NICE National Institute for Health and Care Excellence

HealthTech Programme

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

Draft guidance consultation comments

Included evidence

Comment number	Consultee	Section number	Comment	Response
1	Consultee 1 Annalise.ai	Not specified	We acknowledge the comprehensive nature of NICE's evidence review for AI-based vertebral fragility fracture (VFF) detection tools. However, we believe the guidance could be further strengthened by including studies that particularly validate the performance and clinical utility of Annalise CXR (Enterprise and Container), particularly in detecting vertebral abnormalities associated with fracture risk. Seah, et al. (2021)* We respectfully recommend reconsidering the exclusion of the Seah et al. (2021) paper, which presents robust validation data for CXR device. This large-scale multi-reader, multi-case trial, evaluated the model's performance across all 124 findings included in the algorithm (including spinal wedge fracture). The test set comprised 2,565 cases - 329 of which involved spinal wedge fractures. Importantly, this study demonstrated strong diagnostic accuracy for identifying spinal wedge fractures (AUC = 0.953). Moreover, it showed a significant improvement in radiologist performance: the area under the curve (AUC) increased from 0.719 (without AI assistance) to 0.857 (with AI assistance), and absolute increase of 0.138 (95% CI: 0.093 - 0.182). Excluding this study solely because it assessed additional findings does not seem justifiable, given its clear relevance to VFF detection. Frost, et al (2025)** As noted in our initial submission, a retrospective study (with a forthcoming prospective arm) has been conducted within an NHS trust. The	Thank you for your comment. The external assessment group (EAG) previously reconsidered the inclusion Seah et al. 2021 during the factual accuracy check process for the assessment report (see responses 25 and 26 on pages 8 and 9 of the responses to factual accuracy check comments table in the Supporting Documentation). The EAG noted that this study did not address the decision problem of this assessment. The EAG reviewed the abstract by Frost et al. 2025. It noted that it was published after the original EAG searches were conducted (17 to 19 February 2025). However, the EAG noted that the study does represent evidence for Annalise Container CXR (v2.2). Specifically, the abstract presents retrospective diagnostic data from a UK NHS setting: a specialist cancer hospital where the patient group may be at greater risk of VFFs. The EAG noted that 57% (955/1,669) included

Comment number	Consultee	Section number	Comment	Response
			retrospective phase – recently presented at the 2025 World Congress on Osteoporosis [see abstract P497 by Frost, F. et al] analysed real-world data. The authors reported that lateral projections were performed in 57% of all CXR cases over a 1-month period. Notably, AI detected vertebral fractures in 100 studies, of which 34 were had not been reported in the original radiology report. In 23/34 of the missed cases, at least 4/5 reviewers confirmed this was a true miss. As a consequence, "change management strategies have been implemented, and standardised codes for vertebral compression fractures were introduced for radiologists". * Seah, J.C.Y., Tang, C.H.M., Buchlak, Q.D., Holt, X.G., Wardman, J.B., Aimoldin, A., Esmaili, N., Ahmad, H., Pham, H., Lambert, J.F., Hachey, B., Hogg, S.J.F., Johnston, B.P., Bennett, C., Oakden-Rayner, L., Brotchie, P. & Jones, C.M., 2021. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. The Lancet Digital Health, 3(8), pp.e496-e506. Available at: https://doi.org/10.1016/S2589-7500(21)00106-0 **Frost, F., Simpson, S., Berger, L., Wong, F. & Barker, C., 2025. Retrospective analysis to assess improved identification of vertebral fragility fractures on chest X-ray in a UK cancer hospital. Poster P497, WCO-IOF-ESCEO Congress 2025, [online] The Christie Hospital, Manchester, United Kingdom. Available at: https://www.wco-iof-esceo.org/download/2025/abstract-book	lateral projections; which may be higher than a general population. The EAG highlighted that there was minimal detail on the methods (for example the definition and severity of VFF identified was not reported, dates of patient imaging and radiology reporting were not reported). The reference standard was a group of 5 radiologists/reporting radiographers. The AI flagged 100/955 lateral images as positive for VFFs. Of the 100 flagged VFF images, 76 had 100% agreement with the reference standard (true positive), 4 were 100% classed as negative for VFF by the reference standard (false positive). A total of 34 were flagged by AI but had no mention of VFF in the original clinical reports; on retrospective review of the lateral image, 23/34 (67.6%) had ≥80% consensus with the reference standard. The EAG noted that no patient characteristics were reported, and as clinical review only occurred for VFFs flagged by the AI, it was not possible to determine the proportion of false negatives. Additionally, no information was reported for the 714 frontal images. The EAG concluded that the abstract provided additional evidence for the Annalise Container CXR technology from an NHS setting but would not change the assessment report conclusions nor address the evidence gaps.

Comment number	Consultee	Section number	Comment	Response
				The committee considered the study by Frost et al. 2025. It recognised that it represented relevant evidence for the technology but that it was published after the EAG's original searches. The committee noted that the presented analysis was on lateral chest X-ray images and it therefore did not address its concerns as noted in section 3.12 of the guidance. Please see also the response to comment 12.
2	Consultee 1 Annalise.ai	3.7 Evidence base	We respectfully recommend the committee review the decision to exclude the Seah et al. (2021)* paper from the evidence. This paper robust validation data for the Annalise Enterprise CXR device. Excluding this study solely because it assessed additional findings does not seem justifiable, given its clear relevance to VFF detection. The study demonstrated strong diagnostic accuracy for identifying spinal wedge fractures (AUC = 0.953), and showed a significant improvement in radiologist performance: the area under the curve (AUC) increased from 0.719 (without AI assistance) to 0.857 (with AI assistance), and absolute increase of 0.138 (95% CI: 0.093 - 0.182). *Seah, J.C.Y., Tang, C.H.M., Buchlak, Q.D., Holt, X.G., Wardman, J.B., Aimoldin, A., Esmaili, N., Ahmad, H., Pham, H., Lambert, J.F., Hachey, B., Hogg, S.J.F., Johnston, B.P., Bennett, C., Oakden-Rayner, L., Brotchie, P. & Jones, C.M., 2021. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. The Lancet Digital Health, 3(8), pp.e496-e506. Available at: https://doi.org/10.1016/S2589-7500(21)00106-0	Thank you for your comment, please see the response to comment 1.
3	Consultee 2 Technical University of Munich	Not specified	Has all of the relevant evidence been taken into account? Yes.	Thank you for your comment.

Comment number	Consultee	Section number	Comment	Response
	(affiliated with Bone-screen)			
4	Consultee 2 Technical University of Munich (affiliated with Bone-screen)	1.4 Why the committee made these recommendations	To not include mild osteoporotic fractures may constitute a risk, as also minor fractures are associated with higher morbidity and mortality as well as with higher subsequent fracture risk. PMID: 36819281	Thank you for your comment. In the evidence review, the EAG used the definition and severity of the vertebral fractures from the identified evidence itself (where reported; see section 5.1 in the assessment report). Therefore, evidence on mild osteoporotic fractures was included in the assessment report and was available to the committee. However, many studies did not include mild fractures or did not report the severity of the included fractures altogether. The EAG also highlighted that the training and validation sets of some of the AI technologies may not have included mild fractures, so the evidence for these technologies would not be expected to include mild fractures either.
5	Consultee 3 Bone-screen	Not specified	Has all of the relevant evidence been taken into account? Yes, all relevant evidence seems to have been taken in consideration by the committee.	Thank you for your comment.
6	Consultee 3 Bone-screen	3.8 Diagnostic accuracy	Moderate to severe vertebral fractures (Genant Grades 2 and 3) often show clear deformities and are more likely to be picked up by radiologists or clinicians, even incidentally. Grade 1 (mild) fractures, however, can be subtle and are frequently missed. We believe that detecting grade 1 fractures with high specificity will assist fracture prevention with the highest ROI.	Thank you for your comment. Please see the response to comment 4.

Comment number	Consultee	Section number	Comment	Response
7	Consultee 5 Society & College of Radiographers	Not specified	Has all of the relevant evidence been taken into account? Yes, limited evidence currently available.	Thank you for your comment.
8	Consultee 5 Society & College of Radiographers	3.7 Evidence base	'reported the technologies failure rates' did those studies report what the failure rates were? ie classify in terms of technological / human error (such as positioning, patient movement) etc	Thank you for your comment. The EAG summarised evidence relating to failures within the assessment report (see section 5.1 and subsections for each technology). It noted that the definition of failure varied and included “unable to upload”, “unable to process X-ray”, “incorrectly classifying as unsupported anatomical regions”, “unsuccessful model inference”, “unable to analyse”. In some cases, the reason for failure was not reported. The EAG concluded that there was a lack of detail reported in the published evidence to determine whether there is a common characteristic, cause, or image which fails to be processed or analysed by the AI technologies (see section 5.2 of the assessment report). The committee recognised that the reason for failure represents an evidence gap and recommended evidence generation to address this gap.

