Early Value Assessment report commissioned by the NIHR Evidence Synthesis Programme on behalf of the National Institute for Health and Clinical Excellence – protocol

Title of project: Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer

Name of External Assessment Group (EAG) and project lead

Produced by: Warwick Evidence

Project lead: Chris Stinton

Clinical effectiveness lead: Jill Colquitt

Health economics lead: Mary Jordan

Clinical effectiveness: Jill Colquitt, Emma Loveman, Janette Parr, Iman Ghosh

Health economics: Mary Jordan, Peter Auguste

Information specialist: Rachel Court

Statistics: Mubarak Patel

Correspondence to: Warwick Evidence
Division of Health Sciences
Warwick Medical School
University of Warwick
Coventry CV4 7AL

Date completed:

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

Glossary of terms

AI	Artificial intelligence
ASG	Assessment subgroup
CXR	Chest x-ray
DAC	Diagnostic Advisory Committee
EAG	External Assessment Group
EVA	Early value assessment
NHS	National Health Service
NICE	The National Institute for Health and Care Excellence
UK	United Kingdom

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, so it is often found late. Treatments for cancers that are found late are usually not as helpful as for those that are found early.

Healthcare professionals can find signs of lung cancer on chest x-ray images. Computer software (artificial intelligence) is available that might help with this. This software can find features (e.g., nodules) that need looking at more closely. Finding these might help healthcare professionals decide who needs extra tests, and how quickly. This could help to find and treat lung cancer early when treatments might work better.

The purpose of this project is to find evidence on this type of software and to find gaps in the evidence. This may help guide further research and data collection. We will also make a conceptual model to allow discussion of what evidence is needed to make a full cost-effectiveness model to understand the costs of using the software to find lung cancer.

2. Purpose of the decision to be made

Lung cancer occurs when abnormal cells multiply in an uncontrolled way to form a tumour in the lung.[1] It is one of the most common types of cancer in the UK and each year over 43,000 new cases are diagnosed.[2] In the early stages of the disease people usually do not have symptoms which means lung cancer is often diagnosed late.[3] The five-year survival rate for lung cancer is low, at below 10%.[2] Early diagnosis may improve survival.[3] NICE has identified software that has an artificial intelligence (AI) developed algorithm (referred to hereafter as AI software) as potentially useful in assisting with the identification of suspected lung cancer.

The purpose of this early value assessment (EVA) is to assess the evidence on adjunct AI software for analysing chest x-rays (CXR) for suspected lung cancer, and identify evidence gaps to help direct data collection and further research. A conceptual modelling process will be undertaken to inform discussion of what would be required to develop a full-executable cost-effectiveness model for future economic evaluation.

3. Decision problem

3.1. Population

There are two populations of interest in this EVA: (1) people referred from primary care who are having CXR because they have symptoms suggestive of lung cancer (symptomatic population), and (2) people referred from primary care who are having CXR for reasons unrelated to lung cancer (incidental population).

3.2. Target condition: Lung cancer

Lung cancer is one of the most common causes of cancer in the UK. There approximately 43,000 new cases diagnosed annually.[2] The incidence of lung cancer is highest amongst older people.[4] It is rare in people under the age of 40. More than 40% of people diagnosed with lung cancer are 75 years or older.[3]

Lung cancer occurs when abnormal cells multiply in an uncontrolled way to form a tumour in the lung.[1] Cancer that begins in the lungs is called primary lung cancer. Cancer that begins elsewhere and spreads to the lungs is called secondary lung cancer. There are two main forms of primary lung cancer: non-small-cell lung cancer and small-cell lung cancer. These are named after the type of cell in which the cancer started growing. Non-small-cell lung cancer is the most common type (80-85% of cases) and can be classified into one of three kinds: squamous cell carcinoma, adenocarcinoma, or large-cell carcinoma. Small-cell lung cancer is less common but usually spreads faster than non-small-cell lung cancer.[3] Most cases of lung cancer are caused by smoking. Although people who have never smoked can also develop the condition, smoking cigarettes is responsible for more than 70% of cases.[3] People who smoke are 25 times more likely to get lung cancer than people who do not smoke. Other exposures can also increase the risk of lung cancer. These include radon gas (naturally occurring), occupational exposure to certain chemicals and substances, and pollution.[3]

Symptoms of lung cancer include persistent cough, coughing up blood, and shortness of breath. However, in the early stages of the disease people usually do not have symptoms.[3] This means lung cancer is often diagnosed late. In 2018, more than 65% of lung cancers in England were diagnosed at stage 3. Survival rates for lung cancer are very low. Recent estimates suggest 5-year survival rates of 10%.[3] The NHS Long Term Plan sets out the NHS's ambition to diagnose 75% of all cancers at an early stage by 2028.[5]

3.3. Intervention

AI combines computer science and datasets to enable problem solving. Machine learning and deep learning are sub-fields of AI. They comprise AI algorithms which seek to create expert systems to make predictions or classifications based on data input.[6] Many paradigms of deep learning have been developed but the most used of these is the Convolutional Neural Network.[7]

This assessment covers the use of AI software as an adjunct to an appropriate radiology specialist to assist in the identification of suspected lung cancer. AI technologies subject to this assessment are standalone software platforms developed with deep learning algorithms to interpret CXR. The algorithms are fixed but updated periodically. The AI software automatically interprets radiology

images from the CXR to identify abnormalities or suspected abnormalities. The abnormalities detected and the methods of flagging the location and type of abnormalities differ between different AI technologies. For example, a CXR may be flagged as suspected lung cancer when a lung nodule, lung mass, hilar enlargement, or a combination of these are identified. A technology may classify CXRs into those with and without a nodule, or it may identify several different abnormalities or lung diseases.

3.4. Comparator

The comparator for this assessment is CXR images reviewed by an appropriate radiology specialist (e.g. radiologist or radiographer) without assistance from AI software.

3.5. Diagnostic test

Following CXR, people with suspected lung cancer should be offered a contrast-enhanced chest CT scan to diagnosis and stage the disease (contrast medium should only be given with caution to people with known renal impairment). The liver, adrenals and lower neck should also be included in the scan.[8] If the CT scan indicates there may be cancer, the type and sequence of investigations may vary but typically include a positron emission tomography-CT (PET-CT) scan and an image-guided biopsy. Other methods that may be used include MRI, endobronchial ultrasound-guided transbronchial needle aspiration (EBUS- TBNA), and endoscopic ultrasound-guided fine-needle aspiration (EUS-FNA).[8] The PET-CT scan can show where there are active cancer cells which can help with diagnosis and choosing the best treatment.[3]

3.6. Care pathway

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 lung cancer recommends that people aged 40 and over are offered an urgent CXR (within 2 weeks of referral) if they have two or more symptoms of lung cancer, or if they have ever smoked and have at least one of the following unexplained symptoms: cough, fatigue, shortness of breath, chest pain, weight loss, appetite loss.[9] An urgent CXR 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 CXR findings suggest lung cancer, referral to secondary care should be made using a suspected cancer pathway referral for an appointment within 2 weeks. If the CXR is normal (without any clinically relevant lung abnormalities), high risk patients, i.e., those who present with ongoing, unexplained symptoms, are referred to secondary care. Low risk patients are discharged. In this EVA, AI software is applied to

CXR of patients who are referred from primary care. Referrals for CXR outside primary care are beyond the scope of this project.

3 Objectives

The overall aim of this project is to identify evidence on adjunct AI software for analysing CXR for suspected lung cancer, and identify evidence gaps to help direct data collection and further research. A conceptual modelling process will be undertaken to inform discussion of what would be required to develop a full-executable cost-effectiveness model for future economic evaluation. These will be examined via an EVA. The assessment is not intended to replace the need for a full assessment (Diagnostic Assessment Report) or to provide sufficient detail or synthesis to enable a recommendation to be made on whether AI software can be implemented in clinical practice at the present time.

Based on the scope produced by NICE[10] we defined the following questions to inform future assessment on the benefits, harms, and costs of adjunct AI for analysing on CXR for suspected lung cancer compared to human reader alone:

1. What is the test accuracy and test failure rate of adjunct AI software to detect lung cancer on CXR?
2. What are the practical implications of adjunct AI to detect lung cancer on CXR?
3. What is the clinical effectiveness of adjunct AI software applied to CXR?
4. What are the cost and resource use considerations relating to use of adjunct AI to detect lung cancer?
5. What would a health economic model to estimate the cost-effectiveness of adjunct AI to detect lung cancer look like?

4 Methods for assessing test accuracy, practical implications, and clinical effectiveness questions

The timeline to produce this EVA report is ten weeks, which is substantially shorter than a typical systematic review or rapid review. To achieve the aims within the timeline, pragmatic decisions regarding the methods have been made in collaboration with NICE and clinical experts. These include:

- Targeted searches (section 4.2)
- Methods for screening evidence (section 4.3), extracting data (section 4.4), appraising risk of bias (section 4.5), and dealing with missing data (section 4.6)
- Prioritisation of evidence (section 4.7)

4.1 Inclusion criteria

The inclusion and exclusion criteria for the test accuracy, practical implications, and clinical effectiveness questions are presented in Table 1.

Table 1. Study eligibility criteria

	Key question 1. What are the test accuracy and test failure rates of adjunct AI software to detect lung cancer on CXR? **Sub-questions:** **1a. What is the test accuracy of adjunct AI software to detect lung nodules?** **1b. What is the concordance in lung nodule detection between radiology specialist with and without adjunct AI software**	**Key question 2. What are the practical implications of adjunct AI software to detect lung cancer on CXR?[a]**	**Key question 3. What is the clinical effectiveness of adjunct AI software applied to CXR?**
Population	Adults referred from primary care who are: 1. undergoing CXR due to symptoms suggestive of lung cancer, e.g. cough, fatigue, shortness of breath, chest pain, weight loss, appetite loss, persistent or recurrent chest infection, finger clubbing, supraclavicular lymphadenopathy or persistent cervical lymphadenopathy, chest signs consistent with lung cancer and/or thrombocytosis (symptomatic population) 2. undergoing CXR for reasons unrelated to lung cancer (incidental population) Where data permits, subgroups will be considered based on: • Ethnicity • Age • Sex • Socio-economic status		
Target condition	Lung cancer		
Intervention	CXR interpreted by radiology specialist (e.g. radiologist or radiographer) in conjunction with the following AI software: AI-Rad Companion Chest X-ray (Siemens Healthineers), Annalise CXR (annalise.ai), Auto Lung Nodule Detection (Samsung), ChestLink		

	Radiology Automation (Oxipit), ChestView (GLEAMER), Chest X-ray (Rayscape), ClearRead Xray – Detect (Riverain Technologies), InferRead DR Chest (Infervision), Lunit INSIGHT CXR (Lunit), Milvue Suite (Milvue), qXR (Qure.ai), red dot (behold.ai), SenseCare-Chest DR Pro (SenseTime), VUNO Med-Chest X-Ray (VUNO)		
Comparator	CXR interpreted by radiology specialist without the use of AI software		
Reference standard	For accuracy of lung cancer detection: Lung cancer confirmed by histological analysis of lung biopsy, or diagnostic methods specified in NICE guideline 122,[8] where biopsy is not applicable For accuracy of nodule detection: Radiology specialist (single reader or consensus of more than one reader)	NA	NA
Outcome	Test accuracy for the detection of lung cancer (sensitivity, specificity, positive predictive value, numbers of true positive, false positive, true negative, false negative results, number of lung cancers diagnosed) Test failures (rates, and data on inconclusive, indeterminate, and excluded samples, failure due to any other reason) Characteristics of discordant cancers cases	Practical implications[a] (time to x-ray report, CT scan, diagnosis, turnaround time (image review to radiology report), acceptability of software to clinicians, impact on clinical decision-making, impact of false positives on workflow)	Mortality, morbidity, health-related quality of life

	Test accuracy for the detection of lung nodules Concordance in lung nodule detection between radiology specialist with and without adjunct AI software		
Study design	Comparative study designs		
Publication type	Peer reviewed papers		
Language	English		
Exclusion	Versions of AI software that are not commercially available, are not named in the protocol, or are not specified in the study publication. Computer aided detection that does not include AI software. Non-human studies. Letters, editorials, communications, conference abstracts, qualitative studies. People with a known diagnosis of lung cancer at the time of CXR. Studies of children. Study designs that do not include a control/comparator arm. Simulation studies or studies using synthetic images. Studies not applicable to primary care patients, e.g., neurosurgery, transplant, or plastic surgery patients, people in secure forensic mental health services. Studies where more than 10% of the sample do not meet our inclusion criteria. Studies without extractable numerical data. Studies that provided insufficient information for assessment of methodological quality/risk of bias. Articles not available in the English language. Studies using index tests or reference standards other than those specified in the inclusion criteria. Studies of people who do not have signs and symptoms of cancer or a suspected condition or trauma (i.e., people undergoing health screening). Studies where it cannot be determined if the inclusion criteria are met.		

[a] For the 'acceptability' and 'impact on decision-making' outcomes, the relevant population is the radiologist or radiographer interpreting the CXR of adults defined under 'Population'.

4.2 Search strategy

An iterative approach will be taken to develop the search strategy, making use of relevant records identified during initial scoping searches and from relevant reviews.[11, 12] The strategy will be developed by an information specialist, with input from team members, aiming for a reasonable balance of sensitivity and specificity. Based on scoping work already undertaken, a series of complementary, targeted searches is favoured over a single search to retrieve a manageable number of records to screen (see Appendix 1). Searches will be run in a range of relevant bibliographic databases covering the fields of medicine and computer science. Searches will be limited to studies published in English because studies published in other languages are likely to be difficult to assess in the timescale of this EVA. Where possible and if reliable limits are available, non-human studies, letters, editorials, communications, and conference abstracts will be removed during the searches. A publication date limit may be applied, if it can be safely assumed there were no studies on the technologies of interest published before the chosen date. Database search strings will be developed for MEDLINE and appropriately translated for each of the other databases, considering differences in thesaurus terms and syntax. Example searches are provided in Appendix 1. The following bibliographic databases will be searched:

MEDLINE All (via Ovid), Embase (Ovid), Cochrane Database of Systematic Reviews (Wiley), Cochrane CENTRAL (Wiley), Epistemonikos, ACM Digital Library.

Searches for ongoing trials will be conducted in the following trials register resources:

ClinicalTrials.gov, WHO ICTRP.

Records will be exported to EndNote X9.3, where duplicates will be systematically identified and removed. Reference lists of included studies and a selection of relevant reviews will be checked. Experts and team members will be consulted and encouraged to share relevant studies.

4.3 Review strategy

Titles and abstracts of records identified by the searches will screened by one reviewer. A second reviewer will independently assess a random 20% sample of the titles/abstracts. We will retrieve the full publication of records considered potentially relevant by either reviewer. Full text articles will be assessed against the full inclusion/exclusion criteria by one reviewer, with a random 20% sample assessed independently by a second reviewer. Disagreements will be resolved by consensus, or through discussion with a third reviewer. Records rejected at full text stage (including reasons for exclusion) will be documented.

4.4 Data extraction strategy

Data will be extracted by one reviewer, with a random 20% checked by a second reviewer. All data extraction will be entered into a piloted electronic data collection form. Any disagreements will be resolved by consensus or discussion with a third reviewer.

We will consult clinical advisors in terms of relevance of studies to current UK practice. Studies will be deprioritised if they are not clinically relevant to current practice. Data from deprioritised studies will not be fully extracted, but a summary of these studies will be provided in a table. Summary information on ongoing trials will also be presented in a table.

4.5 Assessment of study risk of bias

Risk of bias of all included studies will be assessed by one reviewer, with a random 20% assessed by a second reviewer. We will use tools that are appropriate to the study design, e.g. those produced by the Joanna Briggs Institute (JBI).[14] Deprioritised studies will not be assessed for risk of bias.

4.6 Methods of analysis/synthesis

Test accuracy results will be presented by testing strategy. Meta-analysis of diagnostic test accuracy studies will be considered if time allows and sufficient data (at least five studies) from reasonably homogenous studies are available.

Meta-analysis of clinical effectiveness will be considered if time allows and sufficient data (at least five studies) from reasonably homogeneous studies are available. This will be guided by study design, population, outcomes, and risk of bias assessment. Homogeneity will be measured using appropriate statistical tests and by assessing the aforementioned study characteristics.

Missing data will be excluded from analyses. Methods of imputation will not be performed, nor will we attempt to contact authors to get pertinent missing data due to a lack of time.

If a meta-analysis is deemed inappropriate in either of the above cases findings from individual studies will be compared narratively.

We will provide a description of the evidence gaps in terms of the outcomes specified in Table 1. Based on this, recommendations on targets for future research will be made.

4.7 Prioritisation of research questions

To adhere to the ten-week timeline for this EVA, the research questions will be prioritised and addressed in the following order. Questions will be addressed one at a time, with subsequent questions

addressed only if time allows. The order of the questions was determined in collaboration with NICE, following advice from members of the Assessment Subgroup (ASG). The ASG highlighted that test accuracy (especially false positive rates) and time to diagnosis were the most clinically important questions at this time.

A. Symptomatic population

1. Key question 1. Test accuracy/test failure to detect lung cancer (lung cancer)
2. Key question 2. Practical implications (lung cancer)
3. Key question 3. Clinical effectiveness (lung cancer)
4. Key question 1, subquestion 1a (nodules)
5. Key question 1, subquestion 1b (nodules). Concordance will only be assessed in the absence of test accuracy data.

B. Incidental population

6. Key question 1. Test accuracy/test failure to detect lung cancer (lung cancer)
7. Key question 2. Practical implications (lung cancer)
8. Key question 3. Clinical effectiveness (lung cancer)
9. Key question 1, subquestion 1a (nodules)
10. Key question 1, subquestion 1b (nodules). Concordance will only be assessed in the absence of test accuracy data.

6. Methods for developing a conceptual cost-effectiveness model

This section describes the process and rationale for the development of a conceptual[15] decision analytic model to inform potential full cost-effectiveness evaluation of AI software for analysing CXR images to identify suspected lung cancer.

Development of a conceptual decision analytical model is intrinsically linked to the current clinical pathway for detection and management of lung cancer and the positioning of AI software within this pathway. Key points throughout the clinical pathway (for adults referred for CXR from primary care) will be identified with reference to Figure 1 in the final NICE scope for this topic,[10] existing guidelines on the diagnostic and care pathway,[8, 9, 16, 17] and close collaboration with clinical experts.

By necessity, this will be an iterative process to achieve a model structure which is pragmatic in its representation of the complex clinical pathways that adults accessing primary care may follow to arrive at a diagnosis of lung cancer.

Once diagnosis is achieved, evidence linkage between intermediate outcomes and long-term outcomes is required to assess cost-effectiveness over a lifetime horizon.

A pragmatic searching approach of the literature will be used to identify existing methods of cost-effectiveness modelling for AI software in CXR and inform parameterisation of a conceptual model. This is intended to support the design and development of the conceptual economic model, but not to perform a full systematic review or define evidence gaps comprehensively. Targeted searches of reviews of cost-effectiveness modelling which look more broadly at the diagnosis of lung cancer (for example both diagnostic and screening models) will likely be required.

Search strategies will be developed iteratively by the lead health economist and senior information specialist and results reviewed by the health economist. Studies identified in these targeted reviews will not be subject to a formal assessment, but will be discussed narratively, focussing on the methods used, assumptions made, availability of evidence to support evidence-linkage approaches and considerations for future modelling and research.

The result of this development process will be a framework on which to identify:

• Direct effects of the AI software
• Indirect effects due to modifying clinical decisions and the impact on further tests/treatments
• Effects on the timing of decisions and actions
• Influence on patient, radiology specialist, and clinician experience

These outcomes will be reported in terms of resource use, costs, health related utilities, and time, in line with the NICE reference case.[18]

The conceptual model will also provide insight into outcome measures which may prove value drivers in the future assessment of cost-effectiveness with the addition of AI software analysis of CXR compared to current standard practice of radiology specialist alone reporting CXR.

Given the time available, the diagnostic component of the model will be the primary focus of this report with priority given to the following considerations:

- Input parameters to populate model – to include consideration of type of evidence required, sources available, and gaps in the evidence.
- Relevant outcome measures to compare cost and clinical effectiveness of AI software in the detection of lung cancer.

- Identification of potential value drivers of model – to include how these can be measured for inclusion in a cost-effectiveness model.

As the outcome of this report is a conceptual model rather than a fully executable model, no data will be run, or results produced, to inform decision making on the cost-effectiveness of AI software for this indication. However, the potential budget impact of implementing AI software based on initial software installation, estimated unit costs per X-ray image analysed, and numbers of adults referred for CXR from primary care in both symptomatic and incidental populations, will be estimated. This will be based at an individual institution level if sufficient relevant data can be retrieved within project timescale. If not, national database sources will be used to estimate budget impact at a national level. Any further inferences cannot be made from this report.

To conclude, the conceptual modelling process will explore both the structure, and evidence requirements for parameter inputs for future model development. It will also facilitate the identification of cost outcomes, potential value drivers for AI software for this indication, and evidence linkage requirements for longer term outcomes where time permits. Costs associated with implementing AI software will also be considered. A pragmatic search strategy will be used to identify evidence from current literature to support these aims and is not intended as a systematic literature review or definitive summary of evidence gaps. This assessment will not include development of an executable cost effectiveness model.

6. Handling of information from manufacturers

All data submitted by the manufacturers/sponsors/stakeholders will only be considered if received by the External Assessment Group by the 6th January 2023. Data received 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.

8. Competing interests of authors and advisors

None of the authors have any competing interests.

9. Timetable/milestones

Milestone	Completion date
Draft protocol	21.11.2022
Final protocol	23.11.2022
Progress report	16.12.2023
Draft assessment report	26.01.2023
Final assessment report	16.02.2023

References

1. Cancer Research UK. *Lung cancer*. 2019. URL: https://www.cancerresearchuk.org/about-cancer/lung-cancer (Accessed 14 November 2022).

2. Clinical Knowledge Summaries. *Lung and pleural cancers - recognition and referral*. National Institute for Health and Care Excellence; 2021. URL: https://cks.nice.org.uk/topics/lung-pleural-cancers-recognition-referral/ (Accessed 14 November 2022).

3. NHS website. *Overview: Lung cancer*. 2022. URL: https://www.nhs.uk/conditions/lung-cancer/ (Accessed 14 November 2022).

4. Cancer Research UK. *Lung cancer incidence by age*. 2021. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence#heading-One (Accessed 23 November 2022).

5. NHS England. *NHS Long Term Plan*. 2021. URL: https://www.longtermplan.nhs.uk/ (Accessed 14 November 2022).

6. IBM Cloud Education. *Artificial Intelligence (AI)*. 2020. URL: https://www.ibm.com/uk-en/cloud/learn/what-is-artificial-intelligence (Accessed 14 November 2022).

7. Farhat H, Sakr GE, Kilany R. Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19. *Mach Vis Appl* 2020;**31**(6):53. http://dx.doi.org/10.1007/s00138-020-01101-5

8. National Institute for Health and Care Excellence. *Lung cancer: diagnosis and management: NICE guideline [NG122]*. 2022. URL: https://www.nice.org.uk/guidance/ng122 (Accessed 14 November 2022).

9. National Institute for Health and Care Excellence. *Suspected cancer: recognition and referral NICE guideline [NG12]*. 2021. URL: https://www.nice.org.uk/guidance/ng12 (Accessed 14 November 2022).

10. National Institute for Health and Care Excellence. *Artificial Intelligence software for analysing chest X-ray images to identify suspected lung cancer: Final scope*. 2022. URL: https://www.nice.org.uk/guidance/gid-dg10065/documents/final-scope (Accessed 23 November 2022).

11. National Institute for Health and Care Excellence, King's Technology Evaluation Centre. *Artificial intelligence for analysing chest X-ray images: Medtech innovation briefing [MIB292]*. National Institute for Health and Care Excellence; 2022. URL: https://www.nice.org.uk/advice/mib292 (Accessed 15 November 2022).

12. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D*, et al.* Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med* 2021;**4**(1):65. http://dx.doi.org/10.1038/s41746-021-00438-z

13. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R*, et al.* A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat Med* 2021;**27**(10):1663-5. http://dx.doi.org/10.1038/s41591-021-01517-0

14. The Joanna Briggs Institute. *JBI Critical appraisal tools*. 2021. URL: https://jbi.global/critical-appraisal-tools (Accessed 14 November 2022).

15. Tappenden P. *Conceptual modelling for health economic model development [HEDS Discussion Paper 12/05]*. White Rose Research Online; 2012. URL: https://eprints.whiterose.ac.uk/74464/1/HEDSDP1205.pdf (Accessed 28 November 2022).

16. 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

17. NHS England. *Updated National Optimal Lung Cancer Pathway – Sept 2020*. British Thoracic Oncology Group; 2020. URL: https://www.btog.org/news/just-released-updated-national-optimal-lung-cancer-pathway-sept-2020/ (Accessed 23 November 2022).

18. National Institute for Health and Care Excellence. *NICE health technology evaluations: the manual: Process and methods [PMG36]*. 2022. URL: https://www.nice.org.uk/process/pmg36/ (Accessed 23 November 2022).

APPENDIX 1

Search strategy: Multi-stranded, targeted approach

Search #	**Search**	**Number of hits in MEDLINE in testing**	**Sources**
1	Intervention (AI and chest x-ray) AND Study type ('Reviews (best balance of sensitivity and specificity)' Clinical Queries limit OR systematic reviews filter (specific filter))	360	Epistemonikos, MEDLINE, Embase, CDSR, a computer science database
2	Intervention [broader] (AI) AND lung cancer or lung nodule AND study type (systematic reviews filter (specific filter))	100	Epistemonikos, MEDLINE, Embase, CDSR, a computer science database
3	Intervention (AI and chest x-ray) AND selected outcomes (lung cancer / lung nodule)	709	MEDLINE, Embase, CENTRAL (inc. trial register records), a computer science database
4	Technology names / companies [look in title, abstract and institution fields] AND (chest x-ray / lung cancer / lung nodule)	137	MEDLINE, Embase, CENTRAL (inc. trial register records), a computer science database
	Check references of relevant reviews and studies found via NICE and team members' scoping or clinical experts	n/a	NICE, team members, clinical experts
	Total	1,284	
	Estimated Total (all databases, after deduplication)	2,000-2,500	

Search # 1.

Ovid MEDLINE(R) ALL <1946 to November 21, 2022>

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/ 161241

2 ai.kf,tw. 39973

3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 124610

4 exp Neural Networks, Computer/ 54027

5 (neural network* or convolutional or CNN or CNNs).kf,tw. 90544

6 exp Diagnosis, Computer-Assisted/ 86388

7 Pattern Recognition, Automated/ 26365

8 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 33598

9 (support vector machine* or random forest* or black box learning).kf,tw. 37695

10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 397256

11 exp Radiography, Thoracic/ 40533

12 X-Rays/ 31127

13 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 66473

14 11 or 12 or 13 121779

15 10 and 14 3872

16 limit 15 to "reviews (best balance of sensitivity and specificity)" 349

17 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 288222

18 (systematic* adj3 (review* or overview* or search or literature)).mp. 328853

19 17 or 18 459889

20 15 and 19 41

21 16 or 20 360

Search #2.

Ovid MEDLINE(R) ALL <1946 to November 21, 2022>

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/ 161241

2 ai.kf,tw. 39973

3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 124610

4 exp Neural Networks, Computer/ 54027

5 (neural network* or convolutional or CNN or CNNs).kf,tw. 90544

6 exp Diagnosis, Computer-Assisted/ 86388

7 Pattern Recognition, Automated/ 26365

8 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 33598

9 (support vector machine* or random forest* or black box learning).kf,tw. 37695

10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 397256

11 exp Lung Neoplasms/ or Solitary Pulmonary Nodule/ 268463

12 ((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 326597

13 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 946

14 (sclc or nsclc).kf,tw. 64513

15 11 or 12 or 13 or 14 398393

16 10 and 15 6758

17 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 288222

18 (systematic* adj3 (review* or overview* or search or literature)).mp. 328853

19 17 or 18 459889

20 16 and 19 100

Search #3.

Ovid MEDLINE(R) ALL <1946 to November 21, 2022>

1 exp Lung Neoplasms/ or Solitary Pulmonary Nodule/ 268463

2 ((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 326597

3 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 946

4 (sclc or nsclc).kf,tw. 64513

5 1 or 2 or 3 or 4 398393

6 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/ 161241

7 ai.kf,tw. 39973

8 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 124610

9 exp Neural Networks, Computer/ 54027

10 (neural network* or convolutional or CNN or CNNs).kf,tw. 90544

11 exp Diagnosis, Computer-Assisted/ 86388

12 Pattern Recognition, Automated/ 26365

13 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 33598

14 (support vector machine* or random forest* or black box learning).kf,tw. 37695

15 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 397256

16 exp Radiography, Thoracic/ 40533

17 X-Rays/ 31127

18 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 66473

19 16 or 17 or 18 121779

20 5 and 15 and 19 709

Search #4.

Ovid MEDLINE(R) ALL <1946 to November 21, 2022>

1 AI-Rad Companion Chest X-ray*.kf,tw,in. 1

2 Annalise CXR*.kf,tw,in. 1

3 Auto Lung Nodule Detection*.kf,tw,in. 0

4 ChestView*.kf,tw,in. 0

5 (Chest X-Ray Classifier* or Quibim*).kf,tw,in. 46

6 CheXVision*.kf,tw,in. 0

7 (ClearRead Xray* adj2 Detect).kf,tw,in. 0

8 InferRead DR Chest*.kf,tw,in. 0

9 JLD-02K*.kf,tw,in. 0

10 Lunit INSIGHT CXR*.kf,tw,in. 4

11 Milvue Suite*.kf,tw,in. 0

12 ChestEye Quality*.kf,tw,in. 0

13 (qXR* or Qure*).kf,tw,in. 6824

14 (red dot* or behold*).kf,tw,in. 1091

15 SenseCare-Chest DR Pro*.kf,tw,in. 0

16 VUNO Med-Chest X-Ray*.kf,tw,in. 0

17 (X1* and Visionairy Health).kf,tw,in. 0

18 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 7966

19 exp Lung Neoplasms/ or Solitary Pulmonary Nodule/ 268463

20 ((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 326597

21 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 946

22 (sclc or nsclc).kf,tw. 64513

23 19 or 20 or 21 or 22 398393

24 exp Radiography, Thoracic/ 40533

25 X-Rays/ 31127

26 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 66473

27 24 or 25 or 26 121779

28 18 and 23 91

29 18 and 27 61

30 28 or 29 137