External Assessment Group's Protocol

GID-HTE10059 Artificial intelligence technologies to aid the opportunistic detection of vertebral fragility fractures: Early Value Assessment

Produced by: Newcastle External Assessment Group (EAG)

Authors: Kim Keltie, Lead Healthcare Scientist, The Newcastle upon Tyne Hospitals NHS Foundation Trust (NuTH);

Elliot Blacklock, Administrator, NuTH;

Humayra Dervin, Clinical Scientist, NuTH;

Paula Leslie, Pre-registrant Clinical Scientist, NuTH;

Katrin Bangel, Pre-registrant Clinical Scientist, NuTH;

Emma Belilios, Project co-ordinator, NuTH;

Fiona Beyer, Principal Research Associate; Newcastle

University;

Gurdeep Sagoo, Health Economics Group Co-Lead, Newcastle University;

Luke Vale, Professor of Health Economics, London School of Hygiene and Tropical Medicine; NuTH;

Andrew Sims, EAG Director, NuTH

Correspondence to: Dr Andrew Sims, Northern Medical Physics and Clinical Engineering, NMPCE (Medical Physics, NCCC Level 2) Freeman Hospital High Heaton, Newcastle upon Tyne NE7 7DN nuth.nmpce.hta@nhs.net

Date of issue: 12/02/2025

Acknowledgement: NICE Information Services team provided the literature search terms which were used in the initial scoping of this topic (Appendix).

Contents

Abbreviations

Term	Definition
AI	artificial intelligence
DXA	dual-energy X-ray absorptiometry
EAG	External Assessment Group
FLS	Fracture Liaison Service
IFU	instructions for use
NICE	National Institute for Health and Care Excellence
VFF	vertebral fragility fracture

1 Background

The topic has been identified by NICE for Early Value Assessment. The objective of an Early Value Assessment is to identify promising technologies in health and social care where there is significant need and potentially enable early conditional access to these while informing further evidence generation. A rapid appraisal of the evidence is undertaken in line with the early value assessment interim statement (NICE, Dec 2022) to determine if these offer plausible value to the NHS. The evidence developed will show if the expected value of the technologies is realised and will be used to inform a subsequent final NICE evaluation when a decision will be made on the routine use of the technologies in the NHS.

1.1 Objectives

The purpose of this evidence assessment is to summarise the evidence for the health technologies included in the Final Scope. The aim is to evaluate the clinical-effectiveness and cost-effectiveness, identify evidence gaps, and highlight any risks associated with the potential use of these technologies in the NHS while further evidence is generated. It should be noted that the purpose of the review is not to compare the technologies with each other. Based on the scope developed by NICE, the following specific primary objectives are proposed:

- To identify, review and summarise evidence of the clinical effects and safety of artificial intelligence (AI) technologies as an aid or adjunct to radiologists to enhance opportunistic vertebral fragility fractures (VFF) detection, when compared with the standard of care.
- To identify, review and summarise the economic evidence of artificial intelligence (AI) technologies as an aid or adjunct to radiologists to enhance opportunistic VFF detection, when compared with standard of care.
- To develop an early economic model to provide an initial assessment of the potential cost-effectiveness of AI technologies for VFF when compared with standard of care.
- To summarise information on the capacity, capabilities and practicalities of implementing AI technologies for VFF detection.

- To identify important evidence gaps and outline what data could be collected to address them.

1.2 Care Pathways

Vertebral fragility fractures (VFFs) are fractures in the spine that occur following a fall from standing height or less but they can also occur spontaneously as a result of day-to-day activities involving very little trauma or stress. VFFs are the most common type of fragility fractures – most are caused by osteoporosis, which results from bone weakness, with an incidence of 12% in women aged 50–79 years, increasing to 20% in women over 80 years old (Curtis et al., 2016). Despite this, they are often detected incidentally on radiographic images, but up to 70% remain undiagnosed (Dalal et al., 2022), leading to missed opportunities for early intervention.

VFFs are a strong predictor of future osteoporotic fractures, particularly hip fractures, with over 55% of hip fracture patients having had prior evidence of VFFs (Gonnelli et al., 2013). VFFs are associated with an eightfold increase in age-adjusted mortality. The costs and morbidity of osteoporotic VFFs and hip fractures are significant. Because of the high prevalence of osteoporosis, particularly in the elderly population, timely identification and management of VFFs are critical for preventing further fractures and reducing healthcare costs.

