



OPEN ACCESS

Original research

Software using artificial intelligence for nodule and cancer detection in CT lung cancer screening: systematic review of test accuracy studies

Julia Geppert,¹ Asra Asgharzadeh,^{2,3} Anna Brown,³ Chris Stinton,¹ Emma J Helm,⁴ Surangi Jayakody,³ Daniel Todkill,³ Daniel Gallacher,³ Hesam Ghiasvand,^{3,5} Mubarak Patel,³ Peter Auguste,³ Alexander Tsertsvadze,³ Yen-Fu Chen ,³ Amy Grove,³ Bethany Shinkins,¹ Aileen Clarke,³ Sian Taylor-Phillips⁶

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/thorax-2024-221662>).

¹Warwick Screening & Warwick Evidence, Warwick Medical School, University of Warwick, Coventry, UK

²Population Health Science, University of Bristol, Bristol, UK

³Warwick Evidence, Warwick Medical School, University of Warwick, Coventry, UK

⁴Department of Radiology, University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK

⁵Research Centre for Healthcare and Communities, Coventry University, Coventry, UK

⁶Warwick Screening, Warwick Medical School, University of Warwick, Coventry, UK

Correspondence to

Dr Yen-Fu Chen;
Y.F.Chen@warwick.ac.uk

JG and AA are joint first authors.

Received 8 March 2024
Accepted 4 September 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY. Published by BMJ.

To cite: Geppert J, Asgharzadeh A, Brown A, et al. *Thorax* Epub ahead of print: [please include Day Month Year]. doi:10.1136/thorax-2024-221662

ABSTRACT

Objectives To examine the accuracy and impact of artificial intelligence (AI) software assistance in lung cancer screening using CT.

Methods A systematic review of CE-marked, AI-based software for automated detection and analysis of nodules in CT lung cancer screening was conducted. Multiple databases including Medline, Embase and Cochrane CENTRAL were searched from 2012 to March 2023. Primary research reporting test accuracy or impact on reading time or clinical management was included. QUADAS-2 and QUADAS-C were used to assess risk of bias. We undertook narrative synthesis.

Results Eleven studies evaluating six different AI-based software and reporting on 19 770 patients were eligible. All were at high risk of bias with multiple applicability concerns. Compared with unaided reading, AI-assisted reading was faster and generally improved sensitivity (+5% to +20% for detecting/categorising actionable nodules; +3% to +15% for detecting/categorising malignant nodules), with lower specificity (−7% to −3% for correctly detecting/categorising people without actionable nodules; −8% to −6% for correctly detecting/categorising people without malignant nodules). AI assistance tended to increase the proportion of nodules allocated to higher risk categories. Assuming 0.5% cancer prevalence, these results would translate into additional 150–750 cancers detected per million people attending screening but lead to an additional 59 700 to 79 600 people attending screening without cancer receiving unnecessary CT surveillance.

Conclusions AI assistance in lung cancer screening may improve sensitivity but increases the number of false-positive results and unnecessary surveillance. Future research needs to increase the specificity of AI-assisted reading and minimise risk of bias and applicability concerns through improved study design.

PROSPERO registration number CRD42021298449.

INTRODUCTION

Early detection, assessment, monitoring and timely intervention of pulmonary nodules are the key approach to reducing lung cancer morbidity and mortality. Lung cancer screening programmes have been established in several countries including the USA, Croatia, Czech Republic and Taiwan

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Artificial intelligence (AI)-based software is increasingly used to assist the detection and measurement of pulmonary nodules as part of lung cancer screening, but its impact on test accuracy and clinical management has not been comprehensively critiqued and summarised.

WHAT THIS STUDY ADDS

⇒ AI assistance in lung cancer screening tends to increase sensitivity (detecting more cancers) but at the cost of reduced specificity (resulting in significant additional surveillance of nodules, which would never develop into cancer).
⇒ Evidence was mostly from retrospective studies conducted in research settings with high risk of bias and applicability concerns.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Adoption of AI software and further research should focus on improving the specificity of AI assistance and prospective collection of evidence from in-practice settings using robust study design.

following growing evidence demonstrating survival benefits.^{1 2} In September 2022, the UK National Screening Committee recommended targeted lung cancer screening using low-dose CT for people aged 55–74 identified as being at high risk of lung cancer.³

Recommendations for nodule management differ across guidelines internationally,⁴ but most rely on measuring the diameter or the volume of the nodule to help determine next steps. Many individuals with nodules are placed under regular CT surveillance to assess whether the nodule is growing. Obtaining an accurate manual measurement of nodules can be challenging; nodules present in a wide range of different shapes and sizes. There is evidence of substantial inter-reader and intra-reader variability, and that variability increases the more complex the nodule morphology is.⁵ In the recently published Dutch–Belgian lung cancer screening trial (NELSON), 9.2% of the CT scans were

indeterminate (ie, showed either a solid nodule with a volume of 50–500 mm³, pleural-based solid nodules with a minimal diameter of 5–10 mm or a solid nodule with a non-solid component with a mean diameter of ≥ 8 mm).⁶ All these individuals required a repeat CT scan in 3 months to calculate volume-doubling time. As the proportion of people with nodules detected on CT scans is high, the accurate measurement and appropriate management of nodules have significant implications for radiologist time and potential patient anxiety.

Computer-aided detection (CAD) systems for assisting radiologists in reading CT scans, which rely on predefined rules, thresholds and patterns, have been available for many years. They were used in the NELSON trial,⁶ the UKLS trial,⁷ the Multicentric Italian Lung Detection trial⁸ and the ongoing Yorkshire Lung Screening Trial.⁹ Different types of software using modern forms of artificial intelligence (AI) capable of automatically detecting and measuring pulmonary nodules have become available and could potentially reduce the screening workload and reading time for radiologists. These operate differently to traditional CAD systems; they do not rely on predefined rules and instead learn task-relevant features and generate algorithms from raw input data.

We aimed to examine the accuracy of CE-marked (compliant with relevant European Union regulations), AI-based software use for automated detection and analysis of pulmonary nodules in chest CT scans as part of lung cancer screening. As secondary outcomes, we analysed the reading time and the provided information on the impact of AI assistance on Lung CT Screening Reporting & Data System (Lung-RADS) categorisation.

METHODS

Protocol and registration

This systematic review is an update of part of a diagnostic technology assessment for the National Institute for Health and Care Excellence.¹⁰ The protocol for the original systematic review was registered with PROSPERO. This paper is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for diagnostic test accuracy studies.¹¹

Data sources

We conducted literature searches on 17–19 January 2022 and updated these on 6 March 2023. The search strategy was based on three themes: lung cancer/nodules, AI and computer tomography/mass screening/early detection of cancer. Databases searched were MEDLINE, Embase, Cochrane Database of Systematic Reviews, Cochrane CENTRAL, Health Technology Assessment (HTA) database (CRD), International HTA database (INAHTA), Science Citation Index Expanded (Web of Science), Conference Proceedings—Science (Web of Science). Endnote V.20 was used to identify and remove duplicate results.

