Evidence overview: Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke

This overview summarises the main issues the diagnostics advisory committee needs to consider. It should be read together with the final scope and the diagnostics assessment report.

1 Aims and scope

Software with artificial intelligence (AI)-derived algorithms can be used to analyse CT brain scan images from people with acute suspected stroke to detect and report irregularities. The result of this analysis is intended to support the scan review and reporting by a trained healthcare professional.

By identifying, quantifying and highlighting stroke-related changes in the brain, the AI-derived algorithms may assist in confirming a stroke and support clinical decisions about suitability of an appropriate time-sensitive treatment such as thrombolysis and thrombectomy. Using the software in the radiology pathway may lead to quicker review of scans by a multi-site clinical team, improved decisions about treatment, expedited patient transfer, faster access to the correct treatment and improved patient outcomes. Some software has features that can prioritise the review of stroke CT scans.

Decision questions

- Decision question 1: Does software-assisted review of non-enhanced CT brain scans for guiding thrombolysis treatment decisions for people with suspected acute stroke represent a clinically- and cost-effective use of NHS resources?
- Decision question 2a: Does software-assisted review of CT angiography brain scans for guiding mechanical thrombectomy treatment decisions for

people with an ischaemic stroke represent a clinically- and cost-effective use of NHS resources?

- Decision question 2b: Does software-assisted review of CT perfusion brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke after a CT angiography brain scan represent a clinically- and cost-effective use of NHS resources?

Populations

A separate population is considered for each decision question:

- Decision question 1: People referred to or attending secondary care with a suspected acute stroke who were well in the last 24 hours
- Decision question 2a: People with an ischaemic stroke who were well in the last 6 hours
- Decision question 2b: People with an ischaemic stroke who were last known to be well in the last 6 to 24 hours who have already had a CT angiography brain scan to confirm ischaemic stroke.

The assessment focused on people aged 18 or over. Depending on the availability of evidence, the following subpopulation may be considered: people over the age of 80 with small vessel disease and calcification of the cerebrovasculature.

Interventions

AI-derived software-assisted CT brain scan review by a healthcare professional using any of the following software or platforms:

- icobrain ct (icometrix)
- Aidoc (Aidoc)
- Aidoc + icobrain (Aidoc and icometrix)
- RapidAI (Ischemaview)
- e-Stroke (Brainomix)

- Viz (Viz.ai)
- qER (Qure.ai)
- Zebra-Med (Zebra Medical Vision)
- CT Perfusion 4D (GE Healthcare)
- Brainscan CT (Brainscan.ai)
- Cercare stroke (Cercare Medical)
- CINA head (Avicenna)
- Accipio (MaxQ AI)
- Biomind (Biomind.ai)

Not all software assesses non-enhanced CT, CT angiography and CT perfusion brain scans, so they are not assessed in all of the decision questions. Further details are in table 2 on pages 17 to 20 of the final scope.

Comparator

The comparator is CT brain scan review by a healthcare professional without assistance from AI-derived software.

Non-enhanced CT brain scans may be reviewed by a radiologist, specialist radiologist, radiographer, stroke physician or emergency medicine physician, depending on availability of staff. CT angiography and CT perfusion brain scans are more likely to be reviewed by a radiologist, neuroradiologist or an interventional neuroradiologist.

Healthcare setting

- Comprehensive stroke centre
- Acute stroke centre

Further details, including descriptions of the interventions, comparator, care pathway and outcomes, are in the final scope for software with artificial intelligence derived algorithms for analysing CT brain scans in people with a suspected acute stroke.

2 Clinical effectiveness evidence

The external assessment group (EAG) did a systematic review to identify evidence on the clinical effectiveness and diagnostic accuracy of software with artificial intelligence (AI)-derived algorithms for analysing CT brain scans in people with a suspected acute stroke. Find the methods and results on pages 39 to 81 of the diagnostics assessment report.

Overview of included studies

There were 22 studies reported in 30 publications that met the inclusion criteria for the systematic review. Five studies were published as conference abstracts only and 2 studies had not yet been peer reviewed.

Of the included studies, 15 were diagnostic accuracy studies and 7 were observational 'before and after' studies. 21 of the studies were retrospective, and in 1 study it was not clear whether recruitment was retrospective or prospective. [redacted] Most studies were done in US. Three studies did not report the study location.

Of the 22 studies, 8 evaluated software on the RapidAI, 9 on the Viz, 4 on the e-Stroke and 1 software on the CINA head platform. No study reported data for more than 1 platform. No eligible studies were found that assessed the CT Perfusion 4D, Biomind, icobrain ct, Aidoc, icobrain + Aidoc, Brainscan CT, Zebra-Med, Cercare stroke, qER or Accipio.

Of the included studies, 4 provided some evidence on the software use for the review of non-enhanced CT brain scans (decision question 1), 18 for the review of CT angiography brain scans (decision question 2a) and 3 for the review of CT perfusion brain scans (decision question 2b). No studies separately reported outcomes in people aged 80 or over who have a small vessel disease and calcification of the cerebrovasculature. Find an overview

of the included studies in tables 3 and 4 on pages 51 to 54 of the diagnostics assessment report.

Study quality

The EAG assessed the quality of the diagnostic accuracy studies using the QUADAS-2 tool. Because of unclear reporting, the risk of bias in the studies was often unclear. In all studies there was high concern over the applicability of the index test to the questions in this assessment. This was because none of the diagnostic accuracy studies evaluated the software for assisting a clinician's' review of the CT brain scans (as per intended use). Instead, they reported using the software as a standalone intervention. Find a summary of the QUADAS-2 assessment in table 5 on page 57 of the diagnostics assessment report.

To assess the quality of the observational 'before and after' studies, the EAG used a checklist developed for this assessment. It found that the reporting of the studies was generally unclear. The primary outcome in all studies was time to treatment but not all studies reported this together with clinical outcomes. Only 3 of the 7 studies (Adhya et al. 2021, Gunda et al. 2020 and Hassan et al. 2020) reported there were no other changes to the care pathway (apart from software implementation) between the 2 time periods assessed. Most studies did not report enough information to assess whether the study participants were comparable before and after the implementation of the software. Only 2 of studies (Al-Kawaz et al. 2021 and Hassan et al. 2020) specified the point in the care pathway where the software was used, and who reviewed the findings. No study described what CT imaging criteria was used to guide decisions on whether people should be offered thrombolysis and thrombectomy.

Decision question 1

The EAG identified 4 studies on software use for the review of non-enhanced CT brain scans that were relevant to decision question 1. Only 1 of these

studies, a manuscript (Mair et al. 2021) which has yet to be peer reviewed, was not a conference abstract.

Intermediate outcomes in people with a suspected acute stroke having a non-enhanced CT brain scan

Test accuracy

Table 1 summarises the accuracy estimates reported in 3 studies that evaluated the diagnostic performance of AI-derived software technologies to detect intracerebral haemorrhage in non-enhanced CT brain scans from people with suspected acute stroke. [redacted]. Find more details in table 7 on page 62 of the diagnostics assessment report.

Table 1 Accuracy to detect intracerebral haemorrhage in non-enhanced CT brain scans in suspected acute stroke

Study	Software	Study size	Sensitivity (95% confidence interval)	Specificity (95% confidence interval)
Barriera et al. (2018)	Viz ICH	284	90.2 (83.9 to 94.2)	100.0 (97.5 to 100.0)
Herweh et al. (2020)	Brainomix (unspecified)	160	91.1 (82.8 to 95.6)	88.9 (80.2 to 94.0)
Mair et al. (2021)	Brainomix (e-ASPECTS)	[redacted]	[redacted]	[redacted]

Effect on time to treatment

One ‘before and after’ study (Gunda et al. 2020) evaluated the effects on time to treatment in a centre where people who needed thrombectomy were

transferred to another unit. It looked at using Brainomix e-ASPECTS (relevant to decision question 1) to assist review of non-enhanced CT scans, and Brainomix e-CTA (relevant to decision question 2a) to assist the review of CT angiography scans. The results were not reported separately for e-ASPECTS and e-CTA.

The mean time from door to thrombolysis treatment was 44 minutes before and 41 minutes after the software implementation. The mean time from the first CT brain scan (non-enhanced) to thrombectomy was 174 minutes before and 145 minutes after the software implementation. The proportion of people who had thrombolysis was 11.5% before the software was being used and 18.1% after. The proportion of people transferred for thrombectomy was 2.8% before and 4.8% after the software was used. The study did not report any clinical outcomes. Find more details in table 8 on page 59 of the diagnostics assessment report.

Clinical outcomes in people with a suspected acute stroke having a non-enhanced CT brain scan

No studies were identified.

Decision question 2a

Intermediate outcomes in people with ischaemic stroke having a CT angiography brain scan

Test accuracy

Eleven studies reporting enough information to calculate diagnostic performance for detecting large vessel occlusions in CT angiography brain scans in people with ischaemic stroke. Nearly all the studies limited evaluation of diagnostic performance to occlusion detection in anterior circulation. Only 1 study, on Viz LVO (Dornbos et al. 2020), included detection of occlusions in the posterior circulation. Most studies included detection of occlusions in

smaller branches of the arteries (M2 or M3 segments of the middle cerebral artery). Two studies, on Viz LVO, did not describe the type of large vessel occlusions to be detected (Shalitin et al. 2020 and Yahav-Dovrat et al. 2021).

Table 2 summarises the single study derived accuracy estimates for Rapid CTA, Rapid LVO, Brainomix e-CTA and Avicenna CINA LVO and the pooled estimate from EAG's meta-analysis of 5 studies for Viz LVO. When the Viz LVO study that included occlusions in posterior circulation was excluded (Dornbos et al. 2020), the summary sensitivity estimate was slightly higher (91.3%; 95% CI 84.9% to 95.1%) but specificity was similar (89.3%; 95% CI 83.5% to 93.2%). The EAG did not calculate summary estimates for Rapid CTA because populations in the studies partially overlapped, nor for Rapid LVO because definitions of large vessel occlusion were different.

Find more details of the accuracy estimates for Rapid CTA and Rapid LVO in table 9 on pages 66 to 67, for Viz LVO in table 11 on page 72 and figures 4 and 5 on page 71, for Brainomix e-CTA in table 13 on page 75, for Avicenna CINA LVO in table 15 on page 77 of the diagnostics assessment report.

Table 2 Accuracy to detect large vessel occlusion in CT angiography brain scans

Study	Software	Study size	Occlusions in smaller branches of arteries included	Sensitivity (95% confidence interval)	Specificity (95% confidence interval)
Amukotuwa et al. (2019a)	Rapid CTA	926	M2 segment of the middle cerebral artery (MCA)	95.4% (92.7% to 97.1%)	79.4% (75.8% to 82.6%)
Amukotuwa et al. (2019b; study population is a subset of the population in Amukotuwa et al. 2019a)	Rapid CTA	477	M2 segment of the MCA	91.5% (84.6% to 95.5%)	81.1% (76.8% to 84.8%)
Dehkharghani et al. (2021)	Rapid LVO	217	Not included	96.3% (90.9% to 98.6%)	98.1% (93.5% to 99.5%)
Paz et al. (2021)	Rapid LVO	151	M2 or M3 segments of the MCA	63.6% (51.6% to 74.2%)	85.9% (76.9% to 91.7%)
Pooled estimate from Barriera et al. (2018a), Chatterjee et al. (2018), Dornbos et al. (2020), Shalitin et al. (2020) and Yahav-Dovrat et al. (2021)	Viz LVO	5,320 (from 5 studies)	M2 segment of the MCA in 2 studies, occlusions in smaller vessels not included in 1 study, definition not provided in 2 studies	88.0% (76.9% to 94.2%)	89.9% (85.5% to 93.0%)
McLouth et al. (2021)	Avicenna CINA LVO	378	M2 segment of the MCA	98.1% (94.5% to 99.3%)	98.2% (95.5% to 99.3%)

Study	Software	Study size	Occlusions in smaller branches of arteries included	Sensitivity (95% confidence interval)	Specificity (95% confidence interval)
Seker et al. (2020)	Brainomix e-CTA	301	Proximal M2 segment of the MCA	83.8 % (77.3% to 88.7%)	95.7% (91.0% to 98.0%)

Comparative test accuracy

Seker et al. (2020), as well as evaluating the diagnostic performance of Brainomix e-CTA, provided some comparative accuracy data for the CT brain scan review done by clinicians alone in a subset of the study population (144 people). The reference standard in this study was a neuroradiologist with more than 10 years of experience, who had access to both the imaging and clinical data (including data on interventional therapy and follow up). Table 3 summarises the accuracy estimates for the clinicians with different amounts of experience. The software, with sensitivity of 84.3% (95% CI 74.0% to 91.0%), missed more large vessel occlusions than any of the clinicians. The comparative specificity was more varied. Find more details in table 14 on page 73 of the diagnostics assessment report.

Table 3 Clinicians' accuracy to detect large vessel occlusion in CT angiography brain scans from Seker et al. (2020)

CT scan reviewer	Sensitivity (95% confidence interval)	Specificity (95% confidence interval)
Neuroradiologist	97.1% (90.2% to 99.2%)	98.6% (92.7% to 99.8%)
Radiology resident	95.7% (88.1% to 98.5%)	91.9% (83.4% to 96.2%)
Neurology resident 1	85.7% (75.7% to 92.1%)	90.5% (81.7% to 95.3%)
Neurology resident 2	91.4% (82.5% to 96.0%)	100% (95.1% to 100%)

Effect on time to treatment

In addition to the study that reported effects on time from non-enhanced CT brain scan to thrombectomy for both Brainomix e-CTA and e-ASPECTS together (see page 8; Gunda et al. 2020), there were 5 ‘before and after’ studies that evaluated effects of implementing software in clinical practice. In 3 studies the effects were assessed at centres where transfer was needed for thrombectomy (Dornbos et al. 2020, Hassan et al. 2020 and Morey et al. 2020a). In 1 study the setting was unclear (Adhya et al. 2021). Table 4 summarises effects on time to treatment. In all the studies, time from hospital entry or CT angiography to groin puncture in thrombectomy reduced after software was adopted. Find more details in table 10 on page 68 and table 12 on page 73 of the diagnostics assessment report.

Table 4 Effects of software-assisted review of CT angiography brain scans on time to thrombectomy and clinical outcomes after thrombectomy

Study	Software	Study size before and after software implementation	Time to thrombectomy before and after software implementation	Odds ratio (95% CI) 90-day mRS score ≤ 2 before versus after	Mean difference (95% CI) 90-day mRS score	Mean difference (95% CI) in days in hospital	Odds ratio (95% CI) complications in hospital	Odds ratio (95% CI) death in hospital
Adhya et al. (2021)	Rapid CTA	74 and 72 (67 for clinical outcomes)	92 to 68 (mean, no variance reported)	1.75 (0.84 to 3.67)	Not reported (mean mRS 4.47 before and 3.9 after software implementation)	-	-	-
Dornbos et al. (2020)	Viz LVO	Not reported	185 to 141 (median, reduction p=0.027)	-	-	-	-	-
Hassan et al. (2020)	Viz LVO	28 and 11	216 to 127 (median, reduction p=0.026)	1.67 (0.45 to 6.23)	-	-2.5 (-4.7 to -0.3)	0.60 (0.06 to 6.28)	1.33 (0.31 to 5.73)
Hassan et al. (2021a)	Viz LVO	86 and 102	206.6 to 119.9 (mean difference -86.7, 95% CI -125.9 to -47.5)	0.88 (0.46 to 1.69)	-	7.0 before and 7.5 after (no significant increase)	0.87 (0.46 to 1.62)	1.10 (0.55 to 2.21)
Morey et al. (2020a)	Viz LVO	29 and 26	161.3 to 146.7 (mean difference -44.6, 95% CI -68.6 to -20.6)	-	-1.0 (-2.1 to 0.1)	-	-	-

Clinical outcomes in people with ischaemic stroke having a CT angiography brain scan

Of the 5 studies that reported reduced time to thrombectomy after implementation of software specifically for assisting CT angiography brain scan review, 4 also reported effects of software use on clinical outcomes after thrombectomy. Table 4 summarises effects reported. The studies reported positive and negative effects on the proportion of people who were functionally more independent (with modified Rankin Scale [mRS] score 2 or less) and length of hospital stay, reduction in the mean 90-day mRS score and lower rate of complications but higher rate of death during hospital stay after software implementation. None of the effects were statistically significant; statistical power was not reported but the EAG advise that the studies were unlikely to have been powered to detect a difference given the study designs used and the relatively small sample sizes

The EAG highlighted that none of the studies provided information about outcomes for people who did not have a thrombectomy. No information was provided about the performance of Viz LVO (if used the way it would be in clinical practice) for identifying people with a large vessel occlusion eligible for thrombectomy. The Rapid CTA study evaluated performance of the software in the context of providing an automated alert system.

Decision question 2b

Intermediate outcomes in people with a large vessel occlusion having a CT perfusion brain scan

Test accuracy

Kauw et al. (2020) used Rapid CTP to analyse retrospective data from 176 people. To calculate accuracy, the EAG used the treatment that people had as the reference standard, that is whether they were treated for a large

vessel occlusion or not. This gave a sensitivity of 95.2% (95% CI 90.0% to 97.8%) and specificity of 80.0% (95% CI 67.0% to 88.8%). This study also reported 11% failure rate in Rapid CTP image processing. Find more details in table 16 on page 80 of the diagnostics assessment report.

Effect on time to treatment

There were 2 ‘before and after’ studies that evaluated effects on time to treatment of implementing AI to assist review of CT perfusion brain scans. Kamal et al. (2017), a conference abstract, used an unspecified Rapid technology and Al-Kawaz et al. (2021) used RapidAI Mobile Application to analyse both CT angiography and CT perfusion scans. The studies reported effect on time to treatment for both algorithms together. Al-Kawaz et al. (2021) reported that the mean time from hospital entry to groin puncture in thrombectomy reduced by 33.2 minutes (95% CI -60.2 minutes to -6.2 minutes) after the software was adopted in a centre that did thrombectomies. Kamal et al. (2017) reported a slight increase (mean difference 2.0 minutes [95% CI -12.9 minutes to 16.9 minutes]) but this was not statistically significant. The setting was not clear. Find more details in table 17 on page 80 of the diagnostics assessment report.

Clinical outcomes in people with a large vessel occlusion having a CT perfusion brain scan

Al-Kawaz et al. (2021) found that mean 90-day mRS for people who had a thrombectomy was 2.9 both before and after software implementation. Kamal et al. (2017) reported that the proportion of people with an mRS score of 3 or less increased after software implementation (odds ratio 1.34, 95% CI 0.66 to 2.74) but this effect was not statistically significant. Size of the study population was a lot smaller after the implementation (41 people of whom 23 had an mRS score of 3 or less) compared with before the implementation (119 people of whom 58 had an mRS score of 3 or less). Find more details in table 17 on page 80 of the diagnostics assessment report.

Health-related quality of life outcomes

No studies were found on the effects of software-assisted review on health-related quality of life outcomes for any of the decision questions.

3 Cost effectiveness evidence

The EAG did a systematic review to identify any published economic evaluations of software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke. Find the full systematic review results on pages 86 to 89 of the diagnostics assessment report. The EAG also constructed a de novo economic model to assess the cost effectiveness of software-assisted review of CT angiography brain scans for guiding mechanical thrombectomy treatment decisions in ischaemic stroke (decision question 2a).

Systematic review of cost-effectiveness evidence

One study met inclusion criteria for the review. Van Leeuwen et al. (2021) used a decision tree combined with a lifetime state transition model to estimate potential cost effectiveness of using AI software aided detection of large vessel occlusions on CT angiography brain scans in ischaemic stroke. The comparator was clinician's CT brain scan review without the software. Analyses were done from a societal UK perspective.

Economic analysis

The EAG developed a de novo economic model to assess cost effectiveness of software-assisted review of CT angiography brain scans for guiding mechanical thrombectomy treatment decisions in ischaemic stroke (decision question 2a). The model was based on the cost-effectiveness analysis by van Leeuwen et al. (2021). The EAG commented that the assessment primarily considers the question of whether AI-derived software-assisted CT brain scan review could more accurately detect large vessel occlusions (in the proximal anterior circulation) than clinician's review alone.

Models were not developed to assess the cost effectiveness of software-assisted CT brain scan review when used as in decision questions 1 and 2b, because the EAG judged that the systematic reviews did not identify enough evidence to support their modelling.

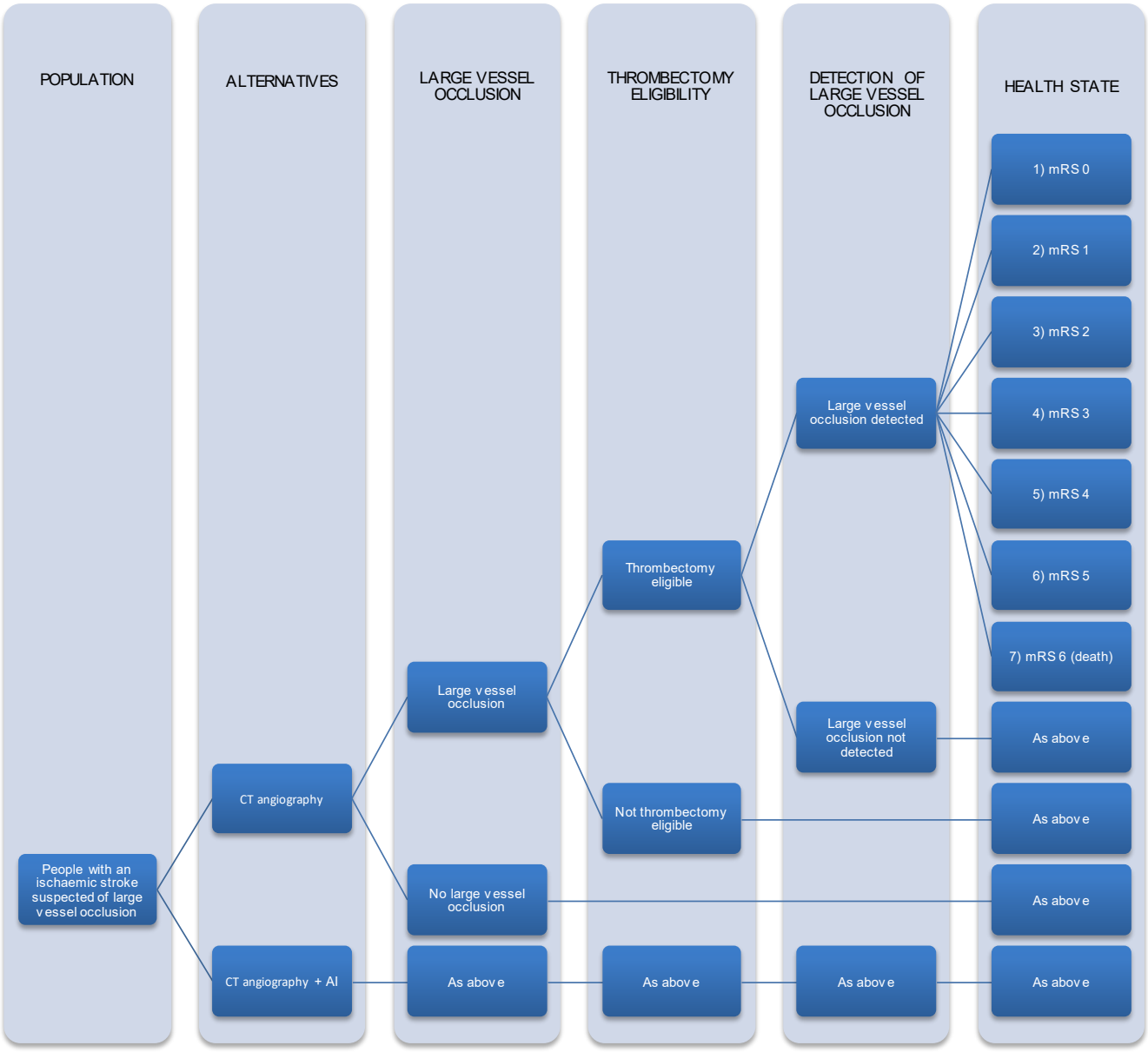
Model structure

The model had 2 parts:

- a short-term decision analytic tree that captured diagnostic pathway and initial costs (in first 90 days after stroke) and
- a long-term state transition model that estimated lifetime costs and quality-adjusted life years (QALYs).

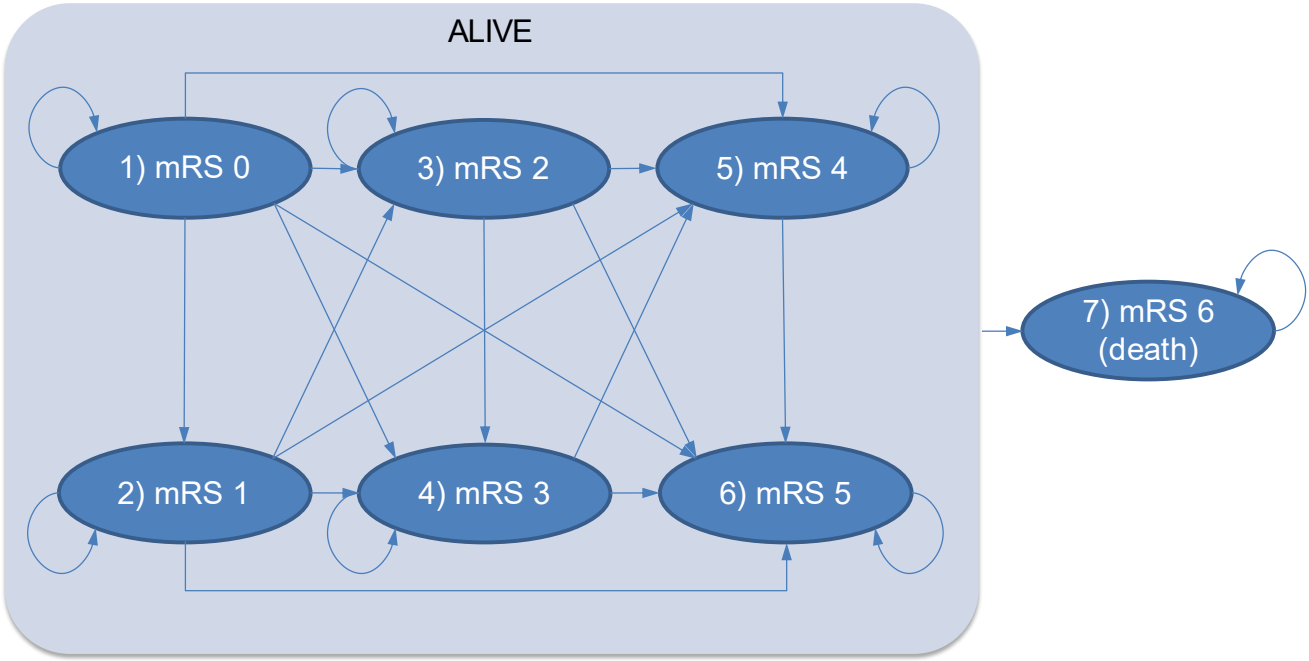
Figure 1 presents the structure of the short-term decision analytic model. People with ischaemic stroke are assessed for large vessel occlusion by a clinician's review of the CT angiography brain scan (comparator) or using the software to help the clinician's scan review (intervention). These diagnostic strategies have different diagnostic accuracy estimates. If a large vessel occlusion is detected, thrombectomy may be suitable. Whether or not people for whom a thrombectomy is suitable actually go on to have one depends on the sensitivity to detect the large vessel occlusion. Depending on the specificity, a large vessel occlusion could also be falsely detected in people who do not have one. But it is assumed that specialist clinicians check findings before doing thrombectomies, so false positive findings from the initial reviews of CT angiography brain scans have only cost consequences in the model. People who do not need a thrombectomy will not have one. Find more details on pages 90 and 91 of the diagnostics assessment report.

Figure 1 Short-term decision analytic model



At the end of the short-term model, people enter the model that represents the rest of their lifetime. Figure 2 shows the structure of this subsequent model. This model includes 7 health states based on mRS scores describing the level of disability and functional independence. The scale runs from perfect health without symptoms (mRS 0) to death (mRS 6). Cycle time in the model is 1 year. After each cycle, people can remain in the same health state, die, or if they have a recurrent stroke, move to a worse health state.

Figure 2 Long-term state transition model



Population

The modelling was done for people with ischaemic stroke and suspected large vessel occlusion. The starting age of the cohort was 66 years.

Comparator

The comparator was CT angiography brain scan review by a healthcare professional without assistance from AI-derived software.

Model inputs

Find the full list of model parameters in table 28 on pages 106 to 108 of the diagnostics assessment report.

Proportion of people with a large vessel occlusion

The modelled proportion of people with a large vessel inclusion was 46.1% (95% CI 43.0% to 49.1%). This was estimated by pooling the prevalence of

large vessel occlusions in the 12 diagnostic accuracy studies identified in the systematic review that did not use a case-control study design.

Proportion of people with a large vessel occlusion eligible for thrombectomy

The proportion of people with a large vessel occlusion eligible for a thrombectomy in the base case was 41.2% (95% CI 40.6% to 41.8%). This was based on a registry-based study by McMeekin et al. (2017) and includes people who presented within 4 hours of symptom onset, people who presented later and people for whom the timing of presentation was unknown. This was the same for people assessed with and without software-assisted review. In a scenario analysis, the EAG increased the proportion of people reviewed by the software who are eligible for thrombectomy to 50% (scenario 2). This potentially reflects use of the software detecting large vessel occlusions earlier, with more people with large vessel occlusion therefore eligible for thrombectomy.

Accuracy to detect large vessel occlusion

In the absence of accuracy data for software used together with clinician review, the diagnostic accuracy estimates for both the intervention and the comparator in the base case were elicited from 5 clinical experts. To inform the estimates, the EAG gave the experts accuracy estimates for each AI-derived software alone from identified studies that best matched the scope of the model. The software accuracy studies were Amukotuva et al. (2019a) for Rapid CTA, Seker et al. (2020) for Brainomix e-CTA and McLouth et al. (2021) for Avicenna CINA LVO, shown in table 2. For Viz LVO, the study given was by Chatterjee et al. (2018) that reported sensitivity of 91.2% (95% CI 77.0% to 97.0%) and specificity of 85.0% (95% CI 64.0% to 94.8%). The experts also received accuracy estimates for clinicians' scan review (shown in table 3) that were available from Seker et al. (2020). One set of accuracy estimates was elicited for software-assisted CT brain scan review and another for the clinician's review. Table 5 shows the expert elicitation-based sensitivity

and specificity estimates used in the base case). Find more details on pages 82 to 82 and 92 to 95 of the diagnostics assessment report.

Table 5 Expert elicited diagnostic accuracy to detect large vessel occlusion in CT angiography scans used in base case model

Strategy	Sensitivity	Specificity
Clinician's CT angiography review (comparator)	93.00	94.09
Clinician's AI software-assisted CT angiography review (intervention)	94.13	93.77

Initial distribution of people across the health states

Table 6 shows the proportion of people assigned to different mRS health states based on whether they had a large vessel occlusion and thrombectomy. Proportions for people with a large vessel occlusion were calculated from data in Román et al. (2018), a meta-analysis of individual patient-level data from 7 randomised controlled trials to estimate mRS scores 90 days after stroke. The proportions for people without a large vessel occlusion were from a prospective study of 15 South Korean stroke centres (Paek et al. 2019). Find more details on pages 96 and 97 of the diagnostics assessment report.

Table 6 Initial distribution of people across the health states

Subpopulation	mRS 0	mRS 1	mRS 2	mRS 3	mRS 4	mRS 5	mRS 6
People with a large vessel occlusion who had a thrombectomy	11.1%	18.1%	18.6%	16.1%	15.9%	5.5%	14.7%
People with a large vessel occlusion who did not have a thrombectomy	6.3%	9.8%	14.2%	16.2%	25.1%	10.9%	17.5%
People who did not have a large vessel occlusion	21.8%	35.2%	23.8%	12.4%	4.7%	2.1%	0.0%

Rate of recurrent stroke

The annual rate for recurrent stroke after ischaemic stroke in the modelled base case was 2.8%, based on Pennlert et al. (2014). This study estimated sex- and age-adjusted annual risk of recurrent stroke at 28 days after an ischaemic stroke in the Swedish population-based Monitoring Trends and Determinants of Cardiovascular Disease (MONICA) stroke incidence registry between 1995 and 1998. Risk was constant over time, and the same across the mRS health states. People could have no more than 1 recurrent stroke a year. An alternative risk of recurrent stroke of 2.01% was used in a scenario analysis (from Mohan et al. 2011). Find more details on page 98 of the diagnostics assessment report.

Mortality

People who were more functionally dependent (people in health states mRS 2 to mRS 5) had a higher risk of death than people who were less functionally dependent (people in health states mRS 0 and mRS 1). Relative risk of stroke-related death increased by mRS health state, up to 2.57 for mRS 5 (Slot et al. 2009). This study used data from the Oxfordshire Community Stroke Project, the Lothian Stroke Register and the First International Stroke Trial in the UK. Find more details on pages 98 and 99 of the diagnostics assessment report.

Costs

Find the full list of costs used in the model in table 28 on pages 106 to 108 of the diagnostics assessment report.

Cost of software

The base case used 1 overall cost for the AI-derived software. To calculate this, the annual license fees for the 4 software platforms that had evidence relevant to the modelled decision question RapidAI (mean cost £57.63 per patient), Viz (mean cost £80.73 per patient), e-Stroke (mean cost £51.53 per patient) and CINA head (mean cost £7.08 per patient) were applied to the UK

setting in terms of the number of comprehensive and primary stroke centres and total number of patients based on the Sentinel Stroke National Audit Programme. The mean of these costs, £49.24 per patient, was used in the base case. Higher costs were explored in scenario analysis. Find more details in table 24 on pages 100 to 101 of the diagnostics assessment report.

Cost of initial stroke treatment

Table 7 summarises the initial stroke treatment costs (within 90 days of stroke) for each branch of the decision tree. The EAG used Lobotesis et al. (2016) as the main source for cost parameters. Costs and resource use for treatments were from the NICE's technology appraisal guidance on alteplase for the treatment of acute ischaemic stroke, Medtronic (a manufacturer of thrombectomy devices) and Personal Social Services Research Unit (PSSRU17), Unit Costs of Health & Social Care. False positive findings of a large vessel occlusion in people who did not have one were assumed to add £559 to the treatment costs (ambulance transfer and 1 day at a stroke unit). A higher estimate of the additional cost was explored in scenario analysis. Find more details in table 25 on pages 102 to 103 of the diagnostics assessment report.

Table 7 Initial stroke treatment costs for each branch of the decision tree

Decision-tree branch	Costs per patient	Source
People with large vessel occlusion who have a thrombectomy	£8,794	Lobotesis et al. (2016) and Van Leeuwen et al. (2020)
People with large vessel occlusion who do not have a thrombectomy	£702	Lobotesis et al. (2016) and Van Leeuwen et al. (2020)
People without large vessel occlusion	£745	Patel et al. (2020)
People without large vessel occlusion incorrectly identified as having a large vessel occlusion (false positive findings)	£1,304	Patel et al. (2020)

Costs of stroke care

Table 8 shows the base-case costs of short (within 90 days of stroke) and long-term stroke care for people in different modified Rankin Scale health states. These costs included hospitalisation, management of adverse events, and the costs of personal social services such as nursing and residential care, and were sourced from Lobotesis et al. (2016). In this study, the long-term costs were estimated using data from the Oxford Vascular Study (OXVASC) and clinician input.

Table 8 Mean short and long-term stroke care costs for people in different health states

Health state	Mean short-term stroke care costs per person	Mean long-term stroke care costs per year per person
mRS 0	£3,145	£2,846
mRS 1	£3,700	£3,348
mRS 2	£4,255	£3,850
mRS 3	£16,409	£13,697
mRS 4	£22,200	£18,532
mRS 5	£26,367	£30,093
mRS 6 (cost of death)	£3,328	-

Health-related quality of life

Table 9 shows the utility values used in the base case for the modified Rankin Scale health states, from Rivero-Arias et al. (2010). This study derived modified Rankin Scale and EQ-5D-3L information from people with stroke or transient ischaemic attack who took part in the Oxford Vascular Study (OXVASC) in the UK. The utility values were adjusted for age. Alternative utility values (from Rebchuk et al. 2020 and Wang et al. 2020) were applied in scenarios. Find more details on pages 99 and 100 of the diagnostics assessment report.

Table 9 Utility values for modified Rankin Scale health states

Health state	Utility value
mRS 0	0.936
mRS 1	0.817
mRS 2	0.681
mRS 3	0.558
mRS 4	0.265
mRS 5	-0.054

Further assumptions

- The assessment primarily considered the claim that AI-derived software-assisted CT brain scan review more accurately detects large vessel occlusions.
- Thrombectomy eligibility was independent of the diagnostic strategy.
- The risk of recurrent stroke was assumed the same across all mRS health states.
- People who had a recurrent stroke could not have a thrombectomy.
- Transitions between health states (mRS 0 to mRS 5) were only possible in case a recurrent stroke happened. After a recurrent stroke, patients could either stay in their mRS health state or move to a more severe mRS health state.
- False positive findings had cost consequences only.

Effect of these assumptions on the base-case model results was explored in sensitivity and scenario analyses.

Base case results

Cost effectiveness of software-assisted review of CT angiography brain scans

The software-assisted review of CT angiography brain scans to detect large vessel occlusions was more effective and more costly compared with review

of the scans without the software. The probabilistic incremental cost-effectiveness ratio (ICER) for the software-assisted review strategy was £3,380 per QALY gained (deterministic ICER £3,490 per QALY gained).

In the short term, the software-assisted review of the scans detected a slightly higher proportion of large vessel occlusions (17.9% compared with 17.6%) and so a slightly larger proportion of people had a thrombectomy. This meant that a slightly higher proportion of people were more functionally independent (with a mRS score 3 or less) and had a better quality of life, although this difference was small (less than 0.1%). The costs of this strategy were higher mainly because of the initial costs of the software and the larger number of thrombectomies.

The results of the probabilistic analysis, shown in table 10, and the deterministic analysis, shown in table 30 on page 111 of the diagnostics assessment report, are similar.

Table 10 Probabilistic cost-effectiveness results

Strategy	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER
Clinician's CT angiography review	£116,273	5.9000	-	-	-
Clinician's AI software-assisted CT angiography review	£116,281	5.9026	£9	0.0025	£3,380

Abbreviations: QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio

Analysis of alternative scenarios

Robustness of the cost-effectiveness results to alternative model assumptions was considered in several scenario analyses. Find the full list of the EAG's scenario analyses on pages 109 and 110 of the diagnostic assessment report.

Scenario analyses

Software-assisted strategy dominated current practice when:

- proportion of people with large vessel occlusion eligible for thrombectomy in the software-assisted review strategy was increased to 50% (from 41.2%; scenario 2)
- sensitivity of the software-assisted review was increased to 96% (from 94.13%; scenario 3)
- mortality was allowed to be lower than general population mortality (scenario 16)
- sensitivity and specificity of current practice strategy was modelled using less experienced scan reviewer (resident reviewer) estimates from Seker et al. (2020) (scenario 20)

When the software costs were increased from £49.24 to £100 per patient, the ICER for the software-assisted review strategy increased to £22,072 per QALY gained (scenario 1).

Current practice dominated software-assisted review strategy when:

- sensitivity of the software-assisted review was decreased to 90% (lower than the sensitivity of the clinician's review; scenario 4)
- sensitivity and specificity of the current practice strategy was modelled using expert scan reviewer (neuroradiologist) estimates from Seker et al. (2020) (sensitivity was increased from 93.00% to 97.1% and specificity from 94.09 to 98.6%; scenario 19)

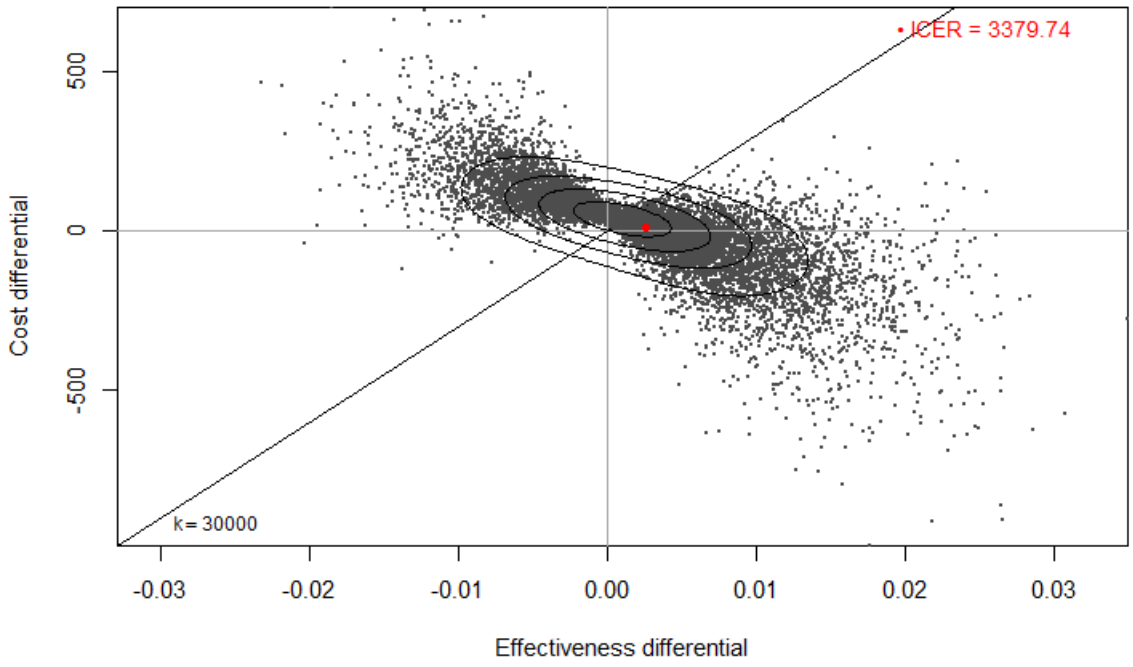
In other scenarios the ICERs for the software-assisted scan review were all below £7,000 per QALY gained. Find more details in table 31 on pages 124 and 125 of the diagnostics assessment report.

Sensitivity analyses

Figure 3 presents probabilistic sensitivity analysis simulations plotted on a cost-effectiveness plane. This shows that the more effective the software-assisted scan review is compared with the current practice, the less it tends to cost.

Based on the cost-effectiveness acceptability curves from the same analysis, at a maximum acceptable ICER of £20,000 per QALY gained, the software-assisted scan review has a 53.6% probability of being cost effective. At £30,000 per QALY gained, this probability is 56.2%.

Figure 3 Cost-effectiveness plane from the probabilistic sensitivity analysis of the base case



Based on sensitivity analyses run, the EAG highlighted the sensitivity of both the software-assisted scan review and the clinician's review alone as important parameters for cost-effectiveness estimates. Further parameters highlighted by the EAG were the proportion of people with large vessel

occlusion for whom thrombectomy was suitable, software costs and stroke care costs for people with the worse health states mRS 4 and mRS 5.

The EAG did 2-way sensitivity analyses between different combinations of software specific model inputs that most strongly drove the model conclusions: the sensitivity of the software-assisted scan review, software costs and the proportion of people with large vessel occlusion for whom thrombectomy was suitable in the software-assisted review strategy. These showed that although sensitivity was the main driver of the results, the software costs and the proportion of people with a large vessel occlusion for whom thrombectomy was suitable in the software-assisted scan review strategy can affect the sensitivity that is needed for this strategy to be cost effective. Find full analysis results on pages 112 to 123 of the diagnostics assessment report.

4 Summary

Clinical effectiveness

Only limited evidence, particularly to inform decision questions 1 and 2b, was found. No eligible studies were found for 9 of the 13 included software platforms. No study reported data for more than 1 software platform.

None of the diagnostic accuracy studies evaluated the software for assisting clinicians' review of the CT brain scans (as per intended use) but rather when used as a standalone intervention.

For decision question 2a (people with ischaemic stroke having a CT angiography brain scan), 1 study (Seker et al. 2020) provided accuracy estimates from clinician alone and software alone (Brainomix e-CTA) in the same population. Sensitivity was lower for the software. The EAG commented that, based on this, the software would only improve performance when used together with a clinician if it detected large vessel occlusions missed by the clinician. Sensitivity estimates from other studies for some of the other

software platforms used alone were higher than clinician alone estimates from Seker et al. (2020) but these were estimated in different populations. The sensitivity estimates of the software as a standalone intervention ranged from 63.6% to 98.1% and the specificity from 79.4% to 98.2%.

The observational 'before and after' studies provided information about the effects of implementing the software but only for people who had a thrombectomy. None of these studies were done in the UK. For people with ischaemic stroke having a CT angiography brain scan (decision question 2a), all 5 studies that reported this outcome showed a reduction in time to thrombectomy after software was adopted. Impact on clinical outcomes (such as mRS score after thrombectomy), where reported, was less clear and no statistically significant changes in clinical outcomes were found.

No studies were found on the effects of software-assisted review on health-related quality of life outcomes for any of the decision questions.

Cost effectiveness

The EAG did not produce cost-effectiveness estimates for using the software according to decision questions 1 and 2b.

The EAG's modelling addressed decision question 2a and investigated the potential benefit of a software-assisted CT angiography brain scan review providing a more accurate diagnosis of large vessel occlusion. A de novo model that used expert elicitation for accuracy inputs estimated an ICER of about £3,400 per QALY gained for the addition of the software to detect large vessel occlusion in CT angiography scans compared with clinician review alone. The EAG did not produce cost-effectiveness estimates for specific software platforms. It used sensitivity and specificity estimates from expert elicitation informed by accuracy estimates from several different company's software, and used an average cost from several different software platforms.

Differences between the 2 strategies, in outcomes and costs, were in general very small. Sensitivity of both the software-assisted scan review and the clinician's review alone were important parameters for cost-effectiveness estimates. Further parameters which had a noticeable impact on the ICERs highlighted by the EAG were the proportion of people with large vessel occlusion for whom thrombectomy was suitable, software costs and stroke care costs for people with the worse health states mRS 4 and mRS 5.

In the base case, the EAG did not include the potential impact of the software on reducing time to thrombectomy. In a scenario analysis, the proportion of people for whom thrombectomy was suitable when software was used was increased to simulate quicker detection of large vessel occlusions meaning more people could have thrombectomy. In this scenario, use of software dominated current practice.

5 Issues for consideration

Clinical effectiveness

Only limited evidence, particularly to inform the decision questions 1 and 2b, was found.

In all the diagnostic accuracy studies, for all the decision questions, the AI-derived software was evaluated as a standalone intervention and not as an aid to human interpretation. Only 1 study, for decision question 2a, provided a direct comparison between accuracy of the software alone and clinician alone in the same population. It is uncertain how using the software alongside the clinician's review will affect the accuracy of the non-enhanced CT (decision question 1), CT angiography (decision question 2a) and CT perfusion brain scans (decision question 2b).

The observational 'before and after' studies provided information about the effects of implementing the software but only for people who had a thrombectomy. The EAG commented that these studies provide no

information about the extent to which the software-assisted CT brain scan review may have missed detecting a target condition (for example a large vessel occlusion) and influenced the eligibility for treatment.

In the observational 'before and after' studies relevant to decision question 2a, time to thrombectomy was reduced but the effect on clinical outcomes after the software introduction was less clear. None of the 'before and after' studies relevant to decision question 1 and 2b reported this data after implementing a software that assisted only the review of non-enhanced CT or CT perfusion brain scans. The EAG highlighted that none of the studies for any of the 3 decision questions evaluated the effects in populations that were both comparable before and after software implementation and where adding the software was the only change to the care pathway. This means that the extent to which any of the effects seen were because of the software implementation is uncertain.

No eligible studies were found that assessed Aidoc (Aidoc), icobrain + Aidoc (Aidoc and icometrix), RapidAI (Ischemaview), e-Stroke (Brainomix), qER (Qure.ai), Zebra-Med (Zebra Medical Vision), Brainscan CT (Brainscan.ai), CINA head (Avicenna), Accipio (MaxQ AI) or Biomind (Biomind.ai) for decision question 1. No eligible studies were found that assessed Aidoc or icobrain + Aidoc for decision question 2a, or icobrain ct, e-stroke, Viz (Viz.ai), CT Perfusion 4D (GE Healthcare) or Cercare stroke (Cercare Medical) for decision question 2b.

Cost effectiveness

No cost-effectiveness models were developed for decision questions 1 and 2b because the evidence available was not enough to support the modelling.

The model developed for decision question 2a used expert elicited accuracy estimates for the base case. The model did not consider differences in performance between different software. No data showing equivalence of performance between software were identified and differing accuracy

estimates across software were reported (although no study assessed more than 1 software in the same population).

Without evidence on the accuracy of the software-assisted scan review compared with clinician's review alone, there is considerable uncertainty over how use of software would affect clinician's judgement. The assumptions made by experts in producing estimates of the accuracy of clinicians alone and clinician plus software from available data are not clear.

The cost of the AI-derived software in the base case was based on an average cost of several software packages. The technology-specific costs are considerably higher or lower than the cost used in the base case. Cost of software was a large driver of the higher incremental costs of the software-assisted CT brain scan review strategy.

The base-case model did not include potential impact of reduced time to thrombectomy caused by the addition of the AI-derived software-assisted review. The EAG noted the results of an individual-patient-data meta-analysis (Saver et al. 2016) and a multicentre RCT (MR CLEAN study, Fransen et al. 2016) show that as time to thrombectomy increases, clinical outcomes get worse. Saver et al. (2016) reported that each hour of reperfusion delay was associated with a reduction in the proportion of patients achieving function independence (mRS 0 to 2), with an absolute risk difference of -5.2% (95% CI: -8.3% to -2.1%). The EAG commented that it remains unclear whether the potential reductions in time to intervention that might be achieved by implementing AI-derived software technologies would translate into improved clinical outcomes in 'real world' settings.

6 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

People who have had a stroke may have impaired cognitive function and physical disability that limits everyday activity. Disability is protected characteristic under the Equality Act 2010. AI algorithms for stroke diagnosis may have lower sensitivity in people over the age of 80 with small vessel disease and calcification of the cerebrovasculature. Ability to assess the performance of AI algorithms in different age groups may be driven by the availability of training data in different age groups. Some people may have limitations in their ability to cooperate with being scanned.

7 Implementation

IT issues

IT compatibility issues and capacity issues within NHS Trusts may be a potential barrier to the implementation of AI software. However, if hospitals within a stroke network purchase the same AI software this may help overcome this barrier.

Clinician confidence

Confidence in the accuracy of the software can influence the choice to use it. False positive results may increase clinician's workload if several reviews must be done before overriding the results from software.

8 Authors

Suvi Härmälä
Topic lead

Thomas Walker (to December 2021), Amy Crossley (from January 2022)
Technical advisers

Glossary

CT angiography brain scan

Imaging done with a special dye to check how blood flows through the vessels in the brain.

CT perfusion brain scan

A non-invasive imaging that shows which areas of the brain are adequately supplied with blood and provides detailed information on delivery of blood or blood flow to the brain.

Large vessel occlusion

A blood clot in a large artery in the brain.

Modified Rankin Scale (mRS)

A functional assessment scale that measures the degree of disability or functional dependence after a stroke. The scale runs from perfect health without symptoms (mRS 0) to death (mRS 6). An mRS score of 2 or less indicates no or only slight disability and functional independence.

Thrombectomy

A procedure to help restore blood flow to the brain in ischaemic strokes that are caused by a large vessel occlusion. It involves using a small device, passed through a catheter into the artery, to remove the blood clot. The procedure can be done under local or general anaesthetic.

Thrombolysis

Use of a drug called alteplase to dissolve blood clots and help restore blood flow to the brain in ischaemic stroke.