Technoleg Iechyd Cymru
Health Technology Wales

Topic Exploration Report [1]

Topic explorations are designed to provide a high-level briefing on new topics submitted for consideration by Health Technology Wales. The main objectives of this report are to:

- Determine the quantity of evidence available for a technology of interest.
- Identify any gaps in the evidence.
- Inform decisions on topics that warrant fuller assessment by Health Technology Wales (HTW).

Topic exploration report number	TER437
Topic	Clinical decision support software with integrated AI for the management of patients with incidental pulmonary nodules identified on chest CT scans.
Summary of findings	Lung nodules can often be found incidentally on CT scans carried out for reasons not related to suspected lung cancer and may be indicative of early lung cancer. Artificial intelligence (AI) tools that may aid diagnosis of lung cancer by detecting and measuring nodules from imaging are available. Optellum Virtual Node Clinic (VNC) is a clinical management system for patients with already identified incidental nodules. The system i) automatically analyses radiology reports to identify patients where incidental nodules have been found, ii) adds patients to a clinical management dashboard where images can be automatically uploaded and patients tracked longitudinally, and iii) incorporates lung cancer prediction (LCP) AI software to assess identified nodules for risk of malignancy to inform further clinical decision making. HTW did not identify any studies evaluating Optellum as a whole, i.e. including all components, or any similar products. One systematic review, four primary studies and one ongoing study evaluated the LCP component of Optellum only, and three additional primary studies evaluated similar AI software. There were no published cost effectiveness analyses of Optellum or its LCP component, but three publications from ongoing or completed studies are expected. Overall, the AI tools appeared to have good sensitivity and specificity in diagnosing lung cancer, and were an improvement to radiologist alone calculation of the Brock or Mayo scores, which may reduce the number of unnecessary scans undertaken by patients and help prompt rapid intervention in early lung cancer.

[1] Cyfieithu dogfennau HTW wedi'u cyhoeddi o'r Saesneg i'r Gymraeg
Translation of published technical HTW documents from English into Welsh

Introduction and aims

Lung cancer is the third most common cancer in the UK, and accounted for 13% of all cancer cases, with 2,513 people diagnosed with lung cancer in Wales in 2016-18. Lung nodules may be an indication of lung cancer and, as such, need rapid additional review to inform further investigations and treatment. Lung nodules are incidentally detected on around 30% of chest CT (computerised tomography) scans which are carried out for reasons other than suspected lung cancer, such as in routine care and emergency or trauma settings. They can also be identified during lung screening of high-risk populations (which has been recommended across the UK by the National Screening Committee and is currently being trialled in Wales). Guidelines in the UK recommend further investigation of nodules which are more than 5mm at their maximum diameter. In the UK, these are usually assessed via the 'Brock model' which is a method of predicting a nodules risk of malignancy. Other calculation models, such as the Mayo model, may also be used in other countries or for certain populations. Over 90% of incidentally detected nodules are found to be benign, but up to 60% of the patients with incidental nodules are not followed up appropriately.

Digital tools, including Artificial intelligence (AI), have been proposed to aid the assessment of nodules. These can be used at two points of the patient pathway. Firstly, computer/AI-assisted diagnostic technologies can help identify nodules, measure their size, and classify whether they appear to be benign or malignant based on the initial scan, prior to radiologist review. These aim to help the reviewing clinician focus on high-risk nodules, and rapidly assess whether or not further investigation may be needed. Technologies performing this activity are not being reviewed within this report and are covered by NICE Diagnostics Guidance 55 (published July 2023). Secondly, digital tools can be incorporated into the clinical management pathway, following initial radiologist review (with or without AI assistance) of the scan and identification and measurement of nodules. It is this latter group that this report focuses on.

Optellum's virtual node clinic (VNC) is an AI-assisted clinical decision support software for lung cancer diagnosis and management. The VNC software includes three components: i) natural language processing (NLP) AI software that scans all radiology reports to identify patients who have reported and measured pulmonary nodules; ii) patient tracking and management VNC dashboard; and iii) LCP convolutional neural network (CNN) AI software that risk stratifies identified nodules to support clinical decision making - this is done in place of the manually calculated Brock model. Nodules can be tracked longitudinally to assess changes in risk over time, overcoming potential issues around lack of identification and follow-up of changing or at-risk nodules. Methods which can better predict the risk of nodules being malignant or benign may help improve consistency of scan review and reduce the staff time taken for manual calculation. If nodules can be identified as likely benign, patients can avoid undergoing further tests and investigations, which saves the patients both time and stress, and reduces spend within the NHS. If nodules are identified as high risk of malignancy, this could result in earlier diagnosis and treatment of very early-stage lung cancer, potentially improving patient outcomes.

Health Technology Wales researchers searched for evidence on the clinical and cost effectiveness of AI assisted clinical management systems for patients with incidental pulmonary nodules identified on chest CT scans.

Evidence overview

HTW did not identify any published evidence evaluating the entirety of the Optellum VNC technology in clinical practice, or any other technologies with similar functionality (i.e. containing a clinical management system that scan reports for incidental pulmonary nodules and adds them to a monitoring system). The topic proposer (TP) submitted some confidential data around the usability of the VNC in 10 healthcare professionals which concluded that the technology was usable, but this has not been included in this report.

HTW found one pre-print systematic review, seven primary studies and one ongoing study, related to the LCP AI component and technologies similar to this, and one conference abstract related to the NLP AI component. There were no Health Technology Assessments identified.

Secondary evidence

HTW identified one pre-print systematic review of AI-assisted diagnostic models compared to other malignancy risk prediction methods, including physician judgement and the Brock and Mayo models (Wulaningsih et al, 2023). They included 20 studies across North America, Europe and Asia. Three of the included studies assessed Optellum's LCP in the USA and UK, with others reviewing a mix of other CNN and non-CNN AI models. Overall, the review included studies comprising of 7,664 participants and 10,128 nodules, 2126 of which were confirmed malignant via histopathological ground truth. The AI models were found to be more sensitive than both physician judgement alone (by 15.8%), and clinical risk models alone (by 35.4%). Pooled specificity of AI was similar to physician judgement alone (0.77 v 0.80) but were 5.5% more specific than clinical risk models such as Brock/Mayo alone. The relative AUCs for the AI were superior when compared to both physician judgement or clinical risk models, at 1.06 and 1.22 respectively. To note, this systematic review included a study by Baldwin et al (2020) that formed part of the IDEAL study, which validated Optellum's LCP retrospectively in a UK data set from six trusts, prior to a prospective cohort study (NCT03753724) which is outlined below in ongoing studies.

Primary evidence

HTW identified four primary studies assessing Optellum's LCP, and three similar AI technologies; Lung-RADS (Nair et al, 2019), Nodule-X (Causey et al, 2018) and PanCan (van Riel et al, 2017). We were also provided with an abstract from the TP, summarising evidence for Optellum's NLP (Dotson et al, 2023).

The Lung-RADS model assessed 6,956 patients with nodules identified as part of the low-risk arm of the National Lung Cancer Screening trial in the US. Patients are grouped into risk categories similar to the Brock model. The AI showed similar sensitivity and specificity to the Brock model, with area under the curve of 0.82 (0.81 to 0.85) for LCP and 0.83 (0.80 to 0.85) for the Brock model. The authors also showed moderate agreement between LCP and the Brock model for classification of patients, with a weighted kappa 0.60 ($p < 0.001$). Causey et al (2018) trained and validated multiple models of Nodule X using over 1000 nodules in images from the The Lung Image Database Consortium and Image Database resource initiative (LIDC, IDRI). Nodule X works by taking a marked point as a region of interest and generating a prediction for the classification of the marked nodule. There were several models tested, achieving a range of AUC varying from 0.68 to 0.99. PanCan was assessed using 300 scans from the Danish Lung Cancer Screening Trial. The AI created a probability score of malignancy which was compared to those calculated by experts. PanCan was found to have similar performance to humans with an AUC of 0.932 versus 0.910 ($p = 0.184$), but humans were superior when it came to differentiating between size-matched malignant and benign nodules (AUC 0.819 vs 0.706, $p < 0.001$).

Evidence overview
One conference abstract was submitted by the TP regarding the VNC NLP AI element, and the TP notes this is due to be published soon. Dotson et al (2023) compared the AI's ability to flag radiology reports which had been annotated by independent reviewers. 813 reports were evaluated across a variety of settings including emergency, outpatient and inpatient, 275 of which had at least one nodule. The NLP algorithm achieved sensitivity of 1.00, specificity of 0.98, positive predictive value of 0.96 and negative predictive value of 1.00 across all reports. These values remained high for all care settings and CT scan types – with and without contrast, and chest with or without pelvis. Of the published studies evaluating Optellum's LCP specifically, one study compared Optellum's LCP to clinical judgements (Heuvelmans et al, 2021) and, one to the Brock model (Chetan et al, 2022). One study reviewed how Optellum's LCP would have affected further clinical investigations (Tsasok et al, 2021), with the final study evaluating how it could be used to assess longitudinal changes to nodules (Paez et al, 2023). Heuvelmans et al (2021) validated the LCP in data from the UK, Germany and the Netherlands. 2106 incidentally identified nodules measuring 5-15mm in 1650 patients recruited into the Early Lung Cancer Diagnosis Using Artificial Intelligence and Big Data (LUCINDA) study. The nodules were located, contoured and labelled by clinicians who had access to the ground truth diagnosis as decided based on the study protocol, which referred to the Fleisher or British Thoracic Surgery guidelines. The overall AUC was 94.5 %, with a high sensitivity of 99%. In 22.1% of the nodules, malignancy could be ruled out, meaning 18.5% patients could avoid follow-up scans. Chetan et al (2022) used data from the National Lung Screening Trial in the USA, including 10,495 nodules across 4,660 patients to compare the accuracy of the Brock model (using both manual and automated measurement of nodules by other AI software) and the LCP. The LCP obtained an AUC of 0.936 compared to the Brock model's AUC of 0.873 for manual measurements, 0.883 for automated axial measurements and 0.896 for automated spherical diameter measurements. However, the authors found that ablating the nodule and parenchyma texture meant the LCP accuracy dropped slightly to an AUC of 0.915, and ablating the nodule leaving parenchyma only results in a large drop in LCP accuracy, to 0.717. The authors concluded that the LCP AI tool was equal to, or better than using the Brock model, even when other automated software is used to provide nodule measurements. They also note that nodule size and morphology played a key role in the LCP model predictions - in a similar way to the importance of such information in the Brock model calculations. Tsasok et al (2021) aimed to determine how using the LCP software would have influenced further investigational management, as compared to standard practice. They did this by performing an observational retrospective study on 190 nodules of 5-15mm identified on baseline scans as part of routine practice at Oxford University Hospital NHS Foundation Trust. LCP thresholds based on percentage score were determined as follows: < 0.56% meant very low risk with no follow-up required, from 0.56% to 5% meant low risk requiring one follow-up, and > 80% meant high risk and required expedited intervention. No action was required for nodules scoring between 5% and 80%. 158 benign nodules scored between 0.1% and 70.8% using LCP software, with a median of 5.5%. 32 malignant nodules scored between 10.1% and 98.7%, with a median of 59%. They found that all 21 CT scans conducted in the (LCP determined) very low risk group and 24 out of 61 CT scans in the low-risk group could have been avoided, resulting in a 18.6% reduction in the number of CT scans. Additionally, they noted that there could be a 3.6-month reduction in time delay for 5 patients who were in the high-risk group. Finally, Paez et al (2023) assessed subjects with incidentally or 'screening detected' nodules measuring 6-30mm in diameter, who had at least 3 CT scans. The study was performed in the USA. The authors evaluated 48 benign and 32 malignant nodules and found that LCP scores

Evidence overview

for malignant nodules increased over time whereas benign nodule scores remained stable, with a difference in trend of 0.111, $p < 0.001$.

Cost effectiveness

No published cost effectiveness evaluations were identified of these technologies. However, there are plans to perform a cost-effectiveness analysis using a decision tree model on the LUCINDA study used in Heuvelmans et al (2021) and as part of the ongoing studies noted below.

Ongoing studies

DOLCE (NCT05389774) is running across the UK and is due to complete in August 2024 with a recruitment target of 2000 across 12 trusts (four currently recruiting). It aims to determine whether Optellum's LCP generates clinical and health-economic benefits over the current standard of care. Outcome measures include effect on time of discharge, overall clinical management, patient outcomes, health economics including costs of healthcare and health related utilities, and effect on adherence to clinical guidelines.

The prospective cohort component of the IDEAL (NCT03753724) study (initial validation component reported by Baldwin et al, 2020) was based in the UK. This focused on the ongoing data collection and evaluation of the 1293 patients recruited who had nodules 5-15mm, it was completed in June 2022 but there are no published results available as yet. The nodules were classified as benign, intermediate or potentially malignant based on initial scan, and the Optellum LCP AI tool was used to review the nodules. The study aims to assess the overall diagnostic performance and health economic benefits of the tool.

Technology classification

Optellum's VNC, including LCP, and other similar AI tools, are digital health technologies and were determined to be a Tier C technology according to the Evidence Standards Framework for Digital Health Technologies. Technologies within this classification provide information that will be used to aid treatment or diagnosis, to triage or identify early signs of a disease or condition or will be used to guide next diagnostics or next treatment interventions. For technologies of this classification, it is recommended that satisfactory evidence for clinical and cost effectiveness and clinical utility is produced in a relevant setting. This includes studies conducted in a setting like the UK health and care system, peer-reviewed studies, prospective studies, and may also include real-world evidence.

Areas of uncertainty

- There was limited published evidence identified about the training, validation and accuracy of the natural language processing system to identify patients.
- Only one study clearly assessed the possible impact of Optellum's LCP software on clinical pathways the patients would likely have followed if it were used in practice, including effects on CT scan avoidance for benign nodules and reducing time-delays for malignant nodules. Further research on this may be useful, especially for carrying out cost- effectiveness analysis.
- No identified published literature reviewed the effect of the LCP or similar AI tools on patient outcomes such as disease grade at diagnosis and overall survival.
- There is no published cost-effectiveness evidence currently, however there is some which is likely to be published in the next 12 months.

Areas of uncertainty
• There was no published evidence around the usability and acceptability of the tools from the healthcare professionals point of view, although some confidential unpublished data from the TP concluded that the VNC software is usable by healthcare professionals. • AI tools are available for both detection and measurement of nodules in the first instance and for use later in the patient pathway (as in Optellum's LCP). Further evaluation of the combination of AI technology at both stages may be useful, as would comparing the technologies to identify which part of the patient pathway AI tools may be most useful and cost effective. • Several of the identified studies were funded or sponsored by Optellum.

Literature search results

Health technology assessments and guidance
No evidence found
Evidence reviews and economic evaluations
Wulaningsih W, Akram A, et al. (pre-print, 2023). Deep learning-based computer-aided diagnostic models *versus* other methods for predicting malignancy risk in CT-detected pulmonary nodules https://doi.org/10.1101/2023.06.06.23291012
Individual studies
Baldwin DR, Gustafson J, Pickup L, et al. (2020). External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. Thorax. 75:306-312. http://dx.doi.org/10.1136/thoraxjnl-2019-214104 Chetan, MR, Dowson, N, et al. (2022). Developing an understanding of artificial intelligence lung nodule risk prediction using insights from the Brock model. European Radiology 32, 5330–5338 https://doi.org/10.1007/s00330-022-08635-4 Heuvelmans MR, van Ooijen PMA, et al. (2021). Lung cancer prediction by Deep Learning to identify benign lung nodules. Lung Cancer. 154: 1-4. https://doi.org/10.1016/j.lungcan.2021.01.027 Paez R, Kammer MN, et al. (2023). Longitudinal lung cancer prediction convolutional neural network model improves the classification of indeterminate pulmonary nodules. Sci Rep. 13(1):6157. https://doi.org/10.1038/s41598-023-33098-y Tsakok MT, Mashar M, et al. (2021). The utility of a convolutional neural network (CNN) model score for cancer risk in indeterminate small solid pulmonary nodules, compared to clinical practice according to British Thoracic Society guidelines. European Journal of Radiology. 137. https://doi.org/10.1016/j.ejrad.2021.109553
Ongoing research
Nottingham University Hospital NHS Trust. DOLCE: Determining the Impact of Optellum's Lung Cancer Prediction (LCP) Artificial Intelligence Solution on Service Utilisation, Health Economics and Patient Outcomes (NCT05389774). Available at: https://classic.clinicaltrials.gov/ct2/show/NCT05389774 [accessed 04/10/2023] Gleeson F. IDEAL: Artificial Intelligence and Big Data for Early Lung Cancer Diagnosis Study (NCT03753724). Available at: https://classic.clinicaltrials.gov/ct2/show/NCT03753724 [accessed 04/10/2023]
Other AI technologies identified via Cochrane and Google Scholar
Causey JL., Zhang J, et al. (2018). Highly accurate model for prediction of lung nodule malignancy with CT scans. *Sci Rep* **8**, 9286. https://doi.org/10.1038/s41598-018-27569-w Nair VS, Sundaram V, Gould MK (2019). Comparing the performance of lung-RADS with the brock model to identify cancer risk in the NLST. American journal of respiratory and critical care medicine. 199(9). Available: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-02076556/full?highlightAbstract=brock%7Cmodel [Accessed 29/09/2023] van Riel SJ, Ciompi F, et al. (2017). Malignancy risk estimation of pulmonary nodules in screening CTs: Comparison between a computer model and human observers. PLoS One. 12(11):e0185032. https://doi.org/10.1371/journal.pone.0185032
Provided by TP

Dotson TL, et al. (2023) Identifying Patients With Pulmonary Nodules From CT Radiology Reports Using Natural Language Processing (NLP). American Journal of Respiratory and Critical Care Medicine. 207:A6516. https://doi.org/10.1164/ajrccm-conference.2023.207.1_MeetingAbstracts.A6516

Date of search	26/09/2023
Concepts used	General: Artificial intelligence; Brock model; CT; Lung or pulmonary nodule or node; Malignancy score or probability or classification or predict; Virtual or remote; Virtual node/ nodule clinic. Technology specific: Optellum; LCP; LCP-convolutional neural network (CNN).

Proposed research question and evidence selection criteria (if selected)

Proposed Research question	The clinical and cost effectiveness of using an AI assisted clinical management system for patients with incidental pulmonary nodules identified on chest CT scans.

	Inclusion criteria	Exclusion criteria
Population	Adults who have received a CT scan of their chest for any reason where nodules of >5-30mm are incidentally detected. This can be via lung screening, or incidental findings on CT scans performed for other reasons	Nodules identified following CT scan for suspected lung cancer
Intervention	AI assisted clinical management system for lung cancer, including functionality to identify chest CT scan reports with incidental pulmonary nodules, manage and track patients and assess malignancy risk of the nodules using AI	Review of PET scans or x-rays AI systems used earlier in the patient pathway to perform identification, measurement and classification of nodules on chest CT scans prior to initial radiologist review.
Comparison/ Comparators	Standard clinical practice, e.g. Brock or Mayo model malignancy risk calculation, clinician review and filing of reports with incidental nodules	
Outcome measures	Sensitivity Specificity PPV NPV AUC Time to diagnosis Number of additional tests carried out/ avoided Time to treatment Overall survival Cost effectiveness System acceptability and usability Health related Quality of life Economic outcomes	
Study design	Any, with priority given to systematic reviews, RCTs, diagnostic test accuracy studies/ validation studies	

Proposed specialities	Cancer; Respiratory System