NICE National Institute for Health and Care Excellence

NICE guidance

Artificial intelligence (AI) technologies to aid opportunistic detection of vertebral fragility fractures: early value assessment

HealthTech guidance

Published: 14 October 2025

www.nice.org.uk/guidance/htg760

Your responsibility

This guidance represents the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, healthcare professionals are expected to take this guidance fully into account, and specifically any special arrangements relating to the introduction of new interventional procedures. The guidance does not override the individual responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or guardian or carer.

All problems (adverse events) related to a medicine or medical device used for treatment or in a procedure should be reported to the Medicines and Healthcare products Regulatory Agency using the Yellow Card Scheme.

Commissioners and/or providers have a responsibility to implement the guidance, in their local context, in light of their duties to have due regard to the need to eliminate unlawful discrimination, advance equality of opportunity, and foster good relations. Nothing in this guidance should be interpreted in a way that would be inconsistent with compliance with those duties. Providers should ensure that governance structures are in place to review, authorise and monitor the introduction of new devices and procedures.

Commissioners and providers have a responsibility to promote an environmentally sustainable health and care system and should assess and reduce the environmental impact of implementing NICE recommendations wherever possible.

Contents

This guidance replaces HTE34.

1 Recommendations

Can be used with evidence generation

1.1 Five artificial intelligence (AI) technologies can be used in the NHS during the evidence generation period as options to aid the opportunistic detection of vertebral fragility fractures (VFFs). The technologies are:

- BriefCase-Triage
- CINA-VCF Quantix
- HealthVCF
- HealthOST
- IB Lab FLAMINGO.

These technologies can only be used:

- within their indicated populations as outlined in their instructions for use and with consideration of the risk groups as recommended in NICE's guideline on assessing the risk of fragility fracture in osteoporosis
- if the evidence outlined in the evidence generation plan is being generated
- as long as they have appropriate regulatory approval, including NHS England's Digital Technology Assessment Criteria (DTAC) approval.

Commissioners should take into account whether a technology is likely to remain available on the UK market and supported by its company when entering into a contract.

1.2 The companies must confirm that agreements are in place to generate the evidence. They should contact NICE annually to confirm that evidence is being

generated and analysed as planned. NICE may revise or withdraw the guidance if these conditions are not met.

1.3 At the end of the evidence generation period (3 years), the companies should submit the evidence to NICE in a format that can be used for decision making. NICE will review the evidence and assess if the technology can be routinely adopted in the NHS.

More research is needed

1.4 More research is needed on the following AI technologies that aid the opportunistic detection of VFFs before they can be funded by the NHS:

- Annalise Enterprise CXR/Annalise Container CXR
- BoneView
- TechCare Spine.

What this means in practice

Can be used with evidence generation

The 5 technologies listed in recommendation 1.1 can be used as an option in the NHS during the evidence generation period (3 years) and paid for using core NHS funding. During this time, more evidence will be collected to address any uncertainties. Companies are responsible for organising funding for evidence generation activities.

Take into account whether a technology is likely to remain available on the UK market and supported by its company before generating evidence to address the evidence gaps. Evidence generation should preferably be on technologies that will still be available in the NHS after the evidence generation period.

After the evidence generation period, NICE will review this guidance and the recommendations may change. Take this into account when negotiating the length of contracts and licence costs.

Potential benefits of use in the NHS during the evidence generation period

- **Clinical benefit:** Clinical evidence suggests that AI technologies can help opportunistically detect VFFs that would otherwise have been missed. This could help identify more people with a VFF who need treatment to improve their quality of life and reduce the risk of future fractures.
- **Resources:** By reducing the risk of further fractures, early detection and treatment of VFFs could reduce the demand on other costly services, such as those needed to manage hip fractures.
- **System benefit:** Using AI technologies can help reduce variation in clinical practice and help healthcare professionals to implement the Royal College of Radiologists' guidance for the recognition and reporting of osteoporotic vertebral fragility fractures.

Managing the risk of use in the NHS during the evidence generation period

- **Clinical subgroups:** There is no evidence to show whether the AI technologies

are equally clinically effective across all age groups. Older age is a risk factor, but there are other risk factors independent of age. It is uncertain whether the opportunistic detection of VFFs in all subgroups represents value for money in the NHS.

- **Resources:** Implementing the AI technologies could have a big impact on radiology services, such as increasing the number of diagnostic images that need to be reviewed by a radiologist and the number of referrals for dual-energy X-ray absorptiometry (DEXA) scans that need to be done.
- **Costs:** Early results from the economic modelling show that the technology could be cost effective. But, there is uncertainty around the cost of some of the technologies and the true cost of implementing them in the NHS. Trusts should take into account the costs of the AI technologies used in this assessment when implementing the technologies. When negotiating with companies, trusts should also consider the upfront costs for implementing a technology and should monitor costs associated with its use in populations at a lower risk of osteoporosis.
- **Clinical risk:** Using AI technologies to help detect VFFs on diagnostic images is considered to have a low clinical risk. This is because the technologies are used in addition to standard care in which healthcare professionals make treatment decisions. AI technologies do not replace the definitive radiology review.
- **Implementation guidance:** Clear local protocols will need to be in place when using AI technologies. This is to ensure that healthcare professionals refer people with a newly identified VFF to the appropriate services.
- **Equality:** There is a risk that the AI technologies may have reduced diagnostic accuracy in different populations. These include younger people who may have risk factors for VFF, people from ethnic minorities and other groups that may have been underrepresented in the AI training set.

NICE has produced tools and resources to support the implementation of this guidance.

More research is needed

There is not enough evidence to support funding the 3 technologies listed in recommendation 1.4 for the purpose of opportunistic detection of VFFs in the NHS.

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

What evidence generation and research are needed

Evidence generation and more research are needed on:

- the diagnostic accuracy of the technologies compared with current NHS standard care, including in key subgroups such as people under 50 and people at a higher risk of a VFF
- the failure rates of the technologies and the reasons for failure
- the impact of identifying additional VFFs on referral rates for other services, including DEXA
- the impact of identifying additional VFFs on treatment
- the impact of introducing the technologies on the workload of healthcare professionals
- the short-term impact on quality of life of identifying and managing a VFF.

The evidence generation plan gives further information on the prioritised evidence gaps and outcomes, ongoing studies and potential real-world data sources. It includes how the evidence gaps could be resolved through real-world evidence studies.

Why the committee made these recommendations

AI technologies can help healthcare professionals spot VFFs on X-ray images and CT scans involving the spine that are done for unrelated conditions (opportunistic detection). Treatment can reduce symptoms and the risk of future fractures, so detecting VFFs early has clear benefits. Preventing future fractures can also reduce the demand on radiology services and save money elsewhere in the NHS.

BriefCase-Triage, CINA-VCF Quantix, HealthVCF, HealthOST and IB Lab FLAMINGO are designed to help detect VFFs on CT scans. Diagnostic accuracy evidence comparing them with a reference standard suggests that they can help detect moderate to severe VFFs. Early economic evidence shows that they could be cost effective. So, these 5 technologies are recommended for use with evidence generation.

Annalise Enterprise CXR/Annalise Container CXR, BoneView and TechCare Spine are designed to help detect VFFs on X-ray images. Clinical evidence for Annalise Enterprise CXR/Annalise Container CXR is uncertain because it is based on studies that included mostly or only lateral chest X-ray images. In the NHS these are not commonly done and are usually only performed in specific groups. So, the evidence may not be generalisable to the NHS and the diagnostic accuracy of the technology in this context is uncertain. The usefulness of BoneView and TechCare Spine is uncertain because they only analyse spine X-ray images, which are usually done for indications relating to back pain and include a thorough review of the spine. This means VFFs are less likely to be missed on these X-ray images, so the technologies may not offer additional benefit. In addition, there is no clinical evidence for TechCare Spine. So, these 3 technologies can only be used in research.

Evidence on the diagnostic accuracy of the technologies compared with standard care in the UK is limited. So, more data should be collected to show how much better they are at detecting additional VFFs in clinical practice. More evidence is also needed on the downstream effects of the technologies for both the people having diagnostic imaging and the healthcare professionals. This should include the effect of the technologies on the rates of referral and on treatment, quality of life and healthcare professional workload. Companies should also address gaps in the evidence around how often their technologies are unable to process an image (the failure rate) and the reasons why this happens.

2 Information about the technologies

2.1 The technologies included in this early value assessment use artificial intelligence (AI) algorithms to assist the opportunistic detection of vertebral fragility fractures (VFFs) on X-ray images and CT scans involving the spine. They are intended to be used as decision aids for healthcare professionals interpreting the X-ray or CT scan.

2.2 Some companies provide the software directly, whereas others provide it through multivendor platforms. The technologies use X-ray or CT scans in digital imaging and communications in medicine (DICOM) format, which are stored on the hospital's picture archiving and communications system (PACS).

2.3 Different technologies report and display results in different ways including as annotated images within PACS Viewers, DICOM Secondary Captures or through standalone applications. Some also have notifications or summary reports. The technologies included in this assessment are shown in table 1.

Table 1 Artificial intelligence technologies

Technology (company)	CE mark	Population	Image type	Compatible imaging
Annalise Enterprise CXR/ Annalise Container CXR (Annalise.AI)	Class IIb	People 16 years and over	Chest X-ray	Anterior–posterior, posterior–anterior or lateral view
BoneView (Gleamer)	Class IIa	People over 2 years	Spinal X-ray	Appendicular skeleton, ribs and thoracic-lumbar spine
TechCare Spine (Milvue)	Class IIa	Not reported	Spinal X-ray	Thoracic or lumbar spine lateral views
BriefCase-Triage (Aidoc Medical)	Class IIa	People 18 years and over	CT	Chest, abdominal
CINA-VCF Quantix (Avicenna.AI)	Class IIb	People 50 years and over	CT	Chest, abdominal

Technology (company)	CE mark	Population	Image type	Compatible imaging
HealthVCF (Nanox AI)	Class IIa	People 50 years and over	CT	Chest, abdominal pelvic showing T1 to L5
HealthOST (Nanox AI)	Class IIa	People 50 years and over	CT	Chest, abdominal pelvic showing T1 to L4
IB Lab FLAMINGO (IB Lab)	Class IIa	People 50 years and over	CT	Thoracic or lumbar spine

2.4 Some of the technologies have additional functionalities, such as vertebral labelling, bone mass density assessment, prioritisation tools or the ability to detect other pathologies. This early value assessment has assessed the clinical and cost effectiveness of the technologies only for the opportunistic detection of VFFs.

3 Committee discussion

The diagnostics advisory committee considered evidence on artificial intelligence (AI) technologies for the opportunistic detection of vertebral fragility fractures (VFFs) from several sources. This included evidence submitted by the companies, a review of clinical and cost effectiveness by the external assessment group (EAG), a resource impact assessment by NICE and responses from stakeholders. Full details are available in the project documents for this guidance.

The condition

3.1 A VFF is a fracture in the spine that happens when bones are weaker than normal. VFFs can happen after a fall from standing height or lower (low-energy trauma) or spontaneously from day-to-day activities involving very little trauma or stress. They are the most common type of fragility fracture caused by osteoporosis, which reduces bone density and strength. Osteoporotic VFFs are common in older people and particularly in women, trans men and non-binary people after menopause. But, they can also be associated with other conditions or factors, such as chronic or long-term corticosteroid or glucocorticoid use or malignancy in the vertebrae. Other risk factors include:

- a history of falls
- family history of hip fracture
- low body mass index
- smoking
- alcohol intake and
- secondary causes of osteoporosis, such as:
  - rheumatoid arthritis
  - inflammatory bowel disease or
  - malabsorption.

Patient experts explained that vertebral fractures can be life changing and emotionally challenging.

Current practice

3.2 VFFs can be identified when a person presents to a healthcare setting with symptoms that suggest a VFF. VFFs can also be detected incidentally on diagnostic images that include the spine but were taken for reasons other than a suspected VFF. This is known as opportunistic detection. Clinical experts explained that there is no clear pathway for people with non-acute VFFs and there is variation across the NHS. Where available, people are referred to fracture liaison services. The Getting It Right First Time (GIRFT) guidance on vertebral fragility fractures outlines best practice in the care of people with a VFF in the NHS.

Unmet need

3.3 People with a VFF often experience deformity, height loss, immobility and pain, which leads to reduced quality of life. The risk of death is also higher. VFFs are also a strong predictor of further osteoporotic fractures, such as hip fractures. The economic cost of fractures to the NHS is substantial. There are effective pharmacological and non-pharmacological treatment options for managing the symptoms associated with a VFF. Treatment can also reduce the risk of further fractures.

3.4 Millions of diagnostic images are taken annually in the NHS for reasons other than VFF detection. These could be used to opportunistically detect VFFs. But the clinical experts noted that despite ongoing efforts to raise awareness of VFFs, most remain undiagnosed. The clinical and patient experts stressed that improving detection and treatment offers a significant opportunity to reduce the burden of VFFs and reduce the risk of further fractures. The committee concluded that there is an unmet clinical need that can be addressed by AI technologies.

Innovative aspects

3.5 The technologies use AI to detect vertebral fractures. This could improve VFF detection rates, leading to more people getting the care they need. The clinical experts highlighted that many AI technologies have been adopted across the NHS and that AI offers significant potential for improving care.

3.6 All of the identified technologies have algorithms that are fixed. Four companies (Aidoc Medical, Annalise.AI, Nanox AI and Avicenna.AI) have said that their technologies have settings to control the AI software's sensitivity and specificity, which are configured at setup or during use. This can help tailor the performance based on a hospital or centre's needs.

Clinical effectiveness

Evidence base

3.7 From the EAG's evidence searches there were 22 studies that met the inclusion criteria for the clinical-effectiveness review. Most studies evaluated HealthVCF (8 studies), Annalise Enterprise CXR/Annalise Container CXR (5 studies), IB Lab FLAMINGO (4 studies) or CINA-VCF Quantix (3 studies). There was 1 study each on BoneView and BriefCase-Triage. During consultation on the draft guidance, 3 additional relevant studies were submitted: 1 on Annalise Container CXR and 2 on HealthOST. Most studies included diagnostic accuracy as an outcome and were retrospective. Eleven of the studies reported the technologies' failure rates. Other relevant outcomes were reported in a minority of studies.

Diagnostic accuracy

3.8 Diagnostic accuracy evidence was available for all of the technologies except TechCare Spine. The majority of the diagnostic accuracy evidence compared the performance of the technologies against a reference standard. Most of the studies demonstrated high sensitivity and specificity for detecting moderate and severe vertebral fractures. Thirteen studies on 5 technologies also compared the

AI software's performance against standard care (most commonly the original radiology report) but most of these were not done in the UK. These studies suggest that the AI softwares could improve the detection rate of VFFs compared with standard care. But the committee noted that these studies may not reflect standard care in the NHS. The clinical experts commented that, in their experience, the detection rate in the UK is low and that some data from UK prevalence studies and databases suggests that many VFFs remain undiagnosed. The committee judged that the evidence on the diagnostic accuracy of standard care in the NHS is very limited and uncertain but that the technologies are likely to improve detection rates.

3.9 The EAG also highlighted that the reference standards varied across the studies and that some may not reflect NHS practice. A specialist committee member explained that the reference standard in the NHS would be at least 1 radiologist with specialist musculoskeletal training reviewing the diagnostic image specifically looking for a VFF. The committee judged the evidence from non-UK studies versus a reference standard to still be informative for the diagnostic accuracy of the AI technologies.

3.10 Most of the evidence is retrospective. The EAG explained that a retrospective study design is appropriate for assessing a technology's diagnostic accuracy because of the risk of participation bias with prospective studies. But, prospective evidence would be better suited to show the impact of the technologies on other outcomes, such as changes to clinical management. The committee agreed that retrospective evidence in this case was appropriate for assessing the diagnostic accuracy of the technologies.

3.11 The committee concluded that, overall, the evidence suggests that the AI technologies can detect additional moderate to severe vertebral fractures (as confirmed by a reference standard) that were not reported in the original radiology report. Generally, the technologies could detect these fractures with a specificity above 90%. But it was uncertain how much the technologies can improve VFF detection in the NHS. So, the committee agreed that more evidence is needed that compares the technologies with standard care in the NHS. No studies were identified for TechCare Spine, so the committee concluded that its diagnostic accuracy is unknown.

Choice of technologies

3.12 TechCare Spine and BoneView both analyse X-ray images of the spine. The clinical experts explained that it is less likely that a VFF would be missed on this type of image. This is because it would usually be taken to investigate back pain, so the spine would be thoroughly reviewed. The clinical experts added that a VFF is also less likely to be missed on a lateral chest X-ray image, so they questioned the value of Annalise Enterprise CXR/Annalise Container CXR. They also noted that although frontal and back-view (anterior–posterior and posterior–anterior) chest X-ray images are routine practice, lateral chest X-ray images are no longer commonly done in the NHS. So, technologies that detect VFFs on lateral X-ray images may be less useful in the NHS. The company Annalise.ai said that lateral chest X-ray images are still routinely taken in some NHS trusts for specific groups of people. It also noted that a lateral chest X-ray image is optional for Annalise Enterprise CXR/Annalise Container CXR, and that the technology can also analyse frontal and back-view chest X-ray images. But the EAG cautioned that the diagnostic accuracy studies for this technology included mostly lateral chest X-ray images, so the diagnostic accuracy of the technology using frontal and back-view images alone is uncertain. The committee acknowledged that a minority of trusts may routinely take lateral chest X-ray images, but those would usually be specialised centres that provide care for relatively small groups of high-risk patients. This is because the added cost and radiation exposure usually outweighs the benefit of the lateral image. The committee also noted that VFFs in only the thoracic, not the lumbar, spine will be detected on a lateral chest X-ray image. So, it concluded that further research using X-ray images applicable to NHS practice is needed on the diagnostic accuracy of these technologies before they can be funded in the NHS.

3.13 The committee heard that HealthOST is an updated technology based on HealthVCF. The company explained that the algorithm used in HealthOST was trained on a similar, but larger, dataset to the algorithm used in HealthVCF. It also explained that, at the time of the second committee meeting, both technologies were available in the NHS. But, the committee noted that the current regulatory approval for HealthVCF is due to expire in 2028, so the technology is unlikely to be available on the UK market after 2028. The company confirmed that it intends to gradually replace HealthVCF with HealthOST. The committee considered the evidence on HealthVCF and HealthOST. It acknowledged that, because the

algorithm had changed, the evidence on HealthVCF was unlikely to be generalisable to HealthOST. The committee concluded that its preference was for evidence to be generated using HealthOST while it is used in the NHS, because this is the technology that would be more widely available in the future. But, the committee noted that HealthVCF would remain available for some time. It cautioned that trusts implementing the technologies should consider whether the technologies are likely to remain available on the UK market and supported by their companies when entering into contracts.

Impact on clinical management

3.14 The committee asked about the impact of false positive results from the AI technologies. The clinical experts explained that this may lead to some unnecessary dual-energy X-ray absorptiometry (DEXA) scans. But, they added that the technologies can only be used as a decision aid, and the healthcare professional reviewing the diagnostic image (the reporting practitioner) may identify some of the false positive results. They also noted that a high number of false positives would have an impact on the workforce if an additional review by a radiologist was needed, as well as affecting the people who would need additional imaging.

3.15 The clinical experts also highlighted the lack of evidence on how many people would be referred or treated after opportunistic identification of their VFF. They also noted the variation in access to fracture liaison services in the NHS. The EAG confirmed that very little evidence was identified on how introducing the technologies affected the clinical management of VFFs, particularly in an NHS setting. The committee concluded that evidence on the impact on referral rates, treatment and the radiology workload of introducing the technologies should be generated for all of the technologies.

Other outcomes

3.16 The committee noted that there was at least 1 study that provided evidence about the failure rates for 7 of the technologies and that they differed between the technologies. The committee queried the definition of failure of the AI

software to interpret a diagnostic image and asked if there was evidence on the causes of failure from the clinical evidence review. The EAG explained that the failure rate included both a failure of the AI software to process a diagnostic image or a failure of the AI software to produce a definitive report. There could be a number of reasons for this, for example unable to upload image, unable to process image, incorrect classification as unsupported anatomical regions, unsuccessful model inference, and unable to analyse image. It also explained that the causes of failure were not reported in most cases. A specialist committee member added that the failure rate could be related to the image quality of the diagnostic image. The committee concluded that the failure rates and the reasons for failure represent evidence gaps, and further evidence should be generated for all of the technologies.

Cost effectiveness

Clinical parameters

3.17 The EAG developed an early economic model to explore the potential cost effectiveness of opportunistic VFF detection with assistance from AI technologies compared with current standard care (reporting radiographer without AI assistance). The committee noted that the sensitivity and specificity values used for the standard care arm in the model were from a small expert elicitation study, so were very uncertain. It recalled that there was a lack of studies comparing the use of AI technologies with standard care in the NHS (see sections 3.8 to 3.11). The committee also queried whether it is known what proportion of people whose VFF is opportunistically detected are already having osteoporosis treatment. This is because identifying a VFF will not provide any added benefit for these people, in terms of future fracture risk reduction. The committee also queried whether the proportion of people who have already had or been referred for a DEXA scan is known. It recalled that there was a lack of evidence on the impact of the technologies on clinical management (see sections 3.14 and 3.15). The EAG said that this may be captured in the model because it assumed that only 15% of people correctly identified as having a VFF would have treatment. But, this was also based on the same expert elicitation study. The committee concluded that there was substantial uncertainty in some of the

clinical parameters in the model because of the lack of data. It recommended that this should be addressed with evidence generation.

Quality of life

3.18 A quality-of-life benefit was modelled for people with a VFF correctly identified in the model. But the EAG noted that this was from a study in people who were diagnosed with a vertebral fracture after presenting with symptoms and so may be experiencing more severe symptoms than someone diagnosed opportunistically. Because of this, the EAG said that the short-term gain from treatment was probably too high. It explained that to better reflect the population in this assessment, it halved the utility value in the base case and explored even smaller utility gains in scenario analyses. These had a large impact on the cost-effectiveness results. The patient experts noted that VFFs can be extremely painful but that getting a diagnosis can be difficult because the symptoms can be mistaken for something else. There may also be pain in other areas, such as the stomach. The clinical experts agreed that people with opportunistically identified VFFs are very likely to still experience symptoms. They stressed that medicines can provide relief and can therefore also improve quality of life. The committee agreed that there was substantial uncertainty around the short-term impact on quality of life of identifying VFFs opportunistically. It concluded that future evidence generation should address this evidence gap.

Cost parameters

3.19 The EAG calculated the cost per scan for each technology, which included product subscription, implementation, integration, training and maintenance costs. Clinical experts noted that in practice implementation costs can vary widely and may depend on the individual centre implementing them. The company-provided costs were commercial in confidence, so the EAG also modelled a hypothetical scenario using a generic AI technology costing £7.36 per scan. The committee noted that the cost per scan of all technologies was similar to or below the cost of a generic AI technology. No cost was provided for BoneView. The EAG used a notional additional cost of £1 per scan to the cost of scanning provided by the company. But the committee noted that it was not

known whether this cost reflected the true cost of the technology. So, it concluded that the potential cost effectiveness of BoneView was even more uncertain. The committee advised that trusts should consider the costs used in this assessment when implementing the technologies.

3.20 The committee noted that the cost for treating and managing a VFF after it has been identified may have been overestimated. This was because it was sourced from a technology appraisal on a treatment for osteoporosis, which was based on people with a diagnosed VFF and included hospitalisation costs. The committee queried whether people who have been diagnosed under standard care may be experiencing more severe symptoms than those identified opportunistically. But, it recognised that, for an assumed level of clinical benefit, using a higher value for the cost of treating and managing a VFF would underestimate the value of AI technologies. This therefore served as a conservative estimate.

Model structure

3.21 The EAG's model was a decision tree with a 1-year time horizon. The committee highlighted that the short time horizon of the model was a major limitation. This is because many of the benefits of detecting a VFF earlier would occur beyond this short time horizon. But, it noted that there are also costs that would be incurred in the future. The EAG said that any longer-term modelling would have been subject to substantial uncertainty because of the lack of evidence on the impact of the technologies on clinical management (see sections 3.14 and 3.15). But it expected that including longer-term costs and benefits would likely improve the cost effectiveness of the technologies. This is because additional relevant costs and effects would be included. For example, a reduced future fracture risk for people whose VFFs are identified earlier and have treatment. A reduction in quality of life and additional costs for people whose VFF is not reported would also be included. The committee concluded that longer-term modelling would be needed in the future to reduce this uncertainty when the recommendations are reviewed after evidence generation.

Plausibility of cost effectiveness

3.22 The committee noted that, in the base case, all of the technologies were more expensive than standard care. But, they also led to quality-of-life gains for the people whose VFFs were identified and treated.

3.23 The committee recalled some of the key uncertainties related to parameters in the model. Among them, the committee recognised the uncertainty of the sensitivity and specificity used in the standard care arm (see section 3.17). But it noted that varying those parameters had a small impact on the results in the sensitivity analyses. That is, unless the diagnostic accuracy of standard care approached that of the AI technologies, in which case the AI technologies were unlikely to be cost effective. The committee also acknowledged the significant uncertainty of the utility gain parameter (see section 3.18). It noted that if the utility gain was smaller than the one used in the base case, the technologies were unlikely to be cost effective. But, it recalled that the EAG did not capture any longer-term benefits of the AI technologies, and this is likely to have underestimated their value (see section 3.21). The committee noted the uncertainty in some of the parameters and the limitations of the model structure. But it concluded that, despite this, it is plausible that the AI technologies could be cost effective if implemented in the NHS.

Risks

3.24 The committee noted the resource impact assessment. It showed that implementing the technologies in the NHS could lead to a significant increase in the number of X-ray images and CT scans that need to be reviewed by a radiologist. This is because many of the diagnostic images with a VFF identified may need an additional review by a specialist radiologist, especially if the first review was done by a reporting practitioner without specialist musculoskeletal training, which is likely to be the case with opportunistic detection. There would also be an increase in the number of referrals for a DEXA scans that need to be done, because most people would be referred after a VFF was identified. The committee heard that using 1 of the technologies in a large NHS trust had led to a significant increase in workload. It also heard that across the NHS there are capacity issues for both acquisition and reporting of DEXA scans. The committee

noted that although immediate large-scale implementation across the NHS is unlikely, introducing these AI technologies could significantly increase the pressure on radiology and other downstream services. The committee also noted that most of the provisionally recommended technologies are only indicated for use in people over 50. So, the resource impact of using those technologies will be smaller than the impact of the technology that is indicated for use in people over 18. The committee recalled that age is an important risk factor for VFFs, but that there are also other risk factors independent of age that can result in VFFs in younger people (see section 3.1). But the AI technologies can only be configured to assess specific diagnostic images based on demographic information, such as age, and not on other risk factors, so they could not be targeted for those specific groups. The committee recognised that the prevalence of VFFs may overall be lower in people under 50, so using the technologies to analyse images in this age group may be less beneficial. The committee also highlighted that, depending on the false positive rate, the resource impact could be much greater if the technologies are used in a wider population. But it noted that the EAG did not identify any evidence to enable any subgroup analyses. So, it was uncertain whether the clinical effectiveness of the technologies would differ in people under 50 or people with another risk factor. It was also uncertain whether it would be cost effective to use the technologies to analyse images in younger age groups or what an appropriate age cut-off might be. The committee concluded that evidence should still be generated across all age and risk groups to establish whether this would be a good use of resources. The committee added that the financial and system risks would need to be managed when generating this evidence. It noted that NICE's guideline on assessing the risk of fragility fracture in osteoporosis can help NHS trusts define groups that are at higher risk, if prioritising specific groups is considered appropriate when implementing the technologies.

3.25 The committee queried whether implementation of the AI technologies could lead to healthcare professionals becoming over-reliant on them and whether this could lead to deskilling in the longer term. The clinical experts explained that it is possible that healthcare professionals would learn from the AI's feedback and there is no imminent risk of deskilling. The committee recalled that the technologies can only be used as a decision aid, so would still always need clinical review and judgement (see section 2.1).

Equality considerations

3.26 The committee recalled that the technologies vary in their indications and that 5 of them are indicated only for people over 50. But, the clinical experts emphasised that VFF risk rises significantly with age and that most VFFs are in people over 50. The committee also recalled that in the clinical evidence review, the mean or median ages of the study populations were generally between 65 and 80 years. The committee noted that 1 of the recommended technologies is indicated for use in people over 18 years. It highlighted that all of the technologies should be used within their indicated populations, as outlined in each technology's instructions for use. But, it added that evidence generation in younger populations could help guide future recommendations on targeting the technologies. The committee recalled that osteoporotic VFFs do happen in younger people and that there are multiple risk factors. In particular, they are more common in women, trans men and non-binary people after menopause, in whom osteoporosis is more common. But there are other risk factors for osteoporosis (see section 3.1). The committee also recalled that VFFs can also be a result of chronic or long-term corticosteroid or glucocorticoid use or malignancy in the vertebrae.

3.27 The committee highlighted that a common limitation of AI technologies is the lack of transparency about the data used to train the algorithm. It thought that the technologies may perform worse for people who may have been underrepresented in the AI training datasets. This could include younger people, ethnic minorities, people with comorbidities or those who have had previous treatment. The EAG remarked that there are limited details about the characteristics of the patient population in the clinical evidence. The committee noted that future evidence generation should include relevant patient characteristics that would allow for analyses to investigate whether the technologies have been tested or validated in diverse patient populations. The committee reiterated that consideration should be given to the data the algorithms were trained on and whether they work as well for all groups. It recommended that companies should be transparent in providing details on this data.

3.28 The committee heard that there are geographical inequalities with regard to access to radiology and bone health services. It is currently unknown whether

implementing the AI technologies could improve or exacerbate those inequalities. The evidence generation plan specifies that, ideally, future research should be done across NHS trusts with and without fracture liaison services and replicated across multiple centres.

4 Committee members and NICE project team

This topic was considered by NICE's diagnostics advisory committee, which is a standing advisory committee of NICE.

Committee members are asked to declare any interests in the technology 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.

Chair

Neil Hawkins
Vice chair, diagnostics advisory committee

NICE project team

Each evaluation is assigned to a team consisting of 1 or more health technology analysts (who act as technical lead(s) for the evaluation), a technical adviser, a project manager and an associate director.

Ivan Maslyankov
Technical lead

Judith Shore
Technical adviser

Deonee Stanislaus
Project manager

Rebecca Albrow
Associate director

Update information

Minor changes since publication

December 2025: Health technology evaluation 34 has been migrated to HealthTech guidance 760. The recommendations and accompanying content remain unchanged.

ISBN: 978-1-4731-7665-2