Diagnostic Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence – Final protocol

Title of project

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images [DAP60]

Name of External Assessment Group (EAG) and project lead

Produced by: Warwick Evidence

Authors: Julia Geppert
Peter Auguste
Asra Asgharzadeh
Anna Brown
Surangi Jayakody
Hesam Ghiasvand
Mubarak Patel
Emma Helm
Dan Todkill
Jason Madan
Chris Stinton
Yen-Fu Chen
Sian Taylor-Phillips

Correspondence to: Dr Yen-Fu Chen
Populations, Evidence and Technologies
Division of Health Sciences
Warwick Medical School
University of Warwick
[redacted]

Tel: [redacted]

Email: [redacted]

Date completed: 6th December 2021 (Revised 8th December 2021)

The views expressed in this protocol are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors. The authors have no conflicts of interest.

Glossary of terms	
A&E	Accident and emergency
AI	Artificial intelligence
BTS	British Thoracic Society
CASP	Critical Appraisal Skills Programme
CEAC	Cost-effectiveness acceptability curve
CT	Computed tomography
DAC	Diagnostic Advisory Committee
EAG	External Assessment Group
EBUS-TBNA	Endobronchial ultrasound-guided transbronchial needle aspiration
EUS-FNA	Endoscopic ultrasound-guided fine-needle aspiration
HR	Hazard ratio
HSROC	Hierarchical summary ROC model
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
LY	Life-years
MRI	Magnet resonance imaging
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PACS	Picture archiving and communication system
PET	Positron emission tomography
PSS	Personal Social Services
PPSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life-years
RCT	Randomised controlled trial
ROC	Receiver operating characteristic
TLHC	Targeted Lung Health Check
UK NSC	UK National Screening Committee
VDT	Volume doubling time
WTP	Willingness-to-pay

1 Plain English Summary

Lung cancer is one of the most common types of cancer in the UK. People in the early stages of the disease may not have symptoms and so lung cancer is often diagnosed late. Lung nodules are small abnormal areas of tissue in the lung. Many of the nodules are harmless but some of them could present early stages of lung cancer.

Detecting lung nodules, examining their features, and monitoring their change in size and appearance over time on a CT (computed tomography) scan that includes the chest can help doctors identify cancer at an earlier stage and deal with it more quickly. However, detecting and monitoring lung nodules that do not develop into cancer can worry people unnecessarily. Detecting and measuring nodules in CT images can be difficult for several reasons, including the small size of the nodules, different shapes, their location, and how close they are to other lung structures. At present, most radiologists (doctors who specialise in medical imaging) or other healthcare professionals detect lung nodules on CT scan images without assistance from any computer software but may use software to measure the size of a nodule after they have spotted it.

Computer software capable of detecting and analysing lung nodules automatically could be used to assist the healthcare professional when reviewing CT scan images that include the chest. Such software could detect lung nodules that might have been overlooked and may measure the nodules more consistently and quickly. The software could also assess whether and how fast a previously identified nodule is growing. This might help doctors to decide whether additional investigation or follow-up imaging is needed or whether no further action is required. The software may also help the healthcare professional to recognise and record specific features of lung nodules that are important for decisions about what to do next and may enable them to complete these tasks more quickly. However, the software may detect more nodules that do not develop into cancer and lead to unnecessary investigations. Findings reported by the software are always checked by a healthcare professional before they are included in the reports but depending on their experience and speciality, their confidence to overrule the software might differ.

This review will look at the evidence on how well software performs in the automated detection and analysis of lung nodules, how good it is at helping radiologists or other healthcare professionals to find, measure, and assess growth of lung nodules. It will also investigate benefits and harms of using such software and whether it offers value for money. This report will cover people who had a CT scan that includes the chest for various reasons, including those who had a CT scan due to symptoms suggestive of lung cancer; for investigating other conditions unrelated to lung cancer; or to monitor previously identified lung nodules. Depending on the upcoming recommendations to be made by the

UK National Screening Committee (UK NSC), this report may additionally cover people who had a chest CT scan as part of a lung cancer screening programme.

2 Decision problem

2.1 Purpose of the decision to be made

This diagnostics assessment focuses on the use of computer software with artificial intelligence (AI)-derived algorithms for automated detection and analysis of lung nodules from computed tomography (CT) scan images that include the chest. AI is a term that broadly refers to "machines that perform tasks normally performed by human intelligence, especially when the machines learn from data how to do those tasks."[1] The technologies included in this diagnostic assessment are defined by the NICE final scope and comprise computer software that has been developed in a process that involves learning from data to detect and analyse lung nodules in CT scan images. The algorithms in the software are fixed but updated periodically.

Lung nodules are small rounded or irregular shaped growths found inside the lung. They vary in size, which is strongly associated with risk of malignancy but in a nonlinear fashion.[2] A nodule with a diameter of less than 3 mm is referred to as a micronodule, the measurement of which is not recommended due to accuracy limitations.[3] Lung nodules with a diameter smaller than 5 mm have low probability of being lung cancer[4] and do not usually require further actions if they are detected incidentally. Therefore, as a general rule, the term 'lung nodules' refers to nodules with a diameter of 5 mm or greater in this protocol unless otherwise specified.

Most lung nodules on a CT scan appear as solid structures, but some are sub-solid. Sub-solid nodules have either a solid part surrounded by a non-solid, cloud-like structure (part-solid nodules) or they appear entirely non-solid (pure ground-glass nodules). While most lung nodules are benign (noncancerous), some may be malignant (cancerous) or may develop into lung cancer.

Lung nodules are found when people are referred for a CT scan that includes the chest because of signs and symptoms suggestive of lung cancer, to investigate other conditions unrelated to lung cancer, or as part of lung cancer screening programmes. People with previously identified lung nodules can also have CT scans as part of surveillance to assess whether the growth of the nodules indicates malignancy (lung cancer) and if further assessment or treatment is needed. Lung nodules may be challenging to detect because of their small size, varying shape, and proximity to other structures in the lung.

Software capable of automatically detecting and analysing lung nodules on chest CT scan images could be used to assist radiologists or other healthcare professionals when reviewing scan images. This could increase the detection of lung nodules that need further investigation or CT surveillance but could also increase the detection of benign nodules and lead to unnecessary follow-up investigations or CT surveillance. The same software could also help in assessing the growth of previously identified nodules which are being monitored with CT surveillance. Use of the software may impact on the recognition and recording of those lung nodule characteristics that are important for decisions on appropriate follow-up. It may also affect the time it takes to review and report the CT scan images. Although the software can automatically detect and analyse lung nodules in a CT scan image, the healthcare professional reporting the scan is still expected to review the findings of the software and therefore no clinical decisions will be based on findings of the software alone. However, healthcare professionals reviewing CT scans may differ in confidence to overrule software depending on their experience and speciality (e.g. thoracic radiologists vs general radiologists).

The External Assessment Group (EAG) will assess if the use of software for automated detection and analysis of lung nodules from CT scan images by radiologists and other health professionals represent a clinically and cost-effective use of National Health Service (NHS) resources.

2.2 Target condition and diagnostic and care pathway

2.2.1 Target condition: Lung cancer

Lung cancer is one of the most common types of cancer in the UK. Its incidence rises steeply from around age 45-49.[5] Lung cancer 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 all 39,267 lung cancers in England were diagnosed at stage 3 (n=7,886) or 4 (n=18,104).[6] The NHS Long Term Plan sets out an ambitious target of diagnosing 75% of all cancers at an earlier stage, stages 1 or 2, by 2028.[7]

While most lung nodules are non-cancerous, in a small number of cases, they can be the first signs of an early cancer in the lung. In the absence of other specific and reliable signs and biomarkers, identification and monitoring lung nodules using CT scans of the chest remain the primary means of detecting lung cancer at earlier stages.

2.2.2 Diagnostic and care pathway

2.2.2.1 Pathway to CT scan due to signs and symptoms suggestive of lung cancer

The identification of people with signs and symptoms suggestive of lung cancer often happens in primary care. The NICE guideline on recognition and referral for suspected cancer[8] recommends that people aged 40 and over are offered an urgent chest X-ray (within 2 weeks of referral) if they have two or more, or if they have ever smoked and have one or more, of the following unexplained symptoms:

- cough;
- fatigue;
- shortness of breath;
- chest pain;
- weight loss or
- appetite loss.

An urgent chest X-ray should also be considered for people aged 40 or over if they have persistent or recurrent chest infection, finger clubbing, enlarged lymph nodes near the collarbone or in the neck (supraclavicular lymphadenopathy or persistent cervical lymphadenopathy), chest signs consistent with lung cancer or increased platelet count (thrombocytosis).

If the chest X-ray findings suggest lung cancer, referral to secondary care should be made using a suspected cancer pathway referral for an appointment within 2 weeks. During scoping, clinical experts noted if the X-ray findings do not show abnormalities but an ongoing suspicion of lung cancer remains, referral to secondary care for a CT scan may also be made. People aged 40 or over who present with unexplained coughing up of blood (haemoptysis) should be referred directly to secondary care without a chest X-ray using the suspected lung cancer referral pathway.

In secondary care, people with known or suspected lung cancer should be offered a contrast-enhanced chest CT scan to further the diagnosis and stage the disease (NICE guideline on diagnosis and management of lung cancer).[9]

2.2.2.2 Lung cancer screening

The UK National Screening Committee (UK NSC) does not currently recommend screening for lung cancer. This recommendation is currently under review, expected to be completed in 2022.[10] NHS England are evaluating the Targeted Lung Health Check programme (TLHC) in some areas of England.[11] In this programme, people aged over 55 years but less than 75 years who have ever smoked are invited to a lung health check. The lung health check involves collecting information

about lung health, lifestyle and family and medical history, and measuring height and weight. Following the lung health check, people assessed as high risk of lung cancer are offered a low-dose CT scan. The use of computer aided detection systems is not a requirement under this protocol, but software is being used as part of the TLHC programme.

2.2.2.3 Initial assessment and CT surveillance of lung nodules

In the NHS, the investigation of identified lung nodules follows the British Thoracic Society (BTS) guidelines for the investigation and management of pulmonary nodules and depends on the composition of the nodule (i.e. solid or sub-solid).[12] The guideline recommends the same diagnostic approach for nodules detected incidentally, due to symptomatic presentation, or through screening (the TLHC Programme also follows the BTS guideline). The guideline recommendations are for lung nodules in adults. During scoping, clinical experts explained that lung nodules in children are very rarely malignant and so lung nodules in children are not currently routinely investigated to avoid unnecessary CT scans.

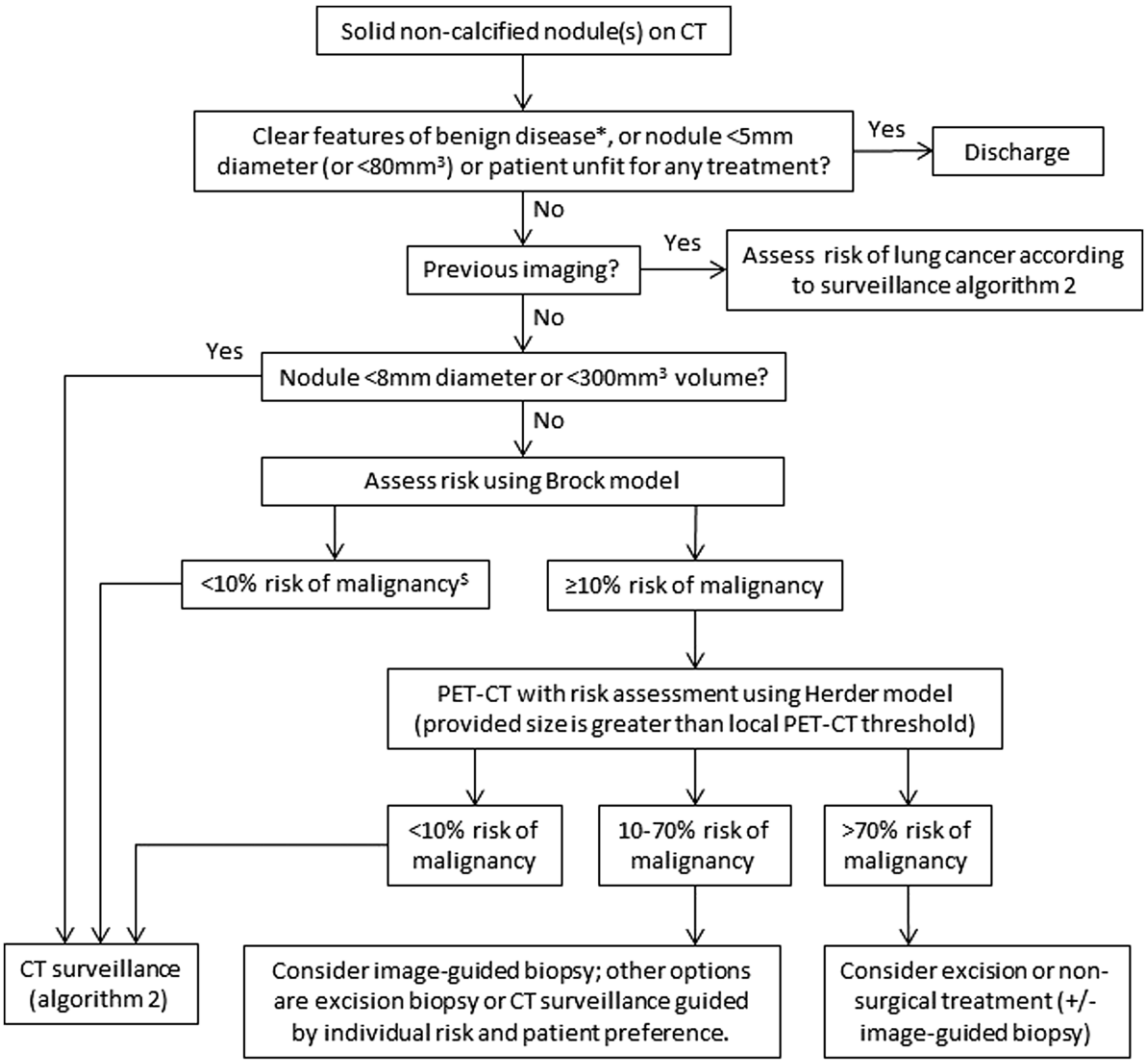

Figure 1. Initial assessment of solid lung nodules *(reproduced from Callister et al. 2015)*[12]

* Some nodules seen may be attached to or very near the lining of the lungs (perifissural nodules), these are often pulmonary lymph nodes.

Figure 1 shows the recommended pathway for the initial assessment of solid lung nodules. When there are multiple nodules, the size of the largest nodule should be considered. For newly identified nodules above a specified size, malignancy risk is estimated using the Brock model.[13] The nodule size (in diameter) and the number of nodules detected are among the inputs to this multivariable prediction model.[14]

The initial assessment of sub-solid nodules (part-solid and ground-glass) follows a similar pathway (see **Figure 2**). But because these nodules can sometimes disappear on their own, the pathway involves a repeat CT scan at 3 months before the use of the Brock malignancy risk model. Herder model[15] is not used for sub-solid nodules.

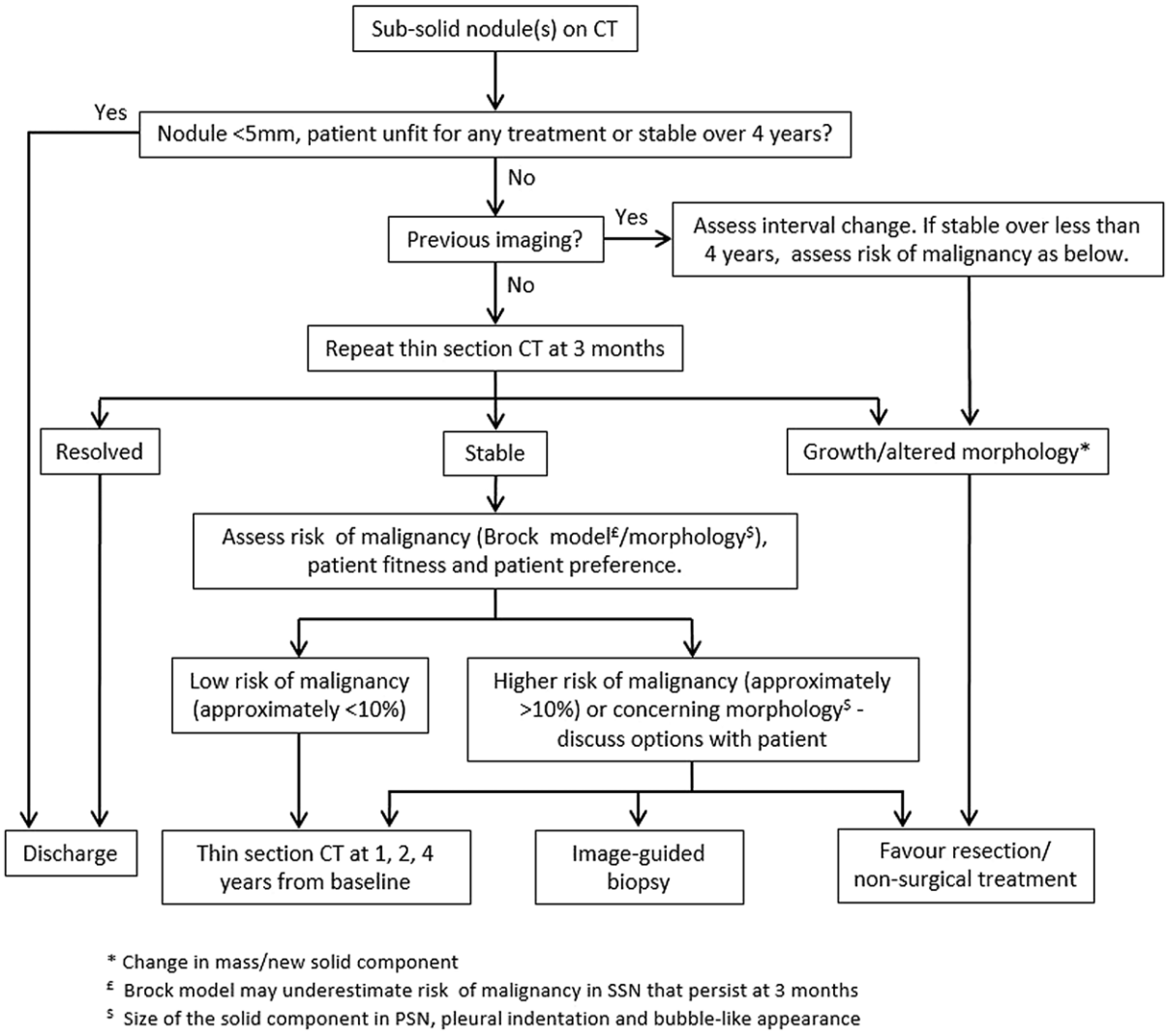

Figure 2. Sub-solid pulmonary nodules algorithm. *(reproduced from Callister et al. 2015)*[12]
PSNs, part-solid nodules; SSN, sub-solid nodules.

Figure 3 shows the recommended pathway for CT surveillance of solid lung nodules. The overall aim of this approach is to use the presence and speed of the nodule growth to discriminate between benign and malignant nodules. The nodule growth should be assessed by estimating its volume doubling time (VDT). The surveillance period for sub-solid nodules is longer (4 years) than for solid nodules (1 year with volume and 2 years with diameter measurements).

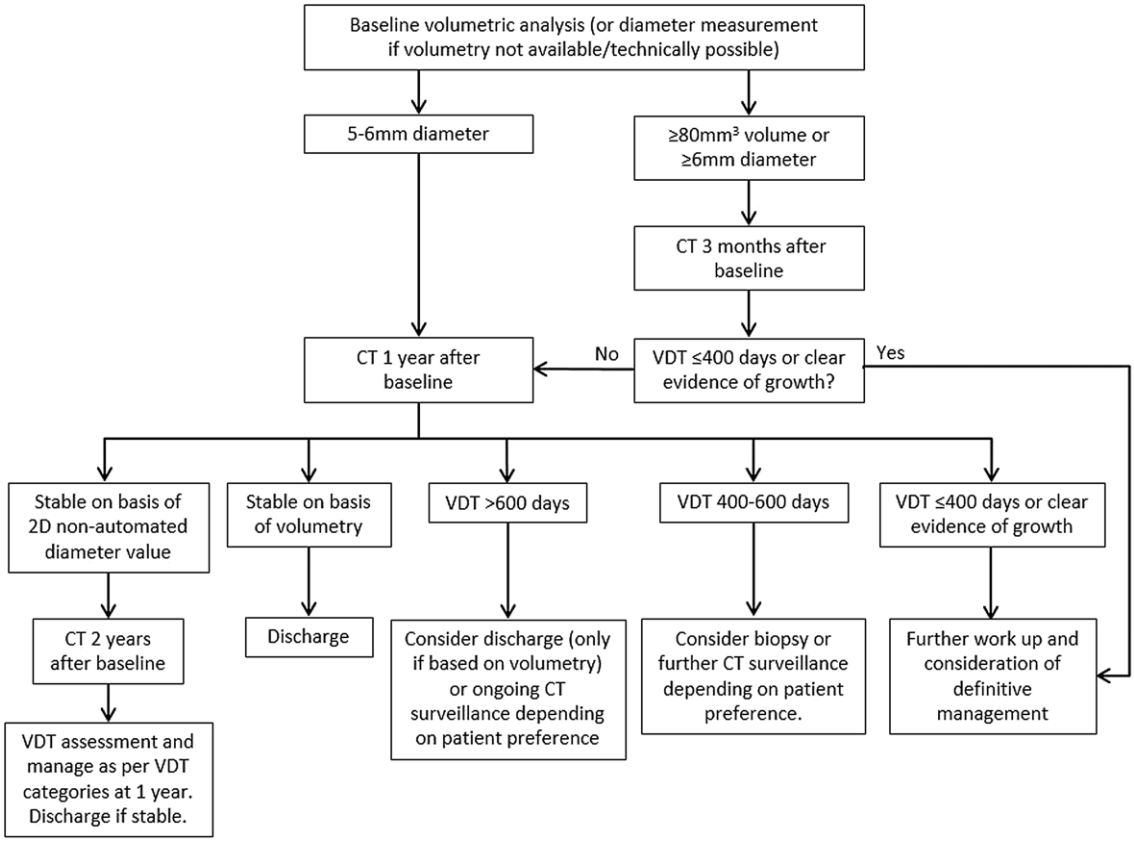

Figure 3. CT surveillance of solid lung nodules *(reproduced from Callister et al. 2015)*[12]

The BTS guidelines are currently being updated.[16]

2.2.2.4 Current methods of detecting nodules and measuring nodule volume and growth on CT scans

Currently, in routine clinical practice in the UK, radiologists or other healthcare professionals such as radiographers detect lung nodules on chest CT scan images without assistance from any software. The healthcare professional reviewing the scan may be a specialist in reviewing chest CT images (such as a thoracic radiologist) or less specialised (such as a general radiologist in an Accident & Emergency [A&E] department).

In the TLHC programme, the healthcare professionals reviewing the scans are radiologists specialised in reviewing chest CT images. They are either radiologists who regularly lead at their local lung cancer multidisciplinary team or radiologists who yearly, as part of their normal clinical practice, report more than 500 thoracic CT scans of which a significant proportion are lung cancer CT scans.[17] Software for the automated detection of lung nodules has been used in the TLHC programme. The British Society of Thoracic Imaging and the Royal College of Radiologists have published a summary of radiology-related considerations for the TLHC, including advice on software.[18]

The 2015 BTS guidelines for the investigation and management of pulmonary nodules[12] recommend that the size of an identified nodule is quantified as the volume of the nodule. To do this, a volumetry software needs to be used. In current practice, this is often a software that is part of the picture archiving and communication system (PACS), or it may be a module on a software that comes with the CT scanner. When measuring the size of the part-solid nodules, the diameter of the solid part of the nodule is considered. In ground-glass nodules, the diameter of the entire nodule is measured.

This volumetry software may or may not have the capability of comparing sequential scans to automatically measure the volume doubling time (VDT). When this feature is not available or not used, the VDT can be calculated by inputting the nodule volume measurements and dates of the 2 scans into the BTS Pulmonary Nodule Risk Calculator.[14] In addition to growth, for ground-glass nodules any later appearance of a solid part is assessed.

Where volumetry software is not available or measuring the nodule volume by the software is not possible because of the quality of the image or the location of the nodule within the lung, the largest diameter of the nodule is measured. The VDT can then be estimated by inputting the diameter measurements and dates of the 2 scans using the BTS Pulmonary Nodule Risk Calculator.[14] During scoping, clinical experts reported that diameter measurements are still widely used in the NHS.

2.2.2.5 Diagnosis and staging of lung cancer

To guide the treatment of lung cancer, information about type and spread of the lung cancer (stage) are needed. The NICE guideline on diagnosis and management of lung cancer [9] recommends choosing investigations that give the most information about diagnosis and staging with the least risk to the person. The type and sequence of investigations may vary, but the investigations commonly include a contrast enhanced CT of the chest, abdomen and pelvis, a positron emission tomography-CT (PET-CT) scan and an image-guided biopsy. Other methods that may be used include magnetic resonance imaging (MRI), endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) and endoscopic ultrasound-guided fine-needle aspiration (EUS-FNA).

2.2.2.6 Treatment for lung cancer

After diagnosis, treatment for lung cancer is based on several factors, such as overall health of the patient and the type, size, position, and stage of the cancer. The treatment may include surgery, chemotherapy, radiotherapy, immunotherapy or other targeted therapy drugs or a combination of these (NICE guideline on diagnosis and management of lung cancer[9]).

The NICE pathway for treatment of lung cancer[19] provides further details on the treatment of both non-small-cell and small-cell lung cancer.

2.3 Population: People who have a CT scan that includes the chest

This diagnostic assessment will include people who have any type of CT scan (e.g. with or without contrast, low-dose or standard dose; excluding PET-CT) that includes part or all of the chest for the following reasons:

1. 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:

- because of signs or symptoms suggestive of lung cancer (symptomatic population);
- for reasons unrelated to suspicion of lung cancer (incidental population);
- who attend lung cancer screening (inclusion of the screening population will be dependent on the upcoming lung cancer screening recommendations of the UK NSC in 2022).

2. People having CT surveillance for a previously identified lung nodule (surveillance population).

Use of the technologies for cancer staging and cancer follow-up (including detection of metastasis to the lung) in people with extrathoracic primary cancers is outside the scope of this assessment.

2.4 Interventions

The software identified for this assessment uses algorithms that have been produced using AI. The algorithms are fixed but updated periodically. Software is included in this diagnostic assessment if it has automated nodule detection and volume measurement capability.

Some of the software can also compare subsequent scans to automatically measure VDT. In some of the software, parameters can be changed to adjust the nodule detection performance (thus varying the sensitivity and specificity for nodule detection). Some include an integrated Brock model calculator.

Some of the software may only be able to analyse images of CT scans that include the entire lung. Some may be indicated for use only with a specific type of CT scan (for example scans without contrast or low-dose CT) or in specified populations (for example people without symptoms suggestive of lung cancer or people aged 18 or older).

Thirteen relevant technologies have been identified by the NICE. The section below describes the specific technologies included in this assessment. The descriptions as well as **Error! Reference source not found.** are reproduced from the final scope issued by NICE.

2.4.1 AI-Rad Companion Chest CT (Siemens Healthineers)

AI-Rad Companion Chest CT is a CE-marked (class IIa medical device) software. It includes Lung-CAD, a tool that can detect and measure solid lung nodules in CT scans that cover the entire lung, with and without contrast. The algorithms are optimised for nodules between 3 mm and 30 mm. Lung-CAD is compatible with slice thickness of up to 2.5 mm. It is indicated for use in both screening and diagnostic protocols in people without diffuse interstitial or airway diseases, severe pneumonia, extensive granulomatous diseases, prior thoracotomy or history of radiation therapy involving the lung parenchyma who are aged 22 and over. The software integrates with the PACS.

2.4.2 AVIEW LCS+ (Coreline Soft)

AVIEW LCS+ is a CE-marked (class IIa medical device) software. It can detect, measure and assess the growth of solid and sub-solid nodules in low-dose chest CT scans. AVIEW LCS+ is indicated for use in adults. Other indications for use include detection of emphysema (damage to the air sacs in the lung) and coronary artery calcification. The software integrates with PACS. The software is commercially available to the NHS.

2.4.3 ClearRead CT (Riverain Technologies)

ClearRead CT is a CE-marked (class IIa medical device) software. It consists of ClearRead CT Vessel Suppress, ClearRead CT Detect and ClearRead CT Compare features. Using these features, the software can detect, measure and assess the growth of solid and sub-solid lung nodules in low-dose and regular dose CT scans where both lungs are visible, with and without contrast. The software is compatible with slice thickness of up to 5 mm. ClearRead CT is indicated for use in people aged 18 and over who are asymptomatic. The software is updated frequently but none of the functionality is expected to be removed in future updates. The software integrates with, and the findings of the software are visible within, PACS. The company expects that training of radiologists on how to use ClearRead CT is usually done within a day. The software is commercially available to the NHS directly from the manufacturer and through partner organisations.

2.4.4 contextflow SEARCH Lung CT (contextflow)

Contextflow SEARCH Lung CT is a CE-marked (class IIa medical device) software. It can detect and measure solid and sub-solid lung nodules in chest CT scans with and without contrast. It is intended for use in clinically stable, symptomatic patients. Other indications for use include identification of lung-specific image patterns related to diseases such as airway wall thickening, bronchiectasis, emphysema and pneumothorax. contextflow SEARCH Lung CT integrates with PACS. The company expects users to attend a training presentation before using the software. The software is commercially available to the NHS.

2.4.5 InferRead CT Lung (Infervision)

InferRead CT Lung is a CE-marked (class IIa medical device) software. It can detect, measure and assess the growth of solid and sub-solid lung nodules in low-dose or regular dose CT scans with and without contrast. The company advises that InferRead CT Lung is intended for use in asymptomatic populations. The company also states that the use is recommended in people aged 18 and over. Users can dismiss nodules found by the automated analysis but editing the findings is not possible. Users can add nodules, but the software will not measure the volume of user-added nodules. A new version of InferRead CT Lung is expected to be released within 18 months. The current version will continue to be supported and available to the NHS. InferRead CT Lung includes rules for reporting that follow the BTS guidelines for the investigation and management of pulmonary nodules.[12] The software integrates with, and the findings of the software are visible within, PACS. The company expects radiologists to complete a 1-hour training session before using the technology. The software is commercially available to the NHS.

2.4.6 JLD-01K (JLK Inc.)

JLD-01K is a CE-marked (class I medical device) software. It can detect and measure solid and sub-solid lung nodules in chest CT scans without contrast. The software was trained in CT scans where nodules were at least 3 mm in diameter. JLK-01K integrates with PACS.

2.4.7 Lung AI (Arterys)

Lung AI is a CE-marked (class IIa medical device) software. It can detect, measure and assess the growth of solid and sub-solid lung nodules in chest CT scans. The nodule detection and segmentation algorithms are optimised for low-dose chest CT scans, but the software will analyse any chest CT scan including regular dose CT scans with contrast without generating an error. Users can add, edit, or dismiss detected nodules with automatic updates to quantitative nodule information. Lung AI integrates with PACS.

2.4.8 Lung Nodule AI (Fujifilm)

Lung Nodule AI is a software that can detect, measure and assess the growth of lung nodules in chest CT scans. The software is currently approved for use in Japan. The company plans to introduce the technology in Europe once required regulatory clearances are obtained.

2.4.9 qCT-Lung (Qure.ai)

qCT-Lung is a CE-marked (class I medical device) software. It can detect lung nodules at least 3 mm in diameter in chest CT scans without contrast. The software can also measure the volume and assess the growth of lung nodules, but these features are currently available for research purposes only. Other indications for use include detection of emphysema. qCT-Lung is intended for use in people aged 18 and over. The software is compatible with slice thickness of up to 6 mm. qCT-Lung integrates with PACS.

2.4.10 SenseCare-Lung Pro (SenseTime)

SenseCare-Lung Pro is a CE-marked (class IIb medical device) software. It can detect, measure and assess the growth of solid and sub-solid lung nodules in chest CT scans without contrast. Other indications for use include detection of pneumonia (including COVID-19) lesions. The software is compatible with slice thickness of up to 5 mm, but the preferred slice thickness is up to 1.5 mm. SenseCare-Lung Pro integrates with PACS.

2.4.11 Veolity (MeVis)

Veolity is a CE-marked (class IIa medical device) software. It can detect, measure and assess the growth of lung nodules in low-dose and regular dose CT scans that include the complete chest, with and without contrast. The software is compatible with slice thickness of up to 3 mm. Veolity is indicated for use in asymptomatic populations. Users can interact with the software by adding and dismissing nodules in the analysis and editing the findings of the software. With input from the user, the software also calculates the malignancy risk of the nodules using the Brock model. Veolity's current detection algorithm only detects solid nodules. A new version of the software (Veolity 2.0) is planned for the beginning of 2022. This version will detect solid and sub-solid nodules. Usually, 2 updates or functional upgrades per year are planned. Existing versions will continue to be supported. Veolity includes rules for reporting following the BTS guidelines for the investigation and management of pulmonary nodules[12] and integrates with the PACS. The company states that usually 4 to 6 hours of training are needed for radiologists to learn how to use Veolity. The software is commercially available to the NHS, distributed in the UK by SynApps Solutions.

2.4.12 Veye Lung Nodules (Aidence)

Veye Lung Nodules is a CE-marked (class IIb medical device) software. It can detect, measure and assess the growth of solid and sub-solid lung nodules in low-dose or standard dose CT scans where both lungs are visible, with and without contrast. The software is compatible with slice thickness of up to 3 mm. Veye Lung Nodules is intended for use in people aged 18 and over. Users can dismiss nodules found by the automated analysis but editing the findings is not possible. Users can add nodules, but the software will not measure the volume of user-added nodules. The software is updated frequently. Veye Lung Nodules includes rules for reporting following the BTS guidelines for the investigation and management of pulmonary nodules.[12] The software integrates with, and findings of the software are visible within, PACS. The company expects radiologists to attend a 1-hour training session before using the technology. The software is commercially available to the NHS.

2.4.13 VUNO Med-LungCT AI (VUNO)

VUNO Med-LungCT AI is a CE-marked (class IIa medical device) software. It can detect, measure and assess the growth of solid and sub-solid lung nodules in low-dose chest CT scans. It is intended for use in lung cancer screening populations. The software integrates with PACS.

Table 1. Summary of the included technologies (reproduced from final NICE scope)

Product name (manufacturer)	CE mark	Available to the NHS	CT scan types	Detection	Volumetry
AI-Rad Companion Chest CT (Siemens)	Class IIa *	To be confirmed	Low dose, regular dose with and without contrast *	Yes *	Yes *
AVIEW LCS+ (Coreline Soft)	Class IIa *	Yes	Low dose *	Yes	Yes
ClearRead CT (Riverain Technologies)	Class IIa	Yes	Low dose, regular dose with and without contrast	Yes	Yes
contextflow SEARCH Lung CT (contextflow)	Class IIa	Yes	With and without contrast	Yes	Yes
InferRead CT Lung (Infervision)	Class IIa	Yes	Low dose, regular dose with and without contrast	Yes	Yes
JLD-01K (JLK Inc.)	Class I	To be confirmed	Without contrast	Yes	Yes
Lung AI (Arterys)	Class IIa *	To be confirmed	Low dose, regular dose with and without contrast *	Yes *	Yes *
Lung Nodule AI (Fujifilm)	To be confirmed	To be confirmed	To be confirmed	Yes	Yes
qCT-Lung (Qure.ai)	Class I *	To be confirmed	Without contrast *	Yes *	Research only *
SenseCare-Lung Pro (SenseTime)	Class IIb *	To be confirmed	Without contrast *	Yes *	Yes *
Veolity (MeVis)	Class IIa	Yes	Low dose, regular dose with and without contrast	Yes	Yes
Veye Lung Nodules (Aidence)	Class IIb	Yes	Low dose, regular dose with and without contrast	Yes	Yes
VUNO Med-LungCT AI (VUNO)	Class IIa *	To be confirmed	Low dose *	Yes *	Yes *

* Information only from public domain.

2.5 Comparator

The comparator for this diagnostic assessment is review of chest CT scan images by a radiologist or another healthcare professional (such as a radiographer) without software for automated detection and analysis of lung nodules. The reviewer of the scan may use software to help measure the volume of an identified lung nodule (see section 2.2.2.4), but this software does not automatically detect or measure lung nodules. When volumetric software is not used, nodule diameter is used to define the nodule size and nodule growth. The healthcare professional reviewing the scan may or may not be specialised in reviewing chest CT images.

During scoping, clinical experts highlighted that the experience of radiologists in reviewing CT scans for lung nodules will vary, for example from general, trauma or thoracic radiologists. They further

commented that the level of expertise of the healthcare professional reviewing the scan may change the impact of the software. For example, less experienced reviewers may be more likely to act on nodules detected by the software, even if they disagree. For this reason, as highlighted in section 2.2.2.4, the standard protocol for the TLHC programme in England stipulates specific requirements for specialised readers reviewing the CT scans in the programme.[17]

3 Decision questions and objectives

The overall objectives of this diagnostic assessment are to assess the clinical and cost-effectiveness of CT image analysis assisted by software capable of automated detection and analysis of lung nodules compared with unassisted CT image analysis in people undergoing CT scans that include the chest due to symptoms suggestive of lung cancer, for purposes unrelated to suspicion of lung cancer, for surveillance of previously identified lung nodules or (depending on upcoming UK NSC recommendations) for lung cancer screening.

The key questions for this diagnostic assessment report (DAR) are provided in the box below.

Key question 1

What is the accuracy of CT image analysis assisted by software for automated detection and analysis of lung nodules in people undergoing CT scans that include the chest due to symptoms suggestive of lung cancer, for purposes unrelated to suspicion of lung cancer, for surveillance of previously identified nodules or (depending on upcoming UK NSC recommendations) for lung cancer screening, and what are the practical implications (e.g. test failure rate, reading time, acceptability) and the impact on patient management (e.g. stage of cancer detected, time to diagnosis, number of people referred to CT surveillance or having biopsy/excision)?

Sub-questions

1. Does the accuracy of CT image analysis assisted by software for automated detection and analysis of lung nodules, its practical implications and impact on patient management differ between CT scans: (1) with contrast and without contrast; (2) using a low-dose and a standard dose; (3) of solid nodules and sub-solid nodules?
2. Does the accuracy of CT image analysis assisted by software for automated detection and analysis of lung nodules, its practical implications and impact on patient management differ by patients' ethnicity?
3. Does the accuracy of CT image analysis assisted by software for automated detection and analysis of lung nodules, its practical implications and impact on patient management differ between general radiologists/health professionals and specialised thoracic radiologists/health professionals?

4. For the incidental population, does the accuracy of CT image analysis assisted by software for automated detection and analysis of lung nodules, its practical implications and impact on patient management differ by reason for CT scan?
5. a) What is the concordance between readers with and without software support to detect and/or measure lung nodules from CT images?

   b) What is the concordance between readers using different software to detect and/or measure lung nodules from CT images?

   c) Does the use of software-assisted CT image analysis impact on intra-observer and inter-observer variability in lung nodule detection and measurement?

Key question 2

What are the benefits and harms of using software for automated detection and analysis of lung nodules from CT images compared with unassisted CT image analysis in people undergoing CT scans that include the chest due to symptoms suggestive of lung cancer, for purposes unrelated to suspicion of lung cancer, for surveillance of previously identified nodules or (depending on upcoming UK NSC recommendations) for lung cancer screening?

Sub-questions

1. Do the benefits and harms of CT image analysis assisted by software for automated detection and analysis of lung nodules differ between CT scans: (1) with contrast and without contrast; (2) using a low-dose and a standard dose; (3) of solid nodules and sub-solid nodules?
2. Do the benefits and harms of CT image analysis assisted by software for automated detection and analysis of lung nodules differ by patients' ethnicity?
3. Do the benefits and harms of CT image analysis assisted by software for automated detection and analysis of lung nodules differ between general radiologists/healthcare professionals and specialised thoracic radiologists/healthcare professionals?
4. For the incidental population, do the benefits and harms of CT image analysis assisted by software for automated detection and analysis of lung nodules differ by reason for chest CT scan?

Key question 3

What is the cost-effectiveness of using software for automated detection and analysis of lung nodules from CT images compared with unassisted CT image analysis in people undergoing CT scans that include the chest due to symptoms suggestive of lung cancer, for purposes unrelated to

suspicion of lung cancer, for surveillance of previously identified nodules or (depending on upcoming UK NSC recommendations) for lung cancer screening?

Sub-questions

1. Does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ between CT scans: (1) with contrast and without contrast; (2) using a low-dose and a standard dose; (3) of solid nodules and sub-solid nodules?
2. Does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ by patients' ethnicity?
3. Does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ between general radiologists/healthcare professionals and specialised thoracic radiologists/healthcare professionals?
4. For the incidental population, does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ by reason for CT scan?

4 Methods for assessing test accuracy and practical implications

Key question 1

What is the accuracy of CT image analysis assisted by software for automated detection and analysis of lung nodules in people undergoing CT scans that include the chest due to symptoms suggestive of lung cancer, for purposes unrelated to suspicion of lung cancer, for surveillance of previously identified nodules or (depending on upcoming UK NSC recommendations) for lung cancer screening, and what are the practical implications (e.g. test failure rate, reading time, acceptability) and the impact on patient management (e.g. stage of cancer detected, time to diagnosis, number of people referred to CT surveillance or having biopsy/excision)?

Evidence required to address the above decision questions will be identified and assessed in a systematic review using methods described below. The review will follow the principles outlined in the Cochrane Handbook of Diagnostic Test Accuracy[20] and the NICE Diagnostic Assessment Programme manual.[21]

Ideally, priority of the assessment will be given to 'end-to-end studies' that follow patients from testing, through treatment, to final health outcomes such as morbidity and mortality. These studies can remove the need for separate searches for model parameters for cost-effectiveness modelling.[21] Such studies could include randomised controlled trials (RCTs), prospective, comparative cohort studies

and historic cohort studies. However, it is likely that no end-to end studies that directly address the questions for this diagnostic assessment are available. If no end-to-end studies are found, we will include and evaluate studies on test accuracy and practical implications, clinical effectiveness, costs and cost-effectiveness separately, and then synthesise the evidence using a linked evidence approach.[21]

4.1 Identification and selection of studies

4.1.1 Search strategy

A comprehensive search will be developed iteratively and undertaken in a range of relevant bibliographic databases. Searches will combine keywords and, where appropriate, thesaurus (MeSH/EMTREE) terms relating to 'AI', 'lung nodules/lung cancer' and 'CT or screening'. Searches will be limited to studies published in English as studies published in other languages are likely to be difficult to assess. A publication date limit may be applied, if a date is identified before which it can be safely assumed there were no studies on the technologies of interest. A draft MEDLINE search strategy is provided in **Appendix 1**. This will be checked by an Information Specialist not otherwise involved in the project for any omissions or errors in spelling, search syntax, structure and use of MeSH headings, before being adapted for the other databases.

Systematic searches will be conducted in the following databases:
MEDLINE All (via Ovid);
Embase (Ovid);
Cochrane Database of Systematic Reviews (Wiley);
Cochrane CENTRAL (Wiley);
Health Technology Assessment (HTA) database (CRD);
International HTA database (INAHTA);
Science Citation Index Expanded (Web of Science)
Conference Proceedings - Science (Web of Science).

Records will be exported to EndNote X9.3, where duplicates will be systematically identified and removed.

In order to capture unpublished or ongoing studies, searches of MedRxiv preprint server (via the medrxivr app) and clinical trials registries (via clinicaltrials.gov and the WHO ICTRP portal) will be undertaken. The trials registry searches will be highly focussed, including search terms for the specific technologies of interest listed in the project scope, and their manufacturing companies. Websites of the technologies and their manufacturers will also be checked for further information, as

will websites of selected organisations and conferences of interest (to include for example NICE, CADTH, FDA, ISPOR, HTAi, European Society of Radiology Congresses, Radiological Society of North America Annual Meetings, SPIE Proceedings and Conference of the IEEE Engineering in Medicine & Biology Society). Reference lists of included studies and a selection of recent, relevant systematic reviews identified via the database searches will be checked. Forwards citation tracking from key publications of included studies (to identify citing papers) will also be undertaken, using Science Citation Index (Web of Science) and Google Scholar.

4.1.2 Study eligibility criteria

Studies that satisfy the following criteria will be included:

Population	<u>All questions</u> People who have no confirmed lung nodules or 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: • for reasons unrelated to suspicion of lung cancer (incidental population); • because of signs or symptoms suggestive of lung cancer (symptomatic population); • as part of lung cancer screening (inclusion of the screening population will be confirmed following the upcoming lung cancer screening recommendations of the UK NSC). People having CT surveillance for a previously identified lung nodule. In case the screening population is not included in the target population, we may still include evidence from adults who are undergoing lung cancer screening as an approximation to the target population if there are insufficient data available from the target population. Where data permits, the following subgroups may be considered: • Patient's ethnicity; • People who have a CT scan: (1) with or without contrast; (2) using a low-dose or a standard dose; (3) of solid nodules or sub-solid nodules; • For the incidental population, by reason for CT scan.

Target condition	<u>All questions</u> Lung cancer
Intervention	<u>All questions</u> CT scan review by a radiologist or another healthcare professional using any of the following 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) * Indication for use specifies use in asymptomatic population, therefore the software cannot be assessed in symptomatic population. ** Indication for use specifies use in symptomatic population, therefore the software cannot be assessed in incidental or screening populations. Please note: specific indications for use for some of the technologies are unclear because only information in the public domain was available. Evidence on the performance of software alone (without review by a radiologist or other trained reader) will be included with applicability concerns highlighted. Where data permits, the following subgroups may be considered: - General radiologist/other healthcare professional with software support versus radiologist/other healthcare professional with thoracic speciality with software support.
Comparator	<u>All questions</u> CT scan review by a radiologist or another healthcare professional without software for automated detection and analysis of lung nodules (using diameter or volume to measure nodule size) or no comparator. Where data permits, the following subgroups may be considered:

	- General radiologist/other healthcare professional without software support versus radiologist/other healthcare professional with thoracic speciality without software support.
Reference standard	Review question 1 and sub-questions 1-4 • Lung cancer confirmed by histological analysis of lung biopsy or health record review; • CT surveillance (imaging follow-up) without significant growth, follow-up without lung cancer; • Experienced radiologist reading (single reader or consensus of more than one reader).
Outcomes	Review question 1 and sub-questions 1-4. • Accuracy to detect nodules (by nodule size and/or by nodule type; this may include for example the accuracy to detect nodules considered potentially significant by judgement of experienced radiologist(s) and the accuracy to detect malignant nodules, respectively); • Accuracy to assess volume of nodule or change in volume (when interventions are used as part of CT surveillance); • Characteristics of detected nodules (e.g. size, type, location, spiculation); • Proportion of detected nodules that are malignant; • Technical failure rate; • Radiologist reading time; • Radiology report turnaround time; • Impact of test result on clinical decision-making; • Number of people having CT surveillance (this may include also for example the number of people with false positive nodules having unnecessary CT surveillance); • Number of CT scans taken as part of CT surveillance (this may include also for example number of unnecessary CT surveillance scans due to false positive nodules); • Number of people having a biopsy or excision (this may include also for example the number of people having a negative biopsy due to false positive nodules); • Number of cancers detected; • Stage of cancer at detection; • Time to diagnosis;

	• Acceptability and experience of using the software. Sub-question 5. • Concordance between readers with and without software; • Concordance between readers using different software; • Concordance between different software without human involvement; • Inter-observer variability (e.g. positive and negative agreement, Cohen's kappa); • Repeatability/reproducibility.
Study design	All questions • Prospective test accuracy studies; • Retrospective test accuracy studies; • Randomised controlled trials; • Cohort studies; • Historically controlled trials; • Before-after studies; • Retrospective multiple reader multiple case studies; • Qualitative studies for user experience/acceptability.
Publication type	All questions • Peer-reviewed papers. • Conference abstracts and manufacturer data will be included. Only outcome data that have not been reported in peer-reviewed full text papers will be extracted and reported.
Language	All questions English

Papers that fulfil the following criteria will be excluded:

- Studies using PET-CT scan images, lung phantom images or where more than 10% of CT scans are performed in patients with a primary cancer outside the lung (staging).
- Studies using index tests other than those specified in the inclusion criteria.
- Studies with no relevant outcomes reported.
- Non-human studies.
- Letters, editorials and communications will be excluded unless they report outcome data that have not been reported elsewhere, in which case they will be handled in the same way as conference abstracts.
- Articles not available in the English language.

4.1.3 Review strategy

Two reviewers (JG/AA/SJ) will independently screen the titles and abstracts of records identified by the searches. Any disagreements will be resolved through discussion, or retrieval of the full publication. Potentially relevant publications will be obtained and assessed independently by 2 reviewers (JG/AA/SJ). Disagreements will be resolved through consensus, with the inclusion of a third reviewer (CS, YFC) if required. Records that are excluded at full text stage will be documented, including the reasons for their exclusion.

4.2 Extraction and study quality

4.2.1 Data extraction strategy

Data will be extracted by one reviewer (JG/AA/SJ) and checked by a second reviewer (JG/AA/SJ). All data extraction will be entered into a piloted electronic data collection form. Any disagreements will be resolved through consensus, with the inclusion of a third reviewer (CS, YFC) if required.

4.2.2 Assessment of study risk of bias

The risk of bias of test accuracy studies will be assessed using a modified QUADAS-2 tool.[22] As recommended by the QUADAS-2 group, an overall quality score will not be determined.[22] The risk of bias of randomised controlled trials will be assessed using the revised (version 2) 'Cochrane risk-of-bias tool for randomized trials' (RoB 2).[23] Risk of bias in non-randomised controlled trials, before-after studies, historically controlled trials and cohort studies will be assessed using the Cochrane risk of bias in non-randomized studies of interventions (ROBINS-I) tool.[24] Qualitative studies will be critically appraised using the Critical Appraisal Skills Programme (CASP) Qualitative Studies Checklist.[25] We will use the NICE preferred appraisal tools for any other study design.[26] Two reviewers (JG/AA/SJ) will independently undertake risks of bias assessment and critical appraisal. Disagreements will be resolved through consensus, with the inclusion of a third reviewer (CS, YFC) if required. The results of each risk of bias item will be presented in table and/or graph form.

4.3 Methods of analysis/synthesis

Study design, population, software use and outcome measures will be summarised in text, tables and/or presented graphically. Test accuracy results will be presented for different testing strategies (e.g. stand-alone software, software-assisted readers) comparing the index tests to the eligible reference standards. Test accuracy results will then be reported stratified by technology for the detection of all lung nodules and nodules of specific characteristics (e.g. actionable nodules ≥5 mm), respectively.

For each software functionality, e.g. nodule detection and classification of nodule morphology (solid vs subsolid nodule), accuracy results will be treated as binary (e.g. nodule present/absent; solid/sub-solid nodule). Original data extracted from the studies will be used to construct 2x2 tables. The resulting pairs of sensitivities and specificities will be plotted on a receiver operator characteristic (ROC) curve to display covariation between sensitivity and specificity. Pairs of sensitivities and specificities will also be displayed in a paired forest plot to demonstrate scatter and uncertainty. Studies will be grouped by software and its role in the workflow (e.g. stand-alone software, software-assisted reader).

Meta-analysis will be considered if sufficient data from reasonably homogeneous studies are available. This will be guided by characteristics of study design, population, nature of comparisons made, outcome measures and findings from risk of bias assessment. If pooling is feasible, pairs of sensitivity and specificity will be meta-analysed using the hierarchical summary ROC model (HSROC) as recommended for Cochrane reviews when heterogeneity in diagnostic thresholds between studies is anticipated. This approach accounts for variability between studies as well as the relationship between sensitivity and specificity. If sufficient data are available, heterogeneity will be investigated by adding covariates to the model, e.g. study design, software, patients' ethnicity, CT scan type, nodule type, reason for CT scan and reader speciality.

Where data available, we will additionally present subgroup data and may undertake subgroup analyses by:

- Patients' ethnicity;
- Reason for CT scan;
- CT scans with vs without contrast;
- CT scans using different radiation doses (e.g. ultra-low-dose, low-dose, standard dose);
- Solid nodules vs sub-solid nodules;
- General radiologist (or other healthcare professional) vs specialised thoracic radiologist (or other healthcare professional).

Where data permit, sensitivity analyses may be performed to explore the impact of potential bias identified in QUADAS assessment, for example by excluding studies using a single reader as the reference standard.

Qualitative evidence will be analysed and summarised thematically.

5 Methods for assessing clinical effectiveness

Key question 2

What are the benefits and harms of using software for automatic detection and analysis of lung nodules from CT images compared with unassisted CT image analysis in people undergoing CT scans that include the chest due to symptoms suggestive of lung cancer, for purposes unrelated to suspicion of lung cancer, for surveillance of previously identified nodules or (depending on upcoming UK NSC recommendations) for lung cancer screening?

The same review searches and methods that will be used for the test accuracy question (see section 4) will be employed to address this question. We will summarise intermediate outcomes (as identified in section 4) that could predict a change in health outcomes (e.g. potential benefits by earlier nodule detection and shorter time to diagnosis; potential harms of increased false positive test results) as well as outcomes from end-to-end studies. The potential impact of these intermediate outcomes on final health outcomes will be modelled using a linked evidence approach with additional types of evidence collected as described in section 6.

5.1 Identification and selection of studies

5.1.1 Search strategy

The same search strategy as described in the methods for test accuracy will be used (see section 4.1 Identification and selection of studies).

5.1.2 Study eligibility criteria

Studies that satisfy the following criteria will be included:

Population	<u>All questions</u> People who have no confirmed lung nodules or 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: • for reasons unrelated to suspicion of lung cancer (incidental population); • because of signs or symptoms suggestive of lung cancer (symptomatic population); • as part of lung cancer screening (inclusion of the screening population will be confirmed following the upcoming lung cancer screening recommendations of the UK NSC).

	People having CT surveillance for a previously identified lung nodule. In case the screening population is not included in this diagnostic assessment following the publication of UK NSC recommendation, we may still include evidence from adults who are undergoing lung cancer screening as an approximation to the target population if there are insufficient data available from the target population. Where data permits, the following subgroups may be considered: • Patients' ethnicity; • People who have a CT scan: (1) with or without contrast; (2) using a low-dose or a standard dose; (3) of solid nodules or sub-solid nodules; • For the incidental population, by reason for CT scan .
Target condition	Lung cancer
Intervention	<u>All questions</u> CT scan review by a radiologist or another healthcare professional using any of the following 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) * Indication for use specifies use in asymptomatic population, therefore the software cannot be assessed in symptomatic population. ** Indication for use specifies use in symptomatic population, therefore the software cannot be assessed in incidental or screening populations.

	Please note: specific indications for use for some of the technologies are unclear because only information in the public domain was available. Where data permits, the following subgroups may be considered: - General radiologist/other healthcare professional with software support versus radiologist/other healthcare professional with thoracic speciality with software support.
Comparator	CT scan review by a radiologist or another healthcare professional without software for automated detection and analysis of lung nodules (using diameter or volume to measure nodule size). Where data permits, the following subgroups may be considered: - General radiologist/other healthcare professional without software support; radiologist/other healthcare professional with thoracic speciality without software support.
Outcomes	• Morbidity (including any adverse events caused by assessment or treatment); • Mortality; • Health-related quality of life; • Patients' acceptability of use of the software.
Study design	• Randomised controlled trials; • Quasi-randomised trials; • Cohort studies (retrospective/prospective); • Before-after studies; • Historical controlled studies. • Qualitative studies (for patient acceptability of the use of the software)
Publication type	• Peer reviewed papers. • Conference abstracts and manufacturer data will be included. Only outcome data that have not been reported in peer-reviewed full text papers will be extracted and reported.
Language	English

Papers that fulfil the following criteria will be excluded:

- Studies using PET-CT scan images, lung phantom images or where more than 10% of CT scans are performed in patients with a primary cancer outside the lung (staging).
- Studies using index tests other than those specified in the inclusion criteria.

- Studies without relevant outcomes.
- Non-human studies.
- Letters, editorials and communications will be excluded unless they report outcome data that have not been reported elsewhere, in which case they will be handled in the same way as conference abstracts.
- Articles not available in the English language.

5.1.3 Review strategy

Two reviewers (JG/AA/SJ) will independently screen the titles and abstracts of records identified by the searches. Any disagreements will be resolved through discussion, or retrieval of the full publication. Potentially relevant publications will be obtained and assessed independently by 2 reviewers (JG/AA/SJ). Disagreements will be resolved through consensus, with the inclusion of a third reviewer (CS, YFC) if required. Records that are excluded at full text stage will be documented, including the reasons for their exclusion.

5.2 Extraction and study quality

5.2.1 Data extraction strategy

Data will be extracted by one reviewer (JG/AA/SJ) and checked by a second reviewer (JG/AA/SJ). All data extraction will be entered into a piloted electronic data collection form. Any disagreements will be resolved through consensus, with the inclusion of a third reviewer (CS, YFC) if required.

5.2.2 Assessment of study risk of bias

The risk of bias of randomised controlled trials will be assessed using the revised (version 2) 'Cochrane risk-of-bias tool for randomized trials' (RoB 2).[23] Risk of bias in non-randomised controlled trials, before-after studies, historically controlled trials and cohort studies will be assessed using the Cochrane risk of bias in non-randomized studies of interventions (ROBINS-I) tool.[24] Qualitative studies will be appraised using the CASP Qualitative Studies Checklist.[25] We will use the NICE preferred appraisal tools for any other study design.[26] Two reviewers (JG/AA/SJ) will independently assess study risks of bias. Disagreements will be resolved through consensus, with the inclusion of a third reviewer (CS, YFC) if required.

5.3 Methods of analysis/synthesis

We will use the following effect measures for final health outcomes:

- Hazard ratio (HR) for time-to-event data (e.g. time to lung cancer specific mortality);
- Risk ratio for dichotomous outcomes (e.g. incidence of lung cancer during CT surveillance);
- Mean difference between arms for continuous outcomes (e.g. health-related quality of life).

If data permits and studies are clinically similar, we will pool the results in meta-analyses stratified by technology. For dichotomous outcomes, we will calculate the risk ratio and the 95% confidence interval for each study and then pool the studies. For time-to-event data we will pool the hazard ratios. For continuous outcomes, we will pool the mean difference and the 95% confidence interval at the end of follow-up if studies measure the outcome on the same scale. If studies measure the outcome using different scales, we will pool using the standardised mean difference and the 95% confidence interval. If data do not permit a pooled analysis, then we will conduct a narrative synthesis stratified by software.

Qualitative evidence will be analysed and summarised thematically.

6 Methods for assessing cost-effectiveness

Key question 3

What is the cost-effectiveness of using software for the automated detection and analysis of lung nodules from CT images compared with unassisted CT image analysis in people undergoing CT scans that include the chest due to symptoms suggestive of lung cancer, for purposes unrelated to suspicion of lung cancer, for surveillance of previously identified nodules or (depending on upcoming UK NSC recommendations) for lung cancer screening?

Sub-questions

1. Does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ between CT scans: (1) with contrast and without contrast; (2) using a low-dose and a standard dose; (3) of solid nodules and sub-solid nodules?
2. Does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ by patients' ethnicity?
3. Does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ between general radiologists/healthcare professionals and specialised thoracic radiologists/healthcare professionals?
4. In the incidental population, does the cost-effectiveness of CT image analysis assisted by software for automated detection and analysis of lung nodules differ by reason for CT scan?

6.1 Identification and selection of studies

6.1.1 Search strategy

The searches carried out for the systematic review of test accuracy and clinical effectiveness (see section 4.1.1) will be centred around the concepts of AI, lung nodules/cancer and CT or screening, without any restrictions in terms of study type filters. They can therefore be expected to also retrieve any studies relating to cost-effectiveness of using AI-based software in lung nodule/cancer CT imaging.

As there are likely to be few, if any, economic evaluations of cost-effectiveness studies of the use of AI-based software for nodule detection and analysis in this specific population and context, broader searches for lung nodules/cancer imaging or screening (without AI terms, and not specifically CT) will be undertaken to identify information on model structures, costs and utility values to inform the economic model. Where appropriate, search filters for economic evaluations and/or cost or HRQoL studies will be applied.

Sources will include:
MEDLINE All (Ovid);
Embase (Ovid);
National Health Service Economic Evaluation Database (NHS EED) (CRD);
Health Technology Assessment (HTA) database (CRD);
International HTA database (INAHTA);
Cost-Effectiveness Analysis (CEA) registry (Tufts Medical Center);
EconPapers (Research Papers in Economics (RePEc));
ScHARRHUD;
targeted web searches (Google);
selected organisations and conferences of interest (to include for example NICE, CADTH, ISPOR, HTAi, International Health Economics Association and Radiological Society of North America Annual Meetings);
reference lists of selected highly relevant papers.

6.1.2 Study eligibility criteria

Studies that satisfy the following criteria will be included:

Population	People who have no confirmed lung nodules or 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:

	• for reasons unrelated to suspicion of lung cancer (incidental population); • because of signs or symptoms suggestive of lung cancer (symptomatic population); • as part of lung cancer screening (inclusion of the screening population will be confirmed following the upcoming lung cancer screening recommendations of the UK NSC). People having CT surveillance for a previously identified lung nodule. Where data permits, the following subgroups may be considered: • Patients' ethnicity; • People who have a CT scan: (1) with or without contrast; (2) using a low-dose or a standard dose; (3) of solid nodules or sub-solid nodules; • For the incidental population, by reason for CT scan.
Target condition	Lung cancer
Intervention	CT scan review by a radiologist or another healthcare professional using any of the included 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) * Indication for use specifies use in asymptomatic population, therefore the software cannot be assessed in symptomatic population.

	** Indication for use specifies use in symptomatic population, therefore the software cannot be assessed in incidental or screening populations. Please note: specific indications for use for some of the technologies are unclear because only information in the public domain was available. Where data permits, the following subgroups may be considered: - General radiologist/other healthcare professional with software support versus radiologist/other healthcare professional with thoracic speciality with software support.
Comparator	CT scan review by a radiologist or another healthcare professional without software for automated detection and analysis of lung nodules (using diameter or volume to measure nodule size). Where data permits, the following subgroups may be considered: - General radiologist/other healthcare professional without software support; radiologist/other healthcare professional with thoracic speciality without software support.
Outcomes	Cost effectiveness (e.g., incremental costs, incremental benefits, incremental cost effectiveness ratio, quality adjusted life years)
Study design	Full economic evaluations (including cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis). Cost minimisation analysis, cost-consequence/outcome description, costs analysis (UK only) and cost description (UK only) may also be included if full economic evaluations are lacking.
Publication type	Peer reviewed papers. Abstracts and manufacturer data will be included, but only outcome data that have not been reported in peer-reviewed full-text papers will be extracted and reported.
Language	English

Exclusion criteria are the same as described in clinical effectiveness review section.

6.1.3 Review strategy

All records retrieved will be screened independently by 2 reviewers (PA/HG) at title/abstract stage, of which potentially relevant records will be further examined at full-text. Any disagreements between the reviewers will be resolved by a discussion, or recourse to a third reviewer (AA or JM) if an agreement cannot be reached.

6.2 Extraction and study quality

6.2.1 Data extraction strategy

Information will be extracted by 2 reviewers (PA/HG) independently, using a pre-piloted data extraction form for the full economic evaluation studies. The data extraction form will be developed to summarise the main characteristics of the studies and to capture useful information for the economic model. From each paper included in the systematic review, we will extract information about study details (title, author and year of study), baseline characteristics (population, intervention, comparator and outcomes), methods (study perspective, time horizon, discount rate, measure of effectiveness current, assumptions and analytical methods), results (study parameters, base-case and sensitivity analysis results), discussion (study findings, limitations of the models and generalisability), other (source of funding and conflicts of interests), overall reviewer comments and conclusion (author's and reviewer's). Each reviewer will cross-check each other's extractions, with any discrepancies resolved by discussion, or recourse to a third reviewer if an agreement cannot be reached.

6.2.2 Assessment of study methodological quality

The quality of any full economic evaluation studies will be assessed using the consolidated health economic evaluation reporting standards (CHEERS) checklist.[27] Any studies using an economic model will be further assessed against the framework for the quality assessment of decision analytic modelling developed by Philips and colleagues.[28]

6.3 Methods of analysis/synthesis

Due to the nature of economic analyses (different aims/objectives, study designs, populations, and methods) these findings from individual studies will be compared narratively, and recommendations for future economic analyses will be discussed.

6.3.1 Evaluation of costs, health-related quality of life and cost-effectiveness

Model structure

The strategy for inclusion in the cost-effectiveness analysis is reading of CT scans by radiologists or other health professionals assisted by software with AI derived algorithms for automated detection and analysis of lung nodules. This will be compared against the current strategy of image analysis without being assisted by software with AI derived algorithms for automated nodule detection and analysis. If appropriate model-based cost-effectiveness studies addressing the review question are not found, a de novo economic model will be constructed. In constructing the economic model, we will

identify from the literature previous economic models that compared different methods used to detect and analyse lung nodules. We will review these studies and their appropriateness to the decision problem, and review their structures and evidence used to populate the model.

The development of the model will be an iterative process. First, we will develop a conceptual model, informed by the clinical effectiveness review and in consultation with clinical experts to capture the current clinical pathway(s) for detecting/classifying lung nodules. The conceptual model will be used to present a simplified representation of the decision problem and assessment pathway and will be used to identify the information required to parameterise the model, and assumptions that are likely to be made as well as highlight any areas of uncertainties. We will follow the current assessment and care pathways for solid lung nodules and sub-solid nodules, comprising initial nodule detection and classification based on morphology and nodule size (diameter or volume) based on the BTS guidelines.[12] Some of the identified nodules will be subject to assessment of risk of malignancy using Brock model and risk assessment using the Herder risk model following PET-CT. The availability of sufficient data to inform modelling of these processes in the assessment pathway is highly uncertain, and simplification and assumptions are likely to be required to link the initial impact of using software-assisted radiologist review to subsequent processes and outcomes. The assessment pathway will be linked to treatment pathway and subsequent outcomes according to the NICE Pathway.[19]

We anticipate that parameterisation will be driven by the findings from the clinical effectiveness systematic review and supported by clinical expert opinion. We anticipate that the overall model is likely to comprise 2 sections. In the first phase of the model will aim to predict the impact of software-assisted CT reading on if and when cancer is detected in those who have one. It will also aim to predict the surveillance and associated costs and disutilities experienced by people who do not have a malignancy. The second phase will use this information as input and estimate the consequential health outcomes and overdiagnosis resulting from the chosen approach to CT scanning. It is likely that the first phase will follow a decision tree structure and the second phase a Markov/discrete event simulation, but this will be finalised once the clinical background and available data have been confirmed.

Proposed position of software in the current pathway

The technologies will be used in the detection and analysis of lung nodules in the target populations.

Information required to populate the model

The model will be populated with clinical information obtained from the clinical effectiveness systematic review and meta-analysis. Clinical information will likely be required on the prevalence (by population) and sensitivity and specificity stratified by ‘actionable nodules’ (defined as nodules

requiring further investigation, surveillance, or treatment) or malignant nodules for radiologist review of CT scans with software support and radiologist review of CT scans without software support; differences in the number of nodules referred for surveillance and in stage of cancer at diagnosis; differences in time required for reviewing each CT scan image; potential level of overdiagnosis; and differences in these parameters between subgroups of interest.

Resource use and costs

As part of the framework to undertake the economic analysis, information will be required about the resource use and costs associated with the testing strategies used to identify lung nodules that would warrant further assessment or monitoring. Additionally, resource use and costs will be required for the long-term management and surveillance of people with lung nodules. Unit costs will be obtained from the published literature, or from national sources [NHS reference costs schedules and Personal Social Services Research Unit (PSSRU) Costs from Health and Social Care].

Costs will be attributed to implementing and using the artificial intelligence-based software. If data permits, we will develop an inventory that may include the cost of purchasing, implementing, running, maintenance, and updating of the software of interest, and training the radiologists to use the software in their ongoing practice. Other costs will be considered: confirmatory diagnostic (biopsy, bronchoscopy, additional CT-scan, MRI, and PET-CT scan), and then therapeutical interventions (partial or radical nodules excision/surgery, radiotherapy, chemotherapy, and palliative care) costs.

Health outcomes

Several outcome measures will be considered in the economic analysis and will be evaluated if suitable data permit: correct diagnosis, time-to-detection, averted lung cancer deaths, life-years (LY) and quality-adjusted life-years (QALYs) gained. LY and QALYs gained will be calculated from survival information, including incidence and survival of lung cancer, and utility values obtained from the literature and other sources (e.g., elicited from experts). QALYs accrued will be derived based on the utility payoff assigned to the health states occupied along the management pathway. Under each strategy the expected mean benefits yielded are summed over the model time horizon and discounted at a 3.5% per annum rate. We will consider harms (e.g., false positives, false negatives and overdiagnosis) associated with these strategies, including potential disutility caused by anxiety and complications arising from interventions.

Proposed evidence linkage

The proposed evidence linkage will be based on the value of the features of radiologist review of CT scans with software support compared to radiologist review of CT scans without software support in the economic model and how the clinical effectiveness evidence propagates through the assessment

and treatment pathway to link with health and cost outcomes. The components of interest include additional costs associated with the use of software support, additional costs that may be associated with increased nodule detection that may require further investigation (following the assessment of risk of malignancy or PET-CT scans), better outcomes based on early detection of lung cancer, and reduction/increase in the proportion of people requiring CT surveillance.

For simplicity, people in the model will be assumed to be diagnosed as having a benign or malignant nodule following all strategies confirmed by histological analysis of lung biopsy or CT surveillance (imaging follow-up) without significant growth. The longer-term impact of each strategy will be modelled using a Markov/discrete event simulation to simulate the disease progression following diagnosis of malignant nodules and benign nodules. Transitions between disease stages will be obtained from the literature. Several simplifying assumptions will be made to have a workable model.

The link to longer-term mortality outcomes will be modelled via transitions to dead state, with mortality risk based on true underlying state (benign or malignant and the stage of diagnosis). Mortality rates for people who are benign will be derived from mortality rates from the UK general population. Mortality rates following treatment will be obtained from the literature. If data permits, we will consider increased mortality for smoking status.

Costs related to treatment will be considered in the economic model. Treatment will include surgery alone, surgery in combination with chemotherapy/radiotherapy or palliative care. We will assume that treatment will be based on the stage of the cancer and its histological type (non-small cell lung cancer and small-cell lung cancer) as recommended by the NICE Pathway for lung cancer.[19] Any adverse events/complications following treatment and palliation will be considered in the model.

We envisaged using cancer-stage specific health-state utility values for people with malignant tumours, which will be obtained from the published literature. If appropriate, we will consult with clinical experts and the literature to determine if there may be disutility associated with benign nodules. We will use age-related utility values from the UK population norms for people with benign nodules and apply age-related disutilities to reflect an ageing population.

Economic evaluation

If data permits, we will undertake a cost-effectiveness analysis where the ratio between the costs incurred and benefits accrued from using healthcare professional-read CT scan with software support compared to radiologist-read CT scan without software support from the NHS and Personal Social Services (PSS) perspective in a secondary care setting. The results of the analysis will be presented in

terms of an incremental cost-effectiveness analysis (ICER) over a short-term time horizon, which will consider intermediate outcomes (e.g., correct diagnosis and time-to-diagnosis) and lifetime horizon, where each strategy will be ranked, excluding options that were dominated or extendedly dominated, with results expressed as cost per QALY and net benefits.

We will use univariate one-way sensitivity analysis to explore the impact of varying one parameter at a time, whilst keeping all other inputs constant to assess the robustness of the model, with results presented in the form of a tornado diagram. Where data permits, we will undertake probabilistic sensitivity analysis to determine the impact of joint parameter uncertainty. In probabilistic sensitivity analysis, model parameters are assigned a distribution reflecting the amount and pattern of its variation, and cost-effectiveness results are calculated by simultaneously selecting random values from each distribution. This process is repeated several times, with the simulations plotted on an incremental cost-effectiveness plane; each point representing uncertainty in the incremental mean costs and QALYs between the strategies being compared. The results from these simulations will be used to obtain cost-effectiveness acceptability curves (CEAC), which illustrate the effect of sampling uncertainty, and presents the probability that an intervention is optimal at a range of willingness-to-pay (WTP) threshold values.[29] Where sufficient evidence permits, we anticipate undertaking scenario analyses for general radiologists and specialist chest specialist. Other scenario analyses will be undertaken as required through model development.

Where there is no data showing impact of the technologies on detection of malignant nodules in surveillance (detect more malignant nodules or detect earlier) and if there are any data comparing resource use with/without use of the software we will conduct an exploratory cost-comparison analysis. A cost-comparison analysis would comprise an analysis of the resource use and costs associated with healthcare professional-read CT scan with software support compared to that of the healthcare professional-read CT scan without software support over a time horizon long enough to capture the important differences between the technologies being compared.

Areas anticipated to be beyond the scope of the assessment

Quantitative evaluation of potential effects of using AI-derived software on workflow, changes in the interactions between health professionals and patients and between different health professionals and impact on workload and staffing is beyond the scope of the current assessment, except where evidence on radiologist's reading time and/or radiology turnaround time related to the use of the software is found, it will be taken into account in the estimation of costs. Evaluation of cost-effectiveness of use of the technologies for lung cancer screening population is dependent on the forthcoming UK NSC recommendation. Implementation of the planned methods described in this

protocol may be restricted by lack of required evidence and/or substantial uncertainties related to available quantity and quality of data.

7 Handling of information from manufacturers

All data submitted by the manufacturers/sponsors/stakeholders will only be considered if received by the EAG 2 months before the submission date for the DAR. Data that arrives after this date will not be considered. We will extract and quality appraise any data that meets the inclusion criteria, as stated in the methods section of this protocol.

Any 'commercial in confidence' data that is provided by manufacturers, academics, clinicians, or stakeholders, and specified as such, will be highlighted in blue and underlined in the assessment report (followed by company name in parentheses). Any 'academic in confidence' data that is provided by manufacturers, and specified as such, will be highlighted in yellow and underlined in the assessment report. All confidential data used in the cost-effectiveness models will also be highlighted. If confidential information is included in the model, we will provide a model with 'dummy variable values' for the confidential values (i.e. using non-confidential values).

8 Competing interests of authors and advisors

None of the authors have any competing interests.

9 Timetable/milestones

Draft assessment protocol	10th November 2021
Final protocol	6th December 2021
Progress report	3rd March 2022
Draft assessment report	3rd May 2022
Final assessment report	31st May 2022

10 Team members' contributions

Warwick Evidence is an External Assessment Group located within Warwick Medical School. Warwick Evidence brings together experts in clinical and cost effectiveness reviewing, medical statistics, health economics and modelling. The team planned for the work include:

Name: Julia Geppert
Title: Research Fellow
Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]
Email: [redacted]
Contribution: Protocol development, clinical effectiveness reviewer, report writing.

Name: Peter Auguste
Title: Research Fellow
Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]
Email: [redacted]
Contribution: Cost effectiveness lead: Protocol development, systematic review of the health economic literature, health economic modelling, and report writing.

Name: Asra Asgharzadeh,
Title: Research Fellow
Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]
Email: [redacted]
Contribution: Protocol development, clinical effectiveness reviewer, report writing.

Name: Anna Brown
Title: Information Specialist
Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]
Tel: [redacted]
Email: [redacted]
Contribution: Develop the search strategies, undertake searches, write the search methods sections of the draft and final versions of the report and manage references.

Name: Surangi Jayakody
Title: Visiting Research Fellow
Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]
Email: [redacted]

Contribution: Protocol development, clinical effectiveness reviewer, report writing.

Name: Hesam Ghiasvand

Title: Research Fellow

Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]

Email: [redacted]

Contribution: Protocol development, systematic review of the health economic literature, health economic modelling, and report writing.

Name: Mubarak Patel

Title: Research Associate

Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]

Email: [redacted]

Contribution: Statistical support.

Name: Emma Helm

Title: Consultant Radiologist

Address: [redacted]

Email: [redacted]

Contribution: Provide clinical advice and help revise the protocol.

Name: Dan Todkill

Title: Clinical Research Fellow

Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]

Email: [redacted]

Contribution: Quality assurance.

Name: Jason Madan

Title: Professor

Address: Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, [redacted]

Email: [redacted]

Contribution: Methodological advisor on cost-effectiveness.

Name: Chris Stinton

Title: Senior Research Fellow

Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]

Tel: [redacted]

Email: [redacted]

Contribution: Methodological advisor, designing and planning the review, support and training for the clinical effectiveness team.

Name: Yen-Fu Chen

Title: Associate Professor

Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]

Tel: [redacted]

Email: [redacted]

Contribution: Project lead, design and planning of the review, implementation and write up.

Name: Sian Taylor-Phillips

Title: Professor of Population Health

Address: Populations, Evidence and Technologies, Division of Health Sciences, Warwick Medical School, University of Warwick, [redacted]

Tel: [redacted]

Email: [redacted]

Contribution: Senior advisor, design and planning of the review, and write up.

11 Expert advisors

Name: Charles Hutchinson

Title: Professor of Clinical Imaging

Address: [redacted]

Email: [redacted]

Contribution: Provide clinical advice.

Name: Ben Glocker

Title: Reader in Machine Learning for Imaging

Address: [redacted]

Email: [redacted]

Contribution: Provide advice on application of AI in imaging.

12 References

1. HM Government. *National AI strategy.* Office for Artificial Intelligence; 2021. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf (Accessed 6 December 2021).

2. Larici AR, Farchione A, Franchi P, Ciliberto M, Cicchetti G, Calandriello L*, et al.* Lung nodules: size still matters. *Eur Respir Rev* 2017;**26**:170025. http://dx.doi.org/10.1183/16000617.0025-2017

3. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. Recommendations for measuring pulmonary nodules at CT: a statement from the Fleischner Society. *Radiology* 2017;**285**(2):584-600. http://dx.doi.org/10.1148/radiol.2017162894

4. Horeweg N, van Rosmalen J, Heuvelmans MA, van der Aalst CM, Vliegenthart R, Scholten ET*, et al.* Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* 2014;**15**(12):1332-41. http://dx.doi.org/10.1016/s1470-2045(14)70389-4

5. Cancer Research UK. *Lung cancer incidence statistics*. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence (Accessed 6 December 2021).

6. National Disease Registration Service. *Staging data in England*. Public Health England; 2020. URL: https://www.cancerdata.nhs.uk/stage_at_diagnosis (Accessed 10 November 2021).

7. NHS. *NHS long term plan*. URL: https://www.longtermplan.nhs.uk/ (Accessed 9 November 2021).

8. National Institute for Health and Care Excellence. *Suspected cancer: recognition and referral. NICE guideline [NG12]*. NICE; 2015. URL: https://www.nice.org.uk/guidance/ng12 (Accessed 1 December 2021).

9. National Institute for Health and Care Excellence. *Lung cancer: diagnosis and management. NICE guideline [NG122]*. NICE; 2019. URL: https://www.nice.org.uk/guidance/ng122 (Accessed 9 November 2021).

10. UK National Screening Committee. *Adult screening programme: lung cancer*. Gov.uk. URL: https://view-health-screening-recommendations.service.gov.uk/lung-cancer/ (Accessed 1 December 2021).

11. NHS England. *Evaluation of the Targeted Lung Health Check programme*. URL: https://www.england.nhs.uk/contact-us/privacy-notice/how-we-use-your-information/our-services/evaluation-of-the-targeted-lung-health-check-programme/ (Accessed 1 December 2021).

12. Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J*, et al.* British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;**70**(Suppl 2):ii1-ii54. http://dx.doi.org/10.1136/thoraxjnl-2015-207168

13. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K*, et al.* Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;**369**(10):910-9. http://dx.doi.org/10.1056/NEJMoa1214726

14. British Thoracic Society. *PN risk calculator*. URL: https://www.brit-thoracic.org.uk/quality-improvement/guidelines/pulmonary-nodules/pn-risk-calculator/ (Accessed 9 November 2021).

15. Herder GJ, van Tinteren H, Golding RP, Kostense PJ, Comans EF, Smit EF, *et al.* Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest* 2005;**128**(4):2490-6. http://dx.doi.org/10.1378/chest.128.4.2490

16. British Thoracic Society. *BTS guidelines for the investigation and management of pulmonary nodules*. URL: https://www.brit-thoracic.org.uk/quality-improvement/guidelines/pulmonary-nodules/ (Accessed 1 December 2021).

17. NHS England National Cancer Programme. *Targeted screening for lung cancer with low radiation dose computed tomography: standard protocol prepared for the NHS England Targeted Lung Health Checks programme* NHS; 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2019/02/targeted-lung-health-checks-standard-protocol-v1.pdf (Accessed 1 December 2021).

18. British Society of Thoracic Imaging, The Royal College of Radiologists. *Considerations to ensure optimum roll-out of targeted lung cancer screening over the next five years* The Royal College of Radiologists; 2020. URL: https://www.rcr.ac.uk/posts/considerations-ensure-optimum-roll-out-targeted-lung-cancer-screening-over-next-five-years (Accessed 1 December 2021).

19. National Institute for Health and Care Excellence. *NICE pathways: lung cancer overview*. NICE; 2021. URL: https://pathways.nice.org.uk/pathways/lung-cancer (Accessed 9 November 2021).

20. Cochrane Screening and Diagnostic Test Methods Group. *Cochrane handbook for systematic reviews of diagnostic test accuracy*. Cochrane Collaboration; 2013. URL: https://methods.cochrane.org/sdt/handbook-dta-reviews (Accessed 9 November 2021).

21. National Institute for Health and Clinical Excellence. *Diagnostics Assessment Programme manual*. Manchester: NICE; 2011. URL: https://www.nice.org.uk/media/default/about/what-we-do/nice-guidance/nice-diagnostics-guidance/diagnostics-assessment-programme-manual.pdf (Accessed 9 November 2021).

22. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**(8):529-36. http://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009

23. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* 2019;**366**:l4898. http://dx.doi.org/10.1136/bmj.l4898

24. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj* 2016;**355**:i4919. http://dx.doi.org/10.1136/bmj.i4919

25. Critical Appraisals Skills Programme (CASP). *CASP qualitative studies checklist*. 2018. URL: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf (Accessed 1 December 2021).

26. National Institute for Health and Care Excellence. Appendix H: appraisal checklists, evidence tables, GRADE and economic profiles. In: *Developing NICE guidelines: the manual.* NICE; 2014. URL: https://www.nice.org.uk/process/pmg20/resources/appendix-h-appraisal-checklists-evidence-tables-grade-and-economic-profiles-pdf-8779777885 (Accessed 6 December 2021).

27. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D*, et al.* Consolidated health economic evaluation reporting standards (CHEERS) statement. *Eur J Health Econ* 2013;**14**(3):367-72. http://dx.doi.org/10.1007/s10198-013-0471-6

28. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R*, et al.* Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**(36).

29. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001;**10**(8):779-87.

13 Appendix

Appendix 1. Draft search strategy

Ovid MEDLINE(R) ALL <1946 to November 30, 2021>

Date searched: 1/12/21

Search Strategy:

1 exp artificial intelligence/ or exp machine learning/ or exp deep learning/ or exp supervised machine learning/ or exp support vector machine/ or exp unsupervised machine learning/ 130247

2 ai.kf,tw. 33295

3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 85516

4 exp Neural Networks, Computer/ 40461

5 (neural network* or convolutional or CNN or CNNs).kf,tw. 71553

6 exp Diagnosis, Computer-Assisted/ 85235

7 ((computer aided or computer assisted) adj1 (diagnosis or detection)).kf,tw. 5908

8 (support vector machine* or random forest* or black box learning).kf,tw. 30273

9 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 [AI search string based on https://dx.doi.org/10.1136/bmj.n1872 , plus some additional terms and .kf,tw instead of .mp, Algorithms/ removed after testing] 316093

10 exp Lung Neoplasms/di, dg or Solitary Pulmonary Nodule/di, dg 56097

11 ((lung or lungs or pulmon* or bronchial) adj3 (nodul* or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 271939

12 ((pulmonary or lung) adj2 lesion*).kf,tw. 14650

13 10 or 11 or 12 299976

14 Tomography, X-Ray Computed/ or exp Tomography, Spiral Computed/ 417132

15 (comput* adj2 tomograph*).kf,tw. 344949

16 (CT or LDCT).kf,tw. 385298

17 (CAT adj2 (scan* or x-ray* or xray*)).kf,tw. 1338

18 Mass Screening/ 111107

19 ((lung or lungs or pulmon*) adj3 (nodule* or cancer* or tumor* or tumour*) adj3 screen*).kf,tw. 4748

20 "Early Detection of Cancer"/ 31301

21 14 or 15 or 16 or 17 or 18 or 19 or 20 886710

22 9 and 13 and 21 2711

23 (aview* lcs* or clearread* ct* or inferread* ct lung* or lung nodule ai* or veolity* or veye).kf,tw. 6

24 ((ai rad companion* and chest) or contextflow* or search lung ct* or "jld 01k*" or qct lung* or sensecare* lung* or visia* ct* or vuno).kf,tw. 6

25 (coreline* or riverain* or infervision* or fujifilm* or mevis* or aidence*).in,kf,tw. 1341

26 (siemens* healthineers* or contextflow* or jlk inc* or arterys* or qureai* or qure ai* or sensetime* or canon medical* or vuno*).in,kf,tw. 1356

27 (25 or 26) and (10 or 11) 153

28 22 or 23 or 24 or 27 2807

29 exp animals/ not humans/ 4923451

30 28 not 29 2791

31 limit 30 to english language 2681

The artificial intelligence search terms (lines 1-4 & 6) are based on those used in:
Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy BMJ 2021; 374 :n1872 doi:10.1136/bmj.n1872 (see online supplementary appendix 1)

Selected lung cancer/nodule search terms (lines 11-12) were informed by those used in:
Duarte A, Corbett M, Melton H, Harden M, Palmer S, Soares M, Simmonds M. EarlyCDT Lung for lung cancer risk classification of solid pulmonary nodules: A Diagnostics Assessment Report. York EAG, 2021. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-dg10041/documents (accessed 9 November 2021)