Evidence overview: Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

This overview summarises the main issues the diagnostics advisory committee needs to consider. It should be read together with the final scope and the diagnostics assessment report.

1 Aims and scope

Software with artificial intelligence (AI)-derived algorithms can be used to detect and analyse lung nodules, small growths inside the lung, on chest CT scan images. The result of this analysis is intended to support the scan review and reporting by a trained healthcare professional.

Automatically detecting and analysing lung nodules on chest CT scans could assist radiologists or other healthcare professionals in reviewing scan images and support clinical decisions about the need for CT surveillance or further investigation. Detecting lung nodules may help to find and treat lung cancer early.

Using the software in the radiology pathway may:

- increase detection of lung nodules that need further investigation or surveillance
- help assess the growth of lung nodules under CT surveillance
- improve reporting of nodule characteristics to support decision making
- reduce the time to review and report CT scans

Decision question

Does the use of software for automated detection and analysis of lung nodules from CT scan images represent a clinically and cost-effective use of NHS resources?

Populations

- People who have no confirmed lung nodules, lung cancer and who are not having staging investigations or follow-up imaging for primary cancer elsewhere in the body, who have a CT scan that includes the chest:
    - because of signs or symptoms suggestive of lung cancer (symptomatic population)
    - for reasons unrelated to suspicion of lung cancer (incidental population)
    - as part of targeted lung cancer screening (screening population)
- People having CT surveillance for a previously identified lung nodule (surveillance population)

Depending on the available evidence, the following subpopulations may be considered: people with different family backgrounds, people who have a CT scan with or without contrast, and in the incidental population, people by reason for the CT scan.

Interventions

Software-assisted chest CT scan review by a healthcare professional using any of the following AI-derived software for automated detection and analysis of lung nodules:

- AI-Rad Companion Chest CT (Siemens Healthineers)
- AVIEW LCS+ (Coreline Soft)
- ClearRead CT (Riverain Technologies)
- contextflow SEARCH Lung CT (contextflow)
- InferRead CT Lung (Infervision)
- JLD-01K (JLK Inc.)
- Lung AI (Arterys)
- Lung Nodule AI (Fujifilm)
- qCT-Lung (Qure.ai)
- SenseCare-Lung Pro (SenseTime)

- Veolity (MeVis)
- Veye Lung Nodules (Aidence)
- VUNO Med-LungCT AI (VUNO)

Not all software is indicated for use in symptomatic, incidental and screening population, so they are not all assessed for all of the included populations. Find more details in table 1 in the appendix.

Comparator

The comparator is a chest CT scan review by a radiologist or another healthcare professional without assistance from AI-derived software.

The healthcare professional reviewing the scan may or may not be specialised in reviewing chest CT images. In the The Targeted Lung Health Checks programme, implementing the 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. The reviewer of the scan may use software to help measure the volume of an identified lung nodule but this software does not automatically detect or measure lung nodules.

Healthcare setting

- Secondary care
- Targeted lung cancer screening settings

Further details, including descriptions of the interventions, comparator, care pathway and outcomes, are in the final scope for software with AI derived algorithms for automated detection and analysis of lung nodules from CT scan images.

2 Clinical effectiveness evidence

The EAG did a systematic review to identify evidence on the clinical effectiveness and diagnostic accuracy of software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from

CT scan images. Find the methods and results on pages 61 to 186 of the diagnostics assessment report.

Overview of included studies

There were 27 studies, reported in 30 publications, that met the inclusion criteria. Table 1 summarises the characteristics of the 21 included studies where software was used as intended for use in the NHS, alongside clinician review of chest CT scans (find the 6 studies where the software was used only as a standalone intervention in table 2 in the appendix).

Of the 21 studies, 10 provided evidence in a screening population, 1 study in a symptomatic population and 2 in a surveillance population. In 7 studies, the indications for the chest CT varied and in 1 study the reason for the scan was unclear. No studies were found where software was used alongside clinician review in an incidental population. No eligible studies were found on JLD-01K, Lung AI, Lung Nodule AI, qCT-Lung or SenseCare-Lung Pro.

Of the included studies, 16 studies reported test accuracy outcomes and 4 studies concordance or agreement outcomes. Other outcomes reported included technical failure rate and radiologist reading time. No studies reported acceptability or experience of using the software, health-related quality of life or clinical outcomes. Find an overview of the included studies in table 2 on pages 71 to 76, and an overview of the outcome reporting in tables 3 to 5 on pages 77 to 79 of the diagnostics assessment report.

Table 1 Included studies on software used alongside clinician review of chest CT scans

Study	Study type	Software	Study population	Country
Abadia et al. (2021)	Retrospective test accuracy and multi-reader multi-case study	AI-Rad Companion Chest CT	Mixed	US
Hwang et al. (2021a)	Before-and-after study	AVIEW LCS+	Screening	South Korea
Hwang et al. (2021b)	Retrospective analysis of prospective cohort study	AVIEW LCS+	Screening	South Korea
Hwang et al. (2021c)	Prospective screening cohort	AVIEW LCS+	Screening	South Korea
Lancaster et al. (2022)	Multi-reader multi-case study	AVIEW LCS+	Screening	Russia
Singh et al. (2021)	Multi-reader multi-case study	ClearRead CT	Screening	US
Lo et al. (2018)	Multi-reader multi-case study	ClearRead CT	Screening	US
Milanese et al. (2018)	Multi-reader multi-case study	ClearRead CT	Unclear	Switzerland
Hsu et al. (2021)	Multi-reader multi-case study	ClearRead CT	Mixed (Screening)	Taiwan
Takaishi et al. (2021)	Multi-reader multi-case study	ClearRead CT	Mixed	Japan
Röhrich et al. (2022)	Multi-reader multi-case study	ClearRead CT	Mixed	Austria
Kozuka et al. (2020)	Multi-reader multi-case study	InferRead CT Lung	Symptomatic	Japan
Liu et al. (2019)	Multi-reader multi-case study	InferRead CT Lung	Mixed	China
Zhang et al. (2021)	Retrospective test accuracy and multi-reader multi-case study	InferRead CT Lung	Screening	China
Cohen et al. (2017)	Multi-reader multi-case study	Veolity	Surveillance	South Korea
Kim et al. (2018)	Multi-reader multi-case study	Veolity	Surveillance	South Korea
Hall et al. (2022)	Retrospective test accuracy study and multi-reader multi-case study	Veolity	Screening	UK
Jacobs et al. (2022)	Multi-reader multi-case study	Veolity	Screening	, US
Hempel et al. (2022)	Multi-reader multi-case study	Veye Lung Nodules	Mixed	Netherlands
Murchison et al. (2022)	Multi-reader multi-case study	Veye Lung Nodules	Mixed	UK
Park et al. (2022)	Multi-reader multi-case study	VUNO Med-LungCT AI	Screening	US

Study quality

The EAG assessed the quality of the 16 diagnostic accuracy studies using the QUADAS-2 tool with the QUADAS-C extension for comparative accuracy studies. The EAG had concerns over high risk of bias because in most studies the scan reviewers assessed the CT images in a controlled environment instead of clinical practice or the nodule detection threshold used in the software was not clearly pre-specified. In most studies, the reference standard consisted of a too small panel of reviewers or not experienced enough reviewers or both, or the reference standard reviewers participated in the software-assisted or unassisted scan review or the reference standard panel was not blind to the software-assisted review results. Many studies excluded more than 10% of participant data from the analysis. For all studies, the EAG had a high concern over their applicability related to the index test, study population or reference standard. Find more details of the QUADAS-2 assessment on pages 83 to 90 of the diagnostics assessment report.

To assess the quality of the 4 studies that reported on concordance or agreement, EAG used the COSMIN Risk of Bias tool for studies on reliability and measurement error of outcome measurement instruments. All studies were judged ‘doubtful’ (unclear risk of bias).

Quality of 1 study (Hwang et al. 2021c) was not assessed because the relevant study outcomes were not related to accuracy, concordance or agreement.

Intermediate outcomes

Accuracy to detect lung nodules

Overall there were 6 studies that compared the accuracy of radiologists’ CT scan review with and without AI software to detect lung nodules, and reported person-level sensitivity or specificity or both.

Three studies reported accuracy to detect nodules of any kind. One study on InferRead CT Lung was done in a symptomatic population and 2 studies, 1 on

InferRead CT Lung and 1 on ClearRead CT, reported results from a screening population.

There were 2 studies that reported accuracy to detect lung nodules that need follow up. The 2 studies, 1 on ClearRead CT and 1 on Veolity, were done in a screening population.

Two studies reported on the accuracy of detecting malignant lung nodules. The 2 studies, 1 on ClearRead CT and 1 on VUNO Med-LungCT AI, were done in a screening population.

All except 1 study (Hall et al. 2022; Veolity) reported a higher sensitivity for the software-assisted CT scan review compared with the review without software. In all except 1 study (Hsu et al. 2021; ClearRead CT), the specificity of the scan review with software was lower than the specificity of the review without software. Although in some the difference was not statistically significant.

Table 2 summarises the comparative, mainly person-level accuracy estimates in the symptomatic, screening and mixed populations. In Hsu et al. (2021) and Lo et al. (2018), the specificity was reported at a person-level but sensitivity at a per-nodule level. There were 4 further relevant comparative accuracy studies but in 1 study the version of the software used did not have nodule detection function (Singh et al. 2021). Three studies (Liu et al. 2019, Murchison et al. 2022, Takaishi et al. 2021), instead of any person-level accuracy estimates, reported only per-nodule results. This is less useful because it can only tell whether nodules were missed or wrongly detected but not tell whether people with nodules were missed or wrongly identified as having nodules. Find more details of comparative accuracy studies in table 8 on pages 93 to 95 of the diagnostics assessment report. Table 9 on pages 96 to 97 of the diagnostics assessment report describes the non-comparative accuracy studies.

Table 2 Comparative accuracy to detect lung nodules (person-level results)

Study	Software	Population	Study size	Detection target	Reference standard	Sensitivity of software-assisted review (95% CI)	Sensitivity without software (95% CI)	Specificity of software-assisted review (95% CI)	Specificity without software (95% CI)
Kozuka et al. (2020)	InferRead CT Lung	Symptomatic	117	Any nodules	Majority expert review (3 readers)	85.1% (79.8% to 89.5%)	68.0% (61.4% to 74.1%)	83.3% (51.6% to 97.9%)	91.7% (61.5% to 99.8%)
Zhang et al. (2021)	InferRead CT Lung	Screening	860	Any nodules	Consensus expert review (2 readers)	99% (97% to 100%)	43% (38% to 49%)	97% (95% to 98%)	100% (99% to 100%)
Hsu et al. (2021)	ClearRead CT	Screening (subset of mixed)	57	Any nodules	Consensus expert review (2 readers)	All reviewers: 79% (76% to 81%) Expert reviewers: 83% (79% to 86%)	All reviewers: 63% (59% to 66%) Expert reviewers: 73% (69% to 77%)	All reviewers: 81% (78% to 84%) Expert reviewers: 88% (85% to 91%)	All reviewers: 77% (74% to 80%) Expert reviewers: 86% (83% to 90%)
Lo et al. (2018)	ClearRead CT	Screening	324	Nodules to follow up	Consensus expert review (3 readers)	72.5% (SD 3.3%)	60.1% (SD 3.3%)	84.4% (SD 2.0%)	89.9% (SD 2.0%
				Malignant nodules	Biopsy findings for malignant, long-term follow-up for benign	80.0% SD 3.9%)	64.7% (SD 3.9%)	84.4% (SD 2.0%)	89.9% (SD 2.0%)

Study	Software	Population	Study size	Detection target	Reference standard	Sensitivity of software-assisted review (95% CI)	Sensitivity without software (95% CI)	Specificity of software-assisted review (95% CI)	Specificity without software (95% CI)
Hall et al. (2022)	Veolity	Screening	770	Nodules to follow up	All nodules identified by radiologists without software and nodules identified by the software-assisted radiographers confirmed by radiologist consensus	71% (65% to 76%)	91% (86% to 95%)	92% (91% to 94%)	97% (95% to 98%)
Park et al. (2022)	VUNO Med-LungCT AI	Screening	200	Malignant nodules	Positive cancer diagnosis in the same year (diagnostic method not reported)	91.6% (81.7% to 96.4%)	85.2% (74.2% to 92.0%)	Not reported	Not reported

Measurement accuracy

No relevant studies comparing nodule measurement accuracy with and without the AI software were found. One study (Kim et al. 2018) on Veolity evaluated the concordance of nodule diameter measurements between scan reviewers with and without AI software in people with previously detected sub-solid nodules. The EAG was not convinced this was a study representing people having CT surveillance because all participants had a surgical resection after the scan. There was no difference between diameter measurements of the whole sub-solid nodule with or without software, but the measurements of solid part only were larger with software.

The EAG also looked at intra- and inter-observer agreement outcomes. Table 3 on pages 77 and 78 of the diagnostics assessment report provides an overview of these and all the other detection and analysis outcomes the EAG's review included. This table also includes information on where to find further details.

Technical failure rate

There were 11 relevant studies that provided some information about the segmentation functioning of the software. Four studies, 3 on AVIEW Lungscreen (previous version of AVIEW LCS+), 1 on ClearRead CT and 2 on Veolity, were done in a screening population. Two studies on Veolity were in people with previously detected sub-solid nodules. Three studies, 2 on Veye Lung Nodules and 1 on contextflow SEARCH Lung CT, were in mixed populations.

Two of the studies reported on failure to process scans and 1 reported technical difficulties. Singh et al. (2021) noted that the reasons for the processing failures were artefacts, thick sections or missing images. The remaining 6 studies reported on issues in extracting the nodule shape or profile for measuring the size (segmentation). Hwang et al. (2021b) noted that segmentation failed more often in subsolid nodules. Cohen et al. (2017) noted inclusion of a blood vessel or a significant part of the chest wall and

inaccurate segmentation of the ground glass part of the nodule as reasons for segmentation failures in subsolid nodules. Hwang et al. (2021c) suggested that the segmentation failures in clinical practice (higher rate than in a controlled environment) were mostly rejections of segmentation results by radiologists and not failures of the software to segment nodules. Jacobs et al. (2021) noted that where the software segmented nodules and manual measuring of the nodule diameter was not needed, reviewers still manually tuned the segmentation parameters in 28% of the nodules. Table 3 shows the reported technical failure rates.

Table 3 Technical failure rates

Study	Software	Population	Study size	Technical failure type	Technical failure rate
Hwang et al. (2021a)	AVIEW LCS	Screening	4,666	Segmentation	13.4% of 4,990 nodules
Hwang et al. (2021b)	AVIEW LCS	Screening	10,424	Segmentation	8.7% of 10,080 nodules
Hwang et al. (2021c)	AVIEW LCS	Screening	3,353	Segmentation	1.1% in retrospective review 14.4% in review in clinical practice (range 0-57% for individual reviewers)
Singh et al. (2021)	ClearRead CT	Screening	150	Scan processing	18%
Hall et al. (2022)	Veolity	Screening	770	Scan processing	1.2% for reviewer 1 2.3% for reviewer 2
Jacobs et al. (2021)	Veolity	Screening	160	Segmentation	1.9% for 1 of the 7 reviewers 1.3% for 2 of the 7 reviewers 0% for 4 of the 7 reviewers
Cohen et al. (2017)	Veolity	Surveillance	73	Segmentation	9.6% using filtered back projection (FBP) algorithm 6.8% using model-based iterative reconstruction (MBIR) algorithm

Kim et al. (2018)	Veolity	Surveillance	89	Segmentation	6.4% of 109 nodules
Hempel et al. (2022)	Veye Lung Nodules	Mixed	50	Segmentation	Scan reviewers found 54.6% and 44.4% volumes not reliable without AI software and 2.4% and 4.5% with AI software
Murchison et al. (2022)	Veye Lung Nodules	Mixed	337	Segmentation	4.9% of 428 nodules
Röhrich et al. (2022)	Contextflow SEARCH Lung CT	Mixed	100	Technical difficulties	0.9% of 216 of scan readings

Radiologist reading time

Nine relevant studies reported on the effect using the software had on radiologist scan review time. One study on InferRead CT Lung was in a symptomatic population and 3 studies, 1 on ClearRead CT and 2 on Veolity, were in screening populations. Five studies, 1 AI Rad Companion Chest CT, 1 ClearRead CT, 1 contextflow SEARCH Lung CT, 1 InferRead CT Lung and 1 Veye Chest (previous name of Veye Lung Nodules), were in mixed populations. All studies suggested that scan review was faster with than without AI software and more time may be saved when scan reviewers are less experienced. One study (Hall et al.) compared the reading times of inexperienced readers using the software to reading times of experienced readers without the software in clinical practice. Table 4 shows the per scan reading time estimates from these studies.

Table 4 Radiologist reading time

Study	Software	Population	Study size	Reading time per scan with software	Reading time per scan without software
Kozuka et al. (2020)	InferRead CT Lung	Symptomatic	117	2.8 minutes (mean, no variance reported)	3.1 minutes (mean, no variance reported)

Lo et al. (2018)	ClearRead CT	Screening	324	98.0 seconds (mean, no variance reported)	132.3 seconds (mean, no variance reported)
Hall et al. (2022)	Veolity	Screening	770	Reader 1: 3 minutes (median, IQR 2 to 5 minutes); Reader 2: 5 minutes (median, IQR 4 to 8 minutes)	10 minutes (median, IQR 5 to 15 minutes)
Jacobs et al. (2021)	Veolity	Screening	160	86 seconds (median, IQR 51 to 141 seconds)	160 seconds (median, IQR 96 to 245 seconds)
Abadia et al. (2021)	AI Rad Companion Chest CT	Mixed	20 (of 143)	35.7 seconds (mean, no variance reported)	2 minutes 44 seconds (mean, SD 54 seconds)
Hsu et al. (2021)	ClearRead CT	Mixed	150	2 minutes 4 seconds (range 82 to 171 seconds)	2 minutes 36 seconds (range 100 to 227 seconds)
Röhrich et al. (2022)	contextflow SEARCH Lung CT	Mixed	108	222 seconds (mean, SD 156 seconds)	279 seconds (mean, SD 209 seconds)
Liu et al. (2019)	InferRead CT Lung	Mixed	271	About 5 to 10 minutes	About 15 minutes

Hempel et al. (2022)	Veye Chest (previous name of Veye Lung Nodules)	Mixed	50	Reader 1: 150.8 seconds (mean, SD 74.2 seconds) Reader 2: 184.2 seconds (mean, SD 125.3 seconds)	Reader 1: 226.4 seconds (mean, SD 113.2 seconds) Reader 2: 320.8 seconds (mean, SD 164.2 seconds)

Effect of test result on clinical decision-making

One study on Veye Chest (previous name for Veye Lung Nodules; Hempel et al. 2022) compared radiologists' accuracy with and without the software to identify when follow up is not needed (based on nodule size, BTS guidelines) in a mixed population (incidental and surveillance). Point estimates for sensitivity and specificity were higher for radiologists with AI software, but the 95% confidence intervals overlapped. This study also compared the agreement on nodule management recommendations between scan reviewers (based on nodule size, BTS guidelines). Agreement between the scan reviewers was better with than without software.

Acceptability and experience of using the software

No relevant studies were found that described software acceptability to patients or clinicians' experience of using the software.

Clinical outcomes

No studies were found on the effects of software-assisted review on clinical outcomes.

Health-related quality of life outcomes

No studies were found on the effects of software-assisted review on health-related quality of life outcomes.

Ongoing studies

The EAG identified 7 ongoing studies. Two of these assess the software as a standalone intervention. Table 5 lists the ongoing comparative studies. Find more details in table 67 in appendix 2 on pages 322 to 324 of the diagnostics assessment report.

Table 5 Ongoing comparative studies

Study	Software	Population	Study size	Country	Outcomes	Estimated completion date
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
NCT04119960 (2019)	InferRead CT Lung	Screening	250	US	Test accuracy	October 2019
NCT02871856 (2021) International Lung Screen Trial (ILST)	Veolity	Screening	4,500	Australia, Canada, Hong Kong, Spain	Test accuracy, radiologist reading time	December 2023
NCT04792632 (2021)	Veye Lung Nodules	Mixed	350	US	Test accuracy, composition classification accuracy, segmentation and growth assessment accuracy (software alone)	July 2021
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

3 Cost effectiveness evidence

The EAG did a systematic review to identify any published economic evaluations of software with AI derived algorithms for automated detection and analysis of lung nodules from CT scan images. The EAG also constructed de novo economic models to assess the cost effectiveness of AI-software assisted detection and analysis of lung nodules from CT scans in symptomatic, incidental and screening populations.

Systematic review of cost-effectiveness evidence

EAG found 2 models that included steps along the lung cancer care pathway. Bajre et al. (2017) assessed the cost-effectiveness of using trained radiographers compared with radiologists to report chest x-rays in people with suspected lung cancer in the NHS. This model also included parameters for further investigations and lung cancer treatment. Adams et al. (2021) compared the costs of lung cancer screening with and without AI-derived lung cancer risk score to support Lung-RADS screening recommendations in US. These models did not evaluate software for automated detection and analysis of lung nodules from CT scans. Find more details on pages 192 to 196 of the diagnostics assessment report.

Economic analysis

To evaluate the cost-effectiveness of software with AI derived algorithms for automated detection and analysis of lung nodules from CT scans, the EAG built a model for the symptomatic, incidental and screening populations. The model also captured CT surveillance. Because there was not enough nodule detection and measurement accuracy evidence on any of the individual technologies, the model used inputs from those technologies that had the most evidence and cost information available. Where the needed data was not available, the EAG generated technology-related model inputs using other sources of information. This means that the model is exploratory and so the results are illustrative.

Model structure

The model had 2 parts:

- a decision tree that captured detecting lung nodules, and the nodule type and size according to the British Thoracic Society (BTS) pulmonary nodule management guidelines
- a decision tree that captured further investigation and follow up of lung nodules according the BTS guidelines, and lung cancer treatment

Figure 1 presents the structure of the first part of the model. Figure 2 shows the the second part of the model for solid nodules and figure 3 for sub-solid nodules. The model structure was the same for symptomatic, incidental and screening populations. CT surveillance is captured in the second part of the model. The model had a time horizon of 10 years. The EAG considered that this would be long enough to capture the costs and benefits of using the software.

Figure 1 Model structure for initial detection of all lung nodules

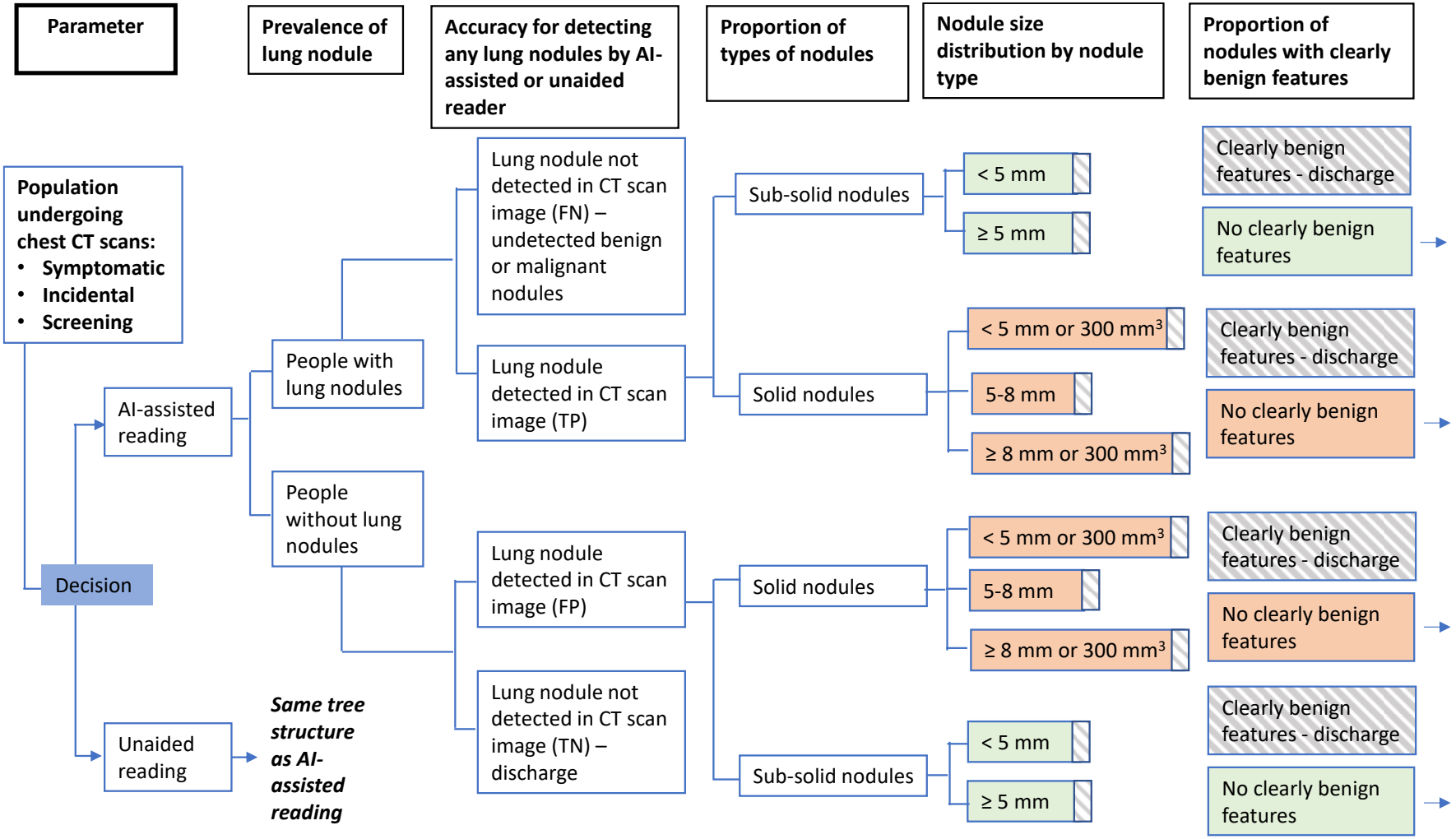


Figure 2 Model structure for investigation and follow up of solid nodules with no clearly benign features

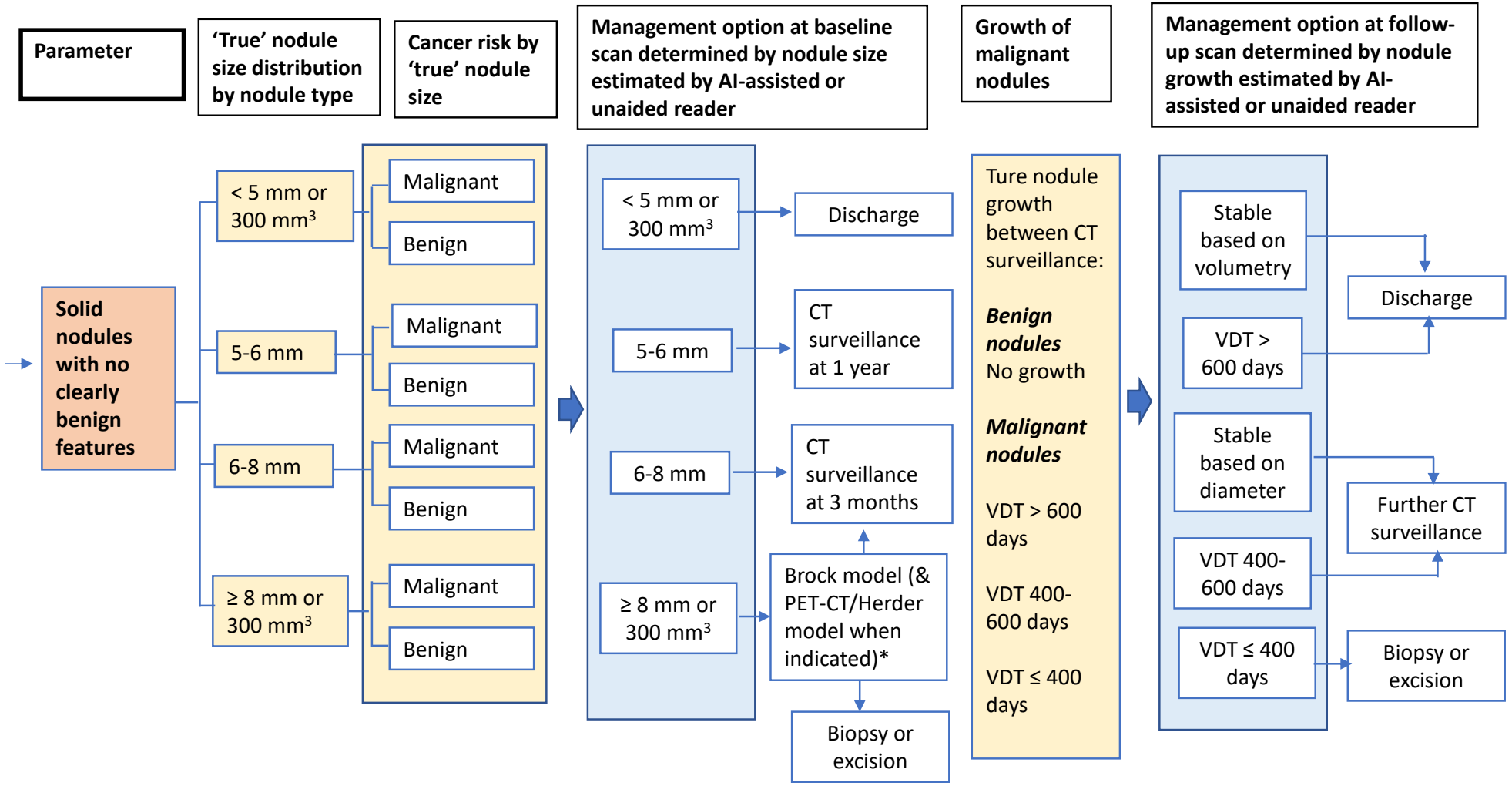
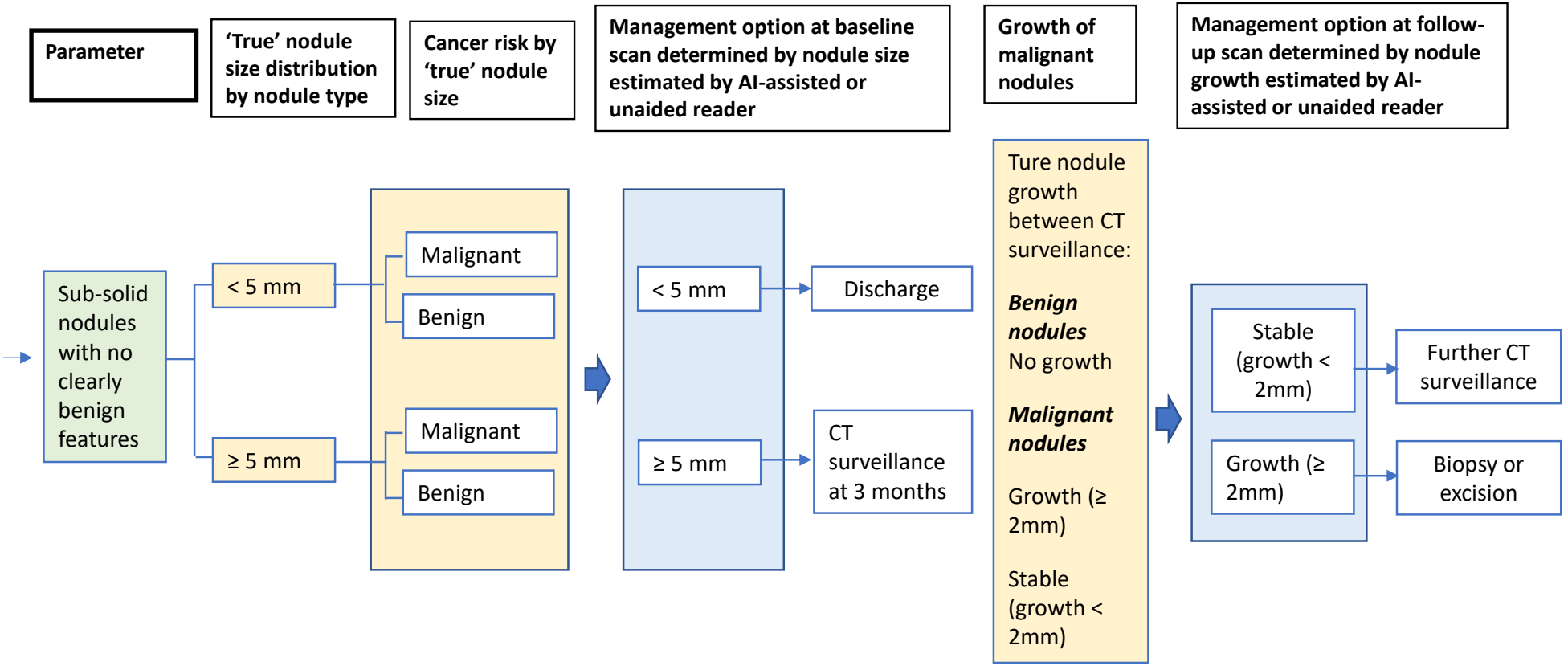


Figure 3 Model structure for investigation and follow up of sub-solid nodules with no clearly benign features



Population

The modelling was done for symptomatic, incidental and screening populations. The model also captured people having CT surveillance. People entering the model were 60 years old.

Comparator

The comparator CT scan review for lung nodules by a healthcare professional without assistance from AI-derived software.

Model inputs

Figures 18 to 20 on pages 214 to 216 of the diagnostics assessment report provide an overview of the model parameters and their sources.

Prevalence of lung nodules

Table 6 shows the prevalence of having at least 1 lung nodule of any kind in the 3 initial detection populations. When there were multiple nodules, the largest nodule was considered in the clinical decision making.

Table 6 Prevalence of lung nodules

Population	Prevalence	Source
Symptomatic	94.9% (95% CI 89.28% to 97.63%)	Kozuka et al. (2020)
Incidental	13% (range from 2% to 24%)	Callister et al. (2015)
Screening	50.9% (95% CI 48.68% to 53.12%)	Field et al. (2016)

Abbreviation: 95% CI, 95% confidence interval

Accuracy to detect lung nodules of any kind

Table 7 shows the accuracy estimates of software-assisted and unassisted CT scan review to detect lung nodules of any kind used in the model. For the screening population, the estimates used were from the experienced chest radiologist group in the study by Hsu et al. Because no data was available for the incidental population, the EAG used the accuracy reported in the symptomatic population study by Kozuka et al. In this study, the scans were reviewed by less experienced radiologists. Find more details about selecting

the accuracy estimates on pages 219 to 222 of the diagnostics assessment report.

Table 7 Accuracy of CT scan review with and without software to detect lung nodules of any kind

Population	Sensitivity software-assisted review (95% CI)	Sensitivity unassisted review (95% CI)	Specificity software-assisted review (95% CI)	Specificity unassisted review 95% CI)	Source (software)
Symptomatic	85.14% (79.80% to 89.50%)	68.02% (61.40% to 74.10%)	83.33% (51.60% to 97.90%)	91.67% (61.55% to 99.88%)	Kozuka et al. (2020; InferRead CT Lung)
Incidental	85.14% (79.80% to 89.50%)	68.02% (61.40% to 74.10%)	83.33% (51.60% to 97.90%)	91.67% (61.55% to 99.88%)	Kozuka et al. (2020; InferRead CT Lung)
Screening	83% (79% to 86%)	73% (69% to 77%)	88% (85% to 91%)	86% (83% to 90%)	Hsu et al. (2021; ClearRead CT)

Proportions of different types of lung nodules

Table 8 shows the proportions of solid and sub-solid nodules. Because no data was available for the incidental population, the EAG assumed that the proportion of solid and sub-solid nodules in this population was the same as in the screening population (study by Hwang et al.). In the model, there was no difference in the nodule type proportions across correctly detected nodules, missed nodules and structures that were incorrectly identified as nodules.

Table 8 Proportions of solid and sub-solid lung nodules

Population	Proportion of solid nodules	Proportion of sub-solid nodules	Source
Symptomatic	77.4%	22.6%	Kozuka et al. (2020)
Incidental	93.9%	6.1%	Hwang et al. (2021b)
Screening	93.9%	6.1%	Hwang et al. (2021b)

Lung nodule size distribution by nodule type

The EAG estimated the proportion of people with lung nodules in each of the size and type group needing different follow up (based on the BTS guidelines). The EAG's size distribution simulation was based on the nodule sizes reported in the screening study by Hwang et al. (2021b) and the symptomatic population study by Kozuka et al (2020). Because data was not reported separately for sub-solid nodules, EAG made assumptions about their median sizes. Because no data was available for the incidental population, the EAG assumed that the nodule size distribution in this population was the same as in the screening population. Table 9 shows the lung nodule size distribution by nodule type in the 3 initial detection populations. Find more details in the addendum to the diagnostics assessment report. Find more details of the simulation methods appendix 8 on pages 431 and 432 of the diagnostics assessment report.

Table 9 Lung nodule size distribution by nodule type

Nodule type	Nodule size in diameter	Prevalence in symptomatic population	Prevalence in incidental population	Prevalence in screening population
Solid	Less than 5 mm	56.5%	61.3%	61.3%
Solid	5 mm to less than 8 mm	24.9%	15.4%	15.4%
Solid	8 mm or larger	56.5%	61.3%	61.3%
Sub-solid	Less than 5 mm	13.9%	24.1%	24.1%
Sub-solid	5 mm or larger	86.1%	75.9%	75.9%

Proportion of nodules with clearly benign features

In the model, 10% of people in each nodule size and type group were discharged because they had a nodule with clearly benign features. This was assumed correct in both strategies.

Proportion of people assigned to different management options based on measured lung nodule size in CT scan review with and without software

The EAG used the lung nodule size distribution and study data on lung nodule size measurement errors to estimate the proportions of people assigned to different management options based on lung nodule size (BTS guidelines) measured by clinicians with and without the AI software on the initial CT scan. The nodule size measurement precision data using AI software for this simulation was available from the study by Martins Jarnalo et al. (2021). This study compared the performance of AI software Veye Chest (previous name of Veye Lung Nodules) used as a standalone intervention with a panel of 3 radiologists in a population with various indications for CT scan (mixed population study). For precision of manual nodule size measurement, the EAG used a lung phantom study by Xie et al. (2013). In this study, manual measurement using electronic calipers underestimated nodule size. The EAG assumed that the proportion of people assigned to different management options based on measured nodule size in CT scan review with and without software distribution in the incidental population was the same as in the screening population. Table 10 shows the proportion of people assigned to different management options in the 3 initial detection populations. Find more details in the addendum to the diagnostics assessment report. Find more details of the simulation methods in appendix 8 of the diagnostics assessment report.

Table 10 Proportion of people assigned to different management options based on measured nodule size in the initial CT scan review with and without software

Nodule type and size	Management option	Proportion of symptomatic population based on software-assisted review	Proportion of symptomatic population based on unassisted review	Proportion of incidental population based on software-assisted review	Proportion of incidental population based on unassisted review	Proportion of screening population based on software-assisted review	Proportion of screening population based on unassisted review
Solid, less than 5 mm	Discharge	50.0%	58.7%	57.1%	60.7%	57.1%	60.7%
Solid, 5 mm to less than 6 mm	CT surveillance in 1 year	31.9%	26.6%	28.2%	25.9%	28.2%	25.9%
Solid, 6 mm to less than 8 mm	CT surveillance in 3 months	12.9%	10.1%	8.9%	8.0%	8.9%	8.0%
Solid, 8 mm or larger	Assessing cancer risk using Brock model	5.1%	4.6%	5.7%	5.4%	5.7%	5.4%
Sub-solid, less than 5 mm	Discharge	22.0%	28.8%	28.1%	35.4%	28.1%	35.4%
Sub-solid, 5 mm or larger	CT surveillance in 3 months	78.0%	71.2%	71.9%	64.6%	71.9%	64.6%

Risk of lung cancer by true lung nodule size

Table 11 shows the risk or prevalence of lung cancer by nodule size in people who had nodules needing follow up. This was based on the Horeweg et al. (2014) study of 7,155 people in the screening group of the Dutch NELSON lung cancer screening trial. The study did not provide separate data for solid and sub-solid nodules so the prevalence of both was assumed the same as the prevalence of both nodule types combined. The prevalence of lung cancer was assumed the same in the symptomatic, incidental and screening populations, and people having CT surveillance populations.

Table 11 Risk of lung cancer by true lung nodule size

Lung nodule type	Lung nodule size (diameter)	Risk (95% CI)
Solid	5 mm to less than 6 mm	0.89% (0.5% to 1.6%)
Solid	6 mm to less than 8 mm	1.1% (not reported)
Solid	Larger than 8 mm	9.4% (not reported)
Sub-solid	5 mm or larger	3.6% (not reported)

Lung nodule growth

The EAG derived nodule growth using a simulation model. It assumed that growth of malignant nodules followed a Gompertz distribution and was conditional on volume-doubling time, based on Treskova et al. (2017). Nearly all solid (99.9%) and sub-solid (99.4% to 99.5%) nodules were expected to stay stable and not grow throughout the CT surveillance period in all the populations. Find more details of the simulation methods in appendix 7 of the diagnostics assessment report.

The time to read and report a CT scan

Table 13 shows the times to read and report a CT scan used in the model. The time to read and report a CT scan without AI software in the screening population was based on the UK Lung Screen Uptake Trial (Hall et al. 2022). Based on the results of the studies included in the EAG's clinical effectiveness review, the reading and reporting was assumed slightly faster when review

was software-assisted. The reading and reporting times for scans in symptomatic and incidental population were assumed longer. This was because clinical experts suggested that in addition to lung nodules, there may be other findings to report. In these settings, there are also likely to be more interruptions to the scan review compared with the screening setting where scans are often reviewed in batches during a protected time.

Table 13 The time to read and report a CT scan

Radiologist time to read and report a CT	Symptomatic population	Incidental population	Screening population
Without AI software	15 minutes	15 minutes	10 minutes
With AI software	12 minutes	12 minutes	8 minutes

Costs

Find the full list of costs used in the model in table 46 on pages 226 and 227 of the diagnostics assessment report.

Cost of detecting lung nodules

Table 14 lists the costs of detecting lung nodules in the initial CT scan.

Table 14 Costs of detecting lung nodules in the initial CT scan

Cost parameter	Symptomatic population	Incidental population	Screening population	Source
CT scan	£145 (single area, with contrast)	£145 (single area, with contrast)	£106 (single area, no contrast)	NHS reference schedule 2020/21
Radiologist time to read and report the scan without AI software (cost per working hour [£147] for a band 9 radiographer used as a proxy)	£36.75 (15 minutes)	£36.75 (15 minutes)	£24.50 (10 minutes)	Personal Social Services Research Unit (PSSRU) 2021
Radiologist time to read and report the	£29.40 (12 minutes)	£29.40 (12 minutes)	£19.60 (8 minutes)	PSSRU 2021

scan with AI software (cost per working hour [£147] for a band 9 radiographer used as a proxy)				
AI software to assist CT scan review	£3.34 per scan	£3.34 per scan	£2.00 per scan	Infervision, Riverain Technologies
Total cost per person	£181.75 without software £177.74 with software	£181.75 without software £177.74 with software	£130.50 without software £127.60 with software	

The EAG used the average per scan cost of InferRead CT Lung in the symptomatic and incidental populations and average per scan cost of ClearRead CT in the screening population because the accuracy estimates of lung nodule detection in the model were from the studies using these technologies. Table 15 shows the available software cost information. Not all software use a per scan pricing model and the per scan cost may depend for example on the expected yearly number of CT scans to be reviewed. The per scan cost estimates are provided by the manufacturers and may not all be based on the same assumptions.

Table 15 AI software costs

Software	Per scan cost
AI-Rad Companion Chest CT	Not known
AVIEW LCS+	Not known
ClearRead CT	£1.50 to £2.50
contextflow SEARCH Lung CT	£2.50 (minimum 5000 scans per year, and an initial set up cost of £5,000)
InferRead CT Lung	£2.67 to £4.00 (and a yearly maintenance fee of £8,000 and an initial set up cost of £3,000)
JLD-01K	Not known
Lung AI	Not known
Lung Nodule AI	Not known
qCT-Lung	Not known
SenseCare-Lung Pro	Not known

Veolity	Per scan cost not known (yearly license fee of £44,000 for 3 concurrent users with a yearly maintenance fee of £8,800 from year 2 onwards, or a monthly license fee of £2,100 for 3 concurrent users with minimum license period 3 years)
Veye Lung Nodules	£4.00 to £6.00 (and a yearly fee of £9,000 and an initial set up cost of £8,000 to £12,000)
VUNO Med-LungCT AI	Not known

Cost of further investigations and treating lung cancer

Table 16 lists the costs of diagnosing, managing and treating lung cancer. Costs from older sources were updated to 2020/21 costs using the Hospital and Community Health Services (HCHS) index from Unit Costs of Health and Social Care 2022.

Cost of each round of CT surveillance was assumed to include the same costs as detecting lung nodules in the initial CT scan (see table 14 on page 27 but CT scan without contrast was used in all populations. Additionally, people having CT surveillance had a multidisciplinary team meeting and a guided needle biopsy if nodule growth with a volume-doubling time of 400 days or faster for solid nodules or growth of at least 2 mm for sub-solid nodules was measured.

Table 16 Costs of further investigations and treating lung cancer

Parameter	Cost	Source
Multidisciplinary team meeting	£146	NHS reference schedule 2020/21
Guided needle biopsy	£1,670	NHS reference schedule 2020/21
PET-CT scan (one area, 19 years and over)	£1,161	NHS reference schedule 2020/21
Treatment for stage 1 lung cancer	£18,705	Cancer Research UK (2014) from Bajre et al. (2017)
Treatment for stage 2 lung cancer	£21,312	Cancer Research UK (2014) from Bajre et al. (2017)
Treatment for stage 3 lung cancer	£23,922	Cancer Research UK (2014) from Bajre et al. (2017)

Treatment for stage 4 lung cancer	£14,909	Cancer Research UK (2014) from Bajre et al. (2017)

QALY decrements and health-related quality of life

The EAG assigned a utility decrement of -0.063 to people who had a falsely detected lung nodule or who were having CT surveillance for nodules that were later diagnosed as benign. They were discharged at the first surveillance scan 3 months or 1 year later. The EAG conducted a scenario analysis in which the utility decrement was assigned to all people having CT surveillance (results of this are in the addendum to the diagnostics assessment report). People with benign nodules and people without lung nodules were assigned a utility value of 0.855, based on Rickets et al. (2020), representing health-related quality of life of the UK general population.

A utility decrement of -0.2 for 3 months was associated with having a biopsy.

Table 17 shows the utility values used in the base case for people with lung cancer, from a Canadian study by Naik et al. (2015).

Table 17 Utility values for lung cancer stages

Lung cancer stage	Utility value
Stage 1	0.81
Stage 2	0.77
Stage 3	0.76
Stage 4	0.76

Mortality

The EAG's model considered lung cancer death and death from other causes. Survival after lung cancer treatment was from Exeter Natural History-Based economic model of Lung cancer screening (ENaBL) model (10-year survival based on cancer stage at diagnosis). For people without lung cancer, the EAG's model used the average of general population mortality for women and men from Office for National Statistics. Most people in the symptomatic and

screening population were assumed to smoke and so have a 30% higher risk of death than the general population (based on Jacobs et al. 1999). Smoking was not assumed to be common in the the incidental population (based on Zhou et al. 2023).

Further assumptions

- All AI software for automated detection and analysis of lung nodules from CT scan images are equal
- No new lung nodules develop after the initial CT scan, any nodules or cancer that are detected later are assumed to have been missed earlier
- No relevant evidence was available in the incidental population, so it was assumed that the population is similar to the screening population for many model inputs
- In the AI software-assisted strategy, 95% of people with stable (not growing), solid nodules are discharged after 1 year and 5% after 2 years in CT surveillance, because BTS guidelines suggest that when nodules are found stable using volumetry, people can be discharged earlier
- In the strategy without the software it was assumed that most growth measurements were done without volumetry and so only 5% of people with stable, solid nodules are discharged after 1 year and 95% people after 2 years in CT surveillance
- In all the 3 initial detection populations, 0.4% of people with a false negative result for nodules of any kind have lung cancer
- There are no cancers caused by radiation exposure

The effect of some of these assumptions on the model results was explored in sensitivity and scenario analyses.

Base case results

Cost effectiveness of software-assisted review of CT scans for lung nodules

Based on the model, in the symptomatic population the software-assisted review of CT scans for lung nodules was more costly and less effective compared with review of the scans without the software. In the incidental population it was slightly less costly but also less effective. In the screening population and those in the screening population having CT surveillance, the software-assisted review was less costly and more effective compared with review of the scans without the software. The EAG did not do surveillance population analysis for the symptomatic and incidental populations. Table 18, 19 and 20 summarise the results of the probabilistic analysis for the 3 initial detection populations. The results of the deterministic analysis were similar. Table 21 summarises the deterministic results for the surveillance population. The EAG did not do probabilistic analysis in this population.

In all the populations, the software-assisted review of the scans detected a larger number of people with nodules that needed follow up compared with scan review without the software. This difference in the number detected was largest in the screening population. In all the populations, the software-assisted review also detected slightly more lung cancers. Find more details of the intermediate model outcomes in table 49 of the diagnostics assessment report.

Table 18 Probabilistic cost-effectiveness results for symptomatic population

Strategy	People with nodules needing follow up detected	People with lung cancer detected	Total costs	Total QALYs	Incremental costs (software-assisted CT scan review compared to clinician's review)	Incremental QALYs (software-assisted CT scan review compared to clinician's review)	ICER
Clinician's CT scan review	333 per 1,000 people	12 per 1,000	£714,680	6350.00	-	-	-
Software-assisted CT scan review	481 per 1,000 people	15 per 1,000	£816,660	6329.80	£101,980	-20.2	Dominated

Abbreviations: QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio

Table 19 Probabilistic cost-effectiveness results for incidental population

Strategy	People with nodules needing follow up detected	People with lung cancer detected	Total costs	Total QALYs	Incremental costs (software-assisted CT scan review compared to clinician's review)	Incremental QALYs (software-assisted CT scan review compared to clinician's review)	ICER
Clinician's CT scan review	42 per 1,000 people	1 per 1,000	£231,370	6573.74	-	-	-
Software-assisted CT scan review	58 per 1,000 people	2 per 1,000	£228,870	6571.26	-£2,500	-2.48	1,008 (South-West ICER)

Abbreviations: QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio

Table 20 Probabilistic cost-effectiveness results for screening population

Strategy	People with nodules needing follow up detected	People with lung cancer detected	Total costs	Total QALYs	Incremental costs (software-assisted CT scan review compared to clinician's review)	Incremental QALYs (software-assisted CT scan review compared to clinician's review)	ICER
Clinician's CT scan review	178 per 1,000 people	7 per 1,000	£470,080	6524.16	-	-	Dominated
Software-assisted CT scan review	223 per 1,000 people	8 per 1,000	£400,200	6532.14	-£69,880	7.98	-

Abbreviations: QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio

Table 21 Deterministic cost-effectiveness results for screening population having CT surveillance

Strategy	Total costs	Total QALYs	Incremental costs (software-assisted CT scan review compared to clinician's review)	Incremental QALYs (software-assisted CT scan review compared to clinician's review)	ICER
Clinician's CT scan review	£921,015	6323.07	-	-	Dominated
Software-assisted CT scan review	£718,813	6365.01	-£201,202	41.94	-

Abbreviations: QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio

Analysis of alternative scenarios

The robustness of the cost effectiveness results to alternative model assumptions was considered in 5 scenario analyses in each of the 3 initial detection populations:

- scenario 1: prevalence of lung nodules
    - in the symptomatic population was dropped to 50% (from 94.9%)
    - in the incidental population was increased to 38% (from 13%)
    - in the screening population was dropped to 33% (from 50.9%)
- scenario 2: reading and reporting a CT scan without the software was assumed to take the same time in both the strategy with and the strategy without software
- scenario 3: reading and reporting a CT scan was assumed to take longer time with than without the software
- scenario 4: people with solid nodules later diagnosed as benign were discharged after 2 years and people with sub-solid nodules later diagnosed as benign after 4 years of CT surveillance in both strategies (instead of 95% people after 1 year and 5% after 2 years in the strategy with the software, and 5% people after 1 year and 95% people after 2 years in the strategy without the software)
- scenario 5: utility decrement of -0.063 assigned to people who had a falsely detected lung nodule or who were having CT surveillance for nodules that were later diagnosed as benign was removed from the model

The EAG did an additional scenario in the screening population using sensitivity and specificity estimates for detecting people with lung nodules needing follow up. Accuracy estimates were taken from the study by Lo et al. (2018; ClearRead CT). This scenario only looked at the initial detection part of the model, not the downstream costs and QALYs. Find more details in section 6 on pages 197 to 207 of the diagnostics assessment report.

The EAG also did 1 scenario analysis for the CT surveillance model. This scenario assumed that all lung cancers missed during surveillance would be diagnosed later as stage 4 cancer instead of stage 1.

Scenario analyses

In the symptomatic population, when the utility decrement of -0.063 assigned to people who had a falsely detected lung nodule or who were having CT surveillance for nodules that were later diagnosed as benign was removed from the model, the ICER for the software-assisted review strategy became £12,709 per QALY gained. In all other scenarios, like in the base case, current practice dominated the software-assisted strategy. Find more details in table 53 on pages 239 to 241 of the diagnostics assessment report.

In the incidental population, removing the utility decrement of -0.063 and increasing the prevalence of lung nodules made the software-assisted strategy dominate current practice. In the other 3 scenarios, current practice dominated the software-assisted strategy. Find more details in table 57 on pages 245 and 246 of the diagnostics assessment report.

In the screening population, software-assisted strategy dominated current practice, like in the base case, in all the 5 scenarios. Find more details in table 61 on pages 250 and 251 of the diagnostics assessment report.

The results of the scenario analysis in the surveillance population were like in the base case. Find more details on page 254 of the diagnostics assessment report.

Sensitivity analyses

The EAG varied values of several model inputs in 1-way sensitivity analyses. Table 22 lists the model parameters that had most influence on cost-effectiveness estimates. Find more details in figure 21 on page 238, figure 22 on page 244 and figure 23 on page 249 of the diagnostics assessment report.

Table 22 Important model parameters for cost-effectiveness estimates based on 1-way sensitivity analyses

Population	Model parameter	Notes
Symptomatic	• sensitivity of CT scan review without software • time to read and report a CT scan with software • time to read and report a CT scan without software	Within the ranges used in the analysis, current practice still dominates the software-assisted strategy
Incidental	• prevalence of lung nodules	Higher prevalence of nodules (towards 24%) is associated with a more favourable ICER for software-assisted strategy
Screening	• time to read and report a CT scan with software • time to read and report a CT scan without software • specificity of CT scan review without software (to detect nodules) • sensitivity of software-assisted CT scan review with software (to detect nodules) • specificity of software-assisted CT scan review with software (to detect nodules)	Within the ranges used in the analysis, software-assisted strategy still dominates current practice

ICER, incremental cost-effectiveness ratio

The EAG also varied some of the model inputs simultaneously in probabilistic sensitivity analysis. Based on these analyses, at a maximum acceptable ICER of £20,000 or £30,000 per QALY gained, software-assisted scan review is very unlikely to be a cost-effective intervention in the symptomatic population, there is high uncertainty of its cost effectiveness in the incidental population, but it is very likely to be a cost-effective intervention in the screening population. Find more details in appendix 9 on pages 446 to 448 of the diagnostics assessment report.

4 Summary

Clinical effectiveness

Limited evidence was found, particularly in the symptomatic, surveillance and incidental populations. In the surveillance population there were 2 studies, in the symptomatic population only 1 study and in the incidental population there were no studies where the AI software was used alongside clinician CT scan review rather than as a standalone intervention. No eligible studies were found on JLD-01K, Lung AI, Lung Nodule AI, qCT-Lung or SenseCare-Lung Pro software. No study reported data for more than 1 software.

Six studies on 4 different software provided person-level data on accuracy to detect nodules comparing clinician CT scan review with and without AI software. But these studies reported on different outcomes, some on accuracy to detect any kind of nodules, some nodules that needed follow up and some nodules that were malignant. Nearly all studies found that CT scan review with AI software was more sensitive but less specific than without the software. In these studies, the reference standards consisted of mostly small panels that may not have been 100% accurate. Reported technical failure rates were generally low and had mostly to do with nodule segmentation. No studies provided data on measurement accuracy.

Eleven studies reported on radiologist reading time. All suggested that scan review was faster with than without AI software and more time may be saved when scan reviewers are less experienced. But none of the comparisons of reading times were fully done in clinical practice.

No studies described acceptability of the software to patients or clinicians experience using it. No studies were found on the effects of software-assisted review on clinical or health-related quality of life outcomes for any populations.

Cost effectiveness

The EAG built a model to assess the cost-effectiveness in the symptomatic, incidental and screening populations and CT surveillance. But because there

was not enough nodule detection and measurement accuracy evidence on any of the individual technologies, the model combined data from various sources to assess a hypothetical software and so the results are only illustrative.

Based on the model, in the symptomatic population the software-assisted review of CT scans for lung nodules was more costly and less effective compared with review of the scans without the software. In the incidental population it was slightly less costly but also less effective. In the screening population and those in the screening population having CT surveillance, the software-assisted review was less costly and more effective compared with review of the scans without the software.

Many of the data inputs for the screening population differed from those from the other two populations, because there were different data sources and more data available including from screening trials. The driving force behind AI assistance estimates being cost effective for screening and not for the other two populations are in the estimated number of false positive results and people undergoing CT surveillance. The ICERs are heavily influenced by the costs and QALY decrements associated with false positive results and CT surveillance.

Sensitivity of CT scan review without software, time to read and report a CT scan with and without software were important model parameters in the symptomatic population. In the screening population, sensitivity and specificity of the CT scan review with and without software, and time to read and report a CT scan with and without software were important. But one-way sensitivity analyses on these parameters did not result in a change in direction of the base-case ICERs. Nodule prevalence was important in the incidental population. Increased prevalence was favourable towards AI-assisted review because it was estimated to have greater sensitivity to detect these nodules in the model.

Key assumption in the model was that having CT surveillance for nodules is associated with a utility decrement of -0.063 if they are diagnosed as benign (based on no growth) at the first surveillance scan 3 months or 1 year later. When this disutility was removed in scenario analyses, cost-effectiveness for symptomatic and incidental populations improved, but the estimates for the screening population became less favourable. This is driven by differing data inputs, for example the screening data suggests AI is more specific whereas the symptomatic and incidental data used suggested the unaided reader was more specific. It is likely to be a more plausible assumption that the disutility associated with having CT surveillance is not limited to having had surveillance for a nodule was not found to be malignant but that having CT surveillance is associated with anxiety of not knowing and so a disutility regardless of its result. Further analysis is provided in the addendum to the diagnostics assessment report.

The model is limited because of the limited data available to populate it.

5 Issues for consideration

Clinical effectiveness

Current evidence available on software to detect and analyse lung nodules on CT scan images is limited in quantity and quality. Table 1 in the appendix provides a summary of the nodule detection accuracy data that was available for each of the technologies in symptomatic, incidental and screening populations.

To better understand the clinical effectiveness of the individual technologies, more data on detection and measurement accuracy and the effect of the test results on clinical decision-making is needed. Better data is also needed to understand the effect of software-assisted review on radiologist reading time.

The committee should consider the following issues:

- study populations

- detection target (detection of any nodules that need follow up or detection of any nodules)
- unit of analysis (person-level versus per-nodule analysis)
- location of the study
- appropriate reference standard
- setting of the study (for example routine clinical practice, particularly relevant to radiologist reading time)

Cost effectiveness

The exploratory modelling shows that the cost-effectiveness of the AI software for detecting and analysing lung nodules on CT scan images depends on whether the benefits of detecting more cancers outweighs the costs and consequences associated with more people having CT surveillance. The number of people having CT surveillance in the EAG's model depends on the prevalence of lung nodules in the populations the software is used for, and the accuracy of detecting and measuring lung nodules. Data on these parameters is weak, leading to uncertainty in the model results.

Only 1 study (Kozuka et al. 2020 on InferRead CT Lung) was available in the symptomatic population, and so the model in this population relies on only 1 study for prevalence and detection accuracy inputs. There was no data on nodule measurement accuracy. This parameter was generated by the EAG using other sources. The model results in the symptomatic population show a large QALY loss due to harms from CT surveillance.

More studies were available in the screening population but only 2 reported on comparative estimates for the detection accuracy parameters in the EAG's model, the sensitivity and specificity of detecting any nodules (Zhang et al. 2021 on InferRead CT Lung; Hsu et al. 2021 on ClearRead CT). Because in Zhang et al. (2021) all the scan reviews with software were done twice and in laboratory conditions but the scan reviews without software were by different readers in clinical practice, the EAG decided to use the expert reader results from the 57-people screening subgroup of the Hsu et al. (2021) study. There

was no data on nodule measurement accuracy. This parameter was generated by the EAG using other sources. Prevalence of any nodules in the screening population in the model was from Field et al. (2016), the largest UK study EAG found to report the prevalence of any nodules. The model results suggested that in the screening population reviewing CT scans using AI software may cost less and be more effective compared with reviewing the scans without the software.

No eligible studies were found in the incidental population and so the most inputs for the model in this population were assumed to be the same as in the screening population. But the accuracy to detect any nodules was taken from the symptomatic population study by Kozuka et al. (2020) because in this study the scans were reviewed by less experienced radiologists. The EAG judged this to be applicable for general radiologists assessing CT images in the emergency department in UK practice. Prevalence of any nodules was from the evidence review for 2015 British Thoracic Society (BTS) pulmonary nodule management guidelines. The model results from the incidental population show a small QALY loss.

The committee should consider the following questions:

- Does the AI software have the potential to be cost effective in the symptomatic, incidental or screening population?
- If so, which technologies have enough data to say they have the potential to be cost-effective in these populations?
- Is the QALY loss seen in the models for symptomatic and incidental populations likely to be seen in clinical practice?

6 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

Lung cancer is considered as a disability and so people with lung cancer are protected under the Equality Act 2010. Incidence rates for lung cancer in the UK are highest in people between ages 85 and 89 (Cancer Research UK 2016-2018). Lung cancer is more common in men than in women. In men, lung cancer is most common in white men and men of Bangladeshi family background. In women, lung cancer is most common in white women. The incidence and mortality of lung cancer are higher in deprived communities.

Some people may find it challenging to lie still and to hold their breath or both during a chest CT scan. Some people may find it difficult to understand the instructions for what to do during the scan.

The AI-derived software may not perform as well in certain populations (such as different ethnic groups or people with lung conditions other than cancer) if these populations were underrepresented in the data used to develop and validate the software.

7 Implementation

Integration into radiologists' workflow

If the software does not fully integrate into the radiologists' workflow within the Picture Archiving and Communication System (PACS) where CT scan images are reviewed and reported, adds steps to the scan review, or does not include rules for reporting lung nodules in the NHS, using the software may increase review time. Clinical experts have raised concerns about how the software integration might affect the stability of the PACS system.

IT capacity and compatibility

There are some concerns about the level of IT support and capacity needed to install and use the software. There are also concerns about the software's ability to analyse images created using different CT scanners and its compatibility with other computer packages or systems. A report generated by an external software may not be compatible with the Radiology Information

System (RIS). It is also possible that the CT scan where the nodule is first identified is done at a different centre than the follow up CT scan. If different lung nodule software are used to assist to review the scans, it may be difficult for the reporting radiologist to compare the scans and assess the nodule growth.

Governance issues

When the software use cloud-based servers for the image analysis, there may be issues about adequate protection of patient data. There may also be questions about what software updates (potentially automatic) might mean for the clinical performance of the software.

8 Authors

Suvi Härmälä
Topic lead

Alice Bell
Topic support

Frances Nixon
Technical adviser

Glossary

CT surveillance

In CT surveillance, people with previously identified lung nodules have further CT scans done to assess whether the growth of the nodules indicates malignancy and if further assessment or treatment is needed.

Lung nodules

Lung nodules are small growths found inside the lung. While most lung nodules are benign, some may be cancerous and develop into lung cancer. Lung nodules can be detected from chest CT scans.

Targeted lung cancer screening

The UK National Screening Committee recommends targeted screening for lung cancer for people aged 55 to 74 identified as being at high risk of lung cancer.

Targeted Lung Health Checks programme

The Targeted Lung Health Checks (TLHC) programme provides a starting point for implementation of targeted lung cancer screening in England. The standard protocol for the Targeted Lung Health Checks programme (TLHC) includes specific requirements for radiologists reviewing the CT scans in the programme.

Volume-doubling time (VDT)

Volume-doubling time is the time in days it takes for a growing lung nodule to double its volume. It is calculated after follow up CT scans in surveillance of lung nodules to assess whether the nodule is likely to be malignant and further assessment or treatment is needed.

Appendix

Table 1 Availability of person-level accuracy and cost data on the technologies in the scope

Product name (manufacturer)	CT scan types	Available cost data	Accuracy data available in symptomatic population	Accuracy data available in incidental population	Accuracy data available in screening population	Accuracy data available in mixed population
AI-Rad Companion Chest CT (Siemens Healthineers)	Low dose, regular dose with and without contrast	No	No	No	No	No
AVIEW LCS+ (Coreline Soft)	Low dose (information from public domain)	No	No	No	No	No
ClearRead CT (Riverain Technologies)	Low dose, regular dose with and without contrast	Cost per scan	No	No	Any nodules, nodules to follow up, malignant nodules (sensitivity at per-nodule level only)	No
contextflow SEARCH Lung CT (contextflow)	With and without contrast	Cost per scan plus initial fee	No	Not applicable (not intended for use)	Not applicable (not intended for use)	No
InferRead CT Lung (Infervision)	Low dose, regular dose with and without contrast	Cost per scan plus initial and yearly maintenance fees	Any nodules	No	Any nodules	No
JLD-01K (JLK Inc.)	Without contrast	No	No	No	No	No

Lung AI (Arterys)	Low dose, regular dose with and without contrast (information from public domain)	No	No	No	No	No
Lung Nodule AI (Fujifilm)	To be confirmed	No	No	No	No	No
qCT-Lung (Qure.ai)	Without contrast (information from public domain)	No	No	No	No	No
SenseCare-Lung Pro (SenseTime)	Without contrast (information from public domain)	No	No	No	No	No
Veolity (MeVis)	Low dose, regular dose with and without contrast	Yearly license and maintenance fees or monthly license fee	No	No	Nodules to follow up	No
Veye Lung Nodules (Aidence)	Low dose, regular dose with and without contrast	Cost per scan plus initial and yearly maintenance fees	No	Np	No	No
VUNO Med-LungCT AI (VUNO)	Low dose (information from public domain)	No	No	No	Malignant nodules (sensitivity only)	No

Please note: specific indications for use for some of the technologies are unclear because only information in the public domain was available

Table 2 Studies of AI software as a standalone intervention

Study	Study type	Software	Study population	Country
Chamberlin et al. (2021)	Accuracy	AI-Rad Companion	Screening	US
Rueckel et al. (2021)	Accuracy	AI-Rad Companion	Incidental	Germany
Wan et al. (2020)	Accuracy	ClearRead CT	Mixed	Taiwan
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Blazis et al. (2021)	Accuracy	Veye Lung Nodules	Mixed	Netherlands
Martins Jarnalo et al. (2021)	Accuracy	Veye Lung Nodules	Mixed	Netherlands