Current Guidelines and Diagnostic Pathway

Currently NICE Clinical Guideline on Osteoporosis assessing the risk of fragility fracture (CG146, 2012 updated 2017) recommends that risk assessments are undertaken in patients who are considered high-risk of VFFs. Assessment of fragility fractures should be considered in all women aged 65 years and over and all men aged 75 years and over, and should also be considered in women aged under 65 years and men aged under 75 years in the presence of certain risk factors, for example: previous fragility fracture; current use or frequent recent use (of more than 3 months) of oral or systemic glucocorticoids; history of falls; family history of hip fracture; other causes of secondary osteoporosis; low body mass index (less than 18.5 kg/m2); smoking; alcohol intake of more than 14 units per week for men and women. Fragility fracture risks are not routinely assessed in people aged under 50 years unless they have major risk factors (for example, current or frequent recent use of oral or systemic glucocorticoids, untreated premature menopause or previous fragility fracture), because they are unlikely to be at high risk.

Radiologists and reporting radiographers are well placed to diagnose VFFs on any imaging modality that includes the spine (including x-ray, barium studies, computed tomography, and magnetic resonance imaging). NICE Clinical Guideline (CG146, 2012 updated 2017) on assessing the risk of VFFs in patients with osteoporosis does not mention the use of Artificial Intelligence technologies to aid detection of VFFs.

The EAG note that there are a number of methods currently used to define VFFs, including the semi-quantitative method described by (Genant et al., 1993). This system involves the visual recognition of a loss of vertebral body height on a lateral projection combined with careful assessment of the vertebral endplates to diagnose a fracture and to exclude non-fracture vertebral deformity. A severity assessment can also be made as follows:

- Grade 1 (Mild) fracture 20-25% loss of vertebral body height.
- Grade 2 (Moderate) fracture 25-40% loss of vertebral body height.
- Grade 3 (Severe) fracture >40% loss of vertebral body height.

Additionally, the radiological guidance for the recognition and reporting of osteoporotic VFFs issued by The Royal College of Radiologists (2021) recommends that when an opportunistic osteoporotic VFF is diagnosed, it is important to look for and comment on:

- The presence/absence of additional VFFs
- The level of fractures
- The severity of fractures
- Evidence of canal/cord/cauda equina compromise (which would necessitate urgent discussion and onward referral)
- Avoidance of repeat imaging examinations because of unnecessary additional ionising radiation applied (The Royal College of Radiologists, 2021).

With regards to reporting, the time taken between imaging acquisition and image interpretation can vary, depending on Trust protocols, clinical priority (urgent or routine), and access to hot reporting (accident and emergency referrals are generally

reported within 24 hours). AI technology may identify additional patients with VFFs and append this data to the local radiology information system. However, processes/protocols will need to be developed within local Trusts to ensure radiological actionable reports and alerts are issued for these patients to access appropriate onward assessment and care. Ideally, radiology reports should include clear statements that there is a newly diagnosed VFF and the referrer must ensure arrangement of appropriate assessment for osteoporosis and fragility fracture and subsequent treatment and management.

Treatment and Management of VFFs

People that have been diagnosed with osteoporosis may be referred to a Fracture Liaison Service (FLS), which comprises multidisciplinary healthcare professionals (nurses, doctors and administrative staff). An audit (undertaken via freedom of information survey) in 2021 reported that 51% of the NHS Trusts (63/123) had access to an FLS (Royal Osteoporosis Society, 2021). In 2019, only 19% of radiology departments in the NHS shared a defined pathway for patients with VFFs (Royal Osteoporosis Society, 2021).

Treatment for osteoporosis involves treating and preventing fractures and using medicines to strengthen bones. The accurate and timely identification of VFFs is needed to enable starting an effective bone health management strategy involving bisphosphonates (denosumab, teriparatide parathyroid hormone 1-34) supported by dietary advice, vitamin D supplementation and exercise fall prevention programmes (Tu et al., 2018).

2 Decision Problem

The decision problem is described in the Final Scope and summarised here.

2.1 Population

Opportunistic detection of VFFs will include patients who have had a radiographic image (X-ray, CT, MRI, DXA) involving the spine taken for reasons other than vertebral fragility fraction (VFF) detection. Subgroup analysis by presence of specific

risk factors (as described in the Final Scope) and by type of VFF will be considered where data allows.

2.2 Intervention

Artificial intelligence (AI) tools listed in the Final Scope used as a decision aid for enhancing radiographic image interpretation in detecting vertebral fragility fractures assessment (which would be subsequently confirmed by radiologist or radiographer with musculoskeletal special interest):

- Annalise Enterprise (CXR) (Annalise.AI)
- Annalise Container (CXR) (Annalise.AI)
- BoneView (Gleamer)
- BriefCase-Triage (Aidoc Medical)
- CINA-VCF Quantix (Avicenna.AI)
- HealthVCF (Nanox AI)
- HealthOST (Nanox AI)
- IB Lab FLAMINGO (IB Lab)
- TechCare Spine (Milvue).

