

**NATIONAL INSTITUTE FOR HEALTH AND CARE
EXCELLENCE**

DIAGNOSTICS ASSESSMENT PROGRAMME

Early value guidance consultation document

**Artificial intelligence-derived software to
analyse chest X-rays for suspected lung
cancer in primary care referrals: early value
assessment**

The National Institute for Health and Care Excellence (NICE) is producing early value guidance on using artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in the NHS in England. The diagnostics advisory committee has considered the evidence and the views of clinical and patient experts. This topic is part of the pilot using the new early value assessment approach. Early value assessment guidance recommendations are conditional while more evidence is collected on the technology to address uncertainty in the evidence base. Once further evidence is collected, this guidance will be reviewed to make a decision on the routine adoption of the technology.

This document has been prepared for public consultation. It summarises the evidence and views that have been considered, and sets out the recommendations made by the committee. NICE invites comments from registered stakeholders, healthcare professionals and the public. This document should be read along with the [evidence](#) (an early value assessment report and the early value assessment report erratum).

The advisory committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?
- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the recommendations sound, and a suitable basis for early value guidance to the NHS?

Equality issues

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the recommendations may need changing to meet these aims. In particular, please tell us if the recommendations:

- could have a different effect on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology
- could have any adverse effect on people with a particular disability or disabilities.

Please provide any relevant information or data you have about such effects and how they could be avoided or reduced.

Note that this document is not NICE's final guidance on artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer. The recommendations in section 1 may change after consultation.

After consultation, NICE will consider the comments received. The final recommendations will be the basis for NICE's early value guidance on using the technology with evidence generation.

Key dates:

Closing date for comments: 30 March 2023

1 Recommendations

1.1 Artificial intelligence (AI)-derived software cannot be used for early routine use in the NHS to analyse chest X-rays alongside clinician review for suspected lung cancer in adults referred from primary care while evidence is generated. The following AI-derived software should only be used in research to help identify abnormal lung features that suggest lung cancer, prioritise review of chest X-rays and speed up referral for a CT scan:

- AI-Rad Companion Chest X-ray (Siemens Healthineers)
- Annalise CXR (Annalise ai)
- Auto Lung Nodule Detection (Samsung)
- Chest link (Oxipit)
- ChestView (Gleamer)
- Chest X-ray (Rayscape)
- ClearRead Xray (Riveraintech)
- InferRead DR Chest (Infervision)
- Lunit INSIGHT CXR (Lunit)
- Milvue Suite (Milvue)
- qXR (Qure.ai)
- Red dot (Behold.ai)
- SenseCare-Chest DR PRO (CADe, SenseTime)
- VUNO Med-Chest X-Ray (VUNO).

1.2 Centres already using AI-derived software for review of chest X-rays in adults referred from primary care may continue to do so but only under appropriate research governance and only alongside clinician review.

1.3 Further research is recommended to assess:

- the impact of the software on clinical decision making and the number of people referred to have a chest CT scan
- how using the software affects healthcare costs and resource use

- the impact on review and reporting time, and time to CT referral and diagnosis
- the diagnostic accuracy of AI-derived software alongside clinician review in detecting nodules or other abnormal lung features that suggest lung cancer, including in people with multiple conditions
- the technical failure rates of the software
- whether the software works in groups that could particularly benefit ([see section 3.9](#)).

Why the committee made these recommendations

There is an unmet need for quicker reporting of chest X-rays in people referred from primary care, particularly when this may support earlier detection of lung cancer. AI-derived software may help to reduce time to diagnosis and treatment by supporting clinician review of images and by prioritising images for review so a CT scan can be done the same day. But, there is insufficient evidence. With the available evidence, it is not possible to assess the clinical and cost benefits or risks of using the technology in the NHS. So, AI-derived software should not be used for clinical decision making in the NHS until more evidence is available.

There is no evidence to show how accurate software-assisted clinician review will be at identifying lung abnormalities compared with clinician review alone in people referred for a chest X-ray by their GP. Using this software could lead to lung cancer being missed or people having unnecessary CT scans, which can cause anxiety. This could also increase costs. More research on the performance of the technologies could show what the benefits and risks are.

Limited evidence suggests that the technologies did not significantly reduce the time to report chest X-rays, but that they could allow same day CT scans when used to prioritise X-rays for clinical review. More research on the impact of the technologies on how lung cancer is diagnosed would allow better understanding of the benefits for patients and the healthcare system.

