NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Draft guidance

AI-derived computer-aided detection (CAD) software for detecting and measuring lung nodules in CT scan images

The National Institute for Health and Care Excellence (NICE) is producing guidance on using AI-derived computer-aided detection (CAD) software for detecting and measuring lung nodules in CT scan images in the NHS in England. The diagnostics advisory committee has considered the evidence and the views of clinical and patient experts.

This document has been prepared for 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 (the external assessment report and the external assessment report addendum and 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 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.

Issue date: February 2023

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 AI-derived computer-aided detection (CAD) software for detecting and measuring lung nodules in CT scan images. The recommendations in section 1 may change after consultation.

After consultation, the committee will meet again to consider the evidence, this document and comments from the consultation. After considering the comments, the committee will prepare its final recommendations, which will be the basis for NICE's guidance on the use of the technology in the NHS in England.

For further details, see the diagnostics assessment programme manual

Key dates:

Closing date for comments: 7 March 2023

Second diagnostics advisory committee meeting: 23 March 2023

1 Recommendations

NICE is aware that companies are reviewing their CE marking in response to changes and advances in regulations for digital health technologies.

1.1 There is not enough evidence to recommend AI-derived computer-aided detection (CAD) software alongside clinician review of CT scan images to detect and measure lung nodules in, or outside of, targeted lung cancer screening.

1.2 Centres already using AI-derived CAD software alongside clinician review as part of targeted lung cancer screening may continue to do so, but only if further data is collected to make sure the benefits for people attending screening and clinicians using the software are realised in practice (see section 1.3).

1.3 Further data collection and research is recommended (see the section on further research) to assess:

- how using CAD software alongside clinician review of CT scan images affects the accuracy of detecting and measuring lung nodules
- how using CAD software alongside clinician interpretation affects scan review and reporting time
- the prevalence of nodules in people who have a chest CT scan because of signs or symptoms that suggest lung cancer.

Why the committee made these recommendations

Using AI-derived CAD software to assist reviewing CT scans could improve the detection of lung nodules.

In targeted lung cancer screening, the evidence is too limited to show which technologies are the most clinically and cost effective. But the model results suggest that using the software alongside clinician review has the potential to be cost

effective. So, although there is not yet enough evidence to recommend the software, centres already using it as part of the targeted lung cancer screening may continue doing so if further data is collected to make sure the benefits of using the software are realised in practice.

Outside of targeted lung cancer screening, using the software could lead to more people being wrongly diagnosed with lung nodules, and having unnecessary CT surveillance. This is very uncertain because there is so little evidence, but having CT surveillance may cause people unnecessary anxiety, so more research is needed.

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 lung nodules – small growths in the lung, which can be cancerous – could help find lung cancer early. Lung nodules can be seen on a chest CT scan. The scan may be done because of signs and symptoms that suggest lung cancer, or as part of targeted lung health checks. Lung nodules can also be detected incidentally on CT scans done for reasons unrelated to lung cancer, such as trauma or heart problems.

2.3 Computer-aided detection (CAD) software with artificial intelligence (AI)-derived algorithms can be used to automatically detect and measure lung nodules on chest CT scan images. This could assist radiologists or other healthcare professionals in reviewing scan images, and support clinical decisions about the need for CT surveillance or further

investigation. Using it alongside clinician review may help to find and treat lung cancer early by:

- increasing detection of lung nodules that need further investigation or surveillance
- helping assess the growth of lung nodules under CT surveillance
- improving reporting of nodule characteristics (such as nodule volume) to support decision making
- reducing the time to review and report CT scans.

The interventions

2.4 The CAD software included in this evaluation has AI-derived automated nodule detection and volume measurement capability. 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, not as standalone interventions. The healthcare professional who reviews the CT scan using the software makes the final reporting decision.

2.5 The technologies in this evaluation are:

- AI-Rad Companion Chest CT (class 2a medical device, Siemens Healthineers)
- AVIEW LCS+ (class 2a medical device [information from public domain], Coreline Soft)
- ClearRead CT (class 2a medical device, Riverain Technologies)
- contextflow SEARCH Lung CT (class 2a medical device, contextflow)
- InferRead CT Lung (class 2a medical device, Infervision)
- JLD-01K (class 1 medical device, JLK Inc.)
- Lung AI (class 2 medical device [information from public domain], Arterys)

- Lung Nodule AI (at the time of writing, Lung Nodule AI does not have a CE mark, Fujifilm)
- qCT-Lung (class 1 medical device [information from public domain], Qure.ai)
- SenseCare-Lung Pro (class 2b medical device [information from public domain], SenseTime)
- Veolity (class 2a medical device, MeVis)
- Veye Lung Nodules (class 2b medical device, Aidence)
- VUNO Med-LungCT AI (class 2a medical device [information from public domain], VUNO).