National screening is outside the scope of this assessment.

2.3 Comparators

The comparator is standard care where incidental identification of VFFs may occur during radiologist or radiographer interpretation of the radiograph without AI assistance, usually within 24 hours of the image being taken within the routine workflow of the NHS organisation. The EAG will consider evidence where other healthcare professionals have interpreted the radiograph and will liaise with Experts to consider the generalisability of the evidence to the UK NHS. The EAG note that incidental identification of VFFs may vary by centre and compliance to the radiological guidance for the recognition and reporting of osteoporotic vertebral fragility fractures (The Royal College of Radiologists, 2021).

The reference standard when considering diagnostic accuracy of the AI would be a radiologist or radiographer with musculoskeletal interest who would interpret the radiograph to confirm the absence or presence of VFF.

2.4 Healthcare setting

Evidence from all healthcare settings will be considered within this Early Value Assessment, including radiological images interpreted by teleradiology reporting services.

2.5 Outcomes

The outcome measures to consider include:

Intermediate outcomes

- Measures of diagnostic accuracy to detect VFFs (which may include vertebral compression fractures)
- Accuracy when used by different healthcare professionals (radiologists, radiographers and other healthcare professionals)
- Failure rate or rate of inconclusive AI reports
- Number of missed fractures
- Rate of missed fracture-related further injury
- Proportion of people that need further imaging
- Intervention related adverse events
- Healthcare professional user acceptability of AI tools.
- Changes to clinical management.

Patient-reported outcomes

- Health-related quality of life

Costs and resource use

Cost of diagnosis for AI and standard of care

- Cost of the AI software
- Staff costs
- Type of healthcare professional (for example, radiographer, radiologist) interpreting the radiograph
- Time to produce a radiographic report
- Time to diagnosis or time to definitive radiology report.
- Training and implementation costs
- Other downstream costs for diagnosis or treatment

Costs consequent on diagnosis:

- Time to further referral or treatment
- Number of treatments and extent of treatments
- Number of hospital appointment/visits, including referrals to fracture clinics and orthopaedic assessment
- Number of hospital admissions

The EAG will extract and report outcomes for each AI technology separately.

3 Methods

3.1 Technology review

The EAG will review the standard request for information forms and instructions for use (IFU) submitted to NICE for each technology within scope in order to develop a technology summary. This will be supplemented by information from company websites and from peer-review publications. Indications and contraindications listed in each IFU will be considered, any evidence identified which has been undertaken in a contraindicated population will be excluded by the EAG. Any missing or incomplete information may be supplemented from information found in the public domain, for

example from company websites, as appropriate. Technology summary tables may be sent to each company to ensure accuracy of content.

3.2 Evidence review

Clinical and economic evidence provided by companies in scope will be supplemented by an independent literature search undertaken by the EAG.

3.3 Search strategy

The search strategy will be developed based on the literature search strategy shared by the NICE Information Specialist team during scoping (Appendix) and identified published literature reviews in the topic area (for example (Aggarwal et al., 2021)), optimised for the decision problem (for example including company and technology names listed in the Final Scope, and older device names as advised by the companies in their completed request for information). The search strategy will be applied to clinical and economic databases separately, using filters as appropriate to identify diagnostic accuracy studies as well as clinical and economic evaluations. The EAG will consider applying limits to the literature search (for example published in English language, date of publication) where appropriate.

Ongoing studies will be identified through searches in World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) and supplemented by information supplied by the companies.

3.4 Study selection

Titles and abstracts will be screened by a single reviewer and 10% sample checked by a second reviewer for relevance to the scope. For those deemed relevant to the scope, full papers will be retrieved and reviewed by a single reviewer for relevance to the scope. Where the technology used is undefined in the title and abstract the EAG will retrieve the full paper for review. Studies including technologies not listed in the Final Scope, or non-commercial products will be excluded. Studies and papers that do not explicitly name the technology in the full paper and have not been sent by the company (where the technology used can be inferred) will be excluded. Studies and results described in company submissions (that is, completed requests for information) which are not available in the public domain or are not provided to the EAG in confidence for checking the information against the source will be excluded

from the assessment. Any exclusions of full papers will have the reason for exclusion tabulated and checked by a second reviewer.

In instances where no evidence directly relevant to the scope is identified for a technology from the literature searching, the EAG may expand the elements of the scope and will consult with clinical experts to determine the generalisability of the included evidence and findings to the UK NHS.