We searched or reviewed websites of selected conference proceedings, health technology assessment organisations, device manufacturers and devices@FDA between 24 January and 16 February 2022. Forward citation tracking from key publications of included studies was also undertaken in May 2022, using Science Citation Index (Web of Science) and Google Scholar. Details of the search strategies are provided in online supplemental material 1. Reference lists of included studies and recent, relevant systematic reviews identified via the database searches were checked.

Study selection

Two reviewers independently reviewed titles and abstracts of all retrieved records and all potentially eligible full-text publications against inclusion criteria. Disagreements were resolved by consensus or discussion with a third reviewer. Studies were eligible for inclusion if they reported test accuracy of AI-based software for automated detection and analysis of lung nodules from CT images performed for lung cancer screening or secondary outcomes relating to the impact on clinical management and practical implications. We included all AI-based software which had (or was anticipated to have) an appropriate regulatory approval (CE mark) across the UK and the EU by December 2021 and was near-market—that is, with anticipated availability for commercial use by 2023. The reference standard for lung nodule presence/absence was experienced radiologist reading. Lung cancer presence was confirmed by histological analysis of lung biopsy or health record review; lung cancer absence was confirmed by CT surveillance (imaging follow-up) without significant nodule growth or follow-up without lung cancer diagnosis. Eligible outcomes included test accuracy for nodule detection and/or risk categorisation based on size (any nodules, actionable nodules and malignant nodules, respectively), impact on clinical management and practical implications. Eligible study designs were test accuracy studies, randomised controlled trials, cohort studies, historically controlled trials, before–after studies and retrospective multireader multicase (MRMC) studies. We included peer-reviewed papers; conference abstracts and manufacturer data were only included if they were related to an eligible peer-reviewed full-text paper and reported additional outcome data.

We excluded studies using PET-CT scan images, lung phantom images or where less than 90% were CT images taken for lung cancer screening. We excluded studies if traditional CAD systems without deep learning were used, or they had no relevant test accuracy or clinical management outcomes, and non-human studies along with letters, editorials and communications unless they reported outcome data not reported elsewhere, in which case they were handled in the same way as conference abstracts. We excluded articles not available in English or published before 2012.

Data extraction and quality assessment

Detailed information related to study design, sampling of patients or CT scan images, AI-based software, reference standard and test accuracy outcomes was collected from each included study. Data allowing construction of 2×2 tables were extracted where possible, to calculate sensitivity and specificity. The unit of analyses (per person or per nodule) and features of detected/missed nodules were noted. Comparative data on the potential or actual impact of AI assistance on clinical management (eg, risk categorisation of lung nodules according to clinical guidelines based on measured nodule sizes) and time required by readers to interpret and report findings of the CT scans were also collected.

One reviewer extracted data into a predesigned electronic data collection form (online supplemental material 2). Data extraction sheets were checked by a second reviewer. Any disagreements were resolved through discussion, with the inclusion of a third reviewer when required. Study quality was assessed independently by two reviewers using QUADAS-2¹² combined with the QUADAS-C tool for comparative studies,¹³ tailored to the review question (online supplemental material 3). Assessment of applicability was based on a UK/EU frame of

reference. Disagreements were resolved through consensus, with the inclusion of a third reviewer if required.

Data analysis

We focused on comparisons between trained human readers (radiologists or other trained healthcare professionals) assisted by AI-based software and those undertaking unassisted reading of CT scan images as this reflects current use of the technology in clinical practice. Supplementary evidence from other comparisons (ie, performance of stand-alone software vs unassisted reading) or non-comparative test accuracy studies (ie, AI-assisted reading or stand-alone software vs reference standard) were also reported where available. We calculated sensitivities and specificities in paired forest plots for the detection of any nodules, actionable nodules and malignant nodules. Where data allowed, we plotted our findings in receiver operating characteristic (ROC) space. Given the substantial heterogeneity in study populations, technologies, reader specialty and experiences, reference standards, test accuracy outcomes used and other study design features, no meta-analysis was carried out and findings are summarised narratively. Secondary outcomes such as reading time and impact on Lung-RADS ratings were summarised narratively.

RESULTS

Study selection

We retrieved 6330 unique results in January 2022, of which 4886 were published since 2012. Nine records were judged to be relevant,^{14–22} and two records were identified from other sources.^{23–24} Update searches in March 2023 yielded an additional 1687 results, only one was identified as potentially eligible²⁵ but was subsequently excluded. Eleven studies were, therefore, included (see online supplemental material 4 for full PRISMA flow diagram). Reasons for exclusions at full-text level are listed in online supplemental material 5.

Study characteristics

Characteristics of included studies are presented in table 1.^{14–24} They comprised 19 770 screened participants. There is potential for overlap as some studies may have sampled the same patients while using the same databases. Two studies used data from the Korean Lung Cancer Screening Project^{15–16} and four studies used US National Lung Screening Trial (NLST) data.^{18–20–22} Three studies were conducted in the USA.^{14–18–20} Two studies reported data from the same screening programme in South Korea.^{15–16} One study was conducted in each of the UK,²³ Taiwan¹⁷ and China.²¹ Two studies conducted in the Netherlands and Denmark²² and in South Korea,¹⁹ respectively, utilised CT scan images from the US NLST. The remaining reader study was conducted in the Netherlands using ultra-low-dose CT images from Russia.²⁴ Eight studies adopted an MRMC design.^{17–24} Two of these used unaided reading originally carried out as part of clinical practice for the comparators.^{21–23} Four studies sampled consecutive patients,^{15–16–21–23} and six used nodule-enriched samples,^{17–20–22–24} while the remaining study adopted random sampling.¹⁴

Six different AI-based software programs were used in the studies: *AI-Rad Companion* (Siemens Healthineers),¹⁴ *AVIEW Lungscreen* (Coreline Soft),^{15–16–24} *ClearRead* (Riverain Technologies),^{17–18–20} *InferRead CT Lung* (Infervision),²¹ *VUNO Med LungCT AI* (VUNO)¹⁹ and *Veolity* (MeVis).^{22–23}

Risk of bias and applicability

The evidence is of low quality. There were problems in most studies in almost all domains in terms of risk of bias and applicability, given the design and operationalisation of the studies and our UK/EU frame of reference (table 2 and online supplemental material 6). Risk of bias according to QUADAS-C was considered ‘high’ in three or more domains in five of the eight comparative studies.^{17–18–20–23–24} These issues included no consecutive or random sampling, test set laboratory studies in which radiologist behaviour is known to differ from clinical practice,²⁶ unpaired design (before/after study or different radiologists with and without AI) and/or suboptimal or biased reference standard.

Test accuracy

AI-assisted reading versus unaided reading

Eight studies reported on AI-assisted reading, where AI-based software was used concurrently (seven studies^{15–18–21–23–24}) or in addition sequentially (also referred to as ‘second-read AI’)¹⁷ to re-interpret images.

One study (described later) compared AI assisted radiographers (without prior experience in thoracic CT reporting) with unaided, experienced radiologists.²³ Across all remaining seven studies, the addition of concurrent AI to trained radiologists increased sensitivity and decreased specificity compared with unaided, trained radiologists. Two studies reported detection of actionable nodules (range: +5% to +13% for sensitivity; –3% to –6% for specificity)^{18–20} and one for detecting malignant nodules (+15% for sensitivity, –6% for specificity).¹⁸ Two studies reported detection of lung cancer through Lung-RADS category ≥ 3 (range, +3% to +7% for sensitivity; –8% to –6% for specificity),^{15–19} see figure 1 and online supplemental material 7. Concurrent AI-assistance also increased sensitivity (+20%) and decreased specificity (–7%) in nodule measurement and categorisation using a volume cut-off of 100 mm³.²⁴ For detection of nodules of any size, including nodules too small to be considered clinically actionable, radiologists’ sensitivity was increased with concurrent AI use (range, +16% to +56%), with an unclear impact on specificity (range, –3% to +4%).^{17–21} One of these studies¹⁷ evaluated both concurrent AI and second-read AI and found very similar sensitivity (79% vs 80%) and specificity (81% vs 82%), see online supplemental material 7 and 8.

For illustrative purposes (ie, the examples given here are plausible but hypothetical, given that test accuracy often changes as the screened population and disease prevalence varies, and the data were based on individual studies that used different AI software), if the changes in sensitivity and specificity for the detection of malignant nodules with concurrent AI assistance was in the range of those observed in the large screening programme reported by Hwang *et al*¹⁵ or in the MRMC study by Lo *et al*,¹⁸ and if the prevalence of lung cancer among the screening population was similar to that observed in the NELSON trial (ie, 0.5%),⁶ AI assistance would allow an additional 150–750 people attending screening with cancers to be detected but an additional 59 700 to 79 600 people attending screening without cancer would be placed on CT surveillance and/or further investigations per million people screened (equivalent to a reduction in positive predictive value of screening from 5% to 3%¹⁵ or from 3% to 2%, respectively¹⁸; online supplemental material 9).

Impact on Lung-RADS categorisation

Three MRMC studies provided comparative data on the impact of AI assistance on Lung-RADS categorisation of nodules.^{19–20–22}

Table 1 Characteristics of included studies

Study, country, software, time period for CT scan	Design, setting, sampling method and sample size	Index test and comparator(s) *	Reader details/reading conditions	Reference standard	Reported outcomes
Chamberlin <i>et al</i> USA AI-Rad Companion, prototype VA10A (Siemens Healthineers) January 2018–July 2019 ¹⁴	Retrospective test accuracy study (non-comparative). Random sample from 1 US centre; 117 LDCT scans.	(A) Stand-alone AI	NA	Nodules: Consensus of two expert radiologists	Accuracy for detecting nodules >6 mm (per person and per nodule analysis)
Hall <i>et al</i> UK Veolity, version 1.2 (MeVis) Date unclear: LSUT November 2015–July 2017 ²³	Retrospective test accuracy study and MRMC study (fully paired). Consecutive sample from UK-based LSUT; 735 LDCT scans.	(C) Concurrent AI (E) Original unaided reader	(C) MRMC: Two radiographers without prior experience in thoracic CT reporting. (E) Clinical practice (LSUT): Five radiologists with 5–28 years of experience in thoracic imaging (5% double reading).	Clinically significant nodules: Original radiologist reading or consensus of two independent radiologists after reviewing discrepant readings between study radiographers and original radiologists. Cancer: NR	Accuracy for detecting clinically significant lung nodules ≥5 mm; accuracy for detecting malignant nodules; reading time
Hsu <i>et al</i> Taiwan ClearRead CT, market version (Riverain Technologies) January–December 2017 ¹⁷	MRMC study (fully paired). Consecutive cases with nodules ≤10 mm or no nodules from one hospital in Taiwan; 150+ CT images (57 LDCT from lung cancer screening).	(B) Second-read AI (C) Concurrent AI (D) Unaided reader	MRMC: Three residents in radiology and three experienced chest radiologists. Two reading sessions with 8-week washout period: first unaided reading followed by second-read AI, then concurrent AI; images in random order.	Nodules: Consensus of two thoracic radiologists with >15 years of experience	Accuracy for detecting any nodules (stratified by seniority of readers)
Hwang <i>et al</i> South Korea AVIEW Lungscreen (Coreline Soft) April 2017–March 2018 ¹⁵	Before-and-after study (unpaired). Consecutive participants from K-LUCAS (11–14 institutions): 1821 participants (before); 4666 participants (after).	(C) Concurrent AI (E) Original unaided reader	Clinical practice: attending thoracic radiologists from 14 institutions (unpaired design)	Cancer: Review of medical records	Accuracy of detecting and categorising actionable nodules to detect lung cancer (Lung-RADS category ≥3)
Hwang <i>et al</i> South Korea AVIEW Lungscreen (Coreline Soft) April 2017–December 2018 ¹⁶	Retrospective analysis of prospective cohort study (non-comparative). 10 424 consecutive participants from K-LUCAS (14 institutions)	(C) Concurrent AI	Clinical practice: 25 radiologists from 14 institutions with 5–38 years of experience; no comparator.	Cancer: Review of medical records; lung cancer diagnosed within 1 year (primary outcome) or any time after LDCT (secondary outcome)	Accuracy of detecting and categorising actionable nodules to detect lung cancer (Lung-RADS category ≥3)
Jacobs <i>et al</i> USA/Denmark/ Netherlands Veolity, version 1.5 (MeVis) August 2002–August 2004 ²²	MRMC study (fully paired). Random sample from NLST (baseline and round 1), 40 cases from each Lung-RADS category; 160 LDCT scans.	(C) Concurrent AI (D) Unaided reader	MRMC: Three radiologists with >5 years of experience and four radiology residents. Two reading sessions with ≥2 weeks washout period: Half with AI support and half unaided per session; images in random order.	NA	Shift in Lung-RADS categorisation; Reading time
Lancaster <i>et al</i> Russia/Netherlands AVIEW LCS, version 1.0.34 (Coreline Soft) February 2017–2018 ²⁴	MRMC study (fully paired). Nodule-enriched: ≥1 solid nodule, no lung cancer diagnosed within 2 years from MLCS baseline scan; 283 ultra-LDCT scans.	(A) Stand-alone AI (C) Concurrent AI (D) Unaided reader	(C) MRMC: Three thoracic radiologists with >7 years of experience in lung cancer screening. (D) MRMC: Two different thoracic radiologists with >7 years of experience in lung cancer screening (using other semi-automated volume measurement software).	Nodules: Independent consensus of three experienced radiologists and one IT technologist	Accuracy of nodule volume measurement and categorisation (<100 mm³, ≥100 mm³)

Continued

Table 1 Continued

Study, country, software, time period for CT scan	Design, setting, sampling method and sample size	Index test and comparator(s) *	Reader details/reading conditions	Reference standard	Reported outcomes
Lo <i>et al</i> USA ClearRead CT (first generation, pre-market) (Riverain Technologies) Date unclear: NLST screened from August 2002 to September 2007 ¹⁸	MRMC study (fully paired). Nodule-enriched: 2 normal cases for each case with nodules from NLST and 2 US hospitals; 324 LDCT scans.	(A) Stand-alone AI (C) Concurrent AI (D) Unaided reader	MRMC: 12 general radiologists with 6–26 years of experience. Two reading sessions with a minimum interval of 37 days (mean, 57 days): first unaided; then AI-assisted.	Actionable nodules: Consensus of three expert thoracic radiologist assisted by corresponding NLST or source documentation. Cancer: Histological findings (presence) or long-term follow-up (absence).	Accuracy for detecting actionable nodules; accuracy for detecting malignant nodules; reading time
Park <i>et al</i> USA/South Korea VUNO Med-LungCT AI, v.1.0.1 (VUNO) Date unclear: NLST baseline screen from August 2002 ¹⁹	MRMC study (fully paired). Nodule-enriched from NLST baseline screens; 200 LDCT scans.	(A) Stand-alone AI (C) Concurrent AI (D) Unaided reader	MRMC: One radiology resident and four radiologists with 1–20 years of experience. Two reading sessions with 6-week washout period: first unaided, then AI-assisted; images in random order.	Cancer: NR (Lung cancer diagnosed within 1 year in the NLST)	Accuracy of detecting and categorising actionable nodules to detect lung cancer (Lung-RADS category ≥ 3) Change in Lung-RADS category
Singh <i>et al</i> USA ClearRead CT, market version (Riverain Technologies) Date unclear: NLST screened from August 2002 to September 2007 ²⁰	MRMC study (fully paired). Nodule-enriched: 100 with SSNs and 23 without SSNs from NLST; 123 LDCT scans.	(A) Stand-alone AI (C) Concurrent AI (vessel suppression only) (D) Unaided reader	MRMC: Two radiologists with 5 and 10 years of thoracic CT experience; sequential interpretation of unprocessed CT images alone, then vessel-suppressed images without washout period.	Sub-solid nodules: Consensus of two experienced thoracic radiologists (11 and 27 years of experience) with adjudication of conflicts by a third radiologist	Accuracy for detecting sub-solid nodules ≥ 6 mm; Change in Lung-RADS category
Zhang <i>et al</i> China InferRead CT Lung, market version (Infervision) November to December 2019 ²¹	Retrospective test accuracy study and MRMC study (fully paired). Consecutive sample from one hospital in China; 860 LDCT scans.	(C) Concurrent AI (MRMC study) (E) Original unaided reader (clinical practice)	(C) MRMC: One resident with 5 years of experience with supervision by one radiologist with 20 years of experience. (E) Clinical practice: One resident drafted report with supervision by radiologist (14 residents and 15 radiologists).	Nodules: Consensus of two radiologists with 20 and 31 years of experience	Accuracy for detecting any nodules (stratified by types and sizes of nodules)

*Index test and comparators: (A) Stand-alone AI: analysis of CT scan image by AI-based software without human input; (B) Second-read AI: CT scan image was first reviewed by an unaided human reader, then was re-interpreted after analysis by AI-based software was shown; (C) Concurrent AI: CT scan image was reviewed by a human reader assisted by concurrent display of analysis by AI-based software; (D) Unaided reader: CT scan image was reviewed by a human reader without assisted by AI-based software; (E) Original unaided reader: CT scan image was interpreted by a human reader as part of clinical practice, and therefore the reader was different from the human reader who interpret the CT scan image in the reader study.

†The study included mixed populations. Only those who underwent CT scans for screening were included in this systematic review.

AI, artificial intelligence; K-LUCAS, Korean Lung Cancer Screening; LDCT, low-dose CT; LSUT, Lung Screen Uptake Trial; Lung-RADS, Lung CT Screening Reporting & Data System; MLCS, Moscow Lung Cancer Screening; MRMC, multi-reader, multi-case study; NA, not applicable; NLST, National Lung Screening Trial.

The proportion of actionable nodules identified (Lung-RADS categories 3–4) was higher when images were assessed with AI assistance in all three studies (66% vs 53%,²² 34.2% vs 28.5%,¹⁹ 55% vs 50%²⁰). However, no reference standards were used, so it is not possible to know whether the additional actionable nodules were malignant.

Impact on CT scan reading time

Three comparative MRMC studies reported on the impact of AI assistance on reading times.^{18–22,23} Reading times were significantly faster with AI assistance compared with unaided readers: median 86 (IQR 51–141) seconds vs 160 (IQR 96–245) seconds ($p < 0.001$)²² and mean 98.0 seconds vs 132.3 seconds per case ($p < 0.01$)¹⁸ for radiologists, and median 3 (IQR 2–5) and 5 (IQR 4–8) min for radiographers using AI in a laboratory (ie, non-clinical) setting vs 10 (IQR 5–15) min for radiologists (unassisted reading in clinical practice).²³

Other methods of using AI (stand-alone AI and supporting less experienced staff)

Studies have also investigated other ways of using AI (comparing stand-alone AI with no human input to unaided radiologists or used AI to support less trained staff) or used non-comparative evidence (eg, AI-assisted reading or unaided reading compared with a reference standard). These are presented in online supplemental material 8.

Across studies and outcomes, stand-alone AI was associated with the highest sensitivity (range 58%–100%) but lowest specificity (62%–82%) when compared with AI-assisted radiologist reading (sensitivity 71%–99%, specificity 74%–97%) and/or unaided radiologist reading (sensitivity 43%–94%, specificity 63%–97%) (online supplemental material 8).^{18–20,24}

One study investigated whether AI assistance would support radiographers to match the accuracy of radiologists.²³ Experienced radiologists were more sensitive (91% vs 71%) and

Table 2 Limitations of the included studies

Reference and country	Applicability concerns regarding European screening population	Design concerns	Laboratory effect ²⁶	Applicability concerns to European screening LDCT reading practice	Applicability concerns—actionable nodule definition	Bias in reference standard	Module detection or measurement only	Accuracy reported per nodule only
Chamberlin <i>et al</i> USA ¹⁴	Yes— 1 US centre; 55–80 years. Excluded images rejected by software due to slice thickness >3 mm or poor image quality.	Yes— Non-comparative (A)	(A) NA	Yes— Stand-alone AI	Yes— >6 mm	Modules: Yes— <3 experienced chest radiologist; not blinded to index test	Yes— Detection of actionable nodules	No
Hall <i>et al</i> , UK ²³	No	Yes— Fully paired CT images but not the same readers with and without AI.	(C) Yes— MIRMC	Yes— (C) Two inexperienced radiographers. (E) 5% double reading.	No	Modules: Yes— Original radiologist report (E) plus radiologist review of scans with additional nodules detected by (C). Cancer: Unclear— Cancers detected either from baseline scan or nodule surveillance.	Yes— Detection of actionable nodules. Detection of malignant nodules ((C) only).	No
Hsu <i>et al</i> , Taiwan ¹⁷	Yes— One centre in Taiwan; 31/57 never smoked. Selected cases with nodules ≤1 cm or no nodules. Exclusion of cases with severe pulmonary fibrosis, diffuse bronchiectasis, extensive inflammatory consolidation, pneumothorax, and massive pleural effusion. 2.5 mm slice thickness.	No	(A) NA (B) (C) (D) Yes— MIRMC	Yes— (A) Stand-alone AI. (B) (C) (D) Three residents in radiology; three experienced chest radiologists.	Yes— No size cut-off used.	Modules: Yes— <3 experienced chest radiologists; not blinded to index test.	Yes— Detection of any nodules.	Yes— Per-nodule sensitivity; per-person specificity.
Hwang <i>et al</i> , South Korea ¹⁵	Yes— 11 (before) and 14 (after) institutions in Korea (K-LUCAS). Only 145/6,487 (2.2%) women.	Yes— (A) Non-comparative. (C) (E) Unpaired non-randomised; not the same readers with and without AI.	(A) NA (C) (E) No	(A) Yes— Stand-alone AI. (C) (E) No	(A) Yes— No size cut-off used. (C) (E) No	Modules: Yes— Single radiologist with second-read AI. Cancer: Yes— Medical record review.	(A) Yes— Nodule detection (any, actionable, malignant). (C) (E) No	(A) Yes—Per-nodule sensitivity. (C) (E) No
Hwang <i>et al</i> , South Korea ³³	Yes— 14 institutions in Korea (K-LUCAS). Only 283/10,424 (2.7%) women.	Yes— Non-comparative (C)	No	No	No	Cancer: Yes— Medical record review.	No	No
Jacobs <i>et al</i> , NL, USA, Denmark ²²	Yes— US NLS (baseline and round 1). Nodule-enriched; Lung-RADS ≥3 120/160 (75%). Slice thickness 1.0 to 3.2 mm.	No	(C) (D) Yes— MIRMC	Yes— Three radiologists with >5 year of experience and four radiology residents (fifth year).	No	NA	NA	NA
Lancaster <i>et al</i> , NL, Russia ²⁴	Yes— Ultra-LDCT (≤1 mSv); 50–80 years. Selected participants who had ≥1 solid nodule and did not develop lung cancer in following 2 years.	Yes— Not the same readers with and without AI.	(A) NA (C) (D) Yes— MIRMC	(A) Yes— Stand-alone AI (C) (D) No	No	Modules size and categorisation: Yes— 2/4 consensus panel readers involved in index test.	Yes— Nodule measurement and risk category.	No

Continued

Table 2 Continued

Reference and country	Applicability concerns regarding European screening population	Design concerns	Laboratory effect ²⁶	Applicability concerns to European screening LDCT reading practice	Applicability concerns—actionable nodule definition	Bias in reference standard	Nodule detection or measurement only	Accuracy reported per nodule only
Lo <i>et al</i> , USA ¹⁸	Yes—US NLST and 2 US hospitals; nodule-enriched (1:2); 3/178 nodules ≥ 3 mm.	No	(A) NA (C) (D) Yes—MVMC	Yes— (A) Stand-alone AI. (C) (D) 12 general radiologists.	No	Nodules: No Cancer: No	Yes— Detection of nodules (actionable, malignant).	Yes— Per-nodule sensitivity; per-person specificity.
Park <i>et al</i> , South Korea, USA ¹⁹	Yes—US NLST; Nodule and cancer-enriched; Prevalence of Lung-RADS ≥ 3 127/200 (64%); lung cancer prevalence 31/200 (16%).	No	(A) NA (C) (D) Yes—MVMC	Yes— (A) Stand-alone AI. (C) (D) 1 of 5 readers was a fourth-year radiology resident.	No	Cancer: Unclear— Same-year positive cancer diagnosis (not stated how diagnosed).	No	No
Singh <i>et al</i> , USA ²⁰	Yes—US NLST; enriched for sub-solid nodules; prevalence of sub-solid nodules 100/123 (81%).	No	(A) NA (C) (D) Yes—MVMC	(A) Yes— Stand-alone AI. (C) Unclear— AI for vessel suppression. (D) No	No	Nodules: No	Yes— Detection of actionable nodules.	Yes— Per-nodule sensitivity; Per-person specificity.
Zhang <i>et al</i> , China ²¹	Yes—One hospital in China (part of NELCIN-B3); general population aged 45–74 years.	Yes— Not the same readers with and without AI.	(C) Yes— MVMC (E) No	Yes— (C) One resident supervised by 1 radiologist; (E) 1 of 14 residents supervised by 1 of 15 radiologists.	Yes— No size cut-off.	Nodules: Yes— <3 experienced chest radiologists; not blinded to index test.	Yes— Detection of any nodules.	No
Legend	No = Random or consecutive screening LDCT images from heavy current or former smokers aged 50–75 years ²⁴ living in Europe; no inappropriate exclusions; ≤ 2 mm slice thickness.	No = Comparative, fully paired design; same readers with and without AI.	No = CT images assessed in clinical practice.	No = Single reading by experienced chest radiologist with (C) or without AI support (D) or (E).	No = In agreement with BTS ²⁵ , Lung-RADS ²⁶ or EUPS ²⁷ guidelines.	No = Nodules: ≥ 3 blinded, experienced chest radiologists. Cancer: Histopathology after biopsy/excision or 2-year follow-up without cancer diagnosis.	No = Accuracy of detection + risk categorisation + recall for lung cancer diagnosis.	No = Per-person sensitivity and specificity.

Index test and comparators: (A) Stand-alone AI: Analysis of CT scan image by AI-based software without human input; (B) Second-read AI: CT scan image first reviewed by an unaided human reader, then re-interpreted after analysis by AI-based software was shown; (C) Concurrent AI: CT scan image reviewed by a human reader assisted by concurrent display of analysis by AI-based software; (D) Unaided reader: CT scan image reviewed by a human reader without AI-based software assistance; (E) Original unaided reader: CT scan image interpreted by a human reader as part of clinical practice; different to the human reader who interpreted the CT scan image in the reader study.

AI, artificial intelligence; BTS, British Thoracic Society; Categ, categorisation; EUPS, European Position Statement; K-LUCAS, Korean Lung Cancer Screening Project; LDCT, low-dose CT; LSUT, Lung Screen Uptake Trial; Lung-RADS, Lung CT Screening Reporting & Data System; MVMC, multi-reader, multi-case study; NA, not applicable; NELCIN-B3, Netherlands-China Big-3 disease screening; NL, Netherlands; NLST, National Lung Screening Trial.

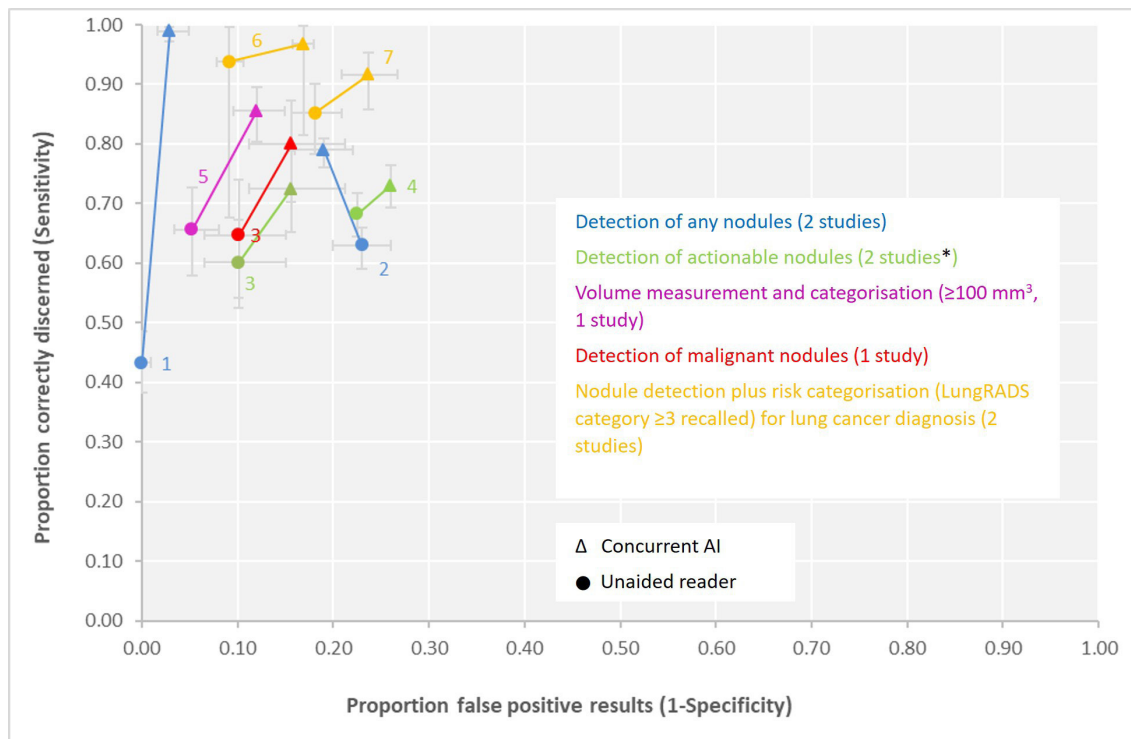


Figure 1 Accuracy of readers (nodule detection; nodule categorisation based on volume measurement; or nodule detection plus risk categorisation and recall decision for lung cancer diagnosis) both with and without concurrent AI use (seven studies with comparative data). Estimates connected with a line are from the same study. ¹ Zhang *et al*²¹; ² Hsu *et al*¹⁷; ³ Lo *et al*¹⁸; ⁴ Singh *et al*²⁰; ⁵ Lancaster *et al*²⁴; ⁶ Hwang *et al*¹⁵; ⁷ Park *et al*.¹⁹ *Data from Hall *et al*²³ are not presented as the study compared AI-assisted reading by radiographers against unaided radiologists, which differed in nature from the other studies. AI, artificial intelligence; Lung-RADS, Lung CT Screening Reporting & Data System.

specific (97% vs 92%) for detecting and categorising actionable nodules than AI-assisted reading by radiographers (without prior experience in thoracic CT reporting) (online supplemental material 8). Further decisions of experienced, unaided radiologists (made during clinical practice) were consistent with British Thoracic Society guidance 71.6% of the time, while the decisions of two radiographers with AI assistance in a laboratory setting were consistent with the guidance 39.7% and 60.7% of the time, respectively.

DISCUSSION

Summary of clinical context

Targeted lung cancer screening programmes are being set up in many countries due to strong randomised controlled trial (RCT) evidence that screening leads to a reduction in lung cancer-specific mortality. This will, however, place enormous pressure on already over-stretched healthcare systems, particularly in terms of scanner capacity and radiologist time. Different types of software using AI-derived algorithms have become available and could potentially reduce the screening workload and reading time for radiologists. These AI-based software, however, also have the potential to cause patient harm or create further workload for radiologists, and evidence is required to determine their performance in a screening context. Here, we have reported the results of a systematic review, synthesising the available evidence on the accuracy, reading time and impact on clinical management.

Statement of principal findings

Our searches yielded 6573 publications, from which 11 heterogeneous studies, reporting on nearly 20 000 patients from six

different countries and using six different AI-based software systems were included. All 11 studies were at high risk of bias with multiple applicability concerns. We used a narrative approach to summarise our results, finding that AI-assisted reading was faster and generally improved sensitivity (range: +5% to +20% for detecting/categorising actionable nodules; +3% to +15% for detecting/categorising malignant nodules), with lower specificity (range: -7% to -3% for correctly detecting/categorising people without actionable nodules; -8% to -6% for correctly detecting/categorising people without malignant nodules) compared with unaided reading. AI assistance tended to increase the proportion of nodules allocated to higher risk categories. If these findings were replicated in a population of a million people attending screening, the impact of AI would be an extra 150–750 cancers detected at the cost of 59 700–79 600 people receiving unnecessary surveillance, reducing positive predictive value.

Strengths and limitations

Our searches were extensive but limited by date (January 2012–March 2023). The 2012 cut-off was introduced after discussion with experts who considered that our definition of AI would not include systems introduced or tested prior to that date. Our searches are also limited to studies published in the English language although this is unlikely to have biased our findings.^{27 28} We aimed to include all AI-based software, which had (or was anticipated to have) appropriate regulatory marking (CE mark) across the UK and the EU, with anticipated availability for commercial use by 2023. However, our searches were inclusive, and we were unlikely to have omitted significant studies from our research because of this inclusion criterion.

QUADAS-2 was used independently by two reviewers¹² combined with the QUADAS-C tool for comparative studies,¹³ which we tailored to the review question to assess risk of bias and applicability. Almost all the studies fell short in key elements of quality, including patient selection, definition of reference standard, index test and flow and timing. The studies we identified were extremely heterogeneous using six different AI-based software systems and from at least six different countries, where the epidemiology of lung cancer, training of radiologists and experience of use of CT screening for lung cancer differ substantially. Therefore, we undertook a narrative review and plotted our findings in ROC space, however if it was possible, meta-analysis would allow for more precise estimates of the accuracy of the addition of AI-based software to CT lung cancer screening. We acknowledge that the potential benefit of AI assistance (150–750 additional lung cancers detected in a screened population of a million people) will depend on the prevalence of lung cancer in the cohort and as such is not generalisable to other populations at higher or lower risk. In addition, software derived from AI potentially allows continuous improvement of performance through learning from expanding sources of data. Although the various softwares evaluated in our review did not involve learning from data in real time, companies may refine their software by retraining their AI models with new datasets and then update the AI-derived algorithms used in the software periodically. Published evaluations on the performance of AI-based software in screening are, therefore, only a snapshot and could be outdated by the time when they are published, and our findings might not completely reflect systems that are currently available. The AI software that we evaluated only processed and utilised data from CT scan images to enhance nodule segmentation, detection and measurement that underpin current practice based on contemporary guidelines. Use of AI software to combine and interrogate additional morphological data from scan images (radiomics) along with a wide range of demographic, histological, proteomic and genomic data for prediction of nodules that are malignant is an area of very active research. These advances could fundamentally change clinical practice in the future. Nevertheless, it is crucial that any claims of improvement in risk stratification and cancer detection with AI software are supported by robust evidence generated from studies with strong designs that address risk of bias and applicability concerns that we highlighted.

Strengths and weaknesses versus other studies

We identified 12 previous systematic reviews on the accuracy of AI for lung nodule/cancer detection and/or malignancy risk prediction in medical images. Nine of these were non-comparative and focused on stand-alone AI performance of algorithms that were not commercially available, so were not informative for our review question (references are reported in online supplemental material 10). One rapid review²⁹ was comparative but focused on the accuracy of AI-based software for the classification of lung nodules into benign or malignant, a software function that was not included in our review.

Two reviews^{30 31} did cover our question but were broader and did not separately report on the screening population or on commercially available software. Li *et al.*³¹ evaluated the impact of AI on physicians' performance in detecting various thoracic pathologies on CT and chest X-ray. The review by Ewals *et al.*³⁰ was more relevant but covered not only the screening population but also the oncologic, symptomatic or mixed populations as well as software that was not commercially available. Of our

11 included papers, only one²⁰ was identified in the review by Li *et al.*³¹ and three^{17 18 21} in the review by Ewals *et al.*³⁰ Despite the broader population in the review by Ewals *et al.*, they found a similar pattern of increased sensitivity and reduced specificity with AI use. However, Li *et al.* found that, across all pathologies and both image types, both sensitivity and specificity generally improved when using AI-based devices. In concordance with our review, a faster reading time was reported with concurrent AI use in both previous reviews.^{30 31}

CONCLUSIONS AND IMPLICATIONS FOR CLINICIANS AND POLICYMAKERS

Our systematic review demonstrates that, when used in population-based lung cancer screening programmes, assistance of AI-based software can increase sensitivity, but at the expense of a reduction in specificity, that is, an increase in false-positive findings. The lung checks in the NHS England Targeted Lung Health Checks programme are already supported by AI,³² and removing AI-based software from existing screening programmes is not a practical policy option. However, the limited available evidence suggests that there is significant scope for improvement in the AI-based software, particularly in specificity. This is particularly important to consider as the screening programme is rolled out in the UK, given the potential increase in false-positive findings and the resulting additional workload for radiologists and anxiety for patients. Furthermore, care must be taken that AI-based software does not contribute to changing disease definitions or referral thresholds as the limited evidence base suggests its measurements and categorisations are more cautious and biased towards greater referral. Finally, more research is needed particularly in clinical settings and around the impact of AI assistance on medical staff with less training. Prospective, comparative, test accuracy studies that measure accuracy of the whole testing pathway with AI assistance integrated in clinical practice and compare it with the accuracy of the pathway without AI assistance are needed.

X Asra Asgharzadeh @Asra_aaa, Hesam Ghiasvand @GhiasvandHesam, Yen-Fu Chen @yfchen12 and Amy Grove @amylogrove

Acknowledgements We thank Pearl Pawson, Eileen Taylor and Sarah Abrahamson for their managerial and administrative support.

Contributors JG, AA, CS and Y-FC undertook the review with assistance from SJ and HG. AB devised the search strategy and undertook the searches in discussion with the other authors. EJH provided clinical advice. DG and MP provided statistical advice. AA, JG, CS, DT, PA, AT, AG, BS, AC, ST-P and Y-FC contributed to the conception of the work and interpretation of the findings. Y-FC, AC, ST-P, BS, JG and CS drafted the manuscript. All authors critically revised the manuscript and approved the final version. Y-FC takes responsibility for the integrity and accuracy of the data analysis. Y-FC acts as guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding This review was funded by the UK National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme (NIHR135325). The funder had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. DT and AG are partly funded by the NIHR West Midlands Applied Research Collaboration. STP is funded by the NIHR through a research professorship (NIHR302434). AG is supported by a NIHR Fellowship (NIHR300060).

Disclaimer 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.