The comparator

CT scan review by a healthcare professional without assistance from AI-derived CAD software

2.6 The comparator is a chest CT scan review by a radiologist or another healthcare professional without assistance from AI-derived CAD software. The healthcare professional reviewing the scan may or may not be specialised in reviewing chest CT images. In the Targeted Lung Health Checks programme (which provides a starting point for the implementation of targeted lung cancer screening in England), the healthcare professionals reviewing scans are radiologists specialised in reviewing chest CT images. In other CT scan settings, levels of specialisation and experience vary.

3 Committee discussion

The diagnostics advisory committee considered evidence on computer-aided detection (CAD) software with artificial intelligence (AI)-derived algorithms for detecting and measuring lung nodules in CT scan 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.

Improving detection of lung cancer

3.1 Clinical experts explained that lung nodules may be challenging for healthcare professionals to detect because of their small size, varying shape, and how close they are to other structures such as blood vessels in the lung. This may mean that sometimes a nodule will be missed by a clinician reviewing CT scans. Although most lung nodules are benign, some may be malignant or develop into lung cancer. Using CAD software to assist reviewing CT scans for lung nodules could improve the detection of lung nodules and improve early diagnosis of lung cancer.

Anxiety while having CT surveillance

3.2 A patient expert said that people who are told they have a lung nodule may experience anxiety, especially if the CT scan was carried out without any expectation of finding lung nodules. Not knowing whether the nodule is benign may further increase the anxiety. Having more information as soon as possible is important for people with lung nodules and their families. The patient expert explained that when CT surveillance is needed to understand the nature of the lung nodule, anxiety can be spread over a long period. If a follow-up scan shows no signs of malignancy, the anxiety may reduce, but will not disappear. Because CT surveillance involves multiple CT scans, people may be further concerned about the repeated radiation exposure.

Need for software in the targeted lung cancer screening setting

3.3 The committee noted that using AI-derived CAD software that automatically detects and measures lung nodules could be helpful for targeted lung cancer screening. The clinical experts explained that it would be useful to understand which technologies are clinically and cost effective in this setting.

Clinical effectiveness

Technologies with no published evidence

3.4 The committee considered the available evidence for each technology. It noted that the external assessment group's (EAG's) review found no relevant published evidence for 5 of the 13 technologies: JLD-01K, Lung AI, Lung Nodule AI, qCT-Lung and SenseCare-Lung Pro, in any of the populations. The committee recommended more research on these technologies (see section 4).

Accuracy of detecting lung nodules in screening

3.5 There were 5 studies on 4 different software (ClearRead CT, InferRead CT Lung, Veolity, VUNO Med-LungCT AI) that compared the accuracy of CT scan review with and without CAD software to detect lung nodules and reported per-person results from a screening population. The committee considered that reporting accuracy results per person instead of per nodule was important. Per-nodule results could only tell whether nodules were missed or wrongly detected, but not whether people with nodules were missed or wrongly identified as having nodules. Nearly all of the studies found that CT scan review with software was more sensitive but less specific than review without the software. In practice, this would mean that more lung nodules and so potentially more cancers would be detected, but more people would go on to have CT surveillance as a result. The clinical experts noted that it was not clear whether this was the case for both solid and subsolid nodules. Subsolid nodules may be especially hard for the software to detect in the low-dose CT scan used for targeted screening. The committee also noted that, because none of the studies looked at more than 1 software, a direct comparison between different software was not possible. The committee concluded that further research was needed on how using AI-derived CAD software alongside clinician review of CT scan images affects the accuracy of detecting and measuring lung nodules.

Generalisability of evidence to clinical practice outside of screening

3.6 Only 1 study (InferRead CT Lung) looked at accuracy in a population with signs and symptoms that suggest lung cancer. The results were similar to those from studies in the screening setting. But the committee noted that because there was only 1 study, performance of the AI-derived CAD software outside of screening was uncertain. It considered whether evidence from screening may be generalisable to routine clinical practice. It noted that the prevalence of lung nodules would be expected to differ between populations. The clinical experts described that people who are referred for a chest CT scan outside of targeted lung cancer screening because of signs or symptoms that suggest lung cancer, or for reasons unrelated to suspicion of lung cancer, are more likely to have other underlying lung conditions (for example asthma, chronic obstructive pulmonary disease [COPD] or granulomatous lung diseases). How common these conditions are may also depend on age and family background. The underlying lung conditions may make it harder for the software to differentiate nodules, especially subsolid nodules, from other nodule-like structures in the lungs, and cause them to be falsely detected as nodules. The clinical experts also noted that, unlike targeted lung cancer screening scans that are always low-dose CT scans without contrast, chest CT scans done for other reasons use a standard dose and may be done with contrast. Further, the healthcare professionals reviewing targeted screening scans are radiologists specialised in reviewing chest CT images for lung nodules, whereas in other settings, levels of specialisation and experience of the healthcare professionals reviewing the scan may vary. The committee concluded that the evidence from a screening population is unlikely to be generalisable to people who have a chest CT scan for other reasons.

