |
| ≤65 | 234 | 289 | 0.983 (0.973-0.990) | 95.7 (93.2-97.9) | 90.7 (87.2-93.8) |
| >65 | 201 | 261 | 0.976 (0.961-0.988) | 92.5 (88.6-96.0) | 93.5 (90.4-96.2) |
| Radiography machine manufacturer | | | | | |
| Agfa | 161 | 194 | 0.974 (0.958-0.985) | 94.4 (90.7-97.5) | 85.6 (80.4-90.2) |
| Carestream | 41 | 47 | 0.988 (0.960-1.000) | 95.1 (87.8-100) | 91.5 (83.0-97.9) |
| Fujifilm | 31 | 74 | 0.998 (0.991-1.000) | 96.8 (90.3-100) | 95.9 (90.5-100) |
| GE Healthcare | 3 | 7 | 1.000 (1.000-1.000) | 100 (100-100) | 100 (100-100) |
| Kodak | 7 | 14 | 0.898 (0.663-1.000) | 71.4 (42.9-100) | 100 (100-100) |
| Konica Minolta | 62 | 78 | 0.973 (0.943-0.993) | 96.8 (91.9-100) | 89.7 (82.1-96.2) |
| McKesson | 37 | 13 | 1.000 (1.000-1.000) | 97.3 (91.9-100) | 100 (100-100) |
| Philips | 48 | 52 | 0.988 (0.967-0.998) | 89.6 (79.2-97.9) | 100 (100-100) |
| Siemens | 20 | 32 | 1.000 (1.000-1.000) | 100 (100-100) | 96.9 (90.6-100) |
| Varian | 23 | 26 | 0.990 (0.955-1.000) | 91.3 (78.3-100) | 100 (100-100) |
| Unknown | 2 | 13 | 1.000 (1.000-1.000) | 50.0 (0.0-100) | 100 (100-100) |
| Patient positioning | | | | | |
| Erect | 290 | 370 | 0.982 (0.972-0.990) | 93.1 (90.0-95.9) | 94.6 (92.2-96.8) |
| Supine | 145 | 180 | 0.976 (0.960-0.988) | 96.6 (93.1-99.3) | 86.7 (81.7-91.7) |
| AP or PA projection | | | | | |
| Without lateral | 311 | 373 | 0.977 (0.968-0.985) | 95.2 (92.6-97.4) | 88.5 (85.3-91.4) |
| With lateral | 124 | 177 | 0.988 (0.971-0.997) | 91.9 (87.1-96.8) | 99.4 (98.3-100) |
| Features of tension | | | | | |
| Absent | 307 | NA | NA | 91.9 (88.6-94.8) | NA |
| Present | 128 | NA | NA | 100 (100-100) | NA |
| Pneumothorax size, cm | | | | | |
| <2 | 164 | NA | NA | 86.6 (81.1-91.5) | NA |
| ≥2 | 271 | NA | NA | 98.9 (97.4-100) | NA |

Abbreviations: AP, anteroposterior; AUC, area under the curve; NA, not applicable; PA, posteroanterior.

greater than 0.95, sensitivity greater than 80%, and specificity greater than 80%. These results are consistent with other FDA-cleared pneumothorax detection models.[10-14,16,17] However, this model is the first FDA-cleared model that distinguishes tension pneumothorax from any pneumothorax.

This study demonstrated consistent accuracy of the AI model in identifying pneumothorax and tension pneumothorax across demographic and technical subgroups, including age, sex, manufacturer, patient positioning, and chest radiograph projection. This performance indicates that the model has generalizability across these subgroups, which hopefully will translate into robust performance as the model encounters further clinical scenarios moving forward. It also achieved sensitivity greater than 80% for detecting pneumothorax regardless of pneumothorax size and whether the features of tension pneumothorax were present or absent. Unsurprisingly, it had a higher sensitivity for detecting a pneumothorax when it was larger (≥2 cm) or had features of tension: the model detected a pneumothorax in all radiographs that had a tension pneumothorax.