Competing interests All authors have completed the ICMJE uniform disclosure. All authors involved in Warwick Evidence are wholly or partly funded by the NIHR. STP and AG are funded by the NIHR on personal fellowships. STP serves as Chair of the UK National Screening Committee Research and Methodology group, but this work is independent research not associated with that role.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iD

Yen-Fu Chen <http://orcid.org/0000-0002-9446-2761>

REFERENCES

- 1 Van Meerbeeck JP, O'Dowd E, Ward B, et al. Lung Cancer Screening: New Perspective and Challenges in Europe. *Cancers (Basel)* 2022;14:2343.
- 2 Lung Cancer Policy Network. Learning from Taiwan: implementing a national lung cancer screening programme. 2023. Available: <https://www.lungcancerpolicy.network.com/learning-from-taiwan-implementing-a-national-lung-cancer-screening-programme/>
- 3 O'Dowd EL, Lee RW, Akram AR, et al. Defining the road map to a UK national lung cancer screening programme. *Lancet Oncol* 2023;24:e207–18.
- 4 Guerrini S, Del Roscio D, Zanoni M, et al. Lung Cancer Imaging: Screening Result and Nodule Management. *Int J Environ Res Public Health* 2022;19:2460.
- 5 Bankier AA, MacMahon H, Goo JM, et al. Recommendations for Measuring Pulmonary Nodules at CT: A Statement from the Fleischner Society. *Radiology* 2017;285:584–600.
- 6 de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med* 2020;382:503–13.
- 7 Field JK, Vulkan D, Davies MPA, et al. Lung cancer mortality reduction by LDCT screening: UKLS randomised trial results and international meta-analysis. *Lancet Reg Health Eur* 2021;10:100179.
- 8 Pastorino U, Silva M, Sestini S, et al. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. *Ann Oncol* 2019;30:1162–9.
- 9 Crosbie PA, Gabe R, Simmonds I, et al. Yorkshire Lung Screening Trial (YLST): protocol for a randomised controlled trial to evaluate invitation to community-based low-dose CT screening for lung cancer versus usual care in a targeted population at risk. *BMJ Open* 2020;10:e037075.
- 10 Geppert J, Auguste P, Asgharzadeh A, et al. Diagnostic Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence – final protocol: software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images [DAP60]. NICE. 2021. Available: <https://www.nice.org.uk/guidance/dg55/documents/final-protocol>
- 11 McInnes MDF, Moher D, Thoms BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018;319:388–96.
- 12 Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- 13 Yang B, Mallett S, Takwoingi Y, et al. QUADAS-C: A Tool for Assessing Risk of Bias in Comparative Diagnostic Accuracy Studies. *Ann Intern Med* 2021;174:1592–9.
- 14 Chamberlin J, Kocher MR, Waltz J, et al. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value. *BMC Med* 2021;19:55.
- 15 Hwang EJ, Goo JM, Kim HY, et al. Implementation of the cloud-based computerized interpretation system in a nationwide lung cancer screening with low-dose CT: comparison with the conventional reading system. *Eur Radiol* 2021;31:475–85.
- 16 Hwang EJ, Goo JM, Kim HY, et al. Optimum diameter threshold for lung nodules at baseline lung cancer screening with low-dose chest CT: exploration of results from the Korean Lung Cancer Screening Project. *Eur Radiol* 2021;31:7202–12.
- 17 Hsu H-H, Ko K-H, Chou Y-C, et al. Performance and reading time of lung nodule identification on multidetector CT with or without an artificial intelligence-powered computer-aided detection system. *Clin Radiol* 2021;76:626.
- 18 Lo SB, Freedman MT, Gillis LB, et al. JOURNAL CLUB: Computer-Aided Detection of Lung Nodules on CT With a Computerized Pulmonary Vessel Suppressed Function. *AJR Am J Roentgenol* 2018;210:480–8.
- 19 Park S, Park H, Lee SM, et al. Application of computer-aided diagnosis for Lung-RADS categorization in CT screening for lung cancer: effect on inter-reader agreement. *Eur Radiol* 2022;32:1054–64.
- 20 Singh R, Kalra MK, Homayounieh F, et al. Artificial intelligence-based vessel suppression for detection of sub-solid nodules in lung cancer screening computed tomography. *Quant Imaging Med Surg* 2021;11:1134–43.
- 21 Zhang Y, Jiang B, Zhang L, et al. Lung Nodule Detectability of Artificial Intelligence-assisted CT Image Reading in Lung Cancer Screening. *Curr Med Imaging* 2022;18:327–34.
- 22 Jacobs C, Schreuder A, van Riel SJ, et al. Assisted versus Manual Interpretation of Low-Dose CT Scans for Lung Cancer Screening: Impact on Lung-RADS Agreement. *Radiol Imaging Cancer* 2021;3:e200160.
- 23 Hall H, Ruparel M, Quaife SL, et al. The role of computer-assisted radiographer reporting in lung cancer screening programmes. *Eur Radiol* 2022;32:6891–9.
- 24 Lancaster HL, Zheng S, Aleshina OO, et al. Outstanding negative prediction performance of solid pulmonary nodule volume AI for ultra-LDCT baseline lung cancer screening risk stratification. *Lung Cancer (Auckl)* 2022;165:133–40.
- 25 Wang D, Cao L, Li B. Computer-aided diagnosis system versus conventional reading system in low-dose (< 2 mSv) computed tomography: comparative study for patients at risk of lung cancer. *Sao Paulo Med J* 2022;141:89–97.
- 26 Gur D, Bandos AI, Cohen CS, et al. The “laboratory” effect: comparing radiologists’ performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 2008;249:47–53.
- 27 Morrison A, Polisenia J, Huseredu D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care* 2012;28:138–44.
- 28 Nussbaumer-Streit B, Klerings I, Dobrescu AI, et al. Excluding non-English publications from evidence-syntheses did not change conclusions: a meta-epidemiological study. *J Clin Epidemiol* 2020;118:42–54.
- 29 Lachance CC, Walter M. *Artificial intelligence for classification of lung nodules: a review of clinical utility, diagnostic accuracy, cost-effectiveness, and guidelines*. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health, 2020. Available: <https://www.cadth.ca/sites/default/files/pdf/htis/2020/RC1228%20AI%20for%20Lung%20Nodules%20Final.pdf>
- 30 Ewals LJS, van der Wulp K, van den Borne BEEM, et al. The Effects of Artificial Intelligence Assistance on the Radiologists’ Assessment of Lung Nodules on CT Scans: A Systematic Review. *J Clin Med* 2023;12:3536.
- 31 Li D, Pehrson LM, Lauridsen CA, et al. The Added Effect of Artificial Intelligence on Physicians’ Performance in Detecting Thoracic Pathologies on CT and Chest X-ray: A Systematic Review. *Diagnostics (Basel)* 2021;11:2206.
- 32 Aidence. Aidence is the preferred AI provider for the NHS England targeted lung health checks. 2021. Available: <https://www.aidence.com/news/aidence-ai-nhse-tlhc/>
- 33 Hwang EJ, Goo JM, Kim HY, et al. Variability in interpretation of low-dose chest CT using computerized assessment in a nationwide lung cancer screening program: comparison of prospective reading at individual institutions and retrospective central reading. *Eur Radiol* 2021;31:2845–55.
- 34 European Commission. European Health Union: a new EU approach on cancer detection – screening more and screening better. 2022. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_5562
- 35 Callister MEJ, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;70 Suppl 2:ii1–54.
- 36 American College of Radiology. Lung CT screening reporting & data system (lung-RADS) version 1.1. 2019. Available: <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads>
- 37 Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. *Lancet Oncol* 2017;18:e754–66.