HealthTech Programme

Artificial intelligence technologies to help detect fractures on X-rays in urgent care

Draft Guidance themed comments

Agreement with guidance

Comment number	Name	Section number	Comment	Response
1	Consultee 5 The Mid Yorkshire NHS Trust	1.6 What evidence generation and research is needed	Agree but there must not be a reliance on Al to provide a diagnosis it is to supplement ones confidence in image interpretation	Thank you for your comment.
2	Consultee 5 The Mid Yorkshire NHS Trust	3.23 Risks	This is an important considerations as not all acute admission in ED for MSK imaging are for injury - some can demonstrate tumours etc	Thank you for your comment.
3	Consultee 6 The Royal Osteoporosis Society	3.2 Patient and carer considerations	The noting of these concerns and reassurance that due to IR(ME)R these technologies cannot be used without human interpretation is very welcome.	Thank you for your comment.
4	Consultee 8 Society & College of Radiographers	1.6 What evidence generation and research is needed	Good point.	Thank you for your comment.
5	Consultee 8 Society &	3.8 System impact	Excellent points and also evidence generation plan. The committee discussion section of the draft guidance offers	Thank you for your comment.

Comment number	Name	Section number	Comment	Response
	College of Radiographers		reassurance of fair and detailed discussions, thank you.	

Clinical evidence

Comment number	Name	Section number	Comment	Response
number 6	Consultee 4	Not specified	Having read the draft guidance document for consultation it is clear that there is a significant lack of non-biased prospective evidence to support to use of AI in the NHS. None of the supporting studies were UK or NHS based, nor did they include the reader demographic to which this document is aimed at providing AI assistance for. The results of the studies are questionable, particularly in relation to the image interpretation accuracy of the reader without the assistance of AI. If this were the case I would be worried about the level of training those readers had previously received and the missed fracture rates. As the studies were retrospective it is unattainable to accurately conclude the cost effectiveness of AI, due to no cost calculations of missed fracture management or increase in orthopedic referrals. The accuracy of each AI provider has not been established in the literature provided,	Thank you for your comment which the committee has considered. The committee acknowledged the limitations of the evidence base (see sections 3.4 to 3.6), but concluded that the risk of implementing AI assistance in the NHS while more relevant evidence is generated is low (see sections 3.20 to 3.24). This assessment evaluated the use of AI as a decision aid for healthcare professionals in urgent care, where clinical experts explained that hot reporting is not always possible (see section 2.2). However, the EAG and clinical experts agreed that the benefit of AI may be different in centres where hot reporting is in place. This has been added to the section on 'system benefit'.
			nor has the impact when this technology fails. If, as the committee has rightly stated	

Comment	Name	Section number	Comment	Response
			readers become reliant upon the technology for initial image review, when the technology fails, this will pose a significant increase in reader inaccuracy and influx of false negative/false positive initial reviews. There are many centers in the UK where by "hot reporting" is available and therefore AI is surplus to requirement in the urgent care setting. As this document states AI is to be supportive to the radiology report and not replace it, what would be its purpose in the "hot reporting" setting? This could be a waste of NHS money in these centers. To conclude, more studies which are designed to provide accurate correlating evidence for the use of AI in UK NHS centers is required to be able to support the use of AI.	
7	Consultee 6 The Royal Osteoporosis Society	1.6 What evidence generation and research is needed	More evidence is needed about the human readers experience to understand who the comparators with AI are in practice. Will the project compare radiologists and reporting radiographers with AI in the studies or ED clinicians and advanced practitioners? Potential bias needs to be eliminated in the studies where better results are shown from studies sponsored by industry compared to non-sponsored studies- there needs to be non-industry sponsored evaluation of the technology. <u>This paper</u> raises concerns on this	Thank you for your comment which the committee has considered. The comparator in this early value assessment is interpretation of X-rays by healthcare professionals in urgent care without AI assistance (see section 2.6). Section 1.6 specifies that evidence generation and more research is needed on the diagnostic accuracy of fracture detection in urgent care by healthcare professionals with and without the help of AI technologies.

Comment	Name	Section number	Comment	Response
			This paper also noted that AI detected ' non- existent fractures which was much less common in human raters' Clarity is also needed on whether AI can detect non-fracture pathology such as tumors	Thank you for highlighting the paper by Husarek et al. We note that the authors did not find an effect of industry funding on the analysis of Al-aided versus unaided interpretation (see Table 5 in the paper). Although the authors report that stand-alone AI detected more non-existent fractures than unaided human interpretation, the use of stand-alone AI is not within the scope of this assessment. When considering only Al-aided versus unaided interpretation, there was no difference in specificity. Detection of non-fracture pathologies is outside the scope of this assessment, however recommendation 1.6 includes "detection of or failure to detect clinically significant non-fracture-related conditions by healthcare professionals with and without the help of AI technologies" to examine whether incidental findings are affected by use of AI
8	Consultee 6 The Royal Osteoporosis Society	3.8 System impact	This is disputed. The Husarek paper cited above suggested AI does in fact generate a relatively high proportion of false positives	Thank you for your comment which the committee has considered. The paper found a higher false positive rate for the use of stand-alone AI versus unaided human interpretation. The use of stand-alone AI is not within the scope of this assessment. The scope of this assessment is to compare the accuracy of X-ray interpretation by



Comment number	Name	Section number	Comment	Response
				healthcare professionals in urgent care with and without the help of AI technologies.
				The findings of the Husarek et al. systematic review were consistent with the conclusions of the committee. It reports significant improvement in sensitivity of fracture detection, without reduced specificity, by healthcare professionals assisted by AI software compared with unassisted interpretation (see Table 5 in the paper).
9	Consultee 6	3.24 Risks	As noted previously, this is not supported by	Thank you for your comment which the
	The Royal		the Husarek paper	committee has considered. Please review
	Society			previous response.