Populations that could particularly benefit from the technologies

3.7 The committee considered groups of people that could particularly benefit from the software being used. It recognised that people with underlying lung conditions, and people whose family background means they are more likely to have subsolid nodules, may be at a higher risk of not having nodules detected and lung cancer missed. If using the software helped to improve detection, it would be particularly beneficial to these groups.

Time to read and report a CT scan

3.8 The EAG's review included 9 studies that looked at time to read and report a CT scan. All of the studies suggested that review was faster with AI-derived CAD software than without. Some studies suggested that less experienced readers may save even more time than experienced readers. But the committee noted that the comparisons in the studies were done at least partly in laboratory-like conditions, rather than in routine clinical practice. The clinical experts explained that in targeted screening, scan review time is protected. But in routine clinical practice, reviewing scans may be less continuous because of interruptions. Using an additional tool could help reviewing, but could also slow it down. The time may also depend on how well the software integrates into the radiologists' workflow within the picture archiving and communication system (PACS), in which CT scan images are reviewed and reported. The EAG pointed out that people reporting scans may behave differently in research conditions than in clinical practice because they know that their decisions will not affect health outcomes. The committee concluded that it is uncertain whether using software would speed up reading and reporting a CT scan outside of targeted screening. There is also a need to confirm whether the suggested time advantage applies in the targeted screening setting.

Cost effectiveness

Exploratory model

3.9 The EAG built a health economic model to evaluate the cost effectiveness of AI-derived CAD software for detecting and measuring lung nodules from CT scan images. The model captured targeted screening, and clinical practice in which people might have a chest CT scan because of signs or symptoms that suggest lung cancer or for other unrelated reasons. Because there was not enough clinical-effectiveness evidence on any of the individual technologies, the model combined data from various sources to assess a hypothetical software. For data that was needed but not available, the EAG used simulation to generate model inputs from other sources of information. The committee concluded that because of the limitations in the data available for modelling, the model should be considered exploratory and the model results indicative only.

Software costs

3.10 The EAG used a software cost of £2 per scan in the model for targeted screening, and £3.34 per scan for chest CT scans because of signs or symptoms that suggest lung cancer or for other unrelated reasons. The model did not include set-up and maintenance fees, but the EAG did a sensitivity analysis on software costs ranging from £1.50 to £6 per scan for targeted screening and from £2.67 to £6 per scan outside of targeted screening. This analysis showed that the software cost was not particularly influential.

Influential model inputs

3.11 The committee recalled that data on accuracy, a model input that had a substantial effect on the model results, was especially limited in people having a chest CT scan because of signs or symptoms that suggest lung cancer or for other unrelated reasons. The clinical experts also pointed out that the prevalence of lung nodules, another key model input, used in

the model for screening (50.9%) and for people with symptoms that suggest lung cancer (94.9%) was higher than they would expect. They stated that the prevalence in people with symptoms that suggest lung cancer was uncertain. The committee concluded that further research is needed to assess the prevalence of lung nodules in people who have a chest CT scan because of signs or symptoms that suggest lung cancer.

Potential for cost effectiveness for targeted screening

3.12 In the EAG's model, for targeted screening, the software-assisted review of the scan images detected a larger number of people with nodules that needed follow up compared with scan image review without the software. It also detected slightly more lung cancers. The software-assisted review was slightly less costly and more effective than scan image review without the software. The committee recalled that, because of a lack of evidence, the model results were only indicative. But it noted that changing different model assumptions did not change the direction of the model results (scan image reviews with software stayed slightly less costly and more effective than scan image reviews without the software). The committee concluded that the software had the potential to be cost effective when used alongside clinician review for targeted screening.

