CLINICAL STUDY DOCUMENT

Study Protocol

EVOLVE-CXR

Official title

EvoXRare-Assisted Chest Radiograph Interpretation in Emergency and Critically Ill Adults: A Nationwide Multicenter Randomized Controlled Trial

Unique protocol ID	WHUH-CXR-AI-RCT-2026-01
Document version	English review draft 1.0
Document date	01 November 2026
ClinicalTrials.gov ID	To be inserted after assignment.
Sponsor	Union Hospital, Tongji Medical College, Huazhong University of Science and Technology
Principal investigator	Heshui Shi, Department of Radiology
Study coordinator	Xiaoyu Han, Department of Radiology
Funding	National Natural Science Foundation of China (82672507); Interdisciplinary Research Program of HUST (2026JCYJ026).

Study synopsis

A pragmatic, nationwide, multicenter, individually randomized, two-arm superiority trial evaluating whether EvoXRare-assisted chest radiograph interpretation reduces patient-level clinically important diagnostic misses compared with routine physician interpretation alone. Approximately 70,000 adults will be randomized 1:1 across 15–20 hospitals in China. The primary endpoint is the patient-level miss rate for prespecified clinically important abnormalities on the index chest radiograph, determined by blinded central adjudication. Key secondary outcomes address critical and rare findings, diagnostic and reporting timeliness, false-positive findings, additional imaging, workflow performance, and selected short-term clinical outcomes.

Document status and provenance

Prepared to accompany the ClinicalTrials.gov registration draft for EVOLVE-CXR. Core design, population, intervention, sample size and endpoint definitions are aligned across the registration, protocol and statistical analysis plan. Administrative fields such as final site list, local approvals, consent/waiver determinations and the assigned NCT number must be verified before activation or public posting. This working draft does not itself establish ethics approval, prospective registration, or site activation.

1 Study identification

Unique Protocol ID	WHUH-CXR-AI-RCT-2026-01
Brief Title	EvoXRare AI-Assisted Interpretation of Emergency Chest Radiographs
Official Title	EvoXRare-Assisted Chest Radiograph Interpretation in Emergency and Critically Ill Adults: A Nationwide Multicenter Randomized Controlled Trial
Acronym and Study Type	EVOLVE-CXR; Interventional randomized clinical trial.
Secondary ID 1	82672507; National Natural Science Foundation of China.
Secondary ID 2	2026JCYJ026; Interdisciplinary Research Program of HUST.

2 Study status and timeline

Recruitment Status	Planned: not yet recruiting; site-specific status updated after activation.
Anticipated Study Start	November 2026.
Anticipated Recruitment	November 2026 through December 2027.
Primary Completion	Anticipated first half of 2028 after final endpoint adjudication.
Study Completion	Anticipated first half of 2028 after prespecified short-term outcome ascertainment and database lock.
Target Centers	Approximately 15–20 hospitals across China.

3 Sponsors, governance and oversight

Responsible Party	Sponsor.
Sponsor	Union Hospital, Tongji Medical College, Huazhong University of Science and Technology
Principal Investigator	Heshui Shi, Department of Radiology
Study Coordinator	Xiaoyu Han, Department of Radiology
Collaborators	Approximately 15–20 participating hospitals across China; final facilities and local investigators are listed after agreements and activation.
Funding	National Natural Science Foundation of China (82672507); Interdisciplinary Research Program of HUST (2026JCYJ026).

3.1 Ethics and regulatory oversight

The lead-center ethics approval is recorded as No. 0964 in the study materials; the exact official number, approval date and English rendering must be verified from the ethics letter. Each participating site must complete the required ethics and institutional approvals before enrolling participants. The ClinicalTrials.gov record, consent approach and protocol must remain consistent with the final approved materials.

Because both groups undergo clinically indicated chest radiography and the intervention modifies the interpretation workflow rather than the examination itself, participating centers may use written informed consent or an ethics-approved waiver/alteration of consent as permitted by local regulations. No center may assume that another center’s waiver applies automatically.

3.2 Data monitoring

An independent clinical and statistical monitoring group will review serious AI-related diagnostic harm, important protocol deviations, technical failure patterns and data quality at prespecified intervals. No formal interim efficacy test is planned unless introduced by a documented protocol amendment. Safety or operational concerns may lead to suspension of a site or of recruitment.

4 Background and rationale

Chest radiography is widely used in emergency departments, resuscitation rooms and intensive care units because it is rapid, portable and available at the bedside. Diagnostic misses can occur when abnormalities are subtle, uncommon, obscured by overlapping anatomy, or encountered under high workload. Clinically consequential examples include pneumothorax, pneumomediastinum, subdiaphragmatic free air, pleural effusion, lobar atelectasis, acute osseous injury, line/tube malposition and other uncommon thoracic abnormalities.

EvoXRare is an AI-assisted chest radiograph interpretation system designed to provide a draft report and high-risk prompts for prespecified clinically important and uncommon findings. The present trial evaluates the added clinical effect of access to EvoXRare during routine interpretation. It is not a model-development study: software version, model weights, preprocessing, thresholds, supported finding set and report format are frozen before formal enrollment, and randomized participants are excluded from model training or outcome-driven recalibration.

The principal hypothesis is that AI-assisted interpretation reduces the proportion of patients with at least one clinically important abnormality that is present on the reference standard but absent from the initial finalized report, without unacceptable increases in false-positive findings or workflow burden.

5 Objectives and hypotheses

5.1 Primary objective

To determine whether EvoXRare-assisted chest radiograph interpretation reduces the patient-level clinically important abnormality miss rate compared with routine physician interpretation alone.

5.2 Secondary objectives

- Evaluate misses and detection for critical Category A findings.
- Evaluate detection of prespecified rare or easily missed findings overall and by finding class.
- Compare time to recognition of critical abnormalities, report turnaround time and critical-result notification time.
- Compare false-positive clinically important findings, radiology recall, repeat radiography and additional imaging.
- Characterize physician acceptance, rejection and editing of AI suggestions and reasons for override when captured.
- Assess technical availability, AI processing failure, latency and workflow adherence.
- Explore selected short-term clinical outcomes including ICU transfer, hospital length of stay and in-hospital mortality without claiming the trial is powered for mortality benefit.

6 Study design

Design	Nationwide, multicenter, prospective, parallel-group, individually randomized, superiority randomized controlled trial.
Allocation	1:1 participant-level allocation.
Randomization	Central computer-generated allocation with variable block sizes, stratified by center and care setting (emergency/resuscitation vs ICU/other critical care). Portable versus fixed radiography may be an additional operational stratification factor if prespecified before activation.
Masking	Open label for interpreting clinicians; blinded central outcome adjudication. Analysis programmers use coded treatment groups until the SAP is finalized and database lock procedures are complete.

Primary unit	Participant/index encounter; one index chest radiograph encounter per participant contributes to the primary randomized analysis.
Planned enrollment	70,000 randomized participants, approximately 35,000 per group.
Sites	Approximately 15–20 hospitals across China.

The trial uses a pragmatic workflow. Randomization occurs before the assigned interpretation workflow is completed. Clinical care must not be delayed for AI processing, research data entry or adjudication. Accidental AI exposure in the control group, AI failure in the intervention group, or physician override of AI advice is recorded but does not remove the participant from the intention-to-treat analysis.

7 Study population

7.1 Inclusion criteria

- Age 18 years or older.
- Undergoing a clinically indicated chest radiograph in a participating emergency department, resuscitation room, intensive care unit, or other prespecified acute/critical-care setting.
- Chest radiograph is technically accessible to the trial workflow before or during routine clinical interpretation.
- Participant has not previously contributed an index encounter to the primary randomized analysis.
- The applicable consent or ethics-approved waiver/alteration of consent requirement has been satisfied at the activated site.

7.2 Exclusion criteria

- Age younger than 18 years.
- Image missing, corrupted, or technically non-diagnostic to a degree that prevents routine clinical interpretation.
- Radiograph obtained solely for research rather than clinical care.
- Image format or transfer route is outside the locked system specification before randomization; such cases may remain in screening logs.
- Previous randomization of the same participant in the primary trial.
- Any local legal, ethical, or institutional restriction preventing required participation or data use.

7.3 Screening and participant identity

Sites maintain a screening log sufficient to distinguish eligible randomized participants, eligible nonrandomized cases and technical exclusions. A trial-specific identifier is assigned before data transfer. Direct identifiers are retained only at the local site according to institutional policy.

8 Randomization, intervention and comparator

8.1 Experimental arm: EvoXRare-assisted interpretation

The same clinically obtained chest radiograph is processed by the locked EvoXRare system on an independent AI workstation or approved deployment pathway. EvoXRare provides a draft chest radiograph report and/or finding-level high-risk prompts. The interpreting physician reviews the original image, verifies all AI output, edits or rejects suggestions as needed, and remains solely responsible for the finalized clinical report and required communication.

8.2 Control arm: standard interpretation

The interpreting physician follows the participating hospital's usual chest radiograph workflow without access to EvoXRare trial output. Routine nontrial tools already in use at the site may continue only if documented before site activation and applied consistently to both trial periods as specified in the site manual.

8.3 Model control and technical failure

- Freeze software version, weights, preprocessing, thresholds, supported findings, user interface and report format before enrollment.
- No online learning, outcome-driven threshold adjustment or retraining with randomized trial participants during the primary trial.
- If AI processing fails, clinical interpretation proceeds without delay. Failure reason and timing are recorded when feasible.
- Physicians may override any AI suggestion. Override does not constitute a protocol violation unless required trial documentation is intentionally omitted.
- Trial AI output is hidden from control clinicians. Accidental exposure is documented as a crossover/protocol deviation.

9 Prespecified findings and reference standard

9.1 Clinically important finding categories

The final finding list and operational definitions are locked in the adjudication manual before primary database analysis. The following categories define the intended framework and may be refined without changing the core endpoint concept if revisions are approved before unblinding.

Category A (critical / potentially urgent): pneumothorax or hydropneumothorax; tension pneumothorax; subdiaphragmatic free air; pneumomediastinum; extensive subcutaneous emphysema; large/new air-space opacity or consolidation associated with acute clinical deterioration; and other prespecified findings requiring urgent communication according to the study manual.

Category B (important): moderate-to-large pleural effusion; lobar or major atelectasis; marked/new cardiomegaly or pulmonary edema where clinically relevant; acute fracture or destructive osseous abnormality; and malposition or complication of clinically important tubes, lines or devices.

Rare/easily missed findings are prespecified in the study manual and analyzed separately. Category membership is identical in both randomized groups.

9.2 Blinded central adjudication

The reference standard is established by blinded central adjudicators using the index radiograph and clinically available confirmatory information, which may include subsequent CT, ultrasound, pathology or surgery when applicable, laboratory findings, clinical course, discharge diagnosis and follow-up imaging. Adjudicators are masked to randomized assignment and EvoXRare output when determining whether a prespecified finding was present on the index radiograph.

At least two qualified reviewers independently assess disputed or potentially primary-endpoint-positive cases according to the adjudication manual; disagreements are resolved by a senior adjudicator or consensus process. The adjudication database records finding category, confidence, supporting source information and whether the case is evaluable for the primary endpoint.

10 Outcome measures

10.1 Primary endpoint

Title	Patient-Level Clinically Important Abnormality Miss Rate on the Index Chest Radiograph
Time Frame	Index interpretation through completion of blinded endpoint adjudication, using clinically available information from the index encounter and prespecified follow-up window; target confirmatory window within 30 days when available.
Event	At least one prespecified Category A or B abnormality judged present by the reference standard but absent from the initial finalized chest radiograph report produced under the assigned strategy.

Primary Denominator	All randomized participants with an evaluable index examination, analyzed according to assigned group. Missing reference-standard status is handled under the SAP rather than coded automatically as negative.
Measure	Adjusted and unadjusted group-specific risks, relative effect and absolute risk difference with 95% confidence intervals.

10.2 Key secondary endpoints

- Patient-level Category A critical finding miss rate.
- Sensitivity and patient-level detection rate for prespecified rare/easily missed findings, overall and by finding class.
- Time from image availability to first documented recognition of a critical abnormality.
- Time from image availability to initial finalized radiology report.
- Time to effective critical-result communication to the responsible clinical team where captured.
- Patient-level false-positive clinically important finding rate.
- Radiology recall, repeat radiography, CT, ultrasound or other additional diagnostic imaging after the index report.
- In the intervention arm: AI suggestion acceptance, rejection, physician edits and high-risk prompts that change the final report.
- AI availability, processing failure, latency and proportion of eligible examinations successfully processed.
- Exploratory short-term clinical outcomes: ICU transfer, hospital length of stay and in-hospital mortality.

11 Sample size

The operational target is 70,000 randomized participants (approximately 35,000 per arm). The planning assumption is a patient-level clinically important abnormality miss rate of 1.5% under usual care and 1.2% under EvoXRare-assisted interpretation, corresponding to an absolute reduction of 0.3 percentage points and a relative reduction of 20%. With two-sided alpha 0.05, 90% power and 1:1 allocation, approximately 62,000 evaluable participants are required under a simple two-sample proportion framework. Allowing approximately 10% for non-evaluable reference-standard status, major data incompleteness or other loss of primary endpoint ascertainment yields approximately 68,900; the operational target is rounded to 70,000.

These event rates are planning assumptions, not trial results. A blinded pretrial or early-recruitment audit may estimate overall abnormality prevalence, reference-standard completeness and aggregate miss-event frequency across several planned centers. Any increase toward approximately 80,000 participants should occur without unblinded comparison of randomized treatment effects and must be documented by protocol amendment before the change is implemented.

12 Data collection and trial workflow

12.1 Required baseline/index data

- Study identifier, center, care setting, age, sex and clinically relevant baseline variables available in routine records.
- Index radiograph acquisition time, image availability time, portable/fixed acquisition status and relevant technical metadata.
- Interpreting physician identifier or coded reader ID, reader experience category when available, report start/submission timestamps and initial finalized report.
- Randomized assignment, AI report availability/view time for the intervention arm, AI output and physician edits/override information when feasible.
- Critical-result recognition and effective notification timestamps when applicable.

12.2 Reference-standard data

Sites provide clinically obtained follow-up information needed for adjudication. The study does not require additional imaging solely to establish a reference standard. Missing or unavailable confirmatory information is recorded transparently. Source information is retained sufficiently to permit central review and audit.

12.3 Data transfer and confidentiality

Before central transfer, direct identifiers are removed and DICOM headers or embedded image text are checked according to the data-management manual. The linkage between local identity and trial identifier remains at the participating site. Central data storage uses role-based access, access logs and approved secure transfer mechanisms.

13 Statistical analysis overview

The Statistical Analysis Plan is finalized before unblinding of treatment-group results and contains the prespecified analysis populations, endpoint derivations, covariate coding, missing-data methods, sensitivity analyses and subgroup definitions. The primary analysis follows the intention-to-treat principle.

The primary binary endpoint will be compared using a regression model that accounts for center and prespecified baseline/operational factors. Because participants are interpreted by multiple physicians within centers, reader clustering and cross-classified case/reader structure will be addressed in prespecified sensitivity analyses when reader identifiers are available. Both relative and absolute treatment effects with 95% confidence intervals will be reported.

Key secondary outcomes will be analyzed according to endpoint type. Time outcomes will use prespecified survival or quantile/regression approaches depending on the observed distribution and censoring structure. Rare-finding analyses emphasize effect estimates and confidence intervals; finding-level analyses account for within-participant correlation. No favorable model or subgroup will be selected after inspecting P values.

14 Safety, protocol deviations and discontinuation

The intervention is diagnostic decision support. Potential harms include false reassurance, false-positive prompts, inappropriate report modification, delayed communication caused by technical workflow, or downstream testing prompted by erroneous output. Physicians retain responsibility for interpretation and may disregard AI suggestions. AI processing must not delay urgent care.

Major deviations include randomization without meeting core eligibility criteria, control-arm exposure to trial AI before report finalization, unavailable assigned AI due to preventable workflow nonadherence, use of an unapproved model version, and material failure to obtain required endpoint data. Deviations are summarized by group but participants remain in the intention-to-treat analysis unless randomization is proven invalid under a prespecified rule.

A participant may withdraw permission for future research data collection as allowed by local regulations. Data already collected before withdrawal are handled according to consent, ethics approval and applicable law. A site may be suspended for safety, persistent protocol noncompliance, data-integrity concerns or failure to maintain required approvals.

15 Quality assurance and monitoring

- Standardized site initiation and training before first enrollment.
- Locked definitions for image availability, report start/finalization, critical finding recognition and notification.
- Central review of screening, randomization balance, missing endpoint data, protocol deviations and AI technical failure.
- Ongoing adjudicator calibration and periodic blinded inter-reviewer agreement checks.
- Source-data verification or targeted audit for a prespecified sample and all serious suspected AI-related harms.
- Version control for protocol, SAP, adjudication manual, eCRF, software and analysis code.

16 Data sharing, publication and reproducibility

The planned IPD-sharing approach is controlled access. After the primary publication and subject to compatible consent, ethics approval and institutional authorization, qualified investigators may request de-identified participant-level analysis data, the data dictionary, protocol, SAP and analysis code. Raw identifying records and unrestricted source imaging are not publicly released.

Primary reporting will follow contemporary randomized-trial reporting standards and relevant AI extensions. The main publication will describe trial flow, center participation, protocol deviations, reference-standard coverage, primary and key secondary outcomes, technical failures and clinically important harms. Prespecified subgroup and technical analyses are labeled clearly as confirmatory or exploratory according to the SAP.

17 Operational schedule

Preparation / activation	October–December 2026: finalize protocol and SAP, prospective registration, ethics review, site agreements, locked AI validation, workstation deployment, eCRF and adjudication manual.
Optional silent workflow run-in	Selected centers may perform a short nonrandomized run-in to verify image transfer, latency, timestamp capture and data completeness. These cases do not enter the randomized primary analysis.
Recruitment	Anticipated November 2026–December 2027; approximately 70,000 unique randomized participants across 15–20 centers.
Endpoint adjudication	Ongoing, blinded, throughout recruitment; final adjudication after the last participant's prespecified reference-standard window.
Database lock / primary analysis	Target first half of 2028 after reconciliation, adjudication and SAP-controlled database lock.

18 References and methodological framework

1. Liu X, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26:1364-1374.
2. Cruz Rivera S, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26:1351-1363.
3. SPIRIT 2025 statement. BMJ. 2025;389:e081477.
4. CONSORT 2025 statement. BMJ. 2025;389:e081123.