





BMJ Open Artificial intelligence (AI) for paediatric fracture detection: a multireader multicase (MRMC) study protocol

Susan C Shelmerdine ^{1,2,3}, Cato Pauling,² Emma Allan,¹ Dean Langan,^{2,4} Emily Ashworth,¹ Ka-Wai Yung,⁵ Joy Barber,⁶ Saira Haque,⁷ David Rosewarne,⁸ Nick Woznitza,^{9,10} Sarim Ather ¹¹, Alex Novak ¹², Kanthan Theivendran,¹³ Owen J Arthurs ^{1,2,3}

To cite: Shelmerdine SC, Pauling C, Allan E, *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

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-084448>).

Received 19 January 2024
Accepted 25 October 2024

ABSTRACT

Introduction Paediatric fractures are common but can be easily missed on radiography leading to potentially serious implications including long-term pain, disability and missed opportunities for safeguarding in cases of inflicted injury. Artificial intelligence (AI) tools to assist fracture detection in adult patients exist, although their efficacy in children is less well known. This study aims to evaluate whether a commercially available AI tool (certified for paediatric use) improves healthcare professionals (HCPs) detection of fractures, and how this may impact patient care in a retrospective simulated study design.

Methods and analysis Using a multicentric dataset of 500 paediatric radiographs across four body parts, the diagnostic performance of HCPs will be evaluated across two stages—first without, followed by with the assistance of an AI tool (BoneView, Gleamer) after an interval 4-week washout period. The dataset will contain a mixture of normal and abnormal cases. HCPs will be recruited across radiology, orthopaedics and emergency medicine. We will aim for 40 readers, with ~14 in each subspecialty, half being experienced consultants. For each radiograph HCPs will evaluate presence of a fracture, their confidence level and a suitable simulated management plan. Diagnostic accuracy will be judged against a consensus interpretation by an expert panel of two paediatric radiologists (ground truth). Multilevel logistic modelling techniques will analyse and report diagnostic accuracy outcome measures for fracture detection. Descriptive statistics will evaluate changes in simulated patient management.

Ethics and dissemination This study was granted approval by National Health Service Health Research Authority and Health and Care Research Wales (REC Reference: 22/PR/0334). IRAS Project ID is 274 278. Funding has been provided by the National Institute for Health and Care Research (NIHR) (Grant ID: NIHR-301322). Findings from this study will be disseminated through peer-reviewed publications, conferences and non-peer-reviewed media and social media outlets.

Trial registration number ISRCTN12921105.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Performing a large multireader study evaluating paediatric fracture detection with and without artificial intelligence (AI) assistance across different medical subspecialties and experience levels will better evaluate which healthcare professionals benefit most from AI assistance.
- ⇒ Our imaging dataset represents a range of children's ages and body parts across multiple National Health Service trusts to comprehensively evaluate performance of a commercially available AI algorithm.
- ⇒ The UK-based population used in this study, lack of patient history with predefined simulated clinical management choices may not exactly mimic real-world practices and outcomes.
- ⇒ Replicating societal and ethical biases, with a comprehensive health economic evaluation of providing AI assistance for fracture detection is difficult to achieve, but our study will provide a guide for future studies.

INTRODUCTION

Approximately, half of all the 12 million children (<16 years) in the UK will fracture a bone during childhood.^{1,2} Radiographic imaging is the first-line imaging tool for assessing the presence and extent of injury. Unfortunately, the interpretation of paediatric fractures is challenging for many healthcare professionals (HCPs) as children sustain different types of injury to adults, which can sometimes be subtle to identify (eg, buckle or Salter Harris fractures), exhibit a wide range of normal appearances across different ages and, sometimes normal physes can be mistaken for injuries. Furthermore, the level of experience of HCPs varies and while patients should, under best practice principles, not be discharged from hospital without a radiology report, in reality, this is not usually available.



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Susan C Shelmerdine;
susan.shelmerdine@gosh.
nhs.uk

In one study, researchers found that misdiagnoses were made in 10% of children's fractures by emergency doctors.³⁻⁶ Unfortunately, due to workforce pressures and staff shortages,⁷ doctors with subspecialist skills in imaging or musculoskeletal injuries are not readily available 24/7 in a busy emergency department. This leads to potential delays in recognising mistakes,⁸⁻¹³ long-term pain and discomfort for the child and, in some situations, missed opportunities for safeguarding (as fractures can be the first sign of inflicted injury).¹⁴

Recently, many artificial intelligence (AI) tools have been developed and demonstrated high diagnostic accuracy rates for the detection of fractures on imaging, in some cases to the same or higher accuracy as a radiologist.^{15 16} Many of these tools, however, have been specifically designed for adults, although encouraging results have been demonstrated when these tools have been applied to children.^{17 18} Within the last year, one AI tool has specifically been approved by the US Food and Drug Administration (FDA) for use in children over the age of 2 years. If implemented clinically, it could potentially improve the quality of paediatric care, streamline orthopedic clinic referrals and reduce the likelihood of medical litigation.¹⁹

Nonetheless, widespread adoption of AI within the National Health Service (NHS) is still nascent, with various barriers to adoption identified, of which lack of sufficient evidence is one major concern.^{20 21} There is, therefore, a vital and crucial need to evaluate how such a tool may help (or hinder) different members of staff in this clinical care pathway and whether the use of such a tool would make any difference to patient management.

