Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke: a systematic review and cost-effectiveness analysis (Addendum 2)

Addendum to a Diagnostic Assessment Report (DAR) commissioned by the National Institute for Health Research (NIHR) Evidence Synthesis Programme on behalf of the National Institute for Health and Care Excellence (NICE)

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Kleijnen Systematic Reviews Ltd in collaboration with Maastricht University

Authors Marie Westwood, Review Manager, Kleijnen Systematic Reviews (KSR) Ltd, United Kingdom (UK)
Bram Ramaekers, Senior Health Economist, Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre (MUMC), The Netherlands
Sabine Grimm, Health Economist, Department of Clinical Epidemiology and Medical Technology Assessment, MUMC, The Netherlands
Nigel Armstrong, Senior Health Economist, KSR Ltd, UK
Caro Noake, Senior Information Specialist, KSR Ltd, UK
Manuela Joore, Associate Professor Health Economics, Department of Clinical Epidemiology and Medical Technology Assessment, MUMC, The Netherlands

Correspondence to Marie Westwood
Kleijnen Systematic Reviews Ltd
[redacted]

Tel: [redacted]
Email: [redacted]

Date completed 09/08/2023

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR135996.

Declared competing interests of the authors: None declared

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

This report should be referenced as follows:

Westwood ME, Ramaekers BLT, Grimm S, Armstrong N, Noake C, Joore MA. Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke: a systematic review and cost-effectiveness analysis (Addendum). Addendum to a Diagnostic Assessment Report. Kleijnen Systematic Reviews Ltd, 2023.

Contributions of authors

Marie Westwood and Nigel Armstrong performed the update systematic review and interpretation of evidence. Bram Ramaekers, Sabine Grimm, Nigel Armstrong and Manuela Joore provided input on the plausibility of further cost-effectiveness analyses. Nigel Armstrong provided the summary and critique of the cost-effectiveness analysis plan, from a stakeholder submission. Caro Noake devised and performed the literature searches and provided information support to the project. Marie Westwood provided senior advice and support to the update systematic review. All parties were involved in drafting and/or commenting on the report.

Total word count (report body): 23,000

Changes to original report

The original EAG report was dated 25th July 2023.

Edits were made in response to Stakeholder comments, and a revised report dated 9th August 2023.

The edits are described in the following table:

Location in report	**Edit made**
Table 6, footnote e)	CSC corrected to PSC
Section 3.2.4, pg 43-44 Section 5.1, pg 58-59	Text amended to reflect additional information provided by OAHSN in response to request from NICE. From: '**However, the key difference, shown on these plots, between the pre- and post-implementation periods was in the numbers of CTA scans perfomed; pre-implementation periods overlapped with the COVID-19 pandemic period and a time period when (according to the plots) very few (0-2 per month) CTA scans were being carried out (the plot for one ASC site indicates an increase in CTA scanning during the pandemic period, the start of which (May/June 2020) coincided with the e-Stroke implementation date at this site), post-implementation CTA rates varied widely from month to month but were substantially higher (4-44 per month for plotted ASC sites and 6-296 for plotted CSC sites).[22]'** To: 'These plots indicated that the pre-implementation periods overlapped with the COVID-19 pandemic period and appeared to indicate that very few (0-2 per month) CTA scans were being carried out during this period.[22] The OAHSN subsequently clarified (in communication with NICE) that the plots in their report showed only the numbers of CTA scans that were processed by e-Stroke only. It is therefore unclear how many CTA scans were conducted in the pre-implementation periods and/or whether and how rates of CTA scanning and MT may have been affected by the COVID-19 pandemic.'
Section 5.1, pg 60	Text deleted to reflect additional information provided by OAHSN in response to request from NICE: '**Crucially, rates of CTA increased during the same period of implementation of e-Stroke so that in fact it might be the case that rates of MT per scan actually decreased.**
Appendix 3, brief description of study and reason for exclusion, pg 95-96	Detail added to entry for Mathieson 2022, to include additional information solicited from Brainomix by NICE: 'Following the submission of the EAG addendum, additional information about this study was provided by Brainomix in response to questions prepared by the NICE team. The Brainomix response indicated that: • Cases were selected to ensure a balanced mix with and without LVO (10 of each) and that LVO cases were selected from a historical registry with occlusion locations to match the distribution in a published IPD meta-analysis of the effectiveness of mechanical thrombectomy. • Twenty patient cases with either stroke or non-stroke diagnoses were selected where the patient was 18 years of age or over, CTA image quality was adequate, demographic information and follow-up imaging were available and there was

	no evidence of ICH, and that (within these selected cases) ten were chosen with LVOs and ten without LVO. These two descriptions appear to be inconsistent and it is difficult to see how a sample selected as described in the second point could have contained sufficient patients with LVO. The response also provided further detail on the randomisation process. *'Readers viewed the 20 images in two sessions, spaced by at least two weeks. At the first session, half of the cases were selected at random to be presented with decision support from AI software (e-CTA), and the remaining half were presented with no decision support. At the second session, the allocation of decision support was reversed, so that all cases were reviewed with and without decision support from e-CTA.'* It was not clear whether this randomisation procedure was performed separately for each reader and only 9/17 readers completed the second read (allocation of decision support reversed). The accuracy estimates presented were calculated by treating each reading, including repeat readings, as an independent data point; thus data from 20 patients were used to derive accuracy [redacted]; no per reader estimates were presented. Cases where the reader was uncertain and needed a second opinion were excluded from the analyses and there were fewer of these cases in the with AI data set. 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).'

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. OBJECTIVE

The overall objective of this addendum was to provide an update to our 2021 Diagnostic Assessment Report (DAR), which assessed the clinical and cost effectiveness of using artificial intelligence (AI)-derived software to support the review of computed tomography (CT) brain scans in acute stroke, in the National Health Service (NHS) setting https://www.nice.org.uk/guidance/gid-dg10044/documents/diagnostics-assessment-report.[1] This addendum is intended to be read in conjunction with that report and **not** as a stand-alone document. The scope[2] and research questions for the assessment are unchanged:

1. Does AI-derived 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?

2a. Does AI-derived software assisted review of CT angiography (CTA) brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke represent a clinically and cost-effective use of NHS resources?

2b. Does AI-derived software assisted review of CT perfusion (CTP) brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke after a CTA brain scan represent a clinically and cost-effective use of NHS resources?

2. BACKGROUND

Our 2021 DAR[1] concluded that the available evidence was not suitable 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. This was because:

- All studies that assessed the diagnostic accuracy of AI-derived software technologies evaluated these technologies as stand-alone interventions, rather than as an adjunct or aid to image interpretation by a healthcare professional (i.e., **not** as AI-derived software technologies would be used in clinical practice, as their use is recommended by the manufacturers and as specified in the inclusion criteria for this assessment). There was, therefore, no information about how the addition of AI-derived software technologies might affect the performance of human readers at the decision points specified in the three research questions.

- Observational 'before and after' studies that assessed the effects of implementing AI-derived software technologies in 'real-world' clinical settings provided limited outcomes data (for treated, test positive patients, mainly in relation to patients undergoing mechanical thrombectomy). These studies reported results indicating that some implementations (particularly automated alert/triage implementations) may be associated with a reduction in time to intervention. However, no study reported information to suggest that these reductions in time to intervention were associated with improvements in clinical outcome and it was, therefore, not clear whether the potential reductions in time to intervention that might be achieved as a result of implementing of AI-derived software technologies would translate into improved clinical outcomes in 'real-world' settings. The lack of accuracy data is also important, in the context of reduced time to intervention, because the implementation of an AI-derived software technology has the potential to change, not only the outcomes of patients who undergo thrombectomy, but also which patients are selected for thrombectomy. No information was available about outcomes for test negative patients and hence there was no information about the extent to which AI-derived software technologies, as implemented, may miss patients with the target condition(s). In addition, no information was available about the outcomes for patients who were false positive, based on the results of image interpretation assisted by AI-derived software technologies, and who would thus have been initially incorrectly flagged as candidates for thrombectomy.

- Our economic analyses did not provide evidence to prefer the AI-derived software strategy over current clinical practice. Our results indicated that if the addition of AI-derived software

assisted review, for guiding mechanical thrombectomy treatment decisions, increased the sensitivity of the diagnostic pathway (i.e., reduced the proportion of undetected large vessel occlusions [LVO's]) this may be considered cost effective. However, the sensitivity of AI-derived software assisted review when added to current clinical practice was uncertain (the estimates used in our modelling were based on clinical expert opinion).

Following consideration of the evidence by the Diagnostic Advisory Committee (23rd February 2022), the issuing of National Institute for Health and Care Excellence (NICE) Diagnostic Guidance was paused to await completion of the NHS England and NHS Improvement Getting It Right First Time (GIRFT) Programme speciality report for stroke,[3] and to allow discussions with NHS England and submission of additional evidence from stakeholders.

This report provides an update to our 2021 DAR;[1] it is an addendum, which is intended to be read in conjunction with that report and **not** as a stand-alone document. A full background to the decision problem, including details of all AI-derived software technologies included in the scope, is provided in our 2021 DAR.[1] In this addendum, we identify and appraise new published and un-published studies, including information submitted by stakeholders, giving particular consideration to the extent to which the evidence gaps (summarised above) have been addressed.

3. ASSESSMENT OF CLINICAL EFFECTIVENESS

Systematic review methods followed the principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care[4], NICE Diagnostics Assessment Programme manual[5] and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.[6]

3.1 Systematic review methods

3.1.1 Search strategy

Searches were first undertaken in 2021,[1] to identify papers examining the use of AI to diagnose acute stroke. Searches were conducted as recommended in the CRD guidance for undertaking reviews in health care and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.[4, 6]

Candidate search terms were identified from target references, browsing database thesauri (e.g., MEDLINE MeSH and Embase). Strategy development involved an iterative approach testing candidate text and indexing terms across a sample of bibliographic databases, so as to reach a satisfactory balance of sensitivity and specificity. Search strategies were developed specifically for each database and the keywords and thesaurus terms were adapted according to the configuration of each database. No restrictions on language, publication status or date were applied.

2023 Update searches

In order to identify any relevant primary studies published since the original strategies were run in July 2021 and updated in October 2021, the main Embase, MEDLINE and Northern Light conference searches (with minor amendments) were rerun in their entirety in May 2023. Results were deduplicated against the original search results and for completeness the medRxiv preprints search was also rerun limiting the date to those papers “posted between "01 Oct 2021 and 23 May, 2023":

- MEDLINE (Ovid): 1946-2023/05/22
- MEDLINE In-Process Citations (Ovid): 1946-2023/05/22
- MEDLINE Daily Update (Ovid): 1946-2023/05/22
- MEDLINE Epub Ahead of Print (Ovid): 1946-2023/05/22
- Embase (Ovid): 1974-2023/05/22
- Northern Light Life Sciences Conference Abstracts (Ovid): 2010-2023/wk18
- MedRxiv (Internet): up to 2023/05/23

Search strategies for all the resources listed above are presented in Appendix 1.

All identified references were downloaded in Endnote software for further assessment and handling. Results for the searches described above search were imported into the original 2021 project library and deduplicated against each other. Individual records within the Endnote reference library were tagged with search information, including the name of the searcher, date searched, database name and host, strategy name and iteration.

3.1.2 Inclusion and exclusion criteria

Separate inclusion criteria were developed, as part of our 2021 assessment,[1] for each of the three research questions. These criteria were unchanged for this update and are summarised in Table 1.

Table 1: Inclusion criteria

Decision question 1	**Is the use of AI-derived software to assist review of non-enhanced CT brain scans to guide thrombolysis treatment decisions for people with suspected acute stroke a clinically effective intervention?**	
Research question	**What is the diagnostic performance of AI-derived software assisted review of plain CT brain scans to rule-out ICH and to rule-in ischaemic stroke in people with suspected acute stroke?**	**What are the clinical effects of using AI-derived software assisted review of plain CT brain scans to guide thrombolysis treatment decisions in people with suspected acute stroke?**
Participants:	Adults (≥18 years old) attending a secondary care stroke centre with suspected acute stroke and who were last known to be well within 24 hours	
Interventions (index test):	AI-derived software assisted review of plain CT brain scan by a healthcare professional other than a neuroradiologist	AI-derived software assisted plain CT brain scan review by a neuroradiologist or other healthcare professional
Comparators:	AI-derived software assisted plain CT brain scan review by a healthcare professional other than a neuroradiologist, using a different AI-derived technology, or unassisted plain CT brain scan review by a healthcare professional other than a neuroradiologist	Unassisted plain CT brain scan review by a neuroradiologist or other healthcare professional
Reference standard:	Unassisted plain CT brain scan review by a neuroradiologist, or by a consensus panel	Not applicable
Outcomes:	Test accuracy (the numbers of true positive, false negative, false positive and true negative test results), for the target conditions ICH and ischaemic stroke. *Where reported, information will also be extracted on technical failure rates, time to intervention and ease of use/acceptability to clinicians	Clinical/patient-perceived outcomes: mortality, function (e.g., mRS), HRQoL, adverse events (e.g., bleed subsequent to thrombolysis), length of hospital stay. *Where reported, information will be extracted on technical failure rates, time to thrombolysis/rate of thrombolysis within the clinically appropriate time window, time in emergency department prior to admission or discharge and ease of use/acceptability to clinicians
Study design:	Diagnostic accuracy studies	All comparative study designs: study designs will be included in a hierarchical manner (RCTs, CCTs, observational studies), i.e., CCTs and observational studies will only be considered for inclusion where no RCTs are identified, or where there are concerns about the applicability (e.g., non-UK settings) or risk of bias for identified RCTs

Decision question 2a	**Is the use of AI-derived software to assist review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?**	
Research question	**What is the diagnostic performance of AI-derived software assisted review of CTA brain scans to guide thrombolysis treatment decisions in people with confirmed ischaemic acute stroke?**	**What are the clinical effects of using AI-derived software assisted review of CTA to guide mechanical thrombectomy treatment decisions in people with confirmed ischaemic stroke?**
Participants:	Adults (≥18 years old) attending a secondary care stroke centre with AIS, who were last known to be well within 6 hours	
Interventions (index test):	AI-derived software assisted CTA brain scan review by a healthcare professional other than a neuroradiologist	AI-derived software assisted CTA brain scan review by a neuroradiologist or other healthcare professional
Comparators:	AI-derived software assisted CTA brain scan review by a healthcare professional other than a neuroradiologist, using a different AI-derived technology, or unassisted CTA brain scan review by a healthcare professional other than a neuroradiologist	Unassisted CTA brain scan review by a neuroradiologist or other healthcare professional
Reference standard:	Unassisted CTA scan review by a neuroradiologist, or by a consensus panel	Not applicable
Outcomes:	Test accuracy (the numbers of true positive, false negative, false positive and true negative test results) for the target condition (LVO/occlusion of the proximal anterior circulation) *Where reported, information will also be extracted on technical failure rates, time to start of interventional procedure (insertion of catheter) and ease of use/acceptability to clinicians	Clinical/patient-perceived outcomes: mortality, function (e.g., mRS), HRQoL, procedure-related adverse events (e.g., bleed subsequent to thrombolysis), length of hospital stay *Where reported, information will be extracted on technical failure rates, time to start of interventional procedure (insertion of catheter), reperfusion rates and ease of use/acceptability to clinicians
Study design:	Diagnostic accuracy studies	All comparative study designs: study designs will be included in a hierarchical manner (RCTs, CCTs, observational studies), i.e., CCTs and observational studies will only be considered for inclusion where no RCTs are identified, or where there are concerns about the applicability (e.g., non-UK settings) or risk of bias for identified RCTs

Decision question 2b	**Is the use of AI-derived software-assisted review of CTP brain scans to guide MT treatment decisions for people with an ischaemic stroke, after a CTA brain scan, a clinically effective intervention?**	
Research question	**What is the diagnostic performance of AI-derived software assisted review of CTA and CTP brain scans to guide thrombolysis treatment decisions in people with confirmed ischaemic acute stroke?**	**What are the clinical effects of using AI-derived software assisted review of CTA and CTP brain scans to guide MT treatment decisions in people with confirmed ischaemic stroke?**
Participants:	Adults (≥18 years old) attending a secondary care stroke centre with suspected acute stroke, who were last known to be well more than 6 hours previously, but within 24 hours, and in whom ischaemic stroke has been confirmed on plain CT	
Interventions (index test):	AI-derived software assisted CTA and CTP brain scan review by a healthcare professional other than a neuroradiologist	1. AI-derived software assisted CTA and AI-derived software assisted CTP brain scan review by a neuroradiologist or other healthcare professional 2. Unassisted CTA and AI-derived software assisted CTP brain scan review by a neuroradiologist or other healthcare professional
Comparators:	AI-derived software assisted CTA and CTP brain scan review by a healthcare professional other than a neuroradiologist, using a different AI-derived technology, or unassisted CTA and CTP brain scan review by a healthcare professional other than a neuroradiologist	Unassisted CTA brain scan review by a neuroradiologist or other healthcare professional and unassisted CTP brain scan review by a neuroradiologist
Reference standard:	Unassisted CTA and CTP scan review by a neuroradiologist, or by a consensus panel	Not applicable
Outcomes:	Test accuracy (the numbers of true positive, false negative, false positive and true negative test results) for the target conditions (LVO/occlusion of the proximal anterior circulation for CTA and presence of salvageable tissue for CTP) *Where reported, information will also be extracted on technical failure rates, time to start of interventional procedure (insertion of catheter) and ease of use/acceptability to clinicians	Clinical/patient-perceived outcomes: mortality, function (e.g., mRS), HRQoL, procedure-related adverse events (e.g., bleed subsequent to thrombolysis), length of hospital stay. *Where reported, information will be extracted on technical failure rates, time to start of interventional procedure (insertion of catheter), reperfusion rates and ease of use/acceptability to clinicians

Study design:	Diagnostic accuracy studies	All comparative study designs: study designs will be included in a hierarchical manner (RCTs, CCTs, observational studies), i.e., CCTs and observational studies will only be considered for inclusion where no RCTs are identified, or where there are concerns about the applicability (e.g., non-UK settings) or risk of bias for identified RCTs
*Secondary outcomes, which are not sufficient to inform decision making in the absence of higher-level outcomes data AI = artificial intelligence; AIS = acute ischaemic stroke; CCT = controlled clinical trial; CT = computed tomography; CTA = computed tomography angiography; CTP = computed tomography perfusion; HRQoL = health-related quality of life; ICH = intracranial haemorrhage; LVO = large vessel occlusion; mRS = Modified Rankin Score; MT = mechanical thrombectomy; RCT = randomised controlled trial; UK = United Kingdom		

3.1.3 Inclusion screening and data extraction

Two reviewers independently screened the titles and abstracts of all reports identified by the 2023 update searches and any discrepancies were discussed and resolved by consensus. Full copies of all studies deemed potentially relevant were obtained and the same two reviewers independently assessed these for inclusion; any disagreements were resolved by consensus. Details of studies excluded at the full paper screening stage are presented in Appendix 3, along with reasons for exclusion.

Studies cited in materials submitted by stakeholders were first checked against the project reference database, in Endnote X20; any studies not already identified by our update searches or included in our 2021 DAR,[1] were screened for inclusion following the process described above. Full details of the data extraction process are provided in our 2021 DAR.[1]

3.1.4 Quality assessment

The methodological quality of observational 'before and after' implementation studies, included in our 2021 DAR,[1] was assessed using a topic-specific checklist, devised by the authors. New observational 'before and after' implementation studies, included in this addendum were assessed using the same checklist. Quality assessment was undertaken by one reviewer and checked by a second reviewer, and any disagreements were resolved by consensus or discussion with a third reviewer.

The results of the quality assessments are summarised and presented in tables (Section 3.2.2).

3.1.5 Methods of synthesis

This addendum provides a narrative synthesis, which follows the structure established in our 2021 DAR,[1] in order to facilitate comparisons between the two documents.

3.2 Update results for the assessment of clinical effectiveness

The update literature searches of bibliographic databases conducted for this addendum identified 1,826 unique references, after deduplication. Following initial screening of titles and abstracts, 54 references were considered to be potentially relevant and ordered for full paper screening; of these, five publications, relating to four studies were included in this addendum.[7-11] One included publication[11] was a full paper relating to a conference abstract[12] which was included in our 2021 DAR,[1] and two[9, 10] were additional publications relating another study[13] included in our 2021 DAR.; all of these new publications reported some additional data. A further eight studies[14-21] were already included in our 2021 DAR.[1] In addition to the publications identified by our searches, one unpublished

report (stakeholder submission to NICE)[22] and a related conference abstract (provided in the Brainomix submission to NICE)[23] were included in this addendum.

All remaining potentially relevant studies cited in documents submitted by stakeholders had already been identified by bibliographic database searches (either the searches undertaken for our 2021 assessment[1] or the update searches undertaken for this addendum), or did not meet the inclusion criteria specified in Table 1. Figure 1 shows the flow of studies through the review process, and Appendix 3 provides details, with reasons for exclusion, of all publications excluded at the full paper screening stage.

Our 2021 DAR included accuracy data from studies where the AI-derived software technology was evaluated as a stand-alone intervention,[1] rather than as an adjunct or aid to human interpretation (as it would be used in clinical practice, as recommended by the manufacturers and as specified in the inclusion criteria for the assessment (Table 1). Data from these studies were included because no other accuracy data were available for these technologies. However, it should be noted that these studies do **not** match the intervention specified in the scope[2] and, as discussed in our 2021 DAR,[1] they cannot provide information about how the addition of AI-derived software technologies might affect the performance of human readers at the clinical decision points specified in the three research questions. Additional data about the accuracy of AI-derived software technologies as stand-alone interventions would not inform the evidence gaps summarised in Section 2; therefore, in consultation with the NICE Diagnostic Assessment Programme team, a pragmatic decision was taken not to include new studies of this type in this addendum. Details of the 15 new studies,[24-38] which reported accuracy data for AI-derived software technologies as stand-alone interventions, are provided in Table 2; no further information from these studies is included in this addendum.

Table 2: Summary of new studies reporting accuracy data for AI-derived software technologies as stand-alone interventions

Study details	AI-derived software technology	Research question and target condition for which accuracy data are reported
Chan 2023[24]	Rapid ASPECTS and Rapid CTA	Q1, accuracy for the detection of infarction and Q2a, accuracy for detection of LVO
Chang 2021[25]	Rapid LVO	Q2a, accuracy for detection of LVO
Chang 2022[26]	Rapid LVO	Q2a, accuracy for detection of LVO
Eldaya 2022[27]	Rapid ICH	Q1, accuracy for detection of ICH
Karamchandani 2023[28]	Viz LVO	Q2a, accuracy for detection of LVO
Leer 2023[29]	Rapid AI	Q2a, accuracy for detection of LVO
Mair 2023[30]	Brainomix e-CTA	Q2a, accuracy for detection of LVO
Mallon 2022[31]	Rapid AI and Brainomix (unspecified)	Q2a, accuracy for detection of LVO

Mannix 2023[32]	Viz LVO	Q2a, accuracy for detection of LVO
Matsoukas 2022[33]	Viz ICH	Q1, accuracy for detection of ICH
Rodrigues 2022[34]	Viz LVO	Q2a, accuracy for detection of LVO
Schlossman 2022[35]	Rapid LVO and CINA LVO	Q2a, accuracy for detection of LVO
Schmitt 2022[36]	Brainomix e-Stroke	Q1, accuracy for detection of ICH
Vella 2023[37]	Brainomix e-CTA	Q2a, accuracy for detection of LVO
Vitellas 2022[38]	Viz LVO	Q2a, accuracy for detection of LVO
AI = artificial intelligence; ASPECTS = Alberta Stroke Program Early CT Score; CTA = computed tomography angiography; ICH = intracranial haemorrhage; LVO = large vessel occlusion; Q = question		

3.2.1 Overview of included studies

Three studies, reported in four publications,[7-10] assessed the effects of Viz LVO on time to intervention or clinical outcomes for patients who received mechanical thrombectomy. Both of these studies were conducted in the United States of America (USA) and concerned patients who were transferred from primary stroke centres (PSCs) to comprehensive stroke centres (CSCs). The remaining study, identified by our update searches, was a single centre study of the effects of implementing Brainomix e-ASPECTS and e-CTA on time to intervention for patients who received intravenous thrombolysis (IVT) or MT; clinical outcomes data were also reported for patients who received thrombectomy.[11] This study was conducted in the Neurology department of a University Hospital in Hungary, which was described as a 'high volume' stroke centre.

The unpublished report, from a stakeholder submission to NICE,[22] concerned a large-scale, UK implementation of Brainomix e-Stroke across 20 Acute Stroke Centres (ASC) and six CSCs across England. The Evidence Assessment Group (EAG) submitted questions (via NICE) with the aim of clarifying the data presented in this report and maximising the potential utility of these data; full details of the questions submitted are provided in Appendix 4. The responses provided,[39] indicated that three sites were excluded from the pre- post-implementation analyses due to small numbers of patients being admitted for stroke at these sites. Sites were grouped into five Integrated Stroke Delivery Networks (ISDN), which were *'reflective but not identical to their natural ISDN footprints.'* The authors of the report also stated that the submitted document was the annual report, written to fulfil their contractual obligations to the funding body (AI in Health and Care Award, managed by the Accelerated Access Collaborative (ACC)), and further noted that: *'This is an interim report to reflect the 2nd year of evaluation findings. For clarity, this report is not intended to be an academic, peer reviewed, scientific research publication. Further reports will be submitted to the ACC in September 2023 (Health Economics and Value report) and in March 2024 (Final evaluation report). Therefore, our work so far in this evaluation is not complete. We are still collecting data and will continue to do so until the end of 2023.'*[39] It was not clear, from the original report, which components of e-Stroke (e-ASPECTS, e-CTA and e-CTP) were implemented at which sites or groups of sites.[22] When asked for

clarification, the report authors provided data on the numbers of cases interpreted by e-Stroke, based on technology usage, for 24 sites; these data indicated that CT scans were available at all 24 sites, CTA at 23/24 sites and CTP at 10/24 sites; Appendix 2 of the same document reported information for 30 *'Networks, Trusts and Hospitals*' indicating that e-ASPECTS and e-CTA were available at all 30 and that e-CTP was available at 9/30.[39] It should be noted that, the authors further stated that: '*Most sites have automatic processing of images, so whilst a scan is processed by e-Stroke that doesn't necessarily mean it has been viewed or used to determine a diagnosis. We have qualitative data on what proportion of cases e-Stroke is used for which can be provided later if deemed useful by the panel.'*[39] The majority of the report focused on the care pathway for patients with LVO and some time to treatment and clinical outcomes data were reported for patients who received MT.[22] Some time to IVT data were also reported, which appeared to be for the wider ischaemic stroke population.[22] The report also included some information, from a survey of clinicians, which may be considered relevant to the outcome 'ease of use/acceptability to clinicians' specified in the inclusion criteria for this assessment (Table 1). The related conference abstract[23] reported 3-month Modified Rankin Score (mRS) data for patients (proportion of MT patients with 3-month mRS ≤2), presenting at one of the participating PSCs, who received MT; time to treatment data were not reported (door-in-door-out [DIDO] and time to referral only).

We did not identify any new studies that evaluated the remaining AI-derived software technologies described in Section 2.2 of our 2021 DAR.[1]

As was the case for our 2021 DAR,[1] our update searches did not identify any studies which reported information about the accuracy of any AI-derived software technology as an adjunct or aid to human interpretation (as it would be used in clinical practice, as recommended by the manufacturers and as defined *a priori* in the scope for this assessment[2] and in the inclusion criteria in the published protocol.[40] The unpublished report, from a stakeholder submission to NICE[22] listed eight evaluations questions, which had been agreed with clinical stakeholders, and these included three questions relating to accuracy or test characteristics:

- *'How accurate is the technology in a real-world deployment environment?'*
- *'In a real-world setting, does the technology perform technically as described in a research setting?'*
- *'Is each module of the technology compliant with the most appropriate reference standard for accuracy and safety?'*

However, the report did not include any measures of accuracy or describe any plan to measure the accuracy of the technology in the context of this implementation. The information included in the report, under these question headings, did not meet the inclusion criteria for outcomes specified in the scope for this assessment:[2]

Under the heading *'How accurate is the technology in a real-world deployment environment?'* the report included only information from a survey of clinicians (number surveyed not reported, total number of questions and survey methods and content unclear, number of respondents 32) for two questions: *'Do you have any concerns about the accuracy of e-Stroke?'* and *'Which functionality do you have concerns about?'*

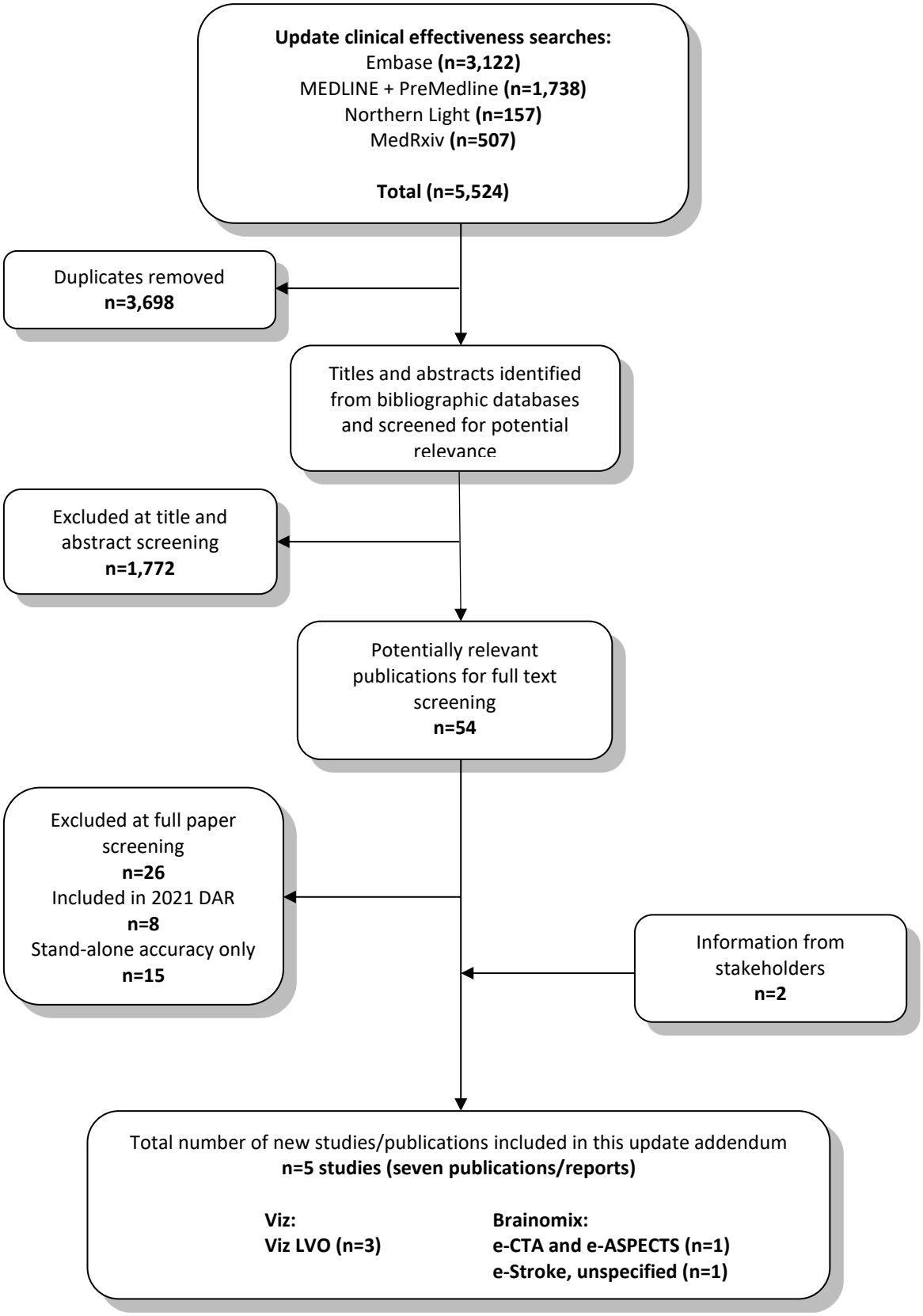
Under the heading *'In a real-world setting, does the technology perform technically as described in a research setting?'* the report included only information from a survey of clinicians (number surveyed not reported, total number of questions and survey methods and content unclear, number of respondents not reported) for a single question *'What difference has Brainomix made to how the images are shared with your stroke network?'*; it is not clear how information from this question would be expected to inform the stated evaluation question.

Under the heading *'Is each module of the technology compliant with the most appropriate reference standard for accuracy and safety?'* the report included only the statement *'e-ASPECTS and e-CTA are certified Class 11a products by the MHRA. ISO/IEC 27001'*

Of the three published studies of Viz LVO, two did not declare any funding sources,[7, 9, 10] and the other had one author who declared research funding from Viz AI.[8] With respect to the published study of Brainomix e-ASPECTS and e-CTA, the authors declared that the e-Stroke Suite was provided by the Angels Initiative (endorsed by the European Stroke Organisation) and that the lead author is the Chief Medical Officer at Brainomix. The authors of the unpublished report indicated, in their response to questions from the EAG,[39] stated that it was an annual evaluation report in relation to an AI in Health and Care Award, that the OAHSN were appointed as the technology specific evaluation team and that Brainomix had no role in the evaluation.

Full details of the characteristics of study participants, study inclusion and exclusion criteria, AI-derived software technologies evaluated, and comparator are reported in the data extraction tables presented in Appendix 2.

Figure 1: Flow of studies through the review process



3.2.2 Study quality

All five of the studies included in this addendum were observational comparisons, either between stroke centres (within a network) with and without implementation of an AI-derived software technology,[8] or of time periods before and after implementation of an AI-derived software technology.[7, 10, 11, 22] The methodological quality of these studies was assessed using a checklist, devised by the authors, and used in our 2021 DAR to assess similar studies.[1] The results of this assessment are summarised in Table 3 and reported. **It should be noted that the authors of the unpublished report[22, 39] have stated that their report is not intended to be an academic, peer reviewed, scientific research publication.** However, the report describes interim findings from a publicly funded, pre- post-implementation, real-world evaluation of e-Stroke and we have therefore included it alongside other pre- post-implementation studies and assessed its methodological quality using the same criteria.

The five new publications identified by our update searches[7-11] were all reports of retrospective studies, which assessed the effects of implementing an AI-derived software technology in real-world settings. Three of these publications were new reports,[9-11] which provided additional data from two studies[12, 13] that were included in our 2021 DAR.[1] The methodological quality assessments for these studies have been revised as indicated by any additional information from the new publications. The unpublished report, from a stakeholder submission to NICE,[22] appeared to be a prospectively planned evaluation of a large-scale, UK implementation an AI-derived software technology, which utilised existing data from Stroke Sentinel National Audit Programme (SSNAP) to inform pre-implementation comparisons. However, the extent to which this study was prospectively planned, with data collection for *a priori* specified outcomes and following a study protocol, was unclear from the submitted report which included only an overview of methods; this overview provided only the following text in relation to quantitative analysis: *'We continue to receive clinical audit data via the Stroke Sentinel National Audit Programme (SSNAP) at a deidentified patient level. We have access to data from January 2019 and are receiving quarterly updates to this dataset. At the time of writing this report we were in receipt of data up to and including September 2022 and the analyses presented here are reflective of this period. We have used a pre-post test approach to our quantitative analysis, focussing on key intervals in the stroke delivery pathway to determine the impact of the technology to both clinical and operational outcomes. Outcome data from this source will also feed into and be used to confirm our cross-case analysis.'*[22] The following statement was included in the response to the EAG's questions: '*Data presented in the annual report has not been standardised. We have largely adopted a pre and post-test approach to analysis. Further analysis will be conducted, including standardisation of data. We applied t-test methodology to the residuals to identify outliers which have been removed from the*

data plots from page 20 onwards. Number of patients included and excluded both post and pre implementation are indicated in the figure legends.'[39] Information about exclusions was only presented for those figures which illustrated change in DIDO time, by implementation site.[22] Information provided in these figures indicated that patients with DIDO time >500 minutes were excluded from the analyses.[22] At its most extreme, this resulted in an analysis which included 11 patients and excluded two patients pre-implementation, and included seven patients and excluded five patients post-implementation.[22] No justification was provided for why for data which were part of the observed, real-world distribution of outcomes were excluded.[22] It was not clear whether there were any exclusions from any other analyses.

The two publications,[7, 8] which reported new studies not included in our 2021 DAR,[1] provided data for time outcomes (time from CTA to groin puncture and recanalization[8] or time from door to groin puncture[7] in patients receiving MT) only. As discussed in our 2021 DAR,[1] time to intervention outcomes alone are not sufficient to inform meaningful estimates of the clinical and cost effectiveness of AI-derived software technologies. 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. One of the new publications included in this addendum[11] provided clinical outcomes (90-day mRS) data for a study where the earlier publication had provided only time to intervention[12] data. The unpublished report, from a stakeholder submission to NICE,[22] provided time to treatment (scan to MT) data, however, these data were only for patients who were admitted directly to one of the six CSCs; no equivalent data were reported for patients who received MT following transfer from any of the 20 participating ASCs. The response to the EAG's questions included the following statement, in relation to this issue: '*We are measuring the effectiveness of e-Stroke to speed up clinical decision time, our specific metrics and measures have been designed with this in mind. DIDO time at acute stroke centres, is considered a better (more refined) measure than scan to MT, as the latter may be confounded by the variation seen in patient transfer time to the thrombectomy centre caused by pressures on ambulance availability. We have included scan to MT for comprehensive centres as a more accurate measure.'* It should be noted that this Diagnostic Assessment seeks to evaluate whether the introduction of AI-derived software technologies will result in changes that translate to changes in outcome for patients. The scope for this topic did not include DIDO or clinical decision times as outcomes.[2] This is because, as described in the response above, changes in these metrics may not be translatable (in the real-world) to clinically meaningful changes in time to treatment. The report included clinical outcomes data (mRS, both at discharge and at 6-months post-MT), however, it was unclear how these data were linked to any changes in time to treatment because it was not clear

which patients were included in the mRS dataset (e.g. all patients who had received MT or only who patients who presented directly to a CSC).[22] The response to the EAG's questions clarified that these data were for all patients who received MT (both transfer patients and those who presented directly to CSCs); it is, therefore, not possible to make any link between the data provided on changes in time to treatment and data provided on clinical outcomes.[39] The report also provided discharge mRS data for the subgroup of patients who received MT and who presented >6 hours after the onset of symptoms, but did not report time from scan to treatment data for this subgroup.[22] Again, the response to the EAG's questions clarified that these data were for all patients (both transfer patients and those who presented directly to CSCs).[39] With respect to the 6-month mRS data, the extent of any loss to follow-up was unclear, because the report did not include the absolute numbers of patients who received MT at participating centres during the periods studied.[22] The response to the EAG's questions indicated that 6-month mRS data were available for 213/666 patients who received MT pre-implementation and 127/652 patients;[39] these numbers represent a loss to follow-up of 68% pre-implementation and 80.5% post-implementation.

With respect to the applicability of the studies included in this addendum[7, 8, 10, 11, 22] to the current decision problem, three studies evaluated the implementation of an AI-derived software technology in the context of providing an automated alert system (i.e., **not** as specified in the scope for this assessment).[7, 8, 10] As discussed in our 2021 DAR,[1] this is a triage-type allocation, where the AI-derived software alone selects patients/images with possible LVO for clinician attention (automated alert), i.e. it is not an application where interpret all scans with the addition of information from the AI-derived software technology. One further study did not provide sufficient detail to determine how the AI-derived software technology had been implemented.[11] The unpublished report, from a stakeholder submission to NICE,[22] did not include any information about which components of e-Stroke were implemented at the participating centres or any details about how the technology was implemented in practice (e.g. who used it, whether the output was used to provide an early alert or was read simultaneously as a decision aid for interpreting clinicians). The response to the EAG's questions included information about which components were implemented at which participating sites (see Section 3.2.1).[39] Details of additional statements on the implementation of e-Stroke are provided below, along with questions for which they were provided:

EAG Q3: Please provide details of which components of e-Stroke (which of the e-Stroke tools) were implemented (at each site if there were differences between sites) – in particular, provide data on the extent of use of AI in interpretation of plain CT, CTA or CTP (number and proportion of eligible patients by site).

'*We can provide data on how many cases have been interpreted by e-Stroke by modality from technology usage data (Additional analysis NICE Assessment AI in Stroke_June23.xls). Most sites have automatic processing of images, so whilst a scan is processed by e-Stroke that doesn't necessarily mean it has been viewed or used to determine a diagnosis. We have qualitative data on what proportion of cases e-Stroke is used for which can be provided later if deemed useful by the panel.'*

EAG Q4: Please provide details of how e-Stroke was incorporated in the care pathway (at each site if there were differences between sites) – please include details of type/seniority of clinician and the number and order of clinicians if more than one involved.

'Seek clarification – do you require information on how e-Stroke has been implemented? The National Optimal Stroke Imaging Pathway (NOSIP) provides guidance on where AI should be incorporated into the pathway

https://www.strokeaudit.org/SupportFiles/Documents/Guidelines/NOSIP-master-version.aspx

Clinician will vary depending on site and we would need to seek further information on this. We do have pathways mapped for almost all sites which indicates when e-Stroke is used in the pathway. A summary of this can be provided later if deemed useful by the panel.'

EAG Q11b: In relation to the table on page 19 of the report ('Rates of MT and DIDO times all ASCs – pre- and post-implementation of e-Stroke'). Noting that DIDO was not amongst the outcomes specified in the scope and protocol for this assessment; please provide an equivalent table reporting time from scan to MT for all ACSs. When providing these data please include details of which components of e-Stroke were being used (at each site, if different) to inform judgements about transfer for MT.

'*With regards to the components used at each site to inform judgements about transfer for MT, we would need to seek further clarification from the CSC clinicians, but we can state that LVO detection (e-CTA) would be broadly used, however as this is a real-world evaluation it would be difficult to know exactly by which clinicians.'*

EAG Q12a: In relation to the table on page 25 ('CSC Scan to MT times'), please provide details of which components of e-Stroke were being used (at each site, if different) to inform judgements about MT.

'*All Comprehensive Stroke Centres have access to all components of e-Stroke. See above answer regarding the use of e-CTA.'*

Observational comparative studies provide a lower level of evidence with respect to the effects of an intervention than randomised controlled trials (RCTs). Where observational study designs are used to provide estimates of effect, it is important to control, as far as possible, for potential confounding

factors (factors other than the intervention that may affect the outcome or outcomes being assessed), for example, by matching participants in the intervention and comparator groups on key risk factors. Four of the studies included in this addendum did not report sufficient information to assess whether participants were comparable before and after the implementation of the AI-derived software technology[7, 11, 22] or between centres with and without the technology,[8] with respect to baseline demographic characteristics, co-morbid conditions and risk factors. One of these studies[11] reported that there were no changes in the care pathway, other than the implementation of then AI-derived software technology, between the two time periods assessed, and two studies did not report information about whether there were any other differences in the care pathway between centres with and without the AI-derived software technology[8] or between the time periods before and after implementation.[7] The report of a large United Kingdom (UK) implementation study[22] described an iterative process and includes the following text, in relation to changes (other than the implementation of the AI-derived software technology) occurring during the study period: *'Despite this being a multi-year real-world evaluation, which relies on an iterative approach to adapt to a fast-changing stroke landscape which continuously sees new challenges and practices evolving, our hypothesis and value proposition have remained unvaried: "e-Stroke aids the evaluation of imaging in patients with suspected acute stroke and decisions for reperfusion therapies. This leads to a reduction in disability and enhanced quality of life with associated cost savings for the Health and Social Care System". Our hypothesis relies on the assumption that the benefits of e-Stroke will be maximised through quality improvement initiatives. This is because clinical outcomes are likely to improve because of faster diagnosis and treatment which are being facilitated by the technology, but also because of improvements across the acute stroke pathway. We have now entered phase three of the evaluation. The closer we get to the end of the project, the more our efforts are being directed towards the identification of promising areas for quality improvement as our goal is to support the Integrated Stroke Delivery Networks (ISDNs) to optimize the benefits of e-Stroke, driving change and maximising impact on operational and clinical outcomes.'* The implantation dates, at individual sites, were proved in response to the EAG's questions,[39] however, the report[22] and response to questions[39] did not include any specific information about what changes to the care pathway (other than the implementation of the AI-derived software technology) may have occurred during the study period, the extent to which any such changes may be related to the implementation of the AI-derived software technology, or the extent to which any changes in the outcomes of interest may be related to components of the intervention other than the AI-derived image interpretation software (e.g. communication aids included in the e-Stroke platform, such as a mobile phone app and web user interface). The related conference abstract, which reported data for one of the participating PSCs,

reported the pre-implementation study period as 1st January 2019 to 28th February 2020 and the post-implementation study period as 1st March 2020 to 31st March 2021 and stated that there were no other changes to the care pathway over this period.[23] The remaining study reported data indicating that here were no significant differences between patients who received MT before and after implementation of the AI-derived software technology, with respect to age, sex, ethnicity, baseline National Institute of Health Stroke Scale (NIHSS) score, receipt of IVT before transfer, or co-morbid conditions (diabetes mellitus [DM], hypertension, atrial fibrillation [AF], history of stroke/transient ischaemic attack [TIA], coronary artery disease [CAD], smoking),[10] but did not report any information about changes in the care pathway, other than the implementation of then AI-derived software technology, between the two time periods assessed.

Of further note, the implementation periods in three of the studies[7, 10, 22] overlapped with the period of the Coronavirus disease 2019 (COVID-19) pandemic, which may be of relevance when considering the potential for confounding changes during the implementation period**. Of particular note, examination of the per site thrombectomy trend plots (pages 21-27 of the OAHSN report) shows a crucial difference between the pre- and post-implementation periods; pre-implementation periods gererally overlapped with the COVID-19 pandemic period and a time period when (according to the plots) very few (0-2 per month) CTA scans were being carried out (the plot for one ASC site indicates an increase in CTA scanning during the pandemic period, the start of which (May/June 2020) coincided with the e-Stroke implementation date at this site).[22]**

The studies included in this addendum were generally poorly reported. No study provided sufficient information to establish both that populations that were comparable (with respect to key baseline characteristics) 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.

Table 3: Summary of quality assessment results for observational comparative studies

Study details	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Figurelle 2023[7]	N	U	U	U	U	Y	N	N	N
Gunda 2022[11]	N	U	Y	U	Y	N	N	Y	Y
Hassan 2022[10]	N	U	Y	Y	U	Y	N	Y	Y
Matsoukas 2023[8]	N	U	U	U	U	Y	N	Y	N
OAHSN and Brainomix[22]	U	U	U	U	N	N	N	Y	Y

Questions (Q): 1. Did the study have a prospective design? 2. Did the study population include an appropriate spectrum of patients? Adults (≥18 years old) attending a secondary care stroke centre with suspected acute stroke and who were last known to be well within 24 hours Adults (≥18 years old) attending a secondary care stroke centre with AIS, who were last known to be well within 6 hours Adults (≥18 years old) attending a secondary care stroke centre with suspected acute stroke, who were last known to be well more than 6 hours previously, but within 24 hours, and in whom ischaemic stroke has been confirmed on plain CT 3. Were the criteria used to select patients for CT imaging similar, before and after the introduction of the AI intervention? 4. Were the study populations, before and after the introduction of the AI intervention, similar with respect to baseline demographic characteristics (e.g., age, male/female), co-morbid conditions (e.g., hypertension, diabetes, AF) and risk factors (e.g., smoking status, previous history)? 5. Other than the availability of AI software, was the care pathway similar before and after the introduction of the AI intervention? 6. Was the implementation of the AI intervention clearly described? (e.g., how and when it was used to assist human readers and what was the level of training and experience of the human readers) 7. Were the CT imaging criteria used to select patients for treatment (e.g., thrombectomy) clearly reported for both the periods before and after the introduction of the AI intervention? 8. In addition to time to intervention outcomes, did the study report the proportion of patients who received treatment (e.g., thrombectomy) before and after the introduction of the AI intervention? 9. In addition to time to intervention outcomes, did the study report clinical outcomes (e.g., 90-day mRS) before and after the introduction of the AI intervention?
AF = atrial fibrillation; AI = artificial intelligence; AIS = acute ischaemic stroke; CT = computed tomography; N = no; Q = question; U = unclear; Y = yes;

The unpublished report, from a stakeholder submission to NICE,[22] also included some results from a mid-term survey of clinicians at participating sites.[41] The authors have stated that their report *'is not intended to be an academic, peer reviewed, scientific research publication'*[39] and we have not undertaken a formal assessment of the methodological quality or risk of bias for the survey. The following text provides a qualitative summary of the key issues, with respect to the reliability of the information provided by the survey and its applicability to the aims of this assessment.

Potential sampling bias

The additional information, provided by the authors of the report in response to questions from the EAG, indicated the distribution of clinical disciplines and participating sites amongst 34 survey respondents (Section 3.2.4).[39] However, no information was provided about survey response rates and it was not clear who/how many people had been asked to participate in the survey or how potential participants had been selected. In addition, the majority (27/34) respondents were from clinical disciplines who do not have expertise in interpreting brain scans or conducting MT.

Potential reporting bias

A copy of the survey was provided,[41] in response to questions from the EAG; the survey included a total of 54 questions. The report included a total of 17 references to information derived from the

survey.[22] Statements about survey results were frequently not linked to specific survey question and/or included interpretations that were not supported by the data presented, e.g. *'We know, from our mid-term survey, that the team at the Royal Sussex use e-Stroke for more than 75% of cases. We can assume that sustained and consistent use of the technology at this site has facilitated and expedited treatment decision.'*[22] The survey included the question: *'What proportion of cases do you use Brainomix for?'*[41] This was presented as a multiple choice question, with the response options *'all cases'*, *'more than 75% of cases'*, *'between 50% and 75% of cases'*, *'between 25% and 50% of cases'*, *'less than 25% of cases'*, or *'never'*[41] and the survey had one respondent from the Royal Sussex.[39] Thus, the definitive statement made about the usage of e-Stroke by the team at Royal Sussex and the assumed positive consequences of this were based on the opinion of a single individual (clinical discipline not reported). Where findings were reported in relation to specific survey questions (Section 3.2.4), reporting was generally incomplete, e.g. percentages only, without absolute numbers of respondents.[22]

Potential bias in survey questions

The mid-term survey included some questions where the choice of question/wording may have introduced bias in favour of e-Stroke.[41] For example, rather than being presented with a neutral question with equal weight being given to potential advantages and disadvantages, participants were asked the following questions with no corresponding question to elicit possible negative effects or opinions:

- *'What are the benefits you are experiencing through use of Brainomix?'*
- *'In your opinion, what positive changes have happened since the introduction of Brainomix?'*
- *'Please tell us below of other ways in which Brainomix has enabled wider improvements in your stroke pathway and services.'*
- *'Please tell us of any ways Brainomix has made a difference to you personally. For example, your work life balance may have improved, or being on call out of hours is easier to manage.'*

Collection of information about usage, workflow and accuracy from a survey of clinicians

The survey included a number of questions where the use of clinical opinion is of questionable value and which would have been better informed by the collection of data during implementation.[41] In some examples, clinicians were asked questions about usage which could have been recorded directly during implementation and where the opinions of small numbers of clinicians have subsequently been used to make firm statements about wider usage patterns as described above:

- *'What proportion of cases do you use Brainomix for?'*
- *'How frequently do you agree with Brainomix?'*

- *'Do you interpret the scans first before referring to Brainomix outputs?'*

There were also a number of questions which sought clinical opinion about hard outcomes, such as rates of intervention and time to decision/intervention, which should be obtained by direct measurement:

- *'In your opinion, has Brainomix helped to identify more eligible patients for mechanical thrombectomy?'*
- *'In your opinion, has Brainomix contributed to a change in the number of CTA scans performed at your hospital?'*
- *'In your opinion, has Brainomix contributed to a change in the number of CTP scans performed at your hospital?'*
- *'In your opinion, has Brainomix reduced the time taken to **reach a decision** to administer thrombolysis?'*
- *'In your opinion, has Brainomix reduced the time taken to **start** thrombolysis?'*
- *'In your opinion, has Brainomix reduced the time taken to **reach a decision** to transfer a patient for / proceed with MT?'*
- *'In your opinion, has Brainomix changed the number of patients having thrombolysis?'*
- *'In your opinion, has Brainomix changed the number of patients undergoing MT?'*

Whilst it may be helpful to elicit clinical opinion about the potential contribution of e-Stoke to any observed change, it is important to first establish (by direct measurement) that a change in e.g., rates of intervention and time to decision/intervention outcomes has occurred which was coincident with the implementation of e-Stroke.

Finally, the survey included the question: '*Do you have any concerns about the accuracy of Brainomix?*' This question cannot provide information to address either of the two agreed evaluation questions, on the theme of accuracy, which were listed in the report:[22]

- *'How accurate is the technology in a real-world deployment environment?'*
- *'In a real-world setting, does the technology perform technically as described in a research setting?'*

3.2.3 Research question 1

Is the use of AI-derived software to assist review of non-enhanced CT brain scans to guide thrombolysis treatment decisions for people with suspected acute stroke a clinically effective intervention?

Two of the studies included in this addendum reported information relevant to research question 1; both studies assessed the effects of implementation of an AI-derived software technology (Brainomix e-Stroke) which was unclearly reported or included multiple components.[11, 22] Detailed study characteristics are provided in Appendix 2.

One study[11] reported an observational 'before and after' study, evaluating the effects on time to treatment of implementing the e-ASPECTS and e-CTA modules of Brainomix e-Stroke in a centre which did not offer thrombectomy (patients requiring thrombectomy were transferred to another unit); the results of this study are summarised in Table 4. The publication stated that *'Although no changes other than the introduction of the e-Stroke Suite were made to service delivery over the duration of the project, we cannot exclude other factors contributing to improved stroke care, such as increased public awareness of stroke and ongoing quality improvement at the department.'* The pre-implementation study period was May to December 2017 and the post-implementation study period was May to December 2018.[11] e-ASPECTS analyses non-contrast CT scans for clot detection, signs of hypodensity and generates a heat map of regional ischaemic change, volume of the change and an automatic ASPECTS score, and e-CTA analyses CTP scans to generate perfusion summary maps, report parameters such as mismatch volume and ratio, hypoperfusion intensity ratio, and assesses eligibility for mechanical thrombectomy, hence, only the implementation of e-ASPECTS is relevant to research question 1. Patients were eligible for IVT if they presented within 4.5 hours of symptom onset, had no evidence of haemorrhage or other contraindications, and did not have definite hypodensity >2/3 of the major cerebral artery (MCA) territory.[11] The proportion of patients receiving thrombolysis was 46/399 (11.5%) before implementation (time to IVT was not recorded for two patients) and 72/398 (18.1%) after implementation;[11] Odds ratio (OR) 1.69 (95% confidence interval [CI]: 1.137 to 2.257). For patients receiving IVT, the mean time from door to treatment was 44±18 minutes before implementation and 42±20 minutes after implementation.[11] There was a statistically significant increase in the proportion of stroke patients who recieved IVT, following the implementation of the AI-derived software technology, (*p=0.009*), however, implementation was not associated with any significant change in the time to delivery of thrombolysis.[11]

The unpublished report[22] reported per site data (for all 26 participating sites) on the median time from scan to IVT pre- and post-implementation of e-Stroke. No ranges or interquartile range (IQR) values were reported for these data; the response to the EAG's request for IQR or range was : '*Data presented in the annual report has not been standardised. We have largely adopted a pre and post test approach to analysis. Further analysis will be conducted, including standardisation of data.'* [39] Overall, the simple comparison of median values presented (with no consideration of variance),

indicated that the time from scan to IVT increased following implementation of e-Stroke at 19/26 sites and decreased at 6/26 sites, with the remaining site having no patients who received IVT in the pre-implementation period. The report also provided per site data (for all 26 participating sites) on the rates of IVT pre- and post-implementation of e-Stroke, however, absolute numbers were not provided, and it was unclear whether the reported rates were for all presenting patients or for scanned patients. The reported rates of IVT ranged from 0% to 23.8% pre-implementation of e-Stroke and from 2.72% to 25.2% post-implementation of e-Stroke; the reported values indicated that the rate of IVT decreased following implementation of e-Stroke at 17/26 sites and increased at 9/26 sites.[22] The response to the EAG's questions[39] indicated that e-ASPECTS and e-CTA were available at all participating sites, but that no data were available about how frequently e-Stroke outputs were viewed or used to determine a diagnosis. Additional information provided about the numbers of presenting patients, in the dataset, who received CT, CTA and CTP indicated that 29,483/67,810 (43%) received CT, 5,463/67,810 (8%) received CTA and 243/67,810 (0.4%) received CTP.[39] The report did not include any information about clinical outcomes, in relation to this data set (all patients who received IVT),[22] and the response to the EAG's questions stated that no information was available regarding haemorrhage rates.[39] The report did not include any information about patient characteristics;[22] the response to the EAG's questions included the following statement: '*It is unlikely that the patient populations at site level would change from pre to post implementation'* and no supporting evidence was provided. With respect to the possible confounding effects of the COVID-19 pandemic on the care pathway (EAG Q6), the response stated: '*We don't have a full picture from all sites at present, this will be reported in our final evaluation report.'*

Table 4: Effects of implementing AI-derived software technologies on time to IVT for patients with ischaemic stroke

Study details	AI-derived software technology	Time to treatment outcome	Pre-implementation	Post-implementation	Difference	Clinical outcome	Pre-imple-mentation	Post-imple-mentation
Gunda 2022[11]	Brainomix e-ASPECTS and e-CTA	Mean (sd) minutes from door to needle, (IVT)	44 (18), (n=44)	42 (20), (n=72)	MD (95% CI) -2 (-9.05, 5.05)[a], *p=0.57*	NR	NA	NA
OAHSN[22, 39]	Brainomix e-ASPECTS and e-CTA	Median minutes from scan to IVT (one data set for each of 26 participating sites)	26.5 (n=166)	25 (n=45)	-1.5	NR	NA	NA
			37 (n=67)	38 (n=95)	1			
			43 (n=111)	48 (n=87)	5			
			49 (n=75)	44.5 (n=126)	-4.5			
			NA (n=0)	30 (n=268)	NA			
			41 (n=168)	47 (n=53)	6			
			46.5 (n=94)	28 (n=214)	-18.5			
			22 (n=48)	19 (n=72)	-3			
			41 (n=83)	51 (n=56)	10			
			27 (n=132)	28 (n=166)	1			
			30.5 (n=68)	50 (n=5)	19.5			
			16.5 (n=200)	17 (n=436)	0.5			
			38 (n=169)	47 (n=74)	9			
			31 (n=165)	32.5 (n=134)	1.5			
			35 (n=159)	29 (n=193)	-6			
			35 (n=130)	36 (n=100)	1			
			44 (n=157)	52.5 (n=34)	8.5			
			41 (n=94)	56 (n=10)	15			
			35 (n=170)	38 (n=74)	3			
			25 (n=135)	33 (n=245)	8			
			33 (n=197)	37 (n=119)	4			
			31 (n=149)	21 (n=315)	-10			
			22 (n=403)	25 (n=112)	3			
			39.5 (n=82)	71 (n=43)	31.5			
			15 (n=371)	23 (n=184)	8			
			16 (n=388)	18 (n=123)	2			

[a] Calculated value; ASPECTS = Alberta Stroke Program Early CT Score; CI = confidence interval; CTA = computed tomography angiography; IVT = intravenous thrombolysis; MD = mean difference; NA = not applicable; NR = None reported; sd = standard deviation

3.2.4 Research question 2a/2b

Is the use of AI-derived software to assist review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?

Is the use of AI-derived software assisted review of CT perfusion brain scans to guide mechanical thrombectomy treatment decisions for people with an ischaemic stroke, after a CTA brain scan, a clinically effective intervention?

All five of the studies included in this addendum reported information relevant to research question 2a.[7, 8, 10, 11, 22] One study reported on a large scale UK implementation of the Brainomix e-Stroke suite, provided information (in response to EAG questions) indicating that e-ASPECTS and e-CTA were available at all 26 participating sites and that e-CTP was available at all six CSCs.[39] However, with respect to the time to the outcomes data reported in this Section, it was not clear which (if any) of the components of e-Stroke were viewed and contributed to the decision to proceed with MT; outcomes data for patients who received MT, provided in this report, may be relevant to either or both of research questions 2a or2b.[22]

Two studies assessed the effects of implementing Brainomix e-ASPECTS and e-CTA[11] or Brainomix e-Stroke,[22] and three studies assessed the effects of implementing Viz LVO.[7, 8, 10]

All five studies reported time to treatment outcomes,[7, 8, 10, 11, 22] three studies also reported clinical outcomes (mRS),[10, 11, 22] and one study provided some additional information about the ease of use/acceptability of the AI-derived software technology to clinicians. [22]

The results of studies in this Section are grouped by AI-derived software technology. Detailed study characteristics are provided in Appendix 2.

Viz LVO

Three observational 'before and after' studies,[7, 8, 10] reported in four publications,[7-10] provided information about the effects of implementing Viz LVO in clinical settings (see Table 5). All of these studies concerned the implementation of Viz LVO, in the context of providing an automated alert system (i.e., **not** as specified in the scope for this assessment). Two studies included patients were patients who were transferred from PSC to CSC for MT, in a 'hub and spoke' system,[8, 10] and the remaining study included both dirrect arrival patients and patients who were transferred as part of a brain emergency telemedicine initiative.[7] Data from all three studies indicated that the implementation of Viz LVO, in this triage-type application, was associated with statistically significant reductions in the time from CTA to arrival at the CSC,[9] time from CTA to groin puncture[8] or time from door to groin puncture[7] (see Table 5). However, only one of these studies reported any clinical outcomes data.[9, 10] For this study, the time from CTA to arrival at CSC data were reported in the conference abstract only and for 28 patients pre-implementation and 15 patients post-implementation;[9] clinical outcomes were reported in both the conference abstract[9] and, for a larger (n=35) post-implementation group, in the full paper.[10] The full publication also reported ORs for clinical outcomes, adjusted for age, use of IVT and AF.[10] There were no statistically significant differences, pre- to post-implementation of Viz LVO, in the rates of good functional outcome (mRS ≤2) at discharge, good reperfusion (modified thrombolysis in cerebral infarction [mTICI] score 2b-3), symptomatic haemorrhage, or discharge mortality, using either adjusted or unadjusted measures, reported in either publication (see Table 5).[9, 10] The conference abstract reported results indicating that the implementation of Viz LVO was associated with small reductions in the length of hospital stay (mean difference [MD] -2.5 [95% CI: -4.712 to -0.288] days) and the length of stay in neurology intensive care unit (ICU) (MD -3.5 [95% CI: -5.205 to -1.795] days).[9] However, the larger data set, in the full paper, reported these outcomes as median (IQR) and found no significant differences pre- to post-implementation. [10]

Table 5: Effects of implementing Viz LVO for the analysis of CTA in patients with AIS, who are potential candidates for thrombectomy

Study details	Time to treatment outcome	Pre-implementation	Post-implementation	MD (95% CI)	Clinical outcome	Pre-implementation	Post-implementation	MD (95% CI) or OR (95% CI)
Figurelle 2023[7]	Median time from door to groin puncture for direct arriving patients, minutes	127	86	*p=0.006*	NR	NA	NA	NA
	Median time from door to groin puncture for telemedicine transfer patients, minutes	42	28	*p=0.036*	NR	NA	NA	NA
Hassan 2022[10]	NR	NA	NA	NA	90-day mRS ≤2	8/28	14/35	Unadjusted OR 0.600 (0.207, 1.174) OR adjusted for age, IVT use and AF 0.743 (0.191, 1.397)
					mTICI score 2b-3	19/28	23/35	Unadjusted OR 0.826 (0.182, 3.75) OR adjusted for age, IVT use and AF 0.434 (0.078, 2.433)
					Symptomatic haemorrhage	2/28	2/35	Unadjusted OR 0.788 (0.104, 5.98)

								OR adjusted for age, IVT use and AF 1.526 (0.852, 3.625)
					Mortality at discharge	6/28	5/35	Unadjusted OR 0.611 (0.165, 2.261) OR adjusted for age, IVT use and AF 0.937 (0.846, 1.037)
					Median (IQR) Hospital stay (days)	9 (6, 12), (n=28)	6.5 (3.5, 11), (n=35)	*p=0.119*
					Median (IQR) Neurology ICU stay (days)	5.5 (3, 8)	4 (2, 7.5)	*p=0.221*
Hassan 2020[9]	Mean (sd) transfer time (time from CTA to arrival at CSC), minutes	171.29 (110.58)	105.27 (62.09)	-66.020 (-117.644, -14.396)[a]	Discharge mRS ≤2	8/28	6/15	OR 1.67 (0.446, 6.232)[a]
					mTICI score 2b-3	24/28	13/15	OR 1.08 (0.174, 6.731)[a]
					Symptomatic haemorrhage	2/28	1/15	OR 0.93 (0.077, 11.165)[a]
					Mortality at discharge	6/28	4/15	OR 1.33 (0.310, 5.727)[a]
					Mean (sd) Hospital stay (days)	9.7 (4.9)	7.2 (2.5)	MD -2.5 (-4.712, -0.288)[a]
					Mean (sd) Neuro ICU stay (days)	6.4 (3.8)	2.9 (1.6)	MD -3.5 (-5.205, -1.795)[a]
Matsoukas 2023[8]	Median (IQR)[b] time from peripheral site CTA to AP, minutes	234 (99.8), (n=40)	146 (53), (n=38)	*P<0.001*	NR	NA	NA	NA

	Median (IQR)[b] time from peripheral site CTA to recanalization	253.5 (86), (n=40)	198 (25), (n=38)	*P<0.001*	NR	NA	NA	NA

[a] Calculated value

[b] Variance measure unclear; single number reported as IQR

AF = atrial fibrillation; AIS = acute ischaemic stroke; AP = arterial puncture; CI = confidence interval; CSC = comprehensive stroke centre; CTA - computed tomography angiography; ICU = intensive care unit; IQR = inter-quartile range; IVT = intravenous thrombolysis; LVO = large vessel occlusion; MD = mean difference; mRS = modified Rankin Score; mTICI = modified thrombolysis in cerebral infarction; NA = not applicable: NC = not calculable; NR = none reported; OR = odds ratio; sd = standard deviation

Brainomix e-Stroke

One study[11] reported an observational 'before and after' study, evaluating the effects on time to treatment of implementing the e-ASPECTS and e-CTA modules of Brainomix e-Stroke in a centre which did not offer thrombectomy (patients requiring thrombectomy were transferred to another unit); the results of this study are summarised in Table 6. The publication stated that *'Although no changes other than the introduction of the e-Stroke Suite were made to service delivery over the duration of the project, we cannot exclude other factors contributing to improved stroke care, such as increased public awareness of stroke and ongoing quality improvement at the department.'* The pre-implementation study period was May to December 2017 and the post-implementation study period was May to December 2018.[11] Patients were eligible for thrombectomy if they had no evidence of haemorrhage or other contraindications, and had an ASPECTS score >5 and occlusion of the intracranial internal carotid artery (ICA), MCA, or basilar artery (BA).[11] The proportion of patients receiving thrombectomy was 11/399 (2.8%) before implementation and 19/398 (4.8%) after implementation,[11] OR 1.77 (95% CI: 0.83 to 3.77). For patients receiving thrombectomy, the mean time from scan to groin puncture was 174±80.5 minutes before implementation and 145±28 minutes after implementation,[11] MD -29.00 (95% CI: -78.21 to 20.21) minutes. Ninety-day mRS data were available for 10/11 patients who received thrombectomy before implementation and for 18/19 patients who received thrombectomy after implementation.[11] The numbers of patients who achieved a good functional outcome (mRS 0 to 2) or an excellent functional outcome (mRS 0 to 1) were reported (Table 6); no information was provided about the distribution of mRS scores in those patients who did not acheive at least a good functional outcome.[11] There were no statistically significant differences in the proprtion of patients who received thrombectomy (*p=0.57*), the time from scan to groin puncture (*p=0.29*), or the proportion of patients achieving a good (*p=1.0*) or excellent (*p=0.55*) functional outcome at 90-days post-thrombectomy, between the periods before and after implementation of the AI-derived software technology.[11]

The unpublished report, from a stakeholder submission to NICE,[22] concerned a large-scale, UK implementation of Brainomix e-Stroke across 20 ASCs and six CSCs. Additional information, provided in response to questions from the EAG, indicated that e-ASPECTS and e-CTA were available at all participating centres (ASC and CSC) and that e-CTP was available at all CSC, but that it was not known in how many cases e-Stroke outputs were viewed or used in clinical decision making.[39] No information was provided about the characteristics of patients in either evaluation period. Changes to stroke care, other than the implementation of the AI-derived software technology, were dynamic and ongoing throughout the evaluation period, as indicated by the statement: *'Despite this being a multi-year real-world evaluation, which relies on an iterative approach to adapt to a fast-changing stroke landscape*

which continuously sees new challenges and practices evolving, our hypothesis and value proposition have remained unvaried: "e-Stroke aids the evaluation of imaging in patients with suspected acute stroke and decisions for reperfusion therapies. This leads to a reduction in disability and enhanced quality of life with associated cost savings for the Health and Social Care System". Our hypothesis relies on the assumption that the benefits of e-Stroke will be maximised through quality improvement initiatives. This is because clinical outcomes are likely to improve because of faster diagnosis and treatment which are being facilitated by the technology, but also because of improvements across the acute stroke pathway. We have now entered phase three of the evaluation. The closer we get to the end of the project, the more our efforts are being directed towards the identification of promising areas for quality improvement as our goal is to support the Integrated Stroke Delivery Networks (ISDNs) to optimize the benefits of e-Stroke, driving change and maximising impact on operational and clinical outcomes."[22] This statement appears to assume that any beneficial changes, accross the acute care pathway, that may occur during or subsequent to the implementation period are attributable to effects of e-Stroke. Additional information indicated that the date of implemenation varied across participating sites, ranging from 25th February 2020 to 25th October 2021 and overlapping with the period of the COVID-19 pandemic.[39] The time periods over which data were collected also varied between sites (from 13 to 33 months for the pre-implementation period and from 12 to 31 months for the post-implementation period).[22] Additional information, provided in response to questions from the EAG, indictated that the overall numbers of patients receiving MT were similar in the periods before and after the implementation of e-Stroke (666 versus 652),[39] however, statistical comparison of overall MT rates was not possible because the total numbers of presenting patients, before and after the implementation of e-Stroke, were not provided. Data presented in the OAHSN report, indicated that there were small increases in the rates of MT, between 0.57% and 3.46%, following implementation of e-Stroke for patients presenting at 14/16 ASC sites and between 1.42% and 6.51% for patients presenting directly to 5/6 CSC sites.[22] However, the numbers of presenting/imaged patients were not reported, for either of these data sets, and there were substantial differences in the absolute numbers of MT, pre- to post-implementation; for the 16 ASC sites for which data were provided the total numbers or MT were 192 pre-implementation and 512 post-implementation, and for the six CSC sites the reported total numbers of MT were 713 pre-implementation and 451 post-implementation.[22] Examination of the per site thrombectomy trend plots (pages 21-27 of the OAHSN report) shows that the numbers of patients presenting with stroke were similar pre- to post-implementation and that the reported numbers were consistant with the rates of thrombectomy presented. These plots indicated that the pre-implementation periods overlapped with the COVID-19 pandemic period and appeared to indicate that very few (0-2 per month) CTA scans were being carried

out during this period.[22] The OAHSN subsequently clarified (in communication with NICE) that the plots in their report showed only the numbers of CTA scans that were processed by e-Stroke only. It is therefore unclear how many CTA scans were conducted in the pre-implementation periods and/or whether and how rates of CTA scanning and MT may have been affected by the COVID-19 pandemic. The per site numbers of MT were not consistent with the information provided about the overall total numbers of MT performed in pre- (n=666) versus post- (n=652) implementation periods.[39] Additional information about the timing of MT, provided in response to questions from the EAG, indictated that the proportion of patients receiving MT who had presented >6 hours after the onset of symptoms increased form the period before 197/666 (29.6%) to the period after the 251/652 (38.5%) the implementation of e-Stroke,[39] OR 1.49 (95% CI: 1.185 to 1.874). The report included some, very limited, information about time from scan to MT; median times were reported, by CSC, before and after implementation of e-Stroke (see Table 6).[22] No estimates of variance were provided and the simple comparison of median values presented (with no consideration of variance), indicated that the time from scan to MT increased following implementation of e-Stroke at 4/6 sites and decreased at 2/6 sites.[22] It should be noted that these data were for patients presenting directly to the CSCs only; the report did not include any data to inform the question of whether the implementation of e-Stroke may be associated with any change in time to treatment metrics for patients who were transferred from ASCs. No clinical outcomes (e.g. mRS) were reported for this patient group; all mRS data, included in the report, were for the whole treated group (patients who received MT, irrespective of whether they presented directly to a CSC or were transferred from an ASC).[22] The mRS data included in the report[22] were generally poorly reported and, when compared with the document[39] and spreadsheet[42] provided in response to questions from the EAG, showed substantial inconsistencies (see Table 6). The mRS rates presented in the report[22] used the numbers of MT patients for whom mRS data were available (rather than the total number of MT patients) as the denominator. In the case of the 6-month mRS data set, for all patients who received MT, the proportion of MT patients for whom these data were missing was very high (453/666 before implementation and 525/652 after implementation)[39] and was higher for the post-implementation period compared to the pre-implementation period (OR 1.94 [95% CI: 1.509 to 2.504]). Based on the data provided in response to questions from the EAG,[39] which were consistent between the document[39] and spreadsheet provided,[42] and using the total number of MT patients as the denominator, the proportion of patients who achieved a good functional outcome (mRS ≤2) 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]), see Table 6. Using the same data set,[39, 42] with the number of MT 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 ≤2) 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). All pre- and post-implementation comparisons of 6-month mRS data should be viewed with extreme caution, given that data for this outcome were missing for the majority of MT patients. The report also included discharge mRS data for the subgroup of patients who received MT >6 hours after the onset of symptoms.[22] As with the 6-month mRS data, there were substantial inconsistencies between the data presented in the report[22] and that which was provided in response to questions from the EAG.[39, 42] As might be expected, the rates of missing data were substantially lower for the discharge mRS data set than for 6-month mRS and were similar (approximately 5%) pre- and post-implementation (see Table 6). Based on the data provided in response to questions from the EAG,[39] which were consistent between the document[39] and spreadsheet provided,[42] and using the total number of patients who received MT >6 hours as the denominator, there were non-significant decreases in both the proportion of patients who achieved a good functional outcome (mRS ≤2) at discharge (OR 0.83 [95% CI: 0.50 to 1.365]) and the mortality rate at discharge (OR 0.81 [95% CI: 0.503 to 1.319]) following implantation of e-Stroke (see Table 6). The conference abstract, which reported data from one of the CSCs involved in this implementation,[23] reported that before implementation 19/22 (86%) of patients referred for MT were transferred and after implementation 21/25 (84%) were transferred. A comparison of the proportion of MT patients who achieved a good functional outcome (mRS ≤2) after 3 months, indicated that this was higher, 10/21 (48%), after implementation than before implementation, 3/19 (16%).[23] However, the numbers of patients involved were very small and the calculated OR was not statistically significant, OR 1.67 (95% CI: 0.340 to 8.175).[23]

The unpublished report, from a stakeholder submission to NICE,[22] also included some information which may be considered relevant to the outcome 'ease of use/acceptability to clinicians', listed in the scope for this assessment;[2] no specific measures were listed for this outcome. The report included some results from a 'mid-term' survey of clinicians at implementation sites.[22] In response to the EAG's questions, the authors of the report provided information that this was a digital survey, undertaken between July and October 2022 (i.e. between 12 and 29 months after implementation, depending upon site).[39] The authors of the report also provided a copy of the full survey[41] and a breakdown of the clinical disciplines of the 34 respondents (stroke consultant n=13, consultant physician n=3, interventional neuroradiologist n=2, radiologist n=5, stroke nurse/stroke advanced care practitioner [ACP] n=8, other n=3);[39] it should be noted that no survey-derived data set, included in the report, was associated with 34 reported responses.[22] It should also be noted that the majority of survey data

(27/34 respondents) were derived from clinical disciplines who do not have expertise in interpreting brain scans or carrying out MT. The response to the EAG's questions did not provide an indication of the survey response rate (how many people were asked to complete the survey), or how survey participants were selected. Additional information, provided by the authors of the report,[39] further indicated that the 34 survey respondents were distributed across 16/26 participating sites, with 9/34 being based at one of the six CSCs.[39] On the question of acceptability to clinicians, one survey questions asked: *'How would you define the general attitude of clinicians towards e-Stroke in your site?'* This was a multiple choice question, which asked respondents to rate attitudes as *'really like'*, *'somewhat like'*, *'neutral'*, *'sceptical'*, or *'don't like at all'*.[41] The report stated that the general attitude towards e-Stroke was positive, with 21/31 (65%) or respondents *'liking or really liking' the technology'*.[22] It should be noted that this question asked survey participants to provide a response on behalf of their site and that the distribution of responses was such that multiple respondents from the same site may have given conflicting responses. When considering responses from those with experience of interpreting brain images and performing MT (radiologists/radiographers and interventional neuroradiologists) the distribution of responses was less positive: 40% of radiologists/radiographers selected *'somewhat like'*, 40% selected *'neutral'* and 20% selected *'sceptical'* (no radiologists/radiographers selected either *'really like'* or *'don't like at all'*) and for interventional radiologists, 50% selected *'really like'* and 50% selected *'sceptical'*; the numbers of radiologists/radiographers and interventional neuroradiologists responding to this question were not reported,[22] and the additional information supplied by the authors of the report[39] did not list radiographers among the respondent categories. Again in relation to the question of acceptability, the survey asked: *'How has e-Stroke affected confidence in your decision making?'*[41] This was a multiple choice question, with response options of *'it has improved my confidence'*, *'I don't use e-Stroke in this way'*, *'not applicable to my role'*, *'there has been no change in my confidence'*, and *'I am less confident'*.[41] The report stated that, due to accuracy challenges in relation to e-ASPECT and e-CTP, *'a minority of clinicians argue that e-Stroke does not improve confidence'*. The survey results were reported as confirming that *'most clinicians'*, 15/23 (65%) felt that e-Stroke had improved their confidence.[22] When considering responses from those with experience of interpreting brain images and performing MT (radiologists/radiographers and interventional neuroradiologists): 50% of radiologists/radiographers chose *'it has improved my confidence'*, 25% *'I don't use e-stroke in this way'* and 25% *'there has been no change in my confidence'* and 100% of interventional neuroradiologists selected *'there has been no change in my confidence'*; the numbers of radiologists/radiographers and interventional neuroradiologists responding to this question were not reported.[22] The survey also included the yes/no question: *'Do you have concerns about the accuracy of e-Stroke?'*[41] and the report

stated that *'just over half of respondents to our survey (17/32) expressed concerns with the accuracy of e-Stroke'*.[22] The bar chart for this question, presented in the report,[22] indicated that for all disciplines (with the exception of consultant physicians) the majority of respondents had concerns about accuracy.[22] Considering responses from those with experience of interpreting brain images and performing MT (radiologists/radiographers and interventional neuroradiologists), 60% of radiologists/radiographers and 100% of interventional neuroradiologists had concerns about accuracy.[22] The survey further asked*: 'Which functionality do you have concerns about?'*[41] However, the report did not provide full information about responses for this question, stating only that: *'When asked which aspect or functionality they were most concerned about, e-ASPECTs overcalling (8/28) and hyperdensity identification (6/28) seemed to be the most concerning;'*[22] no numbers or proportions with concern were reported in relation to LVO identification, haemorrhage identification or perfusion.