Comment	Name	Section number	Comment	Response
number				
10	Consultee 8	1.6 What	An SoR consultant radiographer network	Thank you for your comment which the
	Society &	evidence	member comments:	committee has considered. Please see the
	College of	generation and	Clinical effectiveness & Accuracy:	response to comments 7 and 8.
	Radiographers	research is	The absence of reader experience and	
		needed	characteristics in studies remains an	The committee recognised that more
			unknown and further study with comparison	evidence is needed to clarify the clinical and
			across difference groups of readers with	cost-effectiveness of the AI technologies if
			clarification of reader characteristics in both	used in the NHS, and the outcomes of
			background and experience is required.	interest are outlined in section 1.6 and in the
			Inter-rater studies of human readers also	evidence generation plan. The guidance will
			demonstrate variation. We have just	be revisited once this evidence has been
			reviewed 37000 reporting radiographer	generated to provide final recommendations.
			audited reports with mean sensitivity and	
			specificity of 98% and 99% respectively.	The use of AI in assisting healthcare
			Results centred around 90% for both AI and	professionals to detect fractures on X-rav
			'reader' groups are concerning. The clinical	images could improve the accuracy of
			effectiveness & accuracy require significant	fracture diagnoses in urgent care, which
			clarification going forward and the positivity	would result in fewer complications and
			around current results should be tempered.	additional appointments from missed
			Husarek et al. (2024) published in October of	fractures. The potential benefits of using AI
			this year draws attention to potential bias in	alongside clinical interpretation are stated in
			sponsored studies demonstrating relatively	section 2.4.
			better results for Sen/Spec when compared	
			to unsponsored studies. This has not been	
			acknowledged in these documents.	
			Husarek et al. also draw attention to the fact	
			that there was a 'tendency to detect non-	
			existent fractures, which was much less	
			common in human raters without AI support	
			.' While we might expect an overall net	
			positive outcome of AI in reducing	
			unmanaged fractures, recalls etc. the over	

Comment number	Name	Section number	Comment	Response
			management and over imaging of incorrect calls must also be considered and I have not seen this considered in the recommendations and may be an addition to Decision Problems and Economic considerations. I have not seen any significant consideration of where AI is intended to bring real added diagnostic value i.e. Is AI actually better at detecting certain injuries typically missed by humans or is it just more consistent across certain reader groups and if so with what characteristics? There remain, at least for now, areas of weakness for detection software and these areas must be identified and widely communicated as weaknesses and clearly voiced in all recommendations.	

Place in care pathway

Comment number	Name	Section number	Comment	Response
11	Consultee 3	Not specified	As I understand AI will be used in the ED setting to assist ED practitioners in making an x-ray diagnosis before being reported officially by a radiologist or reporting radiographer. The ED department in the NHS hospital where I work does not have diagnostic quality reporting screens for the	Thank you for your comment which the committee has considered. The aim of this assessment was to determine whether the technologies set out in the scope to help detect fractures on X-rays in urgent care have the potential to be

Comment	Name	Section number	Comment	Response
			ED practitioners to view images on and the training on diagnosing x-rays seems very scant so it is unsurprising that errors in diagnosis are made, and I can only presume that this is a national problem. These are very basic issues which should be addressed before AI is considered as a solution.	clinically and cost effective for use in the NHS, and to identify evidence gaps. Changes to workplace facilities and staff training are beyond the scope of this assessment.
12	Consultee 6 The Royal Osteoporosis Society	1.6 What evidence generation and research is needed	Consideration must also be given to the workforce implications caused by an expected increase in fracture detection, given the current rates of vacancies for radiology and radiography positions - as stated in the supporting evidence. All reports should still be reviewed by a radiologist or reporting radiographer given that Al software will fail to work on a proportion of scans and the potential risk for false positives.	Thank you for your comment which the committee has considered. The committee considered that the risk of increase in referrals due to false positive results was low, because the evidence suggests that the use of AI alongside clinical interpretation does not decrease specificity (see sections 3.4 and 3.24). The EAG note that the assessment did not identify any evidence that evaluated the potential impact on service outcomes of the technology, including whether the introduction of the technology could impact on staff resource. In principle, the evidence suggested that the technology may increase the number of fractures identified on first presentation but may subsequently reduce the number of re-attendances and onward referrals. AI would not replace the definitive X-ray reports, which are made by a radiologist or reporting radiographer. So, there are safety

Comment	Name	Section number	Comment	Response
				net systems in place to identify any potential fractures that may have been missed by Al- assisted interpretation in urgent care. This is stated in the 'managing the risk of use in the NHS with evidence generation' box and in section 3.21.
13	Consultee 6 The Royal Osteoporosis Society	1.6 What evidence generation and research is needed	If this is part of the strategy then this calls into question what value AI can add, if a healthcare professional needs to interpret the result before viewing the AI result.	Thank you for your comment which the committee has considered. The use of AI in assisting healthcare professionals to detect fractures on X-ray images could improve the accuracy of fracture diagnoses in urgent care, which would result in fewer complications and additional appointments from missed fractures. The potential benefits of using AI alongside clinical interpretation are stated in section 2.4. Interpreting X-rays without AI assistance before viewing the AI results is suggested as a method to maintain X-ray interpretation skills for urgent care staff and to minimise the risk of over-reliance on the AI technologies.
14	Consultee 6 The Royal Osteoporosis Society	2.2 Clinical need and practice	This is essential and must remain. Especially as a fail safe for the Al over calling fractures and non fracture pathology, if properly resourced by training sufficient workforce to manage	Thank you for your comment. This guidance does not affect NICE's guideline on non- complex fractures, which remains in place.

Comment number	Name	Section number	Comment	Response
15	Consultee 6 The Royal Osteoporosis Society	3.4 Diagnostic accuracy	It is important to note here that whilst the AI is helping emergency physicians (and presumable advanced practice emergency nurses) to identify fractures they previously missed in ED at the time of patient presentation, these will still need reporting by a radiologist/radiographer.	Thank you for your comment which the committee has considered. It is stated in the 'managing the risk of use in the NHS with evidence generation' box and in section 3.21 that AI technologies do not replace the definitive radiology review.
16	Consultee 6 The Royal Osteoporosis Society	3.21 Risks	Clarity on the value of the technology is needed here if this this is the case. If AI is to be used alongside existing care systems then this suggests the cost effectiveness is not thoroughly understood as a benefit. This should be part of the further assessment/research	Thank you for your comment which the committee has considered. The use of AI in assisting healthcare professionals to detect fractures on X-ray images could improve the accuracy of fracture diagnoses in urgent care, which would result in fewer complications and additional appointments from missed fractures. The potential benefits of using AI alongside clinical interpretation are stated in section 2.4.
17	Consultee 6 The Royal Osteoporosis Society	3.23 Risks	As noted previously, this mitigation calls into question the value that AI can add if human interpretation is still needed.	Thank you for your comment which the committee has considered. Please see response to comment 13.

Comment	Name	Section number	Comment	Response
number				
18	Consultee 8 Society & College of Radiographers	2.4 Clinical need and practice	A reporting radiographer network and special interest discussions at SoR offer anecdotal evidence that AI tools may currently increase the number of unnecessary referrals to fracture clinics. It is hypothesised that is due to lack of training data including a range of images with adapted technique or alternative imaging views. An example has been provided by a consultant radiographer network member in response to this consultation: 'We had an AI fracture tool on trial for just over a year. Internal audit demonstrated:- High number of FP calls especially relating to Neck of Femur (NOF) images (for which it was originally designed) and Paediatric images. (15%) Low number of FN cases around 5% Our in-house review deemed it a safe and effective triage tool but of limited use within the actual imaging department. ED declined to take up the permanent installation of the tool as they believed it would provide limited additional protection. That said our board isn't particularly busy with MSK work and there is no backlog so everything gets reported the following day. Might be useful in a bigger hospital.'	Thank you for your comment. The evidence reviewed by the committee indicated that there was a low risk of decreased specificity when using AI to aid X-ray interpretation. However, this guidance will be reviewed after data is collected following the evidence generation plan, and final guidance issued taking into account the diagnostic accuracy and system impacts observed in NHS practice. This assessment evaluated the use of AI as a decision aid for healthcare professionals in urgent care, where clinical experts explained that hot reporting is not always possible (see section 2.2). However, the EAG and clinical experts agreed that the benefit of AI may be different in centres where hot reporting is in place. This has been added to the section on 'system benefit'.

Implementation

Comment	Name	Section number	Comment	Response
19	Consultee 5 The Mid Yorkshire NHS Trust	1.6 What evidence generation and research is needed	Will significantly impact on the future training of reporting radiographers / advanced Radiographer practitioners in terms of prospective reporting practice and experience.	Thank you for your comment, which the committee has considered. It is possible that the availability of AI as a decision aid could impact the training and development of expertise of radiologists and reporting radiographers. This is outlined in the section "impact on workforce", and it is suggested that healthcare professionals interpret X-rays before viewing AI results to mitigate this issue.
20	Consultee 5 The Mid Yorkshire NHS Trust	1.6 What evidence generation and research is needed	There is the potential to reduce ED attendances with GP patients being able to access open access acute MSK pathways to access diagnostics initially and then get an immediate report. Patients with normal findings / soft tissues injuries can be discharged from radiology with treatment advice and no need to attend ED services	Thank you for your comment. This guidance covers the use of AI to aid X-ray interpretation in all urgent care centres. Direct referrals from GPs to centres that are not considered part of urgent care are outside the scope of the assessment.
21	Consultee 6 The Royal Osteoporosis Society	Not specified	Overall response. The ROS welcomes the trialling of AI technologies to help detect fractures, particularly where the technology can help to improve the detection of vertebral fractures - 70% of which are missed. However, additional consideration needs to be given, as further evidence is gathered, to potential negative workforce implications as well as the challenges of integrating AI within existing NHS systems, processes and IT. Communication has not been considered as	Thank you for your comment which the committee has considered. Implementation costs, including the time and staff to integrate AI within existing NHS systems, are included in the evidence generation plan and outlined in recommendation 1.6. Clinical experts have advised that AI results are DICOM objects and would not be visible on the NHS app. They would be available to radiology when making the definitive report, which is what would be shared with patients.

Comment	Name	Section number	Comment	Response
			a risk in this document, but there are associated risk if patients are entitled to view their results on the NHS app - for example if this includes AI generated false positive results. Consideration must be given as to whether access to these results is appropriate and how these risks will be mitigated. Ultimately the committee must consider whether imaging, reviewed and reported at the time of patient presentation by a trained human, would in fact be clinically more effective or safe than AI. And so if AI can in reality safely mitigate the workforce and resourcing pressures recognised in the supporting documentation.	As outlined in section 2.2, hot reporting is rarely possible, so this was not considered as a comparator in this assessment. However, the assessment evaluated the use of AI as a decision aid, so interpretation of X-rays would still be done by a trained human.
	Consultee 6 The Royal Osteoporosis Society	while more evidence is generated	take place once this evidence has been gathered to inform the final guidance.	After the evidence generation period finishes, the technologies will be reviewed again by NICE in a full guidance process, which will also include consultation with stakeholders.
23	Consultee 6 The Royal Osteoporosis Society	1.6 What evidence generation and research is needed	Consideration must also be given here to how effectively each of the four technologies integrates with existing IT systems and processes. Each hospital will have different IT systems and processes in place	Thank you for your comment. Implementation costs, including the time and staff to integrate AI within existing NHS systems, are included in the evidence generation plan and outlined in recommendation 1.6.
24	Consultee 6 The Royal	3.8 System impact	This section should also consider the potential workforce implications for higher rates of detection and identification of	Thank you for your comment which the committee has considered. The committee concluded that it was unlikely that AI use

Comment	Name	Section number	Comment	Response
number				
	Osteoporosis Society		fractures. As stated above 'AI technologies for fracture detection on X-rays cannot be used without human interpretation'	would lead to an increase in the rate of false referrals. Since all reports should be reviewed by a radiologist or radiographer for definitive reporting, the use of AI would not affect demand for this service. The EAG note that the assessment did not identify any evidence that evaluated the potential impact on service outcomes of the technology, including whether the introduction of the technology could impact on staff resource. In principle, the evidence suggested that the technology may increase the number of fractures identified on first presentation but may subsequently reduce the number of re-attendances and onward referrals.
25	Consultee 7 East Midlands Imaging Network (EMRAD)	1.6 What evidence generation and research is needed	A feedback we got, especially from more experienced radiologists, was that the Al tool eroded their confidence and felt like a threat to their professional authority.	Thank you for your comment, which the committee has considered. It is stated in the 'managing the risk of use in the NHS with evidence generation' box and in section 3.21 that AI technologies do not replace the definitive radiology review.

Comment number	Name	Section number	Comment	Response
26	Consultee 8 Society & College of Radiographers	1.6 What evidence generation and research is needed	An SoR consultant radiographer network member comments: Communication I have not seen communication raised as a potential risk. Communication risk within a system such as a hospital site where the AI analysis/overlay may be visible to a diverse multi- professional group in for example an inpatient stay following an acute admission and where the formal radiology report or electronic notes may not be available to the individual viewing the image and AI output. Effective communication across such a large and diverse workforce including temporary and new staff will be a risk. Communication to the public: Is there a potential responsibility to acknowledge the AI finding and potentially document either agreement or disagreement in patients notes and the formal report? Patients and service users may have their results delivered to their NHS App. Should this include the AI result and if not can we defend this?	Thank you for your comment, which the committee has considered. Implementation considerations such as communication and training of hospital workforces is outside the scope of this assessment. Clinical experts note that the AI technology is a clinical decision support tool, and as such the interpreter holds responsibility for the final decision whether or not they agree with the AI assistance. Please also see response to comment 21.

Cost-effectiveness estimates

Comment number	Name	Section number	Comment	Response
27	Consultee 5 The Mid Yorkshire NHS Trust	3.9 Model structure	Consideration on accuracy of identifying notoriously difficult MSK fractures such as scaphoid would be useful as these fracture carry the highest litigation.	Thank you for your comment, which the committee has considered. The EAG note that, based on the current evidence base, it was not possible to only evaluate the clinical and cost effectiveness of using the technology to aid with the diagnosis of scaphoid fractures. However, it did consider accuracy for hand and wrist fractures, which included scaphoid fractures. The evidence generation plan has been updated to specify fracture type as a subgroup for diagnostic accuracy studies.
28	Consultee 5 The Mid Yorkshire NHS Trust	3.11 Model structure	Agree that equivocal results for hip fracture would require further imaging before surgical intervention	Thank you for your comment.
29	Consultee 6 The Royal Osteoporosis Society	3.15 Plausibility of cost effectiveness	Clarity needs to be given on the comparator here. If radiologists consultant grades are the assumed cost of reporting x-rays this would be an overestimate of cost. Many radiographer grade reporters undertake this in practice, therefore comparable cost effectiveness of AI may be overestimated.	Thank you for your comment. As the initial presentation at A&E, and the X-ray, was common to both AI-assisted and unassisted diagnosis, the EAG excluded the costs of these from the analysis (see the external assessment report, section 8.3.8). Similarly, the definitive radiology report would be required for both arms so this was also excluded. The EAG's analysis compared assisted and unassisted reading with the explicit assumption that the same grade of urgent care staff would be performing the read. The only cost difference therefore was the cost of the AI algorithm (licence fee plus associated fixed costs). The EAG did not consider a scenario where a more

Comment number	Name	Section number	Comment	Response
				highly trained reader may be replaced by a lesser skilled reader with AI assistance.
30	Consultee 7 East Midlands Imaging Network (EMRAD)	1.6 What evidence generation and research is needed	This is crucial, because there are additional costs depending on the Trusts infrastructure (not only integration cost but also the cost of uninstalling if it is no longer needed)	Thank you for your comment. Implementation costs, including the time and staff to integrate AI within existing NHS systems, are included in the evidence generation plan and outlined in recommendation 1.6.
31	Consultee 8 Society & College of Radiographers	1.6 What evidence generation and research is needed	An SoR consultant radiographer network member comments: 1)Scenario analyses does not take into consideration radiographer grade/level reporting and the use of consultant grade in the cost analysis and this may lead to an overestimation of cost efficiency in the UK system. 2)The potential of using AI results to capture inefficient use of imaging and population radiation exposure in the context of low fracture/positive finding incidence allowing for referrer/department/local/national level comparison has not been considered.	Thank you for your comment. As the initial presentation at A&E, and the X-ray, was common to both Al-assisted and unassisted diagnosis, the EAG excluded the costs of these were from the analysis (see the external assessment report, section 8.3.8). Similarly, the definitive radiology report would be required for both arms so this was also excluded. Scenario analyses 5 and 6 assumed that the radiograph would be read by a junior/trainee radiologist or a consultant radiologist to examine the potential benefit of reducing time to interpret radiographs, but staff grade did not form part of the base-case analysis. Please also see response to comment 29. This assessment considers the use of AI to help detect fractures in cases where an X-ray has already been ordered and considered necessary. Auditing the level of X-ray usage and population radiation exposure is beyond the scope of the assessment. The EAG state that no evidence was identified that evaluated the

Comment number	Name	Section number	Comment	Response
				impact of using the technology on the use of further imaging.
32	Consultee 8 Society & College of Radiographers	1.6 What evidence generation and research is needed	It should be clear that is in addition to costs for definitive reporting by reporting radiographer or radiologist.	Thank you for your comment which the committee has considered. Costs for definitive reporting will be required whether or not AI is being used to assist interpretation, so the cost of AI quoted refers only to additional cost over current practice.

Clarification of wording

Comment	Name	Section number	Comment	Response
33	Consultee 5 The Mid Yorkshire NHS Trust	3.7 Children and young people	If there is the suspicion of SPA i.e. fracture presentation does not fit mechanism of injury these patients must be discussed with Radiology immediately and a formal radiological report obtained. The reliance of Al in these cases is not recommended and the subtle findings would be best suited to a trained specialist in this field.	Thank you for your comment. Clinical experts highlighted that there is limited evidence for use of AI where there is suspicion of physical abuse. We have added the following clarification: "Clinical experts noted that cases of suspected physical abuse would be referred for further review by radiology and that the use of AI would not change the escalation pathway in these cases."
34	Consultee 5 The Mid Yorkshire NHS Trust	1.6 What evidence generation and research is needed	Please note many centres use virtual fracture clinics.	Thank you for your comment which the committee has considered. Fracture clinics include both virtual and in-person clinics.

Comment	Name	Section number	Comment	Response
35	Consultee 5 The Mid Yorkshire NHS Trust	1.6 What evidence generation and research is needed	to protect patients from unnecessary radiation further imaging triggered by Al should only be accepted/justified after a formal radiological report has been provided.	Thank you for your comment which the committee has considered. This assessment evaluated use of AI as a decision aid alongside clinician interpretation. It would not be possible to determine if a test ordered by a clinician was influenced by AI assistance, and the scarcity of hot reporting means that currently referrals for further imaging are often made before the radiology report is available. Clinical experts note that IR(ME)R requires justification by referring clinician and practitioner before exposure, so it would not be possible for AI to make a referral for further imaging without human review. The EAG note that no evidence was identified that assessed the impact of using AI as an aid to decision-making on further imaging. This section has been clarified to state that "There is a low risk that that using AI technologies to help detect fractures on X- rays may increase fracture clinic referrals and requests for further imaging such as CT or MRI"
36	Consultee 5 The Mid Yorkshire NHS Trust	1.6 What evidence generation and research is needed	Considerations of overcall of positive findings in paediatrics which may trigger more imaging and normal variants should be considered.	Thank you for your comment. Age is included as a subgroup in the evidence generation plan to examine the accuracy of AI-assisted interpretation in different populations, including paediatrics.
37	Consultee 5 The Mid	1.6 What evidence generation and	I work at a trust with a prevalence of rickets and osteomalacia due to the diversity of the patient population	Thank you for your comment. This has been added to section 3.27.

Comment number	Name	Section number	Comment	Response
	Yorkshire NHS Trust	research is needed		
38	Consultee 5 The Mid Yorkshire NHS Trust	2.2 Clinical need and practice	The majority of acute MSK reporting at my trust is performed by a team of advanced radiographer practitioners with minimal input from the radiologists	Thank you for your comment which the committee has considered. Section 2.2 notes that the reporting can be delivered by a range of different healthcare professionals, which includes radiologists, radiographers or other trained reporters.
39	Consultee 6 The Royal Osteoporosis Society	2.3 Clinical need and practice	 The evidence needs to be clear here on who these fractures are missed by - for example, radiology, radiographer hot reporting or ED staff. The Hussain paper cited found that 25% of missed diagnoses were missed by imaging professionals but more than 1/3 were due to incomplete assessments and so not the fault of imaging professionals. The Kuo paper stated AI performed as well as clinicians 	 Thank you for your comment which the committee has considered. The Hussain paper reports that fractures were the most common condition where diagnostic error occurred in emergency department settings. 437/1,007 (43%) diagnostic errors in fractures were due to inappropriate response to diagnostic imaging. Section 2.3 has been amended to clarify this point. The passage of Kuo relevant to this section states: "Missed or delayed diagnosis of fractures on radiographs is a common diagnostic error, ranging from 3% to 10%."

Comment	Name	Section number	Comment	Response
number				
40	Consultee 8 Society & College of Radiographers	1.1 Can be used while more evidence is generated	What is the definition of adults being used in the recommendations - from what age?	Thank you for your comment which the committee has considered. The wording of the recommendation has been revised in line with updated information on the technologies' indications to better clarify the age range.
41	Consultee 8 Society & College of Radiographers	1.3 Can only be used in research	We suspect that there are other CE marked Al tools that are being tested in clinical practice. We wonder if the expert committee have any further knowledge or intelligence about? Perhaps consider whether to offer advice that any other software products / developed over next 2 years, intended for use of identifying MSK fractures, should be used for research and risks managed. (potentially funded through company/research/non-core NHS funding).	 Thank you for your comment which the committee has considered. A search for appropriate technologies was done during scoping through extensive external and internal engagement. The minimum requirement criteria to identify the technologies are as follows: The technology meets appropriate regulatory requirements (for example, UKCA or CE mark) appropriate to the function. The technology falls within a priority area. The technology is currently being used in the NHS or being planned for uptake in the coming 6 months. Please see the early value assessment interim statement section 2.13. NICE cannot provide guidance on technologies which are not yet available to the NHS. This guidance only covers the technologies named in section 1.

Comment	Name	Section number	Comment	Response
42	Consultee 8 Society & College of Radiographers	1.6 What evidence generation and research is needed	and inequalities in provision of service (for example, out of hours)	Thank you for your comment which the committee has considered. This has been added.
43	Consultee 8 Society & College of Radiographers	2.6 The comparator	In some circumstances those healthcare professionals will be assisted by preliminary clinical evaluation (PCE) from a diagnostic radiographer, prior to the issues of a reporting radiographer or radiologist definitive clinical report. The provision of PCE is variable across organisations.	Thank you for your comment, which the committee has considered. Section 2.1 has been amended for clarity.
44	Consultee 8 Society & College of Radiographers	2.7 The comparator	Might the reference standard also be provided by registrar level radiologist? To note that a reporting radiographer might also be working at the level of enhanced, advanced or consultant level Allied Health Professional. It might be simpler to state that the reference standard is based on the radiologist or reporting radiographer interpretation and report.	Thank you for your comment which the committee has considered. The reference standard for this assessment was defined in the scope as the consultant radiologist or reporting radiographer interpretation and report. However, we recognise that other radiology staff may be involved in providing the definitive report in practice. The evidence generation plan has been amended to state 'experienced radiologist' rather than consultant when defining the reference standard. The EAG note that the evidence review used some flexibility to include evidence with a reference standard that was considered comparable (e.g. definitive report by another specialist trained reporter).

Evidence generation plan

Comment number	Name	Section number	Comment	Response
45	Consultee 1 MILVUE	3.1 3.1 Evidence gaps and ongoing studies	Milvue added upcoming studies taht could support the evidence gap.	Thank you for your comment. We have included the status about your ongoing studies in our evidence generation plan.
46	Consultee 1 MILVUE	3.1 3.1 Evidence gaps and ongoing studies	Use of Augmented Intelligence for the Interpretation of Bone Standard X-rays Prescribed by the Emergency Department (IMMEDIAT Urgences) (IMMEDIAT) - NCT05882435 Medico-economic prospective multicentric randomized cluster multiple period cross- over study with 6 alternate periods (3 with AI, 3 with usual organization) of 1 month in each ED. Diagnostic error rate, Time between x-ray and first diagnostic by radiologist and the emergency physician, Time between x-ray and final diagnostic by the emergency physician, Rate of X-rays interpretation by radiologist without delay, Number of all radiologist at 30 days, Time spent in the ED by the patient, Number of patients invited to come back in the ED, 30 days morbidity, Total cost from the hospital viewpoint Inclusion is completed	Thank you for your comment. Please review previous response.
			Analysis is ongoing, estimated for Q2 2025	

Comment	Name	Section number	Comment	Response
number				
47	Consultee 1 MILVUE	3.1 3.1 Evidence gaps and ongoing studies	Standalone model performance assessment of the performance of TechCare Alert, a software using convolutional neural network techniques for assisting physicians in the diagnosis of bone fractures and elbow joint effusion in adults and pediatric population. Standalone image-level performance of Al compared to the ground truth Patient-level and finding-level performances are secondary objectives. Subgroups analysis will be conducted for age, gender, and ethnicity. Study is completed, under publication process	Thank you for your comment. Please review previous response.
48	Consultee 1 MILVUE	3.1 3.1 Evidence gaps and ongoing studies	Before and after: Evaluating the real-life impact of AI in emergency radiography "Retrospective study of real-life discrepancies between radiologists and emergency physicians in ED, before the use of AI and at year 1 and 2 after the introduction of Milvue Suite. Delta of number of discrepancies between the non AI and the AI periods, and their caracterization based on the impact for the patient (recall, follow-up, no action)." Inclusion is completed Analysis is ongoing, estimated for Q2 2025	Thank you for your comment. Please review previous response.

