

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Early value assessment guidance consultation document

Artificial intelligence technologies to help detect fractures on X-rays in urgent care: early value assessment

The National Institute for Health and Care Excellence (NICE) is producing guidance on using artificial intelligence technologies to help detect fractures on X-rays in urgent care in the NHS in England. The diagnostics advisory committee has considered the evidence and the views of clinical and patient experts. This topic is an early value assessment. Early value assessment guidance recommendations are conditional while more evidence is collected on the technology to address uncertainty in the evidence base. Once further evidence is collected, this guidance will be reviewed to make a decision on the routine adoption of the technology.

This document has been prepared for public consultation. It summarises the evidence and views that have been considered, and sets out the recommendations made by the committee. NICE invites comments from registered stakeholders, healthcare professionals and the public. This document should be read along with the [evidence](#) (the external assessment report).

The advisory committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?
- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the recommendations sound, and a suitable basis for guidance to the NHS?

Equality issues

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the recommendations may need changing to meet these aims. In particular, please tell us if the recommendations:

- could have a different effect on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology

- could have any adverse effect on people with a particular disability or disabilities.

Please provide any relevant information or data you have about such effects and how they could be avoided or reduced.

Note that this document is not NICE's final guidance on artificial intelligence technologies to help detect fractures on X-rays in urgent care. The recommendations in section 1 may change after consultation.

After consultation, the committee will meet again to consider the evidence, this document and comments from the consultation. After considering the comments, the committee will prepare its final recommendations, which will be the basis for NICE's guidance on the use of the technology in the NHS in England.

For further details, see [NICE health technology evaluations: the manual](#).

Key dates:

Closing date for comments: 5 November 2024

1 Recommendations

Can be used while more evidence is generated

1.1 Four artificial intelligence (AI) technologies can be used during the evidence generation period as an option to help healthcare professionals detect fractures on X-rays in urgent care. The technologies are:

- for people of any age:
 - TechCare Alert
- for children and young people 2 years and over, and adults:
 - BoneView
 - RBfracture
- for adults:
 - Rayvolve.

They can only be used if the evidence outlined in the evidence generation plan is being generated.

1.2 The companies must confirm that agreements are in place to generate the evidence (as outlined in NICE's evidence generation plan). They should contact NICE annually to confirm that evidence is being generated and analysed as planned. NICE may withdraw the guidance if these conditions are not met.

1.3 At the end of the evidence generation period (2 years), the companies should submit the evidence to NICE in a form that can be used for decision making. NICE will review the evidence and assess if the technologies can be routinely adopted in the NHS.

Can only be used in research

1.4 More research is needed on qMSK to help healthcare professionals detect fractures on X-rays of adults in urgent care before it can be used in the NHS.

- 1.5 Access to qMSK should be through company, research or non-core NHS funding, and clinical or financial risks should be appropriately managed.

What evidence generation and research is needed

- 1.6 Evidence generation and more research is needed on:
- the diagnostic accuracy of fracture detection in urgent care by healthcare professionals with and without the help of AI technologies
 - costs and clinical outcomes associated with different fracture types and missed fractures
 - fracture clinic referral rates with and without the help of AI technologies
 - any clinically significant changes in treatment decisions for fractures detected with and without the help of AI technologies
 - AI software failure rates and reasons for failure
 - detection of or failure to detect clinically significant non-fracture-related conditions by healthcare professionals with and without the help of AI technologies
 - the diagnostic accuracy of AI technologies to help healthcare professionals detect fractures in different populations
 - implementation costs of AI technologies in different urgent care centres.

Potential benefits of use in the NHS with evidence generation

- **Clinical benefit:** Clinical evidence suggests that the AI technologies may improve fracture detection on X-rays in urgent care without increasing the risk of incorrect diagnoses. This could help reduce the number of fractures that are missed in urgent care, which would reduce the risk of further injury or harm to people during the time between the initial interpretation and treatment decision in urgent care and the radiology review.
- **System benefit:** AI technologies may help reduce variation in standard care by providing a consistent baseline for X-ray interpretation unaffected by differences in staff experience or resources between centres. AI technologies

would also be unaffected by factors such as staff fatigue, distractions, or working outside normal hours.