QALY loss outside of targeted screening

3.13 Based on the EAG's model, outside of targeted screening, review of CT scan images was slightly less effective with than without the AI-derived CAD software. The clinical experts said that some quality-adjusted life year (QALY) loss in this setting could be expected; because of underlying conditions affecting the lungs being more common, the software may more often flag up nodule-like structures as lung nodules. This could lead to more people unnecessarily being offered CT surveillance, which was associated with a disutility in the model to reflect patient anxiety (see section 3.2). The committee noted that like in targeted screening, using the software could also lead to more cancers being detected. But outside

of targeted screening, the clinical benefits of improved detection may not outweigh the disutility associated with more people having CT surveillance. The committee noted that changing some of the model assumptions affected the extent of the QALY loss, mostly reducing it. It recalled the weak clinical-effectiveness evidence in this setting (see section 3.9) and concluded that the QALY loss was uncertain but possible. It further concluded that because there was not enough evidence, it was not possible to say whether using the software outside of targeted screening could be cost effective. The committee recommended further research in these settings (see sections 4.1 and 4.2).

Conclusions

3.14 The committee recalled that the model results indicated that the AI-derived CAD software had the potential to be cost effective in targeted screening. It acknowledged that CAD software that automatically detects and measures lung nodules could be helpful for targeted lung cancer screening. It noted that the limited evidence did not allow the committee to decide which technologies are the most clinically and cost effective in this setting. It considered the need for software, the model results and the robustness of the results to changes in model inputs and assumptions. It concluded that although there was not enough evidence to recommend the software for targeted lung cancer screening, centres already using the software alongside clinician review of CT scan images as part of targeted lung cancer screening may continue doing so if further data is collected to make sure the benefits for people attending screening and clinicians using the software are realised in practice (see section 4). The committee further recalled that outside of targeted screening, it is possible that using the software could lead to more nodule-like structures being falsely detected as lung nodules and so more people unnecessarily having CT surveillance. The patient expert reminded the committee that having CT surveillance may lead to anxiety. The committee concluded that more

research is needed before software is used outside of targeted screening (see section 4).

Research considerations

3.15 The committee recommended more research on the AI-derived CAD software technologies (see section 4). Studies should include groups of people similar to those seen in the NHS. Ideally, studies should compare more than 1 software. The committee noted that the DART project, an ongoing research project that collects data from the targeted lung health checks programme, may be a helpful data source for comparing different technologies for targeted screening. Studies on accuracy of detecting and measuring lung nodules should report per-person results instead of per-nodule results. Studies should also consider reporting data separately in subgroups when using the technologies may be particularly useful or less effective. For example:

- people with subsolid nodules
- scans reviewed by more- and less-experienced reviewers
- outside of targeted screening settings, people who had CT scans:
  - with contrast
  - without contrast.

4 Recommendations for further research

4.1 Research in, and outside of, targeted lung cancer screening is recommended on:

- the accuracy of AI-Rad Companion Chest CT, AVIEW LCS+, ClearRead CT, contextflow SEARCH Lung CT, InferRead CT Lung, JLD-01K, Lung AI, Lung Nodule AI, qCT-Lung, SenseCare-Lung Pro, Veolity, Veye Lung Nodules, VUNO Med-LungCT AI plus clinician review compared with clinician review alone, to detect and measure lung nodules

- scan review and reporting time.

4.2 Research is recommended on the prevalence of lung nodules in people who have a chest CT scan because of signs or symptoms that suggest lung cancer.

5 Implementation

NICE intends to develop tools, in association with relevant stakeholders, to help organisations put this guidance into practice.

In addition NICE will support this guidance through a range of activities to promote the recommendations for further research. The research proposed will be considered for developing specific research study protocols as appropriate. NICE will also incorporate the research recommendations in section 4 into its guidance research recommendations database and highlight these recommendations to public research bodies.

6 Review

NICE will regularly monitor its published technology guidance to check for any new evidence or information that could affect the recommendations. Guidance will not have a fixed review date.

Brian Shine

Chair, diagnostics advisory committee

February 2023

7 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

Phil Crosbie

Senior lecturer in respiratory medicine, University of Manchester

Sujal Desai

Consultant radiologist, Royal Brompton Hospital

Seamus Grundy

Consultant in respiratory medicine, Salford Royal NHS Foundation Trust

Emma O'Dowd

Consultant in respiratory medicine, Nottingham University Hospitals NHS Trust

Nagmi Qureshi

Consultant radiologist, Royal Papworth Hospital NHS Foundation Trust

John Reynolds

Consultant radiologist, University Hospitals Birmingham NHS Foundation Trust

James Shambrook

Consultant cardiothoracic radiologist, University Hospital Southampton

Alan Thomas

Lay specialist

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.

Suvi Härmälä

Topic lead

Frances Nixon and Thomas Walker

Technical advisers

Donna Barnes and Toni Gasse

Project managers

ISBN: