Technology Assessment Report commissioned by the NIHR Evidence Synthesis Programme on behalf of the National Institute for Health and Care Excellence

Title: Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images [DAP060] - Addendum

Produced by	Warwick Evidence, Warwick Medical School, University of Warwick, [redacted]
Authors	Julia Geppert, Research Fellow, Warwick Evidence and Warwick Screening Peter Auguste, Assistant Professor, Warwick Evidence Asra Asgharzadeh, Research Fellow, Warwick Evidence Hesam Ghiasvand, Research Fellow, Warwick Evidence Mubarak Patel, Research Associate, Warwick Evidence Anna Brown, Information Specialist, Warwick Evidence Surangi Jayakody, Assistant Professor, Warwick Evidence Emma Helm, Consultant Radiologist, University Hospitals Coventry and Warwickshire Dan Todkill, Associate Clinical Professor, Warwick Medical School Jason Madan, Professor, Warwick Medical School Chris Stinton, Senior Research Fellow, Warwick Evidence and Warwick Screening Daniel Gallacher, Assistant Professor, Warwick Evidence Sian Taylor-Phillips, Professor, Warwick Screening Yen-Fu Chen, Associate Professor, Warwick Evidence
Correspondence to	Dr Yen-Fu Chen Warwick Evidence, Division of Health Sciences Warwick Medical School, University of Warwick, [redacted] Email: [redacted]

Date completed 16th January 2023

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number 13/53/25.

Declared competing interests of the authors

None.

Acknowledgements

Prof. Ben Glocker (Reader in Machine Learning for Imaging, Imperial College London) provided expert technical advice to the EAG during the preparation of the report.

Prof. Charles Hutchinson (Professor of Clinical Imaging, CSRL, University Hospitals Coventry and Warwickshire) provided expert clinical advice to the EAG during the preparation of the report.

We are grateful to the Specialist Committee members for their advice during our preparation of the economic models.

We thank Sarah Abrahamson for her assistance in editing and proof-reading the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Geppert J, Auguste P, Asgharzadeh A, Ghiasvand H, Patel M, Brown A, Jayakody S, Helm E, Todkill D, Madan J, Stinton C, Gallacher D, Taylor-Phillips S, Chen Y-F. Software with artificial intelligence derived algorithms for automated detection and analysis of lung nodules from CT scan images: A Diagnostics Assessment Report. Warwick EAG, 2022

Contributions of authors

Julia Geppert: Performed the clinical effectiveness review and wrote associated sections of this report.

Peter Auguste: Led the cost-effectiveness components of this project, undertook systematic review of the health economic literature, constructed health economic models and wrote economics sections of this report.

Asra Asgharzadeh: Performed the clinical effectiveness review and wrote associated sections of this report.

Hesam Ghiasvand: Performed the systematic review of the health economic literature, undertook health economic modelling and wrote economics sections of this report.

Mubarak Patel: Performed statistical analyses and wrote associated sections of the report.

Anna Brown: Developed the search strategies, undertook searches, managed references and wrote the search methods sections of this report.

Surangi Jayakody: Performed a review of the literature on overdiagnosis, supported the clinical effectiveness review and report writing of associated sections.

Emma Helm: Provided expert clinical advice and helped development of economic models.

Dan Todkill: Commented on earlier versions of the report and assisted in revising the report.

Jason Madan: Provided methodological advice on economic modelling and revised associated sections of the report.

Chris Stinton: Provided methodological advice and training for the clinical effectiveness team, acted as 3rd reviewer, and assisted in revising the report.

Daniel Gallacher: Provided statistical advice on simulation and assisted in revising the report.

Sian Taylor-Phillips: Contributed to study design and protocol, acted as senior advisor, provided methodological support, and assisted in revising the report.

Yen-Fu Chen: Led the project, its coordination and implementation, and write-up.

Please note that:

Sections highlighted in yellow and underlined are ‘academic in confidence’ (AIC). Sections highlighted in aqua and underlined are ‘commercial in confidence’ (CIC).

Contents

This addendum includes: (1) additional analyses undertaken by the EAG; (2) additional information provided to NICE since EAG's submission of the final diagnostic assessment report, in response to stakeholders' comments received and/or upon request of NICE's technical team.

1. Additional analyses

Additional Scenario analysis 1: assigning QALY decrement to all people under CT surveillance

- **Request**: 'Could you please do a scenario analysis for all the 3 populations (symptomatic, incidental, screening) and surveillance where instead of only people who have CT surveillance for nodules that are later diagnosed as benign, everyone having CT surveillance is assigned the QALY decrement -0.063.

Table 1: Scenario analysis results based on cost per QALY (Decrement of 0.063 assigned to people under surveillance)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
Symptomatic population					
Unaided radiologist reading	715,450	-	6349.06	-	-
AI-assisted radiologist reading (InferRead CT Lung)	816,520	101,080	6328.88	-20.18	Dominated
Incidental population					
AI-assisted radiologist reading (ClearRead CT)	229,210	-	6571.01	-	-
Unaided radiologist reading	231,640	2,430	6573.47	2.46	987
Screening population					
AI-assisted radiologist reading (ClearRead CT)	400,410	-	6531.30	-	-
Unaided radiologist reading	470,630	70,220	6523.37	-7.93	Dominated
Surveillance population					

AI-assisted radiologist reading (ClearRead CT)	699,100	-	6344.80	-	-
Unaided radiologist reading	898,678	199,578	6301.30	-43.50	Dominated
CT, computed tomography; QALY, quality adjusted life-year					

Additional Scenario analysis 2: Use of treatment costs from NSC Exeter model

- Request: Use treatment costs from the Exeter interim report, and re-run probabilistic sensitivity analyses using these costs; the EAG identified two possible sets of treatment costs (see Table 13 of the Exeter interim report): one under the heading "Diagnosis" and the other under the heading "Recurrence". The EAG therefore ran two sets of analyses, one using costs under "Diagnosis" only (see Table 2), and the other using "Diagnosis" plus "Recurrence" (see Table 3)

Table 2: Treatment costs by stage following diagnosis of lung cancer

Stage	Treatment costs following diagnosis	Source
Stage I	£5094	Exeter interim report
Stage II	£5537	
Stage III	£17,999	
Stage IV	£16,456	

Table 3: Treatment costs by stage following diagnosis and recurrence of lung cancer

Stage	Treatment costs following diagnosis and recurrence	Source
Stage I	£20,928	Exeter interim report
Stage II	£29,797	
Stage III	£32,830	
Stage IV	£21,838	

Use treatment costs following diagnosis from the Exeter interim report (Table 2) and re-run probabilistic sensitivity analyses using these costs

Symptomatic population

Table 4: Deterministic sensitivity analysis results based on expected costs and expected QALYs (symptomatic population of 1,000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
Unaided radiologist reading	607,150	-	6349.89	-	-
AI-assisted radiologist reading (InferRead CT Lung)	680,340	73,190	6329.90	-19.99	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Table 5: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (symptomatic population of 1,000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
Unaided radiologist reading	606,650	-	6349.90	-	-
AI-assisted radiologist reading (InferRead CT Lung)	680,550	73,900	6329.90	-20.00	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Incidental population

Table 6: Deterministic sensitivity analysis results based on expected costs and expected QALYs (incidental population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (InferRead CT Lung)	202,660	-	6571.19	-	-
Unaided radiologist reading	208,950	6,290	6573.63	2.44	2,600
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Table 7: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (incidental population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (InferRead CT Lung)	202,490	-	6571.41	-	-
Unaided radiologist reading	208,440	5,950	6573.92	2.51	2,400
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Screening population

Table 8: Deterministic sensitivity analysis results based on expected costs and expected QALYs (screened population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (ClearRead CT)	299,060	-	6532.1	-	-
Unaided radiologist reading	375,260	76,200	6524.1	-7.95	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Table 9: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (screened population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (ClearRead CT)	299,110	-	6532.09	-	-
Unaided radiologist reading	374,990	75,880	6524.16	-7.93	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Surveillance population

Table 10: Deterministic results based on expected costs and QALYs (screening population of 1,000 people undergoing CT surveillance)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (InferRead CT Lung)	523,509	-	6365.01	-	-
Unaided reading	711,501	187,992	6323.07	-41.94	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Use treatment costs following diagnosis and recurrence from the Exeter interim report (Table 3) and re-run probabilistic sensitivity analyses using these costs

Symptomatic population

Table 11: Deterministic sensitivity analysis results based on expected costs and expected QALYs (symptomatic population of 1,000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
Unaided radiologist reading	773,530	-	6349.89	-	-
AI-assisted radiologist reading (InferRead CT Lung)	881,990	108,460	6329.90	-19.99	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Table 12: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (symptomatic population of 1,000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
Unaided radiologist reading	777,010	-	6349.77	-	-
AI-assisted radiologist reading (InferRead CT Lung)	881,030	107,020	6330.02	-19.75	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Incidental population

Table 13: Deterministic sensitivity analysis results based on expected costs and expected QALYs (incidental population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (InferRead CT Lung)	237,040	-	6571.19	-	-
Unaided radiologist reading	238,050	1,010	6573.63	2.44	414
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Table 14: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (incidental population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (InferRead CT Lung)	237,120	-	6571.09	-	-
Unaided radiologist reading	238,330	1,210	6573.66	2.57	470
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Screening population

Table 15: Deterministic sensitivity analysis results based on expected costs and expected QALYs (screening population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (ClearRead CT)	430,290	-	6532.08	-	-
Unaided radiologist reading	497,540	67,260	6524.12	-7.95	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Table 16: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (screening population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (ClearRead CT)	430,440	-	6532.03	-	-

Unaided radiologist reading	497,980	67,530	6524.07	-7.96	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Surveillance population

Table 17: Deterministic results based on expected costs and QALYs (screening population of 1,000 people undergoing CT surveillance)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (InferRead CT Lung)	751,876	-	6365.01	-	-
Unaided reading	955,235	203,359	6323.07	-41.94	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Additional Scenario analysis 3: changing the specificity for AI-assisted and unaided reading for the original EAG base case for the screening population

- Request: In the original EAG base case for the screening population, AI-assisted reading had both better sensitivity and specificity compared with unaided reading. However, as the improved specificity for AI-assisted reading is contested. Consequently, his additional scenario analysis explores the impact of estimates for specificity under a scenario in which AI-assisted reading had worse specificity compared with unaided reading. The alternative values used in this scenario analysis are shown in Table 12. Here we used the upper confidence limit for unaided (0.90) and lower confidence limit for AI-assisted reading (0.85) from the Hsu study.

Table 18: Specificities used in base-case analysis and scenario analyses

Strategy	Base-case	Scenario analysis
AI-assisted radiologist reading	0.88	0.85
Unaided radiologist reading	0.86	0.90

The results are shown in Table 13.

Table 19: Deterministic sensitivity analysis results based on expected costs and expected QALYs (screening population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (ClearRead CT)	402,550	-	6531.15	-	-
Unaided radiologist reading	468,060	65,510	6525.36	-5.79	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Probabilistic ICERs for all the EAG original base case analyses

- Request: Could you please provide probabilistic ICERs for all the base case analyses where you report ICER as cost per QALY?'

In Table 20 through to Table 22, we present the results of the probabilistic ICERS for the symptomatic, incidental and screening population, respectively.

Symptomatic population

Table 20: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (symptomatic population of 1,000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
Unaided radiologist reading	714,680	-	6350.00	-	-
AI-assisted radiologist reading (InferRead CT Lung)	816,660	101,980	6329.80	-20.2	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Incidental population

Table 21: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (incidental population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (InferRead CT Lung)	228,870	-	6571.26	-	-

Unaided radiologist reading	231,370	2,500	6573.74	2.48	1,008
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Screening population

Table 22: Probabilistic sensitivity analysis results based on expected costs and expected QALYs (screening population of 1000 people undergoing CT scan)

Strategy	Expected total costs (£)	Incremental costs (£)	Expected QALYs	Incremental QALYs	ICER (£) per QALY
AI-assisted radiologist reading (ClearRead CT)	400,200	-	6532.14	-	-
Unaided radiologist reading	470,080	69,880	6524.16	-7.98	Dominated
CT, computed tomography; ICER, incremental cost-effectiveness ratio Exact results have been obtained from TreeAge but were rounded by the authors and presented.					

Sensitivity analysis increasing to the upper limit for costs associated with per scan/output

In this sensitivity analysis, we increased the upper limit for costs associated with average cost per scan/output to £6 in order to cover the possible ranges of costs to account for set-up and maintenance costs averaged over expected number of scans performed.

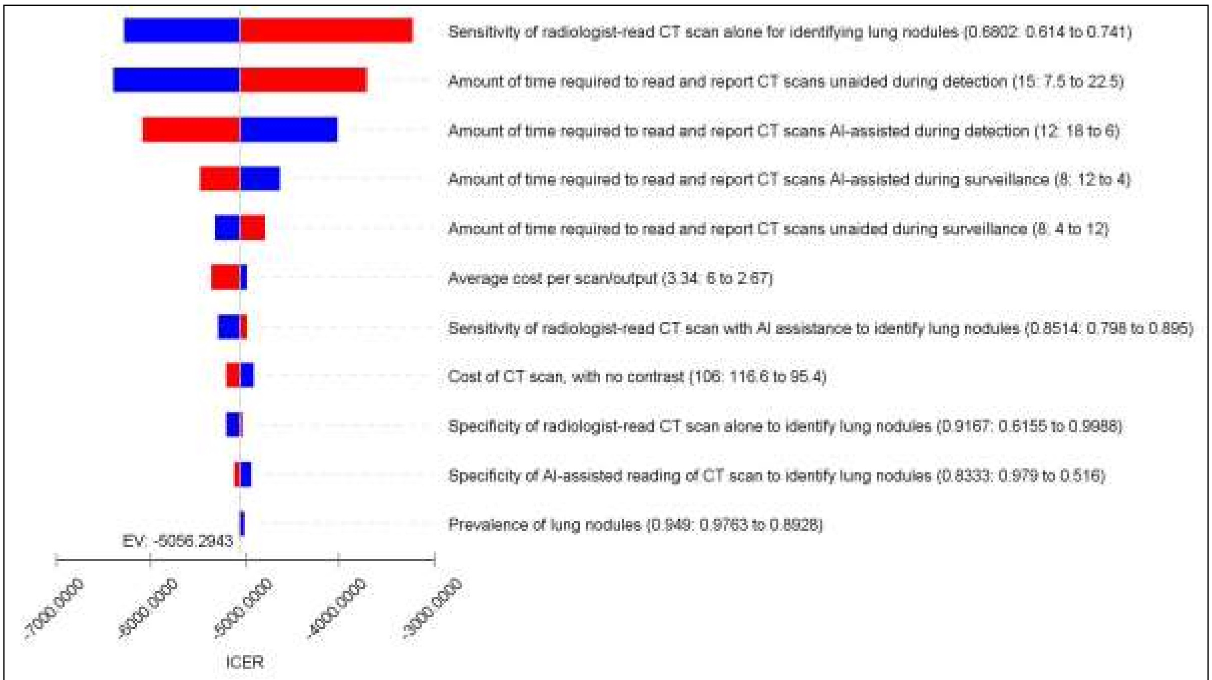

Figure 1: Tornado diagram of the impact to the cost per QALY by changing individual parameters and increasing the upper limit for cost of technology (symptomatic population)

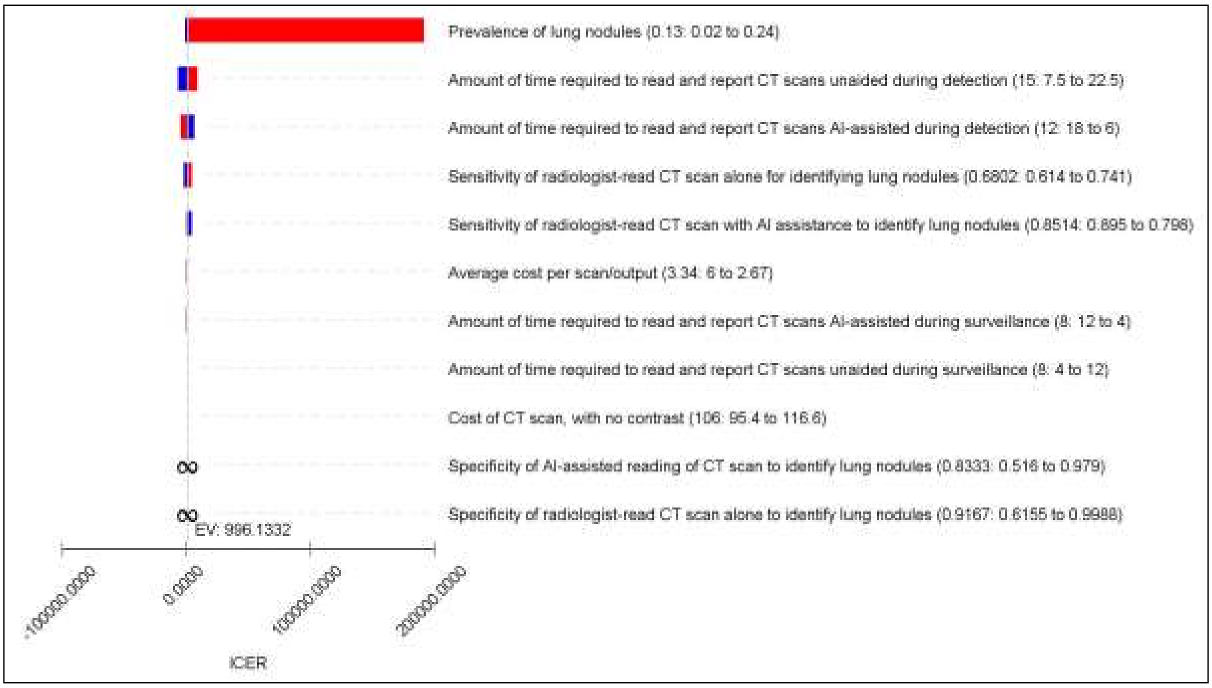

Figure 2: Tornado diagram of the impact to the cost per QALY by changing individual parameters and increasing the upper limit for cost of technology (incidental population)

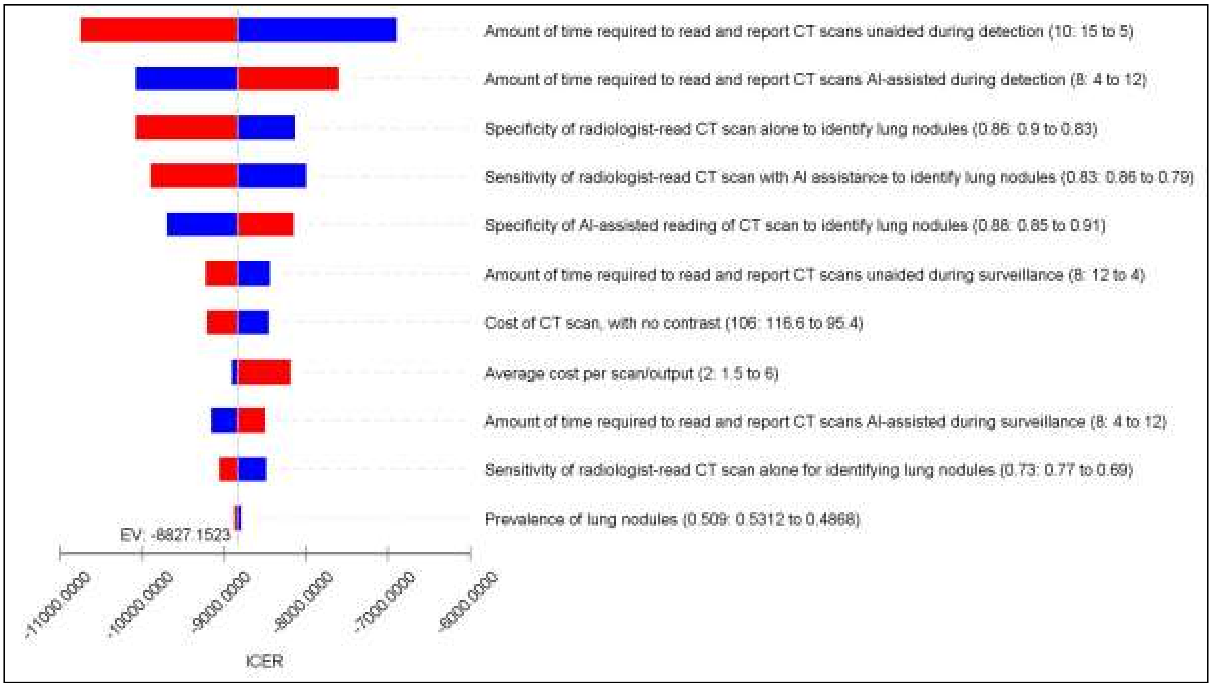

Figure 3: Tornado diagram of the impact to the cost per QALY by changing individual parameters and increasing the upper limit for cost of technology (screening population)

2. Additional information

Proportions of people diagnosed with lung cancer by stage

Request: Please provide the proportions of people that are diagnosed with lung cancer stages 1, 2, 3 and 4 in each strategy in each population

From the cancers missed in the symptomatic population, AI-assisted radiologist reading would identify an additional 3 cancers, which we assumed would present at stage I rather than stage III/IV if detected later. We assumed that 15.84% and 84.16% would present at stage III and IV, respectively. In the incidental population, AI-assisted radiologist reading would identify an additional 0.3411 cancers, which we assumed would present at stage I rather than stage III/IV if detected later. We assumed that 15.84% and 84.16% would present at stage III and IV, respectively. However, in the screening population, AI-assisted radiologist reading would identify an additional 0.5921 cases at stage I. We assumed that 15%, 8%, 22% and 55% would present at stage I, II, III and IV, respectively.

Table 23: Stage shift among additional lung cancer cases detected by AI-assisted radiologist reading strategy for a cohort of 1000 people undergoing CT scans

Stage at diagnosis	Symptomatic population		Incidental population		Screening population	
	AI-assisted radiologist reading	Unaided radiologist reading	AI-assisted radiologist reading	Unaided radiologist reading	AI-assisted radiologist reading	Unaided radiologist reading
Stage I	2.6830	0	0.3411	0	0.5921	0.0794
Stage II	0	0	0	0	0	0.0423
Stage III	0	0.4250	0	0.0540	0	0.1164
Stage IV	0	2.2380	0	0.2871	0	0.2910

Data for model inputs obtained from EAG simulations

Request: Could you please let us know what the following model inputs are:

(A) Proportions of people in different categories of nodule size based on the initial true nodule type and size in each of the 3 populations (symptomatic, incidental, screening)

(B) Proportions of people assigned to different management options in the initial detection phase of the model in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening)

(C) Proportions of people in different growth categories based on true nodule growth at each CT scan in CT surveillance in each of the 3 populations (symptomatic, incidental, screening)

(D) Proportions of people assigned to different management options in CT surveillance in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening)

(E) Proportions of people assigned to the different management option categories after initial detection and in CT surveillance getting the different investigations and treatment (table 46 on costs inputs used in the model, DAR pages 226-227) in each of the 3 populations (symptomatic, incidental, screening)

EAG response: As information on nodule size distribution for the incidental population was not available, we have assumed that it is the same as the nodule size distribution for the screening population. Consequently, the data are presented for symptomatic and incidental/screening population.

Please note that the proportions reported here are applied into the economic model. Prior to this, the economic model applies sensitivity and specificity for the detection of a nodule for both AI-assisted and unaided reading. The better sensitivity of AI-assisted reading detects more nodules which increases the number of people going through surveillance in the economic model.

We provide data for each sub-item below.

(A) Proportions of people in different categories of nodule size based on the initial true nodule type and size in each of the 3 populations (symptomatic, incidental, screening)

Table 24. Proportions of people in different categories of nodule size based on the initial true nodule type and size in each of the 3 populations (symptomatic, incidental, screening)

	Symptomatic		Screening and incidental	
Solid nodules	725,086		527,739	
Size classification				
< 5 mm	409,640	56.5%	323,685	61.3%
≥ 5 mm & < 8 mm	180,701	24.9%	81,448	15.4%
≥ 8 mm	134,745	18.6%	122,606	23.2%
Sub-solid nodules	168,159		57,100	
Size classification				
< 5 mm	23,335	13.9%	13,750	24.1%
≥ 5 mm	144,824	86.1%	43,350	75.9%

(B) Proportions of people assigned to different management options in the initial detection phase of the model in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening)

Table 25. Proportions of people assigned to different management options in the initial detection phase of the model in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening)

	Symptomatic population				**Screening/incidental population**			
Solid nodules	N=725,086				N=527,739			
	Software-assisted		**Unaided**		**Software-assisted**		**Unaided**	
Discharge	362,889	50.0%	425,970	58.7%	301,433	57.1%	320,105	60.7%
3-month CT	231,193	31.9%	192,783	26.6%	148,947	28.2%	136,721	25.9%
1 year CT	93,742	12.9%	73,333	10.1%	47,099	8.9%	42,243	8.0%
MDT	37,262	5.1%	33,000	4.6%	30,260	5.7%	28,670	5.4%
Sub-solid nodules	N=168,159				N=57,100			
	Software-assisted		**Unaided**		**Software-assisted**		**Unaided**	
Discharge	37,054	22.0%	48,388	28.8%	16,047	28.1%	20,237	35.4%
3-month CT	131,105	78.0%	119,771	71.2%	41,053	71.9%	36,863	64.6%

Table 26 on the next page provides further information across the initial detection and subsequent surveillance phases.

Table 26. Proportions of people assigned to different management options in the all phases of the model in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening)

	Symptomatic population				Screening/incidental population			
	Software-assisted		Unaided		Software-assisted		Unaided	
Solid nodules	725,086				527,739			
Discharged	**685,287**	**94.5%**	**689,679**	**95.1%**	**496,069**	**94.0%**	**498,065**	**94.4%**
At baseline	362,889	50.0%	425,970	58.7%	301,433	57.1%	320,105	60.7%
At 3 months	0	0.0%	0	0.0%	0	0.0%	0	0.0%
At 12 months	0	0.0%	1	0.0%	5	0.0%	2	0.0%
At 24 months	322,398	44.5%	263,708	36.4%	194,631	36.9%	177,958	33.7%
Definitive management	**39,799**	**5.5%**	**35,357**	**4.9%**	**31,670**	**6.0%**	**29,674**	**5.6%**
At baseline	37,262	5.1%	33,000	4.6%	30,260	5.7%	28,670	5.4%
At 3 months	1,814	0.3%	1,706	0.2%	1,058	0.2%	806	0.2%
At 12 months	366	0.1%	350	0.0%	183	0.0%	102	0.0%
At 24 months	357	0.0%	301	0.0%	169	0.0%	96	0.0%
CT surveillance (post-baseline)								
At 3 months	231,193	31.9%	192,783	26.6%	148,947	28.2%	136,721	25.9%
At 12 months	323,121	44.6%	264,410	36.5%	194,988	36.9%	178,158	33.8%
5-6 mm (straight to 12m CT)	93,742		73,333		47,099		42,243	
6+ mm (had 3m CT first)	229,379	71.0%	191,077	72.3%	147,889	75.8%	135,915	76.3%
At 24 months	322,755	44.5%	264,059	36.4%	194,800	36.9%	178,054	33.7%
Sub-solid nodules	168,159				57,100			
Discharged	**165,771**	**98.6%**	**164,578**	**97.9%**	**56,285**	**98.6%**	**56,006**	**98.1%**
At baseline	37,054	22.0%	48,388	28.8%	16,047	28.1%	20,237	35.4%
At 3 months	0	0.0%	0	0.0%	0	0.0%	0	0.0%
At 12 months	0	0.0%	0	0.0%	0	0.0%	0	0.0%
At 24 months	0	0.0%	0	0.0%	0	0.0%	0	0.0%
At 48 months	128,717	76.5%	116,190	69.1%	40,238	70.5%	35,769	62.6%

Sub-solid nodules	168,159				57,100			
Definitive management	**2,388**	**1.4%**	**3,581**	**2.1%**	**815**	**1.4%**	**1,094**	**1.9%**
At baseline	0	0.0%	0	0.0%	0	0.0%	0	0.0%
At 3 months	228	0.1%	1,926	1.1%	62	0.1%	386	0.7%
At 12 months	731	0.4%	766	0.5%	286	0.5%	303	0.5%
At 24 months	588	0.3%	437	0.3%	192	0.3%	204	0.4%
At 48 months	841	0.5%	452	0.3%	275	0.5%	201	0.4%
CT surveillance (post-baseline)								
At 3 months	131,105	78.0%	119,771	71.2%	41,053	71.9%	36,863	64.6%
At 12 months	130,877	77.8%	117,845	70.1%	40,991	71.8%	36,477	63.9%
At 24 months	130,146	77.4%	117,079	69.6%	40,705	71.3%	36,174	63.4%
At 48 months	129,558	77.0%	116,642	69.4%	40,513	71.0%	35,970	63.0%

(C) Proportions of people in different growth categories based on true nodule growth at each CT scan in CT surveillance in each of the 3 populations (symptomatic, incidental, screening)

Table 27. Proportions of people in different growth categories based on true nodule growth at each CT scan in CT surveillance in the symptomatic population

Symptomatic population								
	3 months		**12 months**		**24 months**		**48 months**	
Solid nodules	**205,875**		**309,265**		**309,069**		**NA**	
VDT <400 days	1,412	0.7%	195	0.1%	262	0.1%	NA	NA
VDT 400-600 days			12	0.0%	40	0.0%	NA	NA
VDT ≥ 600 days	204,463	99.3%	1	0.0%	28	0.0%	NA	NA
Stable diameter/volume			309,057	99.9%	308,739	99.9%	NA	NA
Sub-solid nodules	**139,208**		**139,132**		**138,720**		**138,708**	
Growth	76	0.1%	412	0.3%	12	0.0%	705	0.5%
Stable diameter	139,132	99.9%	138,720	99.7%	138,708	100.0%	138,003	99.5%

Table 28. Proportions of people in different growth categories based on true nodule growth at each CT scan in CT surveillance in the screening and incidental populations

Screening/incidental population								
	3 months		**12 months**		**24 months**		**48 months**	
Solid nodules	**136,721**		**178,158**		**178,054**		**NA**	
VDT <400 days	806	0.6%	102	0.1%	96	0.1%	NA	NA
VDT 400-600 days			19	0.0%	29	0.0%	NA	NA
VDT ≥ 600 days	135,915	99.4%	2	0.0%	28	0.0%	NA	NA
Stable diameter/volume			178,035	99.9%	177,901	99.9%	NA	NA
Sub-solid nodules	**42,404**		**42,381**		**42,221**		**42,211**	
Growth	23	0.1%	160	0.4%	10	0.0%	248	0.6%
Stable diameter	42,381	99.9%	42,221	99.6%	42,211	100.0%	41,963	99.4%

(D) Proportions of people assigned to different management options in CT surveillance in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening)

Table 29. Proportions of people assigned to different management options in CT surveillance in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening) – solid nodules

	Reader		Baseline		3 months		12 months		24 months		% of total
Symptomatic	Software-assisted	**Discharge**	362,889	50.0%	0	0.0%	0	0.0%	322,398	99.9%	44.5%
		Surveillance at next timepoint	324,935	44.9%	108,378	98.4%	322,755	99.9%	NA	NA	
		Definitive management	37,262	5.1%	1,814	1.6%	366	0.1%	357	0.1%	5.5%
	Unaided	**Discharge**	425,970	58.7%	0	0.0%	1	0.0%	263,708	99.9%	36.4%
		Surveillance at next timepoint	266,116	36.7%	89,049	98.1%	264,059	99.9%	NA	NA	
		Definitive management	33,000	4.6%	1,706	1.9%	350	0.1%	301	0.1%	4.9%
Screening / incidental	Software-assisted	**Discharge**	301,433	57.1%	0	0.0%	5	0.0%	194,631	99.9%	36.9%
		Surveillance at next timepoint	196,046	37.2%	51,571	98.0%	194,800	99.9%	NA	NA	
		Definitive management	30,260	5.7%	1,058	2.0%	183	0.1%	169	0.1%	6.0%
	Unaided	**Discharge**	320,105	60.7%	0	0.0%	2	0.0%	177,958	99.9%	33.7%
		Surveillance at next timepoint	196,964	33.9%	46,895	98.3%	178,054	99.9%	NA	NA	
		Definitive management	28,670	5.4%	806	1.7%	102	0.1%	96	0.1%	5.6%

Table 30. Proportions of people assigned to different management options in CT surveillance in the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening) – sub-solid nodules

			Baseline	3 months		12 months		24 months		48 months		% of total
Symptomatic	**Growth***	Software-assisted	131,105	228	0.2%	731	0.6%	588	0.5%	841	0.6%	1.8%
	Stable**			130,877	99.8%	130,146	99.4%	129,558	99.5%	128,717	99.4%	
	Growth*	Unaided	119,771	1,926	1.6%	766	0.7%	437	0.4%	452	0.4%	3.0%
	Stable**			117,845	98.4%	117,079	99.3%	116,642	99.6%	116,190	99.6%	
Screening / incidental	**Growth***	Software-assisted	41,053	62	0.2%	286	0.7%	192	0.5%	275	0.7%	2.0%
	Stable**			40,991	99.8%	40,705	99.3%	40,513	99.5%	40,238	99.3%	
	Growth*	Unaided	36,863	386	1.0%	303	0.8%	204	0.6%	201	0.6%	3.0%
	Stable**			36,477	99.0%	36,174	99.2%	35,970	99.4%	35,769	99.4%	

*Definitive management; **Further CT surveillance except at 48 months, when people are discharged.

(E) Proportions of people assigned to the different management option categories after initial detection and in CT surveillance getting the different investigations and treatment (table 46 on costs inputs used in the model, DAR pages 226-227) in each of the 3 populations (symptomatic, incidental, screening)

Table 31. Proportions of people assigned to the different management option categories after initial detection and in CT surveillance getting the different investigations and treatment - symptomatic population

Symptomatic	Baseline	Discharged correctly		Discharged incorrectly		3 month CT		12 month CT		24 month CT		48 month CT		MDT correct		MDT incorrect	
Software-assisted																	
Solid	725,086	643,289	88.7%	3,518	0.5%	231,193	31.9%	323,121	44.6%	322,755	44.5%	0	0.0%	14,453	2.0%	25,346	3.5%
Sub-solid	168,159	149,041	88.6%	2,895	1.7%	131,105	78.0%	130,877	77.8%	130,146	77.4%	129,558	77.0%	2,388	1.4%	0	0.0%
Unaided																	
Solid	725,086	654,132	90.2%	3,877	0.5%	192,783	26.6%	264,410	36.5%	264,059	36.4%	0	0.0%	14,027	1.9%	21,330	2.9%
Sub-solid	168,159	150,286	89.4%	1,702	1.0%	119,771	71.2%	117,845	70.1%	117,079	69.6%	116,642	69.4%	3,581	2.1%	0	0.0%

Table 32. Proportions of people assigned to the different management option categories after initial detection and in CT surveillance getting the different investigations and treatment – screening/incidental population

Screening	Baseline	Discharged correctly		Discharged incorrectly		3 month CT		12 month CT		24 month CT		48 month CT		MDT correct		MDT incorrect	
Software-assisted																	
Solid	527,739	469,368	88.9%	2,761	0.5%	148,947	28.2%	194,988	36.9%	194,800	36.9%	0	0.0%	12,131	2.3%	19,539	3.7%
Sub-solid	57,100	51,042	89.4%	799	1.4%	41,053	71.9%	40,991	71.8%	40,705	71.3%	40,513	71.0%	815	1.4%	0	0.0%
Unaided																	
Solid	527,739	472,805	89.6%	3,013	0.6%	136,721	25.9%	178,158	33.8%	178,054	33.7%	0	0.0%	11,623	2.2%	18,051	3.4%
Sub-solid	57,100	51,515	90.2%	520	0.9%	36,863	64.6%	36,477	63.9%	36,174	63.4%	35,970	63.0%	1,094	1.9%	0	0.0%

The EAG provides below additional information on the malignant status of nodules by type and size for people underwent the two strategies for the three populations, as this may assist the interpretation of findings.

Table 33. Malignant status of nodules by size for people underwent the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening) - solid nodules

	Symptomatic				Screening/incidental			
	Software-assisted		Unaided		Software-assisted		Unaided	
All participants	725,086		725,086		527,739		527,739	
Under 5mm	324,553	44.8%	394,461	54.4%	277,763	52.6%	298,384	56.5%
5-8mm	224,956	31.0%	180,812	24.9%	109,573	20.8%	98,797	18.7%
8 mm or more	175,577	24.2%	149,813	20.7%	140,403	26.6%	130,558	24.7%
Under 5mm	324,553	44.8%	394,461	54.4%	277,763	52.6%	298,384	56.5%
Undiagnosed benign	321,607	99.1%	390,881	99.1%	275,309	99.1%	295,742	99.1%
Undiagnosed malignant	2,946	0.9%	3,580	0.9%	2,454	0.9%	2,642	0.9%
5-8 mm	224,956	31.0%	180,812	24.9%	109,573	20.8%	98,797	18.7%
Clear features of benign	22,098	9.8%	17,778	9.8%	10,810	9.9%	9,804	9.9%
No clear features of benign	202,858	90.2%	163,034	90.2%	98,763	90.1%	88,993	90.1%
5-6 mm	93,742	46.2%	73,333	45.0%	47,099	47.7%	42,243	47.5%
Malignant	936	1.0%	743	1.0%	466	1.0%	408	1.0%
Benign	92,806	99.0%	72,590	99.0%	46,633	99.0%	41,835	99.0%
6-8 mm	109,116	53.8%	89,701	55.0%	51,664	52.3%	46,750	52.5%
Malignant	1,241	1.1%	1,068	1.2%	556	1.1%	542	1.2%
Benign	107,875	98.9%	88,633	98.8%	51,108	98.9%	46,208	98.8%
8 mm or greater	175,577	24.2%	149,813	20.7%	140,403	26.6%	130,558	24.7%
Clear features of benign	16,238	9.2%	13,731	9.2%	12,860	9.2%	11,917	9.1%
No clear features of benign	159,339	90.8%	136,082	90.8%	127,543	90.8%	118,641	90.9%
Malignant	12,992	8.2%	12,724	9.4%	11,686	9.2%	11,570	9.8%
Benign	146,347	91.8%	123,358	90.6%	115,857	90.8%	107,071	90.2%

Table 34. Malignant status of nodules by size for people underwent the 2 strategies (software-assisted, unaided) in each of the 3 populations (symptomatic, incidental, screening) – sub-solid nodules

	Symptomatic				Screening/incidental			
	Software-assisted		Unaided		Software-assisted		Unaided	
All participants	168159		168159		57,100		57,100	
Under 5mm	23219	13.8%	35798	21.3%	11,603	20.3%	16,266	28.5%
5mm or more	144940	86.2%	132361	78.7%	45,497	79.7%	40,834	71.5%
Under 5mm								
Undiagnosed benign	23124	99.6%	35238	98.4%	11,545	99.5%	16,119	99.1%
Undiagnosed malignant	95	0.4%	560	1.6%	58	0.5%	147	0.9%
5mm or more								
Clear features of benign	13835	9.5%	12590	9.5%	4,444	9.8%	3,971	9.7%
No clear features of benign	131105	90.5%	119771	90.5%	41,053	90.2%	36,863	90.3%
Malignant	5188	4.0%	4723	3.9%	1,556	3.8%	1,467	4.0%
Benign	125917	96.0%	115048	96.1%	39,497	96.2%	35,396	96.0%

Clarification of people who received multi-disciplinary team (MDT)/biopsy in the model

Question: Is it assumed that everyone having CT surveillance have MDT/biopsy as suggested on page 226 of the DAR? If so, could you do a scenario analysis where only a proportion of those on CT surveillance have MDT/biopsy (those that have VDT<= 400 days?).

EAG response: No. People under CT surveillance in the model only have MDT/biopsy if they have a VDT ≤ 400 days for solid nodules or show a sub-solid nodule growth of ≥ 2 mm.

Time horizon of the model

Question: Please clarify the time horizon used for the model.

EAG response: Cost-effectiveness was assessed over a 10-year time horizon. We chose this time horizon because we thought it would have been long enough to capture the costs incurred and benefits accrued across both strategies.

Starting age and smoking status of model cohort

Question: please clarify the staring age and smoking status of model cohort

EAG response: We assumed that people entering the model were aged 60 years. (p,32 of the DAR) We chose this starting age, which was in line with other previous cost-effectiveness analyses (e.g., Adams et al., 2021; Deppen et al., 2014 and Sutton et al., 2020) We did not specifically model people's smoking status.

We increased the risk of mortality by 1.3 compared to general population mortality for the symptomatic and the screening populations, as we expected that majority would be current smokers. Conversely, for the incidental population, we assumed that majority would not be smokers as can be seen in Zhou et al. (Zhou et al., 2023); hence, we assumed general population mortality.

References:

Adams SJ, Mondal P, Penz E, Tyan CC, Lim H, Babyn P. Development and Cost Analysis of a Lung Nodule Management Strategy Combining Artificial Intelligence and Lung-RADS for Baseline Lung Cancer Screening. J Am Coll Radiol. 2021 May;18(5):741-751. doi: 10.1016/j.jacr.2020.11.014.

Deppen SA, Davis WT, Green EA, Rickman O, Aldrich MC, Fletcher S, Putnam JB Jr, Grogan EL. Cost-effectiveness of initial diagnostic strategies for pulmonary nodules presenting to thoracic surgeons. Ann Thorac Surg. 2014 Oct;98(4):1214-22. doi: 10.1016/j.athoracsur.2014.05.025. Epub 2014 Jul 31. PMID: 25087933; PMCID: PMC4186897.

Sutton AJ, Sagoo GS, Jackson L, Fisher M, Hamilton-Fairley G, Murray A, Hill A. Cost-effectiveness of a new autoantibody test added to Computed Tomography (CT) compared to CT surveillance alone in the diagnosis of lung cancer amongst patients with indeterminate pulmonary nodules. PLoS One 2020;15(9):e0237492. doi: 10.1371/journal.pone.0237492.

Zhou N, Deng J, Faltermeier C, Peng T, Mandl H, Revels S's, et al. The Majority of Patients with Resectable Incidental Lung Cancers are Ineligible for Lung Cancer Screening, JTCVS Open 2023, doi: https://doi.org/10.1016/j.xjon.2022.11.021.

Assumptions associated with proportion of patients with stable nodule size discharged at one year or two years

Question: Could you please clarify why the percentages 95% and 5% were used in this assumption: 'For the AI-assisted reading strategy, we assumed that 95% of people with benign nodules would be discharged at the one-year CT surveillance and 5% would be discharged at the two-year CT surveillance. For the unaided reading strategy, we assumed that 95% of people would be discharged at the two-year CT surveillance and 5% at the one-year CT surveillance.'

EAG response: The statement should have read: 'For the AI-assisted reading strategy, we assumed that 95% of people with **stable** nodules would be discharged at the one-year CT surveillance and 5% would be discharged at the two-year CT surveillance. For the unaided reading strategy, we assumed that 95% of people would be discharged at the two-year CT surveillance and 5% at the one-year CT surveillance.'

This assumption was related to the BTS guideline which recommends that people with a nodule found to be stable at one-year CT surveillance based on volumetry can be discharged, whereas those with a nodule found to be stable at one-year based on 2D diameter values should be followed up again at two years. We assumed that by default AI-assisted reading would provide volumetry but allowed a small proportion (5%) of cases to be measured based on diameter due to technical failure related to AI software. On the other hand, as our clinical expert and participants of the scoping workshop suggested that software (whether involving AI or not) providing volumetry measurements is still uncommon in UK hospitals, we assumed that the vast majority of unaided readings would be based on diameter measurement and only a small proportion (5%) would be measured using volumetry.

Assumption associated with cancer status of nodules during CT surveillance and scenario analysis related to length of CT surveillance

Question: Could you also clarify how during CT surveillance it was determined someone had a benign nodule? And why the 2 and 4 year surveillance lengths were used when testing the effect of the above assumption in the scenario analysis?

EAG response: During CT surveillance, a nodule was assumed to be benign in the model if the nodule was not found to have a VDT < 400 days for a solid nodule or to show a growth of ≥ 2mm for a sub-solid nodule during any surveillance CT scans.

The scenario analysis assuming both strategies (AI-assisted and unaided reading) discharge patients at 2 years for solid nodules and at 4 years for sub-solid nodules (in line with BTS guideline for measurement based on diameter) was carried out to evaluate the impact of (removal of) early discharge conferred by volumetry measurement, which was assumed to be available for 95% of AI-assisted reading and only 5% of unaided reading as explained in Point 9 above.

Ethnicity of populations of the included studies

Question: What did you find out about ethnicity of the populations of the studies included in the review?

EAG response: Ethnicity of the study participants was reported in only two of the 27 included studies: one US study (Chamberlin et al 2021) reported 67% (78/117) of the participants being Caucasian; another UK study (Hall et al. 2022) reported white 84% (628/751), black 10% (74/751) and other 7% (49/751) among the study participants. However, most of the included studies were likely to have fairly homogeneous populations with respective to ethnicity: predominantly white for studies (total n= 16) conducted in the USA (n=7), Western Europe (n=8) and Russia (n=1); predominantly Eastern Asian (total n=11) for studies conducted in China (n=2), Japan (n=2), South Korea (n=5) and Taiwan (n=2). No outcomes stratified by ethnic groups of the study participants were reported among included studies.

Comparative per-person accuracy data and technical failure rate data

Question: Could you please provide a summary of your findings for all the included studies that reported comparative per-person accuracy data or technical failure data or both (except those where software was used as a standalone intervention)?

EAG response: A summary of comparative per-person accuracy data is presented below. No comparative technical failure data were reported; we present technical failure data related to AI-assisted reading.

Table 35. Studies reporting comparative accuracy data using per person analysis (n=6)

Study	Measures	Concurrent AI	Unaided reader	Ethnicity	Key risk of bias and applicability concerns
Hsu 2021, Taiwan Screening ClearRead CT Detection of any nodules	*All readers* Sensitivity * Specificity Sensitivity * Specificity *Junior readers* Sensitivity * Specificity *Senior readers* [b] Sensitivity * Specificity	0.79 (0.76 to 0.81) 0.81 (0.78 to 0.84) 0.80 (0.77 to 0.83) [a] 0.82 (0.79 to 0.84) [a] 0.74 (0.70 to 0.78) 0.74 (0.70 to 0.78) 0.83 (0.79 to 0.86) 0.88 (0.85 to 0.91)	0.63 (0.59 to 0.66) 0.77 (0.74 to 0.80) 0.52 (0.47 to 0.57) 0.68 (0.64 to 0.73) 0.73 (0.69 to 0.77) 0.86 (0.83 to 0.90)	NR (1 hospital in Taiwan)	Taiwan, 1 hospital; Mean age 64 +/- 8 years; 31/57 never smoked; 12 nodules were <3 mm (micronodules); Nodules >10 mm excluded; 2.5 mm section thickness; MRMC study (research setting); 3 radiology residents (junior group) and 3 experienced chest radiologists (senior group); Detection of any nodules ≤10 mm; Per-nodule sensitivity? Reference standard: Consensus reading of 2 experienced chest radiologists unblinded to index test results.
Hsu 2021, Taiwan Mixed ClearRead CT Detection of any nodules	*All readers* Sensitivity * Specificity Sensitivity * Specificity	0.80 (0.79 to 0.82) 0.83 (0.82 to 0.85)	0.64 (0.62 to 0.66) 0.80 (0.78 to 0.81)	NR (1 hospital in Taiwan)	As above, but mixed indication (93 clinical routine, 57 screening)

Study	Measures	Concurrent AI	Unaided reader	Ethnicity	Key risk of bias and applicability concerns
		0.82 (0.80 to 0.84) [a] 0.84 (0.82 to 0.85) [a]			
Zhang 2021, China Screening InferRead CT Lung Detection of any nodules	Sensitivity Specificity	0.99 (0.97 to 1.00) 0.97 (0.95 to 0.98)	0.43 (0.38 to 0.49) 1.00 (0.99 to 1.00)	NR (1 hospital in China)	China, 1 hospital; LDCT screening in 45-74-year olds (age not 55-75 years as in the UK); Concurrent AI: MRMC study (1 resident with supervision by 1 experienced radiologist); Unaided reading: Clinical practice (a total of 14 residents and 15 radiologists); Detection of any nodules; Reference standard: Consensus reading of 2 experienced chest radiologists unblinded to index test results.
Kozuka 2020, Japan Symptomatic InferRead CT Lung Detection of any nodules	Sensitivity Specificity	0.85 (0.80 to 0.90) 0.83 (0.52 to 0.98)	0.68 (0.61 to 0.74) 0.92 (0.62 to 1.00)	NR (1 hospital in Japan)	Japan, 1 hospital; Cases of pneumonia, diffuse lung disease, massive pleural efusion/atelectasis, and severe postoperative complications excluded; MRMC study (research setting) 2 less experienced radiologists; Detection of any nodules ≥3 mm.
Hall 2022, UK Screening Veolity Detection of actionable nodules	Sensitivity Specificity	Inexperienced radiographers 0.71 (0.65 to 0.76) 0.92 (0.91 to 0.94)	Experienced radiologist 0.91 (0.86 to 0.95) 0.97 (0.95 to 0.98)	UK (London) White: 83.2% Black: 9.9% Other: 6.8%	UK screening (London), but age between 60-75 years; Concurrent AI: MRMC study (2 inexperienced radiographers); Unaided reading: Clinical practice (experienced chest radiologists; 5% double reading); Reference standard: Nodules identified by unaided study radiologists, plus consensus radiologist review of any additional nodules identified by the software-assisted radiographers.

Study	Measures	Concurrent AI	Unaided reader	Ethnicity	Key risk of bias and applicability concerns
					Flow & Timing: Accuracy results for 682/770 (88.6%) for R1 and 706/770 (91.7%) for R2 and 716/770 (93.0%) for study radiologists.
Lo 2018, USA Screening ClearRead CT Detection of actionable nodules	Sensitivity * Specificity	0.73 (0.71 to 0.74) 0.84 (0.83 to 0.86)	0.60 (0.58 to 0.62) 0.90 (0.89 to 0.91)	NR (NLST database 84.3%, University Hospital Cleveland 13.9%, University of Maryland Hospital 1.9%)	USA (NLST and 2 hospitals); selected images with and without nodules in ratio 1:2 (case control); 3/178 nodules ≥30 mm; Pre-market version; 12 experienced general radiologists; MRMC study (research setting); Per-nodule sensitivity.
Lo 2018, USA Screening ClearRead CT Detection of malignant nodules	Sensitivity Specificity	0.800 (SD 0.039) 0.844 (SD 0.020)	0.647 (SD 0.039) 0.899 (SD 0.020)	NR (USA; NLST database 84.3%, University Hospital Cleveland 13.9%, University of Maryland Hospital 1.9%)	USA (NLST and 2 hospitals); selected images with and without nodules in ratio 1:2 (case control); 3/95 malignant nodules ≥30 mm; Pre-market version; 12 experienced general radiologists; MRMC study (research setting).
Park 2022, USA, Korea Screening VUNO Med-LungCT AI Detection of malignant nodules	Sensitivity	0.916 (0.817 to 0.964)	0.852 (0.742 to 0.920)	NR (USA; NLST)	USA (NLST dataset); Nodule- and cancer-enriched; MRMC study performed in Korea; 5 readers (1 resident and 4 radiologists with 1, 4, 8, and 20 years of experience in chest radiology); Sensitivity only! Reference standard for lung cancer presence/absence not reported.

MRMC, Multi-reader multi-case; NLST, National Lung Screening Trial; NR, Not reported.

[a] Second read AI; [b] Data used in EAG economic model for the base case of screening population. *The reported sensitivity was based on per nodule analysis.

Table 36. Studies reporting non-comparative test failure data (n=12)

Study	Measures	Concurrent AI	Stand-alone AI	Ethnicity	Population / Nodule characteristics / Slice thickness
Hwang 2021a, Korea Screening AVIEW Lungscreen	Failure of semi-automatic segmentation (clinical practice): All nodules	669/4,990 (13.4%)	NA	NR (Korea, 14 institutions)	K-LUCAS (Korea) 4,666 LDCT taken between April 2017 and March 2018; 4,686 (93.9%) solid 78 (1.6%) part-solid 226 (4.5%) pure ground glass. Non-enhanced CT, slice thickness < 1.5 mm.
Hwang 2021b, Korea Screening AVIEW Lungscreen	Failure of semi-automatic segmentation (clinical practice): All nodules Solid nodules Part-solid nodules Ground glass nodules	874/10,080 (8.7%) 688/9,465 (7.3%) 31/157 (19.7%) 155/458 (33.8%)	NA NA NA NA	NR (Korea, 14 institutions)	K-LUCAS (Korea) 10,424 LDCT taken between April 2017 and December 2018; 9,465 (93.9%) solid 157 (1.6%) part-solid 458 (4.5%) pure ground glass. Non-enhanced CT, slice thickness < 1.5 mm.
Hwang 2021c, Korea Screening AVIEW Lungscreen	Failure of semi-automatic segmentation: 20 radiologists from 14 institutions in clinical practice Central review (1 radiologist, MRMC study)	497/3,452 (14.4%) Range 0 to 57.0% (CV 1.28) 1.1% (107/9,389)	NA NA	NR (Korea, 14 institutions)	K-LUCAS (Korea) 3,353 LDCT taken between April 2017 and December 2017. Non-enhanced CT, slice thickness < 1.5 mm.
Singh 2021, USA Screening ClearRead CT	Software processing failure due to artifacts and/or thick slices (MRMC study)	NR	27/150 (18.0%)	NR (USA, NLST data)	NLST dataset (USA): 150 LDCT - first 125 patients with sub-solid nodules; first 25 patients with no nodules. Non-enhanced CT, slice thickness: 1.2–2 mm.

Study	Measures	Concurrent AI	Stand-alone AI	Ethnicity	Population / Nodule characteristics / Slice thickness
Jacobs 2021, USA, Denmark, Netherlands Screening Veolity	Need to manually tune segmentation parameters (MRMC study) Manual diameter measurement deemed necessary (MRMC study)	28% of nodule segmentations 2/160 (1.3%) nodules (2 readers) 0/160 nodules (4 readers)	NA NA	NR (USA, NLST data)	NLST dataset (USA): 160 LDCT selected by Lung-RADS category; 40 each for Lung-RADS 1 or 2; 3; 4A; 4B. Non-enhanced CT, slice thickness: 1.0 to 3.2 mm.
Hall 2022, UK Screening Veolity Detection of actionable nodules	Issues with the software (no software interpretation, software processing failure) (MRMC study)	R1: 9/770 (1.2%) R2: 18/770 (2.3%)	NA	UK (London) White: 83.2% Black: 9.9% Other: 6.8%	LSUT study (UK): All 770 LDCT with a lung health check appointment between November 2015 and July 2017; 158 with ≥1 nodule (≥5 mm or ≥80 mm^3). Non-enhanced CT, slice thickness: 0.5–1.0 mm; 2 radiographers without prior experience.
Cohen 2017, Korea Surveillance population with applicability concerns Veolity	Failure of semi-automatic segmentation (MRMC study): Sub-solid nodules - FBP Sub-solid nodules - MBIR Manual modifications of nodule segmentation required (MRMC study): Sub-solid nodules - FBP	7/73 (9.6%) 5/73 (6.8%) R1: 27/73 (37.0%) R2: 43/73 (58.9%) (median 35/73, 47.9%). R1: 21/73 (28.8%) R2: 39/73 (53.4%)	NA	NR (Korea, 1 hospital)	1 hospital in Seoul (Korea): 73 patients with preoperative CT scans for sub-solid nodules taken between July 2014 to May 2015; 73 sub-solid nodules. Non-enhanced CT, slice thickness 0.625 mm. Reconstructed with FBP and MBIR, respectively.

Study	Measures	Concurrent AI	Stand-alone AI	Ethnicity	Population / Nodule characteristics / Slice thickness
	Sub-solid nodules - MBIR	(median 30/73, 41.1%).			
Kim 2018, Korea Surveillance population with applicability concerns Veolity	Failure of semi-automatic segmentation (MRMC study): Sub-solid nodules	7/109 (6.4%)		NR (Korea, 1 hospital)	1 hospital in Seoul (Korea): 89 patients with preoperative CT scans for sub—solid nodules taken between November 2014 and July 2016; 109 sub-solid nodules. Non-enhanced CT, slice thickness 0.625 mm.
Roehrich 2022, Austria Mixed Contextflow SEARCH Lung CT	"Technical difficulties" (not further specified, MRMC study)	2/216 (0.9%)	NA	NR (Austria, 1 hospital)	1 hospital in Austria in 2018; first 100 patients with lung pathologies (22 unique, verified diagnoses, but none with lung nodules), first 8 patients without pathological lung findings. Slice thickness: 1 mm.
Hempel 2022, Netherlands Mixed Veye Chest	"Volumetry not deemed reliable" (MRMC study): Relevant nodules that contributed to the reader's management decision	R1: 1/41 (2.4%) R2: 2/44 (4.5%)	NA	NR (Netherlands, 1 hospital)	1 hospital in the Netherlands: 50 chest CT scans taken between July and September 2013 with ≤5 incidentally detected nodules (n=45: 35 with and 10 without prior imaging) or no nodules (n=5) on initial radiology report. Slice thickness: 2.00 mm (n=73) and 3.00 mm (n=12).
Martins Jarnalo 2021, Netherlands Mixed Veye Chest	Failure of semi-automatic segmentation (retrospective study): All 80 nodules correctly detected by stand-alone software	NA	3/80 (3.8%)	NR (Netherlands, 1 hospital)	1 hospital in the Netherlands: Random 145 chest CT scans performed for various indications between December 2018 and May 2020; 91 nodules: 16 sub-solid nodules, 73 solid nodules, 2 mixture of solid/sub-solid. Slice thickness: 1 or 3 mm.

Study	Measures	Concurrent AI	Stand-alone AI	Ethnicity	Population / Nodule characteristics / Slice thickness
Murchison 2022, UK Mixed, Veye Chest	Failure of semi-automatic segmentation (MRMC study): 428 nodules (3-30 mm) from groups [1], [2], [3] and [5]	NA	21/428 (4.9%)	NR (UK, 1 hospital)	1 hospital in Edinburgh (UK): 337 scans of 314 current smokers, ex-smokers and/or those with radiological emphysema between 55-74 years taken between January 2008 and December 2009. [1] 178 without reported nodules; [2] 95 with 1-10 reported nodules; 23 CT images from the same patients with [3] baseline CT scan and [4] follow-up CT scan; [5] 18 with sub-solid nodules. Slice thickness 1.0-2.5mm.

CT, Computed tomography; FBP, Filtered back projection; K-LUCAS, Korean Lung Cancer Screening Project; LDCT, Low-dose computed tomography; MBIR, Model-based iterative reconstruction; NA, Not applicable; NLST, National Lung Screening Trial; NR, Not reported; R1, Reader 1; R2, Reader 2.