- **Resources:** Reducing the number of fractures that are missed at initial interpretation would also reduce the number of people that reattend urgent care after discharge or are recalled to hospital after radiology review. Early results from the exploratory economic modelling show that the AI technologies could be cost effective.
- **Equality:** AI technologies have the potential to reduce geographical inequalities in X-ray interpretation and fracture detection, because they may improve fracture detection in smaller centres with fewer and less-experienced staff.

Managing the risk of use in the NHS with evidence generation

- **Clinical risk:** Using AI technologies to help detect fractures on X-rays in urgent care is considered to have a low clinical risk because they are used in addition to standard care in which healthcare professionals make treatment decisions. Additionally, AI technologies do not replace the definitive radiology review. The available evidence suggests that the AI technologies may improve the accuracy of fracture detection.
- **Implementation guidance:** Clear local protocols will need to be in place when using AI technologies. This is to ensure that healthcare professionals are confident about what action to take when there is disagreement between the healthcare professional and AI technology.
- **Costs:** There is uncertainty around the cost of some software and the true cost of implementation and ongoing post-market surveillance. Costs were estimated at £1 per scan in the exploratory economic modelling. Centres implementing AI to help fracture detection should ensure the cost per scan is similar to the estimated cost. This guidance will be reviewed within 3 years and the recommendations may change. Centres should take this into account when negotiating contract durations and licence costs.

- **Impact on workforce:** If using AI technologies to help fracture detection becomes more widespread and part of the standard diagnostic pathway, there is a risk of over-reliance on the technologies. This could potentially lead to deskilling of the healthcare professionals who interpret the X-rays. It may also reduce the level of scrutiny for non-fracture-related conditions. This risk could be mitigated if healthcare professionals interpret X-rays before viewing the AI results.
- **Resources:** There is a low risk that the AI technologies may increase fracture clinic referrals and requests for further imaging such as CT or MRI. These should be monitored during evidence generation to inform local fracture detection protocols.
- **Limitations of AI for subgroups:** The AI technologies may not be suitable for use in certain groups, for example, children and young people, people with conditions that affect bone health, people with myeloma or people with metastatic bone disease. Centres should ensure that the AI technologies are used within their indications and any limitations are acknowledged and clearly explained to patients.
- **Equality:** There is a risk that the AI technologies may have reduced diagnostic accuracy in different populations, such as people from ethnic minority backgrounds or people with low socioeconomic status. Healthcare professionals should take this into account when interpreting X-rays of people in these groups.

2 The technologies

Clinical need and practice

- 2.1 Fracture assessment and diagnosis in urgent care typically involves triage in which a nurse, advanced clinical practitioner or doctor will do an initial assessment before requesting imaging. X-rays taken by a diagnostic radiographer are usually the first-line imaging approach for non-complex

fractures. Multiple surgical and non-surgical treatment options are available depending on the type of fracture.

2.2 [NICE's guideline on non-complex fractures](#) recommends that a radiologist, radiographer or other trained reporter should review X-rays and provide a definitive report before the injured person is discharged (hot reporting). Clinical experts explained that in practice this is not always possible and reporting delays can occur ranging from days to weeks.

2.3 Missed fractures are reported to be the most common diagnostic error in the emergency department ([Hussain et al. 2019](#)). Missed or delayed diagnosis of fractures on radiographs is reported to occur in around 3% to 10% of cases ([Kuo et al. 2022](#)).

2.4 Artificial intelligence (AI) technologies that can help healthcare professionals detect fractures on X-ray images could improve the accuracy of fracture diagnoses in urgent care. This could help reduce:

- the number of fractures that are missed before a radiologist or reporting radiographer reviews the X-rays
- the number of people being recalled to hospital after radiology review
- the risk of further injury or harm to people during the time between the initial interpretation and treatment decision in urgent care and the definitive radiology report
- the number of unnecessary referrals to fracture clinics.

