Guidance assessment consultation document for GID-HTE10059 Artificial intelligence (AI) technologies to aid opportunistic detection of vertebral fragility fractures: Early Value Assessment

15 July 2025

Guidance development process

NICE early value assessment (EVA) guidance provides recommendations on promising health technologies that have the potential to address national unmet need. NICE has assessed early evidence on these technologies to determine if earlier patient and system access in the NHS is appropriate while further evidence is generated.

EVA guidance recommendations are conditional while more evidence is generated to address uncertainty in their evidence base. NICE has included advice in this guidance on how to minimise any clinical or system risk of early access to treatment.

Further evidence will be generated over the next 3 years to assess if the benefits of these technologies are realised in practice. NICE guidance will be reviewed to include this evidence and make a recommendation on the routine adoption of this technology across the NHS. Find out more on the [NICE webpage on early value assessment (EVA) for medtech](#).

NICE is producing this guidance on AI technologies for the opportunistic detection of vertebral fragility fractures in the NHS in England. The diagnostics advisory committee has considered the evidence and the views of clinical and patient experts.

This document has been prepared for consultation with the stakeholders. It summarises the evidence and views that have been

considered, and sets out the recommendations made by the committee. NICE invites comments from the stakeholders for this evaluation and the public. This document should be read along with the [evidence](#).

The 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?
- Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of age, disability, gender reassignment, pregnancy and maternity, race, religion or belief, sex or sexual orientation?

After consultation:

- Based on the consultation comments received, the committee may meet again.
- If committee meets again it will consider the evidence, this evaluation consultation document and comments from stakeholders.
- The committee will then prepare the final draft guidance, which will go through a resolution process before the final guidance is agreed.

Note that this document is not NICE's final guidance on AI technologies for the opportunistic detection of vertebral fragility fractures. The recommendations in section 1 may change after consultation. More details are available in [NICE's health technology evaluations: the manual](#) and [NICE's early value assessment interim process and methods statement](#).

Key dates:

Closing date for comments: 29 July 2025

1   Recommendations

Can be used with evidence generation

1.1   Four 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
- IB Lab FLAMINGO.

These technologies can only be used:

- according to their indications as outlined in their instructions for use and with consideration of the risk groups as recommended in NICE's guideline on osteoporosis: assessing the risk of fragility fracture
- 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.

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.

What this means in practice

The 4 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.

After this, 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 with evidence generation

- **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, earlier detection and treatment of VFFs could reduce the demand on other costly services, such as those needed to manage hip fracture.
- **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 with evidence generation

- **Clinical subgroups:** There is no evidence to show whether the AI technologies are as 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 radiographs that need to be reviewed by a radiologist and the number of follow-up 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 implementation and ongoing post-market surveillance. 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 with a lower prevalence of osteoporosis.

- **Clinical risk:** Using AI technologies to help detect VFFs on radiographs 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, such as 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.

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
- HealthOST
- TechCare Spine.

> **What this means in practice**
>
> There is not enough evidence to support core NHS funding for the 4 technologies listed in recommendation 1.4 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 years and those at a higher risk of a VFF
- the failure rates of the technologies and the reasons for failure
- the impact on referral and treatment rates of identifying additional VFFs
- the impact on the workload of healthcare professionals of introducing the technologies
- 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 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 4 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 mostly included side-view chest X-ray images. These are not commonly used in UK clinical practice and so the evidence may not be generalisable to the NHS. The usefulness of BoneView and TechCare Spine is uncertain because they only analyse spine X-ray images, which are usually done for back pain and include a thorough review of the spine. This means VFFs are less likely to be missed on these X-ray images and so the technologies may not offer additional benefit. There is no clinical evidence for HealthOST and TechCare Spine. So, these 4 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 people having X-rays and CT scans and healthcare professionals, including the effect of the technologies on referral and treatment rates, 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).

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	X-ray	Chest (anterior-posterior, posterior-anterior or lateral orientation)
BoneView (Gleamer)	Class IIa	People over 2 years	X-ray	Appendicular skeleton, ribs and thoracic-lumbar spine
TechCare Spine (Milvue)	Not yet certified	Not reported	Lateral 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
HealthVCF (Nanox AI)	Class IIa	People 50 years and over	CT	Chest, abdominal pelvic showing T1-L5
HealthOST (Nanox AI)	Class IIa	People 50 years and over	CT	Chest, abdominal pelvic showing T1-L5
IB Lab FLAMINGO (IB Lab)	Class IIa	People 50 years and over	CT	Thoracic or lumbar spine

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 evidence 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 break 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. But, VFFs can also be

detected incidentally on radiographic images that include the spine but that 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 VFFs and there is variation across the NHS. Where available, people are referred to fracture liaison services.

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 symptomatic VFFs. Treatment can also reduce the risk of further fractures.

3.4 Thousands of radiographic images are taken annually 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 set up or during use. This can help tailor the performance based on a hospital or centre's needs. Some of the technologies include additional features, for example for triage or prioritisation.

Clinical effectiveness

Evidence base

3.7   There were 22 studies that met the inclusion criteria for the EAG's 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. Most studies included diagnostic accuracy as an outcome and were retrospective. Ten 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 and HealthOST. 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. Twelve studies on 4 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 software 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 a lot of 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 reviewing the radiograph 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, with generally high specificity. But it was uncertain how much the technologies can improve VFF detection in the NHS. So the committee agreed more evidence is needed comparing the technologies with standard care in the NHS. No studies were identified for TechCare

Spine and HealthOST, so the committee concluded that the diagnostic accuracy for these 2 technologies is unknown.

Choice of technologies

3.12 TechCare Spine and BoneView both analyse X-ray images of the spine. The clinical experts explained that they thought it was less likely that a VFF would be missed on those images. This is because they 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 side-view chest X-ray image and so they questioned the value of Annalise Enterprise CXR/Annalise Container CXR. They also noted that although frontal-view chest X-ray images are routine practice, side-view chest X-ray images are no longer commonly done in the NHS. So, technologies that detect VFFs on side-view X-ray images may be less useful in the NHS. The company Annalise.ai noted that a side-view chest X-ray image is optional for Annalise Enterprise CXR/Annalise Container CXR, and that the technology can also analyse frontal-view chest X-ray images. But the EAG cautioned that the diagnostic accuracy studies for this technology included mostly side-view chest X-ray images, so the diagnostic accuracy of the technology using frontal-view images alone is uncertain. The committee concluded that further research is needed on the diagnostic accuracy of these technologies, using X-ray images applicable to NHS practice.

Impact on clinical management

3.13 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 radiograph 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 additional review by a radiologist was needed.

3.14   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 and treatment rates and radiology workload of introducing the technologies should be generated for all of the technologies.

Other outcomes

3.15   The committee noted that there was at least 1 study that provided evidence about the failure rates for 6 of the technologies and that they differed between the technologies. The committee queried the definition of failure of the AI software to interpret a radiograph 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 radiograph or a failure of the AI software to produce a definitive report. 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 radiograph. 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

3.16 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 and 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). 3.11The committee also queried whether it is known what proportions of people whose VFF is opportunistically detected are already on bone density 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 3.13 and 3.14). 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

3.17 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 with symptomatic vertebral fractures and that the short-term gain from treatment seemed 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 may still be experiencing 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

3.18   The EAG calculated the cost per scan for each technology, which included product subscription, implementation, integration, training and maintenance costs. 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.

3.19   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, 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

3.20   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 3.13 and 3.14). 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

3.21   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.

3.22 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 3.16). 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 3.17). 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, which is likely to have underestimated their value (see section 3.20). 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

3.23 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 radiographic images with a VFF identified may need an additional review by a specialist radiologist, especially if the first review was done by a radiographer who has not had specialist musculoskeletal training. There would also be an increase in the number of follow-up DEXA scans that need to be done. The committee heard that using 1 of the technologies in a large NHS trust had led to a significant increase in workload. 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 services. The committee also noted that most of the provisionally recommended technologies are only indicated for use in people over 50 years. So, the resource impact of using those technologies will be smaller than the impact of the technology which is indicated for use in all adults over 18 years. 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 radiographs 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 years and 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.

3.24   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 that there is no imminent risk of deskilling. The committee recalled that the technologies can only be used as a decision aid and so would still always need clinical review and judgement (see section 2.1).

Equality considerations

3.25   The committee recalled that the technologies vary in their indications and that 5 of them are only indicated for people over 50 years. But, the clinical experts emphasised that VFF risk rises significantly with age and that most VFFs are in people over 50 years. 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.

3.26   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 may not work as well for all groups. It recommended that companies should be transparent in providing details on this data.

3.27   The committee heard that there are geographical inequalities with regard to access to radiology 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