Value of technologies using lateral X-ray images

Comment number	Consultee	Section number	Comment	Response
9	Consultee 1 Annalise.ai	Not specified	Under the Choice of Technologies (Section 3.12) the document states: Clinical experts added that a VFF is less likely to be missed on a side-view chest X-ray and questioned the value of Annalise Enterprise CXR / Annalise Container CXR We respectfully disagree with this conclusion, as it contradicts the evidence provided to the committee. Data from Annalise.ai has demonstrated that detection performance — both by the standalone model and when used as an assistive tool — is significantly improved compared to unassisted detection – indicating a substantial baseline miss rate without AI. The evidence provided (Seah et al. 2021, MGB study*), as well as in the literature (Majumdar et al (2005)**, Mitchell et al (2017)***) indicates that the miss rate of VFF on CXR remains high even when a lateral view is available. Further, we provided as evidence an ongoing study at The Christie NHS Trust [abstract is available here as poster P497 by Frost, F. et al****] the authors reported 23/100 studies with a VFF flagged by the AI were true cases that were missed on the original radiology report (despite including a lateral projection). This suggests a miss-rate of 23% for studies even with a lateral projection within real-world NHS Trust data. The committee raised concerns regarding the relevance of studies involving lateral projections, however we feel this does not diminish the value of our evidence, instead, it underscores it: if VFFs are still missed, even when a lateral projection is available, this highlights both the limitations of current practice, and the added value of the tool. We further believe that the statement “side-view chest X-ray images are no longer commonly done in the NHS” is inaccurate, and thus the assertion that “VFF is less likely to be missed on a side-view chest X-ray” is overstated. To our knowledge, there are a number of trusts across the NHS that routinely obtain lateral projections (albeit for a sub-set of the	Thank you for your comment. The EAG reviewed the additional evidence supplied by the consultee: • Majumdar et al. 2005 reported the number of moderate to severe vertebral fractures identified following clinical review of a sample of 500 patients undergoing chest X-ray (posteroanterior and lateral) for any reason between January and December 2001. Of the 459 included images, vertebral fractures were identified in 72 (15.7%); 29 (40.3%) of which were not included in the original radiologist report. The EAG noted that the dates of original reporting preceded a publication in 2021 by the Royal College of Radiologists which provided guidance for the recognition and reporting of osteoporotic vertebral fragility fractures. Additionally, the EAG noted that a prevalence of 29.7% was used in the economic base case, varied between 7.3% and 52.0% in sensitivity analysis which would cover the prevalence reported in this paper. • Mitchell et al. 2017 reported on the number of VFFs (grade 1, 2 or 3) on retrospective review of 157 patients who had a hip fracture and

Comment number	Consultee	Section number	Comment	Response
			population). For example Gateshead Health NHS Foundation Trust routinely acquires a lateral projection for patients over 40 years of age with haemoptysis; the Christie NHS Foundation Trust routinely performs lateral projections in oncology patients due to increased risk of osteoporosis induced fractures relating to their treatment of and the process of the underlying disease; Leeds teaching hospitals NHS Trust routinely performs lateral projections for all GP referred patients over 40 years of age. Review of our post-market surveillance data has indicated that within Gateshead trust, approximately 5.5% (524/10,024) of all CXRs processed to date include a lateral projection. Across all 10,024 studies, we have identified 561 cases as positive for Spine Wedge Fracture. The additional retrospective study performed by The Christie NHST Trust reported that within a 1-month period, 57% of all CXRs included a lateral projection. * Seah, J.C.Y., Tang, C.H.M., Buchlak, Q.D., Holt, X.G., Wardman, J.B., Aimoldin, A., Esmaili, N., Ahmad, H., Pham, H., Lambert, J.F., Hachey, B., Hogg, S.J.F., Johnston, B.P., Bennett, C., Oakden-Rayner, L., Brotchie, P. & Jones, C.M., 2021. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. The Lancet Digital Health, 3(8), pp.e496-e506. Available at: https://doi.org/10.1016/S2589-7500(21)00106-0 **Majumdar, S.R., Kim, N., Colman, I., Chahal, A.M., Raymond, G., Jen, H., Siminoski, K.G., Hanley, D.A. and Rowe, B.H., 2005. Incidental vertebral fractures discovered with chest radiography in the emergency department: prevalence, recognition, and osteoporosis management in a cohort of elderly patients. Archives of Internal Medicine, 165(8), pp.905–909. Available at: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/486519 ***Mitchell, R.M., Jewell, P., Javaid, M.K., McKean, D., Ostlere, S.J. (2017) 'Reporting of vertebral fragility fractures: can radiologists help reduce the number of hip fractures?', Archives of Osteoporosis, 12(1), p.71. doi: 10.1007/s11657-017-0363-y. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC5547187/ ****Frost, F., Simpson, S., Berger, L., Wong, F. & Barker, C., 2025. Retrospective analysis to assess improved identification of vertebral fragility	involvement of the spine on imaging in a UK NHS hospital in 2013. Of these, VFFs were identified in 41% (65/157); 54% (35/65) of which were not included in the original radiologist report. The EAG noted that 14/65 fractures were grade 1 (mild). The EAG noted that the dates of original reporting preceded a publication of the Royal College of Radiologists which provided guidance for the recognition and reporting of osteoporotic vertebral fragility fractures (published in 2021). Additionally, a prevalence of 29.7% was used in the economic base case, varied between 7.3% and 52.0% in sensitivity analysis which would cover the prevalence described by this paper. • Please see the response to comment 1 regarding Frost et al. 2025. The EAG noted that it was unable to verify the data provided in the 5th paragraph of the consultation comment, but that it indicated a small proportion of all routine chest X-rays include a lateral projection (5.5%) and that the proportion of patients undergoing lateral chest X-rays may differ depending on setting or population. The committee considered the additional evidence provided by the consultee. It

Comment number	Consultee	Section number	Comment	Response
			fractures on chest X-ray in a UK cancer hospital. Poster P497, WCO-IOF-ESCEO Congress 2025, [online] The Christie Hospital, Manchester, United Kingdom. Available at: https://www.wco-iof-esceo.org/download/2025/abstract-book	considered the totality of the evidence on Annalise CXR and noted that the evidence suggests that the technology can identify additional VFFs on lateral chest X-ray images. However, the clinical experts reiterated that these are not commonly done in the NHS and when they are done, this is usually for specific high-risk groups (see also the response to comment 12). The committee noted that the diagnostic accuracy of Annalise CXR is therefore uncertain in an NHS context (because most X-ray images are frontal and back-view). So, it concluded that it is still too uncertain whether the technology is a clinically and cost-effective use of NHS resources and the technology should only be used in a research context.
10	Consultee 1 Annalise.ai	Not specified	In light of the points clarified above – and acknowledged in the document, "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" - we question the committee's position that none of the CXR products have provided adequate evidence. We strongly urge the committee to review the recent evidence we have highlighted, specifically the evaluation conducted within a UK NHS setting (The Christie NHS Foundation Trust by Frost et al. (2025))*. This study suggests that use of Annalise technology could have increased vertebral fracture detection (by 23%) and that this has already impacted clinical management. We respectfully request the committee considers this additional data and re-evaluate its decision. We recommend that similar to the CT products,	Thank you for your comment. Please see the responses to comments 1 and 9.

Comment number	Consultee	Section number	Comment	Response
			Annalise Enterprise CXR / Annalise Container be used with evidence generation, at a minimum in clinical pathways where lateral chest X-rays are performed. This will allow Trusts to assess real-world effectiveness in line with the committee's evidence generation plan. *Frost, F., Simpson, S., Berger, L., Wong, F. & Barker, C., 2025. Retrospective analysis to assess improved identification of vertebral fragility fractures on chest X-ray in a UK cancer hospital. Poster P497, WCO-IOF-ESCEO Congress 2025, The Christie Hospital, Manchester, United Kingdom. Available at: https://www.wco-iof-esceo.org/download/2025/abstract-book	
11	Consultee 1 Annalise.ai	1.4 More research is needed	We appreciate the committee's position as stated in the section on the need for more research on AI technologies aiding opportunistic detection of vertebral fragility fractures. However, in light of evidence provided in the further comments sections, we kindly ask the committee to reconsider its recommendation and support the use of Annalise Enterprise CXR/Container with evidence generation, at a minimum in clinical pathways where lateral chest X-rays are performed.	Thank you for your comment. Please see the response to comment 9.
12	Consultee 1 Annalise.ai	1.4 Why the committee made these recommendations	To our knowledge, there are a number of trusts across the NHS that routinely obtain lateral projections (albeit for a sub-set of the population). For example Gateshead Health NHS Foundation Trust routinely acquires a lateral projection for patients over 40 years of age with haemoptysis; the Christie NHS Foundation Trust routinely performs lateral projections in oncology patients due to increased risk of osteoporosis induced fractures relating to their treatment of and the process of the underlying disease; Leeds teaching hospitals NHS Trust routinely performs lateral projections for all GP referred patients over 40 years of age. Review of our post-market surveillance data has indicated that within Gateshead Trust, approximately 5.5% (524/10,024) of all CXRs processed to date include a lateral projection. Across all 10,024 studies, we have identified 561 cases as positive for Spine Wedge Fracture. The additional retrospective study performed by The Christie NHST Trust reported that within a 1-month period, 57% of all CXRs included a lateral projection.	Thank you for your comment. NICE sought the opinion of healthcare professionals from the listed trusts. Clinicians from Leads Teaching Hospitals NHS Trust confirmed that during X-ray imaging it is standard practice for a lateral projection to be taken in addition to the frontal or back-view projection in people over 50 years of age. But, they acknowledged that the trust may be an outlier in the NHS and that many trusts may not be doing this routinely. This was confirmed by clinical experts and members of the committee. They recognised that lateral chest X-ray images would still be done in the NHS, albeit rarely. But they highlighted that this

Comment number	Consultee	Section number	Comment	Response
				would be for specific high-risk groups (for example, people with cancer or people who have had a pacemaker) for whom the value of the additional projection would outweigh the additional cost and radiation exposure.
13	Consultee 1 Annalise.ai	3.12 Choice of technologies	Frost et al (2025)* from The Christie NHS Trust, indicated 23% of vertebral compression fractures identified by the Annalise CXR device were initially missed by reporters (on studies that included a lateral projection). Gehlbach et al (2000)** reported that only 52% of moderate-to-severe vertebral fractures were noted in radiology reports of lateral chest radiographs. Majumdar et al (2005)*** reported that 40% of vertebral fractures were not reported in the official radiologists' report. *Frost, F., Simpson, S., Berger, L., Wong, F. & Barker, C., 2025. Retrospective analysis to assess improved identification of vertebral fragility fractures on chest X-ray in a UK cancer hospital. Poster P497, WCO-IOF-ESCEO Congress 2025, [online] The Christie Hospital, Manchester, United Kingdom. Available at: https://www.wco-iof-esceo.org/download/2025/abstract-book **Gehlbach, S.H., Bigelow, C., Heimisdottir, M., May, S., Walker, M. & Kirkwood, J.R., 2000. Recognition of vertebral fracture in a clinical setting. Osteoporosis International, 11(7), pp.577–582. doi:10.1007/s001980070078. Available at: https://link.springer.com/article/10.1007/s001980070078 ajph.aphapublications.org ***Majumdar, S.R., Kim, N., Colman, I., Chahal, A.M., Raymond, G., Jen, H., Siminoski, K.G., Hanley, D.A. & Rowe, B.H., 2005. Incidental vertebral fractures discovered with chest radiography in the emergency department: prevalence, recognition, and osteoporosis management in a cohort of elderly patients. JAMA Internal Medicine, 165(8), pp.905-909.	Thank you for your comment. The EAG reviewed the additional evidence supplied by the consultee: • Please see the responses to comments 1 and 9 regarding Frost et al. 2025. • Gehlbach et al. 2000 reported on the number of moderate or severe vertebral fractures on retrospective review of 934 women who had a lateral chest X-ray in a US hospital between 1995 and 1997. No AI was used in this paper. Of these VFFs were identified in 14.1% (132/934); only 1.8% had a discharge diagnosis of vertebral fracture. The EAG noted that the dates of original reporting preceded a publication by the Royal College of Radiologists which provided guidance for the recognition and reporting of osteoporotic vertebral fragility fractures (published in 2021). Additionally, a prevalence of 29.7% was used in the economic model's base case, and was varied between 7.3% and 52.0% in sensitivity analysis, which would cover the prevalence reported by this paper.

Comment number	Consultee	Section number	Comment	Response
			doi:10.1001/archinte.165.8.905. Available at: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/486519	• Please see the response to comment 9 regarding Majumdar et al. 2005.
14	Consultee 5 Society & College of Radiographers	1.4 Why the committee made these recommendations	'based on studies that mostly included side-view chest x=ray images.' presumably that means the training and validation data used in that circumstance is likely to be from an international population (not UK based) which may affect results/transferability when used on the UK population related to diversity of the UK	Thank you for your comment. The committee heard that in most trusts lateral chest X-ray images are not routinely done for all people, except for specific high-risk groups (see the response to comment 12). The EAG reviewed the 2 studies referred to in this comment (Frias 2023, Ghatak et al. 2024) and confirmed that neither were explicitly conducted in the UK or within an NHS setting. Please see the response to comment 9 regarding the committee's conclusions on technologies which detect VFFs on chest X-ray images.

Evidence generation and the evidence generation plan

Comment number	Consultee	Section number	Comment	Response
15	Consultee 2 Technical University of Munich (affiliated with Bone-screen)	1.4 More research is needed	From my point of view, all solutions need more research, as there is no "real world" study available with prospective NHS data.	Thank you for your comment. The committee recognised the need for further research for all of the technologies. The evidence generation plan outlines the real-world evidence that should be collected to address the evidence gaps.
16	Consultee 2 Technical University	1.4 Why the committee made	Companies should also assess the effect of image quality (voxel spacing, noise) on the accuracy.	Thank you for your comment. The committee considered the impact of image quality (voxel spacing, noise) on the accuracy and concluded that this does not

Comment number	Consultee	Section number	Comment	Response
	of Munich (affiliated with Bone-screen)	these recommendations		represent a key evidence gap. The evidence generation recommendations address the essential evidence gaps which need to be addressed by the end of the evidence generation period (3 years) so that the committee is able to make a recommendation when the guidance is reviewed.
17	Consultee 3 Bone-screen	2.1 Impact of the artificial intelligence technologies on health-related quality of life	Suggestion: When collecting the evidence on health-related quality of life as well as cost-effectiveness, please perform a sub-group analysis based on the grade of the detected fracture. E.g. impact of detecting a moderate fracture vs. a severe fracture.	Thank you for your comment. The evidence generation plan lists important subgroups that have been identified already, but this should not be read as an exclusive list. Other important subgroups could be added with input from clinicians while implementing the AI technologies.
18	Consultee 3 Bone-screen	2.1 Failure rates and diagnostic accuracy of the AI technologies ideally compared with NHS standard care	In the context of opportunistic screening, it's essential to gather evidence on diagnostic accuracy at the level of image quality as well. For example: How well does the algorithm perform on low-dose CTs? In patients with implants? In those with severe degenerative changes? These scenarios matter because individuals undergoing imaging often present with co-morbidities or structural changes. And the algorithm must be robust in such real-world conditions.	Thank you for your comment. The committee recommended further evidence generation for all technologies. With regards to the diagnostic accuracy outcomes, the technologies should be compared to current NHS care, therefore these real-world conditions should be included since the evidence is required from NHS care settings.

Comment number	Consultee	Section number	Comment	Response
19	Consultee 4 The Royal Osteoporosis Society	Not specified	**Need for Further Research and Evidence Generation** While optimistic about AI, the ROS agrees that further research is needed before widespread implementation. The use of AI for VFF detection is relatively new, and current evidence, though promising, has limitations. We support NICE's proposal for a managed evidence-gathering period: indeed, the draft guidance conditionally recommends a set of AI technologies for a period of up to three years while additional data are collected. This cautious approach is appropriate given the outstanding questions. We echo the NICE Diagnostics Advisory Committee's view that real-world evidence is required to determine the full impact and cost-effectiveness of these tools. Key research areas should include: diagnostic accuracy of AI vs. standard radiologist practice, failure rates and technical performance in diverse NHS settings, the impact on referral and treatment rates (i.e. do more patients actually get treated for osteoporosis due to AI, and how effectively), effects on healthcare professional workload, and any quality of life improvements for patients identified earlier. Generating robust data on these fronts will reduce uncertainty and inform a potential future move from conditional to routine adoption. Until such evidence is available, ROS supports a measured rollout, deploying AI in a controlled, evaluated manner rather than immediate nationwide implementation. We note that this is aligned with the EVA (Early Value Assessment) approach that seeks to balance early access with patient safety. In summary, our position is one of guarded optimism: we are confident in AI's potential, but we agree that rigorous evaluation is needed to ensure the technology truly delivers the expected benefits in practice and represents a good use of NHS resources.	Thank you for your comment. It has been considered by the committee.
20	Consultee 5 Society & College of Radiographers	1.2 Can be used with evidence generation	Might that also include withdrawal if evidence is indicating (prior to the full 3 year generation period) that models are not performing to an expected level. Is there a mechanism to withdraw if post market surveillance demonstrates the software is performing below what is deemed to be a safe level or threshold?	Thank you for your comment. NICE can withdraw all or part of the guidance if the company does not meet the conditions in section 4 of the evidence generation plan.
21	Consultee 5 Society &	1.3 Can be used with	what is the 'format that can be used for decision making'?	Thank you for your comment. The quoted statement refers to the reports and supporting documentation that will be

Comment number	Consultee	Section number	Comment	Response
	College of Radiographers	evidence generation		submitted to NICE after the 3-year evidence generation period. There are no requirements for specific formats, but the reports and documentations should be usable for decision-making.
22	Consultee 5 Society & College of Radiographers	1.3 Can be used with evidence generation	Is there a timeline available for 'NICE will review the evidence'?	Thank you for your comment. The evidence generation period is 3 years after publication of the guideline. Immediately after publication of the guidance and evidence generation plan, NICE will approach all companies whose technologies have been recommended for use in the NHS with evidence generation.
23	Consultee 5 Society & College of Radiographers	1.4 What evidence generation and research are needed	'the failure rates of the technologies and the reasons for failure' there is potential that the failure rates of the technologies will not be picked up clinically for some time (beyond the EVA timeframe). The technologies are designed to opportunistically find VFF. If the technology does not detect a VFF then it is not likely that there will be follow up imaging to demonstrate that there was a false negative result.	Thank you for your comment. The EAG clarified that within the evidence, the definition of failure varied (including unable to upload, unable to process X-ray, incorrectly classifying as unsupported anatomical regions, unsuccessful model inference, unable to analyse) and was not reported in some cases. The EAG noted that there was a lack of detail reported in the published evidence to determine whether there is a common characteristic, cause, or image which fails to be processed or analysed by the AI technologies. Section 3.16 of the guidance has been amended to clarify the different reasons for failure.
24	Consultee 5 Society & College of	1 1 Purpose of this document	'NICE will withdraw the guidance' Withdraw the guidance altogether or amend the guidance?	Thank you for your comment. The wording of the evidence generation plan has been amended to provide clarity.

Comment number	Consultee	Section number	Comment	Response
	Radiographers			
25	Consultee 5 Society & College of Radiographers	2.1 Impact of the artificial intelligence technologies on health-related quality of life	However, it is noted that not all NHS organisations have access to fracture liaison services. In which case, how will the companies be able to access that data - is it realistic to have this as an essential piece of evidence - will the tool have been used by the organisation at some point in the pathway?	Thank you for your comment. The evidence generation plan notes that the EQ-5D-3L questionnaire will ideally be used to collect health-related quality of life data. Therefore, while this is the preferred tool by NICE, its use is not mandatory. However, the short-term impact on quality of life of identifying and managing a VFF was identified as a key evidence gap and the committee recommended that further evidence needs to be generated to address this gap.
26	Consultee 5 Society & College of Radiographers	2.1 Failure rates and diagnostic accuracy of the AI technologies ideally compared with NHS standard care	'More evidence is needed on the failure rates and diagnostic accuracy' The evidence should specify / provide classification of the failure rates so that the evidence review is informed and able to evaluate the performance of the model.	Thank you for your comment. The reason for failure has been included as a key evidence gap to be addressed by further evidence generation.
27	Consultee 5 Society & College of Radiographers	2.2 Healthcare professional experience and	It would also be useful to record when there is any 'downtime' or when the system is not used/available.	Thank you for your comment. The evidence generation plan has been amended to note this aspect of healthcare professional experience and acceptability of AI technologies.

Comment number	Consultee	Section number	Comment	Response
		acceptability of AI technologies		
28	Consultee 5 Society & College of Radiographers	3.2 3.2 Data sources	The consultation notes that companies will be responsible for collecting and analysing this data. This means that companies must be given access to the data and in a form that they can use to extract the required data. This appears to be a simple process however the experiences reported by for example NHSE AIDF and collection of data for benefits metrics is that data can be difficult to obtain and when obtained it may not be 'clean' data in a form that is easily useable. A clear framework/process for companies to be able to obtain useable data sources in the timeframe set is crucial to their ability to meet the requirements of the EVA.	Thank you for your comment, it has been considered by the committee and NICE's evidence generation team. Recommendations on the process for obtaining data are beyond the remit of NICE's HealthTech programme. Companies should work with NHS trusts to obtain the required data and conduct the necessary analysis.
29	Consultee 5 Society & College of Radiographers	3.3 Real-world before-and-after implementation study	'After an enrolment period, data collection should be long enough for sufficient follow-up' Might that data also be collected retrospectively in this case?	Thank you for your comment. Where available, the data could be collected retrospectively to address the evidence gaps outlined in the guidance and evidence generation plan.
30	Consultee 5 Society & College of Radiographers	3.3 Real-world before-and-after implementation study	'Wider rollout may be possible within the period of evidence generation if, and when it becomes clear that the resource impact of the technologies is manageable.' Who will determine if the resource impact of the technologies is manageable - there is a potential clash of interests given that the companies will be responsible for data collection and analysis whereas the NHS organisation is responsible for decisions about resource impact.	Thank you for your comment. Local commissioners are best suited to follow the resource impact of the technologies and to propose the appropriate speed of roll-out.
31	Consultee 5 Society & College of Radiographers	3.4 Retrospective study	'for example obesity' It is likely that a company will be able to obtain obesity data retrospectively, will that have been recorded?	Thank you for your comment. Companies should work with NHS trusts where they are implementing the technologies to collect available data to address the evidence gaps.

Comment number	Consultee	Section number	Comment	Response
32	Consultee 5 Society & College of Radiographers	3.4 Retrospective study	'number of missed fractures' Number of missed vertebral fragility fractures (other fractures / old remodelled fractures may be present)	Thank you for your comment. The wording of the evidence generation plan has been amended.
33	Consultee 5 Society & College of Radiographers	3.4 Retrospective study	(radiologists, radiographers and other healthcare professionals)' Presumably that refers to suitably trained reporting radiographers rather than diagnostic radiographer/therapeutic radiographer.	Thank you for your comment. The wording of the evidence generation plan has been amended.
34	Consultee 5 Society & College of Radiographers	4 4 Monitoring	'NICE reserves the right to withdraw the guidance if data collection is delayed' or to amend as necessary	Thank you for your comment. The wording of the evidence generation plan has been amended.
35	Consultee 5 Society & College of Radiographers	5	The title of this section is minimum evidence standards but the section does not set iminium evidence standards. It appears to be more of a general overview of the data currently available and focus for future data collection.	Thank you for your comment. The respective section of the evidence generation plan provides a summary of the evidence that was used to recommend technologies for use in the NHS with evidence generation. Further, it lists the evidence that will be required after the 3-year evidence generation period when the technologies will be exiting the EVA process.

Other potentially relevant technologies

Comment number	Consultee	Section number	Comment	Response
36	Consultee 2 Technical University of Munich (affiliated with Bonescreen)	Not specified	Are the recommendations sound and a suitable basis for guidance to the NHS? I miss comments on simultaneous screening for low bone mass (opportunistic QCT). Both screening modalities are available and benefit from each other.	Thank you for your comment. Opportunistic detection of people with low bone mass density is outside the scope of the current assessment. The reference standard for this assessment was determined as reporting practitioner with MSK training interpreting the radiograph. Assessing a technology which detects low bone mass density may require a different reference standard.
37	Consultee 2 Technical University of Munich (affiliated with Bonescreen)	1.1 Can be used with evidence generation	SpineQ (Bonescreen GmbH, Germany) is another AI software for fracture detection in CT; CE-certified since June 2025	Thank you for your comment. SpineQ was not included in this assessment as it is not currently available on the UK market. Clinical experts also advised that the technology may not be in scope due to a potentially different positioning in the care pathway.
38	Consultee 2 Technical University of Munich (affiliated with Bonescreen)	2.3 2 Information about the technologies	the entry: "SpineQ (Bonescreen) / Class IIa / People over 18 years / CT / thoracolumbar spine" is missing in the table.	Thank you for your comment. Please see the response to comment 37.
39	Consultee 2 Technical University	3.23 Risks	This is true for most of the solutions. One possible solution is to opportunistically screen for both low BMD and fractures in a single application. This would readily identify the main risk group (i.e. people with low BMD) incorporating risk factors for low BMD. On the other hand, the	Thank you for your comment. Please see the response to comment 36.

Comment number	Consultee	Section number	Comment	Response
	of Munich (affiliated with Bonescreen)		specificity increases in such a setting as non-osteoporotic deformities are easily excluded, such as old traumatic fractures.	
40	Consultee 3 Bonescreen	Not specified	Are the recommendations sound and a suitable basis for guidance to the NHS? Request NICE to consider recommending Bonescreen's SpineQ. SpineQ has consistently demonstrated superior diagnostic performance compared to all other groups (students, residents, attendings) in detecting any fracture (including grade-1 or mild fracture) and moderate/severe fractures across vertebral, regional, and patient-level analyses. Importantly, it can also extract calibrated bone mineral density which is an important parameter aiding subsequent care.	Thank you for your comment. Companies that wish to increase their technology's visibility to the system should register with the NHS Innovation Service. This will give them access to a collective of organisations including NHS England Specialised Commissioning and the NHS Supply Chain and will provide access to coordinated support and guidance from a variety of support organisations with expertise in the development and widespread adoption of health innovation. Please also see the response to comment 37.
41	Consultee 3 Bonescreen	1.3 Can be used with evidence generation	For opportunistic screening, it is desired that any implemented tool works silently and automatically in the background with zero additional work for the care providers (e.g. we should only send a certain scan to this tool). So, such a tool should have a very high specificity (high true negatives) while also having a high sensitivity. The published results from two tools in your list are as follows: BL Flamingo: Sens (80%); Spec (90%). Avicenna CINA-VCF: Sens (94.4%); Spec (88.7%). Compare this to Bonescreen's SpineQ which has Sens (88%); Spec (99%), which is a better candidate for such opportunistic deployment, of course subject to further validation on NHS benchmarks.	Thank you for your comment. Please see the responses to comments 36 and 37. Further evidence generation on the diagnostic accuracy for all technologies was recommended by the committee. Please see the evidence generation plan for further details.
42	Consultee 3 Bonescreen	2.1 2 Information about the	What is the committees's opinion on also including MRI as a potential modality for opportunistic screening of VFFs? A compression fracture usually occurs on the vertebral body and this is clearly visualised on an	Thank you for your comment. The inclusion of MRI images for opportunistic VFF detection was considered at the scoping stage. However, clinical experts

Comment number	Consultee	Section number	Comment	Response
		technologies	MRI. Moreover, lumbar MRI is also one of the most commonly performed scan.	said that there are currently no commercially available technologies that do this.
43	Consultee 3 Bonescreen	3.16 Clinical parameters	On the flip side, a patient could be osteoporotic (<80mg/mm3 as per ACR guidelines) and not have a fracture yet. Such patients are at a significant risk of first fracture. To tackle such situations, we believe that we need additional information on top of VFF detection. Specifically, if we could extract the bone mineral density, preventive treatments and therapies can be better planned. This is also stated by Expert 5 on Page 198 in the Supporting Documentation (Point 4). Bonescreen's SpineQ extracts volumetric bone mineral density accurately, and is superior to DXA as detecting prevalent and prospective fractures. See evidence here: https://doi.org/10.1007/s00330-020-07655-2 and https://doi.org/10.3389/fendo.2020.586352. Also see https://doi.org/10.3389/fendo.2023.1222041 for evidence on cost effectiveness of opportunistic screening of osteoporosis through BMD.	Thank you for your comment. Please see the response to comment 37.

Impact on long-term outcomes

Comment number	Consultee	Section number	Comment	Response
44	Consultee 5 Society & College of Radiographers	1.3 Can be used with evidence generation	'increasing the number of radiographs' although arguably if the technology is reliable and provides valid results then at some point of acceptance / evidence of trustworthiness there should be no need for follow up radiographs	Thank you for your comment. The committee concluded that the impact of introducing the technologies on long-term outcomes (including referral for follow up imaging) represents a key evidence gap and further evidence should be generated for all of the technologies.
45	Consultee 5 Society & College of	1.3 Can be used with evidence	what are the likely effects on the 'appropriate services' - related to volume of work, resources available currently.	Thank you for your comment. The committee concluded that the impact of introducing the technologies on downstream services represents a key

Comment number	Consultee	Section number	Comment	Response
	Radiographers	generation		evidence gap and further evidence should be generated for all of the technologies.
46	Consultee 5 Society & College of Radiographers	3.13 Impact on clinical management	'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.' also impact on the individual patient having to attend for additional imaging and the radiographic workforce performing the imaging.	Thank you for your comment. The guidance has been amended to note the impact on people receiving further imaging.
47	Consultee 5 Society & College of Radiographers	3.21 Plausibility of cost effectiveness	and likely to result in long-term cost reduction to the NHS (fracture prevention).	Thank you for your comment. This is reflected in section 3.21 of the guidance, which describes that the EAG expected that including longer-term costs and benefits would likely improve the cost effectiveness of the technologies, because additional relevant costs and effects would be included. The committee concluded that the AI technologies could be cost effective if implemented in the NHS, but it is not known whether they would be cost saving in the long run.

Equality considerations

Comment number	Consultee	Section number	Comment	Response
48	Consultee 1 Annalise.ai	Not specified	We agree with the consultation's recognition of important equality considerations and offer some additional reflections: Chest X-ray (CXR) remains a widely accessible imaging modality, especially in healthcare settings where CT availability is limited. The integration of AI-powered CXR decision support tools that leverage existing chest radiographs may help prevent unintended inequalities in fracture detection by making advanced analysis available wherever X-rays are performed, but	Thank you for your comment. The committee considered the availability and accessibility of X-ray services compared with CT services. It considered this along the evidence for technologies which opportunistically detect VFFs on X-ray images. Please see the responses to comments 9, 12 and 13.

Comment number	Consultee	Section number	Comment	Response
			CT is less accessible. Opportunistic AI-based detection using routine chest X-rays offers a low-barrier approach to earlier diagnosis and intervention, reaching populations who might otherwise be missed by conventional screening pathways (e.g people under 50 years). Variation in radiological expertise can impact diagnostic consistency across regions. AI tools serve as valuable adjuncts, helping to standardise detection rates and reduce disparities in care quality.	
49	Consultee 1 Annalise.ai	3.25 Equality considerations	We concur with the committee's acknowledgement that vertebral fragility fractures (VFFs) can occur across a broad range of patient populations, including those under 50. In line with this, we would like to emphasize that we have specifically provided evidence demonstrating enhanced detection of VFFs in such a patient cohort (i.e. The Christie NHS Trust) — a population recognized as being at elevated risk for osteoporosis and consequent vertebral fractures, regardless of age or gender. This real-world evidence highlights the clinical value of VFF detection tools in populations beyond just postmenopausal women or older adults. Given that guidelines already advocate for increased vigilance in such cohorts, these findings support broader consideration of advanced detection technologies within the indicated at-risk populations—beyond strict age limitations.	Thank you for your comment. Please see the response to comment 1.
50	Consultee 2 Technical University of Munich (affiliated with Bone-screen)	Not specified	Are there any equality issues that need special consideration and are not covered in the medical technology consultation document? No.	Thank you for your comment.

Comment number	Consultee	Section number	Comment	Response
51	Consultee 2 Technical University of Munich (affiliated with Bone-screen)	3.26 Equality considerations	On the other hand, there is no evidence of diverse accuracy depending on different patient population. It my rather be the technical parameters of scans (reconstruction algorithm, spatial resolution, noise) that matters.	Thank you for your comment. The committee concluded that the diagnostic accuracy of the technologies compared with current NHS standard care in key subgroups represents a key evidence gap and further evidence should be generated for all of the technologies.
52	Consultee 3 Bone-screen	Not specified	Are there any equality issues that need special consideration and are not covered in the medical technology consultation document? The equality issues are well considered.	Thank you for your comment.
53	Consultee 5 Society & College of Radiographers	Not specified	Are there any equality issues that need special consideration and are not covered in the medical technology consultation document? Not immediately apparent beyond the information contained here and in the supporting documents.	Thank you for your comment.
54	Consultee 5 Society & College of Radiographers	3.25 Equality considerations	'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 which case that is also the majority of that population after the age of 50 (not younger than 50).	Thank you for your comment. Please see the response to comment 51.
55	Consultee 5 Society & College of Radiographers	6 Equalities	'This could worsen regional inequalities, particularly for people living in deprived areas.' This is a valid statement/consideration which appears to be rather passive simply 'considering' - NHS organisations must meet their legal duties with respect to equality and health inequalities. Active mitigation of risks if preferable to consideration.	Thank you for your comment, it has been considered by the committee. When discussing the existing geographical inequalities with regard to access to radiology and bone health services, the committee noted that it is possible that implementing the AI technologies could both improve or exacerbate those

Comment number	Consultee	Section number	Comment	Response
				inequalities (see section 3.28 of the guidance). The evidence generation plan specifies that future evidence generation should be across a variety of trusts, for example those that have and those that don't have access to fracture liaison services.

Technology adoption and implementation

Comment number	Consultee	Section number	Comment	Response
56	Consultee 4 The Royal Osteoporosis Society	Not specified	**Early Detection, Fracture Prevention and FLS Integration** Early detection is essential for effective fracture prevention. Vertebral fractures often go unrecognized until more serious injury occurs, but if identified in time, they serve as a critical warning sign to intervene against osteoporosis progression. We emphasise that detection must be coupled with prompt management. In practice, this means any VFF identified by an AI algorithm should trigger a referral into a Fracture Liaison Service (FLS) or equivalent osteoporosis care pathway. FLS are proven to systematically assess and treat patients after a fragility fracture, reducing the risk of subsequent fractures by up to 40% when robust services are in place. The ROS's vision is that AI-powered detection will feed into FLS programs so that no patient falls through the cracks. Early identification of a VFF, followed by FLS-led intervention (e.g. bone density assessment and medication), can break the fracture cycle and prevent the cascade of hip fractures and other injuries that often follow an initial spinal fracture. We note that opportunistic identification of VFFs via AI aligns perfectly with the "capture the first fracture" principle underpinning FLS – catching osteoporosis in patients as soon as possible after the first sign (in this case, a vertebral break) to stop avoidable suffering. ROS strongly supports this integrated approach and is pleased that the draft NICE guidance recognises the clinical value of ensuring those detected with VFFs receive timely osteoporosis treatment.	Thank you for your comment. It has been considered by the committee.

Comment number	Consultee	Section number	Comment	Response
57	Consultee 4 The Royal Osteoporosis Society	Not specified	**Workforce Planning and Clinical Oversight** In advocating for AI, we must also address the workforce implications. The ROS urges adequate planning to ensure radiographers and other clinicians can support the roll-out of AI detection tools. Importantly, AI is not a replacement for expert judgement – it should be an adjunct to, not a substitute for, radiographers review. We are reassured that the draft NICE guidance emphasises AI must be used alongside clinical judgment, not as an autonomous diagnostic. In practice, any fracture flagged by software will still require verification by a radiographers or appropriately trained clinician. Sufficient radiography capacity is therefore critical. If an AI system generates a high volume of alerts (including some false positives), this could increase image review workload and potentially strain radiology and radiography services. It is essential to plan for this by ensuring we have enough radiologists or reporting radiographers to handle AI outputs within acceptable timeframes. Similarly, FLS teams must be resourced to act on the AI findings. An influx of newly identified vertebral fracture patients will mean more clinic appointments, DXA scans, and treatments initiated by FLS. We note that expert feedback to NICE highlighted uncertainty about how many new referrals or treatments might result from opportunistic AI detection, especially given the current variation in FLS provision across the NHS. ROS stresses that to realize AI's benefits, we need a robust clinical pathway: from scan interpretation (radiology) through to fracture prevention care (FLS). This requires investing in the multidisciplinary workforce, ensuring we have the radiologists to interpret AI results and the FLS nurse specialists, doctors, and therapists to manage the resulting increase in patients at risk. In summary, effective implementation will depend on workforce and capacity planning so that AI-driven findings translate into real-world preventive care rather than unaddressed alerts.	Thank you for your comment. NICE acknowledges the point on broader system readiness and the need for a prepared workforce, but notes that this is not within the remit of the HealthTech programme. However, the committee concluded that the impact of introducing the technologies on the workforce represents a key evidence gap and further evidence should be generated for all of the technologies.
58	Consultee 4 The Royal Osteoporosis Society	Not specified	**Operational Barriers and the Need for Standardisation** As a charity working closely with osteoporosis services, the ROS is aware of operational and funding barriers that arise when new technologies are not yet recommended by NICE. Even highly promising AI tools can face difficulty in adoption; local commissioners and hospitals tend to be hesitant to invest in innovations without formal NICE guidance. This means that, until now,	Thank you for your comment. NICE acknowledges the points on broader system readiness and the need for guidelines and frameworks for implementing AI technologies (for example, covering validation, training for

Comment number	Consultee	Section number	Comment	Response
			some AI-driven fracture detection systems have seen limited uptake in the NHS, as they fall into a gap (awaiting evaluation and approval). We have observed that lack of an official recommendation can translate into funding obstacles, slow procurement, and uncertainty among clinicians about which AI products meet quality and safety standards. The ROS itself has encountered these challenges when trying to support early use of AI to improve fracture care. We are therefore very pleased that NICE is addressing this area through an early assessment and conditional recommendation. This will help legitimise the use of certain AI tools and could unlock funding streams (for example, through NHS England's AI mandate or local innovation budgets) to implement them. Looking ahead, we encourage NICE to explore standardisation measures to support consistency in how AI is deployed for fracture detection. Given multiple AI software products exist (with varying performance), it would be helpful for the system if there were clear standards or even a shortlist of approved technologies. We note that the draft guidance has already identified four specific AI technologies judged to be promising. Providing clarity on which tools are recommended – and under what conditions – will guide NHS providers in adopting solutions that are evidence-based. Moreover, standard setting (for example, defining minimum performance requirements, common metrics for fracture identification, or integration protocols) could ensure a uniform quality of care across regions. Consistency is key: patients across the country should benefit from similar levels of AI-enhanced detection, rather than a postcode lottery of different software or approaches. The ROS suggests that NICE, perhaps in collaboration with professional bodies, consider issuing best-practice guidelines or frameworks for implementing AI in this context (covering validation, training for staff, result reporting formats, information for patients etc.). Such standardisation would build confidence among clinicians and patients, and help embed AI in a sustainable, safe way. We also believe that clear endorsement of certain technologies will encourage industry and healthcare providers to focus on solutions that meet NICE-approved criteria, fostering a cohesive approach to AI in fracture prevention.	staff, result reporting formats), but notes that this is not within the remit of the HealthTech programme. However, NICE's adoption and implementation team will explore potential collaboration opportunities, including with the Royal Osteoporosis Society and other professional organisations, once the guidance is published.

Comment number	Consultee	Section number	Comment	Response
59	Consultee 5 Society & College of Radiographers	1.3 Can be used with evidence generation	'Using AI technologies can help reduce variation in practice' although there is a risk with the current model of only some sites using the technology that variation in practice will be exacerbated by sporadic implementation / access to this technology.	Thank you for your comment. NICE guidance on medical technologies does not carry a mandate for NHS implementation but could result in more commissioning by NHS trusts of technologies that could address an unmet need by NHS trust. An analysis of the impact of using AI technologies on the NHS care pathway during the 3-year evidence generation period may help to address this concern.
60	Consultee 5 Society & College of Radiographers	1.3 Can be used with evidence generation	'should monitor costs associated with its use in populations with a lower prevalence of osteoporosis' who would monitor and how? is it realistic that departments would be able to do that given that earlier in the same section it is noted that there is uncertainty around the cost of some of the technologies.	Thank you for your comment. Local commissioners will have access to the cost of a technology, which represents the key upfront expenditure for the trust. Monitoring this cost against the benefit of identifying additional VFFs is recommended because the committee noted that opportunistic detection of VFFs may be less or not cost-effective in populations at a lower risk of osteoporosis. Key information needed to monitor the costs associated with use in populations at a lower risk of osteoporosis also includes the number of true positive and false positive findings (please also see the evidence generation plan). The quoted statement about the uncertainty around costs relates to the uncertainty of the cost to implement the technology in general.
61	Consultee 5 Society &	3.5 Innovative aspects	'The clinical experts highlighted that many AI technologies have been adopted across the NHS and that AI offers significant potential for improving care.'	Thank you for your comment. It has been considered by the committee.

Comment number	Consultee	Section number	Comment	Response
	College of Radiographers		Yes there is significant potential for improving care however the number of technologies adopted for medical imaging across the NHS is still quite low in comparison to other countries, with hundreds of products available, but little data, evidence, or post market surveillance information available about performance of the various models once implemented in real world settings.	

Configuration of diagnostic accuracy

Comment number	Consultee	Section number	Comment	Response
62	Consultee 1 Annalise.ai	1.3 Can be used with evidence generation	We refer to section: 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 The Annalise device is highly configurable. Operating points (or threshold settings) can be modified at an organisation level to ensure the proportion of cases flagged for review are appropriate and can be adjusted to optimise for positive predictive value where requested, to ensure a low rate of false positives.	Thank you for your comment. Additional functions and configuration settings of the technologies in scope are described in section 2 of the assessment report. Where available, the EAG's evidence review included the configuration settings of the technology in the evidence. But, the EAG noted that there was generally poor reporting of configuration across the included studies.
63	Consultee 3 Bonescreen	3.6 Innovative aspects	If a technology exposes the setting to change the operating point of performance, we argue that the algorithm is no longer "fixed" as stated in the preceding sentence. We also state that changing the operating point is a very consequential step and should not be performed at a centre's level without a thorough sub-group level clinical evaluation.	Thank you for your comment. The EAG explained that it assumed that configuration of diagnostic accuracy settings as stated in the Instructions for Use would have been reviewed by a notified body. The EAG noted that there was generally poor reporting of configuration across the included studies.

Economic evaluation

Comment number	Consultee	Section number	Comment	Response
64	Consultee 2 Technical University of Munich (affiliated with Bone-screen)	Not specified	Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence? I think a 1-year calculation is too short. I would suggest to do this for three to five years, similar to the work done for oQCT: https://pubmed.ncbi.nlm.nih.gov/37576975.	Thank you for your comment. The EAG reviewed the paper by Ruhling et al. 2023 and noted that it compared screening for osteoporosis using opportunistic quantitative CT (derived from non-dedicated routine CT) versus DEXA. The study cited a period of 3 to 5 years as being optimal for best demonstration of reductions in fracture and increases in bone mineral density, and the likelihood of side effects or treatment failures being reported to be the greatest during this period. The EAG clarified that for its model it considered a 1-year time horizon appropriate to capture the immediate impact of opportunistically detecting a vertebral fracture and plausibility of benefit when using AI technologies in addition to standard care (see section 3.21 of the guidance) It considered that due to a lack of data to model longer term impacts, any longer-term results would be too uncertain. The committee noted the 1-year time horizon as a limitation of the economic modelling and concluded that longer-term modelling would be needed in the future to reduce this uncertainty when the recommendations are reviewed after evidence generation.
65	Consultee 2 Technical University of Munich	3.16 Clinical parameters	This is realistic, may be even high compared to other literature outside NHS. PMID: 34648040	Thank you for your comment. The EAG reviewed the paper by Loffler et al. 2022 which summarised the retrospective detection of vertebral fractures (grade 1, 2 or 3) across 718 patients undergoing

Comment number	Consultee	Section number	Comment	Response
	(affiliated with Bone-screen)			cancer staging and angiography CT in a single hospital in Germany between 2008 and 2017. No AI was used in this study. It reported that 30.5% (219/718) showed 1 or more osteoporotic vertebral fractures; with only 24.7% (54/219) included in the original CT reports. The study reported medication (non-steroidal anti-inflammatory drugs, metamizole, opioids, immunosuppressive drugs, glucocorticoids, vitamin D and bisphosphonates) taken in the 3 months prior to the index CT scan, however, it did not report management following detection of vertebral fractures.
66	Consultee 5 Society & College of Radiographers	3.16 Clinical parameters	(reporting radiographer without AI assistance) - also radiologist.	Thank you for your comment. In the economic model the EAG assumed that original reporting of an image conducted in the AI arm would be the same as the standard of care arm; therefore wouldn't alter the relative costs as it is applied to both arms. The EAG used the costs associated with interpreting and reporting the imaging without the use of AI for both arms with a separate cost for radiologist review for all cases identified by the AI (intervention arm) or radiographer (comparator arm). The EAG noted that a secondary review may not be required in all cases (see section 6.2 of the assessment report). Whilst the proportion of images undergoing secondary review was not varied in sensitivity analysis, the impact of changing this proportion can be

Comment number	Consultee	Section number	Comment	Response
				inferred from the sensitivity analysis which varied the proportion of patients flagged as having a VFF undergoing an additional spinal X-ray.
67	Consultee 5 Society & College of Radiographers	3.18 Cost parameters	'The EAG calculated the cost per scan for each technology, which included product subscription, implementation, integration, training and maintenance costs' there are also costs associated with post-market surveillance / collection of benefits metrics data.	Thank you for your comment. The costs associated with post-market surveillance or further data collection have not been incorporated into the costings for each technology within the EAG's economic model, as these costs would typically be covered by the company and as such be subsumed in the price paid by the NHS for the technology.

Requests for clarification of the wording

Comment number	Consultee	Section number	Comment	Response
68	Consultee 5 Society & College of Radiographers	1.1 Can be used with evidence generation	Does 'consideration of risk groups' mean that depending on individual circumstances then the AI may be used for the individuals within the risk groups (beyond the indications outlined in the instructions for use)?	Thank you for your comment. All technologies should only be used as outlined in their Instructions for Use. Risk prioritisation (including based on the recommendations in NICE's guideline) should be done in the eligible populations only. Section 3.1 of the guidance outlines the risk factors for VFFs.
69	Consultee 5 Society & College of Radiographers	1.4 Why the committee made these recommendations	'process an image (the failure rate)' It may be useful to note that there are other reasons for failure. It appears to the reader that the failure rate is only related to inability to process an image when there are multiple potential reasons for failure.	Thank you for your comment. The EAG clarified that within the evidence, the definition of failure varied (including unable to upload, unable to process X-ray, incorrectly classifying as unsupported anatomical regions, unsuccessful model inference, unable to analyse) and was not

Comment number	Consultee	Section number	Comment	Response
				reported in some cases. The EAG noted that there was a lack of detail reported in the published evidence to determine whether there is a common characteristic, cause, or image which fails to be processed or analysed by the AI technologies. Section 3.16 of the guidance has been amended to clarify the different reasons for failure.
70	Consultee 5 Society & College of Radiographers	3.3 Unmet need	'effective pharmacological and non-pharmacological treatment options for managing symptomatic ' this EVA is for opportunistic detection - what are the treatment options in this case - are they different for asymptomatic VFF?	Thank you for your comment. The wording of the respective section has been amended to clarify that all people with a VFF would usually experience symptoms.
71	Consultee 5 Society & College of Radiographers	3.4 Unmet need	'thousands of radiographic images are taken annually' - millions of radiographic images are taken annually in the NHS!	Thank you for your comment. The wording of the respective section has been amended to ensure factual accuracy.
72	Consultee 5 Society & College of Radiographers	3.4 Unmet need	'most remain undiagnosed' - if they remain undiagnosed then how do the experts know that? It is estimated that numbers are much higher?	Thank you for your comment. The committee heard that the prevalence of undiagnosed VFFs has been estimated from large European studies conducted in the past decade. These prevalence estimates are higher than the number of VFFs currently diagnosed in the NHS.
73	Consultee 5 Society & College of	3.8 Diagnostic accuracy	'suggests that a lot of VFFs remain undiagnosed' this is ambiguous, what is 'a lot' likely to represent, for example, tens, hundreds, thousands, hundreds of thousands?	Thank you for your comment. The committee heard that it is estimated that up to 70% of VFFs remain undiagnosed. Clinical experts referred to several studies, including the IMPACT study. However, the

Comment number	Consultee	Section number	Comment	Response
	Radiographers			committee heard that the figure 70% is based on findings from European studies (see also the response to comment 72). So, it was not certain whether this figure is still accurate. But, clinical experts confirmed that many VFFs still remain undiagnosed.
74	Consultee 5 Society & College of Radiographers	3.12 Choice of technologies	This section focuses on the use of 'frontal-view chest xray' in the NHS but it should be noted that is an adapted technique with the majority of projections postero-anterior (back-view).	Thank you for your comment. The wording of the respective section has been amended to ensure factual accuracy.
75	Consultee 5 Society & College of Radiographers	3.23 Risks	'had led to a significant increase in workload' what was that increase, it is not clear what 'significant' is.	Thank you for your comment. The quoted statement reflects the experience of a specialist committee member in 1 trust. The committee recognised that the impact of introducing the technologies on healthcare professional workload represents a key evidence gap and recommended evidence generation to address it.
76	Consultee 5 Society & College of Radiographers	3.23 Risks	especially if the first review was done by a radiographer who has not had specialist musculoskeletal training.' this statement is misleading and inaccurate. Radiographers without specialist musculoskeletal training would not review MSK images, it would be outside the scope of practice of a radiographer if not a trained MSK reporting radiographer.	Thank you for your comment. The quoted statement refers to the case when a radiographer without specialist MSK training reviews a radiograph involving the spine that has been taken for an unrelated condition. The committee noted that this is likely to be the case due to the opportunistic use of the technologies.

Comments in support

Comment number	Consultee	Section number	Comment	Response
77	Consultee 3 Bonescreen	Not specified	Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence? Yes, the summaries and interpretations of the committee are valid.	Thank you for your comment.
78	Consultee 4 The Royal Osteoporosis Society	Not specified	**Broad Support for AI in Vertebral Fracture Detection** The Royal Osteoporosis Society (ROS) welcomes and broadly supports the use of artificial intelligence (AI) to improve the detection of vertebral fragility fractures (VFFs) in clinical practice. As a patient-focused charity, we are encouraged by innovations that can identify spinal fractures earlier and thereby prompt timely osteoporosis treatment. Vertebral fractures are often missed – up to 70% remain undiagnosed – yet they are the most common osteoporotic fractures and a strong predictor of future fractures, including hip fractures. Indeed, more than 55% of hip fracture patients show evidence of prior vertebral fractures, highlighting a major missed prevention opportunity. We therefore support AI technologies that opportunistically flag these “hidden” fractures on routine scans, as this could identify thousands of patients who would otherwise go untreated. By spotting vertebral fractures that clinicians might overlook, AI can help ensure those at risk receive therapy to strengthen their bones, avoiding future fractures and associated costs. In summary, ROS is enthusiastic about the potential of AI to enhance fracture detection as a means to improve patient outcomes and reduce the burden of osteoporosis.	Thank you for your comment. It has been considered by the committee.
79	Consultee 4 The Royal Osteoporosis Society	Not specified	**Conclusion** In conclusion, the Royal Osteoporosis Society is strongly supportive of NICE’s efforts to evaluate and guide the use of AI in detecting vertebral fragility fractures. From a patient perspective, this represents a significant opportunity to diagnose osteoporosis earlier, prevent fractures, and reduce the long-term impact of undiagnosed spinal fractures. We have outlined our key points of support and caution: the importance of early detection linked with FLS-led prevention, the need for adequate workforce planning to verify AI results and manage new patients, the call for further research before full roll-out, and the value of addressing practical barriers through standardisation and guidance. Finally, we wish to clarify that the	Thank you for your comment. It has been considered by the committee.

Comment number	Consultee	Section number	Comment	Response
			ROS does not intend to comment on the technical specifics of the various AI algorithms under review; we recognise that other expert organisations and clinicians are better placed to provide detailed technical input. Our focus, as ever, is on the patient and system benefits.	
80	Consultee 5 Society & College of Radiographers	Not specified	Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence? Yes given the limits of available evidence.	Thank you for your comment.
81	Consultee 5 Society & College of Radiographers	Not specified	Are the recommendations sound and a suitable basis for guidance to the NHS? Yes, careful consideration and reasoning for decisions provided in the consultation.	Thank you for your comment.
82	Consultee 5 Society & College of Radiographers	3.16 Clinical parameters	'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. ' an excellent point.	Thank you for your comment. It has been considered by the committee.

HealthVCF vs HealthOST

Comment number	Consultee	Section number	Comment	Response
83	Consultee 6 Nanox AI	Not specified	From a regulatory perspective, HealthOST represents a natural evolution of HealthVCF: same fundamental clinical objective, improved analytical capabilities, and strengthened compliance with the current EU regulatory landscape.	Thank you for your comment. The committee considered the additional evidence on HealthOST and concluded that the technology can be used as an option for the opportunistic detection of

Comment number	Consultee	Section number	Comment	Response
			The table below shows HealthVCF vs HealthOST highlighting the similarities between them. **Aspect** / **HealthVCF V1.2 (MDD)** / **HealthOST V1.1 (MDR)** / **Evolution Summary** **Regulatory Framework** / MDD 93/42/EEC, Class IIa / MDR 2017/745, Class IIa (Rule 11) / Transitioned to MDR compliance with full Annex II/III documentation, GSPR checklist, PMCF, and PMS upgrades. **Intended Purpose** / Supports clinician decision-making by detecting and quantifying vertebral compression fractures (VCFs) from CT scans. / Provides qualitative & quantitative analysis of spine from CT scans, including vertebral height loss (VCFs) and bone mineral density (osteoporosis/osteopenia). / Same VCF detection core + expanded functionality for bone density assessment and opportunistic osteoporosis screening. **Target Population** / Adults undergoing chest or abdominal CT scans. / Patients ≥50 years old undergoing CT for any indication, covering T1–L4 (height loss) and T11–L4 (bone attenuation). / Similar population focus, now with explicit age criteria and added osteoporosis risk group. **Outputs** / Fracture grading and vertebral labeling report. / Vertebral labeling, % height loss, HU bone density measurement, threshold-based alerts. / Expanded outputs for broader musculoskeletal assessment. **Core Technology** / AI-based detection of VCFs; PACS/DICOM integration; cloud-based processing. / AI-based detection of VCFs + BMD estimation; PACS/DICOM integration; cloud and on-prem options; configurable thresholds. / Maintains same architecture; algorithm extended for new metrics. **Risk Profile** / Class IIa, not significant risk; hazards mainly misclassification or missed detection; mitigated under ISO 14971. / Class IIa, hazards include misclassification of VCFs or osteoporosis, false negatives/positives; mitigated under ISO 14971:2019 with added cybersecurity and usability controls. / Similar core hazards; new risks addressed via MDR-aligned risk management. **Clinical Evidence** / Original CER (2019) + PMS updates; limited use data; no PMCF; performance verified in previous releases. / CER supported by retrospective study (265 subjects), literature on similar devices, SOTA review; PMCF plan; MDR Article 61(10) justification for clinical evidence sufficiency. / Evidence base expanded; performance validated for both VCF and bone density features. **PMS & PMCF** / PMS per legacy procedure; PSUR every 2 years; no PMCF (justified). / MDR-compliant PMS plan; PMCF plan established; performance/safety monitored with CE-mark surveillance. / Strengthened post-market data collection per MDR requirements. **Version History** / V1.0 → V1.1 (filter adjustment) → V1.2 (secondary capture capability). / First EU version V1.1 developed in 2022; builds on VCF detection legacy with new osteoporosis module. / Continuity in core VCF functionality; functional expansion for osteoporosis detection. **Market Focus** / Initially population health/triage for fractures. / Opportunistic screening for chronic musculoskeletal conditions (VCF, osteoporosis). / Strategic focus shift aligns with market demand for comprehensive population health tools. The fracture detection module has not been changed between HealthVCF and HealthOST. In addition, the statistical performance comparison below shows distinct similarity between the two VCF models.  The performance table above shows the diagnostic accuracy for both products side by side as performed on a statistically analogous but distinct data set as typically regulatory bodies frown on re-using identical data sets.	VFFs while further evidence is generated. The wording of the recommendations has been amended to reflect this.

Comment number	Consultee	Section number	Comment	Response
			Ultimately, as the committee looks ahead to the future, products that are MDD regulated such as HealthOST have passed a more rigorous regulatory threshold than those with CE extensions of the MDD regulatory framework. With regard to clinical publications: The NHS funded 2018 ADOPT clinical trial, utilized HealthVCF as the AI product for VCF detection, as that was the only regulated AI from Nanox at the time. Getting direct feedback from the clinical sites using HealthVCF as part of the trial, in addition to other sites, was what drove Nanox to create HealthOST as a next version of the software solution to incorporate the insights from the clinicians and thus create a superior product. Dr. Kassim Javaid, the Oxford Osteoporosis expert who was the PI for the study, was an advocate of the new version, and intended to incorporate it into the ADOPT trial once it had regulatory approval. However, the ADOPT trial ended in February of 2025, whereas HealthOST received CE approval in June 2025 and MHRA approval in the end of July 2025, due to the complexities created by the EU MDD to MDR changes. Nevertheless, we pursued other avenues for data collection with two publications to date which we are attached and will be uploaded to the NICE docs site as well, that validate its performance and utility in detecting patients with VCF, in exactly the same way HealthVCF is utilized. • Queen Elizabeth University Hospital, Glasgow - Mathew et al. 2025 • University of Toronto & Hospital St Michaels form Toronto - Mckeag et al. 2024 The poster from Scotland was published at the BIR AI on the 2024, and also mentioned on the BIR 2025. The paper from St Michaels and Toronto University was published at ECR 2025 in Toronto on multiple digital platforms (like Diagnostics). The NICE committee members have put in tremendous efforts to evaluate the various vendors and AI products for Osteoporosis, but, as expected, they focused on those with regulatory approval for obvious clinical reasons.	

Comment number	Consultee	Section number	Comment	Response
			In light of the incredible importance of the NICE recommendations coupled with the recent regulatory approval of the newer version of the Nanox AI bone solution and the evidence attached, we kindly request that the NICE team evaluate the data we have provided regarding HealthOST when they can and given the added information, consider adjusting the recommendation regarding HealthOST. Meanwhile, we plan to maintain commercial availability of HealthVCF in the UK market until such time that Nice will issue a report recommending HealthOST.	