AI technologies may also improve consistency of X-ray interpretation when the ability of healthcare professionals to interpret X-rays may be reduced, for example, when they are tired, distracted or working outside normal hours.

The interventions

2.5 The technologies included in this early value assessment are standalone software that use AI-derived algorithms to analyse X-ray images to detect

fractures. They are intended to be used as decision aids for healthcare professionals interpreting the X-ray image. Some companies provide the software directly, whereas others provide it through multivendor platforms. The technologies use X-ray radiographs in digital imaging and communications in medicine (DICOM) format which are stored on the hospital's picture archiving and communications system (PACS). Images are then interpreted using proprietary AI-derived algorithms. The technologies included in this assessment are shown in table 1.

Table 1 Interventions

AI technology (manufacturer)	CE marking	Regions covered	Population	Other pathologies detected
BoneView (Gleamer)	Class IIa	Appendicular skeleton, ribs and thoracic-lumbar spine	2 years and over	Dislocations, effusions and bone lesions
qMSK (Qure.ai)	Class IIb	Appendicular skeleton and ribs	Adults	–
Rayvolve (AZmed)	Class IIa	Appendicular skeleton and ribs	Adults	Dislocations, joint effusions and chest pathologies (pneumothoraces, cardiomegaly, pleural effusions, pulmonary oedema, consolidation, nodules)
RBfracture (Radiobotics)	Class IIa	Appendicular skeleton and ribs	2 years and over	Effusion of the knee and elbow, and lipohaemarthrosis of the knee
TechCare Alert (Milvue)	Class IIa	Appendicular skeleton and ribs	No age limit	Dislocations, elbow joint effusion, pleural effusion, pulmonary opacity, pulmonary nodules and pneumothorax

The comparator

2.6 The comparator is standard care for fracture assessment in which the urgent care healthcare professional interprets the X-ray radiograph without AI assistance.

- 2.7 The reference standard is based on the consultant radiologist or reporting radiographer interpretation and report.

3 Committee discussion

The [diagnostics advisory committee](#) considered evidence on BoneView, qMSK, Rayvolve, RBfracture and TechCare Alert from several sources, including an external assessment report and an overview of that report. Full details are in the [project documents for this guidance](#).

Patient and carer considerations

- 3.1 People may be anxious about the certainty of their diagnosis and the risk of being discharged with a missed fracture with or without the use of artificial intelligence (AI). In addition to the pain and potential clinical complications associated with a missed fracture, there are also practical stresses such as taking time off work or taking children out of school to reattend urgent care. Patient experts explained that if AI technologies could help improve diagnostic accuracy and reduce the risk of a misdiagnosis, then this would be a welcome benefit for patients.
- 3.2 Human interaction with a healthcare professional is an important factor for patients to ensure they are informed and reassured about their diagnosis. Patient experts explained that people may have different attitudes towards AI technologies and some people may distrust their use because they could be perceived as replacing human involvement. Clinical experts stated that, in practice, AI technologies would be used as a decision aid to assist healthcare professional fracture detection in urgent care (see [section 2.5](#)). They highlighted that the [ionising radiation \(medical exposure\) regulations \(IR\(ME\)R\)](#) state that clinical evaluation of X-rays requires a trained person. Therefore, AI technologies for fracture detection on X-rays cannot be used without human interpretation and so the level of human interaction would not change. The committee noted that people having X-rays for suspected fractures should be informed that AI software is being used, and the role of healthcare professionals and AI software in

interpreting the X-rays should be explained. Patient and clinical experts also highlighted the importance of educating patients and healthcare professionals to understand the benefits and limitations of the software. The importance of shared decision making after AI-assisted diagnosis was also highlighted.

Clinical effectiveness

Evidence base

3.3 There were 16 studies that met the inclusion criteria for the clinical-effectiveness review. Most studies evaluated BoneView (8 studies) and RBfracture (5 studies), 1 study each for Rayvolve and TechCare Alert, and 1 study covering BoneView, Rayvolve and TechCare Alert together. No studies were identified for qMSK that compared interpretation of X-rays by healthcare professionals with or without use of the technology.

Diagnostic accuracy

3.4 Diagnostic accuracy studies typically found improved sensitivity of fracture detection, without reduced specificity, by healthcare professionals assisted by AI software compared with unassisted interpretation. For example, one of the key studies for BoneView ([Duron et al. 2021](#)), which reported estimates for emergency physicians interpreting mixed fracture types, indicated that sensitivity increased from 61% (unassisted) to 74% (assisted). Similar increases in sensitivity were seen for the other software when used by emergency care staff for mixed fractures. [Bachmann et al. \(2024\)](#) reported an increase in sensitivity from 74% unassisted to 83% when assisted by RBfracture, and [Fu et al. \(2024\)](#) reported an increase from 79% to 94% for Rayvolve. The Suite 2020 study reported a much smaller increase in sensitivity (92% to 95%) when using TechCare Alert, but the readers in this study were radiologists rather than emergency physicians. No key studies reported a decrease in specificity when using AI to assist fracture detection. The committee concluded that the available

evidence suggested that AI technologies have the potential to improve the diagnostic accuracy of healthcare professional fracture detection. The committee noted that there was some uncertainty and wide variance in the estimates of sensitivity and specificity of the AI technologies in the studies because of variation in study designs (see [section 3.3](#)).

- 3.5 Clinical experts explained that the diagnostic accuracy of unassisted fracture detection reported in the studies was lower than would be expected in clinical practice (see [section 2.3](#)). The committee noted that this could overestimate the diagnostic accuracy of the AI technologies and therefore their clinical effectiveness. A clinical expert said that in most reader studies there is usually some heterogeneity in the diagnostic accuracy of unassisted healthcare professional review. So it is unclear what should be considered a normal baseline estimate of unassisted diagnostic accuracy. The committee concluded that further evidence on the diagnostic accuracy of AI-assisted and unassisted fracture detection should be collected as part of a real-world evidence generation plan.
- 3.6 The committee concluded that the included studies were not entirely applicable to using AI technologies to help healthcare professionals detect fractures on X-rays in a UK urgent care setting. Most were retrospective, case-control studies. Clinical experts explained that retrospective studies may not represent the diagnostic accuracy of healthcare professional review in clinical practice. This is because in the studies the readers typically interpret X-rays in isolation rather than alongside the patient or patient history and case notes, as they would in clinical practice. The committee noted that none of the studies were done in a UK urgent care setting. The healthcare workers who interpreted the X-rays in the studies differed from those who would typically interpret X-rays in UK urgent care settings (for example, examining the accuracy of radiologists with or without AI assistance, rather than emergency department healthcare professionals). So, it is uncertain how the technologies would perform in

this setting. A clinical expert also highlighted that some of the studies included the AI software result as part of the reference standard.

Children and young people

3.7 The committee considered the limited evidence base for children and young people. It concluded that, similar to adults, AI technologies have the potential to improve the diagnostic accuracy of healthcare professional fracture detection in this subgroup. Only 2 of the key studies identified by the external assessment group (EAG) reported diagnostic accuracy data for children and young people. These studies indicated the potential of AI technologies to improve sensitivity without reducing specificity compared with unassisted fracture diagnosis. A study by [Nguyen et al. \(2022\)](#) evaluated BoneView and showed an increase in sensitivity from 73.2% (unassisted) to 82.7% (assisted) in a mixed reader group (including radiologists) interpreting mixed fracture types. [Bachmann et al. \(2024\)](#) evaluated RBfracture and showed that sensitivity increased from 78% to 89%, also in a mixed reader group interpreting mixed fracture types. A clinical expert highlighted that in one study the unassisted diagnostic accuracy for children was higher than that reported for adults. They said that this was unusual and suggested that it may indicate a bias in the selection of cases or in the staff involved in interpreting the X-rays, leading to uncertainty in the results. A clinical expert explained that there are important differences in X-ray interpretation and fracture detection between children and young people and adults. There is wide variance in how children's bones can look on X-ray images and this can complicate fracture detection. They also highlighted that there is limited evidence in children younger than 2 years or when there is a suspicion that the injuries are a result of abuse.

System impact

3.8 The committee concluded that although more system-level impact data was needed, the risk of AI technologies negatively affecting the healthcare

system was low. This is because the evidence suggests it is unlikely that AI use would lead to an increase in the rate of false referrals (see [section 3.4](#)). The only evidence on system-level impact was on X-ray reading times with and without AI assistance, which was available for 3 of the technologies (BoneView, Rayvolve and RBfracture). The committee noted that using the AI technologies resulted in changes of only a few seconds (reductions and increases) compared with unassisted reads. Clinical experts explained that reading time estimates from the studies may have limited relevance to clinical practice. This is because in the studies, healthcare professionals interpreting the X-rays may only be looking at the X-ray in isolation (see [section 3.3](#)). In clinical practice they would take time to consider the patient history and may do a more detailed review of the suspected fracture site. The committee felt that other system-level effects, such as fracture clinic referral rates with and without AI assistance, would have more impact, and data on this could be collected as part of the evidence generation plan.

Cost effectiveness

Model structure

- 3.9 The EAG constructed an exploratory economic model to explore the potential cost effectiveness of AI-assisted fracture detection compared with unassisted diagnosis in an urgent care setting. The model consisted of 3 separate sub-models for the fracture sites that were considered to gain the greatest potential benefit from AI-assisted diagnosis, because the costs and clinical outcomes of these fractures differed substantially. These fracture sites were wrist and hand, ankle and foot, and hip. Each model comprised a decision tree incorporating the prevalence, sensitivity and specificity and cost per diagnosis for AI-assisted and unassisted fracture detection.
- 3.10 The committee felt that the EAG's exploratory economic model structure and assumptions likely underestimated the impact of false-negative

diagnoses. It noted that people with a false-negative diagnosis were assumed to reattend urgent care 2 to 4 weeks after their initial presentation, with no further disutilities assumed to occur in that time. Clinical experts explained that this was an oversimplification and did not reflect clinical practice. This is because a delay in treatment could result in changes to the injury, which may change further management. For example, a 2-week delay to treating a wrist fracture may result in callus formation which would then require different surgery, or a missed ankle fracture may require surgery in addition to a brace or cast. There is also a risk of further injury if people are discharged with an undiagnosed fracture, and they may re-present in other settings such as a GP surgery or physiotherapy. The committee noted that costs associated with further management because of delayed treatment were not captured in the economic model beyond the cost of an additional A&E appointment. These assumptions would therefore underestimate the benefit of improving fracture detection using AI technologies in the model results.

- 3.11 The committee also concluded that the model overestimated the impact of false-positive diagnoses of hip fracture. The model assumed that false-positive diagnoses of hip fracture would result in unnecessary surgery. Clinical experts said that this was highly unlikely because further imaging such as CT or MRI would usually be requested if there was any uncertainty in the diagnosis. So the costs for this group are likely overestimated in the model.

Costs and clinical outcomes of fractures

- 3.12 The committee noted that because all the evidence used in the model was from retrospective studies, there was no data on the costs and clinical outcomes associated with misdiagnosed fractures. The EAG explained that because of this lack of evidence the model assumed the only consequence of a missed fracture was pain. The committee concluded that the costs and clinical outcomes associated with missed fractures

were uncertain but likely underestimated (see [section 3.10](#)). Further data on the costs and outcomes associated with fractures in urgent care could be collected as part of the evidence generation plan.

Diagnostic accuracy inputs

3.13 The baseline sensitivity and specificity estimates were taken either from [Bousson et al. \(2023\)](#) for BoneView, Rayvolve and TechCare Alert, or from [Bachmann et al. \(2024\)](#) for RBfracture and unassisted readers. The committee concluded that the model inputs for diagnostic accuracy were uncertain because of the study designs (see [sections 3.3](#) and [3.5](#)), which could have a large impact on the potential cost effectiveness of AI-assisted fracture detection. Bousson et al. was a retrospective study that only included radiologist readers and the reference standard included the AI results. The study by Bachmann et al. was also retrospective and used a case-control design. The committee noted that the accuracy of unassisted readers was lower than expected, so the difference in accuracy between AI-assisted and unassisted fracture detection may have been overestimated (see section 3.5). The committee stated that further evidence was needed on the diagnostic accuracy of AI-assisted and unassisted healthcare professional fracture detection in urgent care.

Cost inputs

3.14 The committee concluded that the true cost of implementing and using AI technologies for fracture detection was uncertain and further evidence was needed on the cost of implementation in different urgent care centres. Some companies did not submit costs for the assessment, so the EAG used a notional cost of £1 per scan in the base case. A clinical expert stated that set-up costs relating to NHS IT time and fees from the picture archiving and communications system (PACS) providers to ensure the new technology works correctly were not included in the economic model. These costs were variable depending on the centre but experts estimated they could be between £1,200 and £120,000. A clinical expert also

explained that there are also ongoing cost and resource requirements associated with post-market surveillance. While this should be supported by companies, it still relies on NHS staff to collect this data. A clinical expert explained that from 2025, there will be additional financial support from the NHS which may help relieve some of the cost impact of implementing AI technologies for fracture detection.

Plausibility of cost effectiveness

- 3.15 The committee said that it is plausible that the AI technologies could be cost effective if implemented in the NHS. This is because the available evidence suggests that they have the potential to improve sensitivity without reducing specificity compared with unassisted fracture diagnosis. In the base case, the committee noted that, overall, BoneView, RBfracture and TechCare Alert were associated with a positive incremental net health benefit compared with unassisted diagnosis at a threshold of £20,000 per quality-adjusted life year (QALY) gained. But in most cases, the 95% confidence intervals crossed zero, both for all separate fracture types and when considered together.
- 3.16 In the EAG's base case, Rayvolve had a negative incremental net health benefit. The committee noted that this was likely because it was modelled as having a lower specificity (67% to 75%) than unassisted fracture detection (87%), resulting in an increase in false-positive results and their associated costs. The diagnostic accuracy estimates used in the base case for Rayvolve were from the study by [Bousson et al. \(2023\)](#). The company (AZmed) stated that diagnostic accuracy estimates for Rayvolve from this study were unreliable because it used an outdated version of the algorithm. The committee considered the diagnostic accuracy estimates from the other key study that used Rayvolve ([Fu et al. 2024](#)) and noted that they showed improved sensitivity and little change in specificity with the AI compared with unassisted reads. The committee concluded that because of the uncertainty in the diagnostic accuracy estimates, it was

reasonable to assume that Rayvolve also had the potential to be cost effective (see [section 3.13](#)).

- 3.17 The committee recalled the uncertainty around the diagnostic accuracy estimates (see [section 3.4](#)). It considered that if the data was significantly overestimating the performance of the technologies, they would be less likely to be cost effective. In the scenario analyses, only the scenarios that changed the diagnostic accuracy significantly affected the model results.
- 3.18 The committee noted that for all fracture sites there was a minimal difference in QALYs between AI-assisted and unassisted diagnosis. The committee said that this is likely because the model underestimates the utility impact of a missed fracture (see [section 3.10](#)) and so may also underestimate the cost effectiveness of the AI technologies.
- 3.19 The committee also recalled the uncertainty in the costs because some companies did not provide a cost per scan, and the variability in estimates of set-up and implementation costs. However, in scenario analyses, the model results were not sensitive to small increases or decreases (less than £3) in the cost per scan. The EAG did a further scenario analysis which included additional installation and set-up costs. This applied a notional one-off set-up cost of £50,000 and assumed a 5-year lifespan of the software. The committee noted that the model results (see [section 3.14](#)) were not significantly affected by the £50,000 additional set-up cost over either a 5-year or 1-year period.

Risks

- 3.20 The committee concluded that although there were risks associated with the implementation of the AI technologies, they were relatively low or could be mitigated during the evidence generation period.
- 3.21 The committee considered that the clinical risk of implementing AI technologies to help detect fractures in urgent care was low because they are used in addition to standard care, in which treatment decisions are

made by healthcare professionals. Also, the definitive X-ray reports are usually made by a radiologist or reporting radiographer, which AI would not replace. So there are safety net systems in place to identify any potential fractures that may have been missed by the AI. Clinical experts explained that there would need to be clear local protocols in place when using AI technologies to ensure that healthcare professionals are clear about what action to take when there is disagreement between the healthcare professional and AI.

- 3.22 The committee said that there was some risk associated with the cost of the AI technologies. This is because 2 companies did not provide pricing information, and there was uncertainty around the true cost of implementation and ongoing post-market surveillance. It noted that small changes to the cost per scan did not have a large effect on model results (see [section 3.19](#)). It said that when centres were implementing the technologies during the evidence generation period, they should consider the notional cost per scan used in the exploratory economic modelling.
- 3.23 Patient and clinical experts highlighted concerns that implementation of AI could lead to healthcare professionals becoming over-reliant on the technologies, and it may also reduce the level of scrutiny for non-fracture-related conditions that can be detected on X-ray. The committee noted that this could potentially be mitigated if healthcare professionals interpreted X-rays unassisted before viewing the AI results.
- 3.24 The committee considered the impact of AI on resource use. It noted that there was a low risk that it may lead to an increase in fracture clinic referrals and requests for further imaging such as CT or MRI. This is because the evidence suggests it is unlikely that AI use would lead to an increase in the rate of false referrals.

Research considerations

3.25 The committee considered that, because the AI technologies are trained on different data sets and use different algorithms, it is likely that they all perform differently. Because there was very little evidence on how the AI technologies differed in terms of diagnostic accuracy (see [section 3.4](#)), it said that comparative, head-to-head studies of the software would be useful to help understand differences in their diagnostic performance.

Equality considerations

3.26 Some of the technologies are not approved for use in children and young people and it is unclear if they are appropriate for use in other subgroups such as older people, and people with conditions that affect bone health. The committee noted that there was limited evidence on the use of AI technologies to help detect fractures in these subgroups. The committee said that the AI technologies should be used within their indications and clinicians should ensure that a technology is appropriate to use for the specific person they are assessing. Failure to do this could result in false reassurances and so increase the risk of a fracture being missed.

3.27 Conditions that can affect bone health may include:

- autoimmune and erosive arthropathies
- fibrous dysplasia
- myeloma
- osteoarthritis
- osteonecrosis
- osteoporosis
- osteogenesis imperfecta
- Paget's disease
- cancer with metastatic bone disease.

- 3.28 Clinical experts stated that the data sets used for training the AI technologies may not be representative of the local patient population. People from low socioeconomic status and or minority groups may not be well represented in these sets and so there is a risk that the diagnostic accuracy of the AI technologies may be reduced in these groups. The committee noted that this was a potential limitation of the technologies and healthcare professionals should take this into account when interpreting X-rays of people in these groups.
- 3.29 A patient expert highlighted the potential for indirect discrimination because of geographical availability and access. They raised concerns around whether the AI technologies would be deployed in smaller minor injuries units in rural areas as well as larger urgent treatment centres and emergency departments in urban areas. However, the committee also considered that AI software may help reduce variation in standard care by providing a consistent baseline for X-ray interpretation which is not affected by differences in staff experience or resource between centres.

Evidence gaps

- 3.30 Evidence gaps identified related to the intervention, the main outcomes including costs, and the population. The committee concluded that there was enough evidence on 4 of the AI technologies to demonstrate their potential benefit when used to help healthcare professionals detect fractures on X-rays in urgent care. It also concluded that the clinical risk of implementation was low (see [sections 3.20 to 3.24](#)). Important evidence gaps for all the AI technologies are:
- **Interventions:** the available evidence suggested that AI technologies have the potential to improve the diagnostic accuracy of healthcare professional fracture detection, but this was uncertain. Also the accuracy of unassisted fracture detection reported in the studies was lower than would be expected in clinical practice. The committee concluded that further evidence on the diagnostic accuracy of AI-

assisted and unassisted healthcare professional fracture detection in urgent care centres was needed. Further evidence is also needed on AI software failure rates and reasons for failure.

- **Outcomes:** there was no evidence on system-level outcomes. The committee considered that the outcome likely to have the largest system-level impact would be fracture clinic referral rates. It highlighted the need for further evidence on fracture clinic referrals with and without AI assistance. To better understand the clinical effectiveness of AI technologies for fracture detection, clinical experts stated that further evidence was needed on clinically significant changes in treatment decisions for fractures detected using AI software. They also stated that evidence was needed on the detection or failure to detect clinically significant non-fracture-related conditions by AI-assisted and unassisted healthcare professionals.
- **Costs:** because the evidence was from retrospective studies, there was no data on the costs and clinical outcomes associated with different fracture types and missed fractures. The true cost of implementing and using AI technologies for fracture detection is uncertain. These costs are important for understanding the financial investment that is needed and also the feasibility and sustainability of integrating AI technologies into routine healthcare. So further evidence is needed on the cost of implementation and use of AI technologies in different urgent care centres.
- **Population:** the committee noted that there was limited evidence on the use of AI technologies to assist with fracture detection in the population subgroups identified in the scope. It highlighted the need for evidence generation on the diagnostic accuracy of AI-assisted healthcare professional fracture detection in different subgroups such as by age, sex, ethnicity, socioeconomic status, and conditions that affect bone health (see [section 3.27](#)).

Ongoing studies

3.31 The committee concluded that although there are several ongoing studies that may provide further evidence on the clinical effectiveness of AI technologies in fracture detection, they will not address all the evidence gaps identified (see [section 3.30](#)). The committee considered 2 ongoing studies evaluating BoneView. [FRACT-AI \(Clinicaltrials.gov, NCT06130397\)](#) is a retrospective multiple-reader, multiple-case study, due to complete in December 2024. A clinical expert explained that an advantage of FRACT-AI is that it will include a range of readers from urgent care healthcare workers in a UK setting. The ['Testing an artificial intelligence tool for childhood fracture detection on X-rays study', ISRCTN12921105](#) is a retrospective, multi-centre, multi-reader observational cohort study evaluating BoneView in paediatric fractures. Because both studies are retrospective, the committee stated that they will be unable to address evidence gaps relating to the post-diagnosis impact of AI-assisted fracture detection. The committee also noted that there were 5 NHS-based real-world data collection studies using RBfracture. Primary outcome measures that will be reported in these studies include increases in productivity through time saving, rates of missed fractures, number of CT scans, inappropriate referrals to fracture clinics, and equivocal findings. These studies are due to complete between late 2024 and late 2025.

4 Committee members and NICE project team

Committee members

This topic was considered by the [diagnostics advisory committee](#), which is a standing advisory committee of NICE.

Committee members are asked to declare any interests in the technologies to be evaluated. If it is considered there is a conflict of interest, the member is excluded from participating further in that evaluation.

The [minutes of each committee meeting](#), which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Additional specialist committee members took part in the discussions for this topic:

Specialist committee members

Amanda Isaac

Consultant musculoskeletal radiologist and honorary associate professor, Guy's and St Thomas' NHS Foundation Trust and King's College London

Azizul Haque

Consultant trauma and orthopaedic surgeon, University Hospitals of Leicester NHS Trust

Claire Hart

Diagnostic radiographer, Hwyl Dda University Health Board and Swansea Bay University Health Board

David Lowe

Consultant in emergency medicine, Queen Elizabeth University Hospital, and honorary professor and clinical director of innovation, University of Glasgow

Garry Swann

Senior locum clinician, University Hospitals Birmingham NHS Foundation Trust

Kate Hawley

Lay specialist committee member

Nick Woznitza

Consultant radiographer, University College London Hospitals NHS Foundation Trust

Sarah Markham

Lay specialist committee member

Susan Shelmerdine

Consultant paediatric radiologist, Great Ormond Street Hospital for Children NHS Foundation Trust

NICE project team

Each diagnostics evaluation is assigned to a team consisting of a technical analyst (who acts as the topic lead), a technical adviser and a project manager.

Simon Webster

Health technology assessment analyst

Jacob Grant

Health technology assessment adviser

Toni Gasse

Project manager

ISBN: