NICE National Institute for Health and Care Excellence

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Section A: Comments on the Diagnostics Assessment Report

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Aidence BV	1	56	1.5	Figure 5 - use case 2 refers to growth assessment/VDT whereas use case 1 does not. It is worth noting that, in use case 1, where a technology is capable of automatic analysis of prior imaging, growth assessment/VDT may still be relevant in cases where a nodule may have been present in an earlier study but not detected or reported. This might be because the scan date predates the introduction of the AI software or it was small enough at that time to warrant no further action. As such, we would recommend inclusion of growth assessment/VDT in use case 1.	Thank you for the information. We did not identify any published evidence related to this function of AI software, and therefore this has no impact on the current assessment but may be useful for consideration in future assessment.
Aidence BV	2	60	1.8	Last paragraph - reference is made here to the 'ideal' availability of 'end-to-end' studies. Whilst we don't disagree that such studies would offer the highest possible standard of evidence, it might be prudent to also mention some of the practical constraints which could limit the execution of such studies. For example, to follow through de novo cases to resultant clinical endpoints (confirmed diagnosis and/or survival) would take at least several years prospectively. The CASCADE study in France is a good example of such a longitudinal approach. Revel M, Abdoul H, Chassagnon G, et al. Lung CAncer SCreening in French women using low-dose CT and Artificial intelligence for DEtection: the CASCADE study protocol. BMJ Open 2022;12:e067263. doi.org/10.1136/bmjopen-2022-067263 In a real-world setting, during this time, the technology itself (algorithms and associated DHT components) and other lung cancer interventions (use of complementary biomarkers, multiomics, new treatments, etc) will have evolved, potentially rendering the results obsolete. It would be extremely difficult in practice to hold all other	Thank you for raising this point and for drawing our attention to this study protocol.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				factors constant (or to adjust post-hoc for such factors statistically) over that period of time.	
Aidence BV	3	200	6.3.1	The value used for lung nodule prevalence used in the 'simple' health economic model is stated as 20.6% for the screening population (based on the UK LSUT results - reference #75). However, the value used in the 'full' model on page 217 (table 40, section 7.4.2) uses a value of 50.9% for the same population (based on the UKLS results - reference #78). Even after excluding Cat 2 nodules (3mm-4.9mm dia.) from the UKLS nodule findings, the prevalence is 26.9% (472+64 / 1994) which is still higher than the UK LSUT results. It is not clear to us why different prevalence values (for the screening population) have been included in the two models. Could this be made explicit in the report? Is it because the 'full' model requires prevalence of any nodule (>=3mm) irrespective of whether or not it is 'actionable'? And, if so, wouldn't it make sense to use the same figure in the 'simple' model, since the outcome of interest there is diagnostic accuracy?	Yes the main difference arose from the different modelling approaches, and hence different parameter inputs for the two models. The prevalence of 20.6% for the simple model was related to actionable nodules, whereas the prevalence of 50.9% for the full model was related to any lung nodules (≥3 mm). We did not use the sum of cat 3 and cat 4 nodules from the study by Field et al. (UKLS, #78) as input for the prevalence of actionable nodules as their definition of cat 3 nodules did not exactly match the BTS 2015 definition (e.g. volume ≥50 mm^3 instead of volume ≥80 mm^3). This might be also one explanation why the observed prevalence of 26.9% cat 3 and cat 4 nodules is higher than the 20.6% observed in the LSUT trial. The test accuracy data (sensitivity and specificity) used in the simple model were related to the detection of actionable nodules, and therefore it would not be appropriate to use the prevalence of 'any lung nodule' in that model.
Aidence BV	4	217	7.4.2	Table 40 indicates a prevalence value of 13% for nodules found incidentally on chest CT (based on evidence review conducted for the BTS guidelines, reference #11). It is worth noting that this value for prevalence of incidental nodules is likely an underestimation because the authors of the evidence review have simply averaged various reports on nodule prevalence found in CT with the various protocols. Cardiac CTs will by definition have less lung nodules because the lungs are only partially in the field of view. Also all/most vendors do not support nodule detection on cardiac or abdomen CTs.	In the absence of more comprehensive data, the EAG considered the values from the BTS guidelines to be appropriate. Cardiac and abdomen CT scan images that only provide partial view of the lung are outside the scope of this assessment. In a scenario analysis, we have varied the prevalence of nodules found incidentally to 38% (see Table 57, page 245).

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Aidence BV	5	217	7.4.2	Table 40 indicates a base case prevalence value of 13% for nodules found incidentally on chest CT (based on evidence review conducted for the BTS guidelines, reference #11). An alternative source for this value might be the 2015 study published by Gould et al, where they observed (via secondary use data in the US) nodule prevalence up to 31%. Gould MK, Tang T, Liu IL et al (2015) Recent trends in the identification of incidental pulmonary nodules. Am J Respir Crit Care Med 192:1208–1214 doi.org/10.1164/rccm.201505-0990OC We note that in figure 22, page 244 (probabilistic sensitivity analysis) the prevalence figure for this population is varied up to 24%. Preliminary results from our current, ongoing INPACT trial, indicate prevalence levels in the incidental population much closer to the 31% reported by Gould et al (2015) and, as the report makes clear "Higher prevalence of lung nodules is associated with more favourable cost-effectiveness for AI-assisted reading" (page 243), we would suggest using an upper limit of 31% in the PSA for the modelling. NB we further note that table 57 (page 245) indicates a potential range of nodule incidence (all nodules, not just actionable nodules) of 13% to 38%, which is different from the tornado diagram in figure 22. So, it is not entirely clear which range was used to inform the final conclusions. Could this be clarified?	Thank you for this alternative source. In one-way sensitivity analysis, our upper limit was 0.24 and we reported the results in the tornado diagram. In scenario analyses, we changed the base-case value from 0.13 to an alternative value of 0.38 to assess the impact to the results. The latter figure of 38% was obtained from a recent Dutch study in a general population without symptomsLancaster HL, Heuvelmans MA, Pelgrim GJ, Rook M, Kok MGJ, Aown A, de Bock GH, van den Berge M, Groen HJM, Vliegenthart R. Seasonal prevalence and characteristics of low-dose CT detected lung nodules in a general Dutch population. Sci Rep. 2021 Apr 28;11(1):9139. doi: 10.1038/s41598-021-88328-y Given that in this scenario the value used is higher than that recommended by the stakeholder, we think that there is no additional gain by undertaking/re-running this one-way sensitivity analysis. For clarity, '(0.1300 to 0.3800)' is not a range. It is a reminder to the reader the base-case input used, and the new input being considered in scenario analysis.
Aidence BV	6	241	8.1.2	States that "Table 54 presents the estimates of the costs and additional people correctly identified with an actionable nodule with the use of AI-assisted radiologist reading compared to unaided radiologist reading in a *symptomatic* population". Should read *incidental* population?	Thank you for identifying this error, which has been corrected in the report erratum.
Aidence BV	7	226	7.4.5	Table 46 input costs (plus also table 37, page 202) - the health economic models are predicated on a unit cost per scan of using the AI software of £2.00 and £3.34. Whilst the results of the probabilistic	Thank you for these helpful comments.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				sensitivity analyses summarised in the various tornado graphs indicate minimal sensitivity to this average cost, we would still urge caution in the publication in the final public report of average costs in this way, unless they are better described. Any modelled price needs to be (a) reflective of the whole cost of ownership of the software product (including any fixed costs such as licensing and hosting) and not just the variable price per scan element, and (b) sustainable in the long term. It is easy, say, for potential new market entrants to offer a heavily discounted price in order to gain market traction but such pricing may not actually be sustainable as the business grows. Ideally, the base case should be pegged to price points which have been proven to be relatively stable over several years so that they are more reflective of the reality of running these systems at scale, than a company with a zero or very low install base.	In response to your comments and upon request of NICE technical team, we have carried out additional sensitivity analyses in which we increased the upper bound to £6 to capture these additional costs that should be considered. The findings are presented in EAG report addendum.
Aidence BV	8	227	7.4.5	Table 46 input costs - the treatment costs by Stage of disease used in the health economic modelling are drawn from a 2014 analysis commissioned by Cancer Research UK (reference #72 refers). The CRUK analysis utilised data from 2010-2014 so the cost information is now at least a decade out of date. You would be hard pressed to find a physician who thought that, based on current available treatments, a Stage I lung cancer patient would cost more than a Stage IV patient, even allowing for disease recurrence in a Stage I case following curative-intent treatment and low survival rates of Stage IV cases. Modelling and cost estimation work undertaken by Dr Neal Navani (clinical lead of the UK lung cancer clinical audit) and presented to the 2nd BTOG lung cancer screening essential update on 21.06.2021 by Prof Mat Callister suggests that, owing to the very high costs associated with the use of immunotherapies, chemo-immunotherapies and tyrosine kinase inhibitors, the average Stage IV treatment costs for lung cancer are now ~ 10X the cost of Stages I/II.	. We thank the stakeholder for their recommendation. We agree that the treatment costs used in the EAG report may not reflect newer therapies currently used. As suggested, we undertook scenario analyses by using treatment costs in the NSC health economic modelling from Exeter for each population of interest to assess the impact on the deterministic results, and separately the probabilistic results (excluding surveillance). We did this using the diagnostic treatment costs, then diagnostic and recurrent costs. The findings are presented in the EAG report addendum.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				This work was fed into the March 2022 preliminary cost-effectiveness analysis undertaken by the Exeter Test Group and Health Economics Group to inform the recent National Screening Committee decision to recommend UK-wide rollout of lung cancer screening. Table 13, page 43, of that report sets out the latest cost estimations for stage-based treatment costs, including costs associated with disease recurrence and ongoing costs. As the Exeter report makes clear (page 39, final paragraph), these costs build upon the approach used in the 2014 CRUK report and were subject to clinical validation, led by Prof David Baldwin with consensus from the NHS Clinical Expert Group for lung cancer. We would therefore urge the Warwick Evidence team to revisit the treatment costs used in their health economic modelling and align them with the cost profile in the NSC health economic modelling from Exeter. It may also be worth including a realistic range of treatment costs in the PSA modelling by population cohort, given that the value (and ICER) derived from earlier detection and stage shift in lung cancer diagnosis will be heavily dependent on treatment costs by stage. Interestingly, Appendix 10 of the DAR (section 13.10, page 450), which compares the Warwick approach to that used by the Exeter team, makes no reference to the discrepancy in input costs used for treatment of lung cancer.	
contextflow	1	55	Table 1.	**For the product name:** contextflow SEARCH Lung CT (contextflow) We confirm that the CT scan types that can be processed are: “Low dose, regular dose with and without contrast.” Instead of only “With and without contrast”	Thank you for the clarification.
Siemens Healthineers	1			We thank NICE for the opportunity to comment on this Diagnostics Assessment Report.	-

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				As we have joined the process only recently we unfortunately missed the opportunity to submit comments or evidence on the draft scope. We apologize for potential inconveniences caused	
Siemens Healthineers	2	71	3.1.2 Table 2	We would like to reference a prospective randomized study that assesses the use of AI-Rad Companion Chest CT by Yacoub et al. (DOI: 10.2214/AJR.22.27598) that we feel falls within the scope of this Diagnostics Assessment Report.	Thank you for drawing our attention to this study. Unfortunately, the paper by Yacoub et al. was published in November 2022. We have only included articles in our review that were published by 31 August 2022. We had a look at the record and would have excluded it as the population included >10% oncologic patients. It would have been described in Table 66 though: **Population:** Single centre, USA; 390 patients who underwent outpatient chest CT; mixed indication (56% and 57% posttreatment surveillance for cancer recurrence in AI-assisted and unassisted groups, respectively). **Index test:** AI-Rad Companion (Siemens Healthineers). 3 experienced cardiothoracic radiologists reading 65 scans each. **Comparator:** Unaided reading; 3 experienced cardiothoracic radiologists reading 65 scans each. **Outcomes:** Interpretation times; mean reduction 22.1%. **Study design:** Prospective, randomised study (clinical practice); scans were randomized using 1:1 between AI-assisted and unaided arms (195 scans per arm).
Siemens Healthineers	3	51	1.4.1	We would like to draw the attention of NICE towards the fact that AI-Rad Companion Chest CT consists of three medical devices that focus on three main parts of the thorax: the lungs [AI-Rad Companion (Pulmonary)], the cardiovascular system [AI-Rad Companion (Cardiovascular)] and the spine [AI-Rad Companion (Musculoskeletal)]. The use case of AI-Rad Companion (Pulmonary) has been assessed in detail by the EAG already. Overall, the usage of AI-Rad Companion Chest CT might provide further benefits to the NHS that have not yet been considered in the assessment report and the economic model.	Thank you for the information. Detection and management of co-morbidities beyond lung nodules/cancer are outside the scope of this assessment. The EAG has highlighted evaluation of the cost-effectiveness of AI software capable of analysing chest CT scans for multiple clinical indications as one of future research priorities.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				We would like to point out additional anticipated benefits that come with the use of AI-Rad Companion Chest CT, especially in incidental populations: AI-Rad Companion (Cardiovascular) is able to measure the heart volume and quantification of coronary calcium volume. Furthermore, segmentation of the aorta and diameter measurements are possible. Both features (coronary calcification and aorta measurements) are listed in the recommendations by the ACR Incidental Findings Committee.[1] AI-Rad Companion (Musculoskeletal) provides labelling and segmentation of thoracic vertebras, measurement of vertebrae heights and quantification of vertebral density (in HU). Publications show that the HU-values of the spine obtained from CT data acquired for other indications can be used to identify osteoporotic patients.[2] This approach is also called “opportunistic screening for osteoporosis”, which can lead to the appropriate initiation of preventive measures and potentially save downstream healthcare costs. According to The International Osteoporosis Foundation “there is strong evidence of widespread under-diagnosis of vertebral fractures”.[3] Despite the reduction in lung cancer mortality due to lung cancer screening with low-dose CT in screening populations, many smokers die of comorbid smoking-related diseases. The identification of CT features associated with these comorbidities could increase the value of screening with minimal impact on lung cancer screening programs. As these smoking-related conditions are not systematically assessed in current lung cancer screening programs, AI can identify individuals with evidence of previously undiagnosed cardiovascular disease, or osteoporosis and offer an opportunity for treatment and prevention.[4]	
Siemens Healthineers	4	51	1.4.1	We would like to clarify that AI-Rad Companion Chest CT is commercially available in the UK.	Thank you for the clarification.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Siemens Healthineers	5	51	1.4.1	We would like to add that no mandatory customer trainings are required, however Siemens Healthineers offers both online training material as well as in-person trainings to familiarize clinicians with the use of AI-Rad Companion Chest CT	Thank you for the additional information.
Siemens Healthineers	6		3.2.1.2 "Use of any prototype software versions that did not later become the commercially available version (e.g. applicability not confirmed by the company)	We can we confirm the applicability as the commercially available software has been used in the studies. The wording "prototype" refers to the wrapper / application of AI-Rad Companion Chest CT. The core LungCAD algorithm used in the studies however is identical to the commercially available and released medical device. The following versions of the nodule detection algorithm were used in the respective publications: Abadia 2021: LungCAD VC30 Chamberlin 2021: LungCAD VC30 Rueckel 2021: LungCAD VD20 The LungCAD currently released and commercially available in the UK is LungCAD VD20. The VD-line is the successor of the VC-line and results can be expected to be superior.	Thank you for the clarification. The EAG would have been able to include the information in the EAG report if it were received earlier when we sought companies' clarifications.
Siemens Healthineers	7		3.2.1.2 "Study did not use a pre-specified nodule size threshold similar to the UK 2015 BTS guidelines (i.e. ≥5mm maximum	Abadia 2021 used a threshold of 4mm, Chamberlin 2021 and Rueckel 2021 used 6mm. The chosen pre-specified lung nodule size thresholds are in line with following guidelines: - Oncology: RECIST guideline (6mm) - Screening: LungRADS guideline (6mm in baseline, 4mm in Follow-Up) - Incidental: Fleischner Guideline: 6mm Although this is a slight deviation from the 2015 BTS guidelines, we want to clarify that Lung CAD VD20 is cleared for nodule size thresholds in line with the UK 2015 BTS guidelines.	Thank you for the clarification. The original statement in the report remains correct.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
			axial diameter or ≥80mm3)"	Lung CAD VD20 is cleared for the following type and average diameter: Solid in the range of 3.0mm to 30.0mm and subsolid (part-solid and ground glass) 5.0mm to 30.0mm.	
Siemens Healthineers	8	168	3.6.13 "AI-RAD Companion Chest (Coreline Soft)"	The manufacturer referred here should be Siemens Healthineers.	Thank you for identifying this error, which has been corrected in the EAG report erratum.
Siemens Healthineers	9	223-7	7.4.4, including Table 46	The full economic model may underestimate improvements in health outcomes and overestimate costs associated with the introduction of AI-assisted lung cancer detection in the different model populations. The model structure may not be adequate to evaluate stage shift which would be one of the key benefits of improved detection (i.e. the capacity to detect cancer at earlier stage, where patients can benefit from interventions with curative intent, as opposed to later stages where interventions will be palliative in nature and likely more costly). Furthermore, the data used to document that benefit is unclear, particularly in terms of the treatments, health outcomes and costs associated with each stage. The cost data documented in Table 46 page 227 is based on research originally conducted in 2014, while significant changes in the management of patients with lung cancer have occurred since. These changes include the introduction of NICE-recommended immunotherapy for patients with locally advanced or metastatic lung cancer (e.g. see TA655, published 21 October 2020, regarding nivolumab for advanced squamous non-small-cell lung cancer after chemotherapy, where the incremental costs of nivolumab over docetaxel were estimated at £31,275, for an ICER of £35,710, in the committee's preferred base case). Anti-PD-1 or PD-L1 checkpoint inhibitors are now standard of care for most patients with non-small cell lung cancer (NSCLC) treated in the NHS in England according to NG122. The costs of immunotherapy cannot be included in the	We thank the stakeholder for their comments. In response, we have undertaken scenario analyses using these costs from the interim report for the Exeter model used by the NSC that includes these novel therapies. The findings are presented in EAG report addendum. As we have assumed that cancer detected by CT scans during the initial or follow-up (surveillance) scans would be at stage I whereas undetected cancer would present at later stages, the 'stage shift' arising from the detection of additional cancer by AI-assisted reading compared with unaided reading would have been captured in our model. Survival following lung cancer diagnosis was estimated according to the stage at diagnosis. We obtained parametric model parameters from the Exeter model, who fitted Weibull models to Kaplan-Meier plots according to the clinical stage of diagnosis, then overall survival was estimated up to the 10-year time horizon. Treatment costs by stage are reported in Table 46 of the EAG report.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				estimates presented in Table 46, as the incremental costs of immunotherapy alone exceeds the value reported for stage IV in that table, and since the marketing authorisations for these drugs were granted after 2014. Conversely, stereotactic ablative radiation therapy (SABR) is now available as an option for patients with earlier stage disease, e.g. for people with stage I–IIA (T1a–T2b, N0, M0) NSCLC who decline lobectomy or in whom it is contraindicated, in an updated to NG112 published in 2019. The introduction of SABR may improve health outcomes while being less resource intensive than surgery. Overall, it is unlikely that the management of patients with earlier stage lung cancer in the NHS is more costly than that of stage IV, as is currently the case in the economic model. A recent cost consequence analysis in the Canadian payer perspective indeed estimated that treatment costs for lung cancer increased with stage at diagnosis, from 84,158.62 CAD at stage I to 178,446.00 CAD at stage IV, and demonstrated that low-dose computed tomography (LDCT) was a cost-saving screening intervention in this setting as it resulted in stage shift, with 75.00% of patients diagnosed at early stages with LDCT screening compared to 31.57% without screening (doi: 10.1016/j.jtocrr.2022.100350). This is consistent with data from randomised controlled trials of lung cancer screening in the UK reviewed by Balata et al. (doi: 10.1016/j.lungcan.2021.09.012) where 47.6% to 85.7% of patients were diagnosed with early stage (I or II) lung cancer in LDCT screening programmes, therefore emphasising the need to precisely estimate the long-term health outcomes and cost implications of AI-assisted lung cancer detection	
Siemens Healthineers	10			References: [1] DOI: 10.1016/j.jacr.2018.04.029 [2] DOI: 10.7326/0003-4819-158-8-201304160-00003; DOI: 10.1016/j.ejrad.2021.109568 [3] https://www.osteoporosis.foundation/sites/iofbonehealth/files/2021-01/2011_VertebralFractureInitiative_PartII_VertebralFractures_Eng.pdf [4] DOI: 10.3390/diagnostics12102435	Thank you for providing these references (two policy documents; two associated with osteoporosis; and one narrative review). None of them would meet our inclusion criteria.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
SynApps Solutions Limited / Mevis Medical Solutions	1	62	Footnote	Regarding the footnote about asymptomatic patients: We think this is your interpretation. What it means (at least from our side) is that it does not need the person to have any symptoms, but that images acquired from people with and without symptoms can be processed. Chapter 2.2.1.1 of the attached user guide defines the intended medical indication: “Veolity is intended to be used for reading CT examinations of the chest, e.g. for screening, diagnosis and monitoring of lung cancer.” For me this actively includes the symptomatic population.’	Thank you for your clarification. We have removed the asterisk linking this footnote with Veolity in the report erratum.
SynApps Solutions Limited / Mevis Medical Solutions	2	54	1.4.11 Veolity (MeVis)	As above we believe the Veolity documentation correctly describes the software is applicable to both symptomatic and Asymptomatic patients.	Thank you for clarification. We have corrected the information in the report erratum.
SynApps Solutions Limited / Mevis Medical Solutions	3	54	1.4.11 Veolity (MeVis)	Minimum amount of training is 1 hour. Longer is only required if the customer wishes to work through additional worked examples. eg. as part of a team training session.	Thank you for the additional information.
The Society & College of Radiographers	1	7	Results	The last paragraph refers to disutilities but it is not clear how that term is being used in this context. Can explanation be provided in the section ‘definition of terms and list of abbreviations’?	Disutility refers to a reduction in valued quality of life due to diseases, symptoms or specific circumstances, for example anxiety that might be caused by not knowing whether an identified lung nodule is malignant or benign.
The Society & College of Radiographers	2	7	Conclusion	The section concludes that ‘AI-assisted image analysis may be cost-effective for the screening population but may be dominated by unaided analysis for the symptomatic population’. It is not clear what dominated by unaided analysis for the symptomatic population means. Does domination refer to the number of research studies available?	The term “dominated” was used as a technical term in economic evaluation and means that AI-assisted image analysis may cost more but produce less health benefit compared with unaided analysis for the symptomatic population.
The Society & College of Radiographers	3	33	Accuracy and reliability	The paragraph states ‘Evidence from one UK study suggests that unaided, experienced radiologists in clinical practice (5% double reading) outperform unexperienced, trained radiographers assisted with concurrent AI who read the same screening CT images as part of a reader study.’	The UK study being referred to was Hall et al. 2022: Hall H, Ruparel M, Quaife SL, Dickson JL, Horst C, Tisi S, et al. The role of computer-assisted radiographer reporting in lung cancer screening programmes. Eur Radiol 2022;32(10):6891-9. http://dx.doi.org/10.1007/s00330-022-08824-1

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Radiographers working in the UK are regulated by the Health and Care Professions Council (HCPC). The title Diagnostic Radiographer or Therapeutic Radiographer are regulated and used to differentiate the type of pre-registration degree path that the individual has followed. It is not clear if the report refers to diagnostic or therapeutic radiographers (different degree programmes and curricula). We presume that it refers to diagnostic radiographers. It is not clear what is meant by 'unexperienced, trained radiographers' – does that refer to newly qualified staff in preceptorship period, post HCPC registration? Alternatively, does that refer to experienced radiographers who have undertaken further postgraduate MSc level reporting courses/training and are working in preceptorship period in CT? Radiologists and radiographers undertake different roles, and it is not clear why the two have been compared in this instance without further context.	The description provided in the paper was: "Two radiographers (denoted as R1 and R2) experienced in thoracic CT acquisition and with prior qualification in chest radiograph reporting, but without prior experience in thoracic CT reporting". The authors did not specify whether the radiographers were diagnostic or therapeutic radiographers, but from the context it was very likely that diagnostic radiographers were involved. We could only provide a very brief description of the radiographers on page 33 as it was one of the summary sections of the report. More details were provided on page 103 of the report. The authors of the study wanted to explore whether trained radiographers assisted with computer software "may offer strategies to optimise the use of radiology resources without loss of sensitivity" in the context of lung cancer screening.
The Society & College of Radiographers	4	46	1.2.4	The paragraph refers to radiographers. As per comment number 3 above, please clarify that this refers to diagnostic radiographers.	Please see our response to comment 3 above.
The Society & College of Radiographers	5	49	1.2.6	'Treatment for lung cancer is based on several factors' – the factors listed here are all clinical. It is important to note that the views and wishes of the individual person with lung cancer must also be adhered to, assuming their capacity to consent. This enacts person-centred care as per the UK Supreme Court decision on consent, the 2015 Montgomery judgment: Values based practice in diagnostic and therapeutic radiography (sor.org)	Thank you for highlighting this important point.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
The Society & College of Radiographers	6	50-55	1.4	At the section 'Description of technology under assessment.' The information provided in this section is limited. Consideration of core assessment criteria / technical aspects of clinical safety, data protection, technical security, and interoperability criteria – as per digital technology assessment criteria for health and social care (NHS England transformation directorate), is best practice for NHS evaluation of each AI system. Was that information available and considered for this review?	Thank you for highlighting this information. We agree that these criteria are very important guidance for the evaluation of digital technologies including AI-derived software, but noted from the NHS England website that the criteria were "designed to be used by healthcare organisations to assess suppliers at the point of procurement or as part of a due diligence process". Our technology assessment focused on evaluating test accuracy, clinical effectiveness and cost-effectiveness, and therefore had different emphasis.
The Society & College of Radiographers	7	202	Table 37	'Costs inputs used in the model'. Society of Radiographers question the validity of 'band 9 radiographer as a proxy for a radiologist' – reporting radiographers are commonly paid at Agenda For Change band 7 not band 9.	We assumed in our model that the CT scan images will be read by a consultant radiologist. However as the source of our cost information obtained from Jones and Curtis 2021 only provided cost information for radiographers of various bands but not for radiologist, we have used the cost for band 9 radiographer to approximate the cost for a consultant radiologist. This does not imply in any way that we assume that CT scan images are read to band 9 radiographers. Jones K, Burns A. Unit costs of health and social care 2021. Personal Social Services Research Unit, University of Kent, Canterbury; 2021. http://dx.doi.org/10.22024/UniKent/01.02.92342
The Society & College of Radiographers	8	205	Figure 15	Please note that there is a problem with the text on the diagram – the first two lines are clear to read. The remainder of the text has become blurred in draft.	We are able to read the different lines in the diagram with similar clarity/resolution in the report, but will be happy to supply a different version of the graph if requested.
The Society & College of Radiographers	9	226	7.4.5	'used a band 9 radiographer by proxy'. It is not clear why a band 9 radiographer was used by proxy. It is very rare for a radiographer to be paid at Agenda For Change pay band 9. See, for example, College of Radiographers: Diagnostic radiography workforce census 2021 (sor.org) and Society of Radiographers: Survey of trainee consultant and consultant radiographers 2020 | SoR	As explained above, we used the cost information to represent the cost for a consultant radiologist and did not assume that the CT scan images are read by radiographers.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
The Society & College of Radiographers	10	231	7.4.10	Time taken to read CT Scans. We note that the Royal College of Radiologists have recently consulted on draft guidance for radiologist home working. Radiologists and reporting radiographers do now report from home; this is an additional variable that does not seem to have been taken into account. For example, will access to AI systems take longer on home networks/slow reporting times if the reporter is waiting for the AI system.	Thank you for highlighting this important issue.
The Society & College of Radiographers	11	260	9	'Integration of the technologies into existing picture archiving and communication system (PACS) and workflow; compatibility with existing CT scanners and workstations' We note that in addition the systems may be required to integrate with Radiology Information System (RIS) and across systems, for example, electronic patient record (EPR).	Thank you for this helpful comment.
UCLH NHS Foundation Trust	1	7, then multiple	8.1.3	The assumption that AI systems overall provide increased specificity in the screening population modelling needs to be more robustly explained. From the evidence review provided, none of the studies (as far as I can see) support the fact that AI provided statistically improved specificity over unaided reading, and in some cases also caused decreased specificity. Given that the assumption of improved specificity drove the dominance of AI for this sub-group.	As we described in Section 7.4.3 of our report, data for the estimated sensitivity and specificity for the base case for the screening population were obtained from Hsu et al. 2021, which was the only study with suitable data identified in our review. We are aware that while point estimates from the study suggested that AI assisted reading had slightly higher specificity compared with unaided reading, the difference was not statistically significant and was associated with substantial uncertainty, which we highlighted throughout our report. We undertook an additional scenario analysis assuming decreased specificity with software, using the lower limit of the 95% confidence interval from the study by Hsu et al. 2021 for AI-aided reading (i.e., 85%) and the upper limit of the 95% confidence interval for unaided reading (i.e., 90%). The findings are presented in EAG report addendum.
UCLH NHS Foundation Trust	2	35 and 143	3.5.1	Would be useful to define whether segmentation failure included unreliable segmentation as well, as this is not clear.	Segmentation failure included unreliable segmentation as well. We state "However, the observed nodule segmentation failure might be mostly due to rejections of segmentation results by radiologists, rather than the inability of the system to segment the nodule."

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
UCLH NHS Foundation Trust	3	55	Table 1	Table 1 should include a column indicating whether the software allows the user to manipulate or add nodules	We were not able to include this information systematically in Table 1 as the information was not provided for several of the software. However, where available we have included the information in the descriptions of the technologies on pages 51-54 of the EAG report.
UK National Screening Committee	1	7		The limitations are noted and expected from the way this technology has advanced, with some products not undergoing adequate testing. More scrutiny in this field is required if we are to be able to be confident their introduction is worthwhile.	Thank you for your comment.
UK National Screening Committee	2	7		The need for ongoing clinical evaluation of the implemented technology is important and should be incorporated into quality assurance.	Thank you for raising this point. We also highlighted on page 270 of the EAG report that “Ongoing audit of potential impact of these updates [of software] on test accuracy and service provision may be desirable”.
UK National Screening Committee	3	7		The lack of cost effectiveness in the clinical setting should not be over-emphasised as this will be heavily influenced by individual service models and is very likely to be the same as for screening in those centres where they have good nodule management services	Thank you for your comment. Our statements were based on findings of our cost-effectiveness analyses, with attendant caveats.
UK National Screening Committee	4	33-35		This is a thorough evaluation and consistent with the current thinking It would be helpful to grade the quality of evidence for each of the products	Thank you for this suggestion. Given the predominantly high risk of bias rating and high applicability concerns across included studies, grading of individual products would lead to designation of low quality / certainty evidence across all products. We have however highlighted the substantial differences in the volume of evidence between individual products on pages 264-265 of the EAG report.
UK National Screening Committee	5	51-55		It would be helpful to give some information as to whether the level of evidence for a given product is commensurate with the medical device classification	Thank you for this suggestion. The EAG considers judging the adequacy of evidence against medical device classification for the software being outside the remit of our assessment, and also noted that the MHRA stated that “existing classification rules are, in some respects, out of step with best international practice - particularly for… software as a medical device” in a recent public consultation : https://www.gov.uk/government/consultations/consultation-

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					on-the-future-regulation-of-medical-devices-in-the-united-kingdom/chapter-2-classification
UK National Screening Committee	6	262		Was there any difference in the impact the AI when considering nodule type - solid, part solid or pure ground glass?	The effect of nodule type on detection accuracy (subquestion 1-3) is described on pages 111-117. Differences in segmentation failure by nodule type (1 study) are reported on pages 143/144. Overall, very limited evidence was found and the impact of AI was hard to estimate. For example, while AI-assisted reading was found to improve the sensitivity for detecting part-solid and pure ground glass nodules more than the detection of solid nodule in isolated studies, the impact on specificity was inconsistent and higher segmentation failure rates were observed for part-solid and pure ground glass nodules compared with solid nodules in another study.

Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images

Diagnostics Assessment Report (DAR) and economic model - Comments

Section B: Comments on the economic model

	Issue	Description of problem	Description of proposed amendment	Result of amended model or expected impact on the result (if applicable)	
Aidence BV	1	The variables included in the 'AI Illustrative structure SCREENING DETECTION ACTIONABLE.trex' model do not correspond to section 6 of the DAR (e.g. the file uses nodule prevalence of 50.9%, not 20.6% as per section 6.3.1, page 200, and the base case test accuracy values are all different). However, the results tables in section 6 (such as table 38, page 204) do agree with the model outputs, suggesting that the model inputs have been incorrectly documented in the report. See also comment no. 3 in Section A above.	Align the report with the corresponding model files and outputs so that there is no discrepancy	Don't know - we haven't updated the model as it's not clear which values were intended to be used as variable inputs	Many thanks for highlighting this discrepancy between the model submitted and the inputs documented in the report. We have now submitted the correct model.