Comment	Name	Section number	Comment	Response
number 49	Consultee 1 MILVUE	3.1 3.1 Evidence gaps and ongoing studies	Enhancing fracture and elbow joint effusion detection in adults using AI : a multi-reader multicentric study Multi-Reader Multi-Case (MRMC)	Thank you for your comment. Please review previous response.
			retrospective study from external multicenter anonymized datasets. Case-level ROC-AUC difference of readers unaided and aided by AI was the primary endpoint. Interpretation time was also recorded. Subgroups analysis will be conducted for age, gender, and ethnicity.	
			 Study is completed, under publication process Abstract submitted to ECR 2025	
50	Consultee 1 MILVUE	3.1 3.1 Evidence gaps and ongoing studies	Enhancing fracture and elbow joint effusion detection in pediatric population using AI : a multi-reader multicentric study Multi-Reader Multi-Case (MRMC) retrospective study from external multicenter anonymized datasets. Case-level ROC-AUC difference of readers unaided and aided by AI was the primary endpoint. Interpretation time was also recorded. Subgroups analysis will be conducted for age, gender, and ethnicity.	Thank you for your comment. Please review previous response.
			Study is completed, under publication process	

Comment number	Name	Section number	Comment	Response
51	Consultee 2	2.2 Effectiveness in different subgroups	Fractures can occur as a result of a non- accidental injury pattern. As such caution should be exercised when managing children, young people, vulnerable adults and those who can not give a clear history as the diagnostic tool and it's interpretation only form one part of the patient episode.	Thank you for your comment. NICE has taken this into account by requesting information on clinical effectiveness on different populations. The physician in charge of management would dictate how patient history is considered. No further action required.
52	Consultee 6 The Royal Osteoporosis Society	2.2 Costs associated with implementing the Al technologies	Consideration must also be given here to how effectively each of the four technologies integrates with existing IT systems and processes. Each hospital will have different IT systems and processes in place. This section should also consider the potential workforce implications for higher rates of detection and identification of fractures. As stated in our feedback to the early value assessment document	Thank you for your comment. NICE expects this circumstance to be accounted when companies disclose the costs of implementation and maintaining the use of the technology across the NHS. No further action required.
53	Consultee 6 The Royal Osteoporosis Society	3.3 Real-world prospective study and embedded qualitative study	This is key if we are to understand the benefits and risk across different organisations and set ups.	Thank you for your comment. No further action required.
54	Consultee 6 The Royal Osteoporosis Society	3.4 Real-world prospective study	This should include a comparator - understanding what standard practice looks like and who is delivering it.	Thank you for your comment. NICE has included healthcare professional unassisted by AI technology as a comparator and consultant radiologist or reporting radiographer interpretation and report as a ground truth. This is stated in section 3.3 of the evidence generation plan. No further action required.



Comment number	Name	Section number	Comment	Response
55	Consultee 7 East Midlands Imaging Network (EMRAD)	Not specified	General Comment Kettering General Hospital (KGH), one of the Partner Trusts in the East Midlands Imaging Network (EMRAD) did a proof of concept trialling RB Fracture with a focus on their A and E department. The scope of practice was for fracture only and did not cover dislocations. The reason for trialling AI for fracture detection was due to the following; - No 24-hour service due to cost - To ascertain whether an AI system could potentially give a sound second opinion for A&E clinicians - To assess whether using AI would reduce cost in comparison to Reporting Radiographers The Audit was for a 6 month period; in the first three months, KGH flagged accuracy gaps while the next three months trialled the upgrade that reflected required adjustments. The first reporter reported outstanding A&E exams and recorded RB's accuracy (TP/TN/FP/FN). The Second reporter repeated this blind to	Thank you for your comment. The EAG note that a report of this evaluation was identified in the EAG's evidence review and evidence from the study was therefore considered in the assessment.
			ine first reporter's record. and the a third	

Comment number	Name	Section number	Comment	Response
			reporter was used if there was a disagreement. The Audit Result A total of 319 exams were audited with a 32/68 positive to negative ratio Sensitivity – 93.7% Specificity – 93.4% Accuracy – 93.5%	
56	Consultee 7 East Midlands Imaging Network (EMRAD)	6	Cultural barriers and priorities mismatch If the key decision makers and users of the technology are not on board (do not see the value of using AI or consider it a threat or have unaddressed misconceptions or do not consider it a priority), regardless of the benefits of the AI, there will be resistance.	Thank you for your comment. We appreciate your opinion on this matter. NICE is committed to embrace technologies that have proved to have clinical and cost- effective benefits. This is to ensure users and clinicians can have reassurance when using the new technology. No further action required.
57	Consultee 8 Society & College of Radiographers	2.1 Clinical and service outcomes	In addition to data collection focused on reduced misdiagnosis rates and the impact of missed fractures, it may be useful to also collect data related to any areas of increased misdiagnosis rates and the impact of incorrect diagnosis (that might be a result of false positives too).	Thank you for your comment. NICE has requested to collect the information you have noted in section 3.4 "Clinical outcomes associated with missed diagnosis or misdiagnosis, for example, unnecessary treatments, further diagnostic procedures, or complications from misdiagnosis, ideally with quality-of-life impact." No further action required.

