

BMJ Open Evaluation of imaging techniques for early detection of intrathoracic cancers in symptomatic patients in primary care: a systematic review

Bogdan Grigore ¹, Jaime L Peters ¹, Wasim Hamad,² Natalia Calanzani ^{3,4}, Lauren Asare,⁵ Fiona M Walter,⁶ Richard Neal⁵

To cite: Grigore B, Peters JL, Hamad W, *et al.* Evaluation of imaging techniques for early detection of intrathoracic cancers in symptomatic patients in primary care: a systematic review. *BMJ Open* 2025;**15**:e091435. doi:10.1136/bmjopen-2024-091435

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-091435>).

Received 20 July 2024

Accepted 17 July 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY. Published by BMJ Group.

¹Exeter Test Group, Department of Health and Community Sciences, University of Exeter, Exeter, UK

²Institute for Health and Primary Care, Queen Mary University of London, London, UK

³Academic Primary Care, University of Aberdeen Institute of Applied Health Sciences, Aberdeen, UK

⁴Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁵University of Exeter, Exeter, UK

⁶Wolfson Institute of Population Health, Queen Mary University of London, London, UK

Correspondence to

Mr Bogdan Grigore;
b.grigore@exeter.ac.uk

ABSTRACT

Objectives Intrathoracic cancers, such as lung cancer, mesothelioma and thymoma, represent diagnostic challenges in primary care. We aimed to summarise evidence on the performance of imaging techniques that could aid the detection of intrathoracic cancers in low prevalence settings.

Design Systematic review and quality appraisal using Quality Assessment of Diagnostic Accuracy Studies-2 and Grading of Recommendations Assessment, Development and Evaluation.

Data sources MEDLINE, Embase and Web of Science were searched with a predesigned search strategy for articles from January 2000 to January 2024.

Eligibility criteria We included studies relevant for primary care, where participants were suspected of having intrathoracic cancer and reported on at least one diagnostic performance measure. We excluded studies where the cancer diagnosis was already established. Data extraction and synthesis screening were conducted independently by two reviewers. Data extraction and quality appraisal were conducted by one reviewer and checked by a second reviewer.

Results Out of 30 539 records identified by the database searches, 13 studies were included. There was heterogeneity in the types of cancers, populations included and reported diagnosis pathways for suspected cancers. Imaging modalities investigated included chest X-ray (three studies), computer tomography (CT, six studies), magnetic resonance imaging (two studies), positron emission tomography CT (two studies), ultrasound (two studies) and scintigraphy (one study). Chest X-ray sensitivity reported for lung cancer ranged from 33.3% to 75.9%, with specificity ranging from 83.2% to 95.5%. For CT, reported sensitivity varied from 58% for pleural malignancy to 100% for lung cancer. One study investigating an artificial intelligence tool to detect lung cancer found poor detection performance in a real-world patient cohort.

Conclusions We found a limited number of studies reporting on the diagnostic performance of usual imaging techniques when used in unselected primary care settings for the diagnosis of intrathoracic cancer in symptomatic patients. There is a need for more studies evaluating such techniques in the general population presenting in primary care, where the prevalence is relatively low.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ The systematic review used a comprehensive search of the literature for studies reporting on diagnostic performance of imaging in primary care patients.
- ⇒ Studies reporting on impact of diagnostic imaging in primary care, as well as studies reporting the use of artificial intelligence in the primary care setting, were also sought.
- ⇒ We used Quality Assessment of Diagnostic Accuracy Studies-2 to evaluate the risk of bias and the Grading of Recommendation, Assessment, Development and Evaluation approach to evaluate the strength and quality of the evidence.
- ⇒ Very few studies focused on primary care symptomatic patients, most of them included small samples and reported limited accuracy data.

A better understanding of the performance could lead to better detection strategies for intrathoracic cancers in primary care. Intrathoracic cancers, such as lung cancer, mesothelioma and thymoma, represent diagnostic challenges in primary care. We aimed to summarise evidence on the performance of imaging techniques that could aid the detection of intrathoracic cancers in low prevalence settings.

BACKGROUND

Intrathoracic cancers including lung cancer, mesothelioma and thymoma remain a diagnostic challenge as they have similar presentation to other conditions. They can be ‘easily missed’ in primary care and are often diagnosed at a late stage.¹ However, many patients experience symptoms and may have detectable imaging findings prior to their diagnosis. Thereby, creating a time window in which imaging may facilitate detection at earlier stages, leading to less invasive treatments and improving survival.²

Imaging is fundamental to the investigation of suspected intrathoracic malignancies.

During their clinical journey, patients with suspected intrathoracic cancer may undergo a number of imaging modalities. Chest X-ray (CXR) is usually the first diagnostic imaging in suspected intrathoracic cancer. The England and Wales National Institute for Health and Care Excellence (NICE) guidelines³ recommend a CXR within 2 weeks when a lung or pleural malignancy is suspected (if the patient is over 40 years old and presents with two or more suggestive symptoms).⁴ While CXR is a relatively accessible investigation in primary care, existing evidence suggests that its sensitivity for lung cancer is limited,⁵ especially for early stages,⁶ with recommendations for general practitioners (GPs) to seek further investigations in high-risk patients with negative CXR.^{5,7}

The second most used imaging technique for diagnosing intrathoracic cancers is CT.⁸ Patients with suspected intrathoracic cancer are often offered contrast-enhanced CT to further the diagnosis or to provide information on the cancer staging.⁴ CT is used as part of different investigative approaches, such as screening,⁹ and biopsy guidance,¹⁰ and cancer staging.^{3,7}

Positron emission tomography CT (PET-CT) is an imaging technique that employs radioactive contrast substances to generate gamma rays that are captured as images.⁸ Many guidelines^{4,11} recommend that all patients with potentially curable cancer have an initial evaluation with PET-CT to evaluate for metastasis and establish cancer stage.

While not recommended for the investigation of primary lung cancer,⁷ MRI is useful in identifying rarer forms of cancer, for example, pleural mesotheliomas¹² alongside CT.⁸ Ultrasound (US) techniques are not usually used for investigating suspected intrathoracic cancers in most settings; however, they can provide valuable information on the condition of the chest wall, pleurae and lymph nodes.⁸

Scintigraphy is an imaging technique where gamma radiation is captured (as in PET-CT);¹³ however, it lacks three-dimensional information. Scintigraphy is a less used method due in part to the greater accessibility of other imaging techniques.¹⁴ Recent studies suggest that scintigraphy with somatostatin analogues may help differentiate between malignant and benign tumours.^{15,16}

There is increasing evidence on the performance of these and other imaging techniques; however, most of this evidence is concerned with populations in secondary and tertiary care, where cancer prevalence is higher than that of the general population. There has been, however, less research into the use of imaging in primary care for cancer detection, where cancer prevalence is much lower. The distinction is important as these techniques may not perform as well in low prevalence populations due to spectrum bias effect;¹⁷ they are likely to have a lower sensitivity and higher specificity in primary care.

General practitioners play a crucial role in investigating or referring symptomatic patients along appropriate diagnostic pathways. With increasing pressure on healthcare systems, there is a need to avoid, where feasible, referring

symptomatic patients where a cancer diagnosis is unlikely. Currently, many countries, including the UK, are implementing CT screening of high-risk asymptomatic populations. However, this will only account for the diagnosis of a small percentage of all intrathoracic cancers, with the majority continuing to be diagnosed through a symptomatic route. Furthermore, accuracy evidence from screening studies, while largely available, may not pertain to symptomatic patients as the underlying risk of cancer is different.

The current study was conducted within the CanTest framework,¹⁸ an initiative with the aim of improving detection of cancer in primary care, in view of more effective implementation into practice.¹⁸ For this systematic review, we aimed to identify and synthesise studies reporting on the performance, feasibility and impact of imaging techniques for the detection/diagnosis of intrathoracic (primary lung/bronchus/trachea, mesothelioma and thymoma) cancers, which have potential relevance to primary care.

METHODS

This systematic review was developed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines¹⁹ and was registered in the PROSPERO database (CRD42023407698).

Search strategy

Scoping database searches suggested that a high number of papers would need to be screened, due to many potentially relevant reports being indexed in non-related categories (eg, papers that may be applicable to primary care but do not mention this setting). Our search strategy was designed to provide a reasonable balance between sensitivity and number of papers to be screened. In order to achieve this, it was conducted in two steps. In the first step, we conducted a database search for studies reporting on diagnostic performance of imaging techniques. We searched MEDLINE, Embase and Web of Science with a predesigned search strategy (using free text and database-specific subject headings, see online supplemental material, appendix 1). The search was restricted to human studies published after 1 January 2000, in any language. Potentially relevant papers identified from other sources (such as newsletters and discussions with colleagues) were also considered. In the second step, we also conducted a search of references of included papers, as well as papers citing these papers. In this step, we also included any available studies reporting on implementation (such as feasibility and acceptability) and impact (such as clinical utility and/or cost-effectiveness).

Inclusion and exclusion criteria are described in [table 1](#).

Study screening

Study screening adhered to PRISMA flowchart.¹⁹

Studies identified from searches were screened for relevance based on title and abstract. Screening was conducted

Table 1 Inclusion and exclusion criteria

	Inclusion	Exclusion
Population	Adults with a suspicion of primary intrathoracic cancers (lung/bronchial/tracheal cancer, mesothelioma and thymoma), but where cancer diagnosis is not established at recruitment.	Adults with established/known cancer diagnosis at recruitment, or in participants with intrathoracic metastases or history of intrathoracic cancers Asymptomatic adults with average risk (ie, screening) Studies in vitro or on animals
Intervention	Imaging techniques that would be suitable for use to aid diagnosis of intrathoracic cancer in primary care.	Biopsy procedures
Comparators	Individuals who do not have a diagnosis of cancer—non-malignant conditions such as benign nodules, COPD or others are eligible comparators	—
Outcome	Diagnostic accuracy/performance of imaging techniques including sensitivity, specificity, positive predictive value, negative predictive value and area under the curve. Where reported, data on the implementation (eg, feasibility and acceptability) of imaging in this context, as well as the impact of imaging, such as clinical utility (eg, reduced time to diagnosis or treatment) and/or cost-effectiveness	
Publication type	Quantitative studies reporting results from primary data (RCTs and observational studies) published in peer-reviewed journals	Abstracts presented at conferences/meetings without an accompanying full-text publication; qualitative studies; reviews, systematic reviews and overviews

COPD, chronic obstructive pulmonary disease; RCTs, randomised controlled trials.

independently by two reviewers using Rayyan.²⁰ A pilot exercise took place between all reviewers for a number of the hits, to ensure inter-reviewer consistency.

Discussion between reviewers or with a third reviewer resolved any disagreement. Full text of studies included at the title and abstract screening stage was then screened independently by two reviewers.

Data extraction

A data extraction form was developed and piloted to ensure consistency in extraction between studies and reviewers. Three included papers were used in the pilot phase and the data extraction form was amended as necessary. The process was iterative, with data being extracted independently by each reviewer, followed by discussion as to the clarity and completeness of the data extraction form and individual questions. Data extraction was then carried out independently by one reviewer and checked by a second. Disagreements and differences in data extraction were resolved by discussion and consensus, where necessary. While the experience and area of expertise of the reviewers in the team varied, two experienced systematic reviewers (BG, JP) ensured the process was methodologically rigorous.

We extracted information on: study characteristics (publication year, country of population of interest, recruitment setting, study aims and design); populations (sample size, age, sex, tumour staging for people with cancer and health status for people without cancer); imaging technique (including technical specifications/characteristics) and summary measures of diagnostic

performance (sensitivity, specificity, positive predictive value, negative predictive value, area under the curve). Where reported, we also extracted data on the implementation (eg, feasibility and acceptability) of imaging in this context, as well as the impact of imaging, such as clinical utility and/or cost-effectiveness.

Quality assessment

To assess the risk of bias (RoB), we used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool²¹; for comparative accuracy, we also used Quality Assessment of Diagnostic Accuracy Studies-Comparative (QUADAS-C).²² QUADAS-2 evaluates the RoB based on four domains (patient selection, index test, reference standard, flow and timing). The first three domains are also assessed in terms of applicability. QUADAS-C is an extension of QUADAS-2, based on the same four domains, but with additional questions relevant to test comparisons.²² For studies reporting outcomes related to implementation or impact, we planned to use the Cochrane Collaboration's 'Risk of Bias' Tool²³ or the ROBINS-I (Risk of Bias in Non-randomised Studies-Interventions) tool.²⁴ For quality appraisal, two reviewers assessed the included papers independently, after piloting the approach on the first three papers; disagreements were discussed and consensus was reached. No automation tool was used in the process.