3.5 Quality assessment strategy

Formal risk of bias assessment will not be completed. Discussion will be included in the EAG report on potential biases in included studies and how the risk of bias could affect key outcomes. The report will explicitly detail the potential sources of bias such as the main confounding factors and will comment on the generalisability of the results to clinical practice in the NHS.

3.6 Data extraction

Data will be extracted from included studies into a bespoke spreadsheet to enable descriptive statistics. Independent, second review of data extraction may be done subject to time and resource availability. Data points to be extracted include information about the study reference, setting, design, population characteristics, intervention characteristics and relevant outcomes as listed in the Final Scope. Any additional outcomes reported in the included evidence will be extracted, if time permits.

3.7 Methods of analysis and synthesis

Clinical evidence will be tabulated in a bespoke spreadsheet and narratively synthesised by technology. Methods and findings from included published economic evidence will be summarised in a tabular format and synthesised in a narrative review. Economic evidence from the perspective of the UK NHS and Personal Social Services will be presented in greater detail.

3.8 Use in the NHS

The EAG will consult with Clinical Experts to determine the patient pathway, clinical context and uptake of the interventions in the NHS. The EAG will consider the

relevance of routinely collected data (for example the Diagnostic Imaging Dataset, and Model Hospital) in support of the evaluation.

3.9 Economic modelling

The EAG will construct an economic model built in either Microsoft Excel or R Programming Language, which will be informed by published economic evaluations describing the diagnostic pathway. This may include learnings from NICE technology appraisal guidance on bisphosphonates for treating osteoporosis (TA464, 2017 updated 2019) and romosozumab for treating severe osteoporosis (TA791, 2022). The EAG will describe the appropriate characteristics of the model (for example structure, setting, input parameters, sources of data, assumptions). The structure of the model and parameters used to populate it will be informed by clinical evidence and economic evidence identified from the EAG review and advice sought from Clinical Experts regarding assumptions and parameter values where evidence is lacking. Targeted searches for economic model inputs may be considered where appropriate.

The EAG will explore the impact of different cost options supplied by companies on the economic model. Where appropriate, and if data and time allow, sensitivity analysis will be undertaken to explore uncertainty. These may include deterministic and probabilistic sensitivity analysis, scenario analyses and subgroup analyses focused on what are believed to be the key characteristics and population subgroups identified in the scope. Costs will be considered from an NHS and Personal Social Services perspective, consistent with the reference case framework (NICE Health Technology evaluations manual, 2022).

3.10 Gap analysis

Evidence gaps identified pertaining to the intermediate and final outcomes from the scope and those pertaining to the economic modelling will be summarised in tabular and narrative form. Key areas for evidence generation will be summarised in tabular form. Narrative text will also address missing clinical evidence for other parts of the scope, such as population, setting and comparators. The EAG will outline potential study designs to address specific research questions to address identified evidence gaps.

4 Handling information

The EAG will consider any data or evidence supplied by the companies or stakeholders involved. If the data meet the inclusion criteria for the review they will be considered. It may not be possible to include data received later than 20 February 2025.

Any 'commercial in confidence' data provided and specified as such will be highlighted in blue and underlined in the EAG Report. Any 'academic in confidence' data provided will be highlighted in yellow and underlined in the EAG Report. Any 'personally identifiable' data provided will be highlighted in pink and underlined in the EAG Report. Any 'confidential price agreements' data provided will be highlighted in green and underlined in the EAG Report. All confidential information, as identified above, will be redacted before publication on the NICE website.

If confidential information is included in any economic models produced, then a version using dummy data or publicly available data in place of confidential data will be provided.

5 Competing interests of authors

None.

6 References

Aggarwal, V., Maslen, C., Abel, R. L., Bhattacharya, P., Bromiley, P. A., Clark, E. M., Compston, J. E., Crabtree, N., Gregory, J. S., Kariki, E. P., Harvey, N. C., Ward, K. A., & Poole, K. E. S. (2021). Opportunistic diagnosis of osteoporosis, fragile bone strength and vertebral fractures from routine CT scans; a review of approved technology systems and pathways to implementation. *Ther Adv Musculoskelet Dis*, *13*, 1759720X211024029. https://doi.org/10.1177/1759720X211024029

Curtis, E. M., van der Velde, R., Moon, R. J., van den Bergh, J. P., Geusens, P., de Vries, F., van Staa, T. P., Cooper, C., & Harvey, N. C. (2016). Epidemiology of fractures in the United Kingdom 1988-2012: Variation with age, sex, geography, ethnicity and socioeconomic status. *Bone*, *87*, 19-26. https://doi.org/10.1016/j.bone.2016.03.006

