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

NICE took the decision to pause publication of its guidance on 'Artificial intelligence (AI) software to help clinical decision making in stroke' in August 2022 to allow discussions with the NHS England Getting It Right First Time (GIRFT) stroke programme to take place. Further evidence has now been submitted by stakeholders and the EAG has updated its searches from the previous Diagnostic Assessment Report and provided an addendum to the main report. This overview summarises the main issues the diagnostic advisory committee needs to consider. It has been written by the NICE technical team and includes information from the addendum document written by the EAG, information from submissions made by stakeholders and results from the EAG's model produced by the NICE team. It should be read with the scope.

1 Aims and scope

Software with 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. Time from onset of stroke symptoms is a factor in deciding whether these treatments are used (see scope section 4.4).

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 have 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

Question 1: People referred to or attending secondary care with a suspected acute stroke who were last known to be well within 24 hours

Question 2a: People with an ischaemic stroke who were last known to be well within 6 hours

Question 2b: People with an ischaemic stroke who were last known to be well within the last 6 -24 hours who have already had a CTA brain scan

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/platforms:

- Accipio
- Aidoc
- Aidoc + icobrain
- Biomind
- Brainscan
- Cercare stroke
- Cina head
- CT Perfusion 4D
- E-stroke
- Icobrain CT
- QER
- RapidAI
- Viz
- Zebra-Med

See the scope for further details on the type of CT scan each software can analyse. The software packages consist of several AI-derived modules which assess different CT scans: non-contrast CT (NCCT), CT angiogram (CTA) and CT perfusion (CTP). The purpose of the software is to support brain scan review and reporting by a trained healthcare profession.

Comparator

An unassisted review of CT brain scans by a healthcare professional

Healthcare settings

Comprehensive stroke centres (CSCs) and acute stroke centres (ASCs).

Further details, including descriptions of the interventions, comparator, care pathway and outcomes, are in the scope.

2 Summary

The EAG highlighted that the addendum it has produced is not intended as a stand-alone document and should be read with the original report, an overview of which can be found here. The protocol for the assessment can be found here.

Clinical effectiveness

The EAG updated their systematic review (see page 21 in the addendum for details) and identified 5 new studies. Three studies assessed the Viz platform (Viz.ai; Figurelle et al. 2023, Matsoukas et al. 2023 and Hassan et al. 2022) and 2 assessed the e-Stroke platform (Brainomix; Gunda et al. 2022 and ongoing work being done by an Academic Health Science Network [AHSN], described below). Other studies mentioned in the addendum related to conference abstracts of the same studies (Hassan et al. 2020 and Gunda et al. 2020) and a conference abstract reporting data from 1 of the participating sites (a primary stroke centre) involved in the AHSN work (Nagaratnam et al. 2021).

All were retrospective studies which assessed the effects of implementing an AI-derived software in real world settings and compared stroke centres with and without use of the AI-derived software, or across time periods before and after implementing the technology (see section 3.2.1 in the addendum for further study details, including quality assessment).

The EAG highlighted that the implementation periods of 3 studies (Figurelle et al. 2023, Hassan et al. 2022 and the AHSN report) overlapped with the COVID-19 pandemic.

One of the included studies was a report of a large UK implementation study provided by the Oxford AHSN, who are evaluating e-Stroke (Brainomix) as

part of the AI in Health and Care Award. The authors explained that this was an interim report to reflect the second year of evaluation findings (as part of a 3-year evaluation), and was not intended to be an academic, peer reviewed, scientific research publication. Their evaluation is currently ongoing, and data will be collected until the end of 2023. Findings presented in the report represent the implementation of e-Stroke in existing stroke pathways at 20 Acute Stroke Centres (ASC), and 6 Comprehensive Stroke Centres (CSC).

The AHSN work included a digital survey of clinicians at implementation sites. The EAG raised concerns about how the survey had been done and commented that some questions sought clinical opinion about outcomes that it considered should be obtained by direct measurement (see page 32 in the addendum). The EAG also noted that all survey data were derived from a total of 34 respondents, distributed across 16 out of 26 participating sites, the majority of whom were from clinical disciplines who do not have expertise in interpreting brain scans or carrying out thrombectomies, that the interpretations of survey results were sometimes not supported by the data presented, and opinions were sometimes from a single respondent at the site.

A submission was made by the NHS England GIRFT Programme to NICE for this assessment.

In April 2022, GIRFT published its national speciality report for Stroke. This included recommendations to provide infrastructure, training and technology to share images between hospitals and clinicians to support image interpretation, with actions including to:

- Increase regional availability of AI decision-support tools and training.
- Provide national support for regional roll-out of AI working closely with ISDN (Integrated Stroke Delivery Networks) footprints.

The EAG said that no study provided sufficient information to establish both that populations were comparable before and after the implementation of the

AI-derived software technology, and that the AI-derived software technology was the only change to the care pathway.

- Gunda et al. (2022) stated that there were no other changes to the care pathway over the study period. The EAG commented that the study authors could not exclude other factors contributing to improved stroke care, such as increased public awareness of stroke and ongoing departmental quality improvement.
- Hassan et al. (2022) reported data indicating that there were no significant differences between patients who received thrombectomy before and after implementation of the AI-derived software technology (age, sex, ethnicity, baseline NIHSS score, receipt of thrombolysis before transfer, or co-morbid conditions) but did not report any information about changes in the care pathway, other than the implementation of the AI-derived software technology.
- The EAG commented that the AHSN work suggested that changes to stroke care, other than the implementation of the AI-derived software technology, were dynamic and ongoing throughout the evaluation period (see page 42 in the addendum). Nagaratnam et al. (2021) stated that there were no changes to the care pathway, other than adoption of the software, in the centre and period they assessed. However, the EAG considered that this did not seem plausible given that the likely place where the study was conducted, the place of work of the first two authors and a third author, the Royal Berkshire Hospital, was also mentioned in another publication, Nagaratnam et al. (2020), which stated that e-Stroke was implemented alongside other changes in the care pathway specifically in response to COVID-19. These included the creation of a WhatsApp group which provided supplementary support to pool the expertise of several specialists (stroke and neurology consultants) during clinical decision making.

The EAG's opinion overall was that the evidence base has not changed substantively since completion of their previous report in 2021. It stated that it

considered the available evidence remains unsuitable to determine the clinical effectiveness of using AI-derived software to support the review of CT brain scans in acute stroke, in the NHS setting.

Ongoing studies

The AHSN work is ongoing, and the group are still collecting data (until the end of 2023) with further analyses planned. The EAG described and critiqued these analyses in the addendum (see sections 4.1 and 4.2 in the addendum).

The EAG also highlighted widespread current use of AI decision support in the NHS (as of April 2023, implementation at 99 of 107 stroke units in England with all other identified centres working on plans to go live before the end of 2023) and that this may limit the potential for commissioning some types of research (see page 69 in the addendum).

No data were found in the initial review on many technologies (Accipio, Aidoc, Biomind, Brainscan CT, Cercare (Perfusion), CT Perfusion 4D, icobrain ct, Neuro Solution or qER). The NHS England GIRFT Programme's submission stated that these are currently not used by any stroke services in England, and that this is a position which is not expected to change any time soon.

Cost effectiveness

The EAG presented no further economic model results in its addendum. It stated that this assessment did not identify any new evidence that would be sufficient evidence to support further modelling.

Recap of previous economic model provided by the EAG

In their previous work, the EAG produced an economic model for population 2a (for detecting large vessels occlusions [LVOs] to guide use of thrombectomy). A decision tree was used to estimate the number of people with an LVO who were eligible for thrombectomy (proportion of ischemic strokes that have LVOs multiplied by proportion of LVOs that are eligible for thrombectomy), and, of these, the proportion that were detected (sensitivity of

assessment to detect LVOs). People with detected LVOs had a thrombectomy which meant they had a lower level of disability at 90 days post stroke (they were in a lower modified Rankin Score [mRS] state), compared to people with an LVO who didn't have a thrombectomy (including false negative cases, where an LVO suitable for thrombectomy was missed). The long-term implications of this were modelled using a Markov model. Further detail on model structure can be found in the EAG's previous report in section 4.2.

The number of people having thrombectomies increased in the model when AI-derived software was added to clinician review if this increased (compared to clinician review alone) the:

- Sensitivity to detect LVOs (as in the base case), or
- Proportion of patients with LVO who are eligible for mechanical thrombectomy, because time to treatment was reduced (done in a scenario analysis).

If an LVO was incorrectly detected (false positive) this was assumed not to result in an unnecessary thrombectomy. It did increase costs, because of unnecessary transfer to a stroke centre qualified to perform thrombectomy. Based on clinical opinion and consistent with a previous model, the EAG assumed a specialist would review the image before agreeing to take patients for thrombectomy and then detect the false positive.

3 Issues for consideration

Key consideration 1: Data on the use of the software to guide thrombolysis decisions (decision question 1)

Description of issue

Less data is available for this use, compared to use of the software to make decisions about thrombectomy. Data from 2 studies identified may give conflicting results about the impact of the software on the number of people

having thrombolysis and time to treatment. It also isn't certain that any changes can be solely attributed to AI-derived software implementation.

Background

Evidence on using AI-software technologies for guiding thrombolysis treatment decisions in the previous review was very limited (section 3.9 of the draft guidance). The EAG did not produce an economic model for this use.

The EAG identified 2 further studies, both assessed the Brainomix e-Stroke package (see section 3.2.3 in addendum for full detail) that met review inclusion criteria. The EAG commented that both studies assessed time from scan to needle and reported no clear difference pre- to post-implementation.

- Gunda et al. (2022) reported a statistically significant increase in the proportion of stroke patients who received thrombolysis after implementation of the AI-derived software technology (e-ASPECTS and e-CTA modules): 11.5% (46/399) before implementation and 18.1% (72/398) after (odds ratio [OR] 1.69; 95% confidence interval [CI] 1.137 to 2.257; p=0.009). There was a non-significant decrease in the time to delivery of thrombolysis: 44 minutes before implementation and 42 minutes after.
- The AHSN report provided interim data from 26 participating sites and found the proportion of people who received thrombolysis had decreased at 17 sites and increased at 9 sites after e-Stroke implementation. The median time to treatment increased at 19 sites and decreased at 6 sites after e-Stroke implementation (1 site had no thrombolysis patients in the pre-implementation period; see table 4 in the addendum for full data).

The AHSN report stated that evaluation sites had suggested that e-Stroke does not have a significant impact on the decision time to administer thrombolysis, but over a third of those that responded to their survey did use e-Stroke to identify potential patients for thrombolysis (detail not included in the EAG's addendum). The EAG commented that this statement was not supported by any further detail.

The EAG commented that neither study reported any data on clinical outcomes.

Question for committee

- What conclusions about the impact of the AI-derived software on decisions about thrombolysis can be made from the available evidence?

Key consideration 2: Accuracy of software used alongside clinician interpretation to guide decisions about thrombectomy (decision question 2)

Description of issue

No studies were again identified that assessed the accuracy of AI-derived software used alongside clinician review (rather than when used alone) and which met the review's inclusion criteria. A stakeholder stated that there is evidence on the impact of adding AI-derived software to clinician review.

Background

In its previous report, the EAG only identified accuracy data from studies where the AI-derived software technology was evaluated as a stand-alone intervention, rather than as an adjunct or aid to human interpretation, as per intended use. In section 4 of the draft guidance, the committee recommended further research be done on the accuracy of the technologies plus clinician compared with clinician alone. For the addendum, the EAG did not discuss studies that assessed AI-derived software as a standalone intervention (detail on such studies can be found in table 2 of the addendum).

No studies were found that fulfilled the inclusion criteria for assessing the accuracy of AI-derived software used alongside clinician interpretation.

- Andralojc et al. (2023) did not meet the inclusion criteria specified in the protocol for this assessment, because the reference standard was only used when e-CTA (e-Stroke, Brainomix) and radiologist disagreed (when

reviewing 300 consecutive scans from Royal Cornwall hospitals NHS Trust) and because the extent of software use was not clear. But the study was discussed by the EAG in the discussion section of the addendum (see page 63). The authors reported results indicating that if decision making always followed e-CTA findings, 1 additional LVO would have been correctly identified, and 32 additional false positives would have been produced, by e-CTA compared to the radiologist review. The prevalence of LVOs in this population was much lower (7%; patients presenting with acute stroke who underwent CTA within the thrombectomy treatment window [unspecified]) than studies previously identified by the EAG, and the prevalence of LVOs used in the EAG's model (46.1%; estimated by pooling prevalence of LVOs from the accuracy studies identified by the EAG as part of the initial report). The study authors suggested the estimate for e-CTA alone was potentially biased in its favour. They also stated that it may be valuable for radiologists to review the output of e-CTA carefully and double-check any areas highlighted (or not highlighted) by the software, and not be unduly influenced by the software should they disagree with it.

The EAG provided a table of studies which were screened for inclusion based on full text publication but did not fulfil 1 or more of the inclusion criteria (see appendix 3 in the addendum). One such study was not discussed further in the addendum but is included in this overview. This is because this study assessed use of AI-derived software alongside clinician review, compared to clinician review alone, with specified software that is in scope of the assessment and provided accuracy estimates for a target condition that was specified in the EAG's inclusion criteria. The reason for exclusion for 'outcomes' was because of limited information on methodology in the poster, and, following completion of the addendum, additional information clarifying study methods was provided by Brainomix. Reason for exclusion for 'population' relates to the negative cases only (the population included non-stroke diagnoses):

- Mathieson et al. (2022; a poster presented at the 2022 European Stroke Organisation Conference submitted by Brainomix) was excluded because of the population assessed and outcomes reported (see page 94 in appendix 3 of the addendum).
  - Seventeen UK radiologists and stroke physicians reviewed the same 20 cases (10 cases selected with LVO and 10 without) provided as vignettes with non-contrast CT reports and CTAs and were asked to interpret the images for presence of LVOs. Negative cases (contributed to specificity) were non-stroke cases (rather than confirmed stroke with no LVO), and positive cases (contributed to sensitivity), were from people with LVOs who were selected from a historical research registry (Grunwald et al. 2019) with occlusion locations to match the distribution in a published individual participant data meta-analysis of the effectiveness of mechanical thrombectomy (Goyal et al. 2016). Half of the CTAs were randomised to be provided with e-CTA (e-Stroke, Brainomix) decision support, and the remaining half with no decision support. Two weeks later, for a subgroup of 9 reviewers the allocation of decision support was reversed so all cases were reviewed with and without decision support. The EAG commented that it was not clear whether this randomisation procedure was performed separately for each reader.
  - The total number of case reviews was 520 (this was incorrectly stated as 220 in the poster) with all readers completing 1 read of the 20 cases, and a subgroup of 9 readers reviewed the cases for a second time. Results from pooled data from both reading points were used to derive accuracy estimates, and that included [redacted] data points for the without AI estimate and [redacted] data points for the with AI estimate; no per reader estimates were presented. Cases where the reader was uncertain and needed a second opinion were not included in accuracy estimates. [redacted]. The EAG commented that the effect of excluding uncertain cases

from the analyses depends upon the distribution of uncertain results between positive and negative cases and how uncertain results would have been classified in a real-world clinical scenario (with a second opinion).

- The authors reported that randomisation to e-CTA improved sensitivity to LVO from 89% to 97% (p=0.007) with a non-significant increase in specificity from 92% to 95% (p=0.3).

The NHS England GIRFT Programme's submission stated that there was evidence on the impact of adding AI-derived software to clinician review and cited 2 papers (Brinjikji et al. 2021 and Grunwald et al. 2019). Both studies were identified by the searches conducted for the original diagnostic assessment but did not meet the inclusion criteria for outcomes reported; see appendix 4 of the initial report) and are not discussed in the addendum. The EAG commented that the studies included only people with an LVO and so could not provide information about accuracy for detecting LVOs.

- Brinjikji et al. (2021) assessed the Brainomix e-ASPECTS software accuracy to detect ASPECTS regions affected in anterior circulation LVO. The GIRFT submission also commented that ASPECTS is the most common criterion used on NCCT for consideration of thrombectomy. Sixteen readers (senior neuroradiologists, junior neuroradiologists and vascular neurologists) interpreted CT scans from 60 participants initially without e-ASPECTS software, and then 2 months later again evaluated the CT scans but with assistance of e-ASPECTS software. Accuracy with and without e-ASPECTS assistance, for each of the 10 regions of the brain assessed by ASPECTS, was provided but only based on combined true negative and true positives reads (that is, no sensitivity or specificity reported). Accuracy was higher in 9 out of the 10 regions when e-ASPECTS was used ($p<0.001$ for 8 regions). Data was also provided split by experience of reviewer. Overall accuracy was higher ($p<0.01$) for senior neuroradiologists, junior neuroradiologists and neurologists with e-ASPECTS software, compared to without. The EAG commented that this

study did not provide accuracy for relevant clinical decision thresholds or use a reference standard specified for this assessment.

- In Grunwald et al. (2019), 3 experienced neuroradiologists independently estimated a CTA collateral score, without and then with knowledge of the e-CTA output. A consensus score was then agreed. Addition of the e-CTA improved the intraclass correlation coefficient between radiologists from 0.58 to 0.77 (a difference of 0.19; 95% CI 0.09 to 0.31; p = 0.003). The EAG commented that the only accuracy data presented were for e-CTA alone and the detection of a favourable collateral score in patients who are eligible for thrombectomy.

Question for committee

- What conclusions about the impact on accuracy of adding AI-derived software to clinician review to identify people suitable for thrombectomy can be made from the available evidence?

Key consideration 3: Impact of the software on time to thrombectomy

Description of issue

Data on the impact of AI-derived software on time to thrombectomy is mixed, and it isn't certain that changes can be solely attributed to AI-derived software implementation. The EAG stated that time to intervention data alone, without outcome data, is insufficient to assess cost effectiveness.

Background

In the initial report, most studies suggested that time to treatment for people who had treatment reduced after implementing the software. But it was unclear if the populations compared had similar characteristics, and whether adding the software was the only change to the care pathway. In the draft guidance, the committee concluded that it was uncertain whether using AI-derived software to help guide treatment decisions in stroke leads to faster

access to thrombolysis or thrombectomy (see section 3.6 of the draft guidance).

The EAG highlighted data (Saver et al. 2016 and the MR CLEAN trial; Fransen et al. 2016) that indicated a negative correlation between time to intervention and functional outcome in patients with LVO who undergo thrombectomy (including 90-day mRS). But it considered that the available evidence is currently not sufficient to support the assumption that the introduction of AI-derived software technologies is associated with clinically meaningful reductions in time to intervention (see page 64 of the addendum).

The EAG also stated that it is important to measure clinical outcomes alongside time to intervention outcomes because it is possible, for example, for the implementation of AI-derived software technologies to reduce time to intervention whilst also being associated with poorer clinical outcomes (for example, if more cases eligible for thrombectomy are missed because of the software). It commented that no study reported information to suggest that reductions in time to intervention associated with the implementation of AI-derived software technologies were also associated with improvements in clinical outcomes.

Some stroke centres don't have an on-site thrombectomy service and need to transfer to stroke centres that can do this. The EAG cautioned that it is unclear to what extent changes in outcomes that relate to decision time or transfer (such as door-in-door-out [DIDO]) translate to clinically meaningful changes in time to treatment.

The NHS England GIRFT Programme's submission stated that: "The effect on clinical outcomes will only be shown through modelling the impact of earlier onset to treatment times through reducing DIDO times for patients with LVO. It will not be possible to do a trial or show impacts over time on clinical outcomes (90-day mRS) when so many other things in the system change, for example, referral criteria and expansion of service availability, changes in

ambulance response times." The AHSN report stated that: "DIDO is a better measure than scan to MT [mechanical thrombectomy], as the latter may be confounded by the variation seen in patient transfer time to the thrombectomy centre caused by pressures on ambulance availability." The EAG considered that the inclusion of factors that affect transfer time in the scan to thrombectomy metric is not an example of confounding but rather is a measure of the true, real-world effects of the intervention. It stated that if any reductions in time to decision making are dominated by delays in transfer, then it is difficult to envisage a time-saving mechanism by which the intervention can be clinically effective in real-world scenarios. The AHSN report noted that ambulance transfer times from ASCs to CSCs are a limiting factor in the delivery of thrombectomy. The EAG also noted that median DIDO times reported in the AHSN report also did not indicate a consistent reduction in time following implementation (median DIDO times increased, following implementation, at 9 sites, were unchanged at 1 site and decreased at 6 sites).

From the addendum:

Viz LVO (see table 5 of the addendum).

- Hassan et al. (2020; from 171 to 105 minutes; $p<0.016$) reported statistically significant reductions after implementation of Viz LVO in median time from CTA at a primary stroke centre to arrival at a CSC for a thrombectomy. Data on clinical outcomes were also reported in Hassan et al. (see key consideration 5).
- Matsoukas et al. (2023; from 254 to 198 minutes; $p<0.001$) reported statistically significant reductions after implementation of Viz LVO in median time from CTA at a PSC to a recanalization procedure.
- Figurelle et al. (2023) reported a statistically significant decrease in median time from door to groin puncture for people arriving directly (from 127 to 86 minutes; $p=0.006$) and for those transferred as part of a telemedicine initiative (from 42 to 28 minutes; $p=0.036$).

e-Stroke (see table 6 of the addendum).

- Gunda et al. (2022) reported a non-significant decrease in mean time from scan (in a centre which did not offer thrombectomy) to thrombectomy after adoption of e-Stroke (174 to 145 minutes; -29.00 minutes; 95% CI -78.21 to 20.21).
- The EAG compared median times from scan to thrombectomy from the AHSN report (data were only available for CSCs) and noted that 4 sites had longer scan to MT time compared to pre-implementation of e-stroke, and 2 sites had reduced time post-implementation. Data on clinical outcomes were also reported in Gunda et al. (2022) and the AHSN report (see key consideration 5).

The EAG highlighted results from the AHSN's survey of clinicians that indicated 25 of 30 (83%) responding clinicians thought that introduction of e-Stroke had reduced the time taken to reach a decision to proceed with thrombectomy (see page 66 of the addendum).

The NHS England GIRFT Programme's submission stated that the assessment should reflect that these tools are diagnostic decision support tools and their impact on patient flow through rapid image transfer as well as diagnostic interpretation. It stated that: "The focus on the software as a diagnostic has possibly rendered other findings inconsequential whereas in terms of delivering an optimal pathway with rapid diagnosis further investigation would be warranted. The review failed to take on board the rapid image transfer functionality of these products so that the CSC stroke physician and/or INR can look at the scans immediately. It is not the diagnostic accuracy that is the MOST VALUABLE feature of the AI products and speaks to the point made by the independent DAR abstract. We don't need more evidence to know this is a benefit. The NHS imaging system isn't going to change to be able to do this in place of these products."

When asked in the AHSN's survey 'what positive changes have happened since the introduction of e-Stroke', the highest scoring response (for both ASCs and CSCs) was improved communication with other sites (data not reported in EAG's addendum). The EAG commented that such data were not included in the EAG's addendum because improved communication or image sharing was not an outcome measure specified in the inclusion criteria for this assessment. It also stated that it is possible that the image sharing components of these technologies may be associated with improvements in workflow, but that image sharing and communications outcomes are not related to the AI-derived image interpretation components of these technologies.

Question for committee

- What conclusions about the impact on time to thrombectomy of introducing AI-derived software can be made from the available evidence?

Key consideration 4: Impact of software on the number of people having a thrombectomy

Description of issue

Some data are now available for the impact of the AI-derived software (e-Stroke) on the number of people having thrombectomy This may show a trend towards increase with software use, but it isn't certain that changes can be solely attributed to AI-derived software implementation.

Background

No studies in the previous review identified impact of the software on incidence of thrombectomy.

The NHS Long Term Plan states that expanding mechanical thrombectomy, from 1% to 10% of stroke patients, will allow 1,600 more people to be independent after their stroke each year.

From the addendum:

- Gunda et al. (2022) reported that the proportion of patients receiving thrombectomy was 11/399 (2.8%) before implementation of the e-ASPECTS and e-CTA (e-Stroke, Brainomix) and 19/398 (4.8%) after (OR 1.77; 95% CI 0.83 to 3.77).
- The AHSN work reported increases in the rates of thrombectomy (the proportion of stroke patients having a thrombectomy), between 0.57% and 3.46%, following implementation of e-Stroke for patients presenting at 14 out of 16 ASC sites and between 1.42% and 6.51% for patients presenting directly to 5 out of 6 CSC sites. The EAG commented that it was unclear how rates of CTA scanning and thrombectomies may have been affected by the COVID-19 pandemic.

  The authors of the AHSN report commented that thrombectomy rates are increasing over time, and that this is due in part is due to many factors such as the impact of national targets, increased skill sets, availability of mechanical thrombectomy services and the establishment of the Integrated Stroke Delivery Networks.
- The AHSN report stated that sites that use e-Stroke and participated in their evaluation had a higher rate (4.21%) of thrombectomy than the national average (2.9%; data not included in the EAG's addendum, the EAG stated this was because it is not a relevant comparator).
- The AHSN report also included responses from their survey in which clinicians were asked if e-Stroke has helped to identify more eligible patients for thrombectomy (data not included in the EAG's addendum, the EAG stated that this was because it was based on opinion about effect on thrombectomy numbers). Of the 29 people that responded, 18 (62%) believed that e-Stroke did help, a further 7 (24.1%) weren't sure and 5 (17.2%) thought that it did not help. The report further stated that clinicians who have specific expertise in either carrying out thrombectomy (Interventional Radiologist) or those that interpret brain scans (Radiologists) see less value in e-Stroke identifying more suitable patients,

whereas Stroke Consultants, Physicians and Specialist Stroke Nurses and Practitioners, see greater value.

- The AHSN report stated the number of thrombectomies done after more than 6 hours increased over time in evaluation sites. It further commented that MRI and CT perfusion (CTP) have allowed treatment window for thrombectomy to be expanded to 24 hours from the onset of symptoms and given that all CSCs have access to CTP imaging, the number of patients now eligible for thrombectomy has increased. The AHSN report also stated that a likely outcome of e-Stroke's rapid image sharing functionality was that ASC teams have more immediate access to thrombectomy specialists, who before the implementation of e-Stroke would be more likely to decline a patient for transfer if they were last known well more than 6 hours ago as they did not have any further information (brain scans) available to them to inform this decision. Information from the AHSN group provided to the EAG indicated that the proportion of people receiving thrombectomy who had presented more than 6 hours after the onset of symptoms increased from 197/666 (29.6%) in the period before, to 251/652 (38.5%) after implementation of e-Stroke (OR 1.49; 95% CI 1.185 to 1.874).

Question for committee

- What conclusions about the impact of introducing AI-derived software on the number of people having a thrombectomy can be made from the available evidence?

Key consideration 5: Impact of software on clinical outcomes

Description of issue

Data on the impact of the software on mRS and mortality was mixed, and it isn't certain that changes can be solely attributed to AI-derived software implementation. There was a lot of loss to follow-up (up to 80%) from the only study that looked at longer-term (6 months) mRS after a stroke (AHSN report) which makes it hard to draw conclusions from these data.

Background

The committee noted that the results from studies identified in the previous review that reported on clinical outcomes (like mRS) were conflicting with some reporting a positive and others a negative impact. The EAG advised that the studies were unlikely to have been appropriately set up to adequately capture any differences in clinical outcomes. Therefore, the reported data are unlikely to show the true effects of implementing the technologies (see section 3.7 of the draft guidance).

Data on additional clinical outcomes can be found in the addendum (tables 5 and 6). The following section summarises data on the impact of software on mRS and mortality.

From the addendum (for full detail see table 6) for the **e-Stroke** platform:

- Gunda et al. (2022) reported a non-significant increase post-implementation of e-ASPECTS and e-CTA (e-Stroke, Brainomix) in the proportion in mRS states 2 or lower (good functional outcome; OR 1.05; 95% CI 0.216 to 5.090) and 1 or lower (excellent functional outcome; OR 2.55; 95% CI 0.414 to 15.653) at 90 days. There was no information about mRS for people who did not achieve at least a good functional outcome.
- The AHSN report included data on mRS. The amount of missing data for 6-month mRS was high (68.0% pre-implementation and 80.5% post-implementation). Results varied depending on whether the values were expressed as a proportion of people with available mRS data (as provided by the AHSN) or the total number of people who had thrombectomy (EAG's preference; see page 44 in the addendum for detail). Using the total number of thrombectomy patients as the denominator, the proportion of patients who achieved a good functional outcome (mRS of 2 or less) after 6-months was significantly lower in the post-implementation period than pre-implementation (OR 0.58; 95% CI 0.407 to 0.815) and there was a non-significant increase in the 6-month mortality post-implementation (OR 1.25;

95% CI 0.609 to 2.551). Using the same data set with the number of thrombectomy patients for whom mRS data were available as the denominator, there was no difference in the proportion of patients who achieved a good functional outcome (mRS of 2 or less) after 6-months in the pre-implementation period (95/213) compared to the post-implementation period (57/127; OR 1.01; 95% CI 0.650 to 1.573). However, the increase in the 6-month mortality rate became statistically significant (OR 2.20; 95% CI 1.043 to 4.626). The EAG cautioned that this data should be viewed with extreme caution because data were missing for the majority of thrombectomy patients.

- For mortality at discharge for people who had a thrombectomy more than 6 hours after onset of symptoms there were non-significant decreases in the mortality rate at discharge, and the impact on proportions at mRS score 2 or less at discharge varied by whether people presented directly to a CSC or were transferred.
- Nagaratnam et al. (2021) reported that the proportions in mRS 2 or lower at 3 months was higher post-implementation of the e-Stroke platform, but not significantly so (48% compared to 16%; OR 1.67; 95% CI 0.34 to 8.18).

For the **Viz LVO** (see table 5 in the addendum):

- Hassan et al. (2022) reported a non-statistically significant increase post-implementation of Viz LVO, in the rates of mRS states 2 or lower (good functional outcome) at 90 days from 8/28 (29%) to 14/35 (40%; p=0.34). Mortality at discharge showed a non-statistically significant decrease post-adoption of Viz LVO from 6/28 (21.4%) to 5/35 (14.3%; p=0.46).
    - Hassan et al. (2020) reported a similar non-significant increase in the number in mRS states 2 or lower at discharge, but mortality was higher, non-statistically significant so, after adoption of Viz LVO (OR 1.33; 95% CI 0.310 to 5.727). It is unclear to what extent data in the 2 Hassan et al. studies were from an overlapping group of patients.

The EAG commented that outcomes related to treatment effectiveness are likely to be affected by which patients are selected for treatment (that is, accuracy). It also said that information provided by studies of this type is limited in that it concerns only treated (that is, test positive) patients. No information is provided about test negative patients. The impact of false negative results (that is, people with an LVO who could have benefited from a thrombectomy if it had been detected) was included in the EAG's model (see table 20 in the initial report).

The EAG's model included the impact of false negative results as a cost impact only. The NHS England GIRFT Programme submission stated that "The false negative risk is really not relevant – it assumes that the products are being used instead of a radiology report – they aren't. If the product isn't there it isn't going to change when and who looks at and reports the scan. So worst case scenario, AI and the clinician looking at the CTA misses an LVO (which an INR would agree needs treatment) and it is picked up by usual reporting. The product itself doesn't introduce any more harm than is there already." The EAG commented that this statement may contradict a further comment made in the GIRFT Programme's submission that an INR would not routinely review AI scans: "There appears to be a failure to acknowledge that under no scenario in the NHS will an INR [interventional radiologist] look at every CTA in acute stroke patients to provide an expert read on whether they have an LVO (and this is not necessary). If the panel is suggesting it is not as good as the experts and not recommending it for the NHS; it is suggesting the NHS needs to get every scan looked at by an INR – this is just not practical nor necessary?".

Question for committee

- What conclusions about the impact of introducing AI-derived software on clinical outcomes (for example, mRS score and mortality) can be made from the available evidence?

Key consideration 6: Cost effectiveness

Description of issue

Based on outputs from the EAG's model, as produced by the NICE team, any AI-derived software caused increase in people having thrombectomies, or impact on mRS, would not need to be large for the technologies to be cost effective. But it may be unclear how likely even such a small change is (see key considerations 4 and 5). Any efficiencies that the technologies cause, compared to current care, that reduce staff time related to assessing scans are not considered in cost effectiveness estimates.

Background

The AHSN report included a section describing methods for an assessment of cost effectiveness that will be done, based on changes in appropriate use of thrombectomy impacting length of stay in hospital, social care costs, and mRS, using a regression analysis. The EAG described this in section 4.1 of the addendum and raised several areas of concern.

The EAG's model (see page 7 for a recap of this) was used by NICE to illustrate the changes in thrombectomies and mRS at 90 days needed for the introduction of AI-based tests to be cost effective. The EAG have not confirmed these outputs.

Adding AI-derived software to clinician review improved health in the model by increasing the number of people who had a thrombectomy. The change in this could be varied in the EAG's model by changing either the sensitivity or the proportion of people who were eligible for thrombectomy for AI plus clinician, compared to clinician review alone. The benefit of greater numbers with eligible LVOs getting thrombectomy was an improvement in mRS at 90 days.

In the EAG's base case (the proportion having thrombectomy if CT scans were reviewed by clinician alone was 17.6%):

- If adding AI-derived software increased the proportion of people with an ischaemic stroke suspected of LVO having thrombectomy by about 0.11 percentage points (that is to about 17.7 %), the test is cost effective at a maximum acceptable ICER of £20,000 per QALY (0.0015 increase in QALYs, £29 increase in cost). This changes the proportion with mRS 0 to 3 at 90 days from 74.74% (clinician alone) to 74.78% (AI plus clinician).

If the prevalence of LVO from Andralojc et al. is used in the EAG's model instead (see key consideration 2), that is 7% prevalence of LVOs instead of 46.1% (the proportion having thrombectomy if CTs were reviewed by clinician alone is now about 2.70%):

- If adding AI-derived software increased the proportion of people with an ischaemic stroke suspected of LVO having thrombectomy by about 0.12 percentage points, the ICER is below £20,000 cost per QALY gained (0.0015 increase in QALYs, £29 increase in cost).

A greater negative impact of introducing the software on the occurrence of false positive result would adversely affect its cost effectiveness. Andralojc et al. reported higher levels of false positive results for the software than clinician review alone (32 more false positives per 300 scans). If specificity for AI plus clinician was lowered to produce this difference in false positives (assuming decision making always followed e-CTA findings), and using prevalence of LVOs at 7%, also from Andralojc et al. (specificity reduced to 82.6%, compared to 94.1% for clinician alone):

- If adding AI-derived software increased the proportion of people with an ischaemic stroke suspected of LVO having thrombectomy by about 0.25 percentage points, the ICER is below £20,000 cost per QALY gained (0.0032 increase in QALYs, £62 increase in cost). This changed the proportion with mRS 0 to 3 at 90 days from 90.40% (clinician alone) to 90.44% (AI plus clinician).

Using the base case prevalence of LVO instead (46.1% rather than 7.0%) and adjusting specificity to give the same increase in false positive results, the change in proportion having thrombectomy needed was similar.

The EAG's model used the mean cost of 4 different software packages in its analysis (£49.24). The cost per patient estimated for e-Stroke alone (Brainomix) was £51.52, and for Viz alone (Viz.ai) was £80.73. At higher costs, the increase in thrombectomies the AI-derived software causes needs to be higher for the technology to be cost effective.

In comments submitted, the NHS England GIRFT Programme stated that: "At the very least there is unlikely to be significant harm and there is significant evidence suggesting the AI support is adding value. You would only need to enable one more MT [mechanical thrombectomy] to easily pay for >1 year licence from a health economic perspective."

The EAG commented in the addendum that an assessment of cost effectiveness based on an increase in the numbers of patients undergoing thrombectomy would be problematic, because it requires assuming that:

- Additional thrombectomies undertaken are appropriate and result in clinical benefit, and that there are no associated detrimental effects for patients who do not have the procedure, and
- It would also be impossible to assess the number of any additional false positives that may be associated with an increased detection rate in the absence of appropriate accuracy data (Andralojc et al. 2023 was cited by the EAG as showing the potential for increase in false positive results).

In the addendum, the EAG stated that there are broadly 3 potential mechanisms by which using AI-derived software to support the review of CT brain scans in acute stroke could provide a cost-effective intervention (described in full in section 5.3 on page 59 of the addendum):

- If it improves the performance of clinicians who routinely interpret brain images, such that more people are correctly classified as suitable for thrombolysis or thrombectomy,
- If it reduces the time from scan to treatment for time critical interventions such as thrombectomy (and assuming it has no impact on clinical decision-making, or any negative effects of this are offset by reduced time to treatment),
- If the addition the AI-derived software technology has no effect on which patients are selected for treatment and insufficient effect on time to treatment to change outcomes but does improve workflow such that there are cost savings for the same outcomes.

Impact on image transfer

The NHS England GIRFT Programme's submission stated that the previous review failed to take on board the rapid image transfer functionality of these products, and this was suggested to reduce costs: "For say 100 referrals it took as a minimum 20 min longer to sit at a screen and wait for the NHS PACS images to come through. So counting just the time of the CSC stroke physician £200 / hour with all on costs/overheads that's about 250 per LVO we look at to get to the 100 so it's a saving of £16,000K pa for a centre taking 100 MT transfers and that's not counting any clinical benefits from more rapid decision making." No impact of the AI-derived software on staff time to interpret scans was included in the EAG's model.

The EAG stated that it cannot draw any conclusions about the availability of evidence to support time savings associated with image sharing because time to treatment decision was not an outcome included in the protocol and EAG's reports. The AHSN report described some results from a survey of clinicians (clinical discipline unspecified) at participating sites, although the EAG highlighted inconsistency in the information provided: "Of the 30 people asked, 25 (83%) said that the introduction of e-Stroke had reduced the time taken to reach a decision to proceed with MT [mechanical thrombectomy]. Of

this, 100% (6/6) CSC staff agreed and 73% (19/26) agreed." The report also included the information that, in response to the question 'In your opinion, what positive changes have happened since the introduction of e-Stroke?' 78% of CSC respondents and 65% of ASC respondents cited 'faster decision to treat'.

Software is provided as a package of different modules

The software packages consist of several AI-derived modules which assess different CT scans: non-contrast CT images (NCCT; for example, Viz ICH and e-ASPECTS), CT angiogram scans (for example, Viz LVO and e-CTA) and CT perfusion scans (for example, Viz CTP and e-CTP). Examples given are modules from software (the Viz platform and e-Stroke platform) that had studies identified in the EAG's updated review; full details of all technologies can be found in the scope.

Less data is available for test impact on thrombolysis use (key consideration 1). The EAG's model related to identifying people for thrombectomy based on software-assisted review of CT angiography brain scans for people who were last known to be well within 6 hours (decision question 2a).

Evidence identified in the initial review on using AI-derived software for thrombectomy treatment decisions for people with ischaemic stroke using CT perfusion after a CT angiography brain scan was very limited (decision question 2b; section 3.9 in draft guidance).

All 5 of the studies included in the addendum reported information relevant to research question 2a (people with an ischaemic stroke who were last known to be well within 6 hours). The AHSN work involved implementation of the Brainomix e-Stroke suite, in which e-ASPECTS and e-CTA were available at all 26 participating sites (ASCs and CSCs) and e-CTP was available at all 6 CSCs. It was not clear which, if any, of the component modules of e-Stroke were viewed and contributed to the decision to proceed with thrombectomy, so data may be relevant for decision question 2a and 2b (software-assisted

review of CT perfusion brain scans for people with an ischaemic stroke who were last known to be well within the last 6 to 24 hours). However, AHSN data indicated that only about 0.4% (243/67,810) of patients presenting with stroke received CTP (see page 35 of the addendum).

The costs of software provided by companies, and used to inform the EAG's model, were based on providing all modules. In submitted comments, the NHS England GIRFT Programme stated that AI is used to support review of non-contrasts CTs and CTA with or without CT perfusion together so trying to deconstruct impact on care per module is artificial, of limited value and it was difficult to see how such a study would be designed.

Questions for committee

- Is use of the AI-derived software likely to introduce efficiencies or save clinician time?
- What conclusions about the likely cost effectiveness of introducing any of the AI-derived software can be made?

4 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.

In the draft guidance, the committee noted that 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.

5 Authors

Anam Khan

Topic lead

Thomas Walker

Technical adviser

6 Glossary

CT angiography (CTA)

Imaging done with a dye to check the health of blood vessels in the brain and how blood flows through them.

CT perfusion (CTP)

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. CTP scans are used to determine a person's eligibility for mechanical thrombectomy more than 6 hours after onset of symptoms.

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.

Non-contrast CT (NCCT)

Imaging done without using a dye to create detailed images of the inside of the body.

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 drugs to dissolve blood clots and help restore blood flow to the brain in ischaemic stroke.