Data synthesis

Extracted data were presented narratively and in summary tables by imaging technique. All relevant outcome

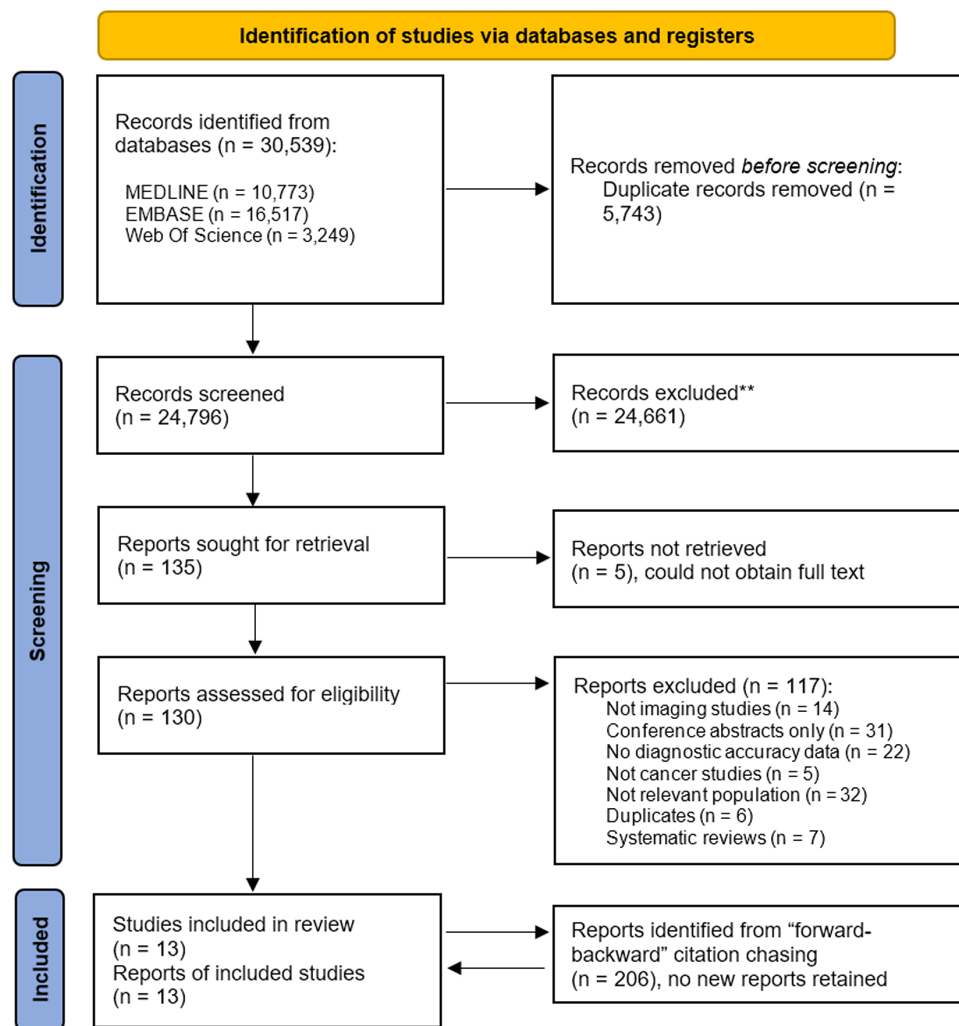


Figure 1 PRISMA diagram of screened papers. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

measures from the included studies were reported. Where available, implementation and impact outcomes (including feasibility, acceptability and cost-effectiveness) were also synthesised using narrative synthesis. Diagnostic results were mainly summarised in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Where these data were not available, but contingency tables could be reconstituted, we reported data from these tables instead. Key critical appraisal aspects were also included in the narrative synthesis.

Heterogeneity across included studies in imaging techniques, populations, study designs and outcome measures precluded meta-analysis.

Quality of GRADE evidence

To evaluate the quality of evidence, we applied the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology²⁵ adapted for tests.²⁶ Two researchers (BG and JP) independently conducted the assessment. Disagreements were discussed and, when necessary, a third researcher (RN) helped the

reach of a consensus. The final quality of evidence was categorised into one of three levels—high, moderate, low and very low.

Patient and Public Involvement (PPIE)

A patient and public involvement and engagement (PPIE) panel was established and consulted throughout the project. In the development of the proposal for funding, and in the development of the protocol, the PPIE panel was very supportive of the need for the work and the approaches that we were taking. Two specific issues were raised. First, the suggestion that we should measure test performance in different ethnic groups led to the potential inclusion of subgroup analysis if the data permitted. The second related to measuring harms as a consequence of testing, as well as issues relating to the process and acceptance to testing (including issues such as waiting times, the need to travel, the need for 'preparation', the stress and anxiety working up to the test). We sought to report on any of these aspects in the included studies.

At the end of the study, we held a PPIE workshop where we discussed the meaning of the results. The PPIE members were supportive of our findings and conclusions, but again raised issues (beyond the scope of the review) about the process and acceptability of testing.

RESULTS

Study screening

Database searches were run in January 2023 and resulted in 21 567 potentially eligible studies. A search update was conducted in January 2024, which resulted in 3399 additional papers. After deduplication, 24 796 studies were identified for title and abstract screening, and 130 papers were assessed in full-text screening. The most likely reasons for exclusion at full-text level were the population not being compatible with our inclusion criteria, the non-availability of full text and the lack of reported accuracy data. 13 papers were included in our review (see figure 1).

An additional search was carried out to identify references and citations of the included papers. 206 papers were identified. After title and abstract screening, two papers were screened in full. No additional papers were included following this step.

Of the 13 studies included, five were conducted in the UK,^{27–31} four in China,^{32–35} two in Germany^{36 37} and one each in France³⁸ and Italy.³⁹ Three studies reported on CXR,^{27–29} six on CT,^{30–32 36–38} two on MRI,^{32 37} two on PET-CT,^{33 35} two on ultrasound^{34 39} and one on scintigraphy.³³ Reference standards used varied across the reports from biopsy and histopathologic analysis,^{30–35 37 38} clinical reports^{28 29 39} to registries^{27 36} and death records.³⁶

Study sample size varied greatly, from 28³⁷ to 8996.²⁸ Most participants were selected based on the symptomatic/clinical suspicion of lung cancer, with the exception of two studies, where asymptomatic, but high-risk (due to asbestos exposure) patients were included,^{36 38} and one study where participants presented with respiratory symptoms, but not specific to cancer.³⁹

A summary of included studies is presented below, starting with those including symptomatic patients, followed by studies including asymptomatic patients (see online supplemental material, appendix 2 for all extracted data). Summaries of the risk of bias are presented in table 2, with estimates of accuracy shown in figure 2 (and online supplemental material, appendix 3). The reporting of accuracy data was heterogeneous. Where possible, 2×2 contingency tables were extracted or reconstructed by the review authors to report estimates of

Table 2 Critical appraisal of studies according to QUADAS-2

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
CXR							
Bhartia <i>et al</i> 2021 ²⁷	⊙	⊙	⊙	?	⊙	⊙	⊙
Bradley <i>et al</i> 2021 ²⁸	⊙	⊙	?	?	⊙	⊙	?
Maiter <i>et al</i> 2023 ²⁹	⊙	⊙	⊙	⊙	⊙	⊙	⊙
Clin <i>et al</i> 2009 ³⁸	?	⊙	?	?	?	⊙	?
CT							
Bai <i>et al</i> 2022 ³²	?	?	?	⊙	?	?	?
Felten <i>et al</i> 2014 ³⁶	?	⊙	⊙	⊙	?	⊙	⊙
Traill <i>et al</i> 2001 ³⁰	?	⊙	⊙	⊙	⊙	⊙	⊙
Tsim <i>et al</i> 2017 ³¹	⊙	⊙	?	⊙	⊙	⊙	⊙
MRI							
Heye <i>et al</i> 2012 ³⁷	?	⊙	⊙	⊙	⊙	⊙	⊙
PET-CT							
Yang <i>et al</i> 2022 ³⁴	⊙	?	⊙	⊙	⊙	⊙	⊙
Zhou <i>et al</i> 2020 ³⁵	?	⊙	⊙	⊙	?	⊙	⊙
Ultrasound							
Rinaldi <i>et al</i> 2020 ³⁹	⊙	?	?	?	⊙	⊙	?
Scintigraphy							
Wang <i>et al</i> 2007 ³³	?	⊙	?	?	?	⊙	?

⊙: low risk; ⊙: high risk; ?: unclear risk.

CXR, chest X-ray; PET-CT, positron emission tomography CT; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2.

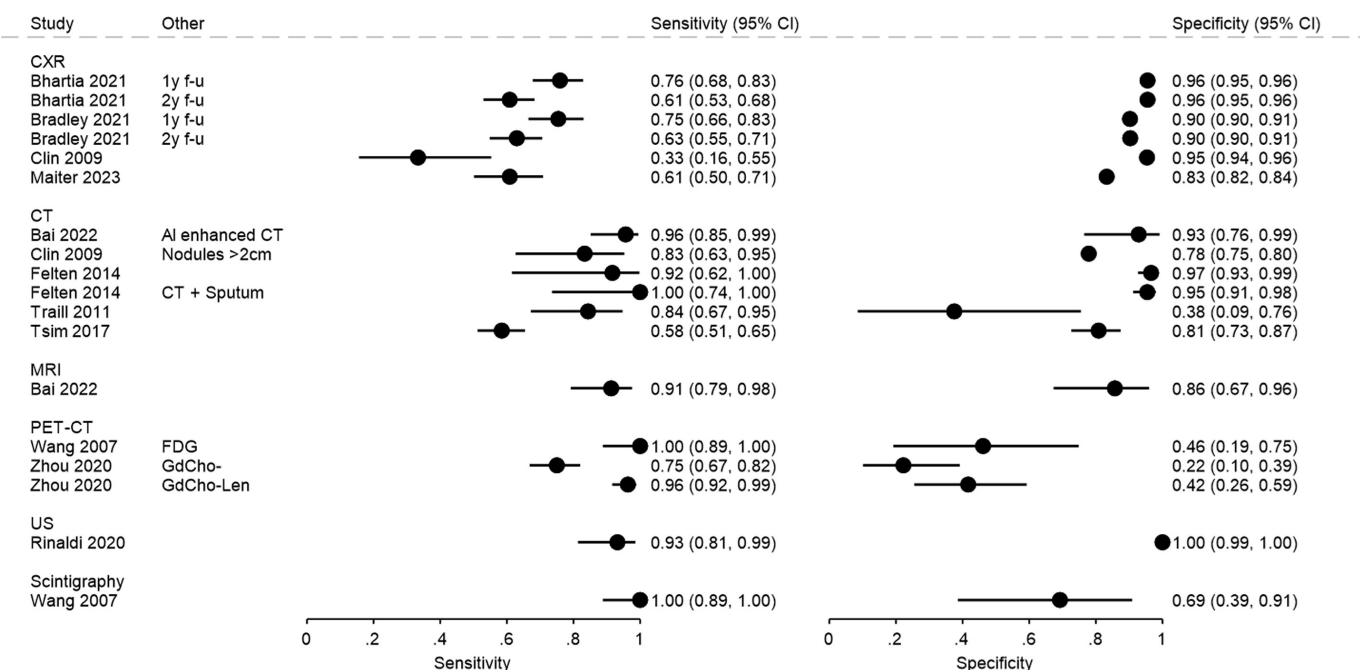


Figure 2 Accuracy data extracted from the included papers.

sensitivity, specificity, PPV and NPV. The majority of the studies did not report area under receiver operator characteristics (AUROC). Only Yang *et al*³⁴ report this data explicitly, while Rinaldi *et al* report AUROC for all diagnoses (of which cancer diagnoses represented only 11%) and Zhou reports a graph indicating the relationship between specificity and sensitivity, without any further detail.

Studies including symptomatic patients

Chest X-ray studies

Bhartia *et al*²⁷ reported on a prospective study where patients with symptoms deemed eligible were able to self-refer for a CXR. 8948 examinations of patients aged 50 years or more were included. The reference standard was diagnosis of any primary intrathoracic cancer or specifically non-small cell lung cancer (NSCLC), within 1 year and 2 years from imaging. There were no significant differences between the sensitivity and specificity of detection for all intrathoracic cancers and NSCLC. This study was appraised as having low risk of bias, with minimal concerns in all appraised domains.

Bradley *et al*²⁸ used a cohort from the same study as Bhartia *et al*²⁷ and explored accuracy of CXR in conjunction with relevant, but non-specific symptoms (eg, cough, haemoptysis, breathlessness, chest pain, weight loss, change in voice) to assist diagnosis of (any) lung cancer. The authors reported a sensitivity of CXR to detect lung cancer within 1 year to be 75.4% (CI 67.5% to 83.3%), and specificity 90.2% (CI 89.6% to 90.9%). These figures are not directly comparable to findings by Bhartia *et al*, as there were slight differences in the samples (8996 vs 8948 participants) due to different exclusion criteria, as well as in the outcomes (lung cancers vs all intrathoracic cancers and NSCLC only, respectively). Bradley *et al* reported PPV

for presence of symptoms in case of a negative CXR and found the PPV of all symptoms for a diagnosis of lung cancer within 1 year of CXR to be <1% for all individual symptoms except for haemoptysis, which had a PPV of 2.9%. This study generally presented low risk of bias and low concerns; however, the risk of bias concerning the reference standard was deemed unclear, due to relying on the hospital database; presumably, if patients moved away, their diagnoses would no longer be recorded.

Maiter *et al*²⁹ evaluated the diagnostic performance of an artificial intelligence (AI) solution for lung nodule detection (Auto Lung Nodule Detection software, ALND; Samsung Electronics, Suwon, South Korea) applied retrospectively to a set of CXRs acquired on the request of GPs. 5722 CXR were included in the study and the detection performance of the AI software was assessed against the cancer diagnosis from a multidisciplinary team. The AI software underperformed against the multidisciplinary team with a PPV of 5.6% (95% CI 4.8% to 6.6%). According to the authors, this may be due to the AI training and validation using “insufficiently representative and generalisable datasets”²⁹ (enriched datasets from secondary centres in Korea, Germany and the USA), reflecting a higher prevalence than in primary care. The study was assessed as having low risk of bias and low applicability concerns.

CT studies

The study conducted by Traill *et al*³⁰ aimed to determine the sensitivity and specificity of contrast-enhanced CT (CE-CT) for the detection of pleural malignancy in patients referred for the diagnosis of a suspected malignant pleural effusion (MPE). 40 consecutive patients participated in the study, of whom 32 had a malignant pleural effusion. Overall, based on figures reported by

the authors, sensitivity for CE-CT was 84.4% for detecting MPE, and the specificity was 62.5%. Risk of bias was assessed as low for the index test, reference standard and patient flow, but high for patient selection, due to limited information reported on this aspect. The small sample reported and the high prevalence of malignancy also raised concerns.

Tsim *et al*³¹ conducted a retrospective analysis of 315 consecutive patients who underwent CE-CT for suspected pleural malignancy. The reference standard was a mix of cytological or histological confirmation, or follow-up of disease progression for at least 6 months. The authors reported CE-CT to have an overall sensitivity of 58% and a specificity of 80% for detecting pleural malignancy; however, only 59 (19%) of the 195 cases detected represented primary mesothelioma, with the rest indicating secondary malignancies. Due to this, the study raised several concerns for high risk of bias regarding the sample included. Another concern was raised by the study using multiple reference standards, varying from histopathological confirmation to consensus over progression in the multidisciplinary team.

CT and MRI studies

Bai *et al*³² explored the diagnostic effect of high-resolution reorganisation of medical images based on deep convolutional neural networks (CNN) in lung cancer diagnostics. The study included 74 patients highly suspected of lung cancer based on symptoms (cough, vocal cord tearing, expectoration and blood in sputum). Patients were examined by CT and MRI, and images from both were enhanced with a deep CNN algorithm. All patients underwent biopsy for diagnostic confirmation. CT-guided diagnostic accuracy was calculated as 94.6%, while MRI-guided diagnosis was 89.2% (see [figure 2](#)). Authors concluded that both techniques were highly accurate, although they performed differently from one another: CT had shorter procedure time, while it involved radiation, and MRI had longer time, while not involving any radiation. The study was appraised as having unclear risk of bias for the patient selection, index test and interpretation of the reference standard, due to limited information reported on how the patient sample was obtained and the context of index test and reference standard interpretation (eg, timing, blinding). There were also some concerns regarding the applicability of the index test, given that only patients highly suspected of cancer were included in the study.

MRI studies

Heye *et al*³⁷ reported on a study that assessed the accuracy of lesion detection by MRI against CT as reference standard. The prospective study included 28 patients with suspected thoracic malignancy and clinical indication of chest diagnostic imaging. All patients underwent MRI and CT. The authors found MRI to be concordant with CT in detecting primary thoracic malignancies, as well as ruling out pulmonary lesions; however, MRI detected only 28%–36% (depending on the sequence used) of

the ≥ 5 mm nodules detected by CT (49%–59%). The study had low risk of bias on the conduct of index test and patient flow, with unclear level of risk of bias for patient selection (due to limited information reported on this step) as well as reference standard (with both MRI and CT conducted by the same observers). There were significant applicability concerns regarding the sample used (26 of the 28 participants had malignancies) as well as the use of CT as reference standard, despite having access to pathophysiological information (ie, which could have offered an arguably better reference standard).

PET-CT studies

In the study by Zhou *et al*,³⁵ 172 patients with suspected lung cancer underwent PET-CT with Gd₂O₃-doped carbon-11-choline-levatinib (GdCho-Len-PET) nanoparticles contrast agent in order to evaluate its diagnostic value. The authors found that GdCho-Len-PET presented with higher sensitivity (96.3%) and specificity (42.7%) compared with GdCho-PET contrast agent (75% and 22.2% respectively). There were some concerns related to the index test used, both in terms of risk of bias and applicability, due to very early use of PET-CT in the diagnostic pathway.

Ultrasound studies

In their 2020 study, Rinaldi *et al*³⁹ explored the utility of role of lung ultrasound (LUS) in the clinical diagnostic algorithm of respiratory diseases. 509 patients admitted for respiratory symptoms were prospectively included and evaluated with LUS. Authors found that LUS findings were highly concordant with the reference standard (41/44 cases were found positive in LUS); however, this was mostly due to the frequent finding of indirect signs such as pleural effusion or thickening, large areas of atelectasis and secondary peripheral lesions. Due to the limited reported information specifically on cancer diagnosis, the risk of bias could not be determined on most domains. There were significant concerns regarding the population included in the study, as it included patients without suspected intrathoracic cancer.

Yang *et al*³⁴ conducted a study exploring the diagnostic capabilities of high-frequency B-mode US and contrast-enhanced US (CEUS) in terms of differentiating between benign and malignant pleural diseases in a study that enrolled 50 consecutive patients. The authors found that both B-mode US and CEUS were able to clearly distinguish between benign and malignant thickened pleurae (area under the curve 0.819 for B-mode US and 0.848 for CEUS). The study was evaluated as having a low risk of bias on patient selection, reference standard and patient flow, but unclear risk of bias on the index test. This was due to the uncertain definition of a positive result. The study was judged as high risk of bias for applicability of the index test and reference standard.

Scintigraphy study

Wang *et al*³³ evaluated the clinical value of tomographic ^{99m}Tc-octreotide scintigraphy in comparison with that of fluorine-18 fluorodeoxyglucose (¹⁸F-FDG) dual-head coincidence imaging (DHC) in a sample of 44 consecutive patients with suspected lung cancer. Diagnosis was confirmed by histopathologic findings or clinical follow-up. The authors concluded that the sensitivity (100%) and specificity (69.2%) of tomographic ^{99m}Tc-octreotide scintigraphy were comparable to those of ¹⁸F-FDG coincidence imaging (100% and 46.2%, respectively) for the primary lesions. The risk of bias could not be determined for most of the domains due to lack of details on the reference standard, length of follow-up and reporting of blinding. There were concerns regarding applicability to a primary care setting as there was a high prevalence of cancer (31 of the 44 patients had cancer).

Studies including asymptomatic patients

CXR and CT study

The study by Clin *et al*³⁸ explored the performance of CXR and CT in screening for lung cancer in an asbestos-exposed population. 972 patients underwent 1230 investigations with CXR and CT to identify early lung cancer. Bronchopulmonary cancer was detected with a rate of 0.82% for CXR and 2.16% for chest CT. At a threshold of 2 mm nodules, CXR had a sensitivity of 33% and a specificity of 95%, while CT had a sensitivity of 83% and a specificity of 78%.

The study exhibited low risk of bias regarding the interpretation of the index tests (as chest radiographs were interpreted independently from the CT scans); however, the risk of bias could not be determined for the other domains, such as the reference standard (only test positive patients having received a biopsy, while data for test negative patients were obtained through follow-up) as well as the relevance of the population (a screening, high-risk population).

CT study

In a study by Felten *et al*,³⁶ 187 high risk, asymptomatic, asbestos-exposed patients were screened for lung cancer with low-dose CT and sputum cytology; participants were followed up for more than 3 years. Reference standard was considered as confirmation biopsy where available, records in an occupational registry or death certificates. On critical appraisal, this study raised considerable concerns on the patient selection domain (with highly selected, but asymptomatic patients), as well as some concerns on the reference standard used (due to reliance on an occupational registry for some of the patient outcomes) and patient flow (due to some patients receiving a different reference standard).

Quality appraisal

Overall, studies varied greatly in the assessed risk of bias and applicability through QUADAS-2 (see table 2 and online supplemental material, appendix 4). The GRADE

assessment downgraded a number of studies, and only Bhartia, Bradley, Maiter, Tsim and Yang were judged as high quality of evidence (see online supplemental material, appendix 5).

DISCUSSION

Our study employed a search strategy that resulted in a significant number of records to screen for eligibility, but only 13 studies met our inclusion criteria. All included studies were diagnostic accuracy studies. Although the search strategy was developed to identify diagnostic accuracy studies, the additional searching undertaken did not identify any relevant feasibility or impact studies. However, some of the included studies mentioned feasibility aspects. For instance, Bai *et al*³² noted an important trade-off of using MRI (longer scanning time, but non-radioactive) compared with CT.

The included studies covered a wide range of imaging techniques to assist the diagnosis of patients suspected of lung cancer. These were mainly on the diagnostic performance of CT, CXR and PET-CT, with a few studies also reporting on MRI, scintigraphy and ultrasound.

There was substantial heterogeneity in types of target cancers, populations included and reported pathways for the diagnosis and management of suspected cancers. In these studies, CXR sensitivity for lung cancer ranged from 33.3%³⁸ to 75.9%,²⁷ with specificity ranging from 83.2%²⁹ to 95.5%.²⁷ For CT, sensitivity varied from 58%³¹ for pleural malignancy to 100%³⁶ for lung cancer (when used in combination with sputum cytology). For MRI, a sensitivity of 91.3% and a specificity of 85.7% for lung cancer were reported.³² Values reported for PET-CT varied for lung cancer sensitivity from 75%³⁵ (using Gd₂O₃-doped carbon-11-choline as contrast substance) to 100%³³ (using ¹⁸F-fluorodeoxy glucose); from the same studies, values for specificity ranged from 22.2%³⁵ to 46.2%.³³ Rinaldi *et al*³⁹ reported a sensitivity of 93.2% and a specificity of 100% for LUS in identifying lung cancer. For scintigraphy, reported sensitivity was 100% and specificity was 69.2%³³ (using ^{99m}Tc-octreotide acetate as contrast).

We identified only one report evaluating the use of an AI system in a primary care context; in this report,²⁹ the particular AI solution used underperformed compared with a multidisciplinary team in the detection of cancer.²⁹ This outcome highlights the broader issue of many AI imaging tools, where datasets used for training and validation do not reflect the prevalence of the general population.

Strengths and limitations

Our review question focused on applicability in the primary care settings. For this reason, we excluded all studies where patients had been explicitly referred to secondary/tertiary care, or where the sample included both suspected and confirmed cancer cases. Due to the very large number of search records obtained, we had to focus the search strategy to a manageable output of

just over 20 000 hits. However, each article was manually screened independently by two reviewers. Consequently, although unlikely, it is possible that relevant studies were missed in this review.

Our analysis focused on symptomatic populations, as symptomatic presentation within primary care remains the predominant route for identifying intrathoracic cancers in health systems with universal healthcare, particularly in the UK.¹ Although there is a growing body of evidence on the performance of various imaging modalities for cancer screening, the fundamental differences between symptomatic and asymptomatic populations mean that this evidence is not directly transferable between the two groups, primarily because of the different performance metrics of the tests in populations with a different incidence of cancer. For this reason, we excluded the more extensive literature on screening studies involving low-risk, asymptomatic populations and confined our analysis to symptomatic detection.

The majority of the studies included cohorts of patients with high suspicion of cancer, suggesting a significantly higher prevalence of intrathoracic cancers compared with that from unselected primary care settings. Due to this, diagnostic performance data may not be generalisable to primary care settings, where many patients may present with mild symptoms.

Many of the included studies had small samples, with just half of the studies including more than 100 participants. Of these, studies by Bhartia *et al*²⁷ and Bradley *et al*²⁸ included a largely overlapping population, derived from the same original cohort. Due to the limited number of patients included across all studies, there is still significant uncertainty as to the performance of these imaging modalities in unselected populations. Furthermore, as the risk of bias assessment highlighted, in a number of studies, it was unclear how patients were identified, whether certain patients were excluded from the studies and whether the way the reference standard was conducted could have biased the study results.

While subjective, the quality of evidence assessed with the GRADE framework indicated that the majority of studies judged as of high quality were focused on CXR, with only one study focused on CT. This again confirms the limited good quality evidence on symptomatic patients in primary care.

We did not conduct a meta-analysis. This was not possible due to the heterogeneity of the included studies; even for studies reporting on the same imaging modality, there were great differences in the diagnostic pathways used, definition of positive cases and populations selected. As mentioned, some of the studies also had overlapping populations.

Relevance to practice

Intrathoracic cancer diagnosis is challenging and not straightforward, with guidelines offering variable recommendations. This review summarised the evidence on imaging performance for intrathoracic cancers

in primary care and highlighted the need for further research in primary care, including future studies with reliable sample sizes and well-defined populations.

For symptomatic patients presenting in primary care, CXR is the most common first imaging investigation. In lesions affecting the thoracic wall (including pleurae), US has also been used, although CT can provide not only information on the areas that cannot be visualised by US, but also better information overall on the presence and nature of intrathoracic lesions than US.

Specifically, the NICE guidelines on lung cancer⁷ offer limited imaging options to primary care practitioners. The guidelines on recognition and referral for lung cancer (NG12) only include recommendations for CXR based on specific symptoms, as well as urgent referral for unexplained haemoptysis or an abnormal CXR. More advanced imaging, such as CT, MRI, PET-CT and scintigraphy, is usually used in secondary care, and rarely, the first imaging technique to be used when intrathoracic cancer is initially suspected in primary care.

Conclusion

There was limited information on the diagnostic performance of usual imaging techniques when used in unselected primary care settings for the diagnosis of intrathoracic cancer in symptomatic patients.

More studies need to focus on the evaluation of such techniques (CT in particular) in the general population presenting in primary care, where the prevalence is relatively low. A better understanding of their performance could inform more efficient detection strategies for intrathoracic cancers in primary care, leading to earlier detection and, ultimately, better outcomes.

X Wasim Hamad @waseem hamad, Natalia Calanzani @NCalanzani and Fiona M Walter @fmw22

Acknowledgements The authors would like to thank Isa Asadi and Gianni Dongo for their contribution to screening and data collection.

Contributors BG assisted in all aspects of the systematic review, including screening, data extraction, critical appraisal, data analysis, drafted and finalised the manuscript. JLP contributed to the drafting of the protocol, screening, data extraction and helped revise the manuscript. WH contributed to the drafting of the protocol, screening, data extraction and helped revise the manuscript. NC contributed to the drafting of the protocol and helped revise the manuscript. LA organised meetings with patient representatives and helped revise the manuscript. FMW contributed to the drafting of the protocol, data analysis and helped revise the manuscript. RN is the guarantor, contributed to the drafting of the protocol, data analysis and helped revise the manuscript. All authors agreed on the final version of the manuscript.

Funding This project is funded by the National Institute for Health and Care Research (NIHR) School for Primary Care Research (project reference 577). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Competing interests FW is member of the BMJ Open Editorial Advisory Board. RN is a co-author of two of the papers included in the review. All authors declare they have no competing interests relating to this study.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

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. All data generated or analysed during this study are included in this published article and its supplementary information files.

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 iDs

Bogdan Grigore <http://orcid.org/0000-0003-4241-7595>

Jaime L Peters <http://orcid.org/0000-0003-1778-3518>

Natalia Calanzani <http://orcid.org/0000-0002-5068-2543>

REFERENCES

- 1 Neal RD, Hamilton W, Rogers TK, *et al*. Easily Missed? *BMJ* 2014;349:g6075.
- 2 Okoli GN, Kostopoulou O, Delaney BC. Is symptom-based diagnosis of lung cancer possible? A systematic review and meta-analysis of symptomatic lung cancer prior to diagnosis for comparison with real-time data from routine general practice. *PLoS One* 2018;13:e0207686.
- 3 National Institute for Health and Care Excellence. Scenario: Referral for suspected lung or pleural cancer. Lung and pleural cancers - recognition and referral, 2021. Available: <https://cks.nice.org.uk/topics/lung-pleural-cancers-recognition-referral/management/referral-for-suspected-lung-or-pleural-cancer>
- 4 National Institute for Health and Care Excellence. Suspected cancer: recognition and referral. NICE guideline [NG12]. 2023.
- 5 Bradley SH, Abraham S, Callister ME, *et al*. Sensitivity of chest X-ray for detecting lung cancer in people presenting with symptoms: a systematic review. *Br J Gen Pract* 2019;69:e827–35.
- 6 ten Haaf K, van Rosmalen J, de Koning HJ. Lung Cancer Detectability by Test, Histology, Stage, and Gender: Estimates from the NLST and the PLCO Trials. *Cancer Epidemiol Biomarkers Prev* 2015;24:154–61.
- 7 National Clinical Guideline, C. Lung cancer: diagnosis and management. NICE clinical guideline NG122: London. 2019.
- 8 Philip B, Jain A, Wojtowicz M, *et al*. Current investigative modalities for detecting and staging lung cancers: a comprehensive summary. *Indian J Thorac Cardiovasc Surg* 2023;39:42–52.
- 9 Snowsill T, Yang H, Griffin E, *et al*. Low-dose computed tomography for lung cancer screening in high-risk populations: a systematic review and economic evaluation. *Health Technol Assess* 2018;22:1–276.
- 10 Birchard KR. Transthoracic needle biopsy. *Semin Intervent Radiol* 2011;28:87–97.
- 11 Sánchez de Cos J, Hernández Hernández J, Jiménez López MF, *et al*. SEPAR Guidelines for Lung Cancer Staging. *Archivos de Bronconeumología (English Edition)* 2011;47:454–65.
- 12 Koyama H, Ohno Y, Seki S, *et al*. Magnetic resonance imaging for lung cancer. *J Thorac Imaging* 2013;28:138–50.
- 13 Bombardieri E, Crippa F, Cataldo I, *et al*. Somatostatin receptor imaging of small cell lung cancer (SCLC) by means of 111In-DTPA octreotide scintigraphy. *Eur J Cancer* 1995;31A:184–8.
- 14 Reid JH, Coche EE, Inoue T, *et al*. Is the lung scan alive and well? Facts and controversies in defining the role of lung scintigraphy for the diagnosis of pulmonary embolism in the era of MDCT. *Eur J Nucl Med Mol Imaging* 2009;36:505–21.
- 15 Kahn D, Menda Y, Kernstine K, *et al*. The utility of 99mTc depreotide compared with F-18 fluorodeoxyglucose positron emission tomography and surgical staging in patients with suspected non-small cell lung cancer. *Chest* 2004;125:494–501.
- 16 Kwekkeboom DJ, Kho GS, Lamberts SW, *et al*. The value of octreotide scintigraphy in patients with lung cancer. *Eur J Nucl Med* 1994;21:1106–13.
- 17 Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ* 2016;353:i3139.
- 18 Walter FM, Thompson MJ, Wellwood I, *et al*. Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. *BMC Cancer* 2019;19:586.
- 19 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
- 20 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.
- 21 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.
- 22 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.
- 23 Sterne JAC, Savović J, Page MJ, *et al*. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.
- 24 Sterne JA, Hernán MA, Reeves BC, *et al*. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
- 25 Guyatt GH, Oxman AD, Vist GE, *et al*. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- 26 Singh S, Chang SM, Matchar DB, *et al*. Chapter 7: grading a body of evidence on diagnostic tests. *J Gen Intern Med* 2012;27 Suppl 1:S47–55.
- 27 Bhartia BSK, Hatton NLF, Aslam R, *et al*. A prospective cohort evaluation of the sensitivity and specificity of the chest X-ray for the detection of lung cancer in symptomatic adults. *Eur J Radiol* 2021;144:S0720-048X(21)00434-4.
- 28 Bradley SH, Hatton NLF, Aslam R, *et al*. Estimating lung cancer risk from chest X-ray and symptoms: a prospective cohort study. *Br J Gen Pract* 2021;71:e280–6.
- 29 Maiter A, Hocking K, Matthews S, *et al*. Evaluating the performance of artificial intelligence software for lung nodule detection on chest radiographs in a retrospective real-world UK population. *BMJ Open* 2023;13:e077348.
- 30 Traill ZC, Davies RJ, Gleeson FV. Thoracic computed tomography in patients with suspected malignant pleural effusions. *Clin Radiol* 2001;56:193–6.
- 31 Tsim S, Stobo DB, Alexander L, *et al*. The diagnostic performance of routinely acquired and reported computed tomography imaging in patients presenting with suspected pleural malignancy. *Lung Cancer (Auckl)* 2017;103:38–43.
- 32 Bai Y, Li D, Duan Q, *et al*. Analysis of high-resolution reconstruction of medical images based on deep convolutional neural networks in lung cancer diagnostics. *Comput Methods Programs Biomed* 2022;217:S0169-2607(21)00666-0.
- 33 Wang F, Wang Z, Yao W, *et al*. Role of 99mTc-octreotide acetate scintigraphy in suspected lung cancer compared with 18F-FDG dual-head coincidence imaging. *J Nucl Med* 2007;48:1442–8.
- 34 Yang H, Zhang Y, Wei D, *et al*. Utility of high-frequency B-mode and contrast-enhanced ultrasound for the differential diagnosis of benign and malignant pleural diseases: a prospective study. *J Thorac Dis* 2022;14:3695–705.
- 35 Zhou T, Hang D, Li Y, *et al*. Role of Gd₂O₃-doped carbon-11-choline-lenvatinib nanoparticles contrast agent PET/CT in the diagnosis of patients with lung cancer. *Oncol Lett* 2020;19:1117–24.
- 36 Felten MK, Knoll L, Schikowsky C, *et al*. Is it useful to combine sputum cytology and low-dose spiral computed tomography for early detection of lung cancer in formerly asbestos-exposed power industry workers? *J Occup Med Toxicol* 2014;9:14.
- 37 Heye T, Ley S, Heussel CP, *et al*. Detection and size of pulmonary lesions: how accurate is MRI? A prospective comparison of CT and MRI. *Acta Radiol* 2012;53:153–60.
- 38 Clin B, Morlais F, Guittet L, *et al*. Performance of chest radiograph and CT scan for lung cancer screening in asbestos-exposed workers. *Occup Environ Med* 2009;66:529–34.
- 39 Rinaldi L, Milione S, Fascione MC, *et al*. Relevance of lung ultrasound in the diagnostic algorithm of respiratory diseases in a real-life setting: A multicentre prospective study. *Respirology* 2020;25:535–42.