Dalal, G., Bromiley, P. A., Kariki, E. P., Luetchens, S., Cootes, T. F., & Payne, K. (2022). Understanding current UK practice for the incidental identification of vertebral fragility fractures from CT scans: an expert elicitation study. *Aging Clin Exp Res*, *34*(8), 1909-1918. https://doi.org/10.1007/s40520-022-02124-w

Genant, H. K., Wu, C. Y., van Kuijk, C., & Nevitt, M. C. (1993). Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*, *8*(9), 1137-1148. https://doi.org/10.1002/jbmr.5650080915

Gonnelli, S., Caffarelli, C., Maggi, S., Rossi, S., Siviero, P., Gandolini, G., Cisari, C., Rossini, M., Iolascon, G., Letizia Mauro, G., Crepaldi, G., Nuti, R., & Group, B. S. (2013). The assessment of vertebral fractures in elderly women with recent hip fractures: the BREAK Study. *Osteoporos Int*, *24*(4), 1151-1159. https://doi.org/10.1007/s00198-012-2119-2

Royal Osteoporosis Society. (2021). *State of the Nation report: Vertebral fracture identification in 2021*. https://strwebprdmedia.blob.core.windows.net/media/fcodhs1m/state-of-the-nation-report-vertebral-fracture-identification-in-2021.pdf

The Royal College of Radiologists. (2021). *Radiological guidance for the recognition and reporting of osteoporotic vertebral fragility fractures (VFFs)* (BFCR(21)5). https://www.rcr.ac.uk/our-services/all-our-publications/clinical-radiology-publications/radiological-guidance-for-the-recognition-and-reporting-of-osteoporotic-vertebral-fragility-fractures-vffs/

Tu, K. N., Lie, J. D., Wan, C. K. V., Cameron, M., Austel, A. G., Nguyen, J. K., Van, K., & Hyun, D. (2018). Osteoporosis: A Review of Treatment Options. *P T*, *43*(2), 92-104. https://www.ncbi.nlm.nih.gov/pubmed/29386866

Appendix: Literature search from NICE Information Specialist team during scoping (Medline)

1 (boneview* or gleamer).af.

2 (CINA-VCF or "avicenna.ai").af.

3 (c-spine or briefcase).af. and (ai or artificial intelligence).tw.

4 aidoc.af. and (spinal or spine or vertebr* or bone or osteop* or fractur*).tw.

5 (Clariosteo or Claripi).af.

6 (Healthost or Healthvcf or "nanox.ai").af.

7 (IB lab or ImageBiopsy Lab).af.

8 ((Annalise adj2 CXR) or "annalise.ai").af.

9 or/1-8

10 exp Osteoporosis/

11 fractures, compression/ or osteoporotic fractures/ or spinal fractures/

12 (osteoporo* or osteop?en*).tw.

13 ((compress* or fragilit* or spine or spinal or vertebr*) adj3 fracture*).ab.

14 or/10-13

15 diagnostic imaging/ or radiography/ or absorptiometry, photon/ or exp magnetic resonance imaging/ or exp tomography, emission-computed/ or x rays/

16 (radiogra* or ct or (comput* adj4 tomogra*) or absorptiometry or dexa or dxa or magnetic resonance or mri or mrs or nmr* or x ray or xray).tw.

17 15 or 16

18 Algorithm*.ti,kf.

19 (algorithm* adj2 (learn* or automate* or detect* or predict* or treatment* or therap* or radiolog* or AI or DL or ML or data or dataset* or base* or classif*)).ab.

20 Artificial Intelligen*.ti,ab,kf.

21 AI.ti,kf.

22 (machine adj2 learn*).ti,ab,kf.

23 machinelearn*.ti,ab,kf.

24 (deep adj2 learn*).ti,ab,kf.

25 deeplearn*.ti,ab,kf.

26 neural network*.ti,ab,kf.

27 (convolutional adj1 network*).ti,ab,kf.

28 automate*.ti.

29 (automate* adj3 (system* or score* or software* or analysis* or analyse* or risk* or evaluat* or tool* or detect* or process*)).ab,kf.

30 (vector machine* or svm*).ti,ab,kf.

31 radiomic*.ti,ab,kf.

32 ((supervised or unsupervised) adj3 (classifier* or prediction*)).ti,ab,kf.

33 or/18-32

34 14 and 17 and 33

35 9 or 34

36 limit 35 to english language

37 limit 36 to (letter or historical article or comment or editorial or news)

38 36 not 37

39 animals/ not humans/

40 38 not 39