The AI model similarly demonstrated consistent model accuracy across most ancillary findings. However, there were some ancillary findings that the model performed less well on. In particular, when the ancillary findings pneumomediastinum, subcutaneous emphysema, and intercostal drain were present, the model had a specificity of less than 80% for detection of pneumothorax. These 3 findings are commonly associated with pneumothorax, and it is possible that the model uses them to identify pneumothorax; hence their presence could contribute to the model determining a radiograph was positive for pneumothorax. It is also possible that their presence suggests a resolved pneumothorax that the model considers positive. From these perspectives, the model is similar to a human reader who might use these ancillary findings to increase sensitivity with a cost to specificity. A follow-on experiment could use the segmentation function of this model to understand where the

Table 3. Performance for Detecting Tension Pneumothorax Across Demographic and Technical Subgroups

| Characteristic | Tension pneumothorax, No. | | AUC (95% CI) | % (95% CI) | |
|---|---|---|---|---|---|
| | Positive | Negative | | Sensitivity | Specificity |
| Overall | 128 | 857 | 0.987 (0.980-0.992) | 94.5 (90.6-97.7) | 95.3 (93.9-96.6) |
| Sex | | | | | |
| Female | 56 | 380 | 0.989 (0.979-0.996) | 94.6 (89.2-100) | 96.6 (94.5-98.2) |
| Male | 72 | 477 | 0.986 (0.977-0.993) | 94.4 (88.9-98.6) | 94.3 (92.0-96.4) |
| Age, y | | | | | |
| ≤65 | 80 | 443 | 0.986 (0.975-0.993) | 93.8 (87.5-98.8) | 94.8 (92.8-96.8) |
| >65 | 48 | 414 | 0.989 (0.981-0.995) | 95.8 (89.6-100) | 95.9 (94.0-97.6) |
| Radiography machine manufacturer | | | | | |
| Agfa | 70 | 285 | 0.980 (0.966-0.990) | 92.9 (85.7-98.6) | 92.6 (89.5-95.4) |
| Carestream | 4 | 84 | 0.991 (0.958-1.000) | 100 (100-100) | 96.4 (91.7-100) |
| Fujifilm | 7 | 98 | 0.991 (0.969-1.000) | 85.7 (57.1-100) | 96.9 (92.9-100) |
| GE Healthcare | 0 | 10 | NA | NA | 100 (100-100) |
| Kodak | 0 | 21 | NA | NA | 95.2 (85.7-100) |
| Konica Minolta | 24 | 116 | 0.998 (0.990-1.000) | 100 (100-100) | 97.4 (94.8-100) |
| McKesson | 7 | 43 | 0.997 (0.970-1.000) | 100 (100-100) | 95.3 (88.4-100) |
| Philips | 5 | 95 | 1.000 (0.987-1.000) | 80.0 (40.0-100) | 98.9 (96.8-100) |
| Siemens | 5 | 47 | 1.000 (1.000-1.000) | 100 (100-100) | 100 (100-100) |
| Varian | 6 | 43 | 0.969 (0.907-1.000) | 100 (100-100) | 86.0 (74.4-95.3) |
| Unknown | 0 | 15 | NA | NA | 100 (100-100) |
| Patient positioning | | | | | |
| Erect | 69 | 591 | 0.988 (0.980-0.994) | 98.6 (95.7-100) | 95.1 (93.4-96.8) |
| Supine | 59 | 266 | 0.989 (0.977-0.996) | 89.8 (81.4-96.6) | 95.9 (93.2-98.1) |
| AP or PA projection | | | | | |
| Without lateral | 108 | 576 | 0.987 (0.979-0.993) | 93.5 (88.9-98.1) | 95.0 (93.1-96.7) |
| With lateral | 20 | 281 | 0.990 (0.979-0.998) | 100 (100-100) | 96.1 (94.0-98.2) |

Abbreviations: AP, anteroposterior; AUC, area under the curve; NA, not applicable; PA, posteroanterior.

Table 4. Performance for Detecting Any Pneumothorax and Tension Pneumothorax in the Presence of Different Ancillary Findings

| | Tension pneumothorax, No. | | | % (95% CI) | |
|---|---|---|---|---|---|
| Finding | Positive | Negative | AUC (95% CI) | Sensitivity | Specificity |
| **Any pneumothorax** | | | | | |
| Pleural effusion | | | | | |
| Present | 174 | 181 | 0.966 (0.947-0.982) | 92.0 (87.9-96.0) | 87.3 (82.3-91.7) |
| Absent | 261 | 369 | 0.985 (0.974-0.993) | 95.8 (93.1-98.1) | 94.3 (91.9-96.5) |
| Rib fracture | | | | | |
| Present | 53 | 49 | 0.945 (0.898-0.980) | 90.6 (83.0-98.1) | 77.6 (65.3-89.8) |
| Absent | 382 | 501 | 0.983 (0.974-0.989) | 94.8 (92.4-96.9) | 93.4 (91.2-95.6) |
| Pneumomediastinum | | | | | |
| Present | 31 | 6 | 0.930 (0.817-0.989) | 100 (100-100) | 33.3 (0.0-66.7) |
| Absent | 404 | 544 | 0.979 (0.970-0.987) | 93.8 (91.3-96.0) | 92.6 (90.4-94.7) |
| Pneumoperitoneum | | | | | |
| Present | 8 | 4 | 1.000 (1.000-1.000) | 100 (100-100) | 100 (100-100) |
| Absent | 427 | 546 | 0.980 (0.971-0.987) | 94.1 (91.8-96.3) | 91.9 (89.6-94.1) |
| Subcutaneous emphysema | | | | | |
| Present | 115 | 36 | 0.904 (0.839-0.954) | 99.1 (97.4-100) | 41.7 (25.0-58.3) |
| Absent | 320 | 514 | 0.982 (0.972-0.990) | 92.5 (89.4-95.3) | 95.5 (93.6-97.3) |
| Focal or diffuse pulmonary anomalies | | | | | |
| Present | 306 | 300 | 0.972 (0.960-0.982) | 93.1 (89.9-96.1) | 89.0 (85.3-92.3) |
| Absent | 129 | 250 | 0.986 (0.966-0.996) | 96.9 (93.8-99.2) | 95.6 (92.8-98.0) |
| Skin fold | | | | | |
| Present | 64 | 128 | 0.964 (0.924-0.991) | 92.2 (85.9-98.4) | 95.3 (91.4-98.4) |
| Absent | 371 | 422 | 0.982 (0.975-0.989) | 94.6 (92.5-96.8) | 91.0 (88.2-93.6) |
| Intercostal drain | | | | | |
| Present | 144 | 47 | 0.931 (0.893-0.961) | 98.6 (96.5-100) | 46.8 (31.9-61.7) |
| Absent | 291 | 503 | 0.982 (0.971-0.990) | 92.1 (89.0-95.2) | 96.2 (94.4-97.8) |
| Evidence of thoracic surgery | | | | | |
| Present | 109 | 102 | 0.945 (0.910-0.974) | 93.6 (89.0-97.2) | 81.4 (73.5-88.2) |
| Absent | 326 | 448 | 0.987 (0.981-0.992) | 94.5 (92.0-96.6) | 94.4 (92.2-96.4) |
| Other lines, tubes, or devices | | | | | |
| Present | 148 | 174 | 0.978 (0.961-0.989) | 95.3 (91.9-98.6) | 88.5 (83.3-93.1) |
| Absent | 287 | 376 | 0.981 (0.970-0.989) | 93.7 (90.9-96.5) | 93.6 (91.0-96.0) |
| **Tension pneumothorax** | | | | | |
| Pleural effusion | | | | | |
| Present | 39 | 316 | 0.984 (0.970-0.993) | 94.9 (87.2-100) | 95.6 (93.0-97.8) |
| Absent | 89 | 541 | 0.990 (0.981-0.995) | 94.4 (89.9-98.9) | 95.2 (93.3-96.9) |
| Rib fracture | | | | | |
| Present | 10 | 92 | 0.973 (0.937-0.995) | 80.0 (50.0-100) | 93.5 (88.0-97.8) |
| Absent | 118 | 765 | 0.989 (0.982-0.994) | 95.8 (91.5-99.2) | 95.6 (94.0-97.0) |
| Pneumomediastinum | | | | | |
| Present | 10 | 27 | 0.989 (0.941-1.000) | 90.0 (70.0-100) | 88.9 (74.1-100) |
| Absent | 118 | 830 | 0.988 (0.981-0.993) | 94.9 (90.7-98.3) | 95.5 (94.2-96.9) |
| Pneumoperitoneum | | | | | |
| Present | 3 | 9 | 1.000 (1.000-1.000) | 100 (100-100) | 100 (100-100) |
| Absent | 125 | 848 | 0.987 (0.980-0.993) | 94.4 (90.4-97.6) | 95.3 (93.8-96.7) |

(continued)

Table 4. Performance for Detecting Any Pneumothorax and Tension Pneumothorax in the Presence of Different Ancillary Findings (continued)

| Finding | Tension pneumothorax, No. | | AUC (95% CI) | % (95% CI) | |
|---|---|---|---|---|---|
| | Positive | Negative | | Sensitivity | Specificity |
| Subcutaneous emphysema | | | | | |
| Present | 23 | 128 | 0.964 (0.911-0.994) | 82.6 (65.2-95.7) | 95.3 (91.4-98.4) |
| Absent | 105 | 729 | 0.990 (0.984-0.994) | 97.1 (93.3-100) | 95.3 (93.7-96.8) |
| Focal or diffuse pulmonary anomalies | | | | | |
| Present | 89 | 517 | 0.982 (0.972-0.990) | 93.3 (87.6-97.8) | 94.4 (92.3-96.3) |
| Absent | 39 | 340 | 0.995 (0.989-0.999) | 97.4 (92.3-100) | 96.8 (94.7-98.5) |
| Skin fold | | | | | |
| Present | 11 | 181 | 0.976 (0.937-0.997) | 90.9 (72.7-100) | 96.1 (93.4-98.9) |
| Absent | 117 | 676 | 0.989 (0.983-0.993) | 94.9 (90.6-98.3) | 95.1 (93.5-96.6) |
| Intercostal drain | | | | | |
| Present | 21 | 170 | 0.976 (0.930-0.998) | 85.7 (71.4-100) | 97.1 (94.1-99.4) |
| Absent | 107 | 687 | 0.989 (0.982-0.994) | 96.3 (92.5-99.1) | 94.9 (93.2-96.4) |
| Evidence of thoracic surgery | | | | | |
| Present | 16 | 195 | 0.987 (0.970-0.997) | 87.5 (68.8-100) | 97.4 (94.9-99.5) |
| Absent | 112 | 662 | 0.988 (0.980-0.994) | 95.5 (91.1-99.1) | 94.7 (93.1-96.4) |
| Other lines, tubes, or devices | | | | | |
| Present | 67 | 255 | 0.987 (0.975-0.995) | 91.0 (83.6-97.0) | 95.3 (92.2-97.6) |
| Absent | 61 | 602 | 0.990 (0.983-0.995) | 98.4 (95.1-100) | 95.3 (93.5-97.0) |

Abbreviation: AUC, area under the receiver operating characteristic curve.

model believes the pneumothorax is; this study used the subsequently FDA-cleared version that included the classification functionality and did not include the segmentation functionality. Similarly, the presence of the ancillary findings rib fracture and evidence of thoracic surgery each had AUC less than 0.95 but still greater than 0.90 for detection of pneumothorax; these results likely reflect the associations of these findings with pneumothorax, albeit not as strong as the associations of pneumomediastinum, subcutaneous emphysema, and intercostal drain.

The cohort selection included 2 strategies: the consecutive radiographs with pneumothorax were identified by a manual review of the original radiology reports from consecutive chest radiographs, while the radiographs with tension pneumothorax were initially identified with natural language processing before the manual review. The first strategy ensured that radiographs reflected the distribution of pneumothorax in the clinical setting, including radiographs with a spectrum of pneumothorax conspicuity. The second strategy was used because of the infrequency of tension pneumothorax among all chest radiographs. One item that we noted was that there were fewer ground truth–positive radiographs compared with the number of original radiology report–positive radiographs. We hypothesized that this discrepancy occurred because of the lack of associated clinical history and the lack of other imaging, such as chest computed tomography, which might have increased the detection rate in the original clinical environment. The ground truth radiologists in this study only had access to the chest radiograph.

## Limitations

This study has some limitations. One key limitation is that it is a retrospective study outside of the clinical workflow. While it demonstrates the accuracy of the AI model in interpreting imaging across many demographic and technical subgroups, it does not do so within the broader clinical environment. Further evaluation will be required to know how the model might impact the clinical workflow, including its impact on radiologists for case prioritization and patient outcomes. Further

evaluation will also be required as the model encounters clinical scenarios beyond the current study, including from new radiographic equipment manufacturers or models.

This study only assessed binary classification performance for the identification of pneumothorax and tension pneumothorax. This limitation was because this model version only provides this output, which is consistent with FDA regulations for a computer-assisted triage device. It is possible that the model detected a different finding on the chest radiograph than the ground truth radiologists, including that it in fact identified a finding on the contralateral side to a pneumothorax. The presence of a segmentation output would allow this possibility to be further investigated. It would be particularly interesting to evaluate the segmentation outputs on false positive radiographs to determine which features caused the model to consider the radiograph to be positive for pneumothorax.

## Conclusions

This diagnostic study assessed an AI model that accurately detected pneumothorax and tension pneumothorax. Its use in the clinical environment may lead to improved care for patients with pneumothorax.

ARTICLE INFORMATION

**Accepted for Publication:** October 17, 2022.

**Published:** December 15, 2022. doi:10.1001/jamanetworkopen.2022.47172



**Corresponding Author:** James M. Hillis, MBBS, DPhil, Data Science Office, Mass General Brigham, 100 Cambridge St, Floor 13, Ste 1303, Boston, MA 02114 (james.hillis@mgh.harvard.edu).

**Author Affiliations:** Data Science Office, Mass General Brigham, Boston, Massachusetts (Hillis, Bizzo, Mercaldo, Chin, Newbury-Chaet, Kalra, Dreyer); Department of Neurology, Massachusetts General Hospital, Boston (Hillis); Harvard Medical School, Boston, Massachusetts (Hillis, Bizzo, Mercaldo, Digumarthy, Gilman, Muse, Kalra, Dreyer); Department of Radiology, Massachusetts General Hospital, Boston (Bizzo, Mercaldo, Digumarthy, Gilman, Muse, Kalra, Dreyer); Annalise-AI, Sydney, Australia (Bottrell, Seah, Jones); Alfred Health, Prahran, Australia (Seah); I-MED Radiology Network, Brisbane, Australia (Jones); Faculty of Medicine and Health, University of Sydney, Sydney, Australia (Jones).

**Author Contributions:** Dr Hillis had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Hillis and Bizzo contributed equally to the work.

*Concept and design:* Hillis, Bizzo, Bottrell, Seah, Jones, Kalra, Dreyer.

*Acquisition, analysis, or interpretation of data:* Hillis, Mercaldo, Chin, Newbury-Chaet, Digumarthy, Gilman, Muse, Kalra, Dreyer.

*Drafting of the manuscript:* Hillis, Mercaldo, Newbury-Chaet, Gilman, Muse, Seah.

*Critical revision of the manuscript for important intellectual content:* Hillis, Bizzo, Mercaldo, Chin, Digumarthy, Bottrell, Jones, Kalra, Dreyer.

*Statistical analysis:* Hillis, Mercaldo.

*Obtained funding:* Hillis, Bizzo, Seah, Dreyer.

*Administrative, technical, or material support:* Bizzo, Chin, Newbury-Chaet, Digumarthy.

*Supervision:* Bizzo, Bottrell, Seah, Jones, Kalra.

**Conflict of Interest Disclosures:** Dr Digumarthy reported receiving grants from GE, Vuno, and QureAI; personal fees from Siemens and Elsevier; and serving as a consultant for Merck, Pfizer, Bristol Myers Squibb, Novartis, Roche, Polaris, Cascadian, AbbVie, Gradalis, Bayer, Zai Laboratories, Biengen, Riverain, Resonance, Biengen, Janssen outside the submitted work. Dr Bottrell reported owning stock in Annalise AI outside the submitted work. Dr Seah reported receiving personal fees from Harrison.ai and having a patent for Annalise CXR Enterprise pending during the conduct of the study. Dr Kalra reported receiving grants from Coreline, Riverain Tech, and Siemens. No other disclosures were reported.

**Funding/Support:** This study was funded by Annalise-AI.

**Role of the Funder/Sponsor:** Annalise-AI was involved in the design of the trial including study protocol development, statistical analysis plan development, and training of ground-truth radiologists together with the Data Science Office at Mass General Brigham. Study conduct, including data collection, management and analysis were performed by the Data Science Office and radiologists from the Mass General Brigham system without involvement of Annalise-AI. Both Annalise-AI and the Data Science Office were involved in the preparation, review and approval of the manuscript and decision to submit the manuscript for publication. Authors who are employees of Mass General Brigham or Massachusetts General Hospital, which received institutional funding from Annalise-AI for the study, or who are employees of Annalise-AI participated in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Data Sharing Statement:** See Supplement 2.

**Additional Contributions:** We thank the broader Mass General Brigham Data Science Office and Annalise teams for their assistance with this project.

## REFERENCES

1. Sahn SA, Heffner JE. Spontaneous pneumothorax. *N Engl J Med*. 2000;342(12):868-874. doi:10.1056/NEJM200003233421207

2. Ebrahimian S, Kalra MK, Agarwal S, et al. FDA-regulated AI algorithms: trends, strengths, and gaps of validation studies. *Acad Radiol*. 2022;29(4):559-566. doi:10.1016/j.acra.2021.09.002

3. Weisberg EM, Chu LC, Fishman EK. The first use of artificial intelligence (AI) in the ER: triage not diagnosis. *Emerg Radiol*. 2020;27(4):361-366. doi:10.1007/s10140-020-01773-6

4. Kitamura G, Deible C. Retraining an open-source pneumothorax detecting machine learning algorithm for improved performance to medical images. *Clin Imaging*. 2020;61:15-19. doi:10.1016/j.clinimag.2020.01.008

5. Wang H, Gu H, Qin P, Wang J. CheXLocNet: automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks. *PLoS One*. 2020;15(11):e0242013. doi:10.1371/journal.pone.0242013

6. Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: a retrospective study. *PLoS Med*. 2018;15(11):e1002697. doi:10.1371/journal.pmed.1002697

7. Cho Y, Kim JS, Lim TH, Lee I, Choi J. Detection of the location of pneumothorax in chest X-rays using small artificial neural networks and a simple training process. *Sci Rep*. 2021;11(1):13054. doi:10.1038/s41598-021-92523-2

8. Feng S, Liu Q, Patel A, et al. Automated pneumothorax triaging in chest X-rays in the New Zealand population using deep-learning algorithms. *J Med Imaging Radiat Oncol*. 2022. doi:10.1111/1754-9485.13393

9. Park S, Lee SM, Kim N, et al. Application of deep learning-based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. *Eur Radiol*. 2019;29(10):5341-5348. doi:10.1007/s00330-019-06130-x

10. US Food and Drug Administration. HealthPNX approval letter (K190362). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K190362.pdf

11. US Food and Drug Administration. Critical Care Suite approval letter (K183182). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf18/K183182.pdf

12. US Food and Drug Administration. red dot Device approval letter (K191556). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K191556.pdf

13. US Food and Drug Administration. AIMI-Triage CXR PTX approval letter (K193300). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K193300.pdf

14. US Food and Drug Administration. Lunit INSIGHT CXR Triage approval letter (K211733). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf21/K211733.pdf

15. US Food and Drug Administration. Annalise Enterprise CXR Triage Pneumothorax approval letter (K213941). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf21/K213941.pdf

16. US Food and Drug Administration. ClearRead Xray Pneumothorax approval letter (K213566). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf21/K213566.pdf

17. US Food and Drug Administration. BriefCase approval letter (K214043). Accessed June 19 2022. https://www.accessdata.fda.gov/cdrh_docs/pdf21/K214043.pdf

18. US Food and Drug Administration. Product classification: radiological computer-assisted prioritization software for lesions. Accessed June 19 2022. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPCD/classification.cfm?id=QFM

19. Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit Health*. 2021;3(8):e496-e506. doi:10.1016/S2589-7500(21)00106-0

**SUPPLEMENT 1.**
**eFigure.** Case Selection Flowchart

**SUPPLEMENT 2.**
**Data Sharing Statement**