2 The diagnostic tests

Clinical need and practice

- 2.1 Lung cancer is one of the most common types of cancer in the UK. It causes symptoms such as persistent cough, coughing up blood, and feeling short of breath. People in the early stages of the disease may not have symptoms and so lung cancer is often diagnosed late. In 2018, more than 65% of lung cancers were diagnosed at stage 3 or 4 (Cancer Research UK). The NHS Long Term Plan sets out the NHS's ambition to diagnose 75% of all cancers at stages 1 or 2 by 2028.
- 2.2 Detecting abnormal lung features such as lung nodules, pleural effusion (excess fluid around the lung), collapsed lung or lung segments, destruction or erosion of bony structures such as the ribs, as well as masses or lymph nodes within the mediastinum could help find lung cancer early. Lung abnormalities can be seen on a chest X-ray. The chest X-ray may be done because of signs and symptoms that suggest lung cancer. Lung abnormalities that suggest cancer can also be detected incidentally on chest X-rays done for reasons unrelated to lung cancer, such as trauma, heart problems, or investigation and monitoring of other lung conditions.
- 2.3 Software with artificial intelligence (AI)-derived algorithms can be used to automatically detect lung abnormalities on chest X-ray images. This could help a radiologist or reporting radiographer with review and interpretation of chest X-ray images and support clinical decisions about the need for a CT scan or further investigations. Using the technology alongside clinician review may help to detect suspected lung cancer by:
- identifying normal or abnormal images and highlighting suspected abnormalities

- prioritising review of chest X-rays and speeding up subsequent referral to CT scan
- increasing the accuracy of suspected lung cancer detection by radiologists or reporting radiographers
- reducing the time to review and report chest X-rays.

The interventions

2.4 The software included in this evaluation have AI-derived algorithms to detect lung abnormalities for suspected cancer. All software technologies in clinical settings use fixed algorithms. They cannot adapt in real time using data from the clinical practice setting in which they are used. In the NHS, AI-based technologies are only used alongside healthcare professionals. The healthcare professional who reviews the chest X-ray using the software makes the final reporting decision. These software could be radiological computer-assisted triage (CAST) software, computer aided detection (CADe) software or computer-aided diagnosis (CADx) software. CAST software is intended to help prioritisation and triage of medical images that are time sensitive for urgent review. CADe and CADx software are technologies that can detect or diagnose an abnormality on the X-ray.

2.5 The technologies in this evaluation are:

- AI-Rad Companion Chest X-ray (class 2a medical device, CADx, Siemens Healthineers)
- Annalise CXR (class 2b medical device, CADe/CAST, Annalise ai)
- Auto Lung Nodule Detection (class 2 a medical device, CADe, Samsung)
- Chest link (class 2b medical device, CADe/CAST, Oxipit)
- ChestView (class 2a medical device, CADe, Gleamer)
- Chest X-ray (class 2a medical device, CADe, Rayscape)
- ClearRead Xray (class 2a medical device, CADe Riveraintech)

- InferRead DR Chest (class 2a medical device, CADe, Infervision)
- Lunit INSIGHT CXR (class 2a medical device, CADe, Lunit)
- Milvue Suite (class 2a medical device, CADe/CAST, Milvue)
- qXR (class 2a medical device, CADe, Qure.ai)
- Red dot (class 2a medical device, CADe/CAST, Behold.ai)
- SenseCare-Chest DR PRO (class 2b medical device, CADe, SenseTime)
- VUNO Med-Chest X-Ray (class 2a medical device, CADe, VUNO).

The comparator

- 2.6 The comparator is a chest X-ray review by a radiologist or reporting radiographer without assistance from AI-derived software. The levels of experience of radiologists or reporting radiographers vary.

3 Committee discussion

The [diagnostics advisory committee](#) considered evidence on software with AI-derived algorithms for detecting and measuring lung nodules and other abnormalities in chest X-ray images from several sources, including an external assessment report and an overview of that report. Full details are in the [project documents for this guidance](#).

Unmet need

There is an unmet need for fast chest X-ray reports

- 3.1 In primary care chest X-ray referrals, there is an unmet need for quicker reporting of chest X-rays. Sometimes, it takes a long time for chest X-ray reports to be returned to a GP, which can have an impact on the time to CT scan, diagnosis and treatment of lung cancer. Factors leading to this delay in chest X-ray reporting include a backlog of chest X-rays that need review and not enough radiologist and reporting radiographer capacity. A clinical expert explained that in some areas lung cancer specialists would not accept a referral for a CT scan unless the chest X-ray is done and

reported. Software that provides triaging of images could prioritise images with abnormal features that suggest lung cancer for urgent review, which could result in faster referral to CT scan, diagnosis and treatment if needed.