Comment number	Name	Section number	Comment	Response
58	Consultee 8 Society & College of Radiographers	2.2 Costs associated with implementing the AI technologies	The education and training of staff with regards to the implementation, validation, and ongoing surveillance/audit of AI tools are important factors that should be considered an integral part of the infrastructure.	Thank you for your comment. NICE has requested to collect the information you have noted in section 3.4 "Costs associated with maintaining the infrastructure needed for the AI technologies, including software, hardware and staff training." No further action required.
59	Consultee 8 Society & College of Radiographers	3.2 3.2 Data sources	It may be necessary to include costing of this process in the data collection for establishment of necessary infrastructure if this is determined to be a necessary ongoing data collection.	Thank you for your comment. NICE has requested to collect the information you have noted in section 3.4 of the evidence generation plan. No further action required.
60	Consultee 8 Society & College of Radiographers	3.3 3.3 Evidence collection plan	to establish a baseline?	Thank you for your comment. NICE has requested to collect data on ground truth by gathering data on the diagnostic accuracy of consultant radiologist or reporting radiographer interpretation and report (section 3.3). No further action required.
61	Consultee 8 Society & College of Radiographers	3.3 3.3 Evidence collection plan	also consider use of the AI tool and difference in performance related to the range of different X-ray imaging systems in use in urgent care centres; including Computed Radiography (CR), Digital Radiography (DR) / the range of different manufacturers and specifications.	Thank you for your comment. NICE has indicated this as a potential barrier for data collection in section 6 and it has suggested that careful consideration about the trusts selected to participate should be given to maximise robust data collection.
62	Consultee 8 Society & College of Radiographers	3.3 Concordance study to assess diagnostic accuracy	it should be clear that this is related to X-ray imaging. Scanning tends to be the terms associated with cross-sectional imaging - CT and MR.	Thank you for your comment. NICE has amended the evidence generation plan in line with this comment about wording.
63	Consultee 8 Society &	3.3 Concordance study to assess	Ground truth may also be subject to discrepancy / error. In current practice if there is discrepancy then review by a third	Thank you for your comment. NICE has suggested different options to collect data

Comment number	Name	Section number	Comment	Response
	College of Radiographers	diagnostic accuracy	party may take place - is that recommended here also?	on the reference standard. No further action required.
64	Consultee 8 Society & College of Radiographers	3.3 Concordance study to assess diagnostic accuracy	for consistency, should that be urgent care centres?	Thank you for your comment. NICE has indicated that data collection is expected in centres that best represent "urgent care centres" in the NHS. No further action required.
65	Consultee 8 Society & College of Radiographers	3.3 Concordance study to assess diagnostic accuracy	for consistency, this was referred to earlier in the text as 'ground truth' - should that be reference standard in the earlier text too?	Thank you for your comment. NICE has amended the wording to improve consistency in the term.
66	Consultee 8 Society & College of Radiographers	3.3 Concordance study to assess diagnostic accuracy	consider that a review by reporting radiographer or radiologist will support evidence analysis at this point - to determine if re-attendance was due to occult fracture with no error by either AI or human.	Thank you for your comment. NICE has indicated that data collection on diagnostic accuracy as well as clinical outcomes on misdiagnoses is required. No further action required.
67	Consultee 8 Society & College of Radiographers	3.3 Concordance study to assess diagnostic accuracy	Good point.	Thank you for your comment.
68	Consultee 8 Society & College of Radiographers	3.3 Real-world prospective study and embedded qualitative study	Staff perspectives can also be captured with focus groups.	Thank you for your comment.
69	Consultee 8 Society & College of Radiographers	3.3 Real-world prospective study and embedded qualitative study	Why thematic analysis specifically? Content analysis may be more appropriate in some cases. It might be simpler to state an appropriate and proportionate approach to data analysis.	Thank you for your comment and suggestion. After further review, thematic analysis is no longer included in the evaluation plan. This change was made to simplify the approach and ensure it aligns with the scope of the evaluation.

Comment	Name	Section number	Comment	Response
number				
70	Consultee 8	3.4 Real-world	including staff education and training prior to	Thank you for your comment. NICE expects
	Society &	prospective	implementation	staff to be educated and trained on the use
	College of	study		of the technology prior to the use of it so that
	Radiographers	-		data collection is reflective of real clinical
	_			practice and meaningful for decision making.
				No further action required.
71	Consultee 8	3.4 Real-world	ongoing audit/service evaluation	Thank you for your comment. NICE has
	Society &	prospective		suggested that a proportion of cases can
	College of	study		ideally be peer reviewed as part of the
	Radiographers			prospective real world study. No further
	_			action required.
72	Consultee 8	3.4 Information	would it be possible to also include details of	Thank you for your comment. NICE has
	Society &	about the	demographics included in the training data	indicated on section 3.4 that information on
	College of	technologies	set - may highlight gaps between training	how the technology was developed should
	Radiographers		and validation data sets, relevant for specific	be provided. No further action required.
			populations.	
73	Consultee 8	4 4 Monitoring	Will there also be mechanisms for clinical	Thank you for your comment. NICE has
	Society &		staff to raise safety concerns with NICE?	indicated that "new safety concerns" need to
	College of			be reported to NICE during the monitoring
	Radiographers			phase, this is stated in section 4 of the
	-			evidence generation plan. No further action
				required.