Objectives

In this study, the aim is to evaluate whether using a commercially available AI tool certified for paediatric

use could help HCPs make better decisions about patient care.

Primary objective:

- Determine differences in diagnostic accuracy rates of HCPs for paediatric fracture detection, before and after using the AI tool.

Secondary objectives:

- Determine whether differences in accuracy rates or effect of adding AI varies according to job role and experience.
- Determine whether user confidence in fracture diagnosis changes when using the AI tool.
- Determine whether management plans are altered following the use of the AI tool.

METHODS AND ANALYSIS

Study design

The overall study design is of a cross-over multireader multicase (MRMC) study, where each reader (a HCP) will act as their own control across two imaging interpretation stages, with an interval washout period of at least 4 weeks duration (figure 1). Differences in diagnostic accuracy rates between stages, changes in reader's confidence rates and simulated patient management will be compared. Subgroup analyses will be conducted according to reader speciality area, experience and body part interpreted.

At the first interpretation stage (1 September 2024–31 October 2024), each reader will review a dataset of 500 paediatric limb radiographs (some normal and some abnormal) without AI assistance, then after the washout period (1 November 2024–30 November 2024), they will proceed with the second interpretation stage (1 December 2024–31 January 2025) where they will each read the same dataset with AI assistance. The order of the radiographs, and thus imaged body parts and those with and without abnormalities, will be randomly ordered

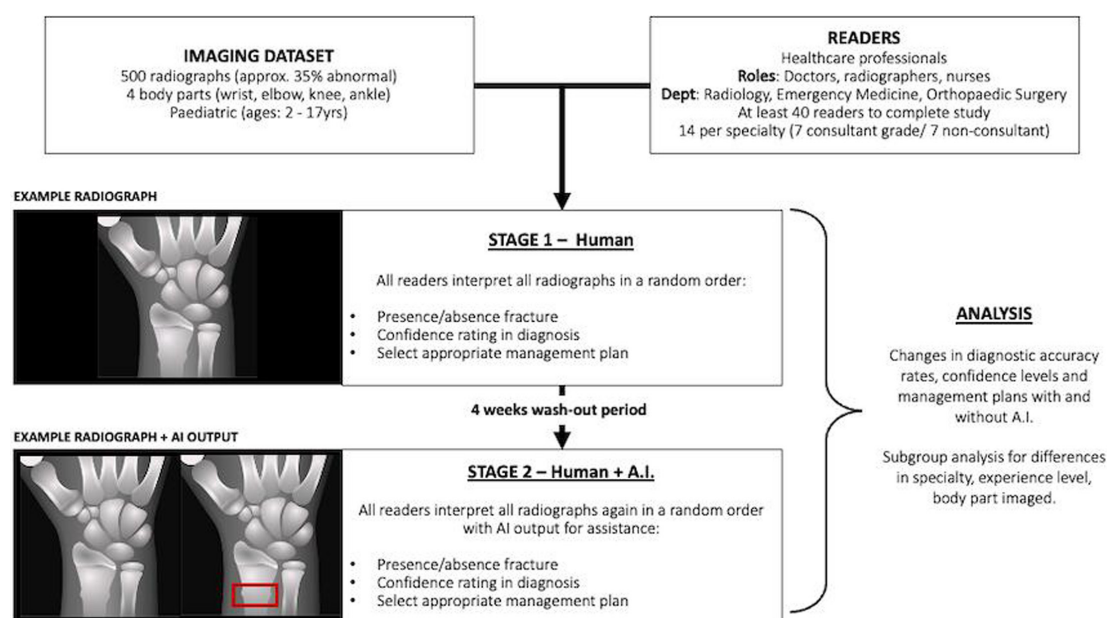


Figure 1 A diagrammatic flow chart of the multireader multicase (MRMC) study outline. AI, artificial intelligence.

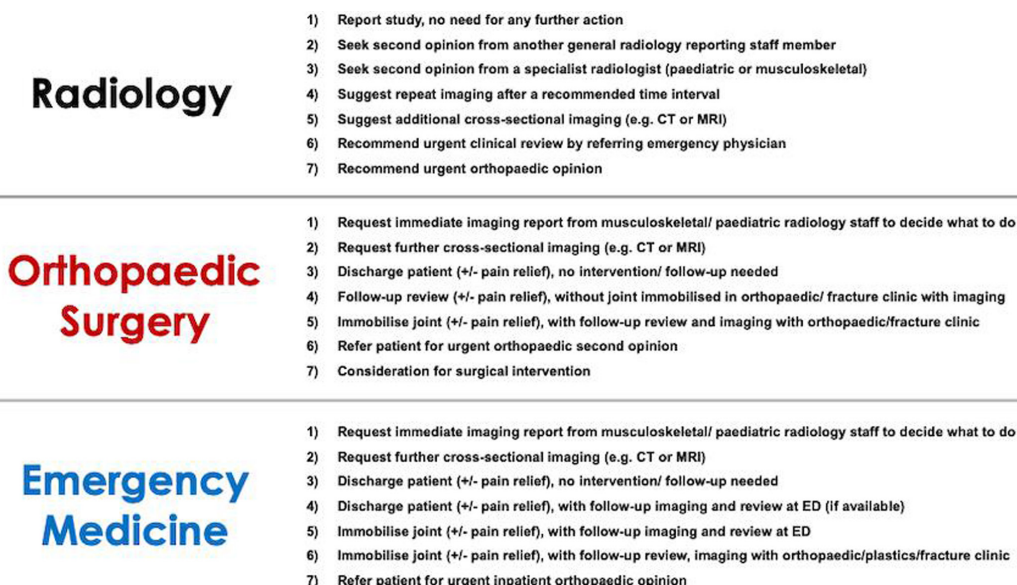


Figure 2 Simulated patient management options provided to different subspecialty readers recruited in this study. For every case reviewed, the reader will be asked to select the single best treatment plan. They will be presented with slightly different options depending on their area of medical expertise. ED, emergency department.

within the dataset for every reader at each interpretation stage to further reduce recall bias.

Readers will complete the imaging interpretations online, via a password-protected and secure imaging platform (details below). They will be given a 2-month period to complete the exercise and are informed not to seek help in the imaging interpretation with anyone else. Clinical information associated with each radiograph will include the age of patient and gender. Mechanism of injury, pain location, history will not be provided. The reader will be asked to assume that there is generalised pain at the joint in question, and no significant medical history (ie, no genetic or metabolic bone disorder or known malignancy).

Feedback requested from each reader for each radiograph will include:

1. Marking the site of a fracture on each image (or selecting 'no fracture').
2. Confidence in their decision using a 5-point Likert scale (1=not confident; 5=absolutely certain).
3. Selecting the most likely management for the patient. This will be done by providing each reader with a drop-down menu of seven predefined plans, tailored to each subspecialty, with the option for free-text comment. Examples of the different options are provided in [figure 2](#).

Inclusion and exclusion criteria (imaging cases)

500 anonymised paediatric radiographic examinations ('cases') will be derived from a larger 5-year retrospective dataset of appendicular radiographs acquired in children from two NHS trusts (King's College Hospital NHS Foundation Trust and St George's University Hospitals NHS Foundation Trust). Both trusts are major trauma centres located in London, UK serving adult and paediatric

cases. All radiographic imaging was acquired as part of routine clinical care. No change in patient management will occur as a result of this study, nor was any change in the usual imaging protocol required for this retrospective data collection.

We will include a mixture of normal and abnormal radiographic examinations according to the minimum ratio determined by our sample size calculation below. Four body parts will be used: wrist, elbow, knee and ankle. These were chosen because limbs account for 81.5% of all paediatric injuries,²² with those occurring at the knee, wrist and elbow being most commonly missed.²³ Wrist and elbow fractures encompass 20.5% of all paediatric fractures and missed fractures in these areas are one of the most common reasons for litigation in children's orthopaedic care.²⁴ Although ankle fractures are less common, they account for up to 25% of all growth plate injuries²⁵ and thus carry a high potential for long-term growth disorders if misdiagnosed.

Our inclusion criteria for all radiographic imaging in the subset of 500 cases include:

- Children aged between 2 and 18 years old (as the intended commercial AI tool is not regulated for children under 2 years of age).

In order to effectively evaluate how well the AI tool could help HCPs identify easily missed fractures, we will include abnormal radiographs that do not contain 'obvious' fractures. Obvious fractures will be defined as any fracture that meets at least one of the following criteria and therefore excluded:

- Any imaging with a 'Red Dot' annotation on the radiograph that cannot be removed (denoting the presence of either a true or false abnormality identified by the radiographer).

- ▶ Any fracture which is angulated by more than 5°.
- ▶ Any fracture which is displaced (>5 mm) or comminuted (multifragmented).
- ▶ Any fracture impacted/shortened by >5 mm.
- ▶ Any fracture with obvious callus formation/sclerosis.

As we will be evaluating the performance of acute fracture detection in an emergency setting, we will exclude healing fractures. Incidental bone lesions will also be excluded from our study, however, normal anatomical variants will be included.

Study cohort characteristics (imaging cases)

Cohort demographic characteristics for whole population and abnormal dataset for this study are available in online supplemental tables 1–3, with demographic data on normal dataset in online supplemental table 4. Overall, 500 different radiographic examinations across 500 different paediatric patients (none included more than once) will be used (comprising 183 fractures in 181 patients). The dataset will consist of 256 boys (97 with fractures) and 245 girls (84 with fractures), with mean age of 10 years (range 2–17 years).

The most common fractured bone in our dataset is the distal radius (41/181, 22.4% fractures), and the three the most common fracture types are Salter Harris 2 type injury (41/183, 22.4%), buckle fracture (39/183, 21.3%) and transverse fracture (28/183, 15.3%).

