



Health Technology Assessment

Volume 28 • Issue 50 • August 2024

ISSN 2046-4924

Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer: an evidence synthesis early value assessment

Jill Colquitt, Mary Jordan, Rachel Court, Emma Loveman, Janette Parr, Iman Ghosh, Peter Auguste, Mubarak Patel and Chris Stinton



Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer: an evidence synthesis early value assessment

Jill Colquitt^{1,†} , Mary Jordan^{2,†} , Rachel Court² ,
Emma Loveman² , Janette Parr² , Iman Ghosh² ,
Peter Auguste² , Mubarak Patel²  and Chris Stinton^{2*} 

¹Effective Evidence, Waterlooville, UK

²Warwick Medical School, University of Warwick, Coventry, UK

*Corresponding author

†Joint lead authors

Published August 2024

DOI: 10.3310/LKRT4721

This report should be referenced as follows:

Colquitt J, Jordan M, Court R, Loveman E, Parr J, Ghosh I, *et al.* Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer: an evidence synthesis early value assessment. *Health Technol Assess* 2024;**28**(50). <https://doi.org/10.3310/LKRT4721>

ISSN 2046-4924 (Online)

Impact factor: 3.6

A list of Journals Library editors can be found on the [NIHR Journals Library website](#)

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 3.6 and is ranked 32nd (out of 105 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), Embase (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nhr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nhr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR135755. The protocol was agreed in November 2022. The draft manuscript began editorial review in February 2023 and was accepted for publication in December 2023. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Copyright © 2024 Colquitt *et al.* This work was produced by Colquitt *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nhr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer: an evidence synthesis early value assessment

Jill Colquitt^{1,†}, Mary Jordan^{2,†}, Rachel Court², Emma Loveman²,
Janette Parr², Iman Ghosh², Peter Auguste², Mubarak Patel² and
Chris Stinton^{2*}

¹Effective Evidence, Waterlooville, UK

²Warwick Medical School, University of Warwick, Coventry, UK

*Corresponding author c.stinton@warwick.ac.uk

†Joint lead authors

Background: Lung cancer is one of the most common types of cancer in the United Kingdom. It is often diagnosed late. The 5-year survival rate for lung cancer is below 10%. Early diagnosis may improve survival. Software that has an artificial intelligence-developed algorithm might be useful in assisting with the identification of suspected lung cancer.

Objectives: This review sought to identify evidence on adjunct artificial intelligence software for analysing chest X-rays for suspected lung cancer, and to develop a conceptual cost-effectiveness model to inform discussion of what would be required to develop a fully executable cost-effectiveness model for future economic evaluation.

Data sources: The data sources were MEDLINE All, EMBASE, Cochrane Database of Systematic Reviews, Cochrane CENTRAL, Epistemonikos, ACM Digital Library, World Health Organization International Clinical Trials Registry Platform, clinical experts, Tufts Cost-Effectiveness Analysis Registry, company submissions and clinical experts. Searches were conducted from 25 November 2022 to 18 January 2023.

Methods: Rapid evidence synthesis methods were employed. Data from companies were scrutinised. The eligibility criteria were (1) primary care populations referred for chest X-ray due to symptoms suggestive of lung cancer or reasons unrelated to lung cancer; (2) study designs that compared radiology specialist assessing chest X-ray with adjunct artificial intelligence software versus radiology specialists alone and (3) outcomes relating to test accuracy, practical implications of using artificial intelligence software and patient-related outcomes. A conceptual decision-analytic model was developed to inform a potential full cost-effectiveness evaluation of adjunct artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer.

Results: None of the studies identified in the searches or submitted by the companies met the inclusion criteria of the review. Contextual information from six studies that did not meet the inclusion criteria provided some evidence that sensitivity for lung cancer detection (but not nodule detection) might be higher when chest X-rays are interpreted by radiology specialists in combination with artificial intelligence software than when they are interpreted by radiology specialists alone. No significant differences were observed for specificity, positive predictive value or number of cancers detected. None of the six studies provided evidence on the clinical effectiveness of adjunct artificial intelligence software. The conceptual model highlighted a paucity of input data along the course of the diagnostic pathway and identified key assumptions required for evidence linkage.

Limitations: This review employed rapid evidence synthesis methods. This included only one reviewer conducting all elements of the review, and targeted searches that were conducted in English only. No eligible studies were identified.

Conclusions: There is currently no evidence applicable to this review on the use of adjunct artificial intelligence software for the detection of suspected lung cancer on chest X-ray in either people referred from primary care with symptoms of lung cancer or people referred from primary care for other reasons.

Future work: Future research is required to understand the accuracy of adjunct artificial intelligence software to detect lung nodules and cancers, as well as its impact on clinical decision-making and patient outcomes. Research generating key input parameters for the conceptual model will enable refinement of the model structure, and conversion to a full working model, to analyse the cost-effectiveness of artificial intelligence software for this indication.

Study registration: This study is registered as PROSPERO CRD42023384164.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR135755) and is published in full in *Health Technology Assessment*; Vol. 28, No. 50. See the NIHR Funding and Awards website for further award information.

Contents

List of tables	xi
List of figures	xiii
List of supplementary material	xv
Glossary	xvii
List of abbreviations	xix
Plain language summary	xxi
Scientific summary	xxiii
Chapter 1 Introduction	1
Purpose of the decision to be made	1
Population	1
Condition	1
Technologies under assessment	2
Comparators	2
Reference standards	2
Care pathway	2
Chapter 2 Decision questions and objectives	5
Chapter 3 Methods	7
Methods for assessing test accuracy, practical implications and clinical effectiveness	7
<i>Search strategy</i>	7
<i>Company submissions</i>	7
<i>Eligibility criteria</i>	8
<i>Review strategy</i>	8
<i>Data extraction</i>	8
<i>Risk of bias</i>	8
<i>Analysis and synthesis</i>	8
<i>Post hoc methods</i>	10
Methods for developing a conceptual cost-effectiveness model	10
<i>Literature review</i>	10
<i>Clinical guidelines</i>	10
<i>Company and clinical expert involvement</i>	11
<i>Methods to assess potential budget impact</i>	11
Chapter 4 Results: test accuracy, practical implications and clinical effectiveness	13
Results of literature searches	13
Study characteristics and key biases of selected excluded studies	13
What are the test accuracy and test failure rates of adjunct artificial intelligence software to detect lung cancer on chest X-rays?	18

What are the practical implications of adjunct artificial intelligence software to detect lung cancer on chest X-rays?	21
What is the clinical effectiveness of adjunct artificial intelligence software applied to chest X-rays?	21
Summary of ongoing trials	21
Chapter 5 Cost-effectiveness	23
Results of literature searches	23
Description of the evidence	23
<i>Bajre et al.</i>	23
<i>Foley et al.</i>	24
<i>Bradley et al.</i>	25
<i>Woznitza et al. (2018)</i>	27
<i>Woznitza et al. (2022)</i>	27
<i>Dwyer-Hemmings and Fairhead</i>	27
Clinical pathway for representation in model	28
<i>Pathway A</i>	30
<i>Pathways B and C</i>	30
<i>Pathways D and E</i>	31
Discussion of inputs to inform model structure	31
<i>Intermediate measures for consideration</i>	31
<i>Summary</i>	37
<i>Question 4</i>	37
<i>Strategies</i>	37
<i>Proposed model structure</i>	38
<i>An example using clinical pathway A</i>	39
<i>Question 5</i>	40
<i>Cost of software</i>	40
<i>Cost of staff training</i>	42
<i>Cost of staff time to read and report chest X-ray</i>	42
<i>Cost of further diagnostic tests</i>	42
<i>Cost of treatment (including costs of any adverse events)</i>	44
<i>Summary</i>	44
Results of potential budget impact assessment	45
<i>Summary</i>	46
Chapter 6 Discussion	49
Statement of principal findings	49
<i>Test accuracy, practical implications and clinical effectiveness</i>	49
<i>Conceptual cost-effectiveness modelling</i>	49
Strengths and limitations	50
<i>Strengths</i>	50
<i>Limitations of the review/early value assessment process</i>	50
<i>Limitations of evidence base</i>	51
Uncertainties	51
Equality, diversity and inclusion	51
Patient and public involvement	52

Chapter 7 Conclusions	53
Test accuracy, practical implications and clinical effectiveness	53
Cost-effectiveness	53
Implications for service provision	53
Suggested research priorities	53
Additional information	55
References	57
Appendix 1 Literature searches	63

List of tables

TABLE 1 Eligibility criteria	8
TABLE 2 Characteristics of summarised (but ineligible) studies	15
TABLE 3 Test accuracy results from summarised (but ineligible) studies	17
TABLE 4 Ongoing studies	22
TABLE 5 Outcome data reported by code	25
TABLE 6 Key outcome data reported by individual codes	26
TABLE 7 Outcome data reported by immediate and standard CXR reporting arms	28
TABLE 8 Input parameters required to populate conceptual model	41
TABLE 9 Costs of AI software by company based on a number of 25,000 CXR images per NHS trust	43
TABLE 10 Costs required for the proposed model	45
TABLE 11 Anticipated budget impact of AI software at NHS trust level for all GP-referred CXR	47
TABLE 12 Anticipated budget impact of AI software at NHS trust level for symptomatic, incidental and whole-population GP-referred CXR	48
TABLE 13 Search strategy summary: multistranded, targeted approach	63
TABLE 14 Bibliographic databases: search summary	64

List of figures

FIGURE 1 Care pathway for the recognition and referral of suspected lung cancer	3
FIGURE 2 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram	14
FIGURE 3 Forest plot of sensitivity	19
FIGURE 4 Forest plot of specificity	20
FIGURE 5 The CXR pathway for people referred from primary care, from the NICE final scope	29
FIGURE 6 Clinical pathways for conceptual model for people referred from primary care	30
FIGURE 7 Illustrative model structure for the detection of lung cancer	38

List of supplementary material

Report Supplementary Material 1 Detailed table of characteristics of summarised (but ineligible) studies

Report Supplementary Material 2 Table of studies excluded at full-text assessment

Supplementary material can be found on the NIHR Journals Library report page (<https://doi.org/10.3310/LKRT4721>).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

Glossary

Artificial intelligence The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.

Deep learning A method in artificial intelligence that teaches computers to process data in a way that is inspired by the human brain. Deep-learning models can recognise complex patterns in pictures, text, sounds and other data to produce accurate insights and predictions.

False-negative value The number of cases in which the index test has wrongly suggested the patient as being disease-free when they do have the disease. FN = c .

False-positive value The number of cases in which the index test has wrongly indicated the patient as having the disease when they do not have the disease. FP = b .

Ground truth The actual nature of the problem that is the target of a machine learning model, reflected by the relevant data sets associated with the use case in question.

Machine learning In artificial intelligence (a subject in computer science), a discipline concerned with the implementation of computer software that can learn autonomously.

Reference standard The test, combination of tests or procedure that is considered the best available method of categorising participants in a study of diagnostic test accuracy as having or not having a target condition.

Sensitivity The proportion of people who test positive for a disease among people who have the disease of interest. The ratio between the true-positive value and (true-positive value + false-negative value).

$$\text{Sensitivity} = \frac{a}{a + c} = \frac{TP}{TP + FN} \quad (1)$$

Specificity The proportion of people who test negative for a disease among people who do not have the disease of interest. The ratio between the true-negative value and (true-negative value + false-positive value).

$$\text{Specificity} = \frac{d}{b + d} = \frac{TN}{TN + FP} \quad (2)$$

Survival rate The percentage of people in a study or treatment group who are still alive for a certain period of time after they were diagnosed with or started treatment for a disease such as cancer.

True-negative value The number of cases in which the index test has correctly indicated the patient as being disease-free. TN = d .

True-positive value The number of cases in which the index test has correctly indicated the patient as having the disease. TP = a .

List of abbreviations

A&E	accident and emergency	MDT	multidisciplinary team
AI	artificial intelligence	NICE	National Institute for Health and Care Excellence
BIA	budget impact analysis	PET-CT	positron emission tomography and computed tomography
CT	computed tomography	PSA	probabilistic sensitivity analysis
CXR	chest X-ray	PSS	Personal Social Services
DAP	Diagnostics Assessment Programme	QALY	quality-adjusted life-year
EAG	External Assessment Group	QoL	quality of life
EDI	equality, diversity and inclusion	RCR	Royal College of Radiologists
EVA	early value assessment	SCM	specialist committee member
GP	general practitioner	SR	standard radiographer
HRQoL	health-related quality of life	TAT	turnaround time
ICER	incremental cost-effectiveness ratio	WHO ICTRP	World Health Organization International Clinical Trials Registry Platform
IR	immediate radiographer		

Note

This monograph is based on the Diagnostic Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed confidential. The full report was used by the Diagnostic Advisory Committee at NICE in their deliberations. The full report with each piece of confidential data removed and replaced by the statement 'confidential information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain language summary

Lung cancer is one of the most common types of cancer in the United Kingdom. Early diagnosis may improve survival, as lung cancer is often diagnosed late. Chest X-rays can be used to identify features of lung cancer. There can be delays in getting X-rays, and sometimes features of lung cancer are not seen on them. Artificial intelligence software may help by finding features of cancer on chest X-rays and highlighting them. A radiologist will look at the X-rays and information from the software. There is a lack of information about how lung cancer diagnosis could change if artificial intelligence software is used and what the costs may be to the National Health Service.

This project looked at the use of artificial intelligence software in the detection of lung cancer in people referred from primary care. Software companies were invited to provide evidence. There were no studies that looked at this topic among people from primary care. We summarised the closest evidence we could find instead. All of this had flaws, so we could not tell if the results were accurate or helpful to this review. It was not clear if artificial intelligence helped to find cancers or improve people's health.

We made a theoretical model to discuss the best way to assess if artificial intelligence software might be cost-effective in detecting lung cancer and what evidence would be needed to do this in a fully working model. Costs and alternative pricing models provided by five companies were used to calculate the cost of adding artificial intelligence software to review chest X-rays in people referred from their general practitioner, for the first 5 years, based on one National Health Service trust.

Future studies are needed to identify the impact of adjunct artificial intelligence on test accuracy, clinical decision-making and patient outcomes (e.g. mortality and morbidity).

Scientific summary

Background

Lung cancer occurs when abnormal cells multiply in an uncontrolled way to form a tumour in the lung. It is one of the most common types of cancer in the UK, and each year over 43,000 new cases are diagnosed. In the early stages of the disease, people usually do not have symptoms, which means that lung cancer is often diagnosed late. The 5-year survival rate for lung cancer is low, at below 10%. Early diagnosis may improve survival. The National Institute for Health and Care Excellence (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. AI combines computer science and data sets to enable problem-solving. Machine learning and deep learning are subfields of AI. They comprise AI algorithms that seek to create expert systems to make predictions or classifications based on data input.

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 on chest X-rays (CXRs). AI technologies subject to this assessment are standalone software platforms developed with deep-learning algorithms to interpret CXRs. 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 or hilar enlargement, or a combination of these, is identified. A technology may classify CXRs into those with and without a nodule, or it may identify several different abnormalities or lung diseases.

Objectives

The overall aim of this early value assessment (EVA) is to identify evidence on adjunct AI software for analysing CXRs for suspected lung cancer and identify evidence gaps to help direct data collection and further research. A conceptual modelling process was undertaken to inform discussion of what would be required to develop a fully executable cost-effectiveness model for future economic evaluation. 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 about whether AI software can be implemented in clinical practice at the present time.

There are two populations of interest in this EVA: (1) people referred from primary care for a CXR because they have symptoms suggestive of lung cancer (symptomatic population) and (2) people referred from primary care for a CXR for reasons unrelated to lung cancer (incidental population). Based on the scope produced by NICE, we defined the following questions to inform future assessment on the benefits, harms and costs of adjunct AI for analysing on CXRs for suspected lung cancer compared with human reader alone in these populations:

1. What is the test accuracy and test failure rate of adjunct AI software to detect lung cancer on CXRs?
2. What are the practical implications of adjunct AI to detect lung cancer on CXRs?
3. What is the clinical effectiveness of adjunct AI software applied to CXRs?
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?

Methods

Data sources

MEDLINE All (via Ovid), EMBASE (via Ovid), Cochrane Database of Systematic Reviews (via Wiley), Cochrane CENTRAL (via Wiley), Epistemonikos, ACM Digital Library, World Health Organization International Clinical Trials Registry Platform, clinical experts, and company submissions.

Eligibility criteria

Population: people referred for CXR from primary care because they have symptoms suggestive of lung cancer, and people referred for CXR from primary care for reasons unrelated to lung cancer.

Intervention: radiology specialist with adjunct AI

Comparator: radiology specialist without adjunct AI

Outcomes: test accuracy, patient management, clinical effectiveness.

Study selection, data extraction and assessment of risks of bias

Titles and abstracts of all identified records were screened by one reviewer against the review eligibility criteria, with a random 20% screened by a second reviewer. Full texts of records considered potentially relevant by either reviewer were retrieved and assessed for inclusion by one reviewer. A random 20% sample were assessed independently by a second reviewer, with any disagreements resolved by consensus or discussion with a third reviewer. We planned to extract data into a piloted form, assess risk of bias and synthesise data using methods described in the research protocol; however, no studies met the inclusion criteria. Post hoc methods were determined following discussions with NICE to select and summarise the closest available evidence to the review inclusion criteria. Studies were selected that assessed eligible AI software in conjunction with radiologists compared with radiologists alone but in which the referral status and symptomatic status of the population was unclear. Data were extracted by one reviewer, with a random 20% checked by a second reviewer. Results were summarised narratively, and key biases were noted.

Data synthesis

A narrative data synthesis was performed.

Modelling

The conceptual modelling process explored both the structure and evidence requirements for parameter inputs, for future model development. An iterative approach was taken to facilitate the identification of cost outcomes, potential value drivers of AI software for this indication and evidence linkage requirements for longer-term outcomes, where time allowed. Costs associated with implementing AI software were also considered.

Information to inform the conceptual model was obtained from a variety of sources including a literature review, current clinical guidelines, discussion with specialist clinical experts and the companies submitting evidence on AI software.

Given the time available, the diagnostic component of the model was the primary focus of the health economics aspect of the report. Priority was given to the following considerations:

- input parameters to populate the model – including consideration of the type of evidence required, sources available and gaps in the evidence
- relevant outcome measures to compare the cost-effectiveness and clinical effectiveness of AI software in the detection of lung cancer
- identification of potential value drivers of the model – with recommendations of how these can be measured for inclusion in a cost-effectiveness model.

Results

Test accuracy, practical implications and clinical effectiveness

No studies met the inclusion criteria of the review. Two ongoing studies with unclear eligibility were identified. In the absence of available evidence, we summarised data from six studies that had unclear populations but included a comparison of CXRs read by readers with and without the use of commercial AI software.

Statistical comparisons were not undertaken in most of the studies, but there was some evidence that sensitivity might be higher among specialist radiologist with AI than among specialist radiologist without AI. This finding was not consistent between studies, however. No significant differences were observed for specificity, positive predictive value or number of cancers detected. None of the studies provided evidence on the clinical effectiveness of adjunct AI software. The summarised studies were small retrospective studies with important methodological limitations, and their generalisability to the UK population is unclear.

Conclusions

There is currently no evidence applicable to this review on the use of adjunct AI software for the detection of suspected lung cancer on CXRs in either people referred from primary care with symptoms of lung cancer or people referred from primary care for other reasons.

Implications for service provision

Lung cancer pathways are complex and contain many routes to diagnosis. Although national guidance and timelines for diagnosis exist, practice varies widely throughout radiology departments and lung cancer teams both within and across NHS trusts. With many ways to achieve these targets, changes in any area of the diagnostic pathway may have a significant impact elsewhere.

There is some evidence for the impact of CXR results on the diagnostic pathway when performed without AI assistance, as is current practice. This is limited, and it is difficult to compare results because of the different study designs used and the different outcomes reported. There is no published evidence to link measures of progression through the diagnostic pathways with long-term outcomes such as stage at diagnosis and survival.

There is currently no applicable evidence to show the impact of the addition of AI software to CXR review on the diagnosis of lung cancer. There may be multiple ways AI software could change measures along this pathway. These could include improved accuracy of lung cancer detection, directing patients along the quickest pathway to diagnosis, quicker report turnaround time to achieve earlier confirmatory testing, or prioritisation of cases for review including those without lung cancer who can be discharged more quickly and free up staff time and resources. AI software may also have a negative impact on pathways by increasing the number of benign lung nodules detected and increasing the number of patients who undergo a computed tomography scan that they might not have needed. This would be detrimental to the patients, with increased exposure to radiation and anxiety due to a positive CXR result, and has cost and resource use implications for the department.

With a lack of evidence on AI software, the impact on service provision is unknown and may have significant implications in terms of progression through diagnostic pathways, resource use, costs and patient outcomes.

Study registration

This study is registered as PROSPERO CRD42023384164.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR135755) and is published in full in *Health Technology Assessment*; Vol. 28, No. 50. See the NIHR Funding and Awards website for further award information.

Chapter 1 Introduction

Sections of this chapter have been reproduced from the protocol,¹ available from the National Institute for Health and Care Excellence (NICE) website. © NICE 2022 *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*. Available from www.nice.org.uk/guidance/hte12/documents/final-protocol All rights reserved. Subject to Notice of rights.

National Institute for Health and Care Excellence guidance is prepared for the NHS in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication.

Purpose of the decision to be made

Lung cancer is one of the most common types of cancer in the UK.² In the early stages of the disease, people usually do not have symptoms, which means that lung cancer is often diagnosed late.³ The 5-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 (CXRs) 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 fully executable cost-effectiveness model for future economic evaluation.

Population

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

Condition

Approximately 43,000 new cases of lung cancer are diagnosed annually in the UK.² The incidence of lung cancer is highest among older people.⁴ It is rare in people aged < 40 years, and > 40% of people diagnosed with lung cancer are aged ≥ 75 years.³

Lung cancer occurs when abnormal cells multiply in an uncontrolled way to form a tumour in the lung.⁵ 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 more 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 > 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 that lung cancer is often diagnosed late. In 2018, > 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 2028.⁶

Technologies under assessment

Artificial intelligence combines computer science and data sets to enable problem-solving. Machine learning and deep learning are subfields of AI. They comprise AI algorithms that seek to create expert systems to make predictions or classifications based on data input.⁷ Many paradigms of deep learning have been developed, but the most used of these is the convolutional neural network.⁸

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 CXRs. 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 or hilar enlargement, or a combination of these, is identified. A technology may classify CXRs into those with and those without a nodule, or it may identify several different abnormalities or lung diseases. Fourteen companies producing AI software for analysing CXR images were included in the NICE scope.⁹

Comparators

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

Reference standards

Following CXR, people with suspected lung cancer should be offered a contrast-enhanced chest computed tomography (CT) scan to diagnose 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.¹⁰ If the CT scan indicates that there may be cancer, the type and sequence of investigations may vary but typically include a positron emission tomography and computed tomography (PET-CT) scan and an image-guided biopsy. Other methods that may be used include magnetic resonance imaging, endobronchial ultrasound-guided transbronchial needle aspiration, and endoscopic ultrasound-guided fine-needle aspiration.¹⁰ The PET-CT scan can show where there are active cancer cells, which can help with diagnosis and choosing the best treatment.³

Care pathway

Figure 1 depicts the care pathway for the recognition and referral of suspected lung cancer as outlined in NICE Guideline NG12.¹¹ The identification of people with signs and symptoms suggestive of lung

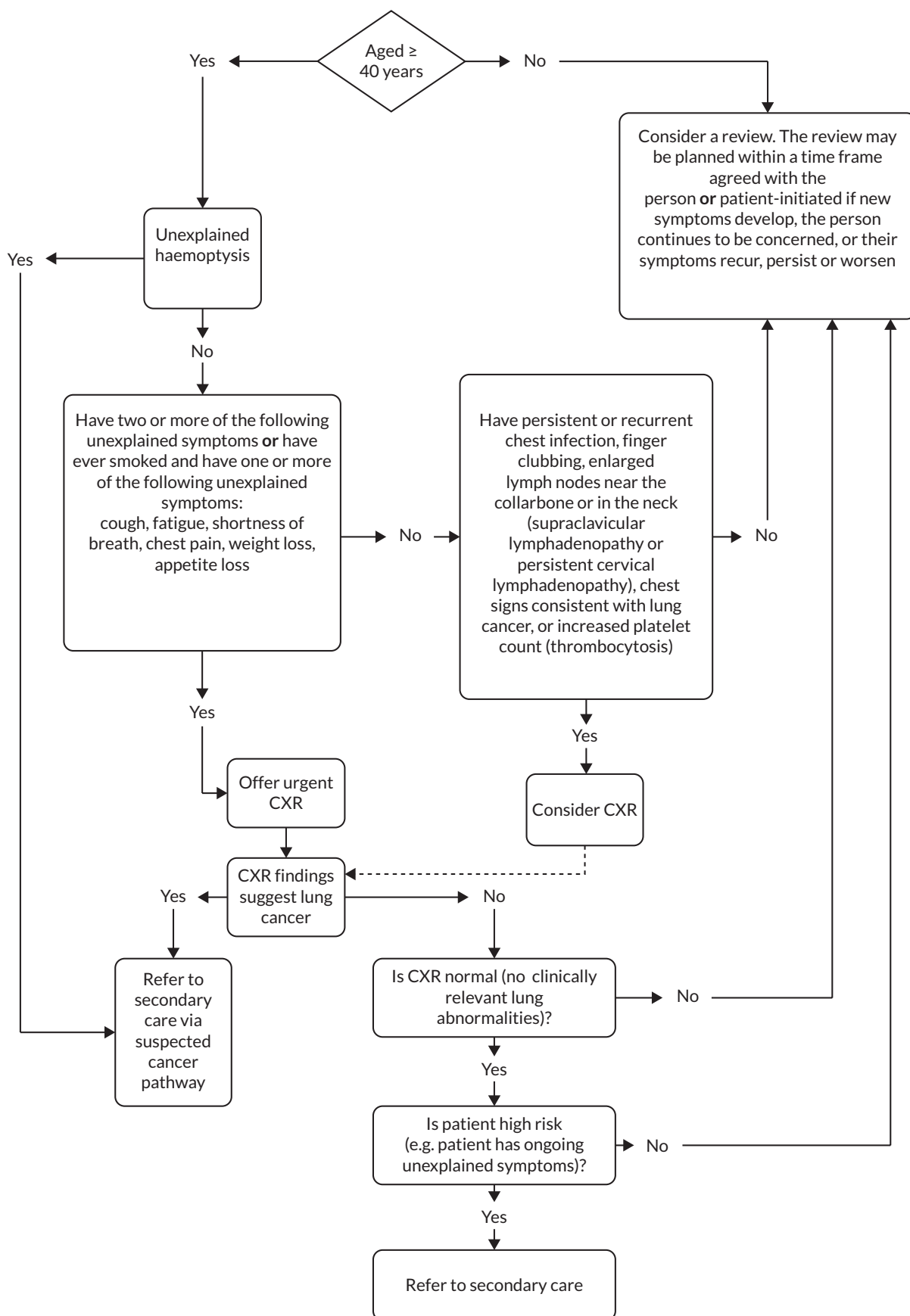


FIGURE 1 Care pathway for the recognition and referral of suspected lung cancer.

cancer often happens in primary care. The NICE guideline on recognition and referral for suspected lung cancer recommends that people aged ≥ 40 years be 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 and appetite loss.¹¹ An urgent CXR should also be considered for people aged ≥ 40 years if they have a 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 an 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, that is 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 CXRs of patients who are referred for CXR from primary care. Referrals for CXR outside primary care are beyond the scope of this project.

Chapter 2 Decision questions and objectives

Sections of this chapter have been reproduced from the protocol,¹ available from the NICE website. © NICE 2022 *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*. Available from www.nice.org.uk/guidance/hte12/documents/final-protocol All rights reserved. Subject to Notice of rights.

National Institute for Health and Care Excellence guidance is prepared for the NHS in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication.

The overall aim of this project was to identify evidence on adjunct AI software for analysing CXRs for suspected lung cancer, and identify evidence gaps to help direct data collection and further research. A conceptual modelling process was undertaken to inform discussion of what would be required to develop a fully executable cost-effectiveness model for future economic evaluation. The available evidence base was examined via an EVA. The assessment was 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 about whether AI software can be implemented in clinical practice at the present time.

Based on the scope produced by NICE,⁹ we defined the following questions to inform future assessment of the benefits, harms and costs of adjunct AI for analysing CXRs for suspected lung cancer compared with human reader alone:

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

Chapter 3 Methods

This report contains reference to confidential information provided as part of the NICE Diagnostic Assessment Process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Sections of this chapter have been reproduced from the protocol,¹ available from the NICE website.

The review is registered on PROSPERO (registration number CRD42023384164), and the protocol is available from the NICE website (www.nice.org.uk/guidance/hte12/history).

The timeline to produce this EVA report was 10 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 were made in collaboration with NICE and clinical experts.

Methods for assessing test accuracy, practical implications and clinical effectiveness

Search strategy

An iterative approach was taken to develop the search strategy, making use of relevant records identified during initial scoping searches and from relevant reviews.^{12,13} The strategy was 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 were favoured over a single search to retrieve a manageable number of records to screen ([Appendix 1](#)). Searches were run in a range of relevant bibliographic databases covering the fields of medicine and computer science. Searches were limited to studies published in English because studies published in other languages were likely to be difficult to assess in the timescale of this EVA. Non-human studies, letters, editorials, communications and conference abstracts were removed during the searches. No date limit was applied to the searches, but only records published in or after 2012 were screened. Database search strings were developed for MEDLINE and appropriately translated for each of the other databases, considering differences in thesaurus terms and syntax. The following bibliographic databases were searched: MEDLINE All (via Ovid), EMBASE (via Ovid), Cochrane Database of Systematic Reviews (via Wiley), Cochrane CENTRAL (via Wiley), Epistemonikos and ACM Digital Library.

A search for ongoing trials was conducted in the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP). A search for ongoing systematic reviews was undertaken in the PROSPERO database.

The full record of searches is provided in [Appendix 1](#).

Records were exported into EndNote X9.3 (Clarivate Analytics, Philadelphia, PA, USA), where duplicates were systematically identified and removed. Reference lists of included studies and a selection of relevant reviews were checked. Experts and team members were consulted and encouraged to share relevant studies.

Company submissions

As part of the Diagnostics Assessment Programme (DAP) process, 14 companies producing AI software were identified by NICE and invited to participate in the EVA and to submit evidence.⁹ The External Assessment Group (EAG) assessed company submissions in exactly the same way as it assessed published evidence, and references lists were examined for relevant studies.

Eligibility criteria

The eligibility criteria for the test accuracy, practical implications and clinical effectiveness questions are presented in [Table 1](#).

Review strategy

Titles and abstracts of records identified by the searches were screened by one reviewer, with a random 20% assessed independently by a second reviewer. Records considered potentially relevant by either reviewer were retrieved for further assessment. Full-text articles were assessed against the full inclusion/exclusion criteria by one reviewer. A random 20% sample was assessed independently by a second reviewer. Disagreements were resolved by consensus or through discussion with a third reviewer. Records rejected at full-text stage (including reasons for exclusion) are reported in [Report Supplementary Material 2](#).

Data extraction

We planned to extract data into a piloted electronic data collection form. Data were to be extracted by one reviewer, with a random 20% checked by a second reviewer, and disagreements resolved by consensus or discussion with a third reviewer. However, no studies met the inclusion criteria.

Risk of bias

We planned to assess the risk of bias of included studies using tools appropriate to the study design, such as those produced by the Joanna Briggs Institute.¹⁴ Risk of bias was to be assessed by one reviewer, with a random 20% assessed by a second reviewer and disagreements resolved through consensus or discussion with a third reviewer. As no studies met the inclusion criteria, no formal risk-of-bias assessment was undertaken.

Analysis and synthesis

Methods of analysis and synthesis were described a priori in the research protocol.¹ However, no studies met the inclusion criteria, so no data synthesis was undertaken.

TABLE 1 Eligibility criteria

	Key question 1. What are the test accuracy and test failure rates of adjunct AI software to detect lung cancer on CXRs? Subquestions: 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 CXRs?^a	Key question 3. What is the clinical effectiveness of adjunct AI software applied to CXRs?
Population	Adults referred from primary care who are: <ol style="list-style-type: none"> undergoing CXR due to symptoms suggestive of lung cancer, for example 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) undergoing CXR for reasons unrelated to lung cancer (incidental population) Where data permit, subgroups will be considered based on: <ul style="list-style-type: none"> ethnicity age sex socioeconomic status 		
Target condition	Lung cancer		

TABLE 1 Eligibility criteria (continued)

	Key question 1. What are the test accuracy and test failure rates of adjunct AI software to detect lung cancer on CXRs? Subquestions: 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 CXRs?^a	Key question 3. What is the clinical effectiveness of adjunct AI software applied to CXRs?
Intervention	CXR interpreted by radiology specialist (e.g. radiologist or radiographer) in conjunction with the following AI software: AI-Rad Companion CXR (Siemens Healthineers), Annalise CXR (annalise.ai), Auto Lung Nodule Detection (Samsung), ChestLink Radiology Automation (Oxipit), ChestView (GLEAMER), CXR (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-CXR (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 NG122, ¹⁰ where biopsy is not applicable For accuracy of nodule detection: radiology specialist (single reader or consensus of more than one reader)	N/A	N/A
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 for any other reason) Characteristics of discordant cancers cases Test accuracy for the detection of lung nodules Concordance in lung nodule detection between radiology specialist with and without adjunct AI software	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
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, for example neurosurgery, transplant or plastic surgery patients, people in secure forensic mental health services. Studies where more than 10% of the sample does 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		
N/A, not applicable.			
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'.			

Post hoc methods

No studies meeting the inclusion criteria were identified. Following discussions with the NICE technical team for this project, we examined the list of excluded studies that were closest to the review inclusion criteria (see [Table 1](#)), that is:

- Interventions: CXRs interpreted by radiology specialist in conjunction with eligible AI software versus radiologists alone and/or reference standard.
- Population: no details provided on the referral status or symptom status [studies that had an explicitly excluded population, for example health screening, preoperative CXRs, inpatients, accident and emergency (A&E), were not selected].
- Outcomes: as defined in [Table 1](#).

Selected studies were tabulated using the approach described in [Review strategy](#) and key biases were noted. Results were summarised narratively.

Methods for developing a conceptual cost-effectiveness model

This section describes the process, methods and rationale for the development of a conceptual¹⁵ decision-analytic model to inform a potential full cost-effectiveness evaluation of adjunct AI software for analysing CXR images to identify suspected lung cancer.

The conceptual modelling process explored both the structure and evidence requirements for parameter inputs, for future model development. This was to facilitate the identification of cost outcomes, potential value drivers of AI software for this indication, and evidence linkage requirements for longer-term outcomes. Costs associated with implementing AI software were also considered.

Information to inform the conceptual model was obtained from a variety of sources including a literature review, current clinical guidelines, discussion with specialist clinical experts and the companies submitting AI software for assessment.

Literature review

A pragmatic search of the literature was used to identify existing methods of cost-effectiveness modelling for AI software in CXRs and inform parameterisation of the conceptual model. It was not intended as a substitute for a systematic literature review or to provide a definitive summary of evidence gaps. This will be required for any future development of an executable cost-effectiveness model.

Following initial scoping searches, we did not expect to find any full economic evaluations of AI software as an adjunct to radiology specialist review of CXRs, particularly in the primary care population. For this reason, a broad search strategy was used across two databases (MEDLINE and Tufts CEA), and broad screening criteria were applied. The primary inclusion criterion was 'lung cancer studies', but, following this, any study that could inform the structure or parameters of a conceptual model was identified at title/abstract level. Full-text assessment of these papers was used to refine screening criteria further into studies that satisfied (1) the primary care referral population, (2) those with specific intention of diagnosis or screening and (3) those most relevant to the UK setting. Reference lists of these studies and publication lists of authorship groups were also screened for any further potentially relevant papers. Studies identified in these targeted reviews were not subject to a formal assessment but discussed narratively. This focused on the methods used, assumptions made, availability of evidence to support evidence linkage approaches and considerations for future modelling and research.

Clinical guidelines

The structure of the decision-analytic model is intrinsically linked to current clinical pathways. Key points throughout the clinical pathway for the detection and management of lung cancer, and the

positioning of AI software within this pathway (for adults referred for CXR from primary care), were identified with reference to [Figure 1](#) in the final NICE scope for this topic,⁹ existing guidelines on the diagnostic and care pathway^{10,11,16,17} and close collaboration with clinical experts.

Company and clinical expert involvement

Information on the relevant AI technologies under review was obtained from company submissions, with requests for additional information sent to companies that registered as stakeholders (Annalise AI, Behold AI, Infervision, Lunit Inc. and Siemens Healthcare).

Using the information gathered from these sources, an iterative process was used to achieve a model structure that is pragmatic in its representation of the complex clinical pathways that adults from primary care populations may follow to arrive at a diagnosis of lung cancer.

Given the time available to conduct this EVA, the primary focus of this report was the diagnostic component of the model. Priority was given to the following:

- input parameters to populate the model – including consideration of the type of evidence required, sources available and gaps in the evidence
- relevant outcome measures to compare the cost-effectiveness and clinical effectiveness of AI software in the detection of lung cancer
- identification of potential value drivers of the model – with recommendations of how these can be measured for inclusion in a cost-effectiveness model.

Once diagnosis is achieved in the model, evidence linkage between intermediate outcomes and long-term outcomes is required to assess cost-effectiveness over a clinically appropriate time horizon. These mainly relate to the mapping of the disease state (i.e. lung cancer) and are not specific to the diagnostic technology being assessed (e.g. utilities, costs and effects of current treatments). Potential sources for the main longer-term outcomes were identified during the literature search, with a focus on those relevant to the UK setting and in line with the requirements of the NICE reference case.¹⁸ An overview is presented in this report as an example of current practices in modelling lung cancer.

Methods to assess potential budget impact

Estimates of the potential budget impact of introducing AI software as an adjunct to radiology specialist review of CXRs were calculated based on methods for a budget impact analysis (BIA) outlined in the NICE evidence standards framework for digital health technologies¹⁹ and International Society for Pharmacoeconomics and Outcomes Research Task Force recommendations.²⁰ These identify six key elements that require inputs for the modelling framework of a BIA:

1. size and characteristics of affected population
2. current intervention mix without the new intervention
3. costs of current intervention mix
4. new intervention mix with the new intervention
5. cost of the new intervention mix
6. use and cost of other health conditions and treatment-related healthcare services.²⁰

Given the limitation of time and scope, a fully comprehensive BIA was not attempted as this would have required data on any changes in resource use and associated cost. The intended outcome of this report was a conceptual model where no outcome data were run and no results were produced. Therefore, estimates on this element were not included. The aim was to approximate the budget impact at an individual institution level, with information sourced from the literature and supplemental information provided by representatives of the institution used as an example.

METHODS

Company submissions to NICE as part of the DAP request for information were screened for cost data. Clarifying questions were sent to all companies (whether or not costings were already submitted) to obtain more granular detail for the purpose of BIA.

Records retrieved from the broad cost search were screened at title/abstract level by one reviewer (MJ) to identify any studies that may have been applicable. These were then retrieved as full text and their suitability for use was assessed. Studies that yielded relevant information were retained, data extracted and authors contacted to obtain further context-specific information.

Chapter 4 Results: test accuracy, practical implications and clinical effectiveness

Results of literature searches

Six of the companies contacted by NICE agreed to participate in the EVA, and five of these provided evidence (Annalise AI, Behold AI, Infervision, Lunit Inc. and Siemens Healthineers). None of the clinical evidence in the company submissions met the review eligibility criteria. An examination of reference lists in the company submissions identified 22 references.

[Figure 2](#) shows the flow of studies through this review. Searches identified a total of 3149 records. Of these, 172 were identified as potentially relevant to the symptomatic population and 104 were identified as potentially relevant to the incidental population. Full texts were obtained and screened. None of the studies met the inclusion criteria specified in [Table 1](#). The eligibility of two ongoing studies was unclear, and these studies are summarised in [Summary of ongoing trials](#).

Reasons for exclusion are described in [Report Supplementary Material 2](#). Among the studies that were potentially relevant to the symptomatic population, the main reasons for exclusion were no eligible AI software or AI not used in conjunction with radiology specialist ($n = 119$), and population not referred from primary care ($n = 30$). Only one identified study was conducted in a population referred from primary care; however, the comparison was not relevant (AI software alone vs. radiologist alone).²¹ Among the studies potentially relevant to the incidental population, the main reasons for exclusion were no relevant outcome ($n = 45$) and no eligible AI software ($n = 28$).

As described in [Analysis and synthesis](#), to provide the closest available evidence to that required in [Table 1](#), we looked for excluded studies that (1) had eligible AI software and (2) compared radiology specialist in conjunction with AI software with radiology specialist alone, but where the referral status of the population was unclear. Studies that had an explicitly excluded population (e.g. a health screening population, preoperative CXRs, inpatients, A&E) remained excluded. Six such studies were identified ([Table 2](#)).

Study characteristics and key biases of selected excluded studies

Characteristics of the summarised studies are described in [Table 2](#). In brief, six studies were summarised²²⁻²⁶ (Siemens 2022, unpublished academic in confidence submission from Siemens Healthineers). The studies employed retrospective designs²²⁻²⁶ and (confidential information has been removed) (Siemens 2022). Four studies were published: two were provided by the companies and not peer reviewed, and one of these is a preprint²² and the other is ongoing (Siemens 2022). The studies were carried out in the USA,²⁶ (Siemens 2022) Germany,²⁶ Republic of Korea²³⁻²⁵ and the UK.²²

Chest X-ray images were obtained from hospital databases,²²⁻²⁵ the Lung Image Database Consortium,²⁶ a health centre database²⁶ or (confidential information has been removed) (Siemens 2022). The number of CXR images included in the studies ranged from 100²⁶ to 434,²⁵ and the number of participants who provided CXR data ranged from 100²⁶ to 400²². No information was provided in any of the studies about the referral route of patients who provided CXR data. It is plausible that the studies include both symptomatic patients and those who underwent CXR for reasons unrelated to lung cancer, as well as those from excluded populations such as people referred from other healthcare settings.

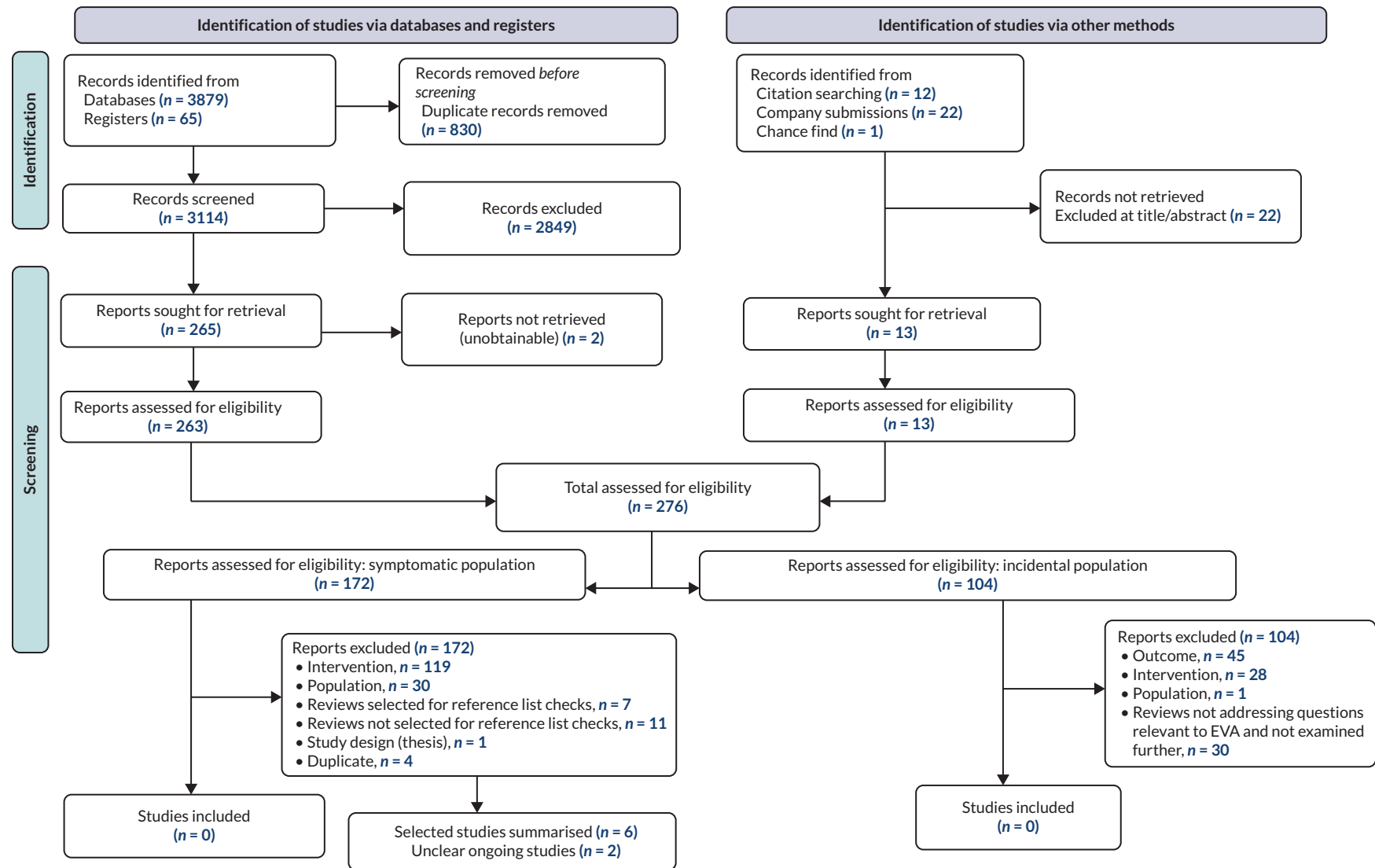


FIGURE 2 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

TABLE 2 Characteristics of summarised (but ineligible) studies

Study (first author and year)	Country	Study design	Population	Index test	Comparator	Reference standard
Dissez 2022 ²²	UK	Retrospective cohort study, one centre	400 CXRs from 400 adults	Red Dot (Behold.ai) + radiologists	Radiologists, radiographers	Blind reads of CXRs by two consultant radiologists
Nam 2020 ²³	Republic of Korea	Retrospective cohort study, one centre	218 CXRs from 218 adults	Lunit INSIGHT version 1.0.1.1 + radiologists	Radiologists	CT scan
Jang 2020 ²⁴	Republic of Korea	Retrospective cohort study, one centre	351 CXRs from 351 adults	Lunit INSIGHT version 1.2.0.0 + radiologists	Radiologists	CXR and CT images
Koo 2021 ²⁵	Republic of Korea	Retrospective cohort study, one centre	434 CXRs from 378 adults	Lunit INSIGHT CXR version 1.00 + radiologist	Radiologists	Consensus from two thoracic radiologists using CXR or CT
Homayounieh 2021 ²⁶	Germany; USA	Retrospective cohort study, two centres	100 CXRs from 100 adults	AI-Rad Companion CXR (Siemens Healthineers) + radiologist	Radiologists	Consensus from two thoracic radiologists using all available clinical data
Siemens 2022	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Prototype AI-Rad Companion CXR algorithm (Siemens Healthineers) + radiologist	Radiologists	Consensus from two thoracic radiologists using CXR or CT

A summary of characteristics of the CXRs assessed by the studies differed both within and across studies ([Table 2](#); more detailed descriptions are provided in [Report Supplementary Material 1](#)). The UK study²² identified random samples of patients who had a clinical text report indicating potentially malignant CXR and a follow-up CT and those with a clinical text report of no urgent findings. Nam *et al.*²³ and Jang *et al.*²⁴ both included a large proportion of confirmed cancer cases with false-negative CXRs prior to diagnosis. Homayounieh²⁶ selected CXRs to ensure that negative and positive cases with different levels of difficulty in detection were included. Siemens 2022 (confidential information has been removed). Koo *et al.*²⁵ included adults with three or fewer nodules on both CXR and CT with at least one nodule pathologically confirmed on biopsy as either benign or malignant.

Images were assessed by a mix of consultant radiologists, board-certified radiologists, radiology trainees and reporting radiographers,²² experienced radiologists,²³ experienced radiologists and radiology residents^{24,25} and senior and junior radiologists.²⁶ This information was not reported in one study (Siemens 2022). The readers had 1²⁴ to 35²⁶ years of experience of reporting CXRs.²²⁻²⁶ One study reported the number of readers with fewer or more than 4 years of experience (Siemens 2022). The number of clinicians included in the studies ranged from 4^{23,25} to 11.²² The accuracy of readers in detecting nodules or lung cancer with and without AI software was each compared with a ground-truth or reference standard, and these varied between the studies ([Table 3](#)). The threshold for defining a positive index test result (i.e. what was considered to be a nodule on CXR) was not defined in any of the studies.

Three studies assessed Lunit INSIGHT,²³⁻²⁵ one assessed Red Dot Behold.ai²² and two assessed AI-Rad Companion Siemens²⁶ (Siemens 2022). It is unclear whether the prototype AI software described in Siemens 2022 is commercially available.

Only a small number of outcomes relevant to the present review were assessed: test accuracy (lung cancer),²² test accuracy (lung nodules)²³⁻²⁶ (Siemens 2022), CT referrals,^{22,24} acceptability of AI to clinicians²² and CXR reading times.^{24,25}

The following risks of bias and applicability concerns were present in the papers:

- Retrospective study designs were used. There is, therefore, the potential for selection bias, missing data and confounding.
- Assessments were conducted on test sets of data interpreted outside clinical practice. Caution is needed in extrapolating from these types of studies, as prior evidence suggests little to no association between performance in this environment and that seen in clinical practice.²⁷
- Only one study was conducted in the UK;²² however, it is unclear if the population from whom the CXRs were taken are reflective of people who would be referred from primary care in a real-world setting. The generalisability of results from the other five studies is similarly limited in this way, and also because populations from the USA and Republic of Korea may differ from the UK population in disease prevalence rates, age and comorbidities, and ethnic diversity.²⁸ There may also be differences in treatment settings and in the training and expertise of radiologists.
- Artificial intelligence software manufacturers were involved in three of the six studies [financial support $n = 2$ ²⁶ (Siemens 2022), employees authors $n = 1$ ²²]. Prior evidence suggests that studies conducted by drug/device manufacturers tend to report more favourable results than non-industry studies.²⁹ Caution in interpretation of these studies is warranted until independent assessment of the AI software is obtained.
- Each radiologist interpreted each CXR with and without AI software. In three studies^{22,24} (Siemens 2022), there was a washout period between readings, whereas in others,^{23,26} the radiologist was aware of their initial decision at the second reading. This is not reflective of UK clinical practice, and there is concern that the first reading could influence the second reading.

TABLE 3 Test accuracy results from summarised (but ineligible) studies

Study name	AI name	Number of patients	Number of CXRs	Number of cancers/nodules	Group	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Lung cancer detection											
Dissez 2022 ²²	Red Dot (Behold.ai)	400	400	72	With AI	55 ^a	82 ^a	17 ^a	246 ^a	77% (75% to 80%)	75% (71% to 77%)
					Without AI	48 ^a	266 ^a	24 ^a	62 ^a	66% (59% to 71%)	81% (77% to 85%)
Nodule detection											
Nam 2020 ^{23,b}	Lunit INSIGHT version 1.0.1.1	N/R	N/R	N/R	With AI	357	36	315	164	53% (49% to 57%)	82% (77% to 87%)
					Without AI	316	44	356	156	47% (43% to 51%)	78% (72% to 84%)
Jang 2020 ^{24,b}	Lunit INSIGHT version 1.2.0.0	351	351	117	With AI	66	19	51	215	56% (47% to 65%)	92% (88% to 95%)
					Without AI	50	24	67	210	43% (34% to 52%)	90% (86% to 94%)
Koo 2021 ²⁵ – per patient any nodule	Lunit INSIGHT version 1.0.0.0	378	434	165	With AI	157 ^a	6 ^a	8 ^a	207 ^a	95% (91% to 98%)	97% (94% to 99%)
					Without AI	152 ^a	15 ^a	13 ^a	198 ^a	92% (87% to 96%)	93% (89% to 96%)
Koo 2021 ²⁵ – per nodule		N/R	N/R	N/R	With AI	N/R	N/R	N/R	N/R	94% (N/R)	N/R (N/R)
					Without AI	N/R	N/R	N/R	N/R	89% (N/R)	N/R (N/R)
Homayounieh 2021 ²⁶	AI-Rad Companion CXR	100	100	N/R	With AI	26.4	2.5	23.6	47.5	55% (48% to 63%)	95% (91% to 9%)
					Without AI	23.6	4.1	26.4	45.5	45% (38% to 53%)	93% (89% to 96%)
Siemens 2022	Prototype AI-Rad Companion CXR algorithm	Confidential information has been removed	Confidential information has been removed	N/R	With AI	N/R	N/R	N/R	N/R	Confidential information has been removed (N/R)	Confidential information has been removed (N/R)
					Without AI	N/R	N/R	N/R	N/R	Confidential information has been removed (N/R)	Confidential information has been removed (N/R)

FN, false negative; FP, false positive; N/R, not reported; TN, true negative; TP, true positive.

^a Values calculated by the EAG using sensitivity, specificity and prevalence data reported in papers.

^b 95% CIs calculated by the EAG using TP, FP, FN and TN. Data are mean values for all readers.

- The threshold for defining a positive index test result was not defined in the studies; therefore, it is not possible to know whether the results of these studies are reflective of how AI would perform under clinical practice conditions, nor is it possible to know whether the results are comparable between studies.
- Where CT referrals were reported, these were hypothetical referrals rather than actual referrals and may not reflect real-world practice.

What are the test accuracy and test failure rates of adjunct artificial intelligence software to detect lung cancer on chest X-rays?

We did not identify any studies that met the inclusion criteria for this question.

Results of six summarised (but ineligible) studies are reported in [Table 3](#). Studies reported test accuracy for individual readers and/or mean values for all readers; the data summarised in [Table 3](#) are the mean values across readers. Forest plots of test accuracy metrics are given in [Figure 3](#) (sensitivity) and [Figure 4](#) (specificity) for studies containing unredacted data.

One study examined the test accuracy of AI software to detect lung cancer on CXRs.²² In this UK study of Red Dot (Behold.ai), sensitivity was significantly higher for the interpretation of CXRs with AI (77%, 95% CI 75% to 80%) than without AI (66%, 95% CI 59% to 71%). No difference was observed for specificity ([Table 3](#)).

Five studies examined the test accuracy of AI software to detect lung nodules on CXRs²³⁻²⁶ (Siemens 2022). Three studies from Republic of Korea²³⁻²⁵ assessed different versions of the Lunit INSIGHT AI software. No statistically significant differences were observed in sensitivity or specificity between readers with and readers without AI in the studies by Nam *et al.*²³ and Jang *et al.*²⁴ ([Table 3](#)). In the third paper,²⁵ an assessment of test accuracy was conducted for *any* nodule and *each* nodule. In the analysis of any nodule, sensitivity was 95.1% for readers with AI software and 92.4% for readers without AI software, and specificity was 97.2% for readers with AI software and 93.1% for readers without AI software. In the analysis of each nodule, sensitivity was 93.9% for readers with AI software and 88.6% for readers without AI software. Specificity was not reported. Instead, false-positive rates were reported to be 3.2% for readers with AI software and 6.3% for readers without AI software. Caution is required in the interpretation of false-positive data, as the paper reports that the false-positive rate is the total number of false positives divided by the number of CXRs, which is a non-standard calculation. It is not possible to know if the above estimates reflect true differences between assessment with/without AI software as no statistical analyses were presented in the paper for any of the above test accuracy metrics, and there were insufficient data to allow us to conduct our own analyses.

Two studies (Homayounieh *et al.*²⁶ and Siemens 2022) assessed versions of the Siemens Healthineers AI-Rad Companion CXR in US and German populations (one study) or in (confidential information has been removed) populations alone ([Table 3](#)). No statistically significant differences were observed in sensitivity or specificity between radiologists with and radiologists without AI software in the study by Homayounieh *et al.*²⁶ In the Siemens 2022 study (confidential information has been removed).

One study²² reported the mean number of cancers detected and found no significant differences with and without AI software (54 cancers, 95% CI 42 to 59 cancers, and 46 cancers, 95% CI 38 to 51 cancers, respectively).

None of the six studies reported AI software test failure.

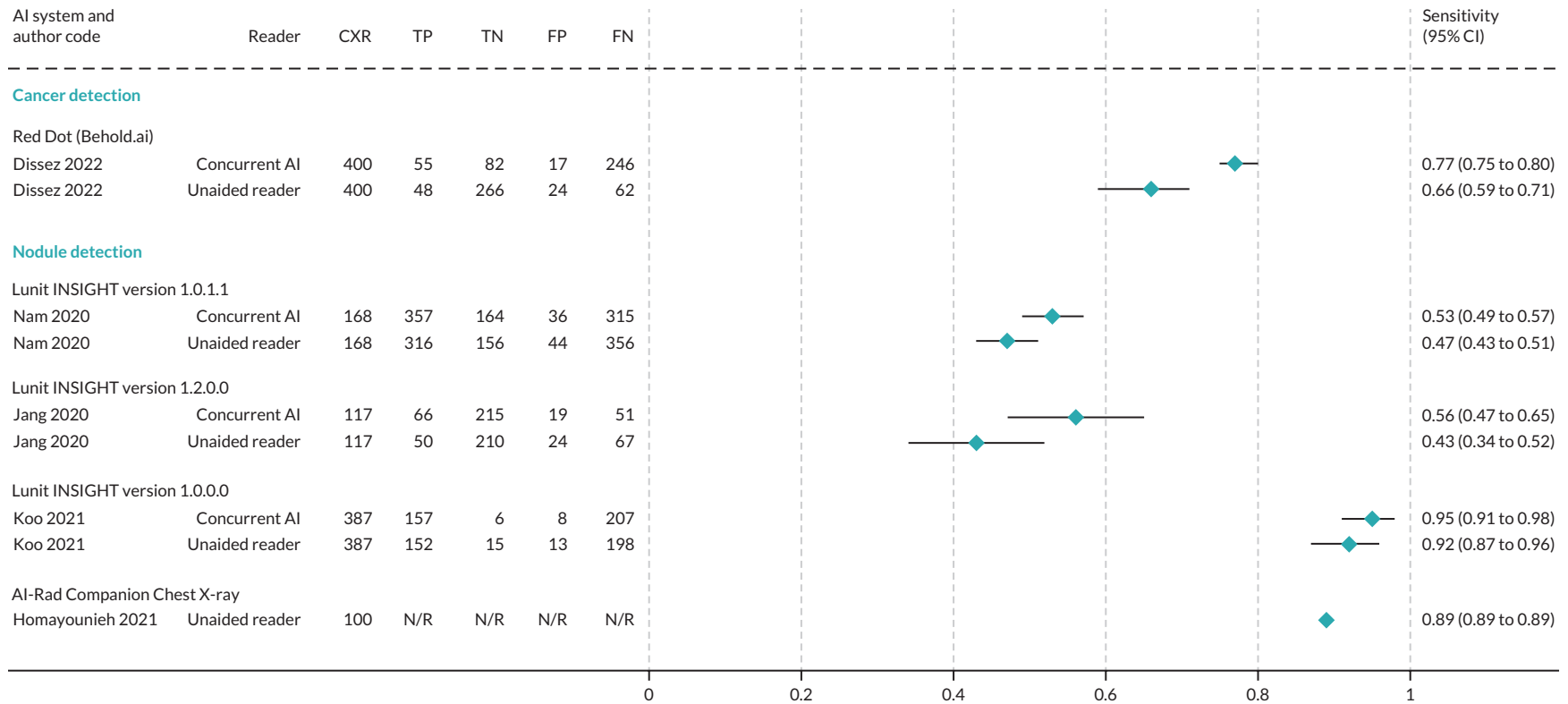


FIGURE 3 Forest plot of sensitivity. Note: it was not possible to calculate CIs for Homayounieh *et al.*²⁶

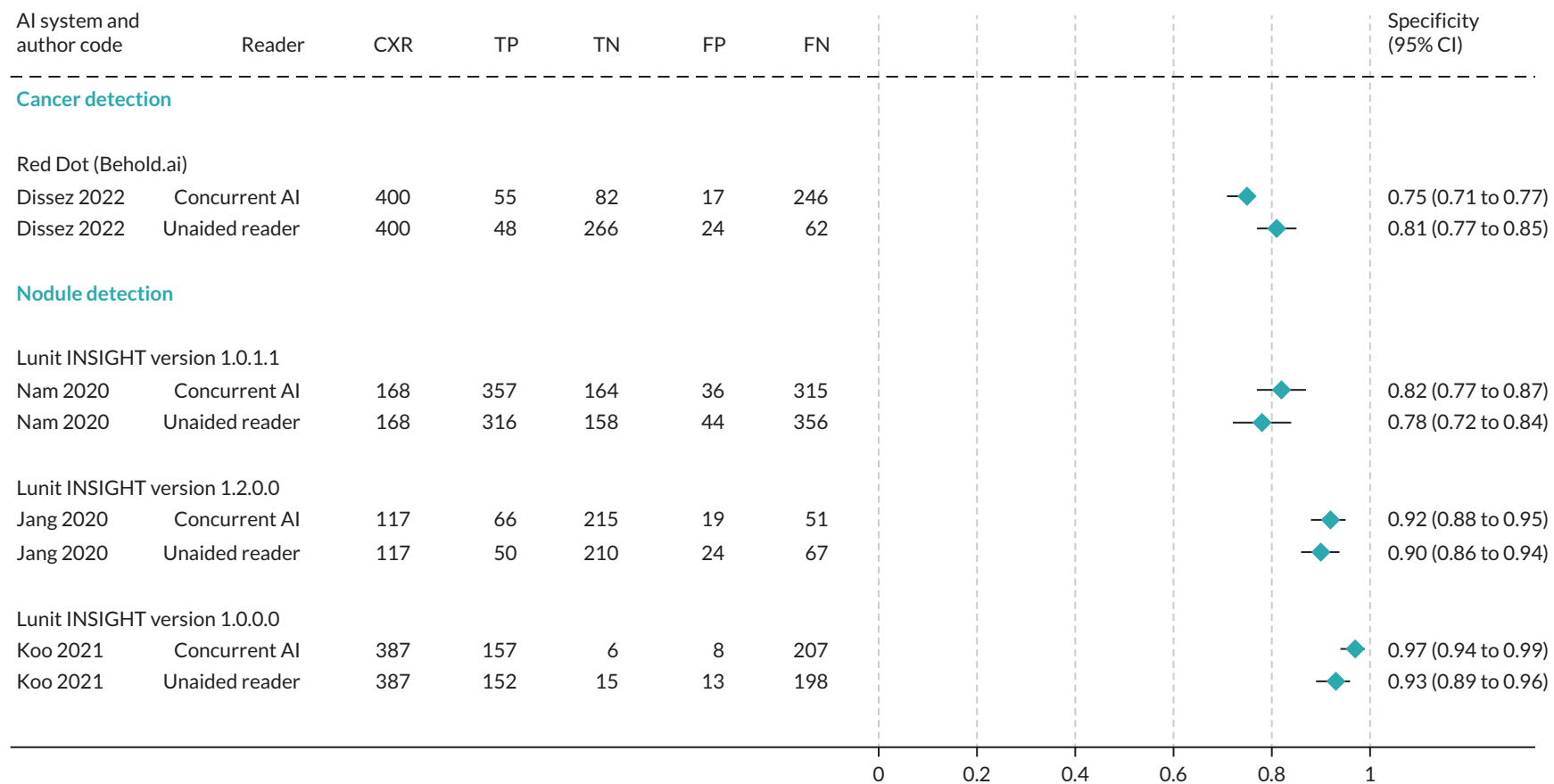


FIGURE 4 Forest plot of specificity.

What are the practical implications of adjunct artificial intelligence software to detect lung cancer on chest X-rays?

We did not identify any studies that met the inclusion criteria for this question.

Two of the summarised (but ineligible) studies^{22,24} provided information on the potential referrals for CT. No statistically significant differences were observed in the number of people who might be recommended for CT follow-up between readers with and readers without AI: Red Dot (Behold.ai) 144 out of 400 (36%) (95% CI 119 to 172) potential referrals with AI and 117 out of 400 (29%) (95% CI 93 to 147) potential referrals without AI;²² Lunit INSIGHT 96 out of 351 (27%, 95% CI 22.8% to 32.3%, calculated by the EAG) with AI and 80 out of 351 (23%, 95% CI 18.5% to 27.5%, calculated by the EAG) patients without AI.²⁴ It is important to note that these are hypothetical referrals. We found no evidence on the impact of AI on the readers' behaviour in real-world clinical practice.

Two studies^{24,25} reported information on reading times. No statistically significant differences were observed in average image reading times between readers with and readers without AI: Siemens Healthineers AI-Rad Companion 22.5 [standard deviation (SD) 40.3] seconds with AI, 24.3 (SD 27.4) seconds without AI, per image;²⁴ Lunit Insight 171 (SD 33.8) minutes with AI, 211.25 (SD 38.4) minutes without AI, to read 434 CXR.²⁵

One study²² reported on the acceptability of Red Dot (Behold.ai) among 10 out of 11 study clinicians. Eight clinicians indicated that reporting was not slowed down by AI, and nine stated that 'the heatmaps (visual display of findings suspicious of lung cancer on CXRs) produced by the AI model were helpful to understand the algorithm's attention points'.²²

What is the clinical effectiveness of adjunct artificial intelligence software applied to chest X-rays?

We did not identify any studies that met the inclusion criteria for this question. None of the six summarised (but ineligible) studies reported clinical effectiveness outcomes.

Summary of ongoing trials

No ongoing trials meeting the inclusion criteria were identified. As described in [Analysis and synthesis](#), we looked for ongoing trials assessing eligible comparisons.

We identified one ongoing trial (KCT0005466)³⁰ comparing Lunit INSIGHT in conjunction with a radiologist with radiologist alone; however, the population is those undergoing CXR for any reason in the outpatient department. It is not known whether the participants underwent CXR for symptoms of cancer or for reasons other than cancer, or if they were referred from primary care.

Details of one ongoing study (NCT05489471),³¹ identified from the Lunit company submission, are unclear. The proportion of general practitioner (GP) referrals, A&E attendances and inpatients is not known, the AI software is not named (but the study is funded by Lunit) and it is not clear whether the comparison is AI software in conjunction with a radiologist versus radiologist alone. This UK-based study is not yet recruiting and has an estimated primary end date of July 2023.

In addition, the Siemens 2022 study provided in the Siemens Healthineers' company submission summarised above is ongoing ([Table 4](#)).

TABLE 4 Ongoing studies

Heading	Details
Trial identifier number	KCT0005466 ³⁰
Title of project	Prospective evaluation of deep-learning-based detection model for chest radiographs in outpatient respiratory clinic
Trial completion date	31 May 2021 (no results posted)
Trial identifier number	NCT05489471 ³¹
Title of project	A study to assess the impact of an AI system on CXR reporting
Trial completion date	Estimated primary end date of July 2023

Chapter 5 Cost-effectiveness

Results of literature searches

A total of 1120 studies were identified through database searches (817 in MEDLINE and 303 in Tufts CEA). Of these, 29 studies were retrieved for full-text assessment (25 from MEDLINE and 4 from Tufts CEA). These covered a wide range of methodologies and research questions. Reference lists of these studies returned four further studies of relevance to this review.

We did not identify any cost-effectiveness studies that directly compared CXR review by radiology specialist with adjunct AI and radiology specialist review without. However, two economic evaluation studies from the database search^{32,33} and an updated analysis of one of these³⁴ found through an authorship search were identified as useful to inform modelling techniques and parameter input sources. Similarly, four studies (one from the database search³⁵ and three from author searches^{36–38}) provided detailed information on radiological and clinical pathways to lung cancer diagnosis in the UK. A systematic review and meta-analysis on the diagnostic performance of CXRs in symptomatic primary care populations³⁹ was also retrieved from the search.

These studies were retained and summarised narratively to include information of relevance to populate the conceptual model. No formal data extraction or quality appraisal was conducted. The studies by Snowsill *et al.*³³ and the Exeter Test Group and Health Economics Group³⁴ were not summarised. Information of the diagnostic component of the conceptual model was prioritised due to project time constraints, whereas these studies^{33,34} pertained more to the longer-term treatment costs and utilities.

Description of the evidence

*Bajre et al.*³²

*Bajre et al.*³² used a decision tree structured model to assess the cost-effectiveness of trained radiographers compared with radiologists for the reporting of CXR in people suspected of having lung cancer. The model simulated a pathway for a hypothetical cohort of 1000 people undergoing CXR for suspected lung cancer, with cost-effectiveness calculations concluding at 5 years. The model started with a cohort of people receiving either a radiologist-reported CXR or a radiographer-reported CXR. The pathway for both strategies was the same. The proportion of those with true disease status was known, characterised by the prevalence of lung cancer. People with lung cancer who had a positive CXR result received a confirmatory test of a CT scan, which also provided staging. The authors included stages I, II, III and IV lung cancers. People with a false-negative result presented later to the A&E department, where they were diagnosed with lung cancer and staged. People who had a false-positive result following CXR received a CT scan that confirmed no lung cancer was present. People who had no lung cancer and had been correctly identified as negative by the CXR received no further testing/imaging.

Information required to populate the model was obtained from the literature and NHS reference costs. The model required information about the prevalence of lung cancer, sensitivity and specificity of radiologist-reported and radiographer-reported CXR to identify lung cancer, and sensitivity and specificity for radiologist-reported CT scan to confirm lung cancer diagnosis and probabilities. Although not explicitly stated, a confirmatory diagnosis was made by the radiologist. The proportion of people diagnosed at first presentation was obtained from statistics published by Cancer Research UK in 2013 (*Bajre et al.*³²). Additionally, information was required about the probability of lung cancer by stage at second presentation following misdiagnosis. All costs included in the model were reported at 2014–5 prices. Costs were required for radiologist and radiographer reading of CXRs, cost of CT scans and total costs of treatment by stage. The authors were not explicit about which treatment people received.

The benefit of the strategies was reported in terms of cases detected at first presentation and quality-adjusted life-years (QALYs) yielded. Utility values by stage of diagnosis were obtained from Naik *et al.*⁴⁰

Several simplifying assumptions were made to give a workable model structure:³²

- Time taken to report CXR is 2 minutes for both radiographers and radiologists.
- False negatives present at A&E at a later date, at which point disease may have advanced a stage (for patients at stages I to III).
- Sensitivity and specificity of radiographer reporting of CXR and radiologist reporting of both CXR and CT scan are independent of disease stage or other patient characteristics such as age.
- Quality of life (QoL) in the year following diagnosis (according to stage at diagnosis) is maintained in subsequent years.
- There is no QoL impact arising from false-positive reporting.
- Findings for non-small-cell lung cancer are representative of lung cancers in general.

The perspective and setting of the economic analysis were not clearly defined, but the analysis appears to be from the NHS and Personal Social Services (PSS) in a secondary care setting, based on the cost inputs. The results of the analysis were presented in terms of an incremental cost-effectiveness ratio (ICER), expressed as cost per QALY. The authors undertook a probabilistic sensitivity analysis (PSA) to assess the joint uncertainty in key model input parameters: prevalence of lung cancer, sensitivity and specificity of radiologist and radiographer reporting of CXRs, lung cancer stage distribution at initial CXR and stage progression following misdiagnosis. The authors stated the sampling distributions for the parameters included in the PSA but have not reported their parameters. The authors undertook a threshold analysis but not a one-way sensitivity analysis.

The authors reported disaggregated results for both strategies. Results were reported on the number of people expected to be diagnosed with lung cancer, QALYs yielded and treatment costs, all by stage. The QALYs yielded appeared to be high, with stage IV expected to yield more QALYs than stage III, and both of these higher than with stage II. There were modest QALY gains by strategy and by stage, with stage I having the greatest expected gain of 2.4 QALYs, favouring radiographer reporting. Radiographer reporting yielded more overall QALYs, but it was unclear with the inputs reported why QALYs yielded by radiologist reporting was greater for stages II and stage IV. Radiographer reporting diagnostic and treatment costs were lower than radiologist reporting costs. Overall results showed that radiographer reporting of CXR dominated radiologist reporting. The PSA results showed that radiographer reporting continued to dominate radiologist reporting in 98% of the iterations. Based on the model structure, its inputs and assumptions, the authors concluded that the use of trained radiographers to report CXR is cost-effective and an increased role for radiographers in the diagnostic pathway would be beneficial to meet hospital waiting time targets for lung cancer diagnosis.

Foley *et al.*³⁵

Foley *et al.*³⁵ conducted a retrospective review of trust audit data (Royal United Hospitals Bath NHS Foundation Trust) to analyse the use of CXR as the first-line investigation in primary care patients with suspected lung cancer. In total, 1488 of the 16,495 primary care referrals received between 1 June 2018 and 31 May 2019 were for suspected lung cancer. CXRs were coded by result as CX1, normal but a CT scan is recommended to exclude malignancy; CX2, alternative diagnosis; or CX3, suspicious for cancer. Outcomes for the study cohort were stratified by CX code and included patient characteristics, number undergoing CT scan, number of lung cancers diagnosed, stage at diagnosis, time from initial CXR to CT scan, time from CT request to CT scan, time to diagnosis, treatment strategy taken and mortality (over an average follow-up period of 322 days in the total cohort). [Table 5](#) shows the results of key outcomes.

Based on these findings, the authors concluded that there was significant delay in lung cancer diagnosis in patients who received a CX1 'normal' initial CXR result ($p < 0.001$) and the majority of patients with a

TABLE 5 Outcome data reported by code (Foley *et al.*³⁵)

Outcome	CXR report code			Statistical significance ($p < 0.05$)
	CX1 (normal but CT scan recommended to exclude malignancy)	CX2 (alternative diagnosis)	CX3 (suspicious for malignancy)	
Total number of CXRs (%)	1056 (75)	288 (20)	72 (5)	–
Number referred for CT (%)	107 (10)	107 (37)	66 (92)	–
Number of lung cancers diagnosed (%)	10 (1)	29 (10)	49 (68)	–
Number diagnosed at advanced stage IIIc/IV (%)	5 (50)	11 (38)	28 (57)	$p = 0.26$
Number of days from CXR to CT ^a	34.6	19.6	1.9	$p < 0.001$
Number of days from CXR to diagnosis ^a	89.7	65.3	30.2	$p < 0.001$
Number receiving treatment with curative intent (%)	4 (40)	14 (48)	13 (27)	$p = 0.14$
Number of deaths in follow-up period (all-cause mortality) (%)	5 (50)	10 (34.5)	27 (55.1)	$p = 0.42$

^a Mean value.

'normal' or an 'abnormal' CXR are diagnosed at an advanced disease stage ($p = 0.26$) with no difference in survival outcomes based on the CXR findings ($p = 0.42$).³⁵

Bradley *et al.*³⁶

Bradley *et al.*³⁶ undertook a retrospective observational study using routinely collected healthcare data from Leeds Teaching Hospitals NHS Trust.

All patients diagnosed with primary lung cancer between January 2008 and December 2015 with a GP-requested CXR in the year before diagnosis were coded based on the result of the earliest CXR in that period. CXR report codes were assigned: (1) suspicion of lung cancer identified/urgent investigation needed; (2) abnormality identified/non-urgent investigation indicated, including diagnoses of pneumonia or consolidation even if repeat imaging was not explicitly suggested; (3) abnormality identified but no further investigation/assessment indicated; and (4) normal CXR, no abnormalities identified.

The sensitivity of CXR was calculated and analyses were performed on time to diagnosis, stage at diagnosis and survival outcomes. Statistical analysis on these outcomes was performed by combining CXR codes 1 and 2 to form a 'positive' result group and codes 3 and 4 to form a 'negative' result group. However, the authors present numerical outcome data for all codes separately as well as for combined groups. Table 6 shows a summary of the key data by individual codes.

Data were also presented on the number of people who had further CXRs requested by their GPs, with median time to second CXR and median times to diagnosis from initial CXR. Of 376 patients with an initial CXR that was 'negative' (codes 3 and 4), 98 (26.1%) had at least one further CXR. Sensitivity calculated based on initial CXR (codes 1 and 2) was 82.3% (95% CI 80.6% to 84.1%).

The authors concluded that the sensitivity results supported previous systematic review findings,⁴¹ and while those with a 'positive' initial CXR finding had a median of 43 days to diagnosis compared

TABLE 6 Key outcome data reported by individual codes (Bradley *et al.*³⁶)

Outcome	Initial CXR code				Total
	1	2	3	4	
Number of CXRs (%)	1383 (65)	370 (17.4)	230 (10.8)	146 (6.9)	2129
Time from CXR to diagnosis, median days (IQR)	36 (23–63)	93 (55–154)	211 (181–296)	193 (87–279)	51 (29–107)
Survival from CXR, median days (IQR)	313 (126–877)	400 (163–964)	408 (238–958)	420 (214–1117)	345 (148–920)
Stage I/II at diagnosis, <i>n</i> (%) [95% CI]	397 (28.7) [26.4 to 31.2]	111 (30) [25.4 to 35.0]	83 (36.1) [30.0 to 42.7]	43 (29.5) [22.4 to 37.7]	634 (29.8) [27.9 to 31.8]
Stage III/IV at diagnosis, <i>n</i> (%) [95% CI]	981 (70.9) [68.4 to 73.3]	259 (70) [65.0 to 74.5]	147 (63.9) [57.3 to 70.1]	103 (70.5) [62.4 to 77.7]	1490 (70) [68.0 to 71.0]
Stage unknown at diagnosis, <i>n</i> (%) [95% CI]	5 (0.4)	0	0	0	5 (0.2)

IQR, interquartile range.

with 204 days for those with 'negative' findings, no direct association with time to diagnosis was found between stage at diagnosis and survival in this study.

Woznitza et al. (2018)

Woznitza *et al.*³⁷ conducted a 4-month feasibility study (November 2016 to March 2017) at a single radiology department at an acute general hospital (Homerton University Hospital, London). The primary outcome was to establish the feasibility of an immediate reporting service for CXRs. Comparison between CXR referrals from general practice that received an immediate and routine report was made to determine the number of lung cancers diagnosed, time to diagnosis, time to CT and number of urgent referrals to respiratory medicine.

From 1687 CXRs of people referred from general practice over the 26-week study period, 36 patients (22 immediate CXR report, 14 routine CXR report) had a CT scan arranged by radiology following a suspicious CXR. This equated to less than one additional unplanned patient per week (mean 0.8 scans per week) accommodated by the CT department. Time from CXR to CT was shorter in the immediate report group, with a mean of 0.9 (SD 2.3) days, than in the routine reporting group, at 10.6 (SD 4.5) days ($p > 0.0001$). No apparent difference was found in time to discussion at multidisciplinary team (MDT).

The study also gave a detailed description of the radiology department demographics and processes for reporting and referral. The results of all CXRs included in the study and pathways taken were explained, including 17 patients with a normal or non-cancer diagnosis at CXR who were subsequently diagnosed with lung cancer.

The authors concluded that it was feasible to introduce a radiographer-led immediate CXR reporting service, but a definitive study assessing outcomes would be needed to determine whether this would have an impact on patient mortality and morbidity.

Woznitza et al. (2022)

Woznitza *et al.*³⁸ conducted a prospective, block-randomised controlled trial (RadioX) at a single acute district general hospital in London (Homerton University Hospital). People referred for CXR from primary care attended sessions that were pre-randomised to either immediate radiographer (IR) reporting or standard radiographer (SR) reporting within 24 hours. Those who received SR reporting were the control group, as this was usual practice in the department. In the intervention group, CXRs were reported while the patient was still in the department, with all patients with CXR findings suggestive of lung cancer offered a same-day CT scan. Those who declined were scheduled for another day.

In total, 8682 CXRs were performed between 21 June 2017 and 4 August 2018, 4096 (47.2%) for IR and 4586 (52.8%) for SR. Lung cancer was diagnosed in 49 patients. [Table 7](#) shows the summary outcome data from trial reporting arms.

The authors stated that a health economic evaluation based on their RadioX trial was to be reported separately.³⁸ The corresponding author was contacted and confirmed that analysis of the data was still under way and they were unable to share any usable information at that time (Nicholas Woznitza, consultant radiographer, University College Hospital London NHS Foundation Trust, 29 November 2022, personal communication).

Dwyer-Hemmings and Fairhead³⁹

The authors performed a systematic review of evidence to inform the diagnostic accuracy of CXR to detect lung malignancy in symptomatic patients presenting to primary care. Nine databases were searched, and data from included studies were extracted to calculate the sensitivity and specificity of CXR where possible. Risk of bias was assessed using the QUADAS-2 tool, with analyses conducted by means of random-effects meta-analyses. Ten studies were included in this review. Summary sensitivity of five studies (those not at high risk of bias) was 81% (95% CI 74% to 87%). Specificity of five studies

TABLE 7 Outcome data reported by immediate and standard CXR reporting arms (Woznitza *et al.*³⁸)

Outcome	Immediate reporting	Standard reporting
Total patients	4096	4586
Previous CXR, <i>n</i> (%)		
Yes	2297 (56.1)	2583 (56.3)
No	1799 (43.9)	2003 (43.7)
Previous CT, <i>n</i> (%)		
Yes	307 (7.5)	334 (7.3)
No	3789 (92.5)	4252 (92.7)
Lung cancer suspected, <i>n</i> (%)		
Yes	1326 (32.4)	1511 (33.0)
No	2757 (67.3)	3062 (66.7)
Known	13 (0.3)	13 (0.3)
Total cancers diagnosed, <i>n</i> (%)	27 (0.7)	22 (0.5)
2WW referral		
Yes	150 (3.7)	189 (4.1)
No	3946 (96.3)	4397 (95.9)
Time from CXR to diagnosis (days)		
Median (IQR)	32 (19, 70)	63 (29, 78) ^a
Mean (SD)	47.2 (35.8)	81.6 (78.5)
Time from CXR to discharge (days) (no cancer diagnosis)		
Median (IQR)	30 (17, 64)	27 (14, 61)
Mean (SD)	54.4 (60.4)	50.3 (63.7)

2WW, 2-week wait referral on suspected lung cancer pathway.
a *p* = 0.03.

was 68% (95% CI 49% to 87%). The authors concluded that there was good evidence regarding sensitivity because they included only those studies that were of similar design and were not at high risk of bias. By contrast, they considered the evidence on specificity to be weaker due to differences in study designs and variability in reported outcomes.³⁹

Clinical pathway for representation in model

The clinical pathway illustrated in [Figure 5](#) was agreed in the NICE final scope.⁹

The development of this pathway was supported by existing guidelines on the diagnostic and care pathway^{10,11,16,17} and collaboration with specialist committee members (SCMs) during the scoping process. Subsequent feedback from SCMs and clinical experts generally supported this as a representation of the multiple pathways patients may follow after primary care referral for a CXR. All emphasised that this was an aspirational pathway, with many alternative routes both in and out through to diagnosis, and that it was not particularly accurate in several trusts.

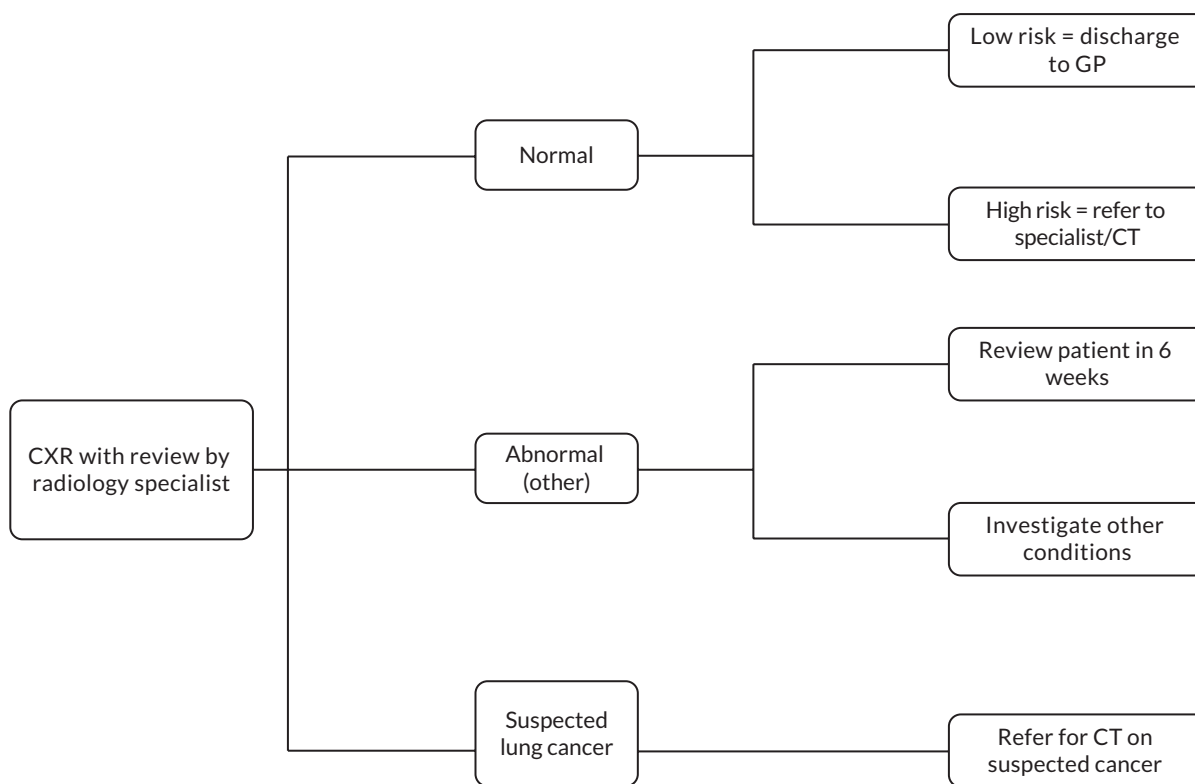


FIGURE 5 The CXR pathway for people referred from primary care, from the NICE final scope.⁹

When critical pathway events were mapped based on the early stages of the National Optimal Lung Cancer Pathway¹⁷ using large cancer databases from two trusts, 83 individual combinations of early pathway events in 1018 suspected lung cancer patients were found.⁴² This highlights the complexity of defining a realistic structure on which to base the clinical component of an economic model. All models by their nature are more simplistic formats of real practice. The balance is to represent the clinical pathway in sufficient detail to capture the main elements, while producing a model that is feasible to construct.

The availability of evidence to inform model parameters also influences the model structure. Where evidence is severely limited, a more simplistic model reduces the number of assumptions relied on to achieve an executable model and reduces the uncertainty introduced.

Two studies identified in the literature search^{35,36} reported data for parameters that had the potential to support multiple differential pathways after CXR results, rather than just a lung cancer suspected and no lung cancer suspected route through model. However, there were limitations in how the data reported from both sources might be applied.

Overall, the EAG determined that the clinical pathway developed during the NICE scoping process was a realistic representation on which to base the conceptual model. Although concerns remained around the feasibility of parameterising the model due to a lack of available evidence and differences in outcome reporting, five differential pathways (A, B, C, D and E) were formulated with feedback from clinical experts and reference to the clinical guidelines.^{10,11,16,17}

Figure 6 shows where each pathway is situated, and each pathway is described in detail below.

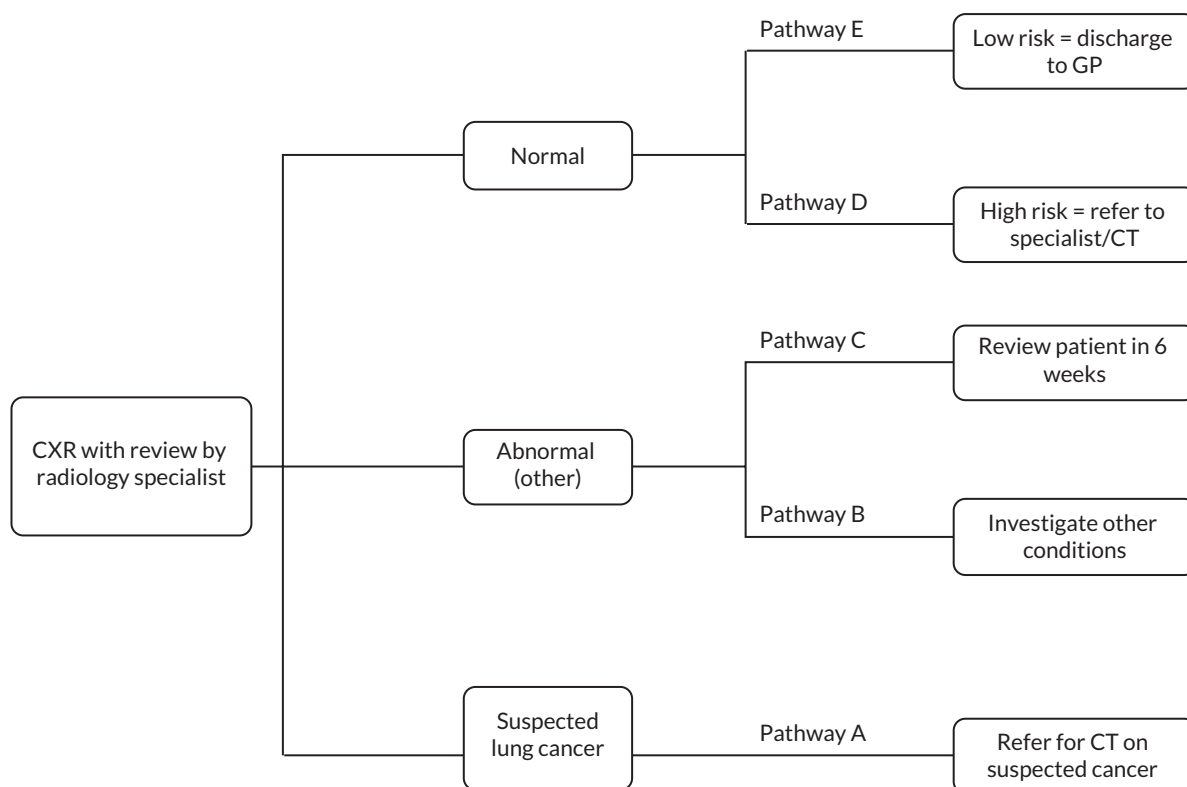


FIGURE 6 Clinical pathways for conceptual model for people referred from primary care.

Pathway A

When CXR findings are suggestive of malignancy, a referral for urgent CT on the suspected lung cancer pathway is made. There is a variation in practice across trusts, but in many institutions highly suspicious CXR findings are flagged to secondary care lung cancer teams who request the CT scan and await referral to the suspected lung cancer clinic from the GP. Once reported, CT scans are triaged by lung cancer team consultants. If they suggest probable lung cancer, an urgent lung cancer team appointment is arranged with appropriate tests for example spirometry, planned biopsy (endobronchial ultrasound) (Alberto Alonso, consultant radiologist, Manchester Hospital NHS Foundation Trust, 19 November 2022, personal communication) for histopathological staging and to inform treatment options at the fast-track lung cancer clinic.¹⁶

If the CT scan appears reasonably normal (despite CXR appearances), then the lung cancer team writes to the patient to inform them of their relatively normal CT appearances and arrange a non-urgent general respiratory clinic (not lung cancer clinic) outpatient appointment (Vidan Masani, consultant respiratory physician and lead for lung cancer, Royal United Hospitals Bath NHS Foundation Trust, 1 February 2023, personal communication). This also includes those who require investigation and management of pulmonary nodules in accordance with British Thoracic Society guidelines.^{16,17}

Pathways B and C

If CXR results are reported as 'abnormal' where findings are indeterminate or suggestive of an alternative diagnosis, people may follow pathway B or C. Here, findings are not sufficient to warrant further urgent investigation, but additional clinical enquiry is required.

Pathway B is taken when an alternative diagnosis is suspected and referral is made by the GP to a secondary care outpatient clinic with relevant expertise for that clinical finding, for example a non-urgent respiratory clinic.

Pathway C is followed when a 6-week repeat CXR is advised in the report. The referral for repeat CXR is made by the GP, and a radiologist or reporting radiographer compares the new image with the previous one. If the abnormality is resolved, then no further action or follow-up is required. If abnormal and suspicious, these cases are 'red-alerted' or 'upgraded' and the lung cancer team and referring GP are notified as per pathway A. The 6-week repeat CXR is used in cases where there is need to exclude infection, try a course of treatment and reassess before considering CT (Jonathan Rodrigues, consultant radiologist, Royal United Hospitals Bath NHS Foundation Trust, 13 November 2022, personal communication).

Pathways D and E

Where CXRs are reported as 'normal', findings may be unremarkable, but several trusts (including Royal United Bath NHS Foundation Trust and Manchester University NHS Foundation Trust) include the following automatic caveat in the report: 'please note that a normal CXR does not exclude malignancy. If there is still a strong suspicion of malignancy (weight loss/unresolved cough/significant or unresolved haemoptysis), referral for a CT scan is advised'. This is to counter false reassurance in a case where clinical suspicion remains high.

People with normal results may, therefore, proceed along pathway D, where their GP considers them at high risk of lung cancer despite nothing being detected on CXR and refers the patient for CT scan and specialist review.

Pathway E is taken when the GP has no further concerns, no further diagnostic testing is requested and management is continued under primary care.

Discussion of inputs to inform model structure

To formulate a final conceptual model, an iterative process was used. This included identifying relevant intermediate and long-term outcome measures for parameterisation and selecting a structure that is most appropriate to support their inclusion.

This section describes the available evidence, gaps in evidence and recommendations for appropriate evidence generation for a range of outcome measures. In this report, these will be classified into intermediate measures (short- to medium-term clinical outcomes encountered during the diagnostic process), long-term clinical outcomes and cost inputs.

Intermediate measures for consideration

- Accuracy in detecting lung cancer

No eligible studies were found in the clinical effectiveness review, but one of the six ineligible studies summarised examined the test accuracy of AI software in detecting lung cancer on CXR.²² In this UK study of Red Dot (Behold.ai), sensitivity was significantly higher for the interpretation of CXR with AI (77% 95% CI 75% to 80%) than without AI (66%, 95% CI 59% to 71%). No difference was observed for specificity (with AI: 75%, 71% to 77%; without AI: 81%, 77% to 85%) (Table 3).

A systematic review and meta-analysis identified in the cost-effectiveness literature review³⁹ provided evidence on the test accuracy of CXR to detect lung cancer in symptomatic patients presenting to primary care. In this population, specifically relevant to this review, summary sensitivity of 81% (95% CI 74% to 87%) was calculated from five studies not at high risk of bias. Summary specificity of 68% (95% CI 49% to 87%) was also obtained from five studies, but evidence was weaker due to their heterogeneous design and variation in reported outcomes.³⁹ Findings of this systematic review were supported by two other studies from the cost-effectiveness search.^{36,41} A retrospective database study

of all primary care referrals for CXR conducted by Bradley *et al.* reported sensitivity of 82.3% (95% CI 80.6% to 84.1%). This was calculated based on an initial CXR coding system that included results suggestive of lung cancer and those with an abnormality identified but no urgent investigation indicated as a 'positive' result for CXRs.

- Turnaround time (TAT; time from start of image review to radiology report)

Turnaround time was identified in the final scope⁹ as a potentially useful outcome measure in this assessment. From a modelling perspective, the review time occurs on the pathway prior to the diagnostic decision outcome. This would be captured in a model as a resource use parameter used to calculate the cost per image, where the rate of radiology specialist's pay is multiplied by the length of time to review scan.

A reduction in cost may be expected where TAT is decreased. However, the direction and magnitude of this relationship are highly uncertain given the lack of evidence found on TAT with AI software assistance and the variation of estimates given for TAT without AI from the literature and clinical expert feedback.

Estimated TAT for CXR varies considerably. As discussed in [What are the practical implications of adjunct artificial intelligence software to detect lung cancer on chest X-rays?](#), of the ineligible studies reported on from the clinical search, two^{24,25} presented information on reading times. No statistically significant differences were observed in average image reading times between readers with and readers without AI: Siemens Healthineers AI-Rad Companion 22.5 (SD 40.3) seconds with AI, 24.3 (SD 27.4) seconds without AI, per image;²⁴ Lunit Insight 171 (SD 33.8) minutes with AI, 211.25 (SD 38.4) minutes without AI, to read 434 CXRs,²⁵ which equates to an average of 23.6 seconds per image with AI, and 29.2 seconds without AI (calculated by the EAG).

No information was given on the methods used for timing. With regard to context, timings were recorded during specified reading sessions under study conditions, so how this would translate to reading times in clinical practice is unknown.

Methods by the Royal College of Radiologists (RCR) to derive guidance on reporting output figures are described comprehensively.⁴³ Eighty reports for plain CXRs per hour (45 seconds per image) is the figure expected on average, over a 6-month minimum period, per in-hours, on-site, non-acute 4-hour reporting session in the NHS.⁴³

Specialist committee members advised average reading times of < 1 to 5 minutes, with an assumption of 2 minutes used in the economic evaluation by Bajre *et al.*³²

Many factors have an impact on reporting output and are well outlined by the RCR.⁴³ Therefore, focusing on this as an outcome measure, without appreciation of real-world context, is of little use unless a reduction in TAT can be shown to have an impact on efficiency of workflow over a sustained period in the NHS environment. This needs to be considered when designing future studies.

Another anticipated benefit of reducing TAT is an increase in the output of radiology specialists performing CXR reviews, thereby addressing the high demand for image reading and inherent limitations on workforce capacity. This is a potential value driver of AI software but would not be captured within the conceptual cost-effectiveness model. The potential value here would be recognised at a system level rather than at the patient level represented in the conceptual model.

- Technical failure rate

Technical failure rate was identified in the final scope⁹ as a potential measure of interest. None of the six studies summarised in the clinical effectiveness review reported any information on technical failure rate in CXR.

- Impact of software output on clinical decision-making

Impact of software on clinical decision-making is the primary measure of importance as the final CXR result is determined by a radiology specialist whether or not AI software is used. Even if the diagnostic accuracy of AI software alone is higher, the outcomes are mediated by human input. The results then determine which clinical pathway a patient will proceed down, affecting the quantity and type of further tests.

No evidence was found on this, and the only extrapolated data were in the form of two studies^{22,24} that provided information on hypothetical referrals to CT. No statistically significant differences were observed in the number of people who might be recommended for CT follow-up between readers with and readers without AI: Red Dot (Behold.ai) 144 out of 400 (36%) (95% CI 119 to 172) potential referrals with AI and 117 out of 400 (29%) (95% CI 93 to 147) potential referrals without AI;²² Lunit INSIGHT 96 out of 351 (27%, 95% CI 22.8% to 32.3%, calculated by the EAG) patients with AI and 80 out of 351 (23%, 95% CI 18.5% to 27.5%, calculated by the EAG) patients without AI.²⁴ It is important to note that these are hypothetical referrals, as CXRs were retrospectively selected from databases in these studies. We found no evidence of the impact of AI on the readers' behaviour in real-world clinical practice.

- Number of people referred for a CT scan

The number of people referred for a CT scan depends on test accuracy and referral decision based on CXR result. As highlighted in the clinical review, evidence to inform these parameters that fall earlier in the clinical pathway was not available in the primary care population for CXR review with adjunct AI.

Computed tomography scans may be requested as a result of initial investigations, usually CXRs, undertaken in any of pathways A, B, C and D (see [Clinical pathway for representation in model](#) for a detailed description). Therefore, the proportion of people referred from each pathway for CXR would be needed for a model.

Only two studies identified in our literature search^{35,37} mentioned the number of people referred for a CT scan. Woznitza *et al.*³⁷ reported that a total of 36 patients out of the 1687 referred for CXR from primary care underwent a CT scan. This included both suspected lung cancer and non-suspected lung cancer populations. The study by Foley *et al.*³⁵ provided much more detailed information ([Table 5](#)) and was specific to the GP-referred population with suspected lung cancer. The number and percentage of CT scans requested by the three CXR result codes were reported: CX3 (suspicious for malignancy), 92% (66/72) had a CT scan; CX2 (abnormal, alternative diagnosis), 37% (107/288) had a CT scan; and CX1 (normal), 10% (107/1056) had a CT scan.

Although limited to only three potential pathways, the data and reporting format from this paper³⁵ are useful to inform conceptual model parameters for current practice with no AI software. Future studies to identify the number of CT referrals made after CXR review with and without AI software assistance, stratified and reported by clinical pathway for both symptomatic (suspected lung cancer) and incidental (no lung cancer suspected) primary care population, are required. Ideally these would be of prospective study design, but hospital-reported data could be used to retrieve this information retrospectively in the incidental primary care population.

- Number of people referred for follow-up CXR

Two studies^{36,37} reported information on follow-up CXR following initial CXR results. In the study by Woznitza *et al.*,³⁷ all patients with CXRs reported as showing pneumonia had a follow-up CXR suggested in 4 to 6 weeks to ensure resolution (17/522, 3%). Where a follow-up CXR was suggested, four (22%) were performed with a mean time from initial to follow-up CXR of 33.8 days (range 10–49 days). In the 13 other cases, follow-up CXRs were not done at the same institution and the authors assumed that they had not been undertaken as no reminders were sent.

Bradley *et al.*³⁶ reported follow-up CXRs performed based on result codes of 2129 initial GP-requested CXRs. Of the 376 patients who had an initial 'negative' result (codes 3 and 4), 98 (26.1%) had at least one further CXR. Of the 370 patients with an initial abnormal finding where non-urgent further review or investigation was advised (code 2), 191 (56.1%) had a second CXR. The median duration to second CXR was 42 days [interquartile range (IQR) 28–57 days]. In total, 324 (15.2%) patients across all CXR result codes (1–4) had at least two CXRs before diagnosis.³⁶

These studies^{36,37} are informative of CXR resource use across multiple pathways, which is useful to consider in future modelling. While Woznitza *et al.*³⁷ had only a relatively small sample size in their feasibility study, it was a prospective design and for this measure reported specifically for those on a clinical pathway following 'abnormal' (other diagnosis) CXR results. This would be pathway B (see [Clinical pathway for representation in model](#)) in the conceptual model. It illustrates that the number of people referred for follow-up CXR does not necessarily equate to resource use, as patient uptake rate is also a factor.

- Number of cancers missed/detected

One (ineligible) study²² from the clinical effectiveness review reported the mean number of cancers detected and found no significant differences with and without AI software (54 cancers, 95% CI 42 to 59 cancers, and 46 cancers, 95% CI 38 to 51 cancers, respectively).

Among the 1687 CXR referrals in the study by Woznitza *et al.*,³⁷ 17 patients were missed who were subsequently diagnosed with lung cancer: 15 were given normal CXR results and 2 were given abnormal (alternative diagnosis) results.

Among 8682 CXR referrals in the Woznitza *et al.*³⁸ study, 48 of the 49 lung cancers diagnosed were detected. The single case that was missed was diagnosed on a subsequent emergency attendance for upper limb deep-vein thrombosis.³⁸

Foley *et al.*³⁵ reported the number of cancers diagnosed by CX code: CX1, 10/1056 (1%); CX2, 29/288 (10%); and CX3, 49/72 (68%). Ten people with lung cancer were given false-negative 'normal' results but were still referred for CT and received diagnosis. Data on the other 949 patients with negative CXR results who were not referred for CT would be informative (although difficult to obtain) to give the total number of false-negative results by CX1 code for use in modelling. Similar information would be required for the CX2 result patients.

Future studies with extended follow-up and use of patient-level hospital-reported data linked to cancer registries would facilitate access to and reporting of this information. These may also provide data on stage at diagnosis.

Both numbers of false negatives (i.e. lung cancers missed) and stage at diagnosis may be important outcome measures for use in evidence linkage. This could be used in modelling to inform any association between time to diagnosis and stage shift and to assign appropriate costs and QoL outcomes by stage at diagnosis.

- Stage of cancer at detection

Bradley *et al.*³⁶ found that 1490 (70%) of the 2129 patients in their study were diagnosed with lung cancer at stage III/IV. Across the four CXR codes used to stratify results of initial CXR, these were reported as (1) 981 (70.9%), (2) 259 (70%), (3) 147 (63.9%) and (4) 103 (70.5%). There was no evidence of a statistically significant association between CXR result and stage at diagnosis.³⁶

Foley *et al.*³⁵ also found no statistical difference between CXR result and stage at diagnosis. Those with advanced stage (IIIc/IV) at diagnosis were reported as CX1, 5 (50%); CX2, 11 (38%); and CX3, 28 (57%) ($p = 0.26$). This was a much smaller sample size than in the Bradley *et al.*³⁶ study, and advanced stage was defined as IIIc/IV³⁵ rather than III/IV.³⁶

Findings from both studies^{35,36} showed that a majority of patients with normal or abnormal CXR results have advanced stage disease at diagnosis.

- Time to CXR report

Time to CXR report was highly dependent on trust and service provided. Most had a same-day reporting facility for GP-requested plain CXR films. Woznitza *et al.*³⁸ reported on the RadioX trial that compared immediate reporting and standard reporting to find median report time to CXR report (termed TAT in this paper).

This may be a more informative measure than TAT per scan as it has a more direct impact on the speed at which a CT scan is requested.

- Time to CT scan

Time from CXR to CT scan was reported in two studies retrieved from the cost-effectiveness search.^{35,37} Foley *et al.*³⁵ found a significant difference in the number of days from CXR to CT by CX result code. The reported mean days were 34.6 for those with CX1, normal but a CT scan is recommended to exclude malignancy; 19.6 for CX2, alternative diagnosis; and 1.9 for CX3, suggestive of cancer.

By contrast, the feasibility study by Woznitza *et al.*³⁷ looked at the time from CXR to CT scan by reporting strategy for those with a CXR result suggestive of lung cancer. Those whose CXR image was reported immediately had a mean of 0.9 days ($n = 22$, SD 2.3 days) until CT scan compared with routine reporting (mean 10.6 days, SD 4.5 days; $p > 0.0001$).

Although these cannot be compared directly, the results of Woznitza *et al.*³⁷ are for the equivalent result population of the CX3 in the Foley *et al.*³⁵ study. This shows significant variation in time from CXR to CT scan as a result of department reporting practices alone. In the Foley *et al.*³⁵ study there were GP reporting sessions for consultants on most days (Jonathan Rodrigues, personal communication), suggesting that this was more in line with the standard reporting process in the study by Woznitza *et al.*³⁷ However, many other procedural variables between the two radiology departments are likely to have an impact on these times.

This highlights the need for the real-world clinical context to be taken into consideration when generating future evidence to inform these measures. This is relevant for studies of outcomes after CXR both with and without AI, as there are only limited data even in current practice, which is difficult to generalise because of variation both within and between NHS trusts.

The results from Foley *et al.*³⁵ are useful for modelling purposes as they establish a difference in time between three diverging clinical pathways, up to the point of confirmatory testing by CT scan. This may support evidence linkage to outcomes further in the lung cancer management pathway.

- Time to diagnosis

Four studies³⁵⁻³⁸ from the cost-effectiveness review reported time to diagnosis. In three,^{35,36,38} this was calculated as the date of the initial CXR to the date the diagnosis was confirmed (either the date of the diagnostic test or the date on which a clinical diagnosis was confirmed by the lung cancer MDT if no pathological sample was taken). In the smallest of the studies, Woznitza *et al.*³⁷ used date of radiological diagnosis confirmed at MDT. The results of histological diagnosis were reported separately, but no data on timing for this were provided.

Foley *et al.*³⁵ found a significant difference in time to diagnosis between CX codes, with a mean of 89.7 days for those with CX1, normal but a CT scan is recommended to exclude malignancy; 65.3 days for those with CX2, alternative diagnosis; and 30.2 days for those with CX3, suggestive of cancer.

Bradley *et al.*³⁶ also reported time to diagnosis by initial CX codes but used median number of days: code 1, suspicion of lung cancer identified/urgent investigation needed, 36 (IQR 23–63) days; code 2, abnormality identified/non-urgent investigation indicated including diagnoses of pneumonia or consolidation even if repeat imaging was not explicitly suggested, 93 (IQR 55–154) days; code 3, abnormality identified but no further investigation/assessment indicated, 211 (IQR 181–296) days; and code 4, normal CXR, no abnormalities identified, 193 (IQR 87–279) days. When calculated by author-defined 'positive' (codes 1 and 2) and 'negative' (codes 3 and 4), time to diagnosis was 43 (IQR 27–78) days and 204 (IQR 105–287) days, respectively.³⁶

Woznitza *et al.*³⁸ presented both mean and median days to diagnosis for those who had IR and those who had SR of their CXR image. The mean number of days was 47.2 (SD 35.8) for IR and 81.6 (78.5) for SR. When median days to diagnosis of 32 (IQR 19–70) for IR and 63 (29–78) for SR were analysed, statistical significance was found ($p = 0.03$).³⁸

Woznitza *et al.*³⁷ also looked at mean time to diagnosis for IR and SR, with study findings of 4.1 and 10.6 days, respectively. However, this was for a small sample of 11 patients, and as discussed this was for radiological diagnosis at MDT only and so did not account for additional waiting time due to biopsy.³⁷

All four studies³⁵⁻³⁸ reported substantial variation in time to diagnosis, demonstrating that this outcome measure can be affected by multiple factors, including CXR result and the subsequent diagnostic pathway followed^{35,36} and different reporting practices in a radiology department.^{37,38} Establishing that AI software has an impact on time to diagnosis beyond fluctuating departmental factors, examining the mechanism by which that impact is produced (by increasing test accuracy, reducing report TAT or other means) and quantifying the impact would require a prospective study, in real-world settings, ideally across multiple sites in the UK.

Once established, change in time to diagnosis may support evidence linkage to outcomes further in the lung cancer management pathway.

- Ease of use/acceptability of the software by clinicians

In the UK study,²² 10 of the 11 clinicians responded to questions about acceptability of the AI. Eighty per cent stated that reporting was not slower when using AI, and 90% stated that the AI 'heatmaps' produced were 'helpful to understand the algorithm's attention points'.

Clinical outcomes for consideration may include:

- morbidity
- mortality.

Costs will be considered from an NHS and PSS perspective.

Costs for consideration may include:

- cost of each AI software available for this indication
- costs of training staff to use software
- costs associated with healthcare professional time to read and report CXR
- costs of diagnostic testing and treatment.

Sources of cost and resource use inputs are discussed in *question 5, What are the cost and resource use considerations relating to the use of adjunct AI to detect lung cancer.*

Summary

Importantly, the EAG did not identify any studies in the clinical effectiveness review that met the inclusion criteria and addressed the outcomes for discussion, highlighting the gap in evidence for all measures of AI software to inform cost-effectiveness analysis at this time.

Four studies^{35–38} from the cost-effectiveness review looking at CXR referral from primary care referrals without AI software informing model parameters, but all had limitations on their applicability due to study type and reported outcomes.

The use of hospital-reported data to conduct retrospective studies shows promise to provide good-quality information on outcomes under current CXR review practices without AI software. The reporting of consistently defined, key clinical pathway outcomes by standardised CXR report codes would allow a comparison between studies and provide more straightforward translation for use in cost-effectiveness modelling.

Coordinated research efforts are required to generate research on all outcome measures identified in for inclusion in the conceptual model. Evidence needs to demonstrate impact on intermediate outcomes over a sustained period of time in the NHS environment to account for differences in outcomes due to the widespread variation in current practices and pathways between individual hospitals sites and trusts. This can be achieved through well-designed studies, with large sample sizes, conducted over a sufficient period to capture the main outcomes of interest. This would reduce the reliance on evidence linkage, which remains particularly weak with regard to impact on stage at diagnosis.

Question 4

What would a health economic model to estimate the cost-effectiveness of adjunct AI to detect lung cancer look like?

This section describes a conceptual model developed by the EAG to identify the structure and components required in any future health economic models estimating the cost-effectiveness of adjunct AI compared with radiologist or reporting radiographer review alone of CXR images to detect lung cancer.

The proposed structure is suitable for both symptomatic and incidental primary care populations referred by their GP for CXR, with certain model parameters varying where appropriate for the specific population. For use in decision-making in the UK setting, an NHS and PSS perspective is adopted.

The conceptual model follows the illustrative pathways shown in [Figure 7](#).

Strategies

For people undergoing a CXR, the CXR image is read by either a radiology specialist alone (current usual practice) or a radiology specialist with adjunct AI software.

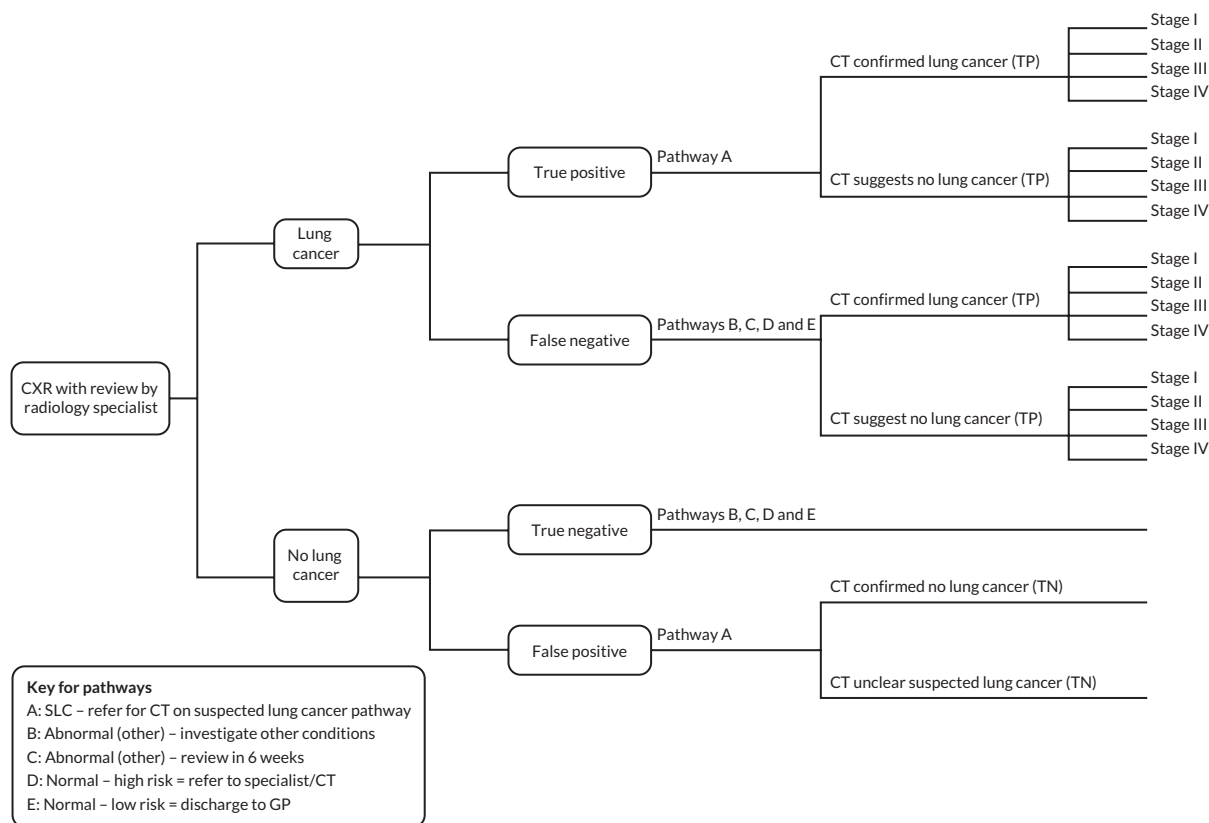


FIGURE 7 Illustrative model structure for the detection of lung cancer. The illustrative pathway for CXR review by radiology specialist with adjunct AI software is identical to the structure presented here for CXR review by radiology specialist alone. If AI was used for triage, an additional step prior to the CXR could be included.

Proposed model structure

A decision tree structure is used to depict the pathway from CXR imaging and review to point of diagnosis. We considered a decision tree structure appropriate to capture the short-term costs and benefits associated with the strategies used to identify people with lung cancer.

A positive CXR result (findings suspicious of lung cancer) follows pathway A, where a CT scan confirms the positive result and provides provisional staging. A utility decrement is applied to a positive result lasting until treatment. Treatment according to stage at diagnosis then commences when utility values for that stage are attributed for true-positive cases. False-positive cases revert to general population utility values.

People with false-negative results follow pathway B, C, D or E depending on whether findings are reported as ‘normal’ or ‘abnormal (alternative diagnosis)’. Either these people eventually undergo a CT scan as part of further clinical investigations along their respective pathways or they are assumed to present at an emergency department later. Any false negatives not detected at first CT scan along any pathway are also assumed to present later as an emergency. These pathways are longer than the most direct route to diagnosis (pathway A), and it is assumed that the delay in time to diagnosis confers a stage shift for a proportion of these people. Treatment then commences by stage at diagnosis.

People who receive a false-positive result at CXR imaging also follow pathway A and go on to receive a CT scan as a minimum further investigation, with a proportion undergoing further testing (e.g. PET scan, biopsy, bronchoscopy) until a true-negative lung cancer diagnosis is confirmed. A temporary utility decrement is applied for a false-positive test result for the duration until a confirmatory test is received showing no lung cancer present. A utility decrement associated with further invasive diagnostic

procedures (biopsy and bronchoscopy) is applied to people with true-positive results and a proportion of those with false-positive results.

As for those people with false-negative results, people with true-negative results follow pathway B, C, D or E depending on whether findings are reported as 'normal' or 'abnormal (alternative diagnosis)'. Additional testing is specific to each pathway.

Pathways A, B, C, D and E are described in detail in [Clinical pathway for representation in model](#). Within the model, separate costs and health-related quality-of-life (HRQoL) outcomes are assigned to each pathway. All pathways that lead to a diagnosis of lung cancer complete the decision tree at a fast-track lung cancer clinic. Total costs and HRQoL outcomes (expressed as QALYs) to point of diagnosis are accrued according to the proportion of people assigned to each pathway as a result of CXR review by the two strategies under comparison.

At the end of the decision tree branches, long-term treatment costs and utility values over a 5-year time horizon are assigned based on stage of lung cancer at diagnosis. These are added to those accumulated during the diagnostic component of the model to provide overall outcomes for each strategy.

The results of a subsequent analysis (in a fully executable model) would be presented in terms of an ICER, where the difference between total costs of CXR review by radiology specialists with and without adjunct AI is divided by the difference between total QALYs for each, to give a cost-per-QALY figure. Prices would be based on the current cost year, with discounting of cost and outcomes applied at 3.5% over the total model time horizon in line with the NICE reference case.¹⁸

For the conceptual model, information is required about the prevalence of lung cancer and the performance of radiology specialists to detect findings indicative of lung cancer on review of CXRs both with and without AI used as an adjunct. These inputs are specific to the population of interest, so figures are required for prevalence and diagnostic accuracy in both symptomatic and incidental primary care populations.

Prevalence figures used in the literature are sourced by Bajre *et al.*³² for use in their economic evaluation from Field *et al.*⁴⁴ and by Geppert *et al.*⁴⁵ for modelling in the DAP060 AI for chest CT diagnostic assessment review from Horeweg *et al.*⁴⁶ Both sources^{44,46} contain estimates of lung cancer prevalence in the screening population. For modelling purposes in Bajre *et al.*³² and Geppert *et al.*,⁴⁵ these prevalence estimates are assumed to be the same for their population of interest. The EAG did not find any more relevant sources, but searches were not exhaustive and more recent estimates of prevalence in the UK population would be advisable for use in future modelling.

For the specific clinical pathways (A, B, C, D and E), people may follow through the decision tree, information on costs and resource use of diagnostic tests and clinical management input is required. The proportion of people taking each pathway and the mean time from initial CXR to diagnosis is also required for each of these pathways, under each strategy.

An example using clinical pathway A

Chest X-ray findings suggestive of malignancy are flagged to secondary care lung cancer teams who request the CT scan and await the formal referral from the patient's GP to the suspected lung cancer clinic. Once reported, CT scans are triaged by lung cancer team consultants. If the scan suggests probable lung cancer, an urgent lung cancer team appointment is arranged with appropriate tests, for example lung function tests and planned biopsy (endobronchial ultrasound) (Alberto Alonso, personal communication). Diagnosis, histopathological staging and treatment options are then discussed at the fast-track lung cancer MDT clinic.¹⁶

This process incurs the cost per person of a CT scan (£153), lung function tests (£285) and biopsy (£1670).⁴⁷

Input to direct these further tests is required by the secondary care lung cancer team on two occasions: (1) to review CXR results, refer for CT scan and notify the GP to make a suspected lung cancer pathways referral; and (2) to review CT scan results and refer for lung function tests and biopsy prior to fast-track lung cancer MDT clinic review. The unit cost of a lung cancer MDT meeting (£146),⁴⁷ or part thereof, would be assigned for encounters with the lung cancer team and the fast-track clinic team. Average times to discuss a case during these meetings are necessary for more accurate costing.

Similarly, utility decrements are also assigned to pathway A. Suspicious lung cancer findings on CXR attract a disutility of -0.063 ⁴⁸ applied over the length of time until confirmatory diagnosis. A disutility of -0.2 is applied for biopsy investigation for a period of 3 months.^{49,50}

The total costs and QALYs accrued are then attributed to the proportion of people in the model who take pathway A as a true-positive or false-positive case ([Table 8](#)).

The conceptual model presented captures the following important outcomes in the diagnostic process.

Clinical outputs from the model:

- number of false positives
- number of additional CT scans
- number of people referred for follow-up CXR
- number of people identified as 'normal' (no lung cancer present) and discharged
- number of cancers missed and detected
- proportion of cancers detected at each stage.

Long-term outcomes from the model:

- total costs per strategy
- total QALYs per strategy
- costs per QALY.

These outcomes would be based on a cohort of 1000 patients entering the model.

Question 5

What are the cost and resource use considerations relating to the use of adjunct artificial intelligence to detect lung cancer?

This section identifies the costs and resource use of adding AI software to CXR review taking an NHS and PSS perspective. Costs are required of each AI software, costs of training staff to use software, resource use and costs associated with healthcare professional time to read and report CXR, and costs of diagnostic testing and treatment. All costs are presented in 2021 prices. Costs obtained from the literature were updated to current prices using the Hospital and Community Health Services index from *Unit Costs of Health and Social Care 2021*.⁵¹ Cost categories are listed with resource use considerations discussed alongside and any potential sources of information identified.

Cost of software

AI software costs were obtained directly from the companies. Five of the 14 companies identified in the final scope⁹ were registered as stakeholders in this EVA and provided cost information to the EAG via NICE communications (Annalise AI, Behold AI, Infervision, Lunit Inc. and Siemens Healthineers).

TABLE 8 Input parameters required to populate conceptual model

Parameter	Population			
	Primary care symptomatic		Primary care incidental	
	Value	Source	Value	Source
Prevalence				
Radiology review without AI software				
Sensitivity				
Specificity				
Radiology review with AI software				
Sensitivity				
Specificity				
Proportion of people following each pathway after CXR review without AI software				
Pathway A				
Pathway B				
Pathway C				
Pathway D				
Pathway E				
Proportion of people following each pathway after CXR review with AI software				
Pathway A				
Pathway B				
Pathway C				
Pathway D				
Pathway E				
Proportion of lung cancers diagnosed at stages I, II, III and IV after CXR review without AI software				
	Stage I	II	III	IV
Pathway A				
Pathway B				
Pathway C				
Pathway D				
Pathway E				
Proportion of lung cancers diagnosed at stages I, II, III and IV after CXR review with AI software				
	Stage I	II	III	IV
Pathway A				
Pathway B				
Pathway C				
Pathway D				
Pathway E				
Utility values for lung cancer diagnosed at stages I, II, III and IV				
	Stage I	II	III	IV
Utility value				

Pricing structures were either fixed annual subscription fees (Annalise AI, Behold AI, Infervision and Siemens Healthineers) or volume-based annual pricing tiers (Infervision and Lunit Inc.). All companies charge a one-off implementation fee in the first year, which covers installation, integration with existing Picture Archive and Communication System/Radiology Information System and staff training. Ongoing subscription costs are renewable on an annual basis, with fees covering software licensing, annual maintenance, support and updates. Pricing is calculated per trust by Annalise AI, Infervision, Lunit Inc. and Siemens Healthineers. By contrast, Behold AI's implementation and subscription fees are per hospital, with a 30,000 annual CXR volume allocation. (Confidential information has been removed.)

The annual subscription cost depends on the volume of CXRs to be processed in either each trust (Annalise AI, Infervision, Lunit Inc. and Siemens Healthineers) or each hospital (Behold AI) annually. The resource use would, therefore, be determined by the number of primary care referrals for CXR for the symptomatic and incidental populations per year.

This information is available through trust databases and has been reported in the literature through retrospective database studies.^{35,36}

[Table 9](#) shows the disaggregated costs of AI software by company based on a number of 25,000 CXR images per NHS trust.

Cost of staff training

Staff training is provided by the AI software companies and the cost is included in the one-off implementation fee (see [Cost of software](#)). Companies reported that the training time for radiologists/reporting radiographers was 1 hour for Lunit and 30 minutes for Infervision. For Behold AI, no training time was given. Instead, the company advised that a training deck is customised for each trust and used to train designated trainers from each organisation, and then the deck is given to the trainers so that they can provide training to their radiologists.

Under the assumption that training is undertaken during protected staff-training time within radiology departments, no further costs would be attributed beyond the implementation fee.

Cost of staff time to read and report chest X-ray

The hourly cost of a radiologist or reporting radiographer was obtained from the literature. Two methods were identified that had been used in previous economic evaluations.^{32,52} In the first of these, Bajre *et al.*³² used the figure of £156 per hour for a radiologist and £53 for a band 7 reporting radiographer. This was originally calculated by Lockwood⁵² based on salary, on-costs and education for the 2015–6 cost year. In the second economic evaluation, the hourly cost of a band 9 radiographer (£147) from *Unit Costs of Health and Social Care 2021*⁵³ was used as a proxy for a radiologist.⁴⁵

The cost of staff time to read and report a single CXR can then be calculated using TAT. Published evidence has suggested no statistically significant difference in reading times of CXRs between readers with and readers without AI.^{24,25} However, these data are from two studies^{24,25} that do not meet the inclusion criteria for the present EVA, and are of uncertain applicability to clinical practice (see [What are the practical implications of adjunct artificial intelligence software to detect lung cancer on chest X-rays?](#)).

Feedback from the SCM suggested timings without the use of AI from 1 minute on average, faster for normal and slower for very abnormal, up to 5 minutes. From the literature, Bajre *et al.*³² assumed a 2-minute reporting time for both radiologists and reporting radiographers.

Cost of further diagnostic tests

Following the initial CXR, further testing may be required. This could include additional CXR, CT scan of chest, CT scan of abdomen (performed with or without contrast), PET scan, bronchoscopy and biopsy, with various combinations of each possible.

TABLE 9 Costs of AI software by company based on a number of 25,000 CXR images per NHS trust

Company, technology name (tech use)	One-off set-up cost/ implementation fee	Annual subscription (based on number of 25,000 images)	Cost per exam	Total first year cost	Indicative cost per image (non-discounted) (5 years average)
Annalise AI, Annalise Enterprise and Triage (CADE and CAST)	£5000–25,000	£51,250 ^a	N/A	£66,250	£2.17
Behold.ai, Red Dot (CADE and CAST)	Confidential information has been removed ^b	Confidential information has been removed ^b	N/A	Confidential information has been removed	Confidential information has been removed
Infervision InferRead DR (CADE)	Confidential information has been removed	Confidential information has been removed (licence fee) Confidential information has been removed (maintenance fee)	N/A Confidential information has been removed	Confidential information has been removed ^c	Confidential information has been removed
Lunit Inc. Lunit, INSIGHT CXR (CADE)	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Siemens Healthineers, AI-RAD (CADx)	£2400	£12,000 ^a	N/A	£14,400	£0.50

CADE, computer-aided detection; CADx, computer-aided diagnosis; CAST, computer-aided simple triage.

^a Based on tier pricing of 'up to' 25,000 images per year.

^b Per-hospital cost.

^c Cost per image price structure (includes one-off implementation fee and annual maintenance fee).

To direct these further tests, clinical input from the GP, respiratory specialists, radiologists and appropriate MDTs is required. The costs of these services can be obtained from *National Schedule of NHS Costs 2020/21*⁴⁷ and *Unit Costs of Health and Social Care 2021*.⁵³

The costs of further tests depend on outcomes along the clinical pathway, including the:

- number of people referred for a CT scan
- number of people referred for follow-up CXR
- number of people identified as 'normal' (no lung cancer present) and discharged
- number of cancers missed/detected
- stage of cancer at detection.

No evidence was identified in the clinical search that addressed these outcomes as a result of AI software assistance in the reading of CXRs. We therefore have no evidence with which to determine whether the use of adjunct AI will increase, decrease, or not affect the number of people requiring additional testing.

Cost of treatment (including costs of any adverse events)

Total treatment costs are assigned according to stage of disease.

Several sources were identified in the literature. Bajre *et al.*³² and Geppert *et al.*⁴⁵ used Cancer Research UK 2014⁵⁴ values originally reported in the 2014–5 price year and included cost of retreatment.

Snowsill *et al.*³³ used figures based on a 2-year costing approach, with index year costs from a UK teaching hospital⁵⁵ and second year costs estimated from the index year using a subsequent year ratio from database analysis in England.⁵⁶ The same authors in an interim update to the UK National Screening Committee³⁴ also used a 5-year microcosting approach with resource use based on the most recent National Lung Cancer Audit secondary care estimates for those in the 55 to 75 years age range to reflect more modern available treatment options, including immunotherapies.

[Table 10](#) summarises the costs required for the proposed model.

For use in any future modelling, all sources where costs are obtained from the literature will need to be updated to current prices at the time using the Hospital and Community Health Services index from the most recent version of the Unit Costs of Health and Social Care.

Summary

Potential sources to inform all unit costs for the cost parameters in the conceptual model proposed by the EAG have been identified. Primarily, these costs can be obtained from the literature and published national index costs and directly from AI software companies.

Evidence to support resource use relating to adjunct AI to detect lung cancer was not identified. Therefore, the total values of cost inputs for all cost parameters could not be calculated.

No evidence was found to determine what, if any, effect AI will have on resource use and in what direction this might take with respect to costs. At this stage, all we are able to determine is that AI represents a new cost, as AI software needs to be purchased and used in addition to the costs and resources consumed in the current clinical pathway.

TABLE 10 Costs required for the proposed model

Parameter	Value (£)	Source
Healthcare professional		
GP consultation	39	<i>Unit Costs of Health and Social Care 2021</i> ⁵³ (per-patient contact of 9.22 minutes)
Radiologist consultation	147	<i>Unit Costs of Health and Social Care 2021</i> ⁵³ [cost per working hour (£147) for a band 9 radiographer as a proxy for a radiologist]
MDT	146	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (CDMT_OTH other cancer MDT meetings)
Other tests		
X-ray	45	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (direct access plain film)
CT scan (single area, with contrast)	153	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (RD21A – computerised tomography scan of one area, with post-contrast, 19 years and over)
CT scan of two areas, without contrast	£127	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (RD23Z – computerised tomography scan of two areas, without contrast)
CT scan of two areas, with contrast	153	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (RD24Z – computerised tomography scan of two areas, with contrast)
Guided-needle biopsy	1670	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (DZ71Z – minor thoracic procedure, guided-needle biopsy)
Lung function tests	285	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (DZ52Z – full lung function testing)
Bronchoscopy	1679	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ (DZ70Z – endobronchial ultrasound examination of mediastinum)
PET scan	1161	<i>National Schedule of NHS Costs 2020/21</i> ⁴⁷ RN01a – PET-CT of one area, 19 years and over
Treatment		
Stage I	20,928	UK National Screening Committee external review: interim report ³⁴
Stage II	29,757	
Stage III	32,830	
Stage IV	21,838	

Results of potential budget impact assessment

Five companies provided cost information to NICE as part of the DAP request for information process, and all responded to the EAG's clarifying questions (Annalise AI, Behold AI, Infervision, Lunit Inc. and Siemens Healthineers). This provided more certainty in the EAG's calculation of the costs of these technologies. In total, there was sufficient information for six different price estimates (Infervision provided two different pricing structure options).

AI software is intended as an adjunct to the existing CXR review process conducted by a qualified radiology specialist. The ultimate diagnostic decision is made by the radiology specialist, the cost of which is assumed to be constant in both current and future practice if AI software were to be implemented. This assumption was made as no evidence was found in our review of any change in resource use due to AI software. This being the case, only the additional costs of AI software are considered here.

As discussed in [Proposed model structure](#), no available evidence was found to inform any changes to progression through the clinical pathway due to the intervention in this population. Therefore, onward health-related service use, diagnostic and treatment costs are assumed to stay the same for the purposes of this analysis. However, for the purposes of any future modelling, costs that may need to be considered include CT scans, CT surveillance for lung nodules detected, further invasive tests, for example biopsy, and treatment for different stages of lung cancer at diagnosis.

Change in test accuracy may result in increased sensitivity with AI software assistance, potentially identifying more cancers/nodules, or decreased specificity (i.e. because of an increase in false positives) wherein more people could be referred for a CT scan, with an associated cost implication.

Several studies were retrieved during the literature search that appeared to provide sufficient data from which the budget impact at an individual institution level could be calculated.³⁵⁻³⁸ These studies have previously been summarised. The study by Foley *et al.*³⁵ was chosen to base the budget impact case on as the trust-wide annual referral number for the appropriate populations (both suspected lung cancer and incidental primary care) was clearly provided. It was also not restricted to those who had a confirmed diagnosis of lung cancer, as in the Bradley *et al.*³⁶ study, and the authors responded to clarifying questions from the EAG on contact to ensure greater accuracy in the interpretation of the study results.

Foley *et al.*³⁵ conducted a retrospective review of 16,945 CXRs referred from primary care and performed across all sites at the Royal United Hospitals Bath NHS Foundation Trust between 1 June 2018 and 31 May 2019; 1488 of these were referred for suspected lung cancer.

On contact with corresponding authors, annual GP referral data for CXR to the Royal United Hospitals Bath NHS Foundation Trust, including a breakdown of those referred for suspected lung cancer, were provided to the EAG for the period January to December, 2019 to 2022, inclusive (Richard Wood, PACS Manager, Royal United Hospitals Bath NHS Foundation Trust, 6 February 2023, personal communication). The EAG intended to calculate budget impact estimates based on these exact numbers, including those reported in the study³⁵ for 2018, as an example of the first 5 years of AI software implementation at single NHS trust. However, due to substantial variation in the numbers referred during this time as services were impacted by the COVID-19 pandemic, the EAG decided to use a conservative assumption whereby the annual referral number from primary care was kept constant at 16,945 over the 5 years for this analysis.

Results are presented in [Table 11](#), with anticipated budget impact at NHS trust level for both symptomatic (suspected lung cancer) and incidental primary care population CXR referrals shown in the final column.

Results are presented in [Table 12](#), with anticipated budget impact over the first 5 years at NHS trust level shown separately for symptomatic, incidental and total primary care population CXR referrals.

The budget impact estimate for the whole primary care population referred for CXR are not expected to be the sum of the symptomatic and incidental populations ([Table 12](#)). This is due to the assumption during calculation that AI software is only approved for use in that specific subpopulation. The use of volume-based pricing structures also means that the cost of AI software implementation and use over 5 years would be the same with (confidential information has been removed) for each of the symptomatic and incidental populations alone as it would be for the whole primary care population.

Summary

Budget impact results vary greatly between companies, but the EAG cautions against direct comparison, as the AI software presented has varying capabilities, and some may be used in different positions early in the diagnostic pathway. For example, Siemens AI software points to a region of interest on the

TABLE 11 Anticipated budget impact of AI software at NHS trust level for all GP-referred CXR

Company, technology name (tech use)	One-off set-up cost/ implementation fee	Annual subscription (based on 16,945 images)	Cost per examination	Total first year cost (VAT applied at 20%)	Cost over first 5 years (non-discounted) (based on 16,945 images per year) (VAT applied at 20%)
Annalise AI, Annalise Enterprise and Triage (CADE and CAST)	£5000–25,000	£51,250 ^a	N/A	£66,250 (assuming mean implementation fee) (£79,500)	£271,250 (£325,500)
Behold.ai, Red Dot (CADE and CAST)	Confidential information has been removed	Confidential information has been removed	N/A	Confidential information has been removed	Confidential information has been removed
Infervision, InferRead DR (CADE)	Confidential information has been removed	Confidential information has been removed (licence fee) Confidential information has been removed (maintenance fee)	N/A Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Lunit Inc., Lunit INSIGHT CXR (CADE)	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Siemens Healthineers, AI-RAD (CADx)	£2400	£12,000 ^a	N/A	£14,400 (£17,280)	£62,400 (£74,880)

CADE, computer-aided detection; CADx, computer-aided diagnosis; CAST, computer-aided simple triage.

^a Based on tier pricing of 'up to' 25,000 images per year.

TABLE 12 Anticipated budget impact of AI software at NHS trust level for symptomatic, incidental and whole-population GP-referred CXR

Company, technology name (tech use)	Cost over first 5 years for symptomatic primary care population	Cost over first 5 years for incidental primary care population	Cost over first 5 years for all primary care population referrals
Annalise AI, Annalise Enterprise and Triage (CADe and CAST)	NDA	£325,500	£325,500
Behold.ai, Red dot (CADe and CAST)	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
InferVision, InferRead DR (CADe)	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Lunit Inc., Lunit INSIGHT CXR (CADe)	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Siemens Healthineers, AI-RAD (CADx)	£26,880	£74,880	£74,880

CADe, computer-aided detection; CADx, computer-aided diagnosis; CAST, computer-aided simple triage; NDA, no data available.

Note

Non-discounted costs, VAT included at rate of 20%. Total population, $N = 16,945$; symptomatic population, $n = 1488$; incidental population $n = 15,457$.

CXR, whereas Annalise AI software identifies a specific location, gives characteristics of the anomaly on CXR and provides a preliminary diagnosis and rating of confidence when used in a concurrent review of images. Similarly, Behold AI and Annalise AI software can provide a triage of CXR images prior to radiology specialist review in order to prioritise reporting, as well as assist with the detection of abnormalities and diagnosis. These differing capabilities may affect the way the AI software is used in practice, with a variety of practical, clinical and cost implications later in the diagnostic pathway. Therefore, without future modelling it is unclear how budget impact estimates for different AI software brands might be comparable.

Chapter 6 Discussion

Statement of principal findings

Test accuracy, practical implications and clinical effectiveness

No studies met the review inclusion criteria. There is currently no evidence, applicable to this review, on the use of adjunct AI software for the detection of suspected lung cancer on CXR in either people referred from primary care with symptoms of lung cancer or people referred from primary care for other reasons. This finding, however, satisfies the secondary aim of this review, which was to identify evidence gaps in this field and inform future research. This is discussed in more detail below.

To provide context to the decision problem, summary results were presented from six studies that did not meet the review inclusion criteria because of unclear populations but were selected for discussion post hoc. The referral status and symptom status of the study participants are unknown, but the studies did provide comparisons of CXRs read by radiologists with and without the use of commercial AI software. Few outcomes were reported in these studies. They provide some insight into two of the key questions of this EVA:

1. What is the test accuracy and test failure rate of adjunct AI software to detect lung on CXR?
2. What are the practical implications of adjunct AI software to detect lung cancer on CXR?

None of the studies provided evidence for the clinical effectiveness of adjunct AI software applied to CXR (question 3).

For question 1, one study reported a higher sensitivity for lung cancer detection by readers with adjunct AI compared with readers alone, with no difference in specificity or cancer detection rate.²² In the four studies for which data are available for publication, no significant between-group differences were found in test accuracy metrics in relation to lung nodules.²³⁻²⁶

For question 2, no significant between-group differences were found for reading time^{24,25} or hypothetical referrals for CT scan.^{22,24} Data from one study indicated that clinicians generally responded positively to the use of AI software.²²

This synopsis of study results is illustrative only of the type of evidence that is currently available on commercial AI to aid the interpretation of CXR. Caution is required in extrapolating from these studies as not only did they not meet the review inclusion criteria, but there were also differences between the studies, and limitations within them. For example, some studies included nodules with differing levels of detection difficulty from easy to challenging, while others excluded images on which nodules were below a certain size; studies used retrospective designs; data were reported from the mean performance across several readers with varying degrees of experience; readers had their findings from the first reading present at the time of the second reading and there was a lack of detailed reporting of key results. There were also differences in the reference standards, making comparisons between studies difficult. Furthermore, generalisability to the UK primary care referred population is unclear in all six summarised studies, and generalisability to the UK population overall is likely to be low in three studies that were conducted in the Republic of Korea.

Conceptual cost-effectiveness modelling

The conceptual modelling process aimed to explore both the structure and evidence requirements for parameter inputs for future model development. There was no evidence available on AI software impact on any of the intermediate outcomes identified to inform parameterisation. Results of EAG searches for evidence to inform these outcomes for the comparator alone (i.e. radiology specialist review of CXR in

the detection of lung cancer in the primary care population) varied by study design and the way outcome measures were reported, limiting the way data could be used.

A simplistic model structure was outlined due to the paucity of evidence and tentative links to long-term outcomes.

Key points:

- Artificial intelligence needs to show changes to intermediate outcomes over a sustained period of time in the NHS environment as pathway variation and clinical practice/structure in radiology departments vary considerably between trusts and individual sites. Unless evidence is produced that is statistically powered to account for a difference in outcomes due to the current variation, evidence linkage to improved outcomes that may demonstrate cost-effectiveness cannot be made. Ideally this would be in the form of a large-scale, multisite, UK-based clinical trial with AI software as an adjunct to radiology specialist review compared directly with existing practice.
- It is not clear that evidence to suggest stage shift in detection of lung cancer can be achieved through CXR identification of suspected lung cancer in any event.

Strengths and limitations

Strengths

- Extensive searches were undertaken, including electronic databases, existing reviews, company submissions and known studies, which reduced the risk of missing studies.
- Clinical experts were involved in the review and asked to provide details of any potentially eligible studies.

Limitations of the review/early value assessment process

- This review employed rapid evidence synthesis methods.⁵⁷ While this approach is used internationally by policy-makers to make expedited assessments of evidence,^{58,59} it is not without risks. In the present review, one reviewer conducted all elements of the review in full (i.e. title/abstracting sifting, full-text assessment), with a second reviewer assessing/checking 20% of each review task. Therefore, 80% of review tasks were only conducted by a single reviewer. Any errors made by the first reviewer relating to this 80% would not have been detected. As a result, there is the possibility that eligible studies were missed.
- We only searched for and included studies published in the English language. Therefore, we do not know if there are relevant papers in other languages.
- Targeted searches were used to retrieve a manageable number of records to screen. Therefore, it is possible that some studies (e.g. broad reviews) were not retrieved. To counter this, we used different combinations of concepts, sources and search methods and tested the overall search strategy's ability to retrieve a set of known studies (found by a variety of methods during the scoping stage).
- Owing to the abridged timescale and limitations in resource for the evidence reviewing processes of this pilot EVA, there was no opportunity to follow up any uncertainties in studies with their authors or to seek further clarification of responses received from the few companies that provided submissions. Additional time was required to clarify the complex eligibility criteria in the scope before the protocol was signed off, and this also impacted on the reviewing timescale.
- As no studies met the eligibility criteria for the review, a pragmatic decision was taken following discussions with NICE to apply additional criteria to the excluded studies to select evidence closest to the review eligibility criteria. This selection process was iterative and involved discussion between two reviewers but was undertaken in the absence of a priori defined criteria. As already discussed,

studies summarised were those where the population referral route and symptom status for the CXR were unknown (not reported). These populations are likely to be no different from those in other excluded studies with better descriptions of their populations. In addition, only summary results were extracted, and no formal risk-of-bias tool was applied to these studies. These results are illustrative only, and the results do not provide evidence on the use of adjunct AI software for the detection of suspected lung cancer on CXR in people referred from primary care.

- The selection of cost-effectiveness studies was undertaken by one reviewer, with wide inclusion/exclusion criteria aimed at the pragmatic identification of literature to support the development of a conceptual model and inform a rudimentary BIA of AI software in the NHS in the UK. Therefore, it is possible that without the rigorous methodology of systematic review processes (including quality appraisal), there may be biases from an individual reviewer, and studies identified in this report may not be fully representative of all those available. Through additional searches of references lists of identified studies, publication bibliographies of relevant authorship, several targeted searches and liaison with specialist committee members and clinical experts, the EAG endeavoured to mitigate the risk of missing pertinent evidence for this report.

Limitations of evidence base

This review found no applicable evidence on which to assess AI software for analysing CXR to identify suspected lung cancer among people referred from primary care.

Uncertainties

This review aimed to assess the test accuracy and test failure rates of adjunct AI software to detect lung cancer or lung nodules on CXR, the practical implications and the clinical effectiveness of adjunct AI software in people referred from primary care, and to develop a conceptual model. No evidence was found on any of these. Therefore, uncertainties remain regarding all review questions.

The evidence that was summarised to provide some insight into the above was limited to 3 of the 14 eligible interventions. No eligible evidence was identified for the following AI software: Annalise CXR (annalise.ai), Auto Lung Nodule Detection (Samsung), ChestLink Radiology Automation (Oxipit), ChestView (GLEAMER), CXR (Rayscape), ClearRead Xray – Detect (Riverain Technologies), InferRead DR Chest (Infervision), Milvue Suite (Milvue), qXR (Qure.ai), SenseCare-Chest DR Pro (SenseTime) and VUNO Med-CXR (VUNO).

Resource use associated with progression through clinical pathways was highly uncertain because of the lack of evidence and difficult to establish for CXR alone (owing to the large number and complexity of clinical pathways possible for a diagnosis of lung cancer). Costs of individual elements in the pathway were sourced from published sources used in previous technology assessments, but without robust resource use data the certainty in overall cost estimates is limited. Long-term treatment costs, calculated by stage at diagnosis, are widely used in the literature, with recent updates to these. However, there is only weak and limited evidence to suggest CXR to stage shift at diagnosis.

Owing to the lack of evidence for all inputs, only a simple conceptual model could be attempted. This by necessity underestimates the complexity of the pathways and creates uncertainty as to whether this would be the optimum modelling to address the practical implications question.

Equality, diversity and inclusion

We know that an equitable, diverse and inclusive research group is a more innovative and successful one. Therefore, we integrate equality, diversity and inclusion (EDI) across our workforce, our review products and our academic output. We embrace diversity of background, perspective, culture

and experience, and, together with our university and health and social care partners, we work to address inequity.

We provide our team a range of opportunities at different career stages and different levels of commitment and provide implicit bias training for all team members. We provide flexible research training and opportunities for innovative methodological design work so that everyone can engage in methods development, irrespective of circumstances and career stage. We expect that all line managers and mentors have supervising/mentoring training and can provide confidential and non-judgmental support.

We have built on our strong institutional inclusion and diversity policies to maximise participation of people from traditionally marginalised groups, and to identify and overcome any barriers to developing a supportive culture for new researchers, including encouragement of flexible work arrangements where relevant.

The University of Warwick holds silver Athena Swan charter status. Our EDI policies are regularly reviewed, and awareness is promoted through newsletters and weekly circulars. Warwick Evidence proactively harnesses the research capacity development resources within the university (e.g. mentoring, reverse mentoring, shadowing, strengths profiling) and aligns these with NIHR Academy systems.

Patient and public involvement

The short timeline of this EVA meant that there was insufficient time to engage patient and public advisors. However, the NICE specialist committee for this assessment included patient representatives who were involved in defining the scope.

Chapter 7 Conclusions

Test accuracy, practical implications and clinical effectiveness

No applicable evidence was found to answer the questions about the impact of adjunct AI on test accuracy, its practical implications or its clinical effectiveness for either of the two populations of interest (people referred from primary care with symptoms of lung cancer and people referred from primary care for other reasons).

Cost-effectiveness

Only a simple cost-effectiveness conceptual model structure was feasible due to a lack of evidence to support all inputs along the lung cancer detection pathway. Complexity and variation in these pathways were found across individual institutions, which necessitated considerable simplification to achieve a nationally representative framework to model.

The conceptual model uses evidence linkage to improved outcomes that may demonstrate cost-effectiveness through stage shift in detection of lung cancer. This is an assumption made in this population, as there is no clear evidence that stage shift can be achieved through CXR identification of suspected lung cancer at this time.

Unit costs for cost parameters in the conceptual model were readily identified; however, evidence to inform resource use change due to adjunct AI to detect lung cancer was not found. Therefore, total values of cost inputs for all cost parameters could not be calculated.

The only determination possible at this point is that AI represents a new cost, as AI software needs to be purchased and used in addition to the costs and resources consumed in the current clinical pathway.

Implications for service provision

There is widespread variation in existing service provision both within and across trusts. Changes in departmental practices alone have been shown to have an impact on outcome measures along the lung cancer diagnostic pathway and have been used positively to try to improve lung cancer diagnosis times.

No evidence was identified in this review to suggest the impact that AI software as an adjunct to CXR review might have on any stage of the diagnostic pathway.

With a complete lack of applicable evidence on AI software, the impact on service provision is unknown but may have significant implications for progression through diagnostic pathways, resource use, costs and patient outcomes.

Suggested research priorities

Given the absence of any eligible evidence on the topic of this EVA, the below research priorities are suggested to enable an assessment of the impact that AI would have on lung cancer detection and longer-term outcomes. They are presented in the order that the outcomes would occur in the patient care pathway. Retrospective study designs could serve as a starting point to determine the potential of adjunct AI software for analysing CXR images to identify suspected lung cancer, but these would not

provide sufficient, unbiased evidence of its impact. This would require prospective studies, the least biased of which are randomised controlled trials.

- Assessment of the test accuracy of specialist radiologist with adjunct AI software compared with specialist radiologist without AI software, conducted with participants who reflect those seen in clinical practice. Evidence within these studies should also provide data on the types/characteristics of cancers and nodules that are detected by AI, and the test failure rates of AI. Ideally, this information would come from prospective studies with follow-up of test-negative cases.
- Assessment of the effects that adjunct AI software has on clinical decision-making, and its acceptability to clinicians.
- Assessment of the effects that adjunct AI software has on intermediate outcomes such as time to CT scan and time to diagnosis.
- Assessment of the clinical effectiveness of adjunct AI software to reduce patient mortality and morbidity and to improve HRQoL.
- Linked assessment of intermediate outcomes along the lung cancer detection pathway from CXR review, time to CT scan, time to diagnosis and stage at diagnosis both with and without the use of AI software. This could be achieved initially for outcomes prior to the introduction of AI software by using retrospective audit data from NHS trusts with sufficient data to link these outcomes in the target population.
- Prospective randomised controlled trials to capture intermediate outcomes would be the favoured methodology to determine the impact of AI software at specified points along the lung cancer pathway, with a study size sufficiently large to account for variations in pathways in current clinical practice and establish support for evidence linkage.
- Studies that evaluate QoL outcomes for people diagnosed with lung cancer by stage of the disease in the UK population.

Additional information

Contributions of authors

Jill Colquitt (<https://orcid.org/0000-0001-5962-2689>) (Honorary Research Fellow) conducted the clinical effectiveness review.

Mary Jordan (<https://orcid.org/0000-0002-0497-8634>) (Research Fellow) contributed to the cost-effectiveness review and developed the conceptual model.

Rachel Court (<https://orcid.org/0000-0002-4567-2586>) (Information Specialist) developed the search strategy and undertook searches.

Emma Loveman (<https://orcid.org/0000-0001-8226-2634>) (Honorary Research Fellow) conducted the clinical effectiveness review.

Janette Parr (<https://orcid.org/0000-0002-0629-1596>) (Research Associate) conducted the clinical effectiveness review.

Iman Ghosh (<https://orcid.org/0000-0002-7073-7468>) (Research Associate) conducted the clinical effectiveness review.

Peter Auguste (<https://orcid.org/0000-0001-5143-3218>) (Assistant Professor) contributed to the cost-effectiveness review and developed the conceptual model.

Mubarak Patel (<https://orcid.org/0000-0001-7573-1447>) (Research Associate) developed the analysis plan.

Chris Stinton (<https://orcid.org/0000-0001-9054-1940>) (Senior Research Fellow) led the project.

All authors were involved in developing the protocol and writing the draft and final version of the report.

Acknowledgements

The authors would like to thank NICE Specialist Committee Members for their expert clinical advice, the radiology department, respiratory medicine department and lung cancer team at the Royal United Hospitals Bath NHS Foundation Trust for their data provision and advice on departmental practices, and Dr Dan Todkill for comments on the report. This Early Value Assessment was conducted against a background of industrial action in universities across the UK. This reduced the time available for this project.

Data-sharing statement

All available data can be obtained by contacting the corresponding author.

Ethics statement

This EVA consists of secondary research; therefore, ethics approval was not required.

Information governance statement

The University of Warwick is committed to handling all personal information in line with the UK Data Protection Act (2018) and the General Data Protection Regulation (UK GDPR). Under the Data Protection legislation, the University of Warwick is the Data Controller, and you can find out more about how we handle personal data, including how to exercise your individual rights and the contact details for our Data Protection Officer here: <https://warwick.ac.uk/services/legalandcomplianceservices/dataprotection/>

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at <https://doi.org/10.3310/LKRT4721>.

Primary conflicts of interest: Janette Parr is funded through a National Institute for Health and Care Research Applied Research Collaboration West Midlands PhD Studentship (number R.MRPT.1110).

References

1. National Institute for Health and Care Excellence. *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*. London: NICE; 2022. URL: www.nice.org.uk/guidance/hte12/documents/final-protocol (accessed 26 January 2023).
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. *Overview: Lung Cancer*. 2022. URL: www.nhs.uk/conditions/lung-cancer/ (accessed 14 November 2022).
4. Cancer Research UK. *Lung Cancer Incidence by Age*. 2021. URL: www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence#heading-One (accessed 23 November 2022).
5. Cancer Research UK. *Lung Cancer*. 2019. URL: www.cancerresearchuk.org/about-cancer/lung-cancer (accessed 14 November 2022).
6. NHS. *NHS Long Term Plan*. 2021. URL: www.longtermplan.nhs.uk/ (accessed 14 November 2022).
7. IBM Cloud Education. *What is artificial intelligence (AI)*. 2020. URL: www.ibm.com/uk-en/cloud/learn/what-is-artificial-intelligence (accessed 14 November 2022).
8. 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**:53. <https://doi.org/10.1007/s00138-020-01101-5>
9. National Institute for Health and Care Excellence. *Artificial Intelligence Software for Analysing Chest X-ray Images to Identify Suspected Lung Cancer: Final Scope*. London: NICE; 2022. URL: www.nice.org.uk/guidance/hte12/documents/final-scope (accessed 23 November 2022).
10. National Institute for Health and Care Excellence. *Lung Cancer: Diagnosis and Management*. NICE Guideline [NG122]. London: NICE; 2022. URL: www.nice.org.uk/guidance/ng122 (accessed 14 November 2022).
11. National Institute for Health and Care Excellence. *Suspected Cancer: Recognition and Referral*. NICE Guideline [NG12]. London: NICE; 2021. URL: www.nice.org.uk/guidance/ng12 (accessed 14 November 2022).
12. National Institute for Health and Care Excellence, King's Technology Evaluation Centre. *Artificial Intelligence for Analysing Chest X-ray Images: Medtech Innovation Briefing [MIB292]*. London: NICE; 2022. URL: www.nice.org.uk/advice/mib292 (accessed 15 November 2022).
13. 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**:65. <https://doi.org/10.1038/s41746-021-00438-z>
14. The Joanna Briggs Institute. *JBICritical 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 Pulmonary Nodule Guideline Development Group. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;**70**(Suppl. 2):1–54. <https://doi.org/10.1136/thoraxjnl-2015-207168>
17. NHS England. *Updated National Optimal Lung Cancer Pathway – Sept 2020*. British Thoracic Oncology Group; 2020. URL: 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]. London: NICE; 2022. URL: www.nice.org.uk/process/pmg36/ (accessed 23 November 2022).
19. York Health Economics Consortium. *Evidence Standards Framework for Digital Health Technologies: Cost Consequences and Budget Impact Analyses and Data Sources*. National Institute for Health and Care Excellence; 2021. URL: www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/budget-impact-guide.pdf (accessed 28 November 2022).
20. Mauskopf JA, Sullivan SD, Annemans L, Caro J, Mullins CD, Nuijten M, *et al.* Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices – budget impact analysis. *Value Health* 2007;**10**:336–47. <https://doi.org/10.1111/j.1524-4733.2007.00187.x>
21. van Beek EJR, Ahn JS, Kim MJ, Murchison JT. Validation study of machine-learning chest radiograph software in primary and emergency medicine. *Clin Radiol* 2022;**78**:25. <https://doi.org/10.1016/j.crad.2022.08.129>
22. Dissez G, Tay N, Dyer T, Tam M, Dittrich R, Doyne D, *et al.* *Enhancing Early Lung Cancer Detection on Chest Radiographs with AI-Assistance: A Multi-Reader Study [Preprint]*. arXiv.org; 2022. URL: <https://arxiv.org/ftp/arxiv/papers/2208/2208.14742.pdf> (accessed 3 January 2023).
23. Nam JG, Hwang EJ, Kim DS, Yoo SJ, Choi H, Goo JM, Park CM. Undetected lung cancer at posteroanterior chest radiography: potential role of a deep learning-based detection algorithm. *Radiol Cardiothorac Imaging* 2020;**2**:e190222. <https://doi.org/10.1148/ryct.2020190222>
24. Jang S, Song H, Shin YJ, Kim J, Kim J, Lee KW, *et al.* Deep learning-based automatic detection algorithm for reducing overlooked lung cancers on chest radiographs. *Radiology* 2020;**296**:652–61. <https://doi.org/10.1148/radiol.2020200165>
25. Koo YH, Shin KE, Park JS, Lee JW, Byun S, Lee H. Extravalidation and reproducibility results of a commercial deep learning-based automatic detection algorithm for pulmonary nodules on chest radiographs at tertiary hospital. *J Med Imaging Radiat Oncol* 2021;**65**:15–22. <https://doi.org/10.1111/1754-9485.13105>
26. Homayounieh F, Digumarthy S, Ebrahimian S, Rueckel J, Hoppe BF, Sabel BO, *et al.* An artificial intelligence-based chest X-ray model on human nodule detection accuracy from a multicenter study. *JAMA Netw Open* 2021;**4**:e2141096. <https://doi.org/10.1001/jamanetworkopen.2021.41096>
27. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, *et al.* The 'laboratory' effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 2008;**249**:47–53. <https://doi.org/10.1148/radiol.2491072025>
28. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, *et al.* *Global Cancer Observatory: Cancer Today*. Lyon: International Agency for Research on Cancer; 2020. URL: <https://gco.iarc.fr/today/home> (accessed 14 February 2023).

29. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2017;2:MR000033. <https://doi.org/10.1002/14651858.MR000033.pub3>
30. Kim Y. *Prospective Evaluation of Deep Learning-Based Detection Model for Chest Radiographs in Outpatient Respiratory Clinic*. 2020. URL: https://cris.nih.go.kr/cris/search/detailSearch.do?sessionId=7C81738AEC5C7255DDD1DE8836613390?seq=20017&search_page=L&search_lang=E&lang=E (accessed 1 March 2023).
31. Avery G. *A Study to Assess the Impact of an Artificial Intelligence (AI) System on Chest X-ray Reporting*. 2022. URL: www.clinicaltrials.gov/study/NCT05489471 (accessed 1 March 2023).
32. Bajre MK, Pennington M, Woznitza N, Beardmore C, Radhakrishnan M, Harris R, McCrone P. Expanding the role of radiographers in reporting suspected lung cancer: a cost-effectiveness analysis using a decision tree model. *Radiography (Lond)* 2017;23:273–8. <https://doi.org/10.1016/j.radi.2017.07.011>
33. Snowsill T, Yang H, Griffin E, Long L, Varley-Campbell J, Coelho H, et al. Low-dose computed tomography for lung cancer screening in high-risk populations: a systematic review and economic evaluation. *Health Technol Assess* 2018;22:1–276. <https://doi.org/10.3310/hta22690>
34. Exeter Test Group and Health Economics Group. *Interim Report on the Cost-Effectiveness of Low Dose Computed Tomography (LDCT) Screening for Lung Cancer in High Risk Individuals*. London: UK National Screening Committee; 2022. URL: <https://view-health-screening-recommendations.service.gov.uk/document/586/download> (accessed 17 January 2023).
35. Foley RW, Nassour V, Oliver HC, Hall T, Masani V, Robinson G, et al. Chest X-ray in suspected lung cancer is harmful. *Eur Radiol* 2021;31:6269–74. <https://doi.org/10.1007/s00330-021-07708-0>
36. Bradley SH, Bhartia BS, Callister ME, Hamilton WT, Hatton NLF, Kennedy MP, et al. Chest X-ray sensitivity and lung cancer outcomes: a retrospective observational study. *Br J Gen Pract* 2021;71:e862–8. <https://doi.org/10.3399/bjgp.2020.1099>
37. Woznitza N, Piper K, Rowe S, Bhowmik A. Immediate reporting of chest X-rays referred from general practice by reporting radiographers: a single centre feasibility study. *Clin Radiol* 2018;73:507.e1–8. <https://doi.org/10.1016/j.crad.2017.11.016>
38. Woznitza N, Ghimire B, Devaraj A, Janes SM, Piper K, Rowe S, et al. Impact of radiographer immediate reporting of X-rays of the chest from general practice on the lung cancer pathway (RadioX): a randomised controlled trial. *Thorax* 2022;78:890–4. <https://doi.org/10.1136/thorax-2022-219210>
39. Dwyer-Hemmings L, Fairhead C. The diagnostic performance of chest radiographs for lung malignancy in symptomatic primary-care populations: a systematic review and meta-analysis. *BJR Open* 2021;3:20210005. <https://doi.org/10.1259/bjro.20210005>
40. Naik H, Howell D, Su J, Qiu X, Brown C, Vennettilli A, et al. Stage specific health utility index scores of Canadian cancer patients. *J Clin Oncol* 2015;33(15_Suppl.):6614. https://doi.org/10.1200/JCO.2015.33.15_SUPPL.6614
41. Bradley SH, Abraham S, Callister ME, Grice A, Hamilton WT, Lopez RR, et al. Sensitivity of chest X-ray for detecting lung cancer in people presenting with symptoms: a systematic review. *Br J Gen Pract* 2019;69:e827–35. <https://doi.org/10.3399/bjgp19X706853>
42. Lawson MH, Underhill S, Chauhan M, Robinson S, Melesi V, Miller S, Beckett P. P13 Mapping the lung cancer pathway. *Thorax* 2021;76(Suppl. 1):A92. <https://doi.org/10.1136/thorax-2020-BTSabstracts.158>

43. Royal College of Radiologists. *Radiology Reporting Figures for Service Planning*. 2022. URL: www.rcr.ac.uk/system/files/publication/field_publication_files/radiology-reporting-figures-2022.pdf (accessed 24 January 2023).
44. Field JK, Duffy SW, Baldwin DR, Brain KE, Devaraj A, Eisen T, *et al*. The UK lung cancer screening trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. *Health Technol Assess* 2016;**20**:1–146. <https://doi.org/10.3310/hta20400>
45. Geppert J, Auguste P, Asgharzadeh A, Ghiasvand H, Patel M, Brown A, *et al*. *Software with Artificial Intelligence Derived Algorithms for Automated Detection and Analysis of Lung Nodules from CT Scan Images: A Diagnostics Assessment Report*. London: National Institute for Health and Care Excellence; 2022. URL: www.nice.org.uk/guidance/GID-DG10046/documents/diagnostics-assessment-report (accessed 14 February 2023).
46. 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**:1332–41. [https://doi.org/10.1016/s1470-2045\(14\)70389-4](https://doi.org/10.1016/s1470-2045(14)70389-4)
47. National Schedule of NHS Costs. *The Main Schedule, Showing Data for the Whole Range of Services Provided by Provider, Including Admitted Patient Care on a Finished Consultant Episode (FCE) Basis*. London: NHS England; 2022. URL: www.england.nhs.uk/wp-content/uploads/2022/07/2_National_schedule_of_NHS_costs_FY20-21.xlsx (accessed 14 February 2023).
48. Mazzone PJ, Obuchowski N, Fu AZ, Phillips M, Meziane M. Quality of life and healthcare use in a randomized controlled lung cancer screening study. *Ann Am Thorac Soc* 2013;**10**:324–9. <https://doi.org/10.1513/AnnalsATS.201301-007OC>
49. Sutton AJ, Sagoo GS, Jackson L, Fisher M, Hamilton-Fairley G, Murray A, Hill A. Cost-effectiveness of a new autoantibody test added to Computed Tomography (CT) compared to CT surveillance alone in the diagnosis of lung cancer amongst patients with indeterminate pulmonary nodules. *PLOS ONE* 2020;**15**:e0237492. <https://doi.org/10.1371/journal.pone.0237492>
50. Stevenson M, Lloyd-Jones M, Morgan MY, Wong R. Non-invasive diagnostic assessment tools for the detection of liver fibrosis in patients with suspected alcohol-related liver disease: a systematic review and economic evaluation. *Health Technol Assess* 2012;**16**:1–174. <https://doi.org/10.3310/hta16040>
51. Jones K, Burns A. The NHS cost inflation index (NHSCII). *Unit Costs of Health and Social Care*. Canterbury: Personal Social Services Research Unit, University of Kent; 2021. URL: www.pssru.ac.uk/pub/uc/uc2021/sourcesofinformation.pdf (accessed 19 April 2024).
52. Lockwood P. An economic evaluation of introducing a skills mix approach to CT head reporting in clinical practice. *Radiography* 2016;**22**:124–30. <https://doi.org/10.1016/j.radi.2015.09.004>
53. Jones K, Burns A. *Unit Costs of Health and Social Care 2021*. Canterbury: Personal Social Services Research Unit, University of Kent; 2021. <https://doi.org/10.22024/UniKent/01.02.92342>
54. Birtwistle M, Earnshaw A. *Saving Lives, Averting Costs: An Analysis of the Financial Implications of Achieving Earlier Diagnosis of Colorectal, Lung and Ovarian Cancer*. Incisive Health, Cancer Research UK; 2014. URL: <https://news.cancerresearchuk.org/2014/09/22/saving-lives-and-averting-costs-the-case-for-earlier-diagnosis-just-got-stronger/> (accessed 14 February 2023).
55. Kennedy MP, Hall PS, Callister ME. Factors affecting hospital costs in lung cancer patients in the United Kingdom. *Lung Cancer* 2016;**97**:8–14. <https://doi.org/10.1016/j.lungcan.2016.04.009>
56. McGuire A, Martin M, Lenz C, Sollano JA. Treatment cost of non-small cell lung cancer in three European countries: comparisons across France, Germany, and England using administrative databases. *J Med Econ* 2015;**18**:525–32. <https://doi.org/10.3111/13696998.2015.1032974>

57. Taylor-Phillips S, Geppert J, Stinton C, Freeman K, Johnson S, Fraser H, *et al.* Comparison of a full systematic review versus rapid review approaches to assess a newborn screening test for tyrosinemia type 1. *Res Synth Methods* 2017;**8**:475–84. <https://doi.org/10.1002/jrsm.1255>
58. Hailey DM. Health technology assessment in Canada: diversity and evolution. *Med J Aust* 2007;**187**:286–8. <https://doi.org/10.5694/j.1326-5377.2007.tb01245.x>
59. Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, *et al.* Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. *Int J Technol Assess Health Care* 2008;**24**:133–9. [Erratum in: *Int J Technol Assess Health Care*. 2008 Summer;**24**(3):369]. <https://doi.org/10.1017/s0266462308080483>

Appendix 1 Literature searches

Sections of the appendix have been reproduced from the protocol,¹ available from the NICE website. © NICE 2022 *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*. Available from www.nice.org.uk/guidance/hte12/documents/final-protocol. All rights reserved. Subject to Notice of rights.

NICE guidance is prepared for the NHS in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication.

Test accuracy, practical implications and clinical effectiveness

TABLE 13 Search strategy summary: multistranded, targeted approach

Search#	Search	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))	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))	Epistemonikos, MEDLINE, EMBASE, CDSR, a computer science database
3	Intervention (AI and chest x-ray) AND selected outcomes (lung cancer/lung nodule)	MEDLINE, EMBASE, CENTRAL (including 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)	MEDLINE, EMBASE, CENTRAL (including trial register records), a computer science database
	Targeted searches for relevant ongoing systematic reviews	PROSPERO
	Targeted searches for relevant ongoing trials	WHO ICTRP
	Check references of relevant reviews and studies found via NICE and team members' scoping or clinical experts	NICE, EAG team members, clinical experts

Bibliographic databases

TABLE 14 Bibliographic databases: search summary

Source(s)	Date searched	Purpose	Description of search	Hits	Notes
MEDLINE (via Ovid)	25 November 2022	Search to identify relevant reviews and primary studies	The four targeted searches run together (see Table 13)	1119	Limited to English language or no language specified. Non-human studies, letters, editorials, comments removed No date limits applied
EMBASE (via Ovid)	29 November 2022	Search to identify relevant reviews and primary studies	The four targeted searches run together (see Table 13)	2198	Limited to English language or no language specified. Non-human studies, letters, editorials, removed No date limits applied
Cochrane Database of Systematic Reviews (via Wiley)	30 November 2022	Search to identify relevant reviews for reference checking	Intervention (AI and CXR) OR Intervention [broader] (AI) AND lung cancer/lung nodule	0	Specialist database for Cochrane systematic reviews
Cochrane CENTRAL (via Wiley)	30 November 2022	Search to identify relevant primary studies	Intervention (AI and CXR) AND lung cancer or lung nodule OR Technology names/companies AND (CXR/lung cancer/lung nodule)	52	Specialist database for trials No date or language limits applied
Epistemonikos	1 December 2022	Search to identify relevant reviews for reference checking	Intervention (AI and CXR) OR Intervention [broader] (AI) AND lung cancer/lung nodule	45	Specialist database for systematic reviews and overviews. Filtered for publication types: <ul style="list-style-type: none"> • systematic review • broad synthesis No date or language limits applied.
ACM Digital Library	1 December 2022	Search to identify relevant reviews and primary studies in a computer science database	Intervention (AI and CXR) OR Intervention [broader] (AI) AND lung cancer/(lung nodule)	12	Limited to Content Type: Review article No date or language limits applied
			Intervention (AI and CXR) AND lung cancer/lung nodule	452	No limits applied
			Technology names/companies AND (CXR)/lung cancer/lung nodule	1	No limits applied

Totals

Total from databases: 3879

Total after duplicates removed: 3049

MEDLINE (via Ovid)

Searched 25 November 2022

Ovid MEDLINE® ALL 1946 to 23 November 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/ 160,931
- 2 ai.kf,tw. 39,919
- 3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 124,190
- 4 exp Neural Networks, Computer/ 53,917
- 5 (neural network* or convolutional or CNN or CNNs).kf,tw. 90,349
- 6 exp Diagnosis, Computer-Assisted/ 86,384
- 7 Pattern Recognition, Automated/ 26,362
- 8 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 33,565
- 9 (support vector machine* or random forest* or black box learning).kf,tw. 37,636
- 10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 [AI] 396,688
- 11 exp Radiography, Thoracic/ 40,528
- 12 X-Rays/ 31,129
- 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. 66,459
- 14 11 or 12 or 13 [CXR] 121,772
- 15 10 and 14 [AI and CXR] 3865
- 16 limit 15 to "reviews (best balance of sensitivity and specificity)" [AI and CXR and Reviews] 349
- 17 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 288,007
- 18 (systematic* adj3 (review* or overview* or search or literature)).mp. 328,557
- 19 17 or 18 459,498
- 20 15 and 19 [AI and CXR and SRs] 40
- 21 16 or 20 [AI and CXR and Reviews/ SRs] 360
- 22 exp Lung Neoplasms/ or Solitary Pulmonary Nodule/ 268,336
- 23 ((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. 326,364
- 24 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 946
- 25 (sclc or nsclc).kf,tw. 64,440
- 26 22 or 23 or 24 or 25 [Lung Cancer/ Nodule] 398,150
- 27 10 and 26 [AI and Lung Cancer/ Nodule] 6749
- 28 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 288,007
- 29 (systematic* adj3 (review* or overview* or search or literature)).mp. 328,557
- 30 28 or 29 [SRs] 459,498
- 31 27 and 30 [AI and Lung Cancer/ Nodule and SRs] 100
- 32 10 and 14 and 26 [AI and CXR and Lung Cancer/ Nodule] 707
- 33 AI-Rad Companion Chest X-ray*.kf,tw,in. 1
- 34 Annalise CXR*.kf,tw,in. 1
- 35 Auto Lung Nodule Detection*.kf,tw,in. 0

- 36 ChestView*.kf,tw,in. 0
- 37 (Chest X-Ray Classifier* or Quibim*).kf,tw,in. 46
- 38 CheXVision*.kf,tw,in. 0
- 39 (ClearRead Xray* adj2 Detect).kf,tw,in. 0
- 40 InferRead DR Chest*.kf,tw,in. 0
- 41 JLD-02K*.kf,tw,in. 0
- 42 Lunit INSIGHT CXR*.kf,tw,in. 4
- 43 Milvue Suite*.kf,tw,in. 0
- 44 ChestEye Quality*.kf,tw,in. 0
- 45 (qXR* or Qure*).kf,tw,in. 6815
- 46 (red dot* or behold*).kf,tw,in. 1090
- 47 SenseCare-Chest DR Pro*.kf,tw,in. 0
- 48 VUNO Med-Chest X-Ray*.kf,tw,in. 0
- 49 (X1* and Visionary Health).kf,tw,in. 0
- 50 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49
[Technology Names/ Companies] 7956
- 51 50 and 14 [Technology Names/ Companies and CXR] 61
- 52 50 and 26 [Technology Names/ Companies and Lung Cancer/ Nodules] 90
- 53 51 or 52 [Technology Names/ Companies and CXR/ Lung Cancer/ Nodules] 136
- 54 21 or 31 or 32 or 53 1190
- 55 limit 54 to english language 1134
- 56 limit 54 to no language specified 0
- 57 55 or 56 1134
- 58 exp animals/ not humans.sh. 5,066,999
- 59 57 not 58 1128
- 60 limit 59 to (comment or editorial or letter) 9
- 61 59 not 60 1119

EMBASE (via Ovid)

Searched 29 November 2022

EMBASE Classic+EMBASE 1947 to 2022 week 47

- 1 exp artificial intelligence/ or exp machine learning/ 373,033
- 2 ai.kf,tw. 55,274
- 3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 146,615
- 4 (neural network* or convolutional or CNN or CNNs).kf,tw. 108,457
- 5 computer assisted diagnosis/ or computer assisted radiography/ 44,996
- 6 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 44,987
- 7 (support vector machine* or random forest* or black box learning).kf,tw. 46,703
- 8 1 or 2 or 3 or 4 or 5 or 6 or 7 [AI] 530,438
- 9 exp thorax radiography/ 230,425
- 10 X ray/ 119,143
- 11 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 107,803
- 12 9 or 10 or 11 [CXR] 379,945
- 13 8 and 12 [AI and CXR] 5577
- 14 limit 13 to "reviews (best balance of sensitivity and specificity)" [AI and CXR and Reviews] 657

- 15 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 414,514
- 16 (systematic* adj3 (review* or overview* or search or literature)).mp. 520,359
- 17 15 or 16 695,121
- 18 13 and 17 [AI and CXR and SRs] 117
- 19 14 or 18 [AI and CXR and Reviews/ SRs] 678
- 20 exp lung tumor/ or lung nodule/ 495,858
- 21 ((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. 493,166
- 22 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 1328
- 23 (sclc or nsclc).kf,tw. 116,762
- 24 20 or 21 or 22 or 23 [Lung Cancer/ Nodule] 655,493
- 25 8 and 24 [AI and Lung Cancer/ Nodule] 12,931
- 26 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 414,514
- 27 (systematic* adj3 (review* or overview* or search or literature)).mp. 520,359
- 28 26 or 27 [SRs] 695,121
- 29 25 and 28 [AI and Lung Cancer/ Nodule and SRs] 313
- 30 8 and 12 and 24 [AI and CXR and Lung Cancer/ Nodule] 1114
- 31 AI-Rad Companion Chest X-ray*.kf,tw,in. 1
- 32 Annalise CXR*.kf,tw,in. 1
- 33 Auto Lung Nodule Detection*.kf,tw,in. 0
- 34 ChestView*.kf,tw,in. 0
- 35 (Chest X-Ray Classifier* or Quibim*).kf,tw,in.57
- 36 CheXVision*.kf,tw,in. 0
- 37 (ClearRead Xray* adj2 Detect).kf,tw,in. 0
- 38 InferRead DR Chest*.kf,tw,in. 0
- 39 JLD-02K*.kf,tw,in. 0
- 40 Lunit INSIGHT CXR*.kf,tw,in. 6
- 41 Milvue Suite*.kf,tw,in. 0
- 42 ChestEye Quality*.kf,tw,in. 0
- 43 (qXR* or Qure*).kf,tw,in. 14,268
- 44 (red dot* or behold*).kf,tw,in. 1520
- 45 SenseCare-Chest DR Pro*.kf,tw,in. 0
- 46 VUNO Med-Chest X-Ray*.kf,tw,in. 0
- 47 (X1* and Visionairy Health).kf,tw,in. 0
- 48 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 [Technology Names/ Companies] 15,850
- 49 48 and 12 [Technology Names/ Companies and CXR] 267
- 50 48 and 24 [Technology Names/ Companies and Lung Cancer/ Nodules] 234
- 51 49 or 50 [Technology Names/ Companies and CXR/ Lung Cancer/ Nodules] 466
- 52 19 or 29 or 30 or 51 2362
- 53 limit 52 to english language 2271
- 54 limit 52 to no language specified 1
- 55 53 or 54 2272
- 56 animal experiment/ not (human experiment/ or human/) 2,472,698
- 57 55 not 56 2263
- 58 limit 57 to (editorial or letter) 65
- 59 57 not 58 2198

Cochrane Database of Systematic Reviews (via Wiley)

Search name: qXR EVA Reviews

Date run: 30 November 2022 19:30:29

ID Search Hits

- #1 [mh "artificial intelligence"] OR [mh "machine learning"] OR [mh "deep learning"] OR [mh "supervised machine learning"] OR [mh "support vector machine"] OR [mh "unsupervised machine learning"] 1540
- #2 ai:ti,ab,kw 5002
- #3 ((artificial OR machine OR deep) NEAR/5 (intelligence OR learning OR reasoning)):ti,ab,kw 3847
- #4 [mh "Neural Networks, Computer"] 217
- #5 (("neural" NEXT network*) OR convolutional OR CNN OR CNNs):ti,ab,kw 1738
- #6 [mh "Diagnosis, Computer-Assisted"] 1943
- #7 [mh ^"Pattern Recognition, Automated"] 193
- #8 ((automat* OR autonomous OR "computer aided" OR "computer assisted") NEAR/3 (detect* OR identif* OR diagnos*)):ti,ab,kw 2092
- #9 (("support vector" NEXT machine*) OR ("random" NEXT forest*) OR "black box learning"):ti,ab,kw 935
- #10 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 13,357
- #11 [mh "Radiography, Thoracic"] 363
- #12 [mh ^X-Rays] 59
- #13 (((chest OR lung* OR thora*) NEAR/3 (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*):ti,ab,kw 5878
- #14 #11 OR #12 OR #13 5948
- #15 #10 AND #14 120
- #16 [mh "Lung Neoplasms"] OR [mh ^"Solitary Pulmonary Nodule"] 8755
- #17 ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) NEAR/3 (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*)):ti,ab,kw 28,597
- #18 ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") NEAR/4 (tumor* OR tumour* OR syndrome*)):ti,ab,kw 17
- #19 (sclc OR nsclc):ti,ab,kw 12,248
- #20 #16 OR #17 OR #18 OR #19 29,193
- #21 #10 AND #20 348
- #22 #15 OR #21 421
- Cochrane Reviews: 0

CENTRAL (via Wiley)

Search Name: qXR EVA Trials

Date run: 30 November 2022 22:52:13

Comment: 30 November 2022

ID Search Hits

- #1 [mh "artificial intelligence"] OR [mh "machine learning"] OR [mh "deep learning"] OR [mh "supervised machine learning"] OR [mh "support vector machine"] OR [mh "unsupervised machine learning"] 1540

- #2 ai:ti,ab,kw 5002
- #3 ((artificial OR machine OR deep) NEAR/5 (intelligence OR learning OR reasoning)):ti,ab,kw 3847
- #4 [mh "Neural Networks, Computer"] 217
- #5 (("neural" NEXT network*) OR convolutional OR CNN OR CNNs):ti,ab,kw 1738
- #6 [mh "Diagnosis, Computer-Assisted"] 1943
- #7 [mh ^"Pattern Recognition, Automated"] 193
- #8 ((automat* OR autonomous OR "computer aided" OR "computer assisted") NEAR/3 (detect* OR identif* OR diagnos*)):ti,ab,kw 2092
- #9 (("support vector" NEXT machine*) OR ("random" NEXT forest*) OR "black box learning"):ti,ab,kw 935
- #10 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 13,357
- #11 [mh "Radiography, Thoracic"] 363
- #12 [mh ^X-Rays] 59
- #13 ((chest OR lung* OR thora*) NEAR/3 (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film* OR CXR*)):ti,ab,kw 5878
- #14 #11 OR #12 OR #13 5948
- #15 [mh "Lung Neoplasms"] OR [mh ^"Solitary Pulmonary Nodule"] 8755
- #16 ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) NEAR/3 (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*)):ti,ab,kw 28,597
- #17 ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") NEAR/4 (tumor* OR tumour* OR syndrome*)):ti,ab,kw 17
- #18 (sclc OR nsclc):ti,ab,kw 12,248
- #19 #15 OR #16 OR #17 OR #18 29,193
- #20 #10 and #14 and #19 47
- #21 ("AI-Rad Companion Chest" NEXT X-ray*) 0
- #22 ("Annalise" NEXT CXR*) 0
- #23 ("Auto Lung Nodule" NEXT Detection*) 0
- #24 ChestView* 0
- #25 (("Chest X-Ray" NEXT Classifier*) OR Quibim*) 0
- #26 CheXVision* 0
- #27 (("ClearRead" NEXT Xray*) NEAR/2 Detect) 0
- #28 ("InferRead DR" NEXT Chest*) 0
- #29 JLD-02K* 0
- #30 ("Lunit INSIGHT" NEXT CXR*) 2
- #31 ("Milvue" NEXT Suite*) 0
- #32 ("ChestEye" NEXT Quality*) 0
- #33 (qXR* OR Qure*) 921
- #34 (("red" NEXT dot*) OR behold*) 71
- #35 ("SenseCare-Chest DR" NEXT Pro*) 0
- #36 ("VUNO Med-Chest" NEXT X-Ray*) 1
- #37 (X1* AND "Visionairy Health") 0
- #38 #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 995
- #39 #14 and #38 4
- #40 #19 and #38 7
- #41 #39 or #40 8
- #42 #20 or #41 53

Trials: 52

Epistemonikos

Searched 1 December 2022

(title:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR superior sulcus OR pulmonary sulcus) AND (tumor* OR tumour* OR syndrome*)))) OR abstract:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR superior sulcus OR pulmonary sulcus) AND (tumor* OR tumour* OR syndrome*))))

Publication type:

Systematic review: 44

Broad synthesis: 1

ACM Digital Library

Searched 1 December 2022

Search for reviews

<https://dl.acm.org/search/advanced>

Selected ACM Guide to Computing Literature

Title:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour*

OR syndrome*)))) OR Abstract:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*))))

Filter by

Content type:

Review article: 12

Searches for primary studies

Searched 1 December 2022

<https://dl.acm.org/search/advanced>

Selected ACM Guide to Computing Literature

Title:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) AND ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*)))) OR Abstract:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) AND ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*))))

452

Searched 2 December 2022

<https://dl.acm.org/search/advanced>

Selected ACM Guide to Computing Literature

Title:(((ChestView* OR "Chest X-Ray Classifier" OR Quibim* OR CheXVision* OR ("ClearRead Xray" AND Detect) OR "InferRead DR Chest" OR JLD-02K* OR "Lunit INSIGHT CXR" OR "Milvue Suite" OR "ChestEye Quality" OR qXR* OR Qure* OR "red dot" or behold* OR "SenseCare-Chest DR Pro" OR "VUNO Med-Chest X-Ray" OR (X1* AND "Visionairy Health"))) AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*)))) OR Abstract:(((ChestView* OR "Chest X-Ray Classifier" OR Quibim* OR CheXVision* OR ("ClearRead Xray" AND Detect) OR "InferRead DR Chest" OR JLD-02K* OR "Lunit INSIGHT CXR" OR "Milvue Suite" OR "ChestEye Quality" OR qXR* OR Qure* OR "red dot" or behold* OR "SenseCare-Chest DR Pro" OR "VUNO Med-Chest X-Ray" OR (X1* AND "Visionairy Health"))) AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (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*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*))))

1

Systematic review register: search summary**PROSPERO**

Searched 15 December 2022

- #1 MeSH DESCRIPTOR Artificial Intelligence EXPLODE ALL TREES 477
- #2 MeSH DESCRIPTOR machine learning EXPLODE ALL TREES 154
- #3 MeSH DESCRIPTOR deep learning EXPLODE ALL TREES 23
- #4 MeSH DESCRIPTOR supervised machine learning EXPLODE ALL TREES 1
- #5 MeSH DESCRIPTOR support vector machine EXPLODE ALL TREES 0
- #6 MeSH DESCRIPTOR unsupervised machine learning EXPLODE ALL TREES 0
- #7 ai 1818
- #8 (artificial or machine or deep) AND (intelligence or learning or reasoning) 1830
- #9 MeSH DESCRIPTOR Neural Networks, Computer EXPLODE ALL TREES 28
- #10 "neural network" or "neural networks" or convolutional or CNN or CNNs 481
- #11 MeSH DESCRIPTOR Diagnosis, Computer-Assisted EXPLODE ALL TREES 15
- #12 MeSH DESCRIPTOR Pattern Recognition, Automated EXPLODE ALL TREES 1
- #13 ((automat* or autonomous or "computer aided" or "computer assisted") AND (detect* or identif* or diagnos*)) 3779
- #14 "support vector machine" or "support vector machines" or "random forest" or "black box learning" 156
- #15 #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 6790
- #16 MeSH DESCRIPTOR Radiography, Thoracic EXPLODE ALL TREES 10
- #17 MeSH DESCRIPTOR X-Rays 29

- #18 ((chest or lung* or thora*) and (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR* 1104
- #19 #18 OR #17 OR #16 1120
- #20 #15 AND #19 96
- #21 MeSH DESCRIPTOR Lung Neoplasms EXPLODE ALL TREES 572
- #22 MeSH DESCRIPTOR Solitary Pulmonary Nodule 6
- #23 (lung or lungs or pulmon* or intrapulmon* or bronch*) AND (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*) 6014
- #24 (pancoast* or "superior sulcus" or "pulmonary sulcus") and (tumor* or tumour* or syndrome*) 5
- #25 sclc or nsclc 896
- #26 #21 OR #22 OR #23 OR #24 OR #25 6062
- #27 #26 AND #15 256
- #28 #27 OR #20 312
- #29 #15 AND #19 AND #26 40

40 sifted online, 2 potentially relevant records sent to reviewers for checking

Trials registers: search summary

WHO ICTRP

Searched 18 January 2023 – targeted search #1

((lung* OR pulmonary OR intrapulmon* or bronch*) AND (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*)) in the Condition

AND

((artificial or machine or deep) AND (intelligence or learning or reasoning)) OR (AI OR "neural network*" or convolutional or CNN or CNNs OR "support vector machine*" or "random forest*" or "black box learning") OR ((automat* or autonomous or "computer aided" or "computer assisted") AND (detect* or identif* or diagnos*)) in the Intervention

AND

Recruitment status is All

32 records for 31 trials found

Searched 18 January 2023 – targeted search #2

((artificial or machine or deep) AND (intelligence or learning or reasoning)) OR (AI OR "neural network*" or convolutional or CNN or CNNs OR "support vector machine*" or "random forest*" or "black box learning") OR ((automat* or autonomous or "computer aided" or "computer assisted") AND (detect* or identif* or diagnos*)) AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*)) in the Intervention

13 records for 13 trials found

After deduplicating with above:

12 records remaining

Searched 18 January 2023 – targeted search #3

((((artificial or machine or deep) AND (intelligence or learning or reasoning)) OR (AI OR “neural network*” or convolutional or CNN or CNNs OR “support vector machine*” or “random forest*” or “black box learning”) OR ((automat* or autonomous or “computer aided” or “computer assisted”) AND (detect* or identif* or diagnos*))) AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*)) in the Title

29 records for 29 trials found

After deduplicating with above: 22

Total from the 3 searches: 65

Sixty-five filtered by the information specialist for basic eligibility (CXR and lung cancer/nodule/ abnormality, or unclear) or duplication with trial records found via other sources. Nine sent to clinical effectiveness reviewer for checking.

Cost-effectiveness searches

CEA Registry

Searched 30 November 2022

<https://cear.tuftsmedicalcenter.org/>

Basic search

Keyword is: lung cancer

Total: 285

Basic search

ICD-10: Malignant neoplasms of respiratory and intrathoracic organs (C30–C39)

Total: 264

Deduplicated in Microsoft Excel

Copied and pasted results from second search into same sheet as the first search then...

using Home > Conditional Formatting > Highlight Cells Rules > Duplicate Values

... and scanned by eye for any unique references in the second search. Kept these and deleted the duplicates.

Total after deduplication: 303

MEDLINE (via Ovid)

Searched 7 December 2022

Ovid MEDLINE® ALL 1946 to 6 December 2022

- 1 exp Radiography, Thoracic/ 40,535
- 2 X-Rays/ 31,182
- 3 ((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray*)).kf,tw. 64,896
- 4 1 or 2 or 3 120,457
- 5 exp Economics/ 653,642
- 6 exp "Costs and Cost Analysis"/ 261,580
- 7 Health Status/ 88,924
- 8 exp "Quality of Life"/ 255,297
- 9 exp Quality-Adjusted Life Years/ 15,263
- 10 (pharmacoeconomic* or pharmaco-economic* or economic* or cost* or price or prices or pricing).ti,ab,kf. 1,054,159
- 11 (expenditure\$ not energy).ti,ab,kf. 36,095
- 12 (value adj1 money).ti,ab,kf. 40
- 13 budget*.ti,ab,kf. 34,691
- 14 (health state* or health status).ti,ab,kf. 78,185
- 15 (qaly* or ICER or utilit* or EQ5D or EQ-5D or euroqol or euro-qol or short-form 36 or short-form 36 or SF-36 or SF36 or SF-6D or SF6D or SF-12 or SF12 or health utilities index or HUI).ti,ab,kf. 311,371
- 16 (markov or time trade off or TTO or standard gamble or SG or hrql or hrqol or disabilit* or disutilit* or net benefit or contingent valuation).ti,ab,kf. 302,967
- 17 (quality adj2 life).ti,ab,kf. 364,802
- 18 (decision adj2 model).ti,ab,kf. 8899
- 19 (visual analog* scale* or discrete choice experiment* or health* year* equivalen* or (willing* adj2 pay)).ti,ab,kf. 81,000
- 20 resource*.ti,ab,kf. 447,554
- 21 (well-being or wellbeing).ti,ab,kf. 130,164
- 22 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 or 18 or 19 or 20 or 21 2,829,367
- 23 exp Lung Neoplasms/ or Solitary Pulmonary Nodule/ 268,862
- 24 ((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. 327,230
- 25 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 946
- 26 (sclc or nsclc).kf,tw. 64,690
- 27 23 or 24 or 25 or 26 399,076
- 28 4 and 22 and 27 817

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health and Care Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library