Evidence overview: Early value assessment – Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer

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

Academic and commercial in confidence information:
Please note that throughout the report academic in confidence information is yellow and underlined and commercial in confidence information is marked blue and underlined.

1 Aims and scope

Software with artificial intelligence (AI)-derived algorithms that are designed to detect and analyse lung abnormalities on chest X-rays are available. These software tools could be used to assist a healthcare professional's review and interpretation of chest X-ray images by identifying images as normal or abnormal, highlighting suspected abnormalities and provide results as heat maps or probability scores

Use of the software may:

- identify abnormal lung features suggestive of lung cancer on a chest X-ray which helps to prioritise review of chest X-rays and speed up subsequent referral to CT scan,
- be used as a decision support tool to increase the accuracy of suspected lung cancer detection by consultant radiologists and reporting radiographers.
- reduce the time to review and report chest X-rays

- help to find and treat lung cancer early.

This topic is presented as an early value assessment. The decision questions that would need to be answered in guidance are presented below.

Decision questions

- Does the use of software with artificial intelligence (AI) derived algorithms for analysing chest X-ray images for suspected lung cancer have the potential to be clinically and cost-effective to the NHS?
- What evidence is available to support the value proposition outlined in the scope (1. identification of lung cancer, 2. triage and prioritisation to improve workflow) and where are the evidence gaps?

Populations

Adults who have a chest X-ray request from primary care because of:

- Symptoms suggestive of lung cancer (symptomatic population)
- Reasons unrelated to suspicion of lung cancer (incidental population)

Depending on the available evidence, the following subgroups will be considered based on:

- Ethnicity
- Age
- Sex
- Socio-economic status

Interventions

AI-derived software-assisted chest X-ray review by a radiologist or reporting radiographer using any of the following software:

- AI-Rad Companion Chest X-ray (Siemens Healthineers)
- Annalise CXR (annalise.ai)
- Auto Lung Nodule Detection (Samsung)

- ChestLink Radiology Automation (Oxipit)
- ChestView (GLEAMER)
- Chest X-ray (Rayscape)
- ClearRead Xray – Detect (Riverain Technologies)
- InferRead DR Chest (Infervision)
- Lunit INSIGHT CXR (Lunit)
- Milvue Suite (Milvue)
- qXR (Qure.ai)
- red dot (behold.ai)
- SenseCare-Chest DR Pro (SenseTime)
- VUNO Med-Chest X-Ray (VUNO)

Comparator

The comparator is chest X-ray image review by an appropriate radiology specialist (radiologist or reporting radiographer) without the assistance from AI-derived software.

Healthcare setting

Primary care

Further details, including descriptions of the interventions, comparator, care pathway and outcomes, are in the final scope for Artificial Intelligence software for analysing chest X-ray images to identify suspected lung cancer.

2 Summary

Clinical effectiveness

The EAG conducted a pragmatic review to identify evidence on the clinical effectiveness and diagnostic accuracy of artificial intelligence derived software for analysing chest X-ray images to identify suspected lung cancer. No studies meeting the predefined inclusion criteria were identified. Post hoc inclusion

criteria were used to identify studies that were closest to meeting the inclusion criteria.

The post hoc inclusion criteria were based on studies that had (1) eligible AI-derived software, and (2) compared radiology specialist in conjunction with AI-derived software to radiology specialist alone, but where the referral status of the population was unclear. Studies that had an explicitly excluded population (for example, a health screening population, pre-operative chest X-ray, inpatients, A&E) remained excluded.

The evidence comprised 6 retrospective studies of which 2 were provided by companies and not peer reviewed, one pre-print and another still ongoing. The total population in all the included studies was 1,597. Only 1 study was conducted in the UK and the rest were from Germany and US (n=1), Korea (n=3), and US (n=1). Three studies assessed Lunit INSIGHT CXR (Lunit), 1 assessed red dot (Behold.ai), and 2 assessed AI-Rad Companion (Siemens Healthineers).

Key heterogeneities and risks of bias were identified in the summarised studies. For example:

<u>Sources of heterogeneity between studies</u>

- chest X-rays were assessed by radiologists or reporting radiographers with various levels of expertise, and the number of clinicians included in the studies ranged from 4 to 11,
- the accuracy of readers in detecting nodules or lung cancer with and without AI software was compared with a ground-truth or reference standard, and these varied between the studies.

<u>Risks of bias</u>

- assessments were conducted on test-sets of data interpreted outside clinical practice,

- software manufacturers were involved in 3 of the 6 studies,
- in 3 studies there was a washout period between readings, whereas in others the radiologist was aware of their initial decision at the second reading,
- 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,
- Where CT referrals were reported, these were hypothetical referrals rather than actual referrals and may not reflect real-world practice.

Diagnostic accuracy

One study examined the test accuracy of AI-derived software to detect lung cancer on chest X-ray (table 1), whereas 5 studies examined the test accuracy of AI-derived software to detect lung nodules on chest X-ray (table 2).

Table 1: Test accuracy for detecting lung cancer

Study	AI-derived software	Number of patients	Sensitivity (95% CI) with software	Sensitivity (95% CI) without software	Specificity (95% CI) with software	Specificity (95% CI) without software
Dissez 2022	Red dot (behold.ai)	400	77% (75% to 80%)	66% (59% to 71%)	75% (71% to 77%)	81% (77% to 85%)

Table 2: Test accuracy for detecting lung nodules

Study	AI-derived software	Number of patients	Sensitivity (95% CI) with software	Sensitivity (95% CI) without software	Specificity (95% CI) with software	Specificity (95% CI) without software
Nam 2020[a]	Lunit INSIGHT version 1.0.1.1 (Lunit)	NR	53% (49% to 57%)	47% (43% to 51%)	82% (77% to 87%)	78% (72% to 84%)
Jang 2020[a]	Lunit INSIGHT version 1.2.0.0 (Lunit)	351	56% (47% to 65%)	43% (34% to 52%)	92% (88% to 95%)	90% (86% to 94%)
Koo 2021 (per patient)	Lunit INSIGHT version 1.0.0.0 (Lunit)	378	95% (NR)	92% (NR)	97% (NR)	93% (NR)

Koo 2021 (per nodule)	Lunit INSIGHT version 1.0.0.0 (Lunit)	NR	94% (NR)	89% (NR)	NR (NR)	NR (NR)
Homayounieh 2021	AI-Rad Companion Chest X-ray (Siemens Healthineers)	100	55% (48% to 63%)	45% (38% to 53%)	95% (91% to 99%)	93% (89% to 96%)
Siemens 2022	Prototype AI-Rad Companion Chest X-ray (Siemens Healthineers)	1018	91% (NR)	76% (NR)	78% (NR)	81% (NR)

[a] 95% confidence intervals (95% CI) calculated by the external assessment group using study data. Data are mean values for all readers.

Impact on clinical decision making

Two of the summarised studies provided information on the hypothetical referrals for CT scans. There were no statistically significant differences in the number of people who might be recommended for CT follow-up between readers with and without use of AI-derived software. Dissez 2022 reported 144 of 400 (36%) potential referrals with use of AI-derived software and 117 of 400 (29%) without. Jang 2020 reported 96 of 351 (27%) potential referrals with use of AI-derived software and 80 of 351 (23%) without.

Reading times and acceptability

Two studies reported information on reading times. No statistically significant differences were observed in average image reading times between readers with and without use of AI-derived software. Jang 20210 reported 22.5 (SD 40.3) seconds per image with use of AI-derived software and 24.3 (SD 27.4) seconds without. Koo 2021 reported 171 (SD 33.8) minutes with use of AI-derived software and 211.25 (SD 38.4) minutes without, to read 434 images.

Dissez 2022 reported on the acceptability of red dot (behold.ai) amongst 10 clinicians. Eight clinicians indicated that reporting was not slowed down by use of AI-derived software, and 9 stated that the heatmaps (visual display of findings) were helpful.

Other outcomes

None of the studies included in the review reported on AI-derived software technical failure or clinical outcomes.

Find the full review results on page 35 of the diagnostics assessment report.

Ongoing studies

The EAG did not identify any ongoing studies that met the inclusion criteria aimed at estimating the clinical effectiveness of adjunct AI-derived software applied to chest X-ray. Two studies were identified based on the post hoc inclusion criteria. One of the ongoing trials (KCT0005466) identified by the EAG compared Lunit INSIGHT in conjunction with radiologist to radiologist alone. The study had an estimated end date of 31/05/2021. It was unknown if patients were referred from primary care and if they had symptoms due to cancer.

The eligibility of one additional ongoing study (NCT05489471) identified from the Lunit company submission is unclear. The population in terms of the proportion of GP referrals, accident and emergency attendances and in-patients is not known, and the comparison (whether it includes AI-derived software plus radiologist versus radiologist alone) and the intervention (the AI-derived software was not named but is funded by Lunit) are not stated. This UK based study is currently not yet recruiting and has an estimated primary end date of July 2023.

Another ongoing trial assessing qXR (Qure.ai) was not described by the EAG but was identified by clinical experts. This study is being conducted in 7 hospitals in England and aims to assess if time taken to diagnosis and reporting of abnormalities on a chest X-ray can be reduced with the assistance of AI-derived software. The population in this study is unclear. The trial will run until June 2023. The study is funded by NHS England (NHSE) and small business research initiative (SBRI) in partnership with the Accelerated Access Collaborative (AAC).

Cost effectiveness

The EAG aimed to develop a conceptual decision analytic model to inform potential future full cost-effectiveness evaluation of AI-derived software for analysing chest X-ray images to identify suspected lung cancer. The EAG conducted a pragmatic review to identify any relevant economic evaluations, clinical guidelines and company submissions, and discussed with specialist clinical experts to inform the conceptual model development. Furthermore, the EAG considered the costs of introducing AI-derived software as an adjunct to radiology specialist review of chest X-ray by developing a simple budget impact analysis.

Evidence to inform conceptual modelling

No cost-effectiveness studies that directly addressed the topic of this review were found. However, 2 economic evaluations and an updated analysis of one of these were identified to inform modelling techniques and parameter inputs. In addition, 4 studies which provided detailed information on radiological or clinical pathways for lung cancer diagnosis in the UK were identified. A systematic review and meta-analysis on the diagnostic performance of chest X-rays in symptomatic primary-care populations was also retrieved from the search. A narrative summary of the identified studies was provided and no formal data extraction or quality appraisal was conducted. The summary prioritised information for the diagnostic component of the conceptual model rather than the longer-term treatment costs and utilities.

Bajre et al. (2017) assessed the cost-effective of trained radiographers compared with radiologists for the reporting of chest X-ray in people suspected of having lung cancer. This study provided information for a decision analytic model structure deemed relevant to this assessment.

Foley et al. (2021) conducted a retrospective review of audit data to analyse the use of chest X-ray as the first-line investigation in primary care patients with suspected lung cancer. The study reported the total number of chest X-

rays, number referred for CT scan, number of lung cancers, number diagnosed at advanced stage, number of days from chest X-ray to CT scan, number of days from chest X-ray to diagnosis, number receiving treatment with curative intent, and all-cause mortality.

Bradley et al. (2021) undertook a retrospective observational study using routinely collected healthcare data. This study provides some information on the sensitivity of chest X-rays in cancer diagnosis. An analysis was performed on time to diagnosis, stage at diagnosis and survival outcomes.

Woznitza et al. (2018) conducted a four-month feasibility study at a single radiology department at an acute general hospital. This study reported on the time to CT scan and time to discussion at the multidisciplinary team. The study also gave detailed description of the radiology department demographics and processes for reporting and referral.

Woznitza et al (2022) conducted a prospective, block-randomised controlled trial at a single acute district general. People referred for chest X-ray from primary care attended sessions that were pre-randomised to either immediate radiographer reporting, or standard radiographer reporting within 24-hours. The outcomes from the study were previous chest X-ray status, previous CT scan status, suspected lung cancer, total cancers diagnosed, 2-week wait referral, time from chest X-ray to diagnosis, and time from chest X-ray to discharge.

Conceptual model development

The EAG outlined a chest X-ray clinical pathway, supported by existing guidelines on the diagnostic and care pathway and collaboration with clinical experts. The report noted that this was an aspirational pathway, with many alternative routes both in and out through to diagnosis. For more detail on the clinical pathway see section 5.3 on page 52 of the assessment report. A model structure based on the final scope and expert consultation was conceptualised by the EAG and is presented in Figure 1.

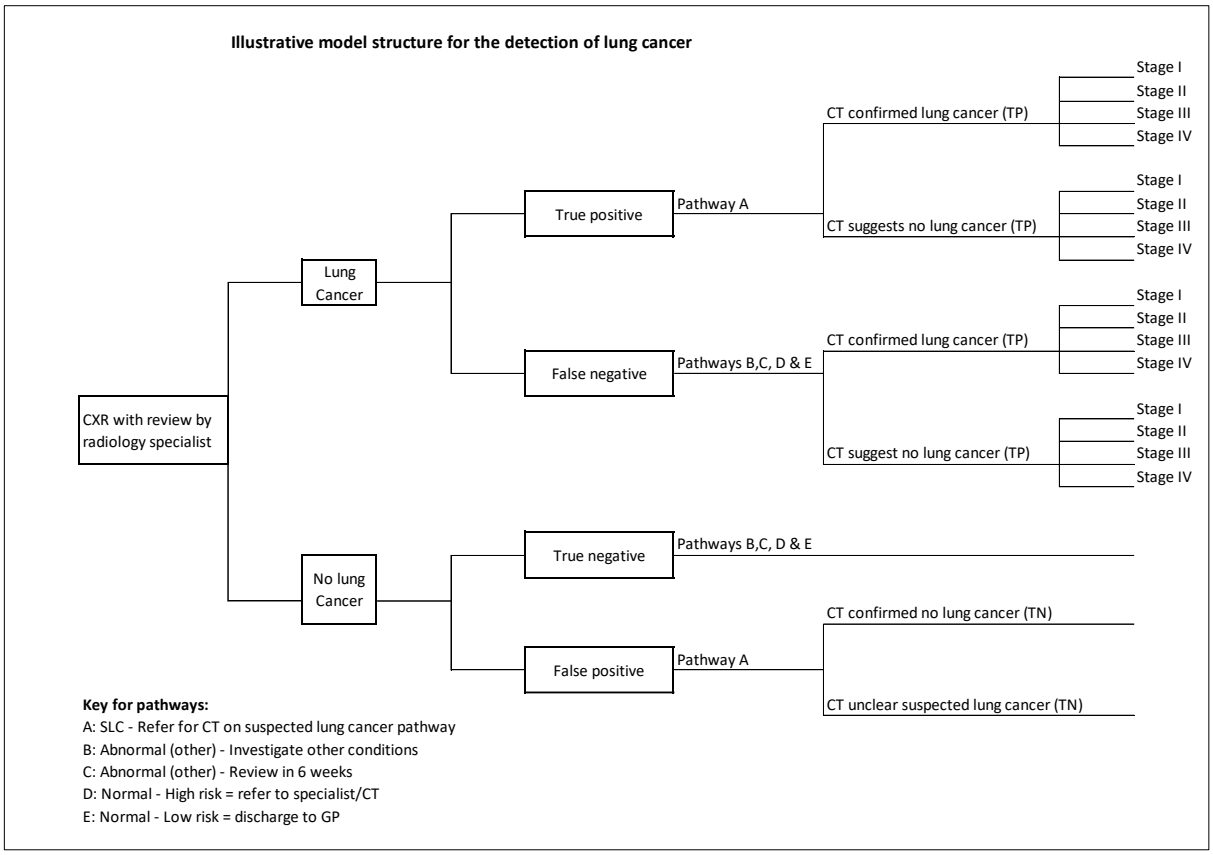

Figure 1: Illustrative model structure for the detection of lung cancer

The illustrative pathway for chest X-ray review by radiology specialist with adjunct AI software is identical to the structure presented here for chest X-ray review by radiology specialist alone. If software were used for triage, an additional step prior to the chest X-ray could be included.

The input parameters identified to have potential to influence health care costs in the model were the intermediate outcomes. These include:

- diagnostic accuracy,
- turnaround time (time from start of image review to radiology report), technical failure rate,
- impact of software output on clinical decision-making,
- number of people referred for CT scan,
- number of people referred for follow-up chest X-ray,
- number of cancers missed/detected,
- stage of cancer detection,
- time to chest X-ray reporting,

- time to CT scan, and
- time to diagnosis.

For more discussion on the inputs to inform the model structure please see section 5.4 on page 57 of the assessment report. The EAG also noted that cost consideration would include:

- cost of each AI-derived software available for this indication (from the company)
- costs of training staff to use software (cost included in the one-off implementation fee)
- costs associated with healthcare professional time to read and report chest X-ray (Literature and Personal Social Services Research Unit (PSSRU))
- costs of diagnostic testing and treatment (National schedule of NHS costs 2020/21 and the PSSRU Unit costs of health and social care 2021)

Cost and resource use considerations

The budget impact analysis considered one-off set up costs, annual subscription fee based on a volume of 16,945 images, total cost per year and the cost over the first 5 years. Because the literature reviews did not provide any evidence to show changes in resource use due to AI-derived software, only the additional costs of AI-derived software were considered. In addition, no evidence found to inform any changes to progression through the clinical pathway due to the use of AI-derived software. Therefore, onward health-related service use, diagnostic and treatment costs were assumed to stay the same.

In any future modelling, costs of CT scans, CT surveillance, further invasive tests, and treatment for different stages of lung cancer at diagnosis would need to be considered. For example, a change in test accuracy may result in:

- increased sensitivity with use of AI-derived software and potentially resulting in more cancers or nodules being identified,

- decreased specificity with use of AI-derived software wherein more people could be referred for a CT scan with an associated cost implication (because of an increase in false positives).

The full details on the budget impact analysis are presented in section 5.5 of the diagnostic assessment report.

Test costs varied between companies (see table 3 and 4), but the EAG cautions against direct comparison, as the AI-derived software presented have varying capabilities and some may be used in different positions early in the diagnostic pathway. For example, some software point to a region of interest on the chest X-ray, whereas others identify a specific location, give characteristics of the anomaly and provide a preliminary diagnosis and rating of confidence. Further, some of the software can provide triage of chest X-ray images prior to radiology specialist review in order to prioritise reporting. However, there is no evidence to indicate that any of these capabilities add value at this point.

Table 3: Anticipated budget impact of AI software at NHS Trust level for all GP-referred chest X-ray

Company **Technology name** **(Tech use)**	**One-off set up cost/ implementation fee**	**Annual subscription (based on volume 16,945 images)**	**Cost per exam**	**Total first year cost** **[VAT applied at 20%]**	**Cost over first 5 years (non-discounted, based on volume 16,945 images per yr)** **[VAT applied at 20%]**
Annalise.ai Annalise CXR (CADe & CAST)	£5,000 - £25,000	£51,250*	N/A	£66,250 (assuming mean implementation fee) [£79,500]	£271,250 [£325,500]
Behold.ai Red dot (CADe & CAST)	[redacted]	[redacted]	N/A	[redacted]	[redacted]
Infervision InferRead DR Chest (CADe)	£3,000	£16,000 (license fee) £6,000 (maintenance fee)	N/A	£25,000 [£30,000]	£113,000 [£135,600]

Annual licence pricing					
Infervision InferRead DR Chest (CADe) Pay-per-scan pricing	£3,000	£6,000 (maintenance fee)	£1.50	£34,418 [£41,302]	£160,088 [£192,106]
Lunit Lunit INSIGHT CXR (CADe)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Siemens Healthineers AI-Rad Companion Chest X-ray (CADx)	£2,400	£12,000*	N/A	£14,400 [£17,280]	£62,400 [£74,880]

*Based on tier pricing of 'up to' 25,000 images per year

Table 4: Anticipated budget impact of AI software at NHS Trust level for symptomatic, incidental and whole population GP-referred chest X-ray

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 CXR	NDA	£325,500	£325,500
Behold.ai Red dot	[redacted]	[redacted]	[redacted]
Infervision InferRead DR Chest Annual licence pricing	£135,600	£135,600	£135,600
Infervision InferRead DR Chest Pay-per-scan pricing	£49,396	£132,342	£188,057
Lunit Lunit INSIGHT CXR	[redacted]	[redacted]	[redacted]
Siemens Healthineers AI-Rad Companion Chest X-ray	£26,880	£74,880	£74,880

NDA = No data available. Non-discounted costs, VAT included at rate of 20%. Total population n = 16,945, symptomatic population n = 1,488, incidental population n = 15,457.

3 Issues for consideration

The following 3 key issues have been highlighted for the committee to discuss.

Key issue 1: Lack of evidence about diagnostic accuracy, technical failure rates, clinical decision making and clinical outcomes

Description of issue

No studies were identified that met the predefined inclusion criteria. This results in uncertainties about the impact of AI-derived software on diagnostic accuracy, technical failure, impact on clinical decision making and clinical outcomes in a primary care population referred for chest X-ray.

Background

The EAG aimed to assess clinical effectiveness in terms of intermediate outcomes and clinical outcomes of AI-derived software to detect suspected lung cancer in chest X-ray images of people referred from primary care.

No studies meeting the inclusion criteria were identified. The post hoc inclusion criteria resulted in the inclusion of 6 studies that had unclear populations. Only 1 study was carried out in the UK.

Results from the UK study assessing diagnostic accuracy of red dot (Behold.ai) to detect lung cancer showed that sensitivity was significantly higher for interpretation with AI-derived software (77%, 95% CI 75% to 80%) than without AI-derived software (66%, 95% CI 59% to 71%). Specificity was slightly lower for interpretation with AI-derived software (75%, 71% to 77%) than without AI-derived software (81%, 77% to 85%) but the difference was not statistically significant. No statistically significant differences in diagnostic accuracy to detect lung nodules between readers with or without AI-derived

software were reported in the other 5 studies (table 2). No studies reported on AI-derived software technical failure or clinical outcomes.

Two studies reported on impact on clinical decision making, but both were conducted retrospectively rather than in a clinical setting. No statistically significant differences were observed between readers with and without use of AI-derived software in terms of the number of people who might be referred for follow-up CT scan. See section 4 on page 28 of the diagnostic assessment report for details.

Questions for committee

- Is the summarised diagnostic accuracy data that has been generated from unclear populations generalisable to the population of interest (people referred from primary care)?
- Does AI-derived software have the potential to be clinically effective if used in the NHS in a population referred for chest X-ray from primary care?
- Are data on diagnostic accuracy and technical failure from a population of people referred from primary care needed? If so, is there a preference for accuracy to detect lung cancer or accuracy to detect lung nodules? Per person or per nodule accuracy? And what should the reference standard be?
- Are there any other key research needs to understand the clinical effectiveness of AI-derived software for analysing chest X-ray images to identify suspected lung cancer in people referred from primary care?

Key issue 2: Lack of evidence about time to read and report and acceptability to clinicians

Description of issue

No studies were identified that met the predefined inclusion criteria. This results in an absence of evidence about the impact of AI-derived on the time to read and report chest X-rays in people referred from primary care.

Background

The EAG aimed to assess practical implications for using the technologies such as time to read and report images and acceptability to clinicians. The EAG did not identify any studies that met the predefined inclusion criteria. The post hoc inclusion criteria resulted in the inclusion of 6 studies that had unclear populations. Only 1 study was carried out in the UK.

Three studies reported on time to read and report and clinician acceptability outcomes. Two reported that no statistically significant differences were observed in average image reading times between readers with and without use of AI-derived software. One study found that clinicians indicated reporting of chest X-rays was not slowed down by the use of AI-derived software and that visual display of abnormal findings was helpful to understand the algorithm's attention points. Additional evidence and feedback provided to the EAG showed significant variability in this measure even without the use of AI software

Questions for committee

- Is the summarised evidence on time to read and report chest X-rays that has been generated from unclear populations generalisable to the population of interest?
- Is there a preference for pursuing symptomatic versus incidental population subgroup?
- Does the summarised evidence indicate that AI-derived software has the potential for equivalent or better chest X-ray reading and reporting times in the population of interest?
- Is more data on time to read and report chest X-rays with and without the use of AI-derived software needed?

Key issue 3: Lack of evidence around the cost impact if AI-derived software were used in the NHS.

Description of issue

The EAG produced the conceptual model agreed in the scope as lack of evidence was anticipated to prevent full exploration of the potential cost and resource use impact of using adjunct AI-derived software in the NHS to analyse chest X-ray images. For this reason, cost and resource use impacts are highly uncertain.

Background

The literature reviews did not identify any evidence to show changes in resource use due to AI-derived software. However, evidence to suggest there are many factors which influence onward resource use at this stage was retrieved, introducing further uncertainty into any further modelling and calculations. The budget impact only used software costs as submitted by companies, with independent calculations undertaken to confirm company estimates, see section on cost and resource use consideration above. Test costs varied between companies, but as the AI-derived software have varying capabilities and some may be used in different positions early in the diagnostic pathway. For example, some software point to a region of interest on the chest X-ray, whereas others identify a specific location, give characteristics of the anomaly and provide a preliminary diagnosis and rating of confidence. Further, some of the software can provide triage of chest X-ray images prior to radiology specialist review in order to prioritise reporting. In any future modelling, costs of CT scans, CT surveillance, further invasive tests, and treatment for different stages of lung cancer at diagnosis will need to be considered.

According to the EAG, coordinated research efforts are required to generate research on all outcome measures identified for inclusion in the conceptual model. Evidence needs to demonstrate impact on intermediate outcomes over

a sustained period 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 of evidence linkage which remains particularly weak with regards to impact on stage at diagnosis.

Questions for committee

- Does AI-derived software have the potential to be cost effective if used in the NHS in a population referred for chest X-ray from primary care?
- What are the key data gaps that need further evidence generation before a full assessment of cost-effectiveness can be made?
- Would a linked-evidence modelling approach that includes diagnostic accuracy, impact on decision making, and lung cancer pathway outcomes from different sources be acceptable for future economic modelling?

4 Equality considerations

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

Potential equality issues relating to the use of AI-derived software include:

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

Question for committee

- Are there any other potential equality concerns relevant to AI-derived software that should be taken into consideration?

5 Implementation

The implementation team gathered some insights from the NHS to establish potential levers and barriers to adopting the technologies. This information was gathered as expert opinion from clinicians within the NHS.

Potential levers to implementation

The insights from the NHS highlighted that the potential levers may include:-

- Minimal training required to use the technologies.
- Expect the technologies can be used on nearly all x-ray images.
- Reports seem to be generally easy to read.
- The technologies may provide additional information to aid decision making regarding treatment / intervention.
- The AI report is produced almost instantly and available on the PACS system.
- Most technologies allow the option to toggle on or off the AI report.
- Several of the technologies can also identify conditions other than lung cancer also.

Potential barriers to implementation

Several potential barriers to implementation were established. According to some experts, these potential barriers may include:-

- Potential clinical governance issues if rolled out widely such as security of sending images and personal data via a cloud based server.
- Capacity of existing IT systems.
- Possible requirement of a dedicated person to read the report.
- Skill / experience / opinion of the person reading the report may affect how the report is ‘interpreted’ and how risk is defined.
- Some clinicians may feel that they are experienced enough to identify lung cancer and that technologies are not needed.

- Funding for NHS Trust to procure the technologies.
- Hospital IT departments may not have the time and ability to install technology.

Integration into radiologists' workflow

If the software does not fully integrate into the radiologists' workflow within the Picture Archiving and Communication System (PACS) where chest X-ray images are reviewed and reported, adds steps to the image review, or does not include rules for reporting lung nodules in the NHS, using the software may increase review time.

IT capacity and compatibility

There are some concerns about the level of IT support and capacity needed to install and use the software.

Governance issues

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

6 Overarching issues for committee consideration

Taking into consideration the issues raised in the previous sections, there are some overarching issues for consideration by the committee. These issues are motivated by the value proposition and early value assessment objectives in general. The table below highlights some potential benefits and risks of conditionally recommending AI-derived software for use while further evidence is generated.

Table 5: Potential benefits and risks of the interventions

Potential benefits	Potential risks
Could be more sensitive to detect cancerous nodules and other abnormalities that suggest cancer • More cancers could be identified at an earlier stage • Could result in improved patient outcomes and quality of life	Specificity to detect cancerous nodules and other abnormalities that suggest cancer could be lower with use of adjunct software than without • Could result in more people being referred for chest CT • This would have cost/resource and disutility implications
Could help workflow triage and reduce the time from initial detection to diagnosis and treatment • Could result in improved patient outcomes and quality of life	One off set-up cost/implementation fee • Cost to get the technology installed at the hospital
Could reduce the time radiologist or diagnostic radiographers spend reviewing and reporting chest X-rays • Potential resource saving	May not reduce the time radiologists or diagnostic radiographers spend reviewing and reporting chest X-rays • Could be an added cost without resource savings in reading/reporting chest X-rays
Anything else?	Anything else?

Questions for the committee

- Are the potential risks and benefits listed in table 5 reasonable?
- Are there any other potential risks and benefits linked to use of AI-derived software?
- Is it acceptable for AI-derived software for analysing chest X-rays to be used in practice in the NHS while further evidence is generated? Or should these software be used only in a research context?
- If conditionally recommended, are there any measures that need to be put in place to mitigate the potential risks?

7 Authors

Peslie Ngambi

Topic lead

Alice Bell

Topic support

Frances Nixon

Technical adviser