We will invite readers from different specialties and experience levels to participate. These will include HCPs working in radiology (including doctors and reporting radiographers), the emergency medicine department (including doctors and senior triage nurses) and orthopaedic surgeons. All experience levels will be welcome to participate, including those with a subspecialty interest within their field (eg, paediatric orthopaedic surgery, paediatric radiology). Readers will, however, be excluded if they do not routinely review paediatric radiographs as part of their expected job role.

Radiologists and reporting radiographers will be recruited voluntarily through society newsletter announcements via 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) as well as Society of Radiographers (SoR). 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). Orthopaedic surgical colleagues will be recruited also via local collaborations, and via the British Society for Children's Orthopaedic Surgery (BSCOS). All announcements and invitations to participate will be also posted via study collaborators through their personal social media channels.

We anticipate a minimum of 40 readers (approximately 14 readers from each specialty) equally split between trainees and consultant-level experience. Each reader will be provided instructions on how to participate in the

study and complete a consent form online asking about their demographic details, job role/specialty and experience level. Prior to any interpretation, a short video and instruction sheet on how to use the online reporting platform will be provided. Reporters will be asked to replicate their usual reporting practices as far as possible (eg, use of a suitable monitor, dim lighting) and would be able to use reference tools (eg, textbooks or websites) that they would normally consult for this task, but just not to consult other reporters.

Intervention (AI tool)

Our 'intervention' will be the use of a commercially available AI tool called 'BoneView' (V.2.3.0) produced by a French AI vendor called Gleamer (Paris, France. <https://www.gleamer.ai/>). The tool received conformity (CE Class 2A EU MDR and FDA) approval for fracture detection in adults and children (aged >2 years old) in March 2022²⁶ on full-resolution Digital Imaging and Communications in Medicine (DICOM) images. This product was chosen as it was the first to achieve FDA approval for use in children and had the greatest evidence basis among all commercial products for fracture detection on radiographs.¹⁹ The full details of how the deep learning algorithm was developed and tested have been described in the existing literature,^{17 27 28} therefore, only a brief overview of how the product was developed is provided here.

The algorithm is a Deep Convolutional Neural Network-based on the object detection framework 'Detectron 2' written and further engineered in Pytorch (V.1.3). It was developed based on a data set of 312 602 radiographs from patients across over 60 radiology departments collected from January 2011 to May 2021. 30% of the radiographs included in the dataset were paediatric (<21 years). When the algorithm confidence surpasses that of a predefined threshold set during algorithm development, the AI tool (BoneView) will create an output of a duplicate radiograph in the imaging examination with either a region of interest on the radiograph with a white square box stating presence of a fracture, a region of interest on the radiograph with a dashed white square box stating 'indeterminate' fracture or no overlay with a note below the image stating no fracture. The data set of the present study does not overlap with any examinations used in the development dataset used to create the AI tool, and nor with any data in this study be used to further train the commercial AI tool.

There have been at least three prior publications evaluating the performance of this AI tool within a French, American and Swiss paediatric dataset across a similar range of body parts proposed in this protocol.^{17 18 29} Two of these studies have included a smaller dataset than this planned study (n=300)^{17 18} with an equal split of normal and abnormal cases (not reflective of clinical practice). One was a MRMC study design using eight radiologist readers.¹⁸ None of the prior studies included a simulated patient management plan component, nor a multidisciplinary team of readers as this study is planning to.

Gleamer has provided its AI tool free of charge for evaluation in this trial but has no involvement in the study design, data analysis, reader recruitment or the decision to publish the final results.

Reference (ground truth) standard

We will use a consensus interpretation by an expert panel of two paediatric radiologists, both with at least 5 years of subspecialist radiology experience as the reference standard (so-called 'ground truth') for this dataset. A bounding box around the entire area of bone fracture on each image (if present) will be assigned where the examination is abnormal. Reference radiologists will have access to the radiographic imaging and original imaging report when setting the ground truth bounding box, as well as any follow-up imaging available; none of this will be available to recruited readers. Disagreements will be resolved by a third musculoskeletal radiologist (with similar experience level).

Data deidentification and secure storage

Scans selected for the study will be deidentified using a software called XNAT V.3.2.4,³⁰ an open source research platform for image-based biomedical research, before being uploaded to a secure image viewing platform for reader interpretation. Access to the platform will be controlled via separate user accounts and passwords for each recruited reader.

All study data generated by the readers' interpretations will be entered into a password-protected and secure database. Individual reader accuracy scores will be anonymised, and the research team will not have access to the identifying link between the participants' personal details and the data. Data about the readers' experience level and subspecialty will be retained to allow group comparisons.

All research staff will comply with the requirements of the Data Protection Act 2018³¹ with regard 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 Good Clinical Practice standards.³²

Statistical methods/data analysis plan

The STARD-AI (Standards for Reporting of Diagnostic Accuracy - Artificial Intelligence) and CLAIM (Checklist for AI in Medical Imaging) guidelines will be adhered to in the reporting of this study.^{33 34} Diagnostic accuracy of the readers (with and without AI assistance) for each body part will be derived (ie, sensitivity, specificity, positive predictive value and negative predictive value). A true positive result will be counted if a mark made by a reader at the site of a suspected fracture falls within the area of the predefined 'ground truth bounding box' area set by the reference radiologists. If the mark made by a reader falls outside a bounding box, it will be counted as a false positive. Where a bounding box was set by the ground truth, but no mark made on the image then a false negative result will be assigned to the reader.

Estimates of these accuracy statistics will be derived through multilevel logistic regression models, with the reader included as a random-intercept. From these models, we will report 95% CIs, and p values (significance level set at 5%). Independent variables will be added, including the reader's job role and experience, to assess their relationship with diagnostic accuracy. We will additionally present these same statistics to the reader, as derived through the random effects of the models, to explore the relationship between sensitivity and specificity. Intraobserver variability for the diagnostic accuracy of fracture detection before and after the use of AI assistance will also be evaluated, with subanalysis conducted to account for different reader medical specialty subgroups.

We will assess for differences in confidence scores for correctly identified images between HCPs with and without AI guidance, and also whether there were significant differences between readers across specialties and experience levels. We will also evaluate how an indeterminate AI reading affects reader decisions.

Changes in clinical management will be compared using descriptive statistics (ie, frequency and percentages) to determine, for example, 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 could provide information to help estimate potential future benefit and cost savings to the NHS at a future clinical implementation stage, where appropriate.

Sample size and power calculation

Using the sample size tables published by Obuchowski *et al* for 'Receiver Operating Characteristic Studies',³⁵ the study has been powered to detect small differences in the AUC (area under the receiver operating curve) 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 at least 20% abnormal cases, our sample size would need to be at least 112 examinations for at least 10 readers, per body part.

In order to ensure better representation of different abnormal findings, we have increased the number of examinations to 125 per body part, with approximately one-third of the cases being abnormal (ie, between 44 and 46 abnormal cases per body part, see online supplemental tables 1–3).

Patient and public involvement

In designing this research protocol and in the application for the funding, the NIHR GOSH Biomedical Research Centre Patient and Public Advisory Groups for research were consulted which included 'The Young Persons Advisory Group (YPAG)' (comprising 24 young people, aged 11–21 years) and 'The Parent and Carer Advisory Group (PCAG)' (comprising 5 parent representatives).³⁶ Many of the children were familiar with the concept of AI,³⁷ and of these, four YPAG and three PCAG members

volunteered to form the 'FRACTURE Patient and Public Involvement & Engagement (PPIE) Steering Committee' for this project and related works.³⁸ Their input has confirmed to us that patients prefer to see how doctors can be helped (rather than replaced) by AI, and therefore, this study aims to understand if AI can enhance current clinical practices and the impact this could have on patient care.

Ethics and dissemination

Human research ethics committee approval

This study was granted approval by NHS Health Research Authority (HRA) and Health and Care Research Wales (HCRW) (REC Reference: 22/PR/0334). IRAS Project ID is 274278. Informed consent was not required for the use of fully anonymised, retrospective imaging data for this study. Written consent will be received by all readers within this study prior to the interpretation exercises.

Intended publications and research dissemination

Datasets generated and/or analysed during the current study are not publicly available due to data confidentiality agreements with data custodians. Results generated by the research will be made publicly available at the summary level. Manuscripts addressing the study aims will be published in peer-reviewed journals and will also be presented at relevant national and international conferences. Findings will also be disseminated via social media and online blogs.

Study outcomes will be disseminated to all relevant clinical and non-clinical stakeholders which include our FRACTURE PPIE Steering Group, the wider Great Ormond Street Hospital YPAG and PCAG members, members of the ESPR, ESSR, BSPR, BSSR, SoR, RCEM, APEM, BSCOS, members of the NIHR Imaging Science Working Group and also the Clinical AI interest group of the Alan Turing Institute. The findings and awareness raised by the study and its dissemination will help inform future AI evaluation for paediatric healthcare, policy decisions and raise awareness of AI training needs for various multidisciplinary subspecialties and HCPs who may encounter such tools as part of their role.

Author affiliations

¹Clinical Radiology, Great Ormond Street Hospital for Children, London, UK

²UCL Great Ormond Street Institute of Child Health, London, UK

³Great Ormond Street Hospital NIHR Biomedical Research Centre, London, UK

⁴Centre of Applied Statistics Courses, University College London, London, UK

⁵Wellcome/ EPSRC Centre for Interventional and Surgical Sciences, London, UK

⁶Clinical Radiology, St George's Healthcare NHS Trust, London, UK

⁷Clinical Radiology, Kings College Hospital NHS Foundation Trust, London, UK

⁸Clinical Radiology, Royal Wolverhampton Hospitals NHS Trust, Wolverhampton, UK

⁹School of Allied Health Professions, Faculty of Medicine, Health and Social Care, Canterbury Christ Church University, Canterbury, UK

¹⁰Clinical Radiology, University College London Hospitals NHS Foundation Trust, London, UK

¹¹Oxford University Hospitals NHS Foundation Trust, Oxford, UK

¹²Emergency Medicine Research Oxford, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

¹³Orthopaedic Surgery, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK

X Susan C Shelmerdine @SusieShells and Kanthan Theivendran @KTheivendran

Acknowledgements We would like to acknowledge Ms Deirdre Leyden, Patient and Public Involvement /Engagement (PPIE) Lead for research at Great Ormond Street Hospital for Children NHS Foundation Trust, and the following persons, who form part of the FRACTURE (Fast Reporting using Artificial intelligence for Children's Traumatic Radiology Examinations) Study Patient and Public Involvement & Engagement (PPIE) Steering Committee for their help in providing feedback for this study: Lauren Lee, Laila Xu, Oceiah Annesley, Maryam Lyden, Becky Harmston, Paul Musticone and Viki Ainsworth. We also acknowledge the kind help from the PACS teams at both St. George's Hospital NHS Foundation Trust and King's College Hospital NHS Foundation Trust, in particular Mr Mukisa Scarinzi and Mr Gregory Stansil. Finally, we acknowledge the assistance on this protocol from Ms Jeanne Ventre and Mr Daniel Jones from Gleamer.

Contributors SS conceived the idea, planned and designed the study protocol. KT, DL, AN, SA and NW contributed to the development of the protocol, study design and methods. CP, EAllan and SS acquired imaging data and checked data for quality control. DL and SS contributed to the data and statistical analysis plan. SS wrote the first draft. All authors (SS, CP, EAllan, DL, EAshworth, K-WY, JB, SH, DR, SA, AN, KT and OJA) critically revised the draft and checked the content for important intellectual content. All authors (SS, CP, EAllan, DL, EAshworth, K-WY, JB, SH, DR, SA, AN, KT and OJA) approved the final written manuscript. The guarantor responsible for the overall content of the article is SS.

Funding SS is funded by a National Institute for Health Research (NIHR) Advanced Fellowship Award (NIHR-301332). CP is funded by the Great Ormond Street Hospital Children's Charity (GOSHCC) (Award Number: VS0618). K-WY is funded by grants from the Wellcome/EPSCRC Centre for Interventional and Surgical Sciences (WEISS) (Award Number: 203145Z/16/Z, NS/A000050/1). NW is funded by research grants awarded by NHSx Accelerated Access Collaborative (AAC) award, SBRI Healthcare and Health Education England (HEE) (Award number not applicable). SA is funded by an NIHR Research for Patient Benefit Grant (NIHR-204982). OJA is funded by a National Institute for Health Research (NIHR) Career Development Fellowship (NIHR-CDF-2017-10-037).

Disclaimer This article presents independently funded research—the views expressed are those of the author(s), and not necessarily those of the National Health Service (NHS), the Department of Health or any of the aforementioned funding bodies.

Competing interests NW reports the following competing interests: consultancy fees from InHealth Reporting, SM Radiology; speaker honoraria fees from AstraZeneca and travel grant award from Qure.ai Technologies. All other authors (SS, CP, EAllan, DL, EAshworth, K-WY, JB, SH, DR, SA, AN, KT and OJA) deny any competing interest statements.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Susan C Shelmerdine <http://orcid.org/0000-0001-6642-9967>

Sarim Ather <http://orcid.org/0000-0001-9614-5033>

Alex Novak <http://orcid.org/0000-0002-5880-8235>

Owen J Arthurs <http://orcid.org/0000-0003-1213-3516>

REFERENCES

- 1 Jones IE, Williams SM, Dow N, *et al*. 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 C, Dennison EM, Leufkens HG, *et al*. Epidemiology of Childhood Fractures in Britain: A Study Using the General Practice Research Database. *J Bone Miner Res* 2004;19:1976–81.
- 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;20:420–4.
- 5 Klein EJ, Koenig M, Diekema DS, *et al*. Discordant radiograph interpretation between emergency physicians and radiologists in a pediatric emergency department. *Pediatr Emerg Care* 1999;15:245–8.
- 6 Al-Sani F, Prasad S, Panwar J, *et al*. Adverse Events from Emergency Physician Pediatric Extremity Radiograph Interpretations: A Prospective Cohort Study. *Acad Emerg Med* 2020;27:128–38.
- 7 Royal College of Radiologists R. RCR Clinical Radiology Workforce Census, 2022. Available: <https://www.rcr.ac.uk/clinical-radiology/rcr-clinical-radiology-workforce-census-2022>
- 8 Halliday K, Drinkwater K, Howlett DC. Evaluation of paediatric radiology services in hospitals in the UK. *Clin Radiol* 2016;71:1263–7.
- 9 McColgan M, Winch R, Clark SJ, *et al*. The changing UK paediatric consultant workforce: report from the Royal College of Paediatrics and Child Health. *Arch Dis Child* 2017;102:170–3.
- 10 Aquino MR, Maresky HS, Amirabadi A, *et al*. After-hours radiology coverage in children's hospitals: a multi-center survey. *Pediatr Radiol* 2020;50:907–12.
- 11 Davies FC, Newton T. Paediatric emergency medicine consultant provision in the UK: are we there yet? *Arch Dis Child* 2015;100:1016–7.
- 12 Radiologists RCo. National Audit of Paediatric Radiology Services in Hospitals, 2015. Available: https://www.rcr.ac.uk/sites/default/files/auditreport_paediatricrad.pdf
- 13 Commission CCQ. Radiology review: A national review of radiology reporting within the NHS in England 2018, Available: <https://www.cqc.org.uk/publications/themed-work/radiology-review> [Accessed 22 May 2020].
- 14 Karmazyn B, Wanner MR, Marine MB, *et al*. The added value of a second read by pediatric radiologists for outside skeletal surveys. *Pediatr Radiol* 2019;49:203–9.
- 15 Kuo RYL, Harrison C, Curran T-A, *et al*. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology* 2022;304:211785:50–62.
- 16 Sheldermine SC. Artificial intelligence for fracture detection and classification in paediatric radiology: a systematic review prospero international prospective register of systematic. 2020.
- 17 Hayashi D, Kompel AJ, Ventre J, *et al*. Automated detection of acute appendicular skeletal fractures in pediatric patients using deep learning. *Skeletal Radiol* 2022;51:2129–39.
- 18 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. *Pediatr Radiol* 2022;52:2215–26.
- 19 Pauling C, Kanber B, Arthurs OJ, *et al*. Commercially available artificial intelligence tools for fracture detection: the evidence. *BJROpen* 2023;6.
- 20 Eltawil FA, Atalla M, Boulous E, *et al*. Analyzing Barriers and Enablers for the Acceptance of Artificial Intelligence Innovations into Radiology Practice: A Scoping Review. *Tomography* 2023;9:1443–55.
- 21 Royal College of Radiologists R. Overcoming Barriers to AI Implementation in Imaging: Outcome of an RCR Expert Stakeholder Day 2022, Available: https://www.rcr.ac.uk/media/05mp1eda/overcoming_barriers_to_ai_implementation_in_imaging_v3.pdf [Accessed 5 Jan 2024].
- 22 Jones S, Tyson S, Young M, *et al*. Patterns of moderate and severe injury in children after the introduction of major trauma networks. *Arch Dis Child* 2019;104:366–71.
- 23 Molins CM, Martinez M. Common Missed Radiographic Findings. *Emerg Med Rep* 2016;37.
- 24 Atrey A, Nicolaou N, Katchburian M, *et al*. A review of reported litigation against English health trusts for the treatment of children in orthopaedics: present trends and suggestions to reduce mistakes. *J Child Orthop* 2010;4:471–6.
- 25 Crawford AH, Al-Sayyad MJ. Fractures and Dislocations of the Foot and Ankle. Philadelphia: Saunders, 2003:516–37.
- 26 FDA. Gleamer BoneView v2.5.0 FDA documentation Ref, Available: https://www.accessdata.fda.gov/cdrh_docs/pdf21/K212365.pdf [Accessed 24 Feb 2023].
- 27 Guermazi A, Tannoury C, Kompel AJ, *et al*. Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence. *Radiology* 2022;302:627–36.
- 28 Duron L, Ducarouge A, Gillibert 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;300:120–9.
- 29 Altmann-Schneider I, Kellenberger CJ, Pistorius S-M, *et al*. Artificial intelligence-based detection of paediatric appendicular skeletal fractures: performance and limitations for common fracture types and locations. *Pediatr Radiol* 2024;54:136–45.
- 30 Marcus DS, Olsen TR, Ramaratnam M, *et al*. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 2007;5:11–34.
- 31 Gov uk. The Data Protection Act, 2018. Available: <https://www.gov.uk/data-protection#:~:text=Under%20the%20Data%20Protection%20Act,access%20personal%20data>
- 32 Authority NHR. Good Clinical Practice, 2020. Available: <https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/good-clinical-practice>
- 33 Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol Artif Intell* 2020;2:e200029.
- 34 Sounderajah V, Ashrafian H, Aggarwal R, *et al*. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. *Nat Med* 2020;26:807–8.
- 35 Obuchowski NA. Sample size tables for receiver operating characteristic studies. *AJR Am J Roentgenol* 2000;175:603–8.
- 36 GenerationR. Generation R - Young People Improving Research 2023, Available: <https://generationr.org.uk/about/>
- 37 Visram S, Leyden D, Annesley O, *et al*. Engaging children and young people on the potential role of artificial intelligence in medicine. *Pediatr Res* 2023;93:440–4.
- 38 Study F. FRACTURE study website. 2020. Available: <https://fracturestudy.com/>

SUPPLEMENTARY MATERIAL

Table S1 - Complete imaging dataset demographics
Demographic characteristics for all cases (both normal and abnormal) by body part and NHS trust (n = 500).
KCH – King’s College Hospital NHS Foundation Trust; SGH – St. George’s University Hospitals NHS Foundation Trust

	Laterality		Gender		Total Cases	Age Range (years)	Median Age (years)	Average Age (years)
	R	L	M	F				
Wrist								
KCH	26	33	24	35	59	2 – 15	11	10
SGH	32	34	40	26	66	4 – 16	11	12
Combined Wrist	58	67	64	61	125	2 – 16	11	11
Elbow								
KCH	26	31	26	31	57	2 – 14	7	7
SGH	35	33	37	31	68	2 – 14	8	8
Combined Elbow	61	64	63	62	125	2 – 14	8	8
Knee								
KCH	34	26	29	31	60	2 – 15	10	10
SGH	34	31	37	28	65	3 – 16	11	11
Combined Knee	68	57	66	59	125	2 – 17	11	10
Ankle								
KCH	25	27	27	27	52	3 – 14	10	10
SGH	39	34	36	36	73	3 – 16	10	10
Combined Ankle	64	61	63	63	125	3 – 16	10	10
TOTAL CASES	244 (48.8%)	257 (51.4%)	256 (51.2%)	245 (49.0%)	500 (100%)	2 – 17	10	10

Table S2 - Abnormal imaging dataset demographics
Demographic characteristics for abnormal radiographic examinations, separated by body part and NHS trust (n = 181).
KCH – King’s College Hospital NHS Foundation Trust; SGH – St. George’s University Hospitals NHS Foundation Trust

	Laterality		Gender		Total Examinations	Age Range (years)	Median Age (years)	Average Age (years)
	R	L	M	F				
Wrist								
KCH	12	13	11	14	25	3 – 15	10	10
SGH	11	9	14	6	20	8 – 16	12	12
Combined Wrist	23	22	25	20	45	3 – 16	11	11
Elbow								
KCH	11	12	12	11	23	3 – 14	5	6
SGH	14	9	13	10	23	3 – 14	6	6
Combined Elbow	25	21	25	21	46	3 – 14	6	6
Knee								
KCH	13	9	9	13	22	3 – 15	10	9
SGH	12	12	18	6	24	3 – 16	11	11
Combined Knee	25	21	27	19	46	3 - 17	10	10
Ankle								
KCH	6	8	6	8	14	3 – 13	10	9
SGH	17	13	14	16	30	3 – 16	10	9
Combined Ankle	23	21	20	24	44	3 – 16	10	9
TOTAL CASES	96 (53.0%)	85 (47.0%)	97 (53.5%)	84 (46.4%)	181 (100%)	3 - 17	9	9

Table S3 - Abnormal imaging dataset pathologies
Abnormal fracture pathologies in our imaging dataset by fracture type and body part (n, %).

	Buckle	Bowing	Salter Harris 1	Salter Harris 2	Salter Harris 3	Salter Harris 4	Greenstick	Avulsion	Transverse	Spiral	Supracondylar	Epiphyseal	Vertical	TOTAL
WRIST														
Distal radius	25	1	1	8	1	0	3	0	2	0	0	0	0	41 (22.4)
Distal ulna	0	0	0	0	0	0	0	0	3	0	0	0	0	3 (1.6)
Metacarpus	0	0	0	0	0	0	0	0	2	0	0	0	0	2 (1.1)
ELBOW														
Proximal radius	1	0	0	4	0	0	0	1	6	0	0	0	0	12 (6.6)
Proximal ulna	0	0	0	0	0	0	0	0	5	0	0	0	0	5 (2.7)
Distal humerus	0	0	0	0	0	0	0	2	1	0	26	0	0	29 (15.8)
KNEE														
Distal femur	2	0	0	3	1	0	0	1	2	0	0	0	0	9 (4.9)
Proximal tibia	6	0	0	5	0	1	0	8	0	0	0	0	0	23 (12.6)
Proximal fibula	0	0	0	0	0	0	0	0	1	0	0	0	0	2 (1.1)
Patella	0	0	0	0	0	0	0	6	6	0	0	0	1	13 (7.1)
ANKLE														
Talus	0	0	0	0	0	0	0	1	0	0	0	0	0	1 (0.5)
Distal fibula	5	0	1	12	0	0	0	2	0	0	0	0	0	26 (14.2)
Distal tibia	0	0	0	9	2	1	0	0	0	5	0	0	0	17 (9.3)
TOTAL CASES	39 (21.3)	1 (0.5)	2 (1.1)	41 (22.4)	4 (2.2)	2 (1.1)	3 (1.6)	21 (11.5)	28 (15.3)	5 (2.7)	26 (14.2%)	10 (5.5)	1 (0.5)	183

Table S4 - Normal imaging dataset demographics
Imaging dataset – demographic characteristics for normal cases by body part and NHS trust (n = 319)
KCH – King’s College Hospital NHS Foundation Trust; SGH – St. George’s University Hospitals NHS Foundation Trust

	Laterality		Sex		Total Cases	Age Range (years)	Median Age (years)	Average Age (years)
	R	L	M	F				
Wrist								
KCH	14	20	13	21	34	2-15	11	10
SGH	21	25	26	20	46	4-16	10	11
Combined Wrist	35	45	39	41	80	2-16	11	11
Elbow								
KCH	15	19	14	20	34	2-13	8	8
SGH	21	24	24	21	45	2-14	10	9
Combined Elbow	36	43	38	41	79	2-14	9	9
Knee								
KCH	17	21	20	18	38	2-15	10	10
SGH	19	22	19	22	41	3-16	11	10
Combined Knee	36	43	39	40	79	2-16	11	10
Ankle								
KCH	19	19	19	19	38	4-14	10	10
SGH	22	21	23	20	43	3-16	10	10
Combined Ankle	41	40	42	39	81	3-16	10	10
TOTAL CASES	148 (46.4%)	171 (53.9%)	158 (49.8%)	161 (50.5%)	319 (100%)	2 - 16	10	10