NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

# Artificial intelligence software to help detect fractures on X-rays in urgent care: An Early Value Assessment

Caroline Farmer,[1*] Helen Coelho,[1] Madhusubramanian Muthukumar,[1] Sophie Robinson,[1] Robert Meertens,[2] Obioha C Ukoumunne,[1,3] Val Santo,[1] Niamh Gale,[2] Jonathan T Evans,[4] Jonathan P Evans,[4] Jenny Lowe,[1] GJ Melendez-Torres[1] and Edward CF Wilson[1]

[1]Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter, UK
[2]Department of Health and Care Professions, University of Exeter Medical School, Exeter, UK
[3]NIHR Applied Research Collaboration South-West Peninsula (PenARC), University of Exeter Medical School, Exeter, UK
[4]Department of Public Health and Sport Sciences, University of Exeter Medical School, Exeter, UK

[*]**Corresponding author** c.farmer@exeter.ac.uk

Published May 2026
DOI: 10.3310/GJCF0822

## Scientific summary

Artificial intelligence software to help detect fractures on X-rays in urgent care: An Early Value Assessment

Health Technology Assessment 2026; Vol. 30: No. 33
DOI: 10.3310/GJCF0822

NIHR Journals Library www.journalslibrary.nihr.ac.uk

# Scientific summary

## Background and objectives

Plain film radiography or X-ray is the most common medical imaging approach used to detect fractures in urgent care settings, including the emergency department, urgent treatment centre, and minor injuries units. X-rays are typically read in urgent care settings by healthcare professionals who are not radiology specialists or are inexperienced at interpreting X-rays, which may increase the likelihood of errors in decision-making, particularly in busy healthcare centres when staff are under significant pressure. Reduced staff numbers, such as outside normal working hours, may also influence the risk of errors in diagnosis. A definitive diagnosis of the injury will be produced by a consultant radiologist or reporting radiographer, although there may be a delay before this is available, meaning that this may arrive after people have been treated and/or discharged from urgent care. Delays vary across settings, and may be longer for children due to availability of specialist in paediatrics.

Artificial intelligence (AI) algorithms have been developed to support clinicians in diagnosing fractures, with the intention to improve the diagnostic accuracy of clinicians reviewing X-rays. Improving diagnostic accuracy means reducing the number of missed fractures (false-negative diagnoses) and the number of people treated for a fracture who do not have one (false-positive diagnoses).

The purpose of this rapid early value assessment (EVA) was to identify the existing evidence base for the technology and to assess whether there was a prima facie case for the technology to represent a value-for-money investment for people in the NHS. A rapid evidence review was conducted followed by 'light touch' early economic modelling to explore whether a plausible case could be made for cost-effectiveness at the prices charged by the companies. The approach was not suitable for a definitive assessment of the cost-effectiveness of one AI-algorithm against another, but rather to inform whether or not the NHS should consider adopting the technology while further evidence is collected.

## Evidence review: clinical and service use outcomes

A broad evidence review was conducted to identify the existing evidence base for clinical, and service outcomes associated with the technology. Searches were conducted in June and July 2024. The review identified 16 studies that evaluated the diagnostic accuracy of the technology as an aid to diagnosing fractures (i.e. when used to assist reading clinicians, and not as a stand-alone diagnostic tool). Evidence was available for four of the eligible technologies: BoneView (Gleamer, Paris, France), Rayvolve® (AZmed, Paris, France), RBfracture™ (Radiobotics, Copenhagen, Denmark) and TechCare Alert™ (Milvue, Paris, France). None of the included studies were conducted in the UK, and all were associated with limitations, including risks of bias and uncertain generalisability to the NHS. Few studies evaluated the technologies when used by clinicians who would typically provide the initial diagnosis in urgent care settings, with most evaluating readers who were clinicians specialising in radiology or among a varied of group of clinicians with varying levels of reading experience. Data were reported for a general sample of people with types of fractures that were eligible for consideration by the technologies. Subgroup data were also available for pre-specified fracture subgroups, for children and for 'less obvious' fractures. None of the included studies reported clinical outcomes associated with use of the technology and, aside from the reading time per scan, no service outcomes were reported. As compared to the list of outcomes specified in the NICE scope (National Institute for Health and Care Excellence. *Artificial Intelligence Software to Help Detect Fractures on X-Rays in Urgent Care: Final Scope*. 2024. URL: www.nice.org.uk/guidance/hte20/documents/final-scope) for this assessment, there was a major gap in the evidence base.

There was unexplained heterogeneity in the results reported across studies. To aid with interpretation of the results, where multiple results were reported by studies according to staff experience, the EAG grouped the data according to reader experience (as described by the included publications). This approach was considered imperfect and did not completely resolve the heterogeneity in the data. The EAG considered that the quality limitations with the evidence base may have contributed to the heterogeneity, and it may also be possible that additional, unidentified effect

modifiers were present across the evidence base. The EAG conducted a feasibility assessment to determine if meta-analysis of the data was possible, but where sufficient numbers of studies were available, these were considered too heterogeneous to pool. Notably, clinical advice to the EAG was that the diagnostic accuracy of unassisted readings in the included studies appeared lower than was expected by the EAG's clinical advisors, which adds uncertainty to the generalisability of the evidence base.

Overall, given the limitations in the evidence base and the heterogeneity in the study results, the EAG did not consider that the evidence base was suitable to determine reliable estimates of the diagnostic accuracy of the technologies for assisting in the diagnosis of fractures. However, based on evaluation of the evidence base as a whole and specifically in studies reporting outcomes for clinicians based in emergency care settings, the EAG identified a general trend for the technology to result in an improvement in sensitivity (i.e. a reduction in missed fractures) with no or minimal improvement in specificity (i.e. no change in false-positive diagnoses). Use of the technology was still associated with varying levels of missed fractures, however, particularly in 'less obvious' fractures, where the technology was considered to be of most potential value. Further evidence, is needed to determine robust evidence for any improvement in sensitivity, and to establish whether the additional fractures identified would result in meaningful clinical benefits for patients.

## Economic evidence and analysis

The evidence review identified no available economic evaluations for the technology. The EAG constructed a simple decision model to establish whether there was a prima facie case for AI-assisted diagnosis to represent a value-for-money investment for NHS patients. As the long-term costs and outcomes for different fractures were substantially different, the EAG divided the analysis into three decision problems, concerning ankle and foot, wrist and hand, and hip fractures. These were chosen on the basis of availability of data and their different downstream costs and consequences. An overall estimate of the costs and consequences for a typical urgent care setting was estimated based on case mix for the three fracture types, with extension to all fractures considered in scenario analysis.

The decision model was a decision tree incorporating prevalence, sensitivity, specificity and cost per scan for each of the five diagnostic strategies (four AI algorithms and unassisted diagnosis). Estimates of the long-term costs and quality-adjusted life-years (QALYs) accrued from a true and false positive and negative for each fracture type were extracted from the literature. The tree was rolled back to estimate the expected cost and QALYs accrued from each diagnostic strategy. Scenario analyses explored key uncertainties.

Overall, most of the AI-assisted algorithms were associated with a positive incremental net health benefit (INHB) at willingness-to-pay thresholds of £20,000 and £30,000 per QALY gained. Due to data limitations, the EAG did not consider the analysis appropriate to compare technologies against each other, although this would be required in a more thorough analysis in future to ensure that the diagnostic accuracy of each algorithm was matched to its price.

The results were mostly robust to the scenario analyses considered with the exception of diagnostic accuracy, where none of the algorithms were associated with a positive INHB (compared with unassisted diagnosis) in the pessimistic scenario.

## Key points for decision-makers

- While a reasonable number of studies have evaluated the diagnostic accuracy of the technology as an aid to the identification of fractures, very few studies are specific to emergency care settings and all were associated with significant limitations due to risk of bias or uncertain generalisability. The existing evidence base was not sufficient to determine an approximate estimate of the diagnostic accuracy of the technology for its intended use.
- Across the evidence base as a whole, there was a trend for the technology to reduce missed fractures without any change to false-positive diagnoses. However, based on the existing evidence, the EAG was unclear whether the

additional fractures identified would translate into meaningful benefits for patients. While there are some fractures that, if missed, can result in significant harm to patients, stakeholders to this assessment also considered it plausible that the technology would improve diagnosis of more subtle fractures that may not require a change in management.
- The evidence suggested that use of the technology would not eradicate the risk of missed fractures, meaning it was likely that health services would need to continue to take precautions to avoid the risk of a missed fracture in clinical practice (e.g. precautionary treatment of high-risk suspected fractures). This means that use of the technology had an unclear impact on healthcare resource use.
- On average, based on a simple decision model for this EVA, most of the AI algorithms considered represent a positive INHB compared with unassisted diagnosis at NICE's conventional threshold of £20,000–£30,000 per QALY. The evidence base was not sufficient to compare different algorithms against one another.
- Economic results were mostly robust to scenario analyses testing different methods for implementing the technology; however, the results were sensitive to variations in the diagnostic accuracy of the technology.

## Study registration

This study is registered as PROSPERO CRD42024574393.

## Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136024) and is published in full in *Health Technology Assessment*; Vol. 30, No. 33. See the NIHR Funding and Awards website for further award information.

# Health Technology Assessment

NICE TAR and DAR

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

**Criteria for inclusion in the *Health Technology Assessment* journal**
Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

## HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

## This article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR136024. The protocol was agreed in June 2024. The draft manuscript began editorial review in December 2024 and was accepted for publication in September 2025. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.



Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).