BMJ Open Respiratory Research

Employment of artificial intelligence for early lung cancer diagnosis: a retrospective cohort study

Nathan R Hill,[1] Travis Dotson,[2] Sarah E Maus,[2] Christina Bellinger,[2] Vishisht Mehta ,[3] Annika Rings ,[4] Lyndsey C Pickup,[4] Noah Waterfield Price ,[4] Václav Potěšil,[4] David P Carbone ,[5] Jasleen Pannu[6]

To cite: Hill NR, Dotson T, Maus SE, *et al*. Employment of artificial intelligence for early lung cancer diagnosis: a retrospective cohort study. *BMJ Open Respir Res* 2025;**12**:e003225. doi:10.1136/bmjresp-2025-003225

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjresp-2025-003225).

Received 17 February 2025
Accepted 11 October 2025

[1]Bristol-Myers Squibb Co R&D Lawrenceville, Lawrenceville, New Jersey, USA
[2]Pulmonary/Critical Care, Atrium Health Wake Forest Baptist, Winston-Salem, North Carolina, USA
[3]Comprehensive Cancer Centers of Nevada, Henderson, Nevada, USA
[4]Optellum Ltd, Oxford, UK
[5]The Ohio State University Comprehensive Cancer Center, Columbus, Ohio, USA
[6]The Ohio State University Wexner Medical Center, Columbus, Ohio, USA

Correspondence to
Dr Annika Rings;
annika.rings@optellum.com

ABSTRACT

Aim This study aims to evaluate the effectiveness of combined artificial intelligence (AI)-based tools for early patient identification, risk stratification and tracking in increasing the follow-up rate of incidentally detected lung nodules, potentially leading to earlier diagnoses of lung cancer, particularly non-small cell lung cancer (NSCLC).

Patients and methods We conducted a retrospective cohort study involving all patients who underwent CT scans at an academic medical centre over an 8-month period. Real-world practice was compared with modelling of a hypothetical intervention with AI tools. This study was complemented by a multi-reader multi-case analysis to enhance the robustness of our findings.

Results The implementation of AI tools significantly increased the rates of guideline-concordant follow-up for detected nodules, rising from 34% without the tool to 94% with the AI intervention (p<0.0001, McNemar's test). Furthermore, the median time to diagnosis of NSCLC was reduced from 129 days to 25 days (p<0.001, Wilcoxon signed-rank test).

Conclusion These findings provide compelling evidence that AI tools can enhance the follow-up rates for patients with incidentally detected lung nodules and expedite the diagnosis of lung cancer. The integration of AI in clinical practice may significantly improve patient outcomes in lung cancer detection and management.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Previous studies have demonstrated that approximately 40% of incidentally detected pulmonary nodules are followed-up in accordance with established clinical guidelines.

WHAT THIS STUDY ADDS

⇒ This study employs Natural Language Processing to identify lung nodules within radiology reports from a large patient cohort. Additionally, it uses diagnostic and procedural codes to reconstruct patient care pathways. The observed standard of care is then compared against a hypothetical intervention (as this study is retrospective) that integrates combined artificial intelligence (AI) tools for patient identification, tracking and risk stratification. Outcomes associated with the use of AI are modelled based on a multi-reader multi-case study.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings of this study suggest that the incorporation of AI tools into the standard care protocol for managing pulmonary nodules has the potential to enhance adherence to clinical guidelines, thereby reducing the time to lung cancer diagnosis. A prospective study would further validate these findings.

INTRODUCTION

Lung cancer remains the leading cause of cancer mortality,[1 2] with an estimated 124 730 deaths in the USA in 2025 alone and a dismal 28% 5-year survival rate.[3] Early diagnosis and treatment hold great promise in improving outcomes through curative interventions such as surgical resection combined with immunotherapies/chemotherapies. Current statistics suggest that 63% of patients with localised non-small cell lung cancer (NSCLC) who undergo treatment can expect to live at least 5 years after diagnosis, with this number rising to 90% for the smallest Stage IA tumours.[4]

The optimum opportunity to drive early intervention lies among the more than 1.5 million new US patients each year[5] who have a lung nodule detected incidentally on CT scans acquired for other reasons (eg, emergency room scans, cardiac CT scans). While most nodules are benign, a subset are cancerous. Unfortunately, the current clinical management pathway suffers from deficient care coordination and risk stratification, resulting in 6 out of 10 incidental nodule patients lost to follow-up,[6] invasive procedures on benign nodules[7] and delayed lung cancer diagnosis, leading to suboptimal outcomes for patients with cancer.

The primary goal of this study is to evaluate whether the utilisation of AI tools could lead to more early-stage lung cancer diagnoses,

that is, bringing about stage shift by ensuring no patient misses their appropriate nodule follow-up, and by speeding up time to diagnosis. The overarching aim is to improve patient outcomes and help with allocation of healthcare resources through better risk stratification.

METHODS

Study design and patient population

In this retrospective cohort study, we compare current clinical practice at a large academic medical centre (AMC) to modelled results based on a reader study from a hypothetical intervention with a commercially-available artificial intelligence (AI)-enabled, computer-assisted diagnosis (CADx) tool. The AI tool comprises three key modules.

Patient discovery

Automated Natural Language Processing (NLP) of radiology reports, deployed across an entire hospital/health system.[8] The goal is to identify and capture more patients with suspicious nodules and to speed up referral to specialists, especially for high-risk patients.

Clinical decision support

An imaging AI/'Radiomics'-based digital biomarker of malignancy risk is computed from the CT image pixels.[9 10] It has been shown to increase the accuracy of physicians' malignancy assessment, reduce variation between physicians and help physicians make more appropriate clinical decisions.[11]

Management and tracking

Assists physicians in longitudinally tracking decisions and follow-up procedures, and reminds the thoracic care team when a patient misses a prescribed follow-up CT.

All patients who underwent a CT scan at the AMC for reasons other than lung cancer screening between July 2017 and February 2018 were included in this study, provided they met the specified inclusion criteria. The retrospective dataset of radiology reports was processed using the AI tool's NLP-based Patient Discovery module to automatically identify the relevant subset of patients.

Inclusion criteria

- Patients must have a CT report from the AMC within the 8-month study period.
- The patient discovery AI must identify a mention of a clinically relevant solid or semi-solid nodule, measuring between 5 and 30 mm, that was incidentally detected in the radiology report.

Definition of 'clinically relevant'

A nodule is considered clinically relevant if it meets one of the following criteria:

- A nodule measuring 8–30 mm that is not fully calcified and for which the radiology report does not explicitly state that follow-up is unnecessary.
- A nodule measuring 5–7 mm that has an explicit follow-up recommendation in the radiology report.

Exclusion criteria

- Patients with a pre-existing cancer diagnosis at the time of the scan.

Additional exclusion criteria for the reader study

- Patients for whom no outcome regarding nodule malignancy or benignity can be determined within 2 years following the reported scan (eg, due to death or relocation outside the healthcare system).
- Scans that do not meet the technical criteria specified in the US Food and Drug Administration (FDA)-approved Indications for Use (IfU) of the CADx tool.

Study objectives

Primary objective

Nodule patient follow-up: to determine whether the implementation of AI-enabled tools can increase the number of clinically relevant lung nodules that receive appropriate follow-up compared with current clinical practices.

Secondary objectives

Patient with cancer follow-up: to evaluate whether the use of AI-enabled tools would increase the number of patients with lung cancer followed-up in accordance with established clinical guidelines.

Time to diagnosis: to assess the potential reduction in median time to diagnosis for lung cancers when comparing current clinical practices with the employment of the AI-enabled care tool.

Exploratory objectives

Size at diagnosis: to investigate whether the AI-enabled tool would contribute to a reduction in the mean size of primary tumours at the time of diagnosis.

Stage shift: to evaluate whether the AI-enabled tool would facilitate a shift in the (sub) stage of cancers at diagnosis compared with current clinical practices.

Data collection

Clinical data meeting the following criteria were collected

Reports

8 months of CT reports spanning July 2017 to Feb 2018 (inclusive) of all CTs performed at the AMC (65 039 radiology reports).

Clinical metadata

Clinical metadata containing information about patient visits from the beginning of the 8-month study period until 2 years after the end of the study period (July 2017 to March 2020) was collected for all patients meeting the inclusion criteria. It includes (and is not limited to)

specific clinical visits, appointment notes, diagnosis notes, International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes and current procedural terminology (CPT) codes. For patients with a known primary lung cancer, diagnosis dates, histology reports, size and stage at diagnosis were collected.

CT DICOM data

All CT digital imaging and communications in medicine (DICOM) data from the patients whose reported lung nodules were deemed to have a known outcome and to be eligible for the reader study was collected.

To select the cases appropriate for the reader study, we excluded the patients not meeting the FDA-cleared IfU of the CADx score for one of the reasons outlined in the detailed Methods section of the online supplemental material.

All collected data were de-identified by the AMC and identified only by hashed ID values.

Data markup and curation

Processing by patient discovery

The retrospective consecutive dataset of radiology reports was processed by the NLP-based patient discovery AI module to automatically identify the subset of patients meeting the inclusion and exclusion criteria above.

Reports where patient discovery detected a measured nodule in the text are referred to as 'Patient Discovery positive'.

Markup of radiology reports

Clinical data specialists with a medical qualification reviewed the report text of the patient discovery-positive subset to determine clinical relevance of the nodule and follow-up recommendations by the radiologist as detailed in the online supplemental material, and the patient discovery negative subset to identify potential false negatives.

We determined outcomes (cancer/benign) and the follow-up procedures the patients received from ICD-10-CM codes, CPT codes and clinical visit notes as described in the online supplemental material.

Reader study workflow

The reader study was conducted as a multi-reader, multi-case study, wherein physician readers were selected based on their level of experience and area of clinical practice, specifically focusing on pulmonologists. To ensure complete separation between the readers and actual patient care, physicians from the AMC where the data were collected were not invited to participate. The CADx score component of the AI tool has been evaluated separately in a previous reader study.[11] A detailed description of the reader study workflow is provided in the online supplemental material.

Modelling approach

The follow-up each patient would have received with AI is modelled based on the most frequent recommendation from the reader study. For patients not included in the reader study, the closest matching patient was determined based on similarity of the Mayo risk score,[12] which has previously been shown to be predictive of guideline-concordant follow-up.[13] Using an aggregate score summarising multiple covariates has previously been suggested as a similarity measure for matching.[14] Basic patient characteristics were similar between the reader study cohort and the full patient cohort (online supplemental figure S1A–C). Further assumptions, validations and data handling are explained in the detailed methods section of the supplement, which introduces all specific modelling methodology used to derive follow-up decisions and outcomes from the reader study results for the hypothetical AI intervention is outlined in the online supplemental materials.

The time to cancer diagnosis with AI was estimated by assigning average waiting times for the diagnostic tests recommended by the readers in the reader study, using standard waiting times reported by expert clinicians at the AMC. If further follow-up imaging was recommended, it was assumed that after the recommended follow-up CT scan, the measured growth of the nodule would have prompted a biopsy recommendation. In this model, the follow-up was assumed to be carried out as recommended by the pulmonologist without being negatively impacted by resource limitations.

The model to determine the size at diagnosis with AI was based on the estimated time to diagnosis and the volume-doubling time calculated for each nodule individually using an exponential growth model and consecutive CT scans.

The stage at diagnosis with AI was modelled based on the estimated time to diagnosis and stage progression times determined by fitting a stage progression model to experimental data. The relative time cancers spent in each stage was constrained to represent the proportions of cancers in each stage from the literature and the expected proportion of adenocarcinomas,[15] using the modelled time-to-diagnosis as described above. The details of this model are outlined in the online supplemental methods.

Statistical analysis

Statistical analysis was conducted according to the prespecified analysis plan. A brief description is provided here, while the detailed statistical analysis methods and sensitivity analyses (online supplemental tables S1 and S2) are described in the supplemental material.

The study employed a paired design, in which the outcome for each patient was compared between the standard of care and the modelled with-AI scenario.

For comparing proportions, the McNemar test was used. Numerical data were compared using the Wilcoxon

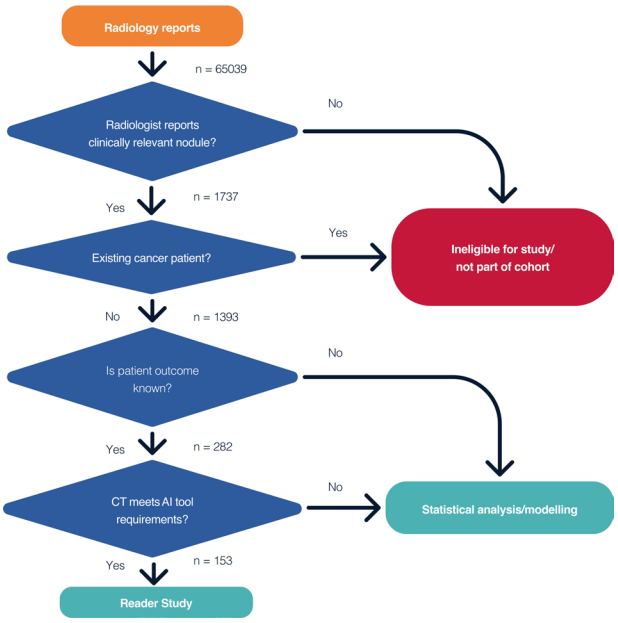

Figure 1 Overall study flowchart. AI, artificial intelligence.

signed-rank test. Time-to-event data was compared using the log-rank test. Analyses involving multiple readers were carried out using bootstrapping.

Patient involvement

This study highlights issues with lung nodule follow-up, which is a key concern for patient safety and suggests improvements to patient care. However, patients were not directly involved in designing this study.

RESULTS

Patient identification using patient discovery software

We obtained all radiology reports from July 2017 to February 2018 (inclusive) for all CT scans performed at the AMC for any indication other than lung cancer screening. A total of 65 039 reports were collected (see study flow diagram in figure 1).

The patient discovery AI was configured to exclude patients with known cancer and those undergoing lung cancer screening. From the remaining eligible reports, the Patient Discovery software identified mentions of a nodule in 4095 cases. Ultimately, 1393 patients met the inclusion criteria (figure 1). The primary reasons for exclusion were an existing cancer diagnosis or the nodule not meeting the 'clinically relevant' definition. Of those included, 153 had CTs suitable for the reader study, all of which were presented to the readers (figure 1). Exclusions from the reader study were due to indeterminate diagnoses or scans not meeting the technical specifications of the CADx tool (figure 1).

Annotation by human data experts on reports identified as negative by Patient Discovery revealed six false negative reports that referenced pulmonary nodules. However, two of these reports pertained to known patients with cancer undergoing follow-up. The remaining reports referenced 'micronodules', 'scattered peribronchocentric ground-glass opacities', 'scattered foci of nodular airspace opacity involving all lobes bilaterally' and 'tiny left lower lobe pulmonary nodules', none of which met the criteria for pulmonary nodules requiring follow-up under the Fleischner guidelines.[16]

Primary objective

Nodule patient follow-up

Under current clinical practice, we determined that 471 out of 1393 (34%) clinically relevant nodule patients received follow-up according to guidelines, while 41 (3%)

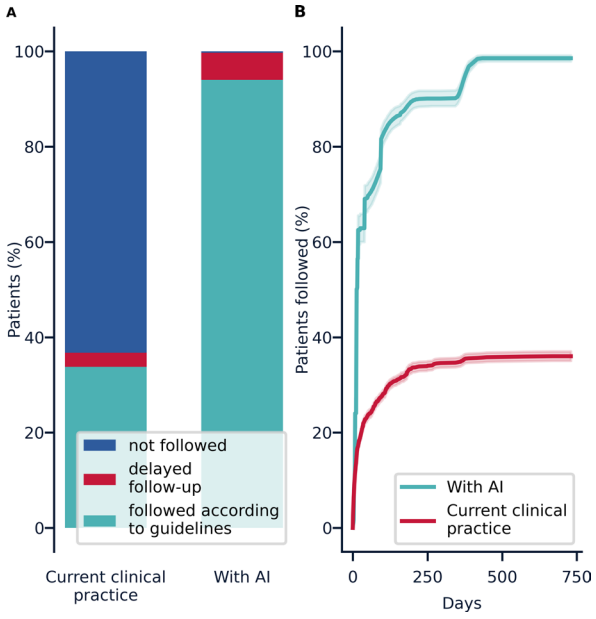

Figure 2 Primary objective. (A) Proportion of clinically relevant nodule patients followed according to guidelines in observed clinical practice versus simulated care with AI. 34% were followed-up according to guidelines in the current clinical practice versus 94% with AI (p<0.0001, bootstrapped McNemar test, 10 000 bootstraps, n=1393 patients). Followed according to guidelines is defined as followed as recommended by Fleischner Society guidelines,[14] or according to radiologist recommendations if different from guidelines. Delayed follow-up is defined as nodule-specific follow-up that occurred with a delay longer than recommended by guidelines. Not followed means the patient had no nodule-related visits at the AMC in the 2-year period after the CT scan date for which data was collected in the study. Note that the modelled 'with AI' scenario represents the number of patients that would have been offered follow-up with AI and does not take into account the possibility of patients refusing treatment or being unable to attend follow-up appointments for other reasons. (B) Time-to-event analysis of time to first follow-up under observed clinical care and with AI tools. Patients were censored at 2 years if no follow-up occurred during this period, as data collection ended after 2 years. Shaded areas represent 95% CIs obtained by bootstrapping. Bootstrapped log-rank test: p<0.0001, n=1393 patients. AI, artificial intelligence; AMC, academic medical centre.

were followed up with delays, and 881 (63%) received no follow-up.

With the implementation of AI, the number of patients who would have been offered guideline-concordant follow-up increased to 1308 (94%), while only 4 (0.3%) were not recommended for follow-up (figure 2A). The median bootstrapped McNemar p value (10 000 bootstraps) when comparing guideline-concordantly followed patients with and without AI was <0.0001.

In a time-to-event analysis comparing the interval from the initial CT to the first follow-up visit, patients were followed-up more rapidly with AI (p<0.0001, log-rank test, figure 2B).

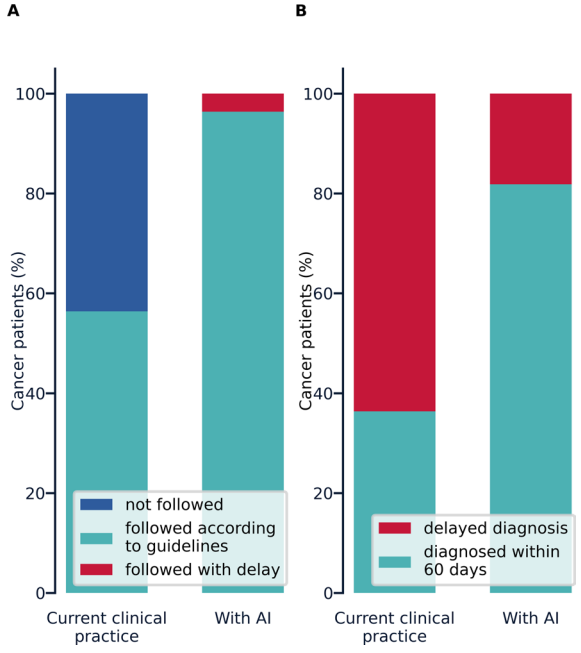

Figure 3 Secondary objective. (A) Proportion of patients with cancer followed according to guidelines in observed clinical care versus with AI tools. 56% of cancers were followed-up according to guidelines in current clinical practice, versus 96% with AI (p<0.0001, McNemar test, n=55 cancers). Followed according to guidelines was defined as receiving the follow-up care recommended by Fleischner Society guidelines or by the radiologist reviewing the initial CT, if different from guidelines. Delayed follow-up was defined as receiving nodule-related follow-up later than recommended by guidelines. Not followed was defined as the patient never had any more nodule-related visits within the 2-year observation period. Note that all 31 patients with cancer with known cancer outcomes were followed-up appropriately by the AMC. The additional 24 patients with cancer are undiagnosed patients assumed to have cancer based on cancer prevalence in the underlying population. (B) Proportion of confirmed patients with cancer diagnosed within 60 days of initial CT. 36% of cancers were diagnosed within 60 days in current clinical practice and 82% with AI (p<0.01, McNemar test, n=22 cancers). Only NSCLC with confirmed histology provided by the AMC is included in this analysis. AI, artificial intelligence; AMC, academic medical centre; NSCLC, non-small cell lung cancer.

To identify differences in patient characteristics among those followed according to guidelines, those followed with delays and those not followed, we analysed the distribution of nodule size, age and sex across the three groups (online supplemental figure S1). Patients with smaller nodules were more likely to be followed with delays or not followed at all (online supplemental figure S1D). This finding was considered when estimating cancer probability in this cohort (see detailed online supplemental methods).

Secondary objectives

Patient with cancer follow-up

For the first secondary objective, we compared the estimated number of patients with cancer lost to follow-up under current clinical practice versus hypothetical care with AI derived from the reader study results.

Under current clinical practice, 31 out of 55 (56%) patients with cancer received follow-up according to guidelines. With AI tools, this number increased to 53 (96%) (p<0.0001, McNemar test, figure 3A). It is important to note that all patients with cancer identified by the AMC were followed according to guidelines in current clinical practice. The additional patients with cancer who did not receive follow-up, as per our evaluation, are undiagnosed patients assumed to have cancer based on population statistics.

As an additional measure of guideline-concordant follow-up for patients with cancer, we compared the proportion of confirmed patients with cancer who received histological confirmation within the recommended period of ≤60 days[17] in current clinical practice versus hypothetical care with AI (figure 3B). In current clinical practice, 36% of confirmed patients with cancer were diagnosed within 60 days, compared with 82% with AI (p<0.01, McNemar test).

Time to diagnosis

For the second secondary objective, the time to diagnosis under current clinical practice was determined from the electronic medical records provided by the AMC. The mean time to diagnosis under current clinical practice was 282 days (SD=361 days), with a median of 129 days (IQR 41–383 days). With AI, the modelled mean time to diagnosis was 53 days (SD=35 days), and the median was 25 days (IQR 25–50 days) (p<0.001, Wilcoxon signed-rank test, figure 4A).

Exploratory objectives

Size at diagnosis

The mean and median size of confirmed NSCLCs at diagnosis under current clinical practice were 22 mm (SD=10.3 mm) and 22 mm (IQR 16.3–24.8 mm) for maximal axial diameter, respectively. The modelled mean and median size at diagnosis with AI were 19 mm (SD=7.6 mm) and 18 mm (IQR 11.8–26.5 mm) for

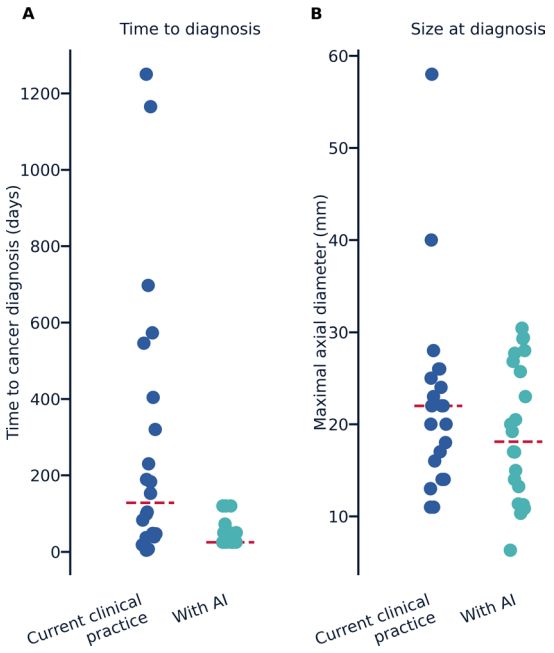

Figure 4 Time to and tumour size at cancer diagnosis. (A) Time from incidental CT to cancer diagnosis for confirmed patients with cancer. The median time-to-diagnosis (dashed line) was 129 days in observed clinical practice and was inferred to be 25 days with AI tools ($p<0.001$, Wilcoxon signed-rank test). 22 histologically confirmed NSCLCs for which the study site provided the diagnosis date are included in this analysis. (B) Size at diagnosis in observed clinical practice and with AI (based on an exponential growth model and measured volume-doubling time) for the same cancers as in (A). The median size at diagnosis (dashed line) was reduced from 22 to 18 mm with AI. AI, artificial intelligence; NSCLC, non-small cell lung cancer.

maximal axial diameter, respectively ($p<0.01$, Wilcoxon signed-rank test, figure 4B).

Stage shift

The stage of primary cancers at diagnosis under current clinical practice was compared with the modelled stage at diagnosis with AI. We limited the analysis to NSCLCs with known stage at diagnosis, resulting in 22 cancers. Details on how the stage shift was modelled are provided in the methods section and online supplemental material. The full set of observed and modelled stages is presented in table 1.

Table 1 Stage distribution of cancers (n=22) at diagnosis in current clinical practice versus with AI

Stage @ diagnosis	Number of cancers	
	Current clinical practice	With AI
I	15	15
II	1	2
III	3	4
IV	3	1

AI, artificial intelligence.

Out of the 22 cancers identified, 7 were diagnosed at a stage higher than Stage I. Notably, two Stage IV cancers would have been diagnosed at a lower stage with the application of AI tools. This indicates that two out of the seven potentially stage-shiftable cancers could have been down-staged. Although the small number of Stage II-IV cancers limits the statistical significance of this finding, it suggests potential for earlier diagnosis.

DISCUSSION

The results of this retrospective data analysis demonstrate that the use of AI tools can significantly impact lung nodule management, increasing the proportion of nodules receiving guideline-concordant follow-up from 34% to 94% with the introduction of AI. Secondary and exploratory objectives reveal earlier and more appropriate follow-up for patients with lung cancer, a reduction in median time to diagnosis and a decrease in tumour size at diagnosis.

The observation that approximately 60% of patients are never followed for their incidental nodules aligns with previous reports[6 18] and underscores the substantial potential for improvement through the implementation of NLP-based AI tools for the automatic identification of at-risk patients. While enhanced follow-up would benefit patients with lung cancer, the increased patient volume may overwhelm healthcare providers. Therefore, it is crucial to combine patient identification with effective risk stratification tools to avoid unnecessary procedures for benign cases.[9] The risk stratification component of this AI tool has been assessed separately in previous studies, showing correct reclassification of benign cases to low-risk, which were classified as intermediate or high-risk in the Mayo model[9] or the Brock model.[10] A previous reader study[11] showed an improvement in readers' accuracy with the AI tool, quantified by an improvement in the receiver-operating characteristic (ROC) curve area under the curve (AUC) score. Despite these benefits, a remaining concern is that the increase in follow-up with AI might increase costs and patient anxiety. Preliminary results quantifying these costs and weighing them against the benefits of life years gained show a favourable incremental cost-effectiveness ratio,[19] and more in-depth studies should now be conducted. Achieving the full benefits of automatic patient identification necessitates the integration of AI into a comprehensive nodule management programme, supported by well-qualified, dedicated nurse navigators.

A limitation of this retrospective study is the potential cohort selection bias, as patients with nodules were identified by the AI's NLP component. We addressed this bias by manual annotations of the reports by clinical data specialists with medical qualifications, and found that there were almost no false negatives. A future study needs to assess the false negative rate and false positive rate of the NLP component systematically.

The study period of July 2017 to February 2018 was chosen to allow for a 2-year follow-up period in all cases, without intersecting with practice changes due to the COVID-19 pandemic. Nodule programmes, guidelines and even scanner technologies have evolved in the meantime, so future studies should focus on current real-world implementation of these AI technologies.

The limited data available at the AMC regarding patients who were never followed prevented us from determining the reasons for their lack of follow-up. It is possible that some patients were followed elsewhere, died from unrelated causes, were deemed untreatable or refused treatment for personal or financial reasons. In such cases, the AI tool would not have improved follow-up. Additionally, some AMC records may have been incomplete, despite our best efforts to obtain all available data. Our sensitivity analysis (see online supplemental file 1) shows the effect of different levels of loss-to-follow-up for reasons outside the scope of the AI tool. While this reduces the effect in the primary objective, the effect remains significant even in hypothetical scenarios in which a large proportion of patients refuses follow-up.

In this retrospective study, we could only model clinician recommendations based on reader study findings, without evaluating the likelihood of patients adhering to those recommendations. The use of AI might influence communication between pulmonologists and patients, potentially affecting patients' decisions to pursue treatment or attend follow-up visits—an effect we could not assess. Consequently, our results regarding follow-up with AI represent an ideal scenario in which all patients offered follow-up actually received it. This limitation highlights the need for validation in a before-and-after study within a real clinical setting.

A more critical question is the rate at which the AI tool facilitates faster follow-up for patients with undiagnosed malignancies, potentially shifting late-stage cancer diagnoses to earlier stages where curative treatment options, such as neoadjuvant therapy and surgery for Stage II cancers, are appropriate, rather than resorting to palliative care for Stage IV disease. Our data suggest a need for additional studies to determine whether employing AI would indeed result in a stage shift. However, a limitation of this study is that it included only patients with lung nodules up to 30 mm, the majority of which are Stage I. The full potential impact can only be assessed when larger lung masses are included.

We observed that for some patients, time to diagnosis was exceptionally rapid in observed clinical practice (within 7 days). Our modelling, which assumed a minimum time of 25 days to a biopsy, did not account for such rapid diagnosis times. However, it is unlikely that the introduction of AI tools would slow down the diagnostic process for patients who were immediately recommended for biopsy with the AI tool. The slower diagnosis observed in this small subset of patients with AI should therefore be regarded as a modelling artefact.

The exploratory objective of stage shift could not be directly addressed in this retrospective study, relying instead on modelling assumptions. The three mechanisms through which we expect AI tools to facilitate stage shift are:

Patient identification

Identifying patients whose nodules require management.

Patient tracking

Ensuring patients are not lost to follow-up.

Patient stratification

Ensuring patients receive the most appropriate types of follow-up on an appropriate schedule.

For this exploratory objective, we could only consider cancers for which outcomes were known, that is, patients followed-up in current clinical practice. This restricts our analysis to only one of the three mechanisms (patient stratification). The modelled stage shift results from the readers recommending earlier and more invasive procedures for cancers in the reader study compared with what the patients received in real clinical practice. The model excludes the potential effect of identifying and following patients with cancer earlier with the NLP tool, thereby underrepresenting the full stage-shift potential of AI tools, which could only be fully assessed through real-world implementations and comparisons in a before-and-after study.

The retrospective design of the study enabled us to assess the effect of the AI tool without altering patient management or introducing risk to patients. However, this design also limited our ability to establish a definitive diagnosis for patients who were lost to follow-up. We estimated that approximately 2.7% of these patients were patients with lung cancer. Consequently, our analyses on time to diagnosis, size at diagnosis and stage shift do not include these patients, leading to a likely underestimation of the AI's effect.

Our modelling outcomes depend on model parameters, which carry inherent uncertainty. Our sensitivity analysis (see online supplemental material) indicates that the outcomes of the study are robust to variations in these parameters within a reasonable range.

This study was conducted at an AMC, where highly-trained experts typically make appropriate decisions for most patients. The greatest unmet need in lung cancer diagnosis exists in regional medical centres and community settings,[20] where CT scans are interpreted by generalists rather than subspecialty-trained thoracic radiology experts and treatment decisions are made by pulmonologists and cardiothoracic surgeons without lung cancer specialisation. AI tools have been shown to reduce variability among readers[11] and potentially may be particularly valuable in settings where physicians are less specialised.

The unique value proposition of this combined AI tool is to identify patients that would otherwise have been missed by flagging them with the NLP tool and bringing them to the attention of the pulmonologist, helping pulmonologists assess the patients' risk of malignancy more accurately than existing risk models[9 10] or the clinician alone,[11] and warning the pulmonologist if patients have missed their appointments. Our model suggests that the combination of these tools could lead to earlier cancer diagnoses. Real-world studies should be conducted to confirm these theoretical findings.

In summary, our results suggest that the introduction of combined AI tools for patient identification, risk stratification and tracking could significantly increase the number of clinically relevant lung nodules followed and lead to earlier diagnoses of NSCLC. This, in turn, would enable better curative treatment options and improve patient survival.

Acknowledgements This work was presented at the SITC conference 2025.[21]

Contributors NRH, AR, LCP, NWP and VP conceived and designed the study. TD, SEM, CB and VM acquired data. AR, LCP, NWP and VP analysed data. All authors were involved in the interpretation of the data and results. AR and LCP drafted the original manuscript. All authors reviewed and critically revised the manuscript. All authors approved the final version of the manuscript to be published. AR is the guarantor.

Funding The data collection and analysis was supported by funding from Bristol Myers Squibb. Trained research staff from Bristol Myers Squibb were involved in the study design and in reviewing the manuscript. Bristol Myers Squibb was not involved in the collection or analysis of the data.

Competing interests NRH is an employee of Bristol Myers Squibb and holds shares in the company. AR, LCP, NWP and VP are employees of Optellum Ltd. and LCP and VP hold shares in the company. DPC consults for AbbVie, Amgen, Arcus Biosciences, AstraZeneca, BMS, Genentech, Glaxo-Smith Kline, Iovance Biotherapeutics, Johnson & Johnson, Lilly, Mirati, MSD, Novartis, Novocure, OncoHost, Pfizer, Regeneron and Roche. All other authors have no competing interests to declare.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the institutional review board of Wake Forest School of Medicine (IRB00070519). A waiver of informed consent was granted by the institutional review board of Wake Forest School of Medicine (IRB00070519).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. In accordance with the Wake Forest Institutional Review Board requirements, the research data may not be publicly shared. Reasonable data requests should be addressed directly to Wake Forest School of Medicine and may be subject to IRB approval procedures.

ORCID iDs
Vishisht Mehta https://orcid.org/0000-0002-2154-802X
Annika Rings https://orcid.org/0000-0002-4065-5330
Noah Waterfield Price https://orcid.org/0000-0002-7097-4383
David P Carbone https://orcid.org/0000-0003-3002-1921

REFERENCES

1 Ferlay J, Colombet M, Soerjomataram I, *et al*. Cancer statistics for the year 2020: An overview. *Intl Journal of Cancer* 2021;149:778–89.
2 Sung H, Ferlay J, Siegel RL, *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J Clinicians* 2021;71:209–49.
3 Surveillance, epidemiology, and end results (seer) program (www.seer.cancer.gov). seer*stat database. Available: https://seer.cancer.gov/statfacts/html/lungb.html [Accessed 4 Nov 2025].
4 Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 2010;17:1471–4.
5 Gould MK, Tang T, Liu I-LA, *et al*. Recent Trends in the Identification of Incidental Pulmonary Nodules. *Am J Respir Crit Care Med* 2015;192:1208–14.
6 Pannu JK, Venious G, Gallagher R, *et al*. P1.03-01 Do We Follow Incidental Lung Nodules Appropriately? A Retrospective Study. *J Thorac Oncol* 2022;17:S100.
7 Mazzone PJ, White CS, Kazerooni EA, *et al*. Proposed Quality Metrics for Lung Cancer Screening Programs. *Chest* 2021;160:368–78.
8 Dotson TL, Gasimova A, Watkins J, *et al*. Identifying Patients With Pulmonary Nodules From CT Radiology Reports Using Natural Language Processing (NLP). D31. IMAGING, BIOMARKERS, AND MACHINE LEARNING IN LUNG CANCER. *American Thoracic Society* 2023.:A6516.
9 Massion PP, Antic S, Ather S, *et al*. Assessing the Accuracy of a Deep Learning Method to Risk Stratify Indeterminate Pulmonary Nodules. *Am J Respir Crit Care Med* 2020;202:241–9.
10 Baldwin DR, Gustafson J, Pickup L, *et al*. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax* 2020;75:306–12.
11 Kim RY, Oke JL, Pickup LC, *et al*. Artificial Intelligence Tool for Assessment of Indeterminate Pulmonary Nodules Detected with CT. *Radiology* 2022;304:683–91.
12 Swensen SJ, Silverstein MD, Ilstrup DM, *et al*. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997;157:849–55.
13 Pannu JK, Venious GJ, Gallagher R, *et al*. Incidental Pulmonary Nodule Detection and Factors Influencing Follow-Up: A Retrospective Study in an Academic Medical Center. *CHEST Pulmonary* 2025;100187.
14 Greifer N, Stuart EA. Matching Methods for Confounder Adjustment: An Addition to the Epidemiologist's Toolbox. *Epidemiol Rev* 2021;43:118–29.
15 Vachani A, Carroll NM, Simoff MJ, *et al*. Stage Migration and Lung Cancer Incidence After Initiation of Low-Dose Computed Tomography Screening. *J Thorac Oncol* 2022;17:1355–64.
16 MacMahon H, Naidich DP, Goo JM, *et al*. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* 2017;284:228–43.
17 RAND Corporation. Quality of Care for Oncologic Conditions and HIV: A Review of the Literature and Quality Indicators. Santa Monica, CA, 2000.
18 Pyenson BS, Bazell CM, Bellanich MJ, *et al*. No Apparent Workup for most new Indeterminate Pulmonary Nodules in US Commercially-Insured Patients. *J Health Econ Outcomes Res* 2019;6:118–29.
19 Godfrey CM, Leech A, Zhu J, *et al*. Artificial Intelligence-assisted Versus Clinician Only Evaluation of Indeterminate Pulmonary Nodules: A Comparative Effectiveness Study. D14. CUTTING EDGE CONCEPTS IN BEHAVIORAL SCIENCES AND HEALTH SERVICES RESEARCH. *American Thoracic Society* 2023.:A6211.
20 Petrelli NJ. A community cancer center program: getting to the next level. *J Am Coll Surg* 2010;210:261–70.