

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

Jill Colquitt,^{1,†} Mary Jordan,^{2,†} Rachel Court,²
Emma Loveman,² Janette Parr,² Iman Ghosh,²
Peter Auguste,² Mubarak Patel² and Chris Stinton^{2*}

¹Effective Evidence, Waterlooville, UK

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

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

†Joint lead authors

Published August 2024

DOI: 10.3310/LKRT4721

Scientific summary

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

Health Technology Assessment 2024; Vol. 28: No. 50

DOI: 10.3310/LKRT4721

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

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

This assessment covers the use of AI software as an adjunct to an appropriate radiology specialist to assist in the identification of suspected lung cancer on chest X-rays (CXRs). AI technologies subject to this assessment are standalone software platforms developed with deep-learning algorithms to interpret CXRs. The algorithms are fixed but updated periodically. The AI software automatically interprets radiology images from the CXR to identify abnormalities or suspected abnormalities. The abnormalities detected and the methods of flagging the location and type of abnormalities differ between different AI technologies. For example, a CXR may be flagged as suspected lung cancer when a lung nodule, lung mass or hilar enlargement, or a combination of these, is identified. A technology may classify CXRs into those with and without a nodule, or it may identify several different abnormalities or lung diseases.

Objectives

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

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

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

Methods

Data sources

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

Eligibility criteria

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

Intervention: radiology specialist with adjunct AI

Comparator: radiology specialist without adjunct AI

Outcomes: test accuracy, patient management, clinical effectiveness.

Study selection, data extraction and assessment of risks of bias

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

Data synthesis

A narrative data synthesis was performed.

Modelling

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

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

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

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

Results

Test accuracy, practical implications and clinical effectiveness

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

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

Conclusions

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

Implications for service provision

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

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

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

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

Study registration

This study is registered as PROSPERO CRD42023384164.

Funding

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

ISSN 2046-4924 (Online)

Impact factor: 3.6

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

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

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

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

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

Criteria for inclusion in the *Health Technology Assessment* journal

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

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

HTA programme

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

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

This article

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

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

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

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

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

