FAIR Trial- A pragmatic randomized trial of AI-assisted fracture detection in emergency department radiography



Principal Investigator
Dr. Martin BREITWIESER, PhD MBA, BSc University Hospital Salzburg Department of Orthopedic Surgery and Traumatology
Study Site
University Hospital Salzburg Müllner Hauptstr. 48 5020 Salzburg, Austria

Co-Investigators	Co-Investigator
a.o. Univ. Prof. Dr. Thomas FREUDE University Hospital Salzburg Department of Orthopedic Surgery and Traumatology	Dr. Matthias SCHAFFERT, MSc University Hospital Salzburg Department of Pediatric and Adolescent Surgery
Co-Investigator	**Biostatistician**
a.o. Univ. Prof. Dr. Klaus HERGAN University Hospital Salzburg Department of Radiology	Konstantin Emil Thiel, MSc Paracelsus Medical University Salzburg Core Facility Biostatistics

Last Update: 05.05.2026

Content

1 Wider research context 4
2 Objectives 5
2.1 Primary Outcome 6
2.2 Key Secondary and Supportive Implementation Outcomes 6
2.3 Secondary Clinical Outcomes 6
2.4 Covariates, Subgroup and Sensitivity Analyses 7
2.5 Power Considerations 7
3 Methods 7
3.1 Research Methodology and Type of Study 8
3.2 Patient Recruitment and Eligibility 8
3.3 Inclusion and Exclusion Criteria 8
3.4 Randomization and Allocation 8
3.5 Intervention: AI-Assisted Fracture Detection 9
3.6 Data Collection 9
3.7 Ground Truth Establishment 9
3.8 Statistical Analyses 10
3.9 Methods to Minimize Bias 11
4 Level of originality / innovation 11
5 Researchers and scientific qualifications 12
5.1 Principal Investigator and Study Coordination 13
5.2 Lead Investigators at Participating Study Sites 13
5.3 Radiology Departments 13
5.4 Biostatistics and Research Support 14
5.5 Independence and AI Software Implementation 14
6 Work and Timeline Planning 14
6.1 Completed Phase: Trial Preparation and Data Collection 14
6.2 Analysis Phase: Data Entry and Curation 15
6.3 Analysis Phase: Expert Adjudication 15
6.4 Analysis Phase: Statistical Analysis 15

6.5 Manuscript Preparation and Dissemination 15
6.6 Summary Timeline 15
7 Ethical, Safety, and Regulatory Considerations 16
7.1 Ethical Approval and Oversight 16
7.2 Informed Consent and Patient Rights 16
7.3 Patient Safety and Risk Management 16
7.4 Data Protection and Confidentiality 17
7.5 Regulatory Compliance and AI Transparency 17
8 Gender and Diversity Aspects in Research 17
Annex 1: References 19
Annex 2: Financial Aspects ..... **Error! Bookmark not defined.**
Annex 3: Clinical Trial Synopsis 21
Annex 4: Scientific CVs ..... **Error! Bookmark not defined.**
Annex 5: Statement on Environmental Sustainability ..... **Error! Bookmark not defined.**

1 Wider research context

Skeletal fractures are among the most common emergency department (ED) presentations, and accurate detection at the index visit is essential to emergency care. Across European EDs, missed fracture rates of 1–3.1% of all presenting fractures have been reported, with most errors occurring between 20:00 and 02:00, when specialist radiology support is least available [1,2]. In the United Kingdom, missed fractures are the most frequent diagnostic error attributed to ED clinicians, with initial radiograph miss rates of 3.3–8.7% among frontline staff. Radiograph misinterpretation is also the most common cause of successful negligence claims arising from emergency care, with harms including death, disability, and chronic pain [3,4].

This problem is not limited to one healthcare system. A European Society of Emergency Radiology survey found that approximately one in five EDs lacks 24/7 staff radiologist coverage, and emergency radiology remains unevenly developed across continental Europe, including the DACH region [5]. Most acute fractures are initially assessed by clinicians with limited formal imaging training, while high patient volume, time pressure, and variable reader experience create conditions prone to perceptual error [1,2,6].

Over the past decade, retrospective studies have shown that deep-learning algorithms can detect fractures on plain radiographs with pooled sensitivities and specificities of approximately 89–92% and AUC values frequently exceeding 0.95 [7-9]. Husarek et al. (2024), including 38,978 radiographs across commercial AI fracture detection products, confirmed that diagnostic accuracy is highest when AI is used with human readers rather than autonomously [7]. However, this evidence base remains limited by enriched datasets with fracture prevalences above routine ED practice, heterogeneous reference standards, evaluation outside clinical workflow, lack of clinician access to AI output at the time of care, and absence of patient-relevant outcomes [7,8,10]. False positives may also increase unnecessary imaging or consultations, while overreliance on AI raises concerns about automation bias [6,8,11].

Recent prospective evidence highlights these limitations. Luiken et al. (2025) evaluated three commercial CE-marked AI algorithms in 1,037 consecutive adult patients across 22 anatomical regions and found that no algorithm exceeded 91% accuracy for acute fractures; performance declined with osteosynthesis hardware, and accuracy for multiple fractures was poor (AUC 64–73%) [12]. The authors concluded that none of the systems was suitable for standalone clinical use and called for prospective multicenter trials integrating AI directly into clinical workflows [12]. A German single-center prospective PACS workflow study found improved resident sensitivity and specificity with AI support, but likewise emphasized the need for prospective controlled evidence under routine conditions [13].

Multi-reader multi-case (MRMC) evidence provides more clinically informative data than standalone algorithm testing. Bachmann et al. (2024) showed in a fully crossed MRMC study of 15 non-specialist ED-affiliated readers that AI assistance reduced missed fractures by 29% and false positives by 21% without increasing reading time [14]. Gains were greatest for non-obvious fractures, hip fractures, and less experienced readers, the frontline group most relevant to ED practice in Austria, Germany, and comparable

European systems [11,14]. A multicenter retrospective study protocol further extends the implementation perspective by examining whether AI-supported diagnostic radiographers from three countries could safely discharge patients without acute skeletal injury, using a 3.1% missed-fracture safety benchmark [15].

Several prospective implementation studies are now underway in the United Kingdom. FRACT-AI (ISRCTN19562541) evaluates AI-assisted fracture detection across six NHS reader groups and four hospitals, including automation bias [3]. The NIHR-funded FRACTURE study (ISRCTN12921105) evaluates AI in 500 pediatric radiographs with multidisciplinary readers and simulated management decisions [4]. The SAMURAI-Fracture trial (ISRCTN23087950) is a prospective cluster-randomized crossover trial across four NHS trusts, measuring real-world endpoints including unnecessary reattendances, additional imaging, and inappropriate immobilization under alternating AI-on and AI-off conditions [16].

At the health-technology assessment level, NICE published an Early Value Assessment on AI technologies for fracture detection on X-rays in urgent care (HTG739, 2025), conditionally recommending four CE-marked systems for NHS use during an evidence generation period, while explicitly identifying prospective real-world evidence on diagnostic accuracy, clinical outcomes, system-level impacts, and implementation costs as essential requirements before routine adoption can be confirmed [17].

Collectively, these studies and assessments demonstrate strong international momentum toward real-world evaluation of AI fracture detection. However, the currently most advanced prospective implementation evidence is still either ongoing, unpublished, simulated, or situated within a single healthcare system. This leaves a clear gap for completed prospective randomized implementation evidence from continental European emergency care systems, particularly from Austrian and German multicenter clinical workflows.

The FAIR Trial addresses this gap by evaluating, in a prospective, multicenter, pragmatic randomized controlled design across three sites in two countries, whether AI-assisted fracture detection delivered to frontline emergency physicians as decision support affects ED throughput, clinician diagnostic confidence, additional imaging during the index visit, and secondary clinical outcomes including missed fractures and diagnostic accuracy.

2 Objectives

The FAIR (Fracture detection with AI in emergency Radiography) Trial addresses whether AI-assisted fracture detection, deployed as decision support for frontline emergency physicians under real-world conditions, improves clinically relevant implementation outcomes in routine emergency care. Rather than focusing primarily on retrospective diagnostic accuracy, the trial evaluates whether AI support affects ED workflow, clinician diagnostic confidence, imaging behavior, duration of care, and clinically relevant diagnostic outcomes under pragmatic conditions.

This represents a deliberate shift away from the diagnostic-accuracy paradigm that has dominated AI fracture detection research. Diagnostic accuracy remains scientifically important and will be assessed as a

secondary clinical outcome, but the central question is whether AI-assisted fracture detection produces measurable value when integrated into real ED workflows.

2.1 Primary Outcome

Time from triage to discharge is the primary outcome of the FAIR Trial. Triage time is defined as the documented time of formal ED triage, marking assignment to the emergency care pathway and avoiding bias from variable pre-triage waiting times. Discharge time is defined as the documented completion of ED treatment for the index presentation, reflecting the end of active emergency care and clinical decision-making rather than physical departure from the ED. Logistical delays after treatment completion, such as transfer to the operating room or ward or waiting for relatives or ambulance transport, are not included.

Time from triage to discharge was selected as the central implementation outcome because it captures the effect of AI integration on ED throughput under real-world conditions. AI support may influence throughput by reducing diagnostic uncertainty, changing additional imaging behavior, or altering escalation to more experienced colleagues. This endpoint therefore provides a measurable basis for assessing workflow, resource use, and imaging behavior, and may inform future health-economic analyses.

2.2 Key Secondary and Supportive Implementation Outcomes

The FAIR Trial evaluates two key secondary implementation outcomes: clinician diagnostic confidence and additional imaging requested during the index visit.

Clinician diagnostic confidence is recorded on a 0–10 ordinal scale using the standardized study form. Confidence is documented separately for the initial treating physician and, where consultation occurs, for the senior physician. The main analysis will focus on the initial treating physician, because this rating is documented before senior escalation and best reflects the direct effect of AI support on frontline decision-making. Senior-physician confidence will be analyzed separately in escalated cases.

Additional imaging during the index visit includes additional radiographs, CT, MRI, and sonography requested during the ED visit. This binary case-level outcome will also be described by imaging modality.

2.3 Secondary Clinical Outcomes

Secondary clinical outcomes assess whether implementation effects are accompanied by clinically meaningful diagnostic effects. These outcomes are important for interpretation but are no longer the primary study driver.

Missed fractures are defined as fractures not diagnosed by the treating physician at the index ED assessment but subsequently identified through expert adjudication combined with limited retrospective chart review where required. Missed fractures will be analyzed as a binary outcome related to diagnostic accuracy.

Diagnostic accuracy will be assessed as supportive secondary evidence against the expert-adjudicated reference standard. Sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy will be reported descriptively to contextualize the implementation findings.

2.4 Covariates, Subgroup and Sensitivity Analyses

Main analyses will adjust for physician experience, anatomical region, and study site, as these factors may influence workflow, confidence, imaging behavior, and diagnostic performance independently of group allocation. Physician experience will be grouped as less than 1 year, 1–6 years, and more than 6 years of specialist training. Anatomical region will be categorized as thorax/ribs, spine, upper extremity, lower extremity, and pelvis. Study site will distinguish University Hospital Salzburg, regional hospital Hallein, and University Hospital Nuremberg.

Exploratory subgroup or interaction analyses may assess whether AI effects differ by physician experience, anatomical region, or study site. These analyses will be interpreted as hypothesis-generating. Sensitivity analyses will address missing data and model assumptions. Time from triage to diagnosis, documented only in a subset of cases, will be analyzed as a predefined supportive time endpoint where available, because it may more directly reflect diagnostic decision-making before treatment steps such as immobilization, reduction, or casting.

2.5 Power Considerations

The sample size is fixed by the accrued prospective dataset. Across all sites, approximately 3,100 cases were enrolled, of which approximately 2,800 are expected to constitute the eligible analysis dataset after application of predefined inclusion and exclusion criteria.

The original trial planning assumed a larger recruitment yield; however, the completed dataset remains substantial for a pragmatic randomized clinical implementation study in emergency radiography. The revised endpoint structure centers on time from triage to discharge as the primary endpoint, with clinician diagnostic confidence and additional imaging as key secondary implementation outcomes. This reflects the scientific aim of the resubmission: to determine whether AI-assisted fracture detection produces measurable value in real ED workflows rather than only measuring algorithmic performance.

As a transparent benchmark, assuming two groups of equal size (n = 1,400 per group), a standard deviation of 30 minutes, and a true mean difference of at least 3.5 minutes, the power to detect this difference at a two-sided significance level of 0.05 is approximately 87% using a two-sample t-test in R (version 4.2.2). The primary analysis will use regression-based models adjusting for pre-specified covariates and repeated observations at physician level where appropriate, as described in Section 3.8.

3 Methods

The FAIR Trial is a prospective, multicenter, pragmatic randomized controlled trial with a completed prospective dataset ready for scientific evaluation.

3.1 Research Methodology and Type of Study

The FAIR Trial evaluates the real-world clinical impact of AI-assisted fracture detection on emergency department workflow, frontline physician decision-making, diagnostic confidence, imaging behavior, and secondary clinical outcomes. The study follows CONSORT-AI reporting principles for randomized trials involving AI interventions.

The trial was conducted at three sites: University Hospital Salzburg, Austria; regional hospital Hallein, Austria; and University Hospital Nuremberg, Germany. All sites used an identical protocol, workflow structure, eligibility logic, data collection instrument, and randomization procedure. Final prospective data collection was completed on April 30, 2026.

3.2 Patient Recruitment and Eligibility

Eligible patients were identified at triage by trained nurses using a structured screening procedure applied consistently across all sites. After eligibility confirmation, treating physicians obtained written informed consent before imaging. Pediatric patients underwent age-appropriate consent and assent procedures. Patients declining consent received standard care without study-related procedures or data collection.

Across all three sites, approximately 3,100 cases were enrolled, of which approximately 2,800 are expected to constitute the eligible analysis dataset. The unit of analysis is the case or encounter rather than the individual patient, because a new injury presentation during the study period constitutes a new eligible case.

3.3 Inclusion and Exclusion Criteria

Patients were eligible if they were aged 2 years or older, presented after trauma with suspected fracture requiring plain radiography, and provided informed consent. Multiple fractures within the same anatomical injury location were included. No exclusions were applied on the basis of image quality. Low-quality radiographs were retained because they reflect routine emergency radiography and excluding them would introduce spectrum bias.

Patients with injuries involving multiple body regions were excluded because parallel imaging pathways and complex clinical trajectories complicate attribution of diagnostic and workflow outcomes. Head and cervical spine injuries were excluded because they fall outside the approved indication scope of the AI algorithm. Further exclusions were age under 2 years, prior imaging or medical consultation for the same injury, contraindications to X-ray imaging such as pregnancy, and lack of informed consent.

3.4 Randomization and Allocation

Patients were randomized 1:1 to the AI-assisted group or control group. The randomization sequence was generated by the principal investigator in R (version 4.2.2) using sample() and set.seed(611), ensuring reproducibility. Allocation was concealed from treating physicians until patient assignment and was independent of age, anatomical region, injury characteristics, physician seniority, and study site.

Intervention patients were assessed in a dedicated room with simultaneous access to radiographs and BoneView™ output. Control patients were assessed in a separate room without visible AI output. Physicians were instructed not to share AI findings across groups. Residual learning effects cannot be fully excluded in an open pragmatic trial, but physical and procedural separation minimized contamination risk.

3.5 Intervention: AI-Assisted Fracture Detection

The AI algorithm used was BoneView™ version 1.0.1, developed by Gleamer (Paris, France). BoneView™ is a CE-marked Class IIa medical device (Notified Body: BSI, n°2797), approved for patients aged 2 years and older, and intended as a decision-support tool rather than a sole basis for clinical decisions.

At Salzburg and Hallein, BoneView™ was delivered via the Aidoc aiOS platform (version aios.3.24.0), already installed at these sites and reactivated for the trial. At Nuremberg, BoneView™ was implemented directly. The software analyzes DICOM radiographs transmitted from PACS and returns fracture probability classification and annotated overlay images simultaneously with the radiographic study.

In the intervention group, physicians reviewed radiographs together with AI output before documenting their diagnosis. In the control group, radiographs were interpreted without AI support. In both groups, all final clinical decisions remained with the treating physician.

3.6 Data Collection

Each patient was documented using a standardized paper-based trial form applied uniformly across all sites. The form captured demographics, injury characteristics, physician diagnosis, diagnostic confidence on a 0–10 scale, time of diagnosis where documented, additional imaging, and escalation to senior physicians. In the AI-assisted group, AI output classification and any documented influence on the final diagnosis were also recorded.

Triage and discharge times are extracted from the electronic health record. Time of diagnosis was entered manually by treating physicians and is available only in a subset; it is therefore used as a predefined supportive time endpoint. Additional imaging during the index visit was recorded by modality, including additional radiographs, CT, MRI, and sonography.

3.7 Ground Truth Establishment

The reference standard will be established through expert adjudication independently of the treating physician’s assessment. All cases are reviewed independently by one senior radiologist and one senior orthopedic trauma surgeon. Agreement constitutes the reference standard; disagreements are resolved by discussion, with a third independent expert reviewer making the final determination if consensus cannot be reached.

For missed-fracture analysis, the orthopedic trauma reviewer additionally performs limited retrospective chart review where required to determine whether a fracture not diagnosed at the index visit was recognized later. Prof. Klaus Hergan, Department of Radiology, University Hospital Salzburg, serves as Lead

of Imaging Adjudication and Ground-Truth Review. Exploratory analyses of discordant AI and clinician cases may be performed where supported by the final dataset.

3.8 Statistical Analyses

All analyses will be conducted using the eligible randomized analysis dataset and will follow the allocated study group (AI-assisted vs. control). All primary and secondary outcomes as well as pre-specified covariates will be summarized descriptively by study group using appropriate summary measures according to variable type (e.g., mean, median, standard deviation, and interquartile range for continuous variables; counts and percentages for categorical variables). All effect estimates will be reported with 95% confidence intervals and two-sided p-values.

The primary endpoint, time from triage to discharge, is a positive continuous variable that may show right-skewness. The primary comparison between study groups will be performed using a mixed-effects linear regression model. If appropriate based on the observed distribution, log-transformation of the endpoint will be applied prior to modeling. The main fixed effect of interest will be study group. To account for relevant clinical and organizational heterogeneity, the model will additionally include physician experience category, anatomical region, and study site as covariates. Since individual physicians contribute repeated observations, a random intercept for physician will be included to account for within-physician clustering. Because the number of study sites is small, study site will be modeled as a fixed effect rather than as a random effect. Results will be presented as adjusted between-group effect estimates with 95% confidence intervals. If log-transformation was applied prior to modeling, back-transformed estimates will be reported on the original time scale to facilitate interpretation. If model diagnostics indicate that model specification is inadequate, an alternative robust model for positive and potentially skewed outcomes will be used as a sensitivity analysis.

The secondary outcome of clinician diagnostic confidence, recorded on a 0-10 ordinal scale, will be analyzed using a proportional-odds cumulative logit model with a random intercept to model repeated measures per physician. The model will include study group as the primary explanatory variable and will adjust for physician experience category, anatomical region, and study site. If the proportional-odds assumption is clearly violated, a less restrictive ordinal model or an appropriate sensitivity analysis will be applied. The main analysis will focus on the confidence rating of the initial treating physician. Confidence ratings of senior physicians will be analyzed separately and descriptively within the subgroup of cases involving escalation to the senior physician.

The secondary outcome of additional imaging during the index visit will be analyzed primarily as a binary case-level outcome (any additional imaging: yes/no). The main analysis will use a logistic regression model, again including study group as the primary explanatory variable and adjusting for physician experience, anatomical region, and study site; where appropriate, physician-level clustering will be accounted for by a random intercept. Adjusted odds ratios with 95% confidence intervals will be reported. Additional imaging

modalities (additional radiographs, CT, MRI, sonography) will be summarized descriptively by study group; exploratory modality-specific analyses may be performed where event numbers permit stable estimation.

The secondary clinical outcome of missed fractures will also be analyzed as a binary outcome using an adjusted logistic regression model with the same covariate structure.

Diagnostic accuracy against the expert-adjudicated reference standard will be reported as secondary supportive evidence. Sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy will be estimated separately by study group with corresponding 95% confidence intervals. Where clinically meaningful and statistically feasible, between-group comparisons for binary accuracy components (e.g., sensitivity, specificity, false-negative rate, false-positive rate) may be supplemented by simple 2×2 analyses, including unconditional exact methods in sparse settings. These analyses are considered supportive and will primarily serve to contextualize the implementation-oriented findings.

Because the study specifies one primary endpoint, no multiplicity adjustment is planned for the primary confirmatory analysis. Secondary and exploratory analyses will be interpreted as supportive unless additional confirmatory hypotheses are explicitly defined for a subsequent manuscript. In that case, an appropriate multiplicity adjustment method (for example, the Bonferroni-Holm procedure) may be applied to the family of tests in order to preserve an appropriate overall type I error rate.

All analyses will be performed in R (version 4.2.2).

3.9 Methods to Minimize Bias

Allocation concealment was maintained until patient assignment, minimizing selection bias. Randomization was independent of patient age, anatomical region, injury characteristics, physician seniority, and study site.

Control-group physicians had no access to AI output. Expert adjudicators will be blinded to group assignment, with AI overlays and metadata removed before review. Physical and procedural separation of AI-assisted and control workflows reduced contamination risk. Inclusion of low-quality radiographs and multiple same-location fractures preserves the pragmatic routine-care case mix. Pre-specified subgroup analyses by physician experience, anatomical region, and study site will assess robustness across clinically relevant strata.

4 Level of originality / innovation

AI-assisted fracture detection tools are commercially available and increasingly deployed across European emergency departments, including in Austria and Germany. Their technical performance has been described in retrospective reader studies and prospective real-world evaluations. Most recently, Luiken et al. prospectively evaluated three CE-marked commercial algorithms, including BoneView™, on consecutive adult patients across multiple anatomical regions and concluded that none was robust enough for

standalone clinical use; integration of AI outputs into clinical workflows at the time of care delivery was identified as the next critical step [12].

This is the gap addressed by the FAIR Trial. Existing evidence, including ongoing NHS studies and European registry data, remains strongly focused on technical performance, reader sensitivity, specificity, AUC, and diagnostic accuracy. These measures are important, but they do not answer the implementation question facing clinicians, hospitals, and healthcare systems: whether AI-assisted fracture detection changes emergency care delivery.

The central innovation of the FAIR Trial is its prospective randomized evaluation of AI-assisted fracture detection beyond diagnostic accuracy, focusing on clinician-centered and system-centered outcomes under real-world conditions. The primary outcome, time from triage to discharge, directly addresses ED throughput. Clinician diagnostic confidence and additional imaging requests capture effects on physician behavior and resource utilization at the point of care. Missed fractures and diagnostic accuracy remain secondary clinical outcomes, ensuring that implementation effects are interpreted alongside clinically consequential diagnostic performance.

This shift is relevant beyond clinical science. Hospitals and healthcare systems cannot base sustainable procurement and deployment decisions solely on accuracy metrics from controlled datasets. They require evidence on whether AI improves workflow, supports clinicians, influences imaging behavior, and maintains clinical safety in routine care. The FAIR Trial therefore generates evidence relevant to clinicians, researchers, hospital administrators, payers, and health-technology assessment bodies.

Additional originality derives from the pragmatic randomized design, patient-level allocation, physical separation of AI-assisted and control workflows, inclusion of a real clinical case mix, and multicenter structure across three sites in two countries. This design avoids the artificiality of enriched reader studies while preserving internal validity and generating evidence directly applicable to Central European emergency care systems. The expert adjudication process, led by Prof. Klaus Hergan and supported by limited retrospective traumatological chart review where required, provides a robust basis for evaluating both implementation outcomes and clinically relevant secondary endpoints.

5 Researchers and scientific qualifications

The FAIR Trial is conducted by a multidisciplinary team in orthopedic trauma surgery, pediatric surgery, emergency medicine, diagnostic radiology, and biostatistics across three sites in two countries. More than 60 emergency physicians, orthopedic surgeons, and pediatric surgeons contributed to data collection, ensuring that the trial reflects frontline emergency care rather than a controlled research setting.

Since the original submission, the project has been embedded in the SMART (Salzburger Medical AI Research in Traumatology) group, led by Dr. Martin Breitwieser and officially registered at Paracelsus Medical University Salzburg. SMART provides the institutional framework for AI-assisted traumatology research at University Hospital Salzburg and the scientific home for the FAIR Trial and related projects.

5.1 Principal Investigator and Study Coordination

Dr. Martin Breitwieser, PhD, MBA, BSc, is a senior attending physician in the Department of Orthopedics and Traumatology at University Hospital Salzburg, head of the SMART research group, and principal investigator of the FAIR Trial. He designed the study, secured the institutional bridge funding that enabled prospective data collection, and served as lead investigator for the orthopedic and trauma surgery component at Salzburg. He also coordinated trial operations across all three sites, including protocol training, randomization management, form logistics, physician instruction, and inter-site communication, alongside full clinical duties. This ensured protocol consistency despite the absence of dedicated research support during recruitment.

Dr. Breitwieser has published work in natural language processing applied to clinical data and AI-based decision support in emergency settings. His combined expertise in orthopedic trauma surgery and applied AI research directly matches the clinical and scientific scope of the FAIR Trial. He will lead the analysis phase, coordinate expert adjudication, and take primary responsibility for manuscript preparation and dissemination.

5.2 Lead Investigators at Participating Study Sites

Each site had a dedicated lead investigator responsible for departmental oversight. At University Hospital Salzburg, Prof. Dr. Thomas Freude, Head of the Department of Orthopedics and Traumatology, serves as institutional sponsor in accordance with ethics approval EK Nr. 1135/2024, ensuring institutional accountability and regulatory compliance. Dr. Martin Breitwieser led the orthopedic and trauma surgery component, while Dr. Matthias Schaffert led the pediatric surgery component and was responsible for protocol implementation, staff coordination, and data collection in pediatric surgery.

At regional hospital Hallein, Dr. Sebastian Filipp oversaw protocol implementation, staff training, and data collection after ethics approval was expanded to include Hallein in September 2025. At University Hospital Nuremberg, Dr. Thomas Reuter led the orthopedic and trauma surgery component, and Dr. Julia Scholsching led the pediatric surgery component. Both ensured protocol adherence at the Nuremberg site.

5.3 Radiology Departments

Prof. Dr. Klaus Hergan, Head of the Department of Radiology at University Hospital Salzburg, serves as Lead of Imaging Adjudication and Ground-Truth Review. He provides senior radiological oversight, supervises independent radiological review, co-leads the consensus procedure, and ensures radiological quality assurance across the dataset. His involvement directly addresses the need for expert radiological leadership in a trial whose clinical secondary outcomes depend on a reliable adjudicated reference standard.

At University Hospital Nuremberg, Prof. Dr. Michael Lell, Head of Radiology, supported data collection. Ground-truth adjudication for the full dataset will be conducted centrally at University Hospital Salzburg under Prof. Hergan’s leadership.

5.4 Biostatistics and Research Support

Konstantin Emil Thiel, MSc, Core Facility Biostatistics at Paracelsus Medical University Salzburg, serves as trial biostatistician. He will oversee statistical analyses, contribute to interpretation, and support reporting of the primary and secondary outcomes, reflecting the complexity of the multicenter randomized dataset.

The analysis phase will be supported by Heike Aigner as research assistant (50% position, 6 months), funded through this grant. Her tasks include data entry and curation, database management, adjudication logistics, and coordination between the research team and biostatistician. This is the first dedicated research support position in the FAIR Trial; prospective data collection was completed without funded research assistance.

5.5 Independence and AI Software Implementation

The study team worked with Gleamer and Aidoc only for technical implementation of BoneView™ before trial commencement: via Aidoc aiOS at Salzburg and Hallein, and directly via Gleamer at Nuremberg. No financial compensation or other benefits were exchanged. The study is fully independent of both companies, with no external influence on study design, data collection, adjudication, analysis, interpretation, or reporting.

6 Work and Timeline Planning

The FAIR Trial followed a structured progression from preparation to completed prospective data collection.

6.1 Completed Phase: Trial Preparation and Data Collection

A mock trial was conducted at University Hospital Salzburg in May 2025 to test workflow procedures, identify logistical issues, and refine data collection before launch. Salzburg site preparation followed in September 2025, including structured training for physicians, nurses, and administrative staff on AI-assisted workflow integration, recruitment, randomization, and documentation. Data collection at Salzburg began in October 2025 and was completed on January 31, 2026.

Preparation at regional hospital Hallein and University Hospital Nuremberg took place in November 2025 using the same training and onboarding program. Data collection at both sites began in January 2026 and was completed on April 30, 2026.

All three sites used an identical protocol, workflow structure, eligibility logic, randomization procedure, and paper-based data collection form. Cross-site coordination was managed by the principal investigator alongside clinical duties, with local support from triage nurses and emergency department physicians. This phase was completed without dedicated funded research support and within the institutional bridge funding. Across all sites, approximately 3,100 cases were enrolled, of which approximately 2,800 are expected to constitute the primary analysis dataset.

6.2 Analysis Phase: Data Entry and Curation

The first two funded months will focus on data entry, cleaning, and curation. Heike Aigner, appointed as research assistant (50% position), will enter approximately 2,800 paper-based trial forms into the study database, verify completeness, cross-check entries, and flag ambiguous documentation for resolution by the principal investigator, who will supervise the process.

6.3 Analysis Phase: Expert Adjudication

Expert adjudication will run in parallel with data entry from Month 2 to Month 4. Under the leadership of Prof. Klaus Hergan, one senior radiologist and one senior orthopedic trauma surgeon will independently review all cases. Agreement constitutes the expert-adjudicated reference standard; discrepancies will be resolved by discussion or, if required, by a third independent reviewer.

Based on three minutes per case across approximately 2,800 eligible cases, radiological review requires approximately 140 hours. Including a 20% reserve for discrepant or complex cases, the requested radiological review volume is 168 hours. Orthopedic trauma review, including limited retrospective chart review for missed-fracture analysis where required, will be performed by the principal investigator and covered by the requested protected research time.

6.4 Analysis Phase: Statistical Analysis

Formal statistical analysis will begin in Month 4–5 after completion of data cleaning and adjudication. Analyses will address the primary endpoint, time from triage to discharge, the key secondary implementation outcomes of diagnostic confidence and additional imaging, and secondary clinical outcomes including missed fractures and supportive diagnostic accuracy measures. Exploratory subgroup and sensitivity analyses will be performed where appropriate.

6.5 Manuscript Preparation and Dissemination

Manuscript preparation will begin in Month 5 and continue through Month 6, with submission targeted for early 2027. The primary manuscript will report the implementation-oriented trial results, including time from triage to discharge, diagnostic confidence, additional imaging, and secondary clinical outcomes. Target journals will include international journals in emergency medicine, radiology, orthopedic trauma surgery, digital health, or implementation science. Additional analyses will be prepared subsequently where scientifically justified. Results will also be disseminated at relevant national and international congresses in 2027.

6.6 Summary Timeline

PHASE	ACTIVITY	TIMEFRAME
COMPLETED	Mock trial and site preparation	May-November 2025
COMPLETED	Prospective data collection at all sites	October 2025 – April 30, 2026
ANALYSIS PHASE	Data entry, cleaning, and curation	Month 1–2 post-funding

ANALYSIS PHASE	Expert adjudication and limited chart review	Month 2–4 post-funding
ANALYSIS PHASE	Statistical analysis	Month 4–5 post-funding
ANALYSIS PHASE	Manuscript preparation and submission	Month 5–6 post-funding
DISSEMINATION	National and international congress presentations	September - October 2027

7 Ethical, Safety, and Regulatory Considerations

The FAIR Trial is a prospective randomized controlled study involving human participants and was conducted in accordance with ethical, safety, and regulatory requirements. It introduced no additional examinations, radiation exposure, or follow-up visits beyond standard emergency department care. The only difference between study arms was the availability of AI-assisted fracture detection in the intervention group. All final clinical decisions were made by the treating physician in both arms.

7.1 Ethical Approval and Oversight

The trial received ethical approval from the Ethikkommission für das Bundesland Salzburg under approval number EK Nr. 1135/2024. Approval was subsequently expanded to include regional hospital Hallein in September 2025 under the same reference number.

The trial was conducted in compliance with the Declaration of Helsinki, Good Clinical Practice (GCP), GDPR, and applicable national data protection legislation. It is registered on ClinicalTrials.gov under NCT06754137. Registration and ethics approval were obtained under the name SMART Fracture Trial; the scientific identity has since been updated to the FAIR Trial for this resubmission, while the approved protocol, design, and conduct remain unchanged.

7.2 Informed Consent and Patient Rights

Written informed consent was obtained prospectively from all participants before imaging and data collection across all sites. Age-appropriate consent and assent procedures were used for pediatric patients: parental consent for children aged 2–6 years; parental consent plus age-appropriate child assent for children aged 6–10, 10–14, and 14–18 years; and standard adult consent for participants aged 18 years and older. Participants received written and verbal information on the study purpose, procedures, risks, and right to withdraw without consequences for clinical care. Patients declining consent received standard emergency care without study-related procedures or data collection.

7.3 Patient Safety and Risk Management

The FAIR Trial introduced no invasive procedures or additional patient risks beyond standard emergency care. No additional imaging, radiation exposure, or follow-up was required solely because of participation. Risks related to AI-associated diagnostic uncertainty, automation bias, and workflow disruption were mitigated through physician training on AI output interpretation and limitations before trial start.

If retrospective expert adjudication identifies a clinically relevant missed fracture that may require further action, the finding will be reviewed by the responsible clinical investigator and managed according to the approved safety procedure, including patient notification and follow-up where indicated. This applies regardless of study group allocation.

7.4 Data Protection and Confidentiality

All patient data are pseudonymized before entry into the study database using unique study identifiers. No personally identifiable information is linked to the research dataset. Data are stored in encrypted, access-controlled systems available only to authorized study personnel. Data retention and deletion follow legal and regulatory requirements. All data handling complies with GDPR and applicable Austrian and German data protection legislation.

7.5 Regulatory Compliance and AI Transparency

The AI algorithm used is BoneView™ version 1.0.1, developed by Gleamer (Paris, France), a CE-marked Class IIa medical device (Notified Body: BSI, n°2797) used strictly within its approved indication and age range for patients aged 2 years and older. At Salzburg and Hallein, BoneView™ was delivered via Aidoc aiOS version aios.3.24.0; at Nuremberg, it was implemented directly. BoneView™ was used solely as diagnostic decision support and did not replace physician judgment or the definitive radiological report.

Gleamer and Aidoc were involved only in technical software implementation before data collection. No financial compensation or other benefits were exchanged. The study is fully independent of both manufacturers, with no external influence on study design, data collection, adjudication, analysis, interpretation, or reporting. Findings, including AI performance and limitations, will be reported transparently in peer-reviewed open-access publications.

8 Gender and Diversity Aspects in Research

The FAIR Trial enrolled patients of all genders across a broad age range, from children aged 2 years to elderly adults, with no exclusions based on gender, ethnicity, or socioeconomic background. This pragmatic recruitment strategy reflects the diversity of patients presenting with suspected fractures in Austrian and German emergency departments.

Sex- and age-related aspects are relevant because fracture patterns, bone density, radiographic appearance, and diagnostic difficulty may differ between patient groups and may influence the performance and clinical usefulness of AI-assisted fracture detection. The study therefore allows exploratory assessment of whether implementation and diagnostic outcomes differ by sex and age group, where event numbers and data completeness permit meaningful interpretation.

AI diagnostic tools also raise questions of algorithmic equity, as performance may vary across demographic groups depending on training data composition. Although the FAIR Trial is not powered to detect small subgroup differences, its pragmatic multicenter dataset allows monitoring of relevant differences in routine clinical use. Any sex- or age-specific findings will be reported transparently as exploratory and hypothesis-generating to guide future validation work on equitable AI implementation in emergency care.

Annex 1: References

1. Hallas, P.; Ellingsen, T. Errors in fracture diagnoses in the emergency department - characteristics of patients and diurnal variation. *BMC Emerg Med* **2006**, *6*, 4, doi:10.1186/1471-227X-6-4.

2. Mattijssen-Horstink, L.; Langeraar, J.J.; Mauritz, G.J.; van der Stappen, W.; Baggelaar, M.; Tan, E. Radiologic discrepancies in diagnosis of fractures in a Dutch teaching emergency department: a retrospective analysis. *Scand J Trauma Resusc Emerg Med* **2020**, *28*, 38, doi:10.1186/s13049-020-00727-8.

3. Novak, A.; Hollowday, M.; Espinosa Morgado, A.T.; Oke, J.; Shelmerdine, S.; Woznitza, N.; Metcalfe, D.; Costa, M.L.; Wilson, S.; Kiam, J.S.; et al. Evaluating the impact of artificial intelligence-assisted image analysis on the diagnostic accuracy of front-line clinicians in detecting fractures on plain X-rays (FRACT-AI): protocol for a prospective observational study. *BMJ Open* **2024**, *14*, e086061, doi:10.1136/bmjopen-2024-086061.

4. Shelmerdine, S.C.; Pauling, C.; Allan, E.; Langan, D.; Ashworth, E.; Yung, K.W.; Barber, J.; Haque, S.; Rosewarne, D.; Woznitza, N.; et al. Artificial intelligence (AI) for paediatric fracture detection: a multireader multicase (MRMC) study protocol. *BMJ Open* **2024**, *14*, e084448, doi:10.1136/bmjopen-2024-084448.

5. Scaglione, M.; Basilico, R.; Delli Pizzi, A.; Iacobellis, F.; Dick, E.; Wirth, S.; Linsenmaier, U.; Calli, C.; Berger, F.H.; Nieboer, K.H. The practice of emergency radiology throughout Europe: a survey from the European Society of Emergency Radiology on volume, staffing, equipment, and scheduling. *European Radiology* **2021**, *31*, 2994–3001, doi:10.1007/s00330-020-07436-x.

6. Fathi, M.; Eshraghi, R.; Behzad, S.; Tavasol, A.; Bahrami, A.; Tafazolimoghadam, A.; Bhatt, V.; Ghadimi, D.; Gholamrezanezhad, A. Potential strength and weakness of artificial intelligence integration in emergency radiology: a review of diagnostic utilizations and applications in patient care optimization. *Emerg Radiol* **2024**, *31*, 887–901, doi:10.1007/s10140-024-02278-2.

7. Husarek, J.; Hess, S.; Razaeian, S.; Ruder, T.D.; Sehmisch, S.; Muller, M.; Liodakis, E. Artificial intelligence in commercial fracture detection products: a systematic review and meta-analysis of diagnostic test accuracy. *Sci Rep* **2024**, *14*, 23053, doi:10.1038/s41598-024-73058-8.

8. Duron, L.; Ducarouge, A.; Gillibert, A.; Laine, J.; Allouche, C.; Cherel, N.; Zhang, Z.; Nitche, N.; Lacave, E.; Pourchot, A.; et al. Assessment of an AI Aid in Detection of Adult Appendicular Skeletal Fractures by Emergency Physicians and Radiologists: A Multicenter Cross-sectional Diagnostic Study. *Radiology* **2021**, *300*, 120–129, doi:10.1148/radiol.2021203886.

9. Bluthgen, C.; Becker, A.S.; Vittoria de Martini, I.; Meier, A.; Martini, K.; Frauenfelder, T. Detection and localization of distal radius fractures: Deep learning system versus radiologists. *Eur J Radiol* **2020**, *126*, 108925, doi:10.1016/j.ejrad.2020.108925.

10. Herpe, G.; Nelken, H.; Vendeuvre, T.; Guenezan, J.; Giraud, C.; Mimoz, O.; Feydy, A.; Tasu, J.P.; Guillevin, R. Effectiveness of an Artificial Intelligence Software for Limb Radiographic Fracture Recognition in an Emergency Department. *J Clin Med* **2024**, *13*, 5575, doi:10.3390/jcm13185575.

11. Meetschen, M.; Salhofer, L.; Beck, N.; Kroll, L.; Ziegenfuss, C.D.; Schaarschmidt, B.M.; Forsting, M.; Mizan, S.; Umutlu, L.; Hosch, R.; et al. AI-Assisted X-ray Fracture Detection in Residency Training:

Evaluation in Pediatric and Adult Trauma Patients. *Diagnostics (Basel)* **2024**, *14*, 596, doi:10.3390/diagnostics14060596.

12. Luiken, I.; Lemke, T.; Komenda, A.; Marka, A.W.; Kim, S.H.; Graf, M.M.; Ziegelmayer, S.; Weller, D.; Mertens, C.J.; Bressem, K.K.; et al. Evaluation of commercial AI algorithms for the detection of fractures, effusions, and dislocations on real-world clinical data: A prospective registry study. *Radiography (Lond)* **2025**, *31*, 103189, doi:10.1016/j.radi.2025.103189.

13. Oppenheimer, J.; Luken, S.; Hamm, B.; Niehues, S.M. A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow. *Life (Basel)* **2023**, *13*, 223, doi:10.3390/life13010223.

14. Bachmann, R.; Gunes, G.; Hangaard, S.; Nexmann, A.; Lisouski, P.; Boesen, M.; Lundemann, M.; Baginski, S.G. Improving traumatic fracture detection on radiographs with artificial intelligence support: a multi-reader study. *BJR Open* **2024**, *6*, tzae011, doi:10.1093/bjro/tzae011.

15. Bachmann, R.; Nexmann, A.; Mudannayake, J.; Jensen, J.; Lauvsnes, S.S.; Lundemann, M. Protocol for Radiographer x AI led discharge. *medRxiv* **2025**, 2025–2010, doi:10.1101/2025.10.10.25337734

16. ISRCTN Registry. *SAMURAI-Fracture Trial: Medical utility of artificial intelligence for fracture detection in the emergency department. ISRCTN23087950. Registered January 2026. Available from: https://www.isrctn.com/ISRCTN23087950. Accessed 18 May 2026.*

17. National Institute for Health and Care Excellence (NICE). *Artificial intelligence (AI) technologies to help detect fractures on X-rays in urgent care: early value assessment. NICE HealthTech guidance 739. London: 2025. Available from: https://www.nice.org.uk/guidance/htg739. Accessed 18 May 2026.*

Annex 2: Clinical Trial Synopsis

1. TITLE OF CLINICAL TRIAL

FAIR (Fracture detection with AI in emergency Radiography) Trial — A pragmatic randomized trial of AI-assisted fracture detection in emergency department radiography

2. GRAPHICAL OVERVIEW

2.1 TIMELINE

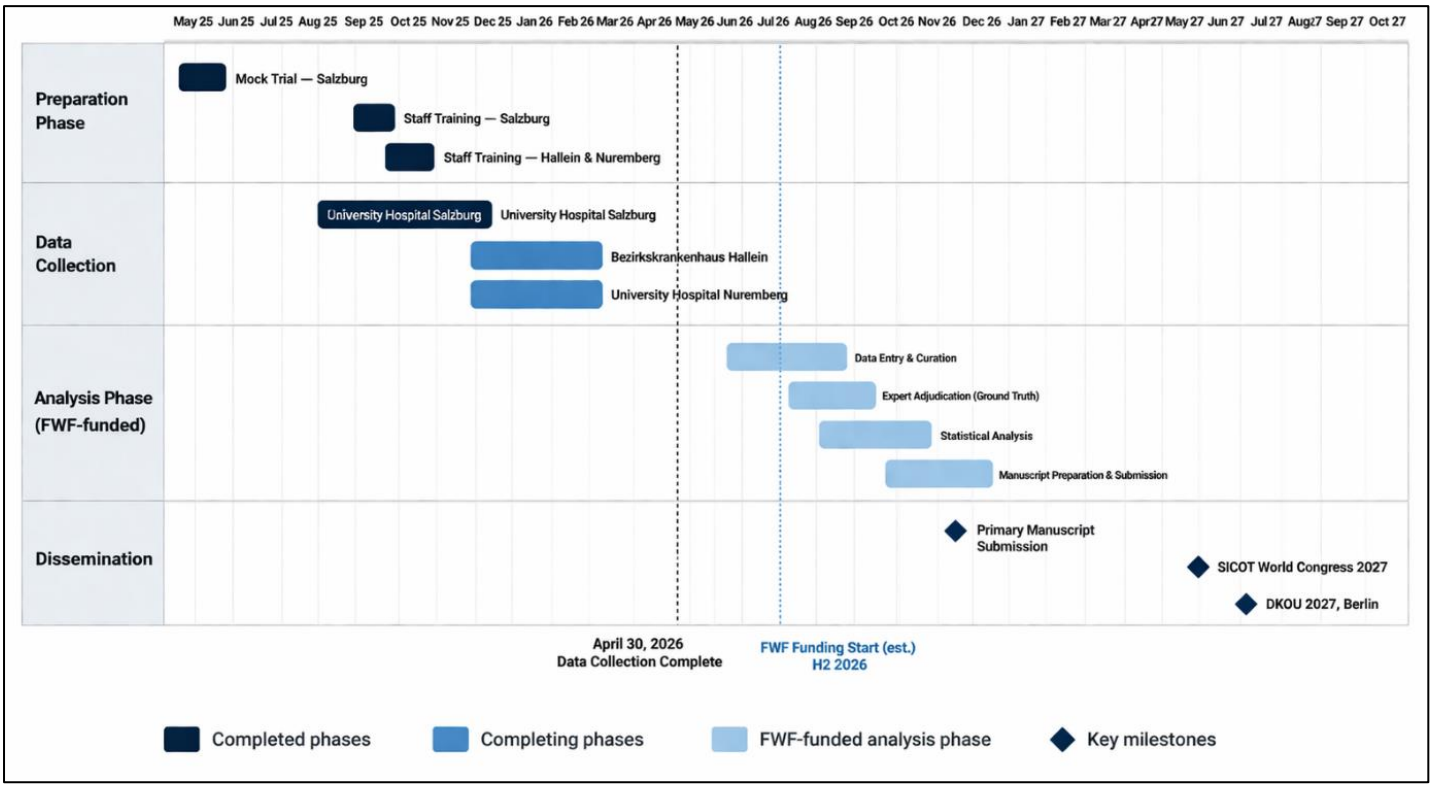


Figure 1. Estimated Timeline of the FAIR Trial

2.2 STUDY FLOWCHART

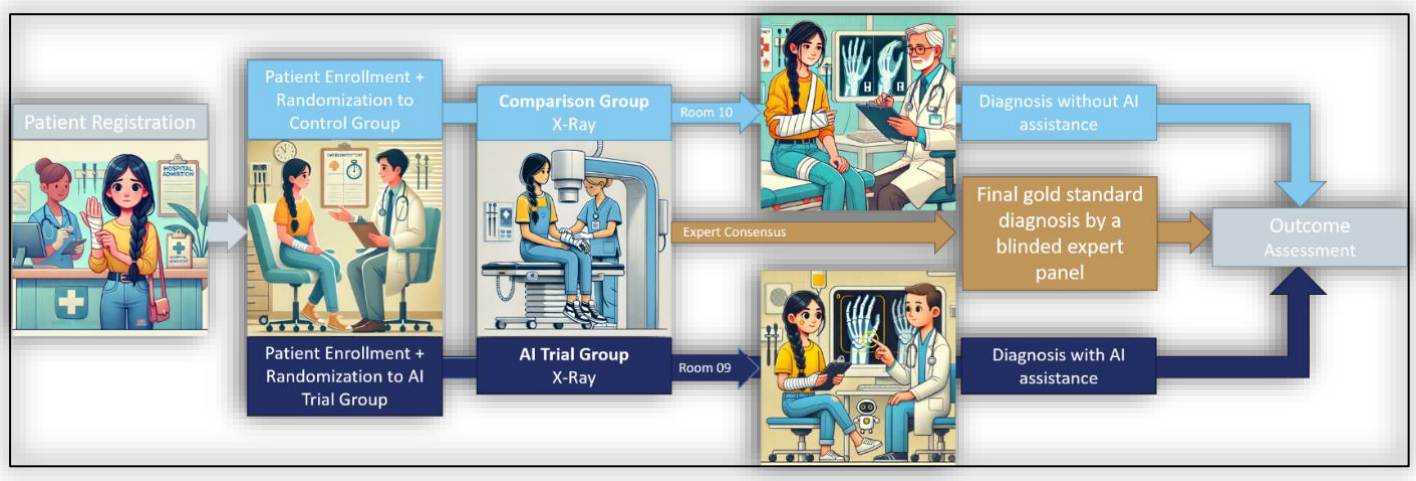


Figure 2. Study Flowchart from Patient Registration to Outcome Assessment

3. APPLICANT

Dr. Martin Breitwieser, PhD MBA BSc, University Hospital Salzburg, Department of Orthopedics and Traumatology, Müllner Hauptstr. 48, 5020 Salzburg; m.breitwieser@salk.at

4. CLINICAL TRIAL TYPE

Prospective, multicenter, pragmatic randomized controlled trial evaluating the real-world clinical impact of AI-assisted fracture detection on emergency department throughput, clinician diagnostic confidence, imaging behavior, and secondary clinical outcomes in routine emergency care.

5. OBJECTIVES

5.1 PRIMARY OBJECTIVE

- To determine whether AI-assisted fracture detection affects time from triage to discharge compared with standard physician-only interpretation in routine emergency department care.

5.2 KEY SECONDARY AND SUPPORTIVE OBJECTIVES

- To evaluate whether AI assistance influences clinician diagnostic confidence.
- To assess whether AI assistance changes the rate of additional imaging requested during the index emergency department visit.

5.3 SECONDARY CLINICAL OBJECTIVES

- To assess missed fractures through expert adjudication combined with limited retrospective chart review where required.
- To report descriptive diagnostic accuracy metrics against the expert-adjudicated reference standard.

6. INTERVENTION

Patients in the intervention group were assessed in a dedicated AI-supported workflow. Treating physicians reviewed the radiograph together with BoneView™ version 1.0.1 (Gleamer, Paris, France), a CE-marked Class IIa AI fracture detection algorithm approved for patients aged 2 years and older. At Salzburg and Hallein, BoneView™ was delivered through the Aidoc aiOS platform; at Nuremberg, it was implemented directly. Patients in the control group were assessed without AI output. All final clinical decisions were made by the treating physician.

7. KEY INCLUSION AND EXCLUSION CRITERIA

Inclusion criteria were age ≥2 years, trauma presentation with suspected fracture requiring plain radiographic imaging, mono-region injury within the approved anatomical scope of BoneView™, and informed consent. Exclusion criteria were multi-region trauma, head or cervical spine injuries, age <2 years, prior imaging or medical consultation for the same injury, contraindications to X-ray imaging, and lack of informed consent. Low-quality radiographs and multiple fractures within the same anatomical injury location were not excluded.

8. PRIMARY AND SECONDARY ENDPOINTS

The primary endpoint is time from triage to discharge. Key secondary implementation endpoints are clinician diagnostic confidence on a 0-10 scale and additional imaging requested during the index visit. Secondary clinical endpoints are missed fractures and descriptive diagnostic accuracy measures including sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy.

9. SAMPLE SIZE, STATISTICAL ANALYSES, AND POWER CONSIDERATIONS

Approximately 3,100 cases were enrolled, with approximately 2,800 expected to constitute the eligible randomized analysis dataset. The primary endpoint will be analyzed using regression-based models adjusting for physician experience, anatomical region, and study site, with physician-level clustering accounted for where appropriate. Confidence will be analyzed using ordinal regression, and binary outcomes such as additional imaging and missed fractures using logistic regression. The available sample size provides approximately 87% power to detect a 3.5-minute between-group difference in time from triage to discharge, assuming 1,400 cases per group, a standard deviation of 30 minutes, and a two-sided alpha of 0.05.

10. TRIAL DURATION

The prospective data collection phase was completed through institutional bridge funding.

Phase	Activity	Timeframe
Completed	*Mock trial, trial preparation, Data collection in Salzburg, Hallein & Nuremberg*	*May 2025 - April 30, 2026*
Analysis phase	Data entry, expert adjudication, limited chart review, statistical analysis, and manuscript preparation	Second half 2026 - early 2027 (subject to funding)
Dissemination	Publication & congress presentation	September–October 2027

11. PARTICIPATING CENTERS

- **University Hospital Salzburg, Austria**

Dept. of Orthopedics and Traumatology, Dept. of Pediatric and Adolescent Surgery, Dept. of Radiology

- **Regional Hospital Hallein, Austria**

Dept. of Orthopedics and Traumatology

- **University Hospital Nuremberg, Germany**

Dept. of Orthopedics and Traumatology, Dept. of Pediatric and Adolescent Surgery, Dept. of Radiology