Anxiety while waiting for final diagnosis

3.2 A patient expert said that people who are told they have a lung nodule or other abnormality may experience anxiety, especially if the chest X-ray was done without any expectation of finding nodules or abnormalities suggesting cancer. Having more information as soon as possible is important and reduces anxiety for people with suspected lung cancer and their families. The patient expert explained that most people are happy to have all tests done quickly, because immediate results would turn a passive period of waiting into something more proactive and less uncertain. According to the patient expert, people would be happy to have AI-derived software used as part of the diagnostic pathway as long as there was evidence to show that it is accurate.

Software capability

Different software have different capability

3.3 AI-derived software have different capabilities. Some technologies only provide CAdE/CAdx, that can detect or diagnose an abnormality on the X-ray, whereas others provide both CAdE/CAdx and CAST ([see section 2.4](#)). The committee considered that software that can prioritise image review with abnormalities suggesting lung cancer might have the greatest benefit, because this could reduce time to CT scan, diagnosis and treatment ([see section 3.1](#)). It heard that AI-derived software for chest X-rays are not able to compare a previous chest X-ray with a new one. Comparison with previous chest X-rays would be helpful because it could help determine whether a lung abnormality was of any clinical concern.

Clinical effectiveness

Technologies with no published evidence

3.4 The committee considered the summarised evidence. The external assessment group's (EAG's) review found no relevant published evidence for the population referred from primary care. The committee acknowledged the EAG's expanded review criteria, that included studies in populations where the referral criteria were unclear. It noted that there was no evidence for 11 of the 14 technologies: Auto Lung Nodule Detection, ChestLink Quality, ChestView, Chest X-ray, ClearRead X-ray Detect, InferRead DR Chest, Lunit INSIGHT CXR, Milvue Suite, qXR, SenseCare-Chest DR PRO and VUNO Med-Chest X-Ray. The committee recommended more research on these technologies (see section 1.3).

Generalisability of evidence to clinical practice

3.5 There were no studies that looked at accuracy in the population of interest, that is, those referred from primary care. There were 6 studies that compared the accuracy of X-ray review by a radiology specialist alongside AI-derived software with review by a radiology specialist alone, but it was unclear where the population had been referred from. The primary care referral population is likely to be different to populations from other settings, such as inpatients and people presenting to the emergency department. These populations could have a higher prevalence of disease and present at a more advanced stage, making it easier to detect abnormalities that indicate cancer. In most primary care referral cases, chest X-rays requested by GPs are done to rule out lung cancer because it is unlikely that the person has lung cancer. So, AI-derived software trained and evaluated in populations from other settings or mixed settings are likely to perform differently in people referred from primary care. The committee agreed that the accuracy data was unlikely to be generalisable to a primary care population. The committee also acknowledged that only 1 UK study was identified and that diagnostic pathways in different

countries may be different. The UK study assessed the diagnostic accuracy of red dot (Behold.ai), but the population was unclear.

Accuracy for detecting lung cancer and nodules

3.6 Only 1 study, assessing red dot (Behold.ai) in the UK from populations with unclear referral, reported accuracy to detect lung cancer. This study found AI-derived software to have a statistically significantly higher sensitivity than X-ray review without the software. The study found AI-derived software had lower specificity but the difference was not statistically significant. The other 5 studies that reported accuracy to detect nodules did not report any statistically significant differences in accuracy. In practice, if specificity is lower, this would mean more false positive results. That is, more people who do not have lung cancer would go on to have CT scans, which are associated with anxiety and costs to the NHS. If sensitivity to detect lung cancer is higher, this would mean more true positive lung cancer cases would be detected and referred for CT scans, which could lead to treatment at an earlier disease stage and improved outcomes. In contrast, if sensitivity to detect lung cancer is lower, this could result in missing the opportunity to detect cancer early. Cancer would be identified at a later stage when the disease is more advanced, which may be associated with worse outcomes and more costly treatment. No studies reported technical failure rates. Because none of the studies looked at more than 1 software, a direct comparison between different software was not possible. Clinical experts highlighted the need for a benchmark test that could be used to compare the different software and identify those that reach a specified level of accuracy. The committee concluded that further research was needed on how using AI-derived software alongside clinician review of chest X-rays affects the accuracy of detecting lung cancer. Research is also needed on the technical failure rates of the software.

Time to read and report a chest X-ray

3.7 The EAG's review included 2 studies that looked at time to read and report a chest X-ray. The comparisons in the studies were done at least partly in laboratory-like conditions, rather than in routine clinical practice. So, results may not be generalisable to current practice in the NHS. No studies suggested that read and report was faster with AI-derived software than without. Because AI-derived software would be used in addition to review by a radiologist or reporting radiographer, it may not reduce time to read and report chest X-rays. The time may also depend on how well the software integrates into the radiologists' workflow within the picture archiving and communication system (PACS). AI-derived software may support less experienced or trainee radiologists and reporting radiographers. This is because the software may identify nodules or abnormalities that the less experienced reader may not pick up, and this could help to improve their skills initially. The committee was uncertain about whether using AI-derived software would speed up reading and reporting a chest X-ray in a UK clinical setting. Further evidence should be collected on the impact of AI-derived software on time to read and report chest X-rays and if this is different between readers with different levels of experience.

Triaging of images

3.8 There were no studies that reported time to CT referral or time to diagnosis. Some software can triage chest X-rays, identifying images that have features that suggest lung cancer, which can be prioritised for review, and identify other images that are likely to be normal and could potentially be reviewed faster. This could help improve workflow by focusing radiology resources on the more urgent chest X-rays, and potentially speeding up time to CT referral, diagnosis and treatment when needed. It could also enable a same day CT scan pathway in the population referred from primary care. The committee considered studies

highlighted by behold.ai that reported on the accuracy of red dot software to identify high confidence normal X-rays and X-rays containing features suspicious for lung cancer. The EAG noted that these studies contain a mixed population rather than a population referred from primary care only. The committee concluded that some software could meet the unmet need for faster chest X-ray reports, leading to faster CT referral, diagnosis and treatment if needed. It further concluded that more research is needed on how using AI-derived software to triage chest X-rays from people referred from primary care affects time to CT referral and time to diagnosis.

Populations that could particularly benefit from the technologies

3.9 The committee considered groups of people that could particularly benefit from the software. It recognised that detecting lung nodules and other abnormalities can be difficult in people with underlying lung conditions such as asthma, chronic obstructive pulmonary disease (COPD), people whose family background means they may be at a higher risk of having lung cancer, and younger women who do not smoke. If using the software helped to improve lung cancer detection, it would be particularly beneficial to these groups.

Cost effectiveness

Conceptual model structure

3.10 The EAG developed a conceptual decision analytic model to inform potential future full cost-effectiveness evaluation of AI-derived software to analyse chest X-rays to identify suspected lung cancer. A model structure was developed based on the chest X-ray clinical pathway. The committee agreed that the conceptual model was a good basic framework. But it highlighted some issues that would need clarifying, for example, whether people with lung cancer and no abnormalities identified on a chest X-ray would be picked up at a later point, such as presentation at an emergency department. Triage and prioritising images with abnormal features that

suggest lung cancer, and the impact of this on time to CT scan, diagnosis and treatment, would need to be captured in a future model. The committee agreed that a linked-evidence approach to economic modelling would be acceptable. That is, using diagnostic accuracy and time to diagnosis data linked to long-term outcome data from separate studies.

Software costs

3.11 The EAG considered the costs of introducing AI-derived software alongside radiology specialist review of chest X-rays by developing a simple budget impact analysis. The budget impact analysis considered one-off set up costs, annual subscription fee based on a volume of 16,945 images, total cost per year and the cost over the first 5 years. Because the literature reviews did not provide any evidence to show changes in resource use because of AI-derived software, these were not considered in the budget impact calculations. Test costs varied between companies, but the EAG cautioned against direct comparison, because the AI-derived software presented have varying capabilities and some may be used in different positions early in the diagnostic pathway.

Conclusions

Potential benefits and risks

3.12 The potential benefits associated with using AI-derived software could include:

- A reduction in the time radiologist or diagnostic radiographers spend reviewing and reporting chest X-rays, which could release staff resources and reduce the time to return a chest X-ray report to a GP.
- Improved workflow, resulting in a reduction of time from chest X-ray to diagnosis and treatment, which could enable a same day CT pathway, improve patient outcomes and quality of life, and save resources in the NHS.

- Higher sensitivity to detect cancerous nodules and other abnormalities that suggest cancer, which could result in more cancers being identified and treated at an earlier stage, improving patient outcomes and quality of life and saving resources in the NHS.

The potential risks associated with using AI-derived software could include:

- The cost of the AI-derived software, which includes a one-off set-up cost, may not be offset by cost and resource savings later in the pathway.
- Lower specificity to detect cancerous nodules and other abnormalities that suggest cancer could result in more people without cancer having CT scans, which would have cost and disutility implications.
- Lower sensitivity to detect cancerous nodules and other abnormalities that suggest cancer could result in lung cancer being missed and identified at a more advanced disease stage, which could lead to more costly treatment and worse patient outcomes.
- Software may not reduce, or may increase, the time radiologists or diagnostic radiographers spend reviewing and reporting chest X-rays, and so the workload and the time to return a chest X-ray report to a GP would not be reduced.

The clinical effectiveness is unknown

3.13 The committee recalled that there was no evidence on the accuracy of AI-derived software to detect suspected lung cancer in a population referred from primary care and no evidence on the technical failure rates. It noted that the evidence summarised came from populations with unclear referral criteria and the results may not be generalisable to a primary care population. The committee concluded that the accuracy of AI-derived software to identify suspected lung cancer on chest X-rays from people referred from primary care is uncertain. Therefore, the

committee could not determine whether AI-derived software is clinically effective, and recommended further evidence is generated on diagnostic accuracy and technical failure rates.

The cost impact is unknown

3.14 The diagnostic accuracy data on AI-derived software to detect suspected lung cancer in a population referred from primary care is uncertain. So, it is also unknown how these technologies would impact the number of people referred for chest CT scans and the number of lung cancer cases that would be identified. The committee was therefore unable to understand whether the benefits would outweigh the risks and was concerned that AI-derived software may not be cost effective. It recommended further research on the impact of the software on clinical decision making, the number of people referred for chest CT scans, and how AI-derived software impacts healthcare costs and resource use.

Addressing unmet need

3.15 The unmet need was around the speed of chest X-ray reports being returned to the referring GP, particularly for X-rays that show abnormalities that suggest lung cancer because a delay can impact on time to diagnosis, treatment and patient outcomes. There was no evidence on the impact of AI-derived software on the time to CT scan and time to diagnosis from a population referred from primary care. So, the committee was uncertain whether any of the AI-derived software could meet the unmet need, but it noted that some of the software probably could reduce delays. The committee recommended further research on the impact of AI-derived software on review and reporting time and time to CT referral and diagnosis.

Use only in research

3.16 The committee recalled that the clinical effectiveness and cost impact of AI-derived software could not be determined, and it was uncertain

whether the software would address the unmet need. Also, the committee could not determine whether the benefits would outweigh the risks if the software was adopted for use in the NHS alongside evidence generation. It was concerned that AI-derived software for chest X-rays could be cost incurring and may not improve clinical outcomes, and should not currently be used to inform clinical care in the NHS. So, the committee decided that AI-derived software should only be used in research to resolve some of the uncertainty and allow re-assessment in the future when further data is available. The committee was aware that some centres are already using AI-derived software and concluded that these centres may continue, but only under appropriate research governance. AI-derived software should only be used alongside clinician review.

4 Diagnostics advisory committee members and NICE project team

Committee members

This topic was considered by the [diagnostics advisory committee](#), which is a standing advisory committee of NICE.

Committee members are asked to declare any interests in the test to be evaluated. If it is considered there is a conflict of interest, the member is excluded from participating further in that evaluation.

The [minutes of each committee meeting](#), which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Additional specialist committee members took part in the discussions for this topic:

Specialist committee members

Alan Thomas

Lay specialist

Alberto Alonso

Consultant radiologist, Manchester University NHS Foundation

Christopher Johns

Consultant thoracic radiologist, Northern General Hospital, Sheffield

Emma O'Dowd

Consultant in respiratory medicine, Nottingham University Hospitals NHS Trust

Robin Alcock

Consultant radiologist, Salisbury District Hospital

Fergus Gleeson

Consultant radiologist, Oxford University Hospitals

NICE project team

Each diagnostics evaluation is assigned to a team consisting of a technical analyst (who acts as the topic lead), a technical adviser and a project manager.

Peslie Ngambi

Topic lead

Alice Bell

Associate technical analyst

Frances Nixon

Technical adviser

Toni Gasse

Project manager

ISBN: