FULL/LONG TITLE OF THE STUDY

External Validation of an Artificial Intelligence Tool for Paediatric Fracture Detection

SHORT STUDY TITLE / ACRONYM

External Validation of AI for Paediatric Fractures

PROTOCOL VERSION NUMBER AND DATE:

26 December 2023, Version 1.4

RESEARCH REFERENCE NUMBERS

IRAS Number:	274278
GOSH R&D Number:	20PC07
FUNDERS Number:	Funding reference: NIHR301322 NIHR Funding and Awards Search Website

SIGNATURE PAGE

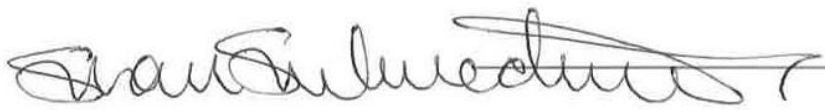
The undersigned confirm that the following protocol has been agreed and accepted and that the Chief Investigator agrees to conduct the study in compliance with the approved protocol and will adhere to the principles outlined in the Declaration of Helsinki, the Sponsor's SOPs, and other regulatory requirement.

I agree to ensure that the confidential information contained in this document will not be used for any other purpose other than the evaluation or conduct of the investigation without the prior written consent of the Sponsor

I also confirm that I will make the findings of the study publically available through publication or other dissemination tools without any unnecessary delay and that an honest accurate and transparent account of the study will be given; and that any discrepancies from the study as planned in this protocol will be explained.

Chief Investigator:

Signature:



Date:
...../...../.....

Name: (please print): **Dr. SUSAN SHELMERDINE**

KEY STUDY CONTACTS

Chief Investigator	Dr. Susan Shelmerdine Consultant Radiologist Department of Clinical Radiology, Great Ormond Street Hospital, London, UK Susan.shelmerdine@gosh.nhs.uk +44 7989 984 955
Study Co-ordinator	Ms Bhavna Saini Radiology Dept Research Co-ordinator Bhavna.saini@gosh.nhs.uk
Funder(s)	National Institute for Health Research, NIHR Grant number 301322
Key Protocol Contributors	**Dr. Susan Shelmerdine** Consultant Paediatric Radiologist, GOSH Honorary Associate Professor, UCL GOS Institute of Child Health, London, UK susan.shelmerdine@gosh.nhs.uk +44 7989 984 955 **Dr. Nicholas Woznitza** Consultant Radiographer & Clinical Academic, UCLH & Canterbury Christ Church University Nichoals.woznitza@nhs.net **Prof. Kanthan Theivendran** Consultant Orthopaedic Surgeon, Sandwell and West Birmingham Hospitals NHS Trust Hon. Prof. Engineering & Applied Science, Aston University k.theivendran@nhs.net **Dr. Sarim Ather** Consultant Academic Radiologist, National Consortium for Intelligent Medical Imaging, University of Oxford, UK Oxford University Hospitals NHS Foundation Trust sarim@raigc.com **Professor Alex Novak** Prof. Emergency Medicine, Emergency Medicine Research Oxford (EMROx) Oxford University Hospitals NHS Foundation Trust Alex.novak@ouh.nhs.uk

STUDY SUMMARY

Study Title	External Validation of an Artificial Intelligence Tool for Paediatric Fracture Detection
Internal ref. no. (or short title)	External Validation of AI for Paediatric Fractures
Study Design	Multi-case, multi-reader (MCMR) diagnostic accuracy trial. Using a multi-centric dataset, the performance of a commercially available AI tool (BoneView, Gleamer) will be compared to healthcare professionals alone (and with AI) across different subspecialties and experience levels.
Study Participants	Children aged 2 – 16 years old
Planned Size of Sample (if applicable)	Multi-case, multi-reader study (using >30 readers across different disciplines) evaluating a subset of 500 paediatric traumatic radiographs of top 4 commonly missed fracture body parts.
Follow up duration (if applicable)	No follow-up as these are retrospective data studies.
Planned Study Period	March 2022 – September 2026 (Length of NIHR funding) For this study we will attempt to 'go live' with the multi-readers study by January 2024, lasting for 3 months. Data interpretation and write-up by 6 month's time point.
Research Question/Aim(s)	1. Does a commercially available fracture detection tool using artificial intelligence improve the performance (and confidence levels) of healthcare professionals who regularly assess paediatric radiographs? 2. What are the simulated changes in management plan that would ensue with the addition of AI, and possible estimated cost effectiveness and patient impact this tool may lead to in a clinical pathway?

FUNDING AND SUPPORT IN KIND

FUNDER(S)	FINANCIAL AND NON FINANCIALSUPPORT GIVEN
NIHR	NIHR 301322 Advanced Fellowship Award

ROLE OF STUDY SPONSOR AND FUNDER

The sponsor is Great Ormond Street Hospital for Children NHS Foundation Trust. They assume overall responsibility for the initiation and management of the study.

This study is funded by the National Institute for Health Research (NIHR) through an Advanced Fellowship grant awarded to the chief investigator, Dr. Susan Shelmerdine. The funder has no role or control over the study design, conduct, data analysis and interpretation, manuscript writing, and dissemination of results and decision regarding publication of the results.

ROLES AND RESPONSIBILITIES OF STUDY MANAGEMENT COMMITEES/GROUPS & INDIVIDUALS

Study Steering Groups

The Great Ormond Street Hospital Young Persons Advisory Group (YPAG) has been involved in the design of this study and several members of the YPAG group and the associated 'parent and carers group' have volunteered to take part in steering group meetings during the life cycle of this study to ensure that decisions made regarding design, output, dissemination are acceptable and understandable to parents and their families.

For more information on how public and patient involvement has helped shape this study please visit the following links:

GOSH YPAG Group:
https://www.gosh.nhs.uk/our-research/our-research-infrastructure/nihr-great-ormond-street-hospital-brc/patient-and-public-involvement/

Case Impact Report detailing the FRACTURE Study (See pages 10 onwards):
https://media.gosh.nhs.uk/documents/PPI__Impact_Case__Studies_Report.pdf

KEY WORDS

Artificial Intelligence, Fracture, Children, Bone, Trauma

STUDY PROTOCOL
External Validation of an Artificial Intelligence Tool for Paediatric Fracture Detection

1. LAY BACKGROUND & RATIONALE

Approximately half of all the 12 million children (<16 years) in the UK will break a bone (a fracture) during childhood[1,2]. A bone X-ray helps doctors see the fracture and make decisions on treatment.

There are three problems with identifying children's fractures:
1. Growing bones can look like fractures.
2. Children have different types of fractures than adults (e.g. buckle fractures)
3. Fractures in children are not always obvious

Without an expert opinion from a radiologist (a specialist who looks at medical images), mistakes are made in diagnosing 10% of children's fractures by emergency doctors[3-6]. Unfortunately, we don't have enough specialist children's radiologists in the NHS to check all X-rays immediately, meaning a delay in recognising mistakes[7-12]. If fractures are missed, children will be left in pain and won't get the right treatment, leading to long term problems with discomfort and disability. In some cases, a fracture can be a sign of abuse, and crucial opportunities to arrange for the child's safety are missed.

Artificial intelligence (AI) has the potential to be used to find fractures on children's X-rays as accurately as a radiologist. AI is a computer programme trained to carry out actions usually done by humans – in this case to find fractures on X-rays. This could reduce the number of missed fractures, ensure children get the right treatment quickly, reduce inefficiencies for parents and children, and save money for the NHS.

2. RESEARCH QUESTION/AIM(S)

In this study we want to investigate whether using an already commercially available artificial intelligence solution can recognise fractures on X-rays in children with suspected bony injury. We also want to know how much more accurate this commercially available tool will enable healthcare professionals to be across different disciplines and experiences, and whether this may lead to changes in their management.

The principal research question is:
Does using an AI tool improve patient outcomes (i.e. reduced error rates, improved reader confidence) in a **simulated** emergency care implementation study?

We would like to see whether using such a software could potentially help doctors make better decisions about patient care by calculating how many injuries the tool may potentially help to identify (or miss) if it was to be used clinically in the future; and also how management plans may differ with the use of an AI tool.

Hypothesis

A commercially available AI algorithm can be used for accurate paediatric fracture detection, and potentially improve clinical decision making in a simulated implementation study.

3. STUDY DESIGN and METHODS

Objectives:

1. To determine the differences in diagnostic accuracy rates of human readers before and after using the AI tool for paediatric fracture detection (and whether this varies according to reader job role/specialty).
2. To determine changes in confidence rates in fracture diagnosis in human readers before and after using the commercially available AI tool.
3. Determine whether the use of the AI tool would lead to changes in patient management.
4. Determine if there is any improvement in speed of reporting when using an AI tool

Methodology:

Data:

The dataset used in this study will be derived from the FRACTURE study data set (see Appendix 1 for details) and include a subset of the images collected.

Two versions of the same subset of radiographic images will be created:

1) one version will be only the plain radiographs, and
2) the second version will include the AI assisted diagnosis which will include bounding boxes overlaid onto the plain radiographs (where there is an abnormality) with a data label (i.e. fracture / effusion).

These datasets will be both fully anonymised and only accessed by the study readers via a secure, password protected image viewer/ reader platform. Each reader will have their own dedicated password for entry onto the image interpretation platform.

According to sample size calculations (see below), the dataset should include at least 112 examinations each body part being evaluated. In this study we intend to include 125 examinations for each of 4 body parts – these include the top **4 most commonly missed/ injured body parts** in children with significant clinical consequences. These include the forearm/wrist, elbow, ankle and knee.

Limbs are the most commonly fractured body parts (accounting for 81.5% of all paediatric injuries[13]), with those of the tibial plateau (knee), radial head (elbow) and scaphoid (wrist) being most commonly missed[14], with incorrect diagnosis of paediatric elbows and wrist radiographs (20.5% of all fractures) being commonest reasons for

litigation in the NHS[15]. Although ankle fractures are less common, they account for up to 25% of all growth plate injuries[16] and thus carry a high potential for long term growth disorders if misdiagnosed.

This therefore means we will have a **dataset of 500 examinations (125 x 4 body parts),** where 35% of the examinations per body part will be abnormal to better balance rate of disease distribution in clinical practice whilst ensuring sufficient abnormal cases for evaluation (**i.e. 125 images per body part; 44 abnormal; 81 normal per body part).** Many prior studies have included cases which were 50% abnormal which has come under criticism for not reflecting 'real world' prevalence before.

The cases will only include children aged >2years old (as the commercial tool is not regulated for under 2 years old) and abnormal examinations will exclude any 'obvious' fractures. Obvious fractures will be defined as any fracture that meets at least one of the following criteria:

- Any imaging with a 'Red Dot' annotation on the radiograph that cannot be removed
- Any fracture which is angulated by more than 5 degrees
- Any fracture which is displaced (>5mm) or communited
- Any fracture impacted/shortened by >5mm
- Any fracture with obvious callus formation/ sclerosis

Our outcome measure will be related to acute fractures therefore healing fractures will be excluded.
Bone lesions will be excluded from our study, however normal anatomical variants will be included.

Ground truth / Reference standard is set as a double reading of the radiographs by two consultant radiologists, with at least 5 years experience and subspecialist skills in either paediatric or musculoskeletal radiology. Any conflicts in outcome findings will be resolved by a third radiologist (who may also be either subspecialised in paediatric or musculoskeletal radiology).

Reader Interpretation:

Readers will be recruited to participate in this study. Each reader will interpret first the radiograph (without AI assistance), and then after a washout period of 4 weeks they will read the same dataset of the same radiograph with AI assistance. The body parts and those with and without abnormalities will be randomly ordered within the dataset for every reader at each reading session (i.e. they wont read all the ankles, followed by all the wrists etc).

The readers do not need to complete all readings in one sitting, but we will expect them to complete the full reading of the dataset within a 2-month period and to not discuss or consult each other on any of their findings. There is no time limit for their radiographic interpretation, although this will be timed on the image viewing platform per image interpretation.

Information associated with each radiograph will include the age of patient and gender. Mechanism of injury, pain location, past history will not be provided but the reader will be asked to assume that there is generalised

pain at the joint in question and no significant past medical history (i.e. no genetic or metabolic bone disorder or known malignancy etc).

Feedback by each reader for each radiograph will include:

1. Presence/Absence of fracture/effusion
2. Confidence in their decision using a 5 point Likert scale (scored from 1 - 5, 1 = not confident; 5 = absolutely certain)
3. Suggested most likely management assuming healthy patient only complaining of pain (drop down menu with free text options provided, choice will depend on the specialist's role and include following options:

Radiology Staff (radiologists and reporting radiographers):

1) Report study, no need for any further action
2) Seek second opinion from another general radiology reporting staff member
3) Seek second opinion from a specialist radiologist (paediatric or musculoskeletal)
4) Suggest repeat imaging after a recommended time interval
5) Suggest additional cross-sectional imaging (e.g. CT or MRI)
6) Recommend urgent clinical review by referring emergency physician
7) Recommend urgent orthopaedic opinion

Emergency Dept Staff (nurse or doctor):

1) Request immediate imaging report from musculoskeletal/ paediatric radiology staff to decide what to do
2) Request further cross-sectional imaging (e.g. CT or MRI)
3) Discharge patient (+/- pain relief), no intervention/ follow-up needed
4) Discharge patient (+/- pain relief), with follow-up imaging and review at ED (if available)
5) Immobilise joint (+/- pain relief), with follow-up imaging and review at ED
6) Immobilise joint (+/- pain relief), with follow-up review and imaging with orthopaedic/plastics/fracture clinic
7) Refer patient for urgent inpatient orthopaedic opinion

Orthopaedic Dept Staff:

1) Request immediate imaging report from musculoskeletal/ paediatric radiology staff to decide what to do
2) Request further cross-sectional imaging (e.g. CT or MRI)
3) Discharge patient (+/- pain relief), no intervention/ follow-up needed
4) Follow-up review (+/- pain relief), without joint immobilised in orthopaedic/ fracture clinic with imaging
5) Immobilise joint (+/- pain relief), with follow-up review and imaging with orthopaedic/fracture clinic

6) Refer patient for urgent orthopaedic second opinion
7) Consideration for surgical intervention

Readers:

We will invite readers from different specialities and experience levels to participate. These will include radiologists (both general, paediatric and musculoskeletal radiologists – of any experience level), reporting radiographers, emergency department staff (i.e. physicians and senior triage nurses) and orthopaedic surgeons.

We anticipate to include at minimum 30 readers (6 readers from each group) with little (<5 years), moderate (5-10 years) and significant (>10 years) experience in their field of expertise.

Radiologists from different experience levels will be recruited voluntarily through radiology society newsletter announcements and the chief investigator's international links with the European Society of Paediatric Radiology (ESPR), European Society of Skeletal Radiology (ESSR) and the British Societies of Paediatric Radiology and Skeletal Radiology (BSPR, BSSR).

Emergency medical staff will be recruited through existing local collaborations, via the Royal College of Emergency Medicine (RCEM), and Association of Paediatric Emergency Medicine (APEM). Surgical colleagues will be recruited also via local collaborations, and via the British Society for Children's Orthopaedic Surgery (BSCOS).

Each reader will complete a consent form online (see appendix 2), with demographic details about themselves and their experience level. All collaborating readers will be included as collaborating authors on the final manuscript and if requested, will get a letter of participation for their appraisal documentation. Readers will also be provided with a participant information sheet (see appendix 3) informing them of the study.

Sample Size: Multi-reader Prospective Dataset

In order to determine number of images required for this validation, and how many readers would be required, advice was provided both by Dr. Dean Langan, Fellow in Statistics, GOS Institute of Child Health and the NIHR Research Design Services (Dr Rafael Gafoor).

Using the sample size tables published by Obuchowski N et al for 'Receiver Operating Characteristic Studies'[17] the study has been powered to detect small differences in the AUC of 0.05, with power of 80% and type 1 error rate of 5% between reader and AI algorithm performance.

Assuming that the dataset will be representative of clinical practice with approximately 20% abnormal cases, then our sample size **would need to be at least: 112 examinations for ten readers, per body part**. Given previous successful recruitment (>35 readers for a rib fracture detection study by our research group)[18],

recruitment of readers will not be a challenge for this study. In order to ensure better representation of abnormal findings, we will increase the number of examinations to 125 per body part.

Commercial AI Tool

The commercial AI tool being used is called 'BoneView' produced by a French AI company called Gleamer (details: https://www.gleamer.ai/). We will be using their latest version of this AI tool. It has received by CE and FDA approval for fracture detection in adults and children (aged > 2 years old).

The algorithm is a Deep Convolutional Neural Network (DCNN) based on the object detection framework "Detectron 2" written and further engineered in Pytorch (version 1.3). The DCNN follows the paradigm of two-stage object detectors. The first stage receives as input the DICOM image without any preprocessing or rescaling. It extracts intermediate feature maps corresponding to different spatial resolutions using a Feature Pyramid Network (FPN). A region proposal network (RPN) generates candidate boxes from these feature maps and attributes a score to each box based on the likelihood that it contains an object. If the score gets above a fixed threshold, the anchor is considered a region of interest (ROI). The features associated with the bounding box coordinates predicted by the RPN at the corresponding resolution are fed to the second stage that refines the results of the RPN. When the confidence score exceeds the predefined operating threshold, the software highlights the ROI by a rectangular box on the images.

The algorithm was developed based on a data set of 300,000+ radiographs of patients from 60+ radiology departments collected from January 2011 to May 2021. Thirty percent of the radiographs included in this data set were paediatric (<18 years old). The development data set was randomly split into 70% training, 10% validation and 20% internal test sets. The data set of the present study does not overlap with the AI development data set.

Existing evidence of the commercial AI tool: Use in children

Gleamer have already previously published the performance of their AI tool in an external validation study in different populations of children across at least three different countries.

Reference:	**Nguyen T et al[19]**	**Hayashi D et al[20]**	**Altmann-Schneider I et al[21]**
Country:	France	USA	Switzerland
Patient ages:	2 – 21 years	2 – 21 years	0 – 18 years
Sample size:	300 examinations	300 examinations	3000 radiographs
Body parts (number):	5	5	3
Body parts (location):	foot/ankle, knee/leg, hand/wrist,	hand/wrist, elbow/upper arm, shoulder/clavicle, foot/ankle, leg/knee	Lower leg Elbow Forearm

	elbow/forearm, shoulder/clavicle		
Normal : abnormal ratio:	50:50 60 images per body part, half abnormal (of these half obvious and half subtle abnormalities)	50:50 60 images per body part	N/A – as per clinical practice 1000 radiographs per body part included
Ground Truth:	Consensus reading of two musculoskeletal radiologists.	Consensus reading of two musculoskeletal radiologists.	Re-read of the original report by paediatric radiologist
Readers:	5 radiology residents and 3 paediatric radiologists.	Not applicable	Not applicable
Diagnostic accuracy:	Standalone sensitivity of AI = 91%, sensitivity of 90%.	Sensitivity per patient of 91.3% (85.6, 95.3%), specificity per patient of 90.0% (84.0, 94.3%). Sensitivity per fracture of 92.5% (87.0, 96.2%). False positive rate per patient in patients without fracture of 0.11.	Lower leg: Sensitivity 87.5%, specificity 87.5%, PPV 98.3%, and NPV 98.3%. Forearm: Sensitivity 92.9%, specificity 98.1%, PPV 98.4%, and NPV 91.7%. Elbow: Sensitivity 80.5%, specificity 94.9%, PPV 93.3%, NPV 84.7%.
Compared to humans:	Mean sensitivity for all groups was 73.3% without AI vs. 82.8% with the aid of AI (9.5% gain (95% CI: [7.05, 11.95], p<0.001). Residents: Sensitivity increased from 71.9% to 82.2%, p<0.001. Expert readers: 75.6% to 83.8% (not significant)	N/A	N/A

Limitations:	Patients with closed physes included, Only radiologist readers	No comparison with human performance	No comparison with human performance

Statistical Analysis

Diagnostic accuracy of the readers (with and without AI guidance) for each body part will be derived (i.e. sensitivity, specificity, PPV, and NPV, and 95% confidence intervals) and statistical significance between groups using AI (versus no AI) will be compared for using a two-sided McNemar test. Interobserver variability between readers (with and without AI guidance) will be derived using unweighted kappa statistics.

We will also assess for differences in confidence scores for correctly identified images between staff with and without AI guidance, and also whether there were significant differences between readers across specialties and experience levels.

Changes in clinical management will be compared using descriptive statistics (i.e. frequency and percentages) to determine how many children would be discharged with a missed fracture, or unnecessary second opinions/ additional imaging sought for cases with and without AI assistance. This will help estimate the potential future benefit and cost savings to the NHS for implementation.

Outcome Measures

In this study, for the 4 body parts with the most commonly missed paediatric fractures we will derive outputs for:

1. Diagnostic accuracy rate of fracture detection (with and without AI, and AI alone)
2. Confidence scores in fracture detection (with and without AI)
3. Inter-observer agreement rate in detection (with and without AI)
4. Which medical specialties do the AI help most (in terms of greatest improvement in accuracy – both in terms of medical specialty, experience levels)
5. Changes in clinical management/referral conferred with and without AI assistance (and therefore potential cost implications).
6. Differences in reader speed in image interpretation (with and without AI)

4. ETHICAL AND REGULATORY COMPLIANCE

Assessment and management of risk

Patient Harm:
Data used in this study will not generate any change to patient management. All data has been retrospectively collected in an anonymised fashion. Patient treatment has already been performed and will not be amended based on our results.

Data Protection:
Data from collaborating institutions will be anonymised prior to transfer, stored as encrypted data within a secure data facility, with appropriate governance oversight. No personal identifiable information will be shared with the commercial AI companies. The products used will be CE marked and FDA approved.

Research Ethics Committee (REC) and other Regulatory review & reports

HRA approval has already been granted and REC approval waived for the collection of the retrospective multicentric dataset. Ethical approval is not required for the multi-reader case study of healthcare professionals.

- If there are deviations from this study protocol, substantial amendments that require review by NHS REC will not be implemented until that review is in place and other mechanisms are in place to implement at site.
- All correspondence with the REC will be retained.
- The Chief Investigator will produce the annual reports as required.

Regulatory Review & Compliance

- Before any site will retrieve data for the study, the Chief Investigator will ensure that appropriate approvals from participating organisations are in place.

- For any amendment to the study, the Chief Investigator or designee, in agreement with the sponsor will submit information to the appropriate body in order for them to issue approval for the amendment. The Chief Investigator or designee will work with sites (R&D departments at NHS sites as well as the study delivery team) so they can put the necessary arrangements in place to implement the amendment to confirm their support for the study as amended.

Peer review

The funder, National Institute for Health Research (NIHR) has reviewed the study grant and protocol for the FRACTURE study. This study is using the same methodology as the FRACTURE study, with the only difference being in the usage of a CE marked commercially available AI tool. This has already been discussed and

approved by senior advisors on the FRACTURE study research panel. The Chief Investigator was interviewed and questioned on various specifics of this grant and study protocol prior to award of funding.

5. PATIENT AND PUBLIC INVOLVEMENT & ENGAGEMENT

Patients and the public have been vital in co-designing this research, ensuring that the aims and questions are relevant and important to them, and the study is feasible and acceptable.

In designing this research protocol and during the application for the study grant, two virtual meetings were held with the NIHR GOSH Biomedical Research Centre (BRC) Patient and Public Advisory Groups for research:

- The Young Persons Advisory Group (YPAG), part of the Generation R affiliation (24 young people attended, aged 11-21 years)
- The Parent and Carer Advisory Group (5 parent representatives attended)

Attendees already had a keen awareness of digital technologies, previously attending workshops at GOSH DRIVE, assessing digital and robotic tools for children's healthcare (https://generationr.org.uk/health-techworkshop-event-report/). Feedback for this study was positive, particularly given the strong emphasis on patient safety and potential to reduce errors.

Four YPAG and three parent and carer advisory group members have agreed to form a 'FRACTURE Study' Steering group for this project. This Steering Group will meet virtually, every 6 months for the duration of the study specifically addressing:

- Development of strategies for troubleshooting issues with data collection/ analysis
- Development of educational tools explaining AI to patients, radiologists and NHS managers
- Assistance in writing of academic papers, reviewing and revising drafts (including lay summaries)

6. PROJECT MANAGEMENT

NIHR GOSH BRC has a well-established clinical research and financial governance structure in place which will help manage this project, including remote data transfer and set up. Project-specific Research Governance will be led by the Radiology Research Lead, with quarterly governance review meetings with the extended research team to ensure that the project stays on track.

Protocol compliance

Accidental protocol deviations will be adequately documented on the relevant forms and reported to the Chief Investigator and Sponsor immediately.

Deviations from the protocol which are found to frequently recur are not acceptable, will require immediate action and could potentially be classified as a serious breach.

Data protection and patient confidentiality

All investigators and study site staff will comply with the requirements of the Data Protection Act 1998 with regards to the collection, storage, processing and disclosure of personal information and will uphold the Act's core principles. Data will be collected and maintained according to ICH-GCP standards. Any source documents that relate to individual research participants will be anonymised and maintained securely within the hospital (locked room). Personnel who have access to the database are legally bound by the confidentiality agreement in their contract of employment.

Source documents will be maintained for 15 years and will be available for inspection by authorised staff including the Chief Investigator, Study Coordinator and statistician. Source documents will be made available if requested for monitoring and audit purposes to the Ethics and Research and Development departments and for inspection by regulatory bodies.

Indemnity

There is no risk of any direct or physical harm to any patients as a result of this study due to the observational nature of the project and lack of any intervention or change in clinical practice. GOSH and NHS indemnity processes are in place. GOSH has non-negligent harm insurance and is also covered by the NHS Clinical Negligence Scheme.

Access to the final study dataset

All imaging data and any potential source documents will be maintained for 15 years and will be available for inspection by authorised staff including the Chief Investigator and Study Coordinator.

Personnel outside of the immediate research team will not be allowed access to the final dataset, in order to ensure that the overall results are not disclosed by an individual study site prior to main publication.

Source documents will be made available if requested for monitoring and audit purposes to the Ethics and Research and Development departments and for inspection by regulatory bodies.

7. DISSEMINIATION POLICY

Dissemination policy

The study protocol is planned to be published in a peer reviewed journal.

On completion of the study, the data will be analysed and tabulated and a Final Study Report prepared. This can be accessed on a dedicated 'FRACTURE Study' website (www.fracturestudy.com) where all announcements regarding the outputs, newsletters and publications will be uploaded to share with the wider public and research team.

This study aims to publish the mature results in well renowned, peer reviewed journals. All presentations and publications require authorisation from the Chief Investigator.

Participating investigators and industrial collaborators who have taken part and provided a substantial contribution to the work will have rights to co-authorship and acknowledgement in the study outputs.

Funding support from the NIHR and GOSH Biomedical Research Centre will be acknowledged within publications, however the funders and sponsor will not have any role in analysis of the data or decision on what will be published.

There are **no plans** for algorithm code and statistical code for generating the results to be made publicly available.

Authorship eligibility guidelines and any intended use of professional writers

There are no plans to use professional paid writers for any work/output arising from this study.
Criteria for individually named authors or group authorship will follow the rules as stated by The International Committee of Medical Journal Editors (ICJME).

APPENDIX 1

FRACTURE STUDY DATASET

A large multicentric dataset of anonymised paediatric radiographic imaging (in DICOM format) is being collected from several paediatric major trauma units and emergency departments as part of an NIHR funded study – entitled the FRACTURE study (ongoing).

Current collaborators (in addition to GOSH) include:

- St George's Hospital, London
- King's College Hospital, London
- Nottingham University Hospitals
- Birmingham Children's Hospital
- University Hospitals Bristol/ Bristol Children's Hospital
- St Mary's Hospital, London / Imperial
- Noah's Ark Children's Hospital for Wales, Cardiff
- Sheffield Children's Hospital, Sheffield
- Alder Hey Children's Hospital, Liverpool

As of 1 February 2023, data from three sites (St George's Hospital, GOSH and King's College Hospital) has been collected, and is pending from the remainder of the sites.

Radiographs in this dataset will include all body parts, and comprise of those referred for imaging where the child has been assessed for potential trauma (this includes suspected physical abuse and paediatric skeletal surveys (i.e. whole body series of radiographs)).

Each radiographic examination of the limbs and spine will comprise of at least two views per patient (i.e. at least a frontal and lateral view). Those of the chest and abdomen are likely to include only one view (frontal view) if part of an acute admission, or multiple views if part of the skeletal survey.

Inclusion criteria

- All patient genders
- All patient ages <16 years old
- All ethnic and socio-economic groups
- Radiographs (i.e. X-rays) of any body part acquired within the last 6 years (1 Jan 2017 – 1 Jan 2023) due to suspicion of trauma (including skeletal surveys acquired for suspected physical abuse).

Exclusion criteria

- Outside of stated age range.
- Outside stated of retrospective time period
- Imaging with any identifiable patient information which cannot be removed such as annotations on the imaging made by radiographers or markers overlying the bones imaged.

Data Transfer & Storage

Anonymised data is uploaded directly from the Picture Archiving and Communication System (PACS) at collaborating sites by the local radiologist collaborator, onto a secure, open source, cloud based imaging platform (XNAT.org[22,23]) where a second level of anonymisation takes place. This ensures that **absolutely no patient identifiable information** will be kept in the metadata or attached to the imaging tests. We will collect information on patient age and gender, but this will not include details of the date of birth.

XNAT has been widely utilised for data sharing in multicentre research studies, and requires little computing knowledge – only installation of a desktop client locally to upload the anonymised files. Data storage, use of a virtual machine and ability to run algorithm pipelines are possible through this platform, which will be set up and managed by our research team IT support staff at the UCL Centre for Medical Image Computing (CMIC).

All research data will be stored securely both via XNAT, and also on-disk encrypted hard-drives (using Advanced Encryption Standard (AES) with a 256-bit key) housed within an ISO27001 compliant data centre, with access limited to research team members. Our data handling facility conforms to NHS HSCIC Information Governance Statement of Compliance Toolkit.

Data Labelling

Each of the anonymised radiographic examinations will be reviewed and labelled by two consultant paediatric radiologists from within GOSH (each with >5 years of paediatric radiology experience). Each study will be ‘double read’, with discrepancies resolved by a consensus review if needed to provide the ‘ground truth’ (i.e. reference standard).

This standard was chosen given the similar level of interrogation required for reporting images for medico-legal proceedings (i.e. suspected physical abuse in children)[74]. In usual practice, emergency radiographs only require a report by one reporter (not necessarily a paediatric radiologist).

Consent

This is a retrospective review of imaging data that has **already been collected at collaborating centres** for usual clinical practice. This study will not be analysing or assessing real patient outcome data, nor will any patient identifiable information be required. As such, direct consent for use of the retrospective imaging data is unnecessary.

APPENDIX 2

Participant Consent Form (To complete prior to doing the readings)

Thank you for agreeing to take part in this study. We will be conducting a multi-reader study to evaluate whether healthcare professionals change their opinion and management options about paediatric fracture detection with and without the assistance of AI.

You will be asked to read 500 examinations across 4 different body parts. All images are from children, some will be normal and some abnormal.

You can do the readings in your own time (you don't have to do all cases in one day but can if you want to), but please complete the exercise before the stated deadline. You will be sent a separate email in due course with instructions on how to use the online reader platform to perform your readings.

We want to appropriately acknowledge everyone involved and have your consent to participate. All readers in this study will be included as co-authors under the collaborative group title of 'FRACTURE Study AI Reporting Research Group' and named individually in the final publication.

Please do fill out this form as accurately and completely as you can.
If you have any questions please contact: susan.shelmerdine@gosh.nhs.uk

Personal Details & Demographics

- First name/s (as you would like to appear in a publication):
- Surname (as you would like to appear in a publication):
- Preferred email contact address:
- Institutional affiliation (state hospital, city, country – you may list more than one)
- Do you have any possible conflicts of interest? (please list or state none)
- What is your gender:
    - Male, female, non-binary/gender fluid, don't want to say, other
- What is your age:
    - 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61+, don't want to say

Job Role Questions:

- What department/specialty do you work in?:
    - Radiology, Orthopaedics, Emergency Department
- What is your job role:
    - SHO (e.g. F2, ST1/2), registrar (ST3+), consultant, nurse, radiographer
- Do you consider yourself a paediatric specialist? (i.e. are you a paediatric radiologist, paediatric orthopaedic surgeon or paediatric emergency care specialist)?
    - Yes / No
- How many years experience do you have in your role?: <state>
- How confident do you already feel looking at paediatric radiographs?
    - 1 (not confident) – 5 (very confident)
- Do you have any prior experience using AI tools in your clinical practice?

- Yes Lots / Yes Some / Yes A Little / No

- On a scale of 1 to 5, how do you feel about the use of AI for paediatric fracture detection?
  (1) I am sceptical and don't think it will be helpful at all
  (2) I will need a lot of evidence to be convinced
  (3) I have no opinion either way
  (4) I am open to the idea, as long as there is some evidence it works
  (5) I am ready to start using it today for all my cases !

Consent Questions:

- I confirm that I have read and understood the invitation for the study, and have had the opportunity to ask any questions
  - I confirm
- I understand that my participation is voluntary, I am free to withdraw at any time without giving any reasons.
  - I confirm
- I understand that by taking part in this study I will form part of the 'FRCR Rapid Reporting AI Research Group' and be acknowledged as such in publications relating to this trial.
  - I confirm
- I confirm that I will not copy or share any patient images pertaining to this study on my personal electronic devices for teaching or research purposes not pertaining to this study.
  - I confirm
- I confirm that I will return my imaging interpretations for this study within the expected time frame allowed to ensure swift processing of data collection and interpretation and failure to do so will result in my exclusion.
  - I confirm

APPENDIX 3

Participant Information Sheet (circulated prior to the study/readings)

Background
BoneView (Gleamer) is a FDA cleared and CE marked AI tool for the interpretation of musculoskeletal plain radiographs and can detect and localise a variety of pathologies (e.g. fracture, effusion, dislocation, bone lesions) in adults and children aged 2 years and above.

It does this by placing a bounding box around the area of suspected abnormality and providing a certainty score (i.e. certain versus doubtful) for a particular abnormality or stating next to the image that there is no pathology present. This AI tool is intended to support healthcare professional for clinical diagnosis not to act autonomously or recommend any treatment plans or course of action. Several studies have been performed of this tool in a paediatric population showing high accuracy levels of the AI acting alone, but few have determined how the tool affects the accuracy/ performance of human readers and whether it makes any impact on the subsequent treatment of patients.

Given the challenges faced in interpreting paediatric bones on radiographs (e.g. many anatomical variants, age related changes, subtle abnormalities), this study is aiming to determine whether using an AI tool could potentially help healthcare professionals make better decisions about patient care.

What is the purpose of this study?
The purpose is to compare the impact of using an artificial intelligence tool in the detection of fractures on children's limb radiographs.

We want to know specifically whether different subspecialties and staff of different experience levels benefit equally from using the AI tool (e.g. improved confidence, improved accuracy of fracture detection), and whether this would theoretically make a difference to the care of children presenting with a suspected acute fracture.

Why have I been invited?
We are aiming to recruit readers who regularly review paediatric musculoskeletal radiographs in their job role of any level across different medical specialties – namely radiology, emergency medicine and orthopaedics. You may be an allied healthcare professional (e.g. nurse practitioner, reporting radiographer), junior doctor (e.g. foundation trainee, registrar or fellow) or consultant level in any of the aforementioned subspecialties.
If you fit into any of these staff categories then you will be most welcome to join this study.

Do I have to take part?
No, you do not. This is entirely voluntary and you may withdraw from the study at any time.

If you decide to withdraw from or stop participating in the study after already starting the work, information already submitted will be used in the analysis. We may also ask you if you'd be happy to tell us why you chose to withdraw so we can improve the experience for other participants in similar studies in the future. You will not be obliged to provide an answer if you do not wish.

What will happen if I decide to take part?
If you would like to take part you will need to complete an online consent form and survey which will be provided to you via a weblink. Your main activities will be as follows:

1) Round 1: You will be asked to complete an online baseline assessment of paediatric limb radiographs – this includes 500 cases, some normal and some abnormal. This will be using a pre-designed platform that allows you to use all the regular features you would normally expect with a PACS image viewer tool

(e.g. magnification, changing brightness etc). You will have the link to the platform through an email that will be sent to the account that you provided us in your acceptance email.

2) Round 2: Once the first round of readings has been completed, there will be a washout period of 4 weeks before you will perform the second round. In this round, you will review the 500 paediatric radiographs again in a random order, but this time there will also be an output of an AI enhanced image using the BoneView algorithm. Again, you will use the pre-designed image viewer platform to report your findings. You will have the link to the platform through an email that will be sent to the account that you provided us in your acceptance email.

3) You will be asked to complete a qualitative survey before the first reading phase (this is already included as part of the consent form), and one more after the second reading phase, which will aim to obtain feedback and comments on your project experience and monitor any changes during the study.

4) Having completed both reading phases plus the surveys, the research team that has invited you to the project will carry out the analysis and processing to generate final results for dissemination.

What training will I receive before using the imaging viewing platform?
You will be given a login and account details prior to joining the study to allow you access to review and score the radiographs. You will also be provided with a short demo video and instructions on how to use the platform with a few test cases for practice before scoring the study cases.

What should I consider?
Any healthcare professional of any experience is welcome to join this study, however we do not want to include complete novices with no experience who would not normally be asked to review paediatric radiographs as part of their job role.

You will only be able to take part in this study if you regularly review (or are expected to review) paediatric radiographs as part of your usual job. You will not be provided with any training in radiograph interpretation so it is important you already have some baseline experience in image interpretation for whatever job role where you'd be expected to carry this out.

Are there any possible disadvantages or risks from taking part?
There are no anticipated risks to taking part in the study. The time taken to complete the surveys and review scans will vary according to your expertise and experience in reviewing such cases – however you will be given a 2 month window at each reading session to complete your answers.

Are there any possible benefits from taking part?

By taking part, you will:

1) Gain experience in reading paediatric musculoskeletal radiographs
2) Gain experience interacting with a commercially available AI tool for fracture detection
3) Gain a certificate of participation for your annual appraisal
4) Be included as a 'collaborating author' on the final manuscript of the study results
5) Be helping inform the future of paediatric radiology AI in healthcare matters.

Will my taking part in the study be kept confidential?
Yes, all information provided to the study team will be kept confidential and for use only in contacting yourself for matters concerning this study. No personally identifiable information will be presented in our published results (apart from your name as a collaborating author on the main study publication), nor will any data be shared with any third parties.

What will happen to the results of the study?
The results of the study will be published in peer reviewed journals and in conference presentations/posters. Individual participant performances in the study will not be personally identified.

What if I have a problem during the study?
Contact details will be provided to you in an instruction sheet prior to starting the study with details of who to contact for any technical issues during the study or clarification of the instructions unclear.

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, you should contact the Chief Investigator of the study (details below).

Have patients and the public been involved in this study?
Yes, this study was designed in conjunction with our FRACTURE Study steering group committee made up of several children/ young persons and parent/carer representatives. Further information about who they are and their involvement is available via the study website: www.fracturestudy.com

If your questions are not adequately answered then please contact the Chief Investigator, Dr Susan Shelmerdine with your queries. Thank you for considering to take part.
Email: susan.shelmerdine@gosh.nhs.uk

REFERENCES

1. Jones IE, Willimas SM, Dow N, Goulding A. How Many Children Remain Fracture-Free During Growth? A Longitudinal Study of Children and Adolescents Participating in the Dunedin Multidisciplinary Health and Development Study. *Osteoporos Int* 2002; **13**: 990-5.
2. Cooper CL, Dennison EM, Leufkens HGM, Bishop N, van Staa TP. Epidemiology of Childhood Fractures in Britain: A Study Using the General Practice Research Database. *J Bone Miner Res* 2009; **19**(12): 1976 - 81.
3. Eakins C, Ellis WD, Pruthi S, et al. Second Opinion Interpretations by Specialty Radiologists at a Pediatric Hospital: Rate of Disagreement and Clinical Implications. *AJR Am J Roentgenol* 2012; **199**: 916-20.
4. Taves J, Skitch S, Valani R. Determining the clinical significance of errors in pediatric radiograph interpretation between emergency physicians and radiologists. *Cjem* 2018; **20**(3): 420-4.
5. Klein EJ, Koenig M, Diekema DS, Winters W. Discordant radiograph interpretation between emergency physicians and radiologists in a pediatric emergency department. *Pediatric emergency care* 1999; **15**(4): 245-8.
6. Al-Sani F, Prasad S, Panwar J, et al. Adverse Events from Emergency Physician Pediatric Extremity Radiograph Interpretations: A Prospective Cohort Study. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine* 2020; **27**(2): 128-38.
7. Halliday K, Drinkwater K, Howlett DC. Evaluation of paediatric radiology services in hospitals in the UK. *Clin Radiol* 2016; **71**(12): 1263-7.
8. McColgan M, Winch R, Clark SJ, Ewing C, Modi N, Greenough A. The changing UK paediatric consultant workforce: report from the Royal College of Paediatrics and Child Health. *Archives of disease in childhood* 2017; **102**(2): 170-3.
9. Aquino MR, Maresky HS, Amirabadi A, et al. After-hours radiology coverage in children's hospitals: a multi-center survey. *Pediatr Radiol* 2020.
10. Davies FC, Newton T. Paediatric emergency medicine consultant provision in the UK: are we there yet? *Archives of disease in childhood* 2015; **100**(11): 1016-7.
11. Radiologists RCo. National Audit of Paediatric Radiology Services in Hospitals. 2015. https://www.rcr.ac.uk/sites/default/files/auditreport_paediatricrad.pdf (accessed 24 May 2020.
12. Commission CCQ. Radiology review: A national review of radiology reporting within the NHS in England. 2018. https://www.cqc.org.uk/publications/themed-work/radiology-review (accessed 22 May 2020.
13. Jones S, Tyson S, Young M, Gittins M, Davis N. Patterns of Moderate and Severe Injury in Children After the Introduction of Major Trauma Networks. *Archives of disease in childhood* 2019; **104**(4): 366-71.
14. Molins CM, Martinez M. Common Missed Radiographic Findings. *Emergency Medicine Reports* 2016; **37**(4).
15. Atrey A, Nicolaou N, Katchburian M, Norman-Taylor F. A review of reported litigation against English health trusts for the treatment of children in orthopaedics: present trends and suggestions to reduce mistakes. *Journal of children's orthopaedics* 2010; **4**(5): 471-6.
16. Crawford AH, Al-Sayyad MJ. Fractures and dislocations of the foot and ankle. Philadelphia: Saunders; 2003.
17. Obuchowski NA. Sample size tables for receiver operating characteristic studies. *AJR Am J Roentgenol* 2000; **175**(3): 603-8.
18. Shelmerdine SC, Langan D, Hutchinson JC, et al. Chest radiographs versus CT for the detection of rib fractures in children (DRIFT): a diagnostic accuracy observational study. *The Lancet Child & adolescent health* 2018; **2**(11): 802-11.
19. Nguyen T, Maarek R, Hermann A-L, et al. Assessment of an artificial intelligence aid for the detection of appendicular skeletal fractures in children and young adults by senior and junior radiologists. *Pediatric Radiology* 2022; **52**(11): 2215-26.
20. Hayashi D, Kompel AJ, Ventre J, et al. Automated detection of acute appendicular skeletal fractures in pediatric patients using deep learning. *Skeletal Radiology* 2022; **51**(11): 2129-39.
21. Altmann-Schneider I, Kellenberger CJ, Pistorius SM, et al. Artificial intelligence-based detection of paediatric appendicular skeletal fractures: performance and limitations for common fracture types and locations. *Pediatr Radiol* 2023.

22. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The Extensible Neuroimaging Archive Toolkit: An Informatics Platform for Managing, Exploring, and Sharing Neuroimaging Data. *Neuroinformatics* 2007; **5**(1): 11-34.
23. XNAT Open Source Imaging Informatics. https://www.xnat.org/about/ (accessed 10 June 2020.