Table 6: Effects of implementing Brainomix e-Stroke for the analysis of CTA in patients with AIS, who are potential candidates for thrombectomy

Study details	Time to treatment outcome	Pre-implementation	Post-implementation	Difference	Clinical outcome	Pre-implementation	Post-implementation	MD (95% CI) or OR (95% CI)
Gunda 2022[11]	Mean (sd) minutes from CT to groin puncture, (MT)	174 (80.5), (n=11)	145 (28), (n=19)	MD (95% CI) -29 (-78.210, 20.210)[a]	90-day mRS ≤2	6/10 (60%)	11/18 (61.1%)	OR 1.05 (0.216, 5.090)[a]
					90-day mRS ≤1	2/10 (20%)	7/18 (38.9%)	OR 2.55 (0.414, 15.653)[a]
OAHSN 2023[22, 39]	Median[b] time from scan to MT (one data set for each of 6 participating CSCs), for patients presenting directly to CSCs only (no transfer patients)	171 (n=63)	173 (n=65)	2	NR	NA	NA	NA
		73 (n=19)	63 (n=129)	-10				
		107 (n=115)	112 (n=56)	5				
		67 (n=40)	82 (n=37)	15				
		96 (n=155)	117 (n=95)	21				
		102 (n=147)	98 (n=69)	-4				
	NR	NA	NA	NA	6-month mRS ≤2, for patients who received MT at any of the six participating CSCs (direct presentation and transfer)	95/666 (14.3%)[c]	57/652 (8.7%)[c]	OR 0.58 (0.407, 0.815)[a]
						91/190 (47.9%)[d]	179/354 (50.6%)[d]	OR 1.11 (0.782, 1.583)[a]
					6-month mRS 6 (mortality), for patients who received MT at any of the six participating CSCs (direct	14/666 (2.1%)[c]	17/652 (2.6%)[c]	OR 1.25 (0.609, 2.551)[a]
						14/190 (7.4%)[d]	35/354 (9.9%)[d]	OR 1.38 (0.723, 2.633)[a]

					presentation and transfer)			
					6-month mRS **missing**, for patients who received MT at any of the six participating CSCs (direct presentation and transfer)	453/666 (68%)[c]	525/652 (80.5%)[c]	OR 1.94 (1.509, 2.504)[a]
	NR	NA	NA	NA	discharge mRS ≤2, for patients who received MT at any of the six participating CSCs (direct presentation and transfer) and who presented >6 hours after symptom onset	35/197 (17.8%)[c]	38/251 (15.1%)[c]	OR 0.83 (0.5, 1.365)[a]
						45/223 (20.2%)[d]	110/484 (22.7%)[d]	OR 1.16 (0.788, 1.718)[a]
					discharge mRS 6 (mortality), for patients who received MT at any of the six participating CSCs (direct presentation and transfer)	39/197 (19.8%)[c]	42/251 (16.7%)[c]	OR 0.81 (0.503, 1.319)[a]
						42/223 (18.8%)[d]	75/484 (15.5%)[d]	OR 0.79 (0.521, 1.198)[a]

					and who presented >6 hours after symptom onset			
					discharge mRS **missing**, for patients who received MT at any of the six participating CSCs (direct presentation and transfer) and who presented >6 hours after symptom onset	10/197 (5.1%)[c]	13/251 (5.2%)[c]	OR 1.02 (0.438, 2.381)[a]
Nagaratnam 2021[e23]	NR	NA	NA	NA	3-month mRS ≤2, for patients who received MT at one of the participating CSCs	3/19 (16%)	10/21 (48%)	OR 1.67 (0.340, 8.175)[a]

[a] Calculated value
[b] No measure of variance reported
[c] Values taken from the information provided in response to the EAG's questions
[d] Values taken from the original report
[e] Publication of data from one PSC included in the OAHSN evaluation
AIS = acute ischaemic stroke; CI = confidence interval; CSC = comprehensive stroke centre; CT = computed tomography; CTA - computed tomography angiography; EAG = Evidence Assessment Group; MD = mean difference; mRS = modified Rankin Score; MT = mechanical thrombectomy; NA = not applicable: NR = none reported; OR = odds ratio; sd = standard deviation

4. ASSESSMENT OF COST EFFECTIVENESS

The up-date searches, conducted for this addendum, did not identify any new costs or cost effectiveness studies which met the inclusion criteria for this assessment.

The additional studies, described in Section 3 of this addendum, do not provide any new data that are sufficient to support further cost effectiveness modelling.

4.1 Summary of OAHSN proposal for economic assessment of AI technology

The OAHSN report contains a section that describes what might be regarded as the methods for an assessment of cost effectiveness, Section 3.7 of the report, where it is stated that: *'We are still in the process of gathering important information on both evaluation questions relating to value, which will be addressed the health economics report due in September 2023.'*[22] It seems that the approach that will be taken is to calculate incremental net benefit (INB) in monetary terms: *'…calculated as: Impact of the technology on outcomes; multiplied by the Value of a unit improvement in those outcomes; less any extra Costs associated with the use of AI.'*[22]

Those 'outcomes' are listed as:

- 'Reductions in costs for acute care if there is a reduction in length of stay due to faster time to surgery following a stroke that would benefit from MT.

- Reductions in costs for social care if there is an increase in the proportion of patients being supported to stay at home.

- Improvements in wellbeing if there is an increase in the mRS score of patients at 90 days and subsequently due to an increase in appropriate use of MT.'

The effect of e-Stroke is stated to be estimated using regression analysis on these outcomes, reinterpreted as:

- 'Reduced length of stay - value per patient is equal to change in number of bed days per patient multiplied by cost per bed day (assessed from Hospital Episode Statistics [HES] data, which is available by site by time period).

- Higher proportion staying at home – value per patient is equal to change in proportion of patients supported to stay at home multiplied by saving in social care when patient is supported to stay at home. In terms of social care costs, we will draw on research presented

in Jones et al 2022[16] [the Personal Social Services Research Unit [PSSRU] Unit costs manual] and Luengo-Fernandez et al 2013.[17]

- Wellbeing value of improved mRS - calculated as the unit value of an improvement in mRS scores due to appropriate MT multiplied by the average improvement in mRS after implementation of appropriate MT. In doing so, we will draw on estimates in Lobotesis et al on average QALY scores per patient for outcomes at 1 year, 2 years, 5 years and lifetime (i) for stent retriever plus IV-tPA; and (ii) for IV-tPA alone.'

Although three outcomes are listed, in table on page 40 of the OAHSN report,[22] as the means of calculating INB, two more are added for estimation using regression analysis to estimate the effect of AI, time to decision and number treated with MT.[22] This table also lists *'Issues taken into account'*, which one might presume are independent variables in the regression analyses: site characteristics (IMD, rurality, scale, ASC/CSC), transfer time, patient characteristics (age, gender, co-morbidities), context (COVID, regional ambulance times), pathway variations, and AI implementation.

Five years is stated to be the time horizon.

Under the heading 'Cost information' are listed the following categories common to with or without AI: MT and inpatient care within 24 hours, which has been costed as £11,780. The "AI-augmented approach" also lists software, equipment, staff training and *'required/recommended changes to standard operating practice'* page 40 of the OAHSN report.[22]

The report also specifies *'Scenarios of value',* expressed as combinations of '*conditions*' (previous or current) and '*implementation of e-Stroke*'. The options for the latter are: '*no implementation'* or some expression of the degree of implementation of e-Stroke, the base-case using the phrase *'substantial implementation of e-Stroke – though not achieving its full potential'* pages 40-41 of the OHSN report.[22] There are also *'Optimistic'* and *'Utopian'* scenarios where implementation is described as: *'relatively feasible improvements to pathways and fuller implementation of e-Stroke'* and *'ideal improvements to pathway arrangements'* respectively, page 41 of the OAHSN report.[22]

4.2 DAG critique OAHSN proposal for economic assessment of AI technology

There is generally a lack of clarity in the proposed economic assessment, the first being that the nature of the intervention is expressed in two ways, either *'AI technology'* or *'e-Stroke'*, the latter being more

than AI and including methods of viewing images and communication. There are also multiple problems with the method of calculating value:

1) The choice of the proposed outcomes has not been justified.

2) The items are expressed in such a way as to presume that AI results in improvement. Indeed, the third item seems to omit the possibility of there being inappropriate MT due to decreased specificity, although there is no mention of accuracy of AI use at all.

3) The value of improvement in mRS appears to be in different units to that of the first two measures of outcomes i.e., quality-adjusted life years (QALYs) as opposed to monetary units, which precludes the calculation of INB as stated.

In terms of estimating the independent effect of AI, a regression-based approach might be reasonable to try to control for the effects of confounding, given the lack of randomisation to intervention and control groups. However, there are multiple problems with the approach described:

1) Time to decision is not defined, although one might assume that the decision is for a MT or transfer to CSC for MT, based on diagnosis of LVO.

2) Number treated with MT is not the same as number treated with MT appropriately: indeed, it is unclear how appropriateness could be determined. Also, given that it is to be estimated by logistic regression, presumably it should be expressed as a proportion as opposed to *'no.'*.

3) Proportion staying at home has no specified follow-up time.

4) mRS is stated to be estimated by linear regression, which is inappropriate given the categorical (non-continuous) nature of the variable: instead, it could be analysed by logistic regression using the dichotomy of good and bad mRS categories.

5) Those apparent independent variables *'Pathway variations'* and *'AI implementation'* are not explained.

Despite the stated method of regression-analysis to estimate outcome, the report also states that no improvement due to e-Stroke might be observed in length of stay and mRS, but that there is benefit in the number of appropriate MT treatments, which leads to an improvement in mRS. This assertion is then followed by a statement that: *'The approach set out here enables us to take into account (a) the outcomes of those that undergo MT using Brainomix and (b) to make allowance for those that could have undergone thrombectomy but did not,'* on page 40 of the OAHSN report.[22] It is unclear what

is meant by this statement. Most worryingly, there is a suggestion of bias in the following statement: *'A further point to note is that while we may not find a direct link when examining changes in mRS, we may find a statistically significant link in terms of time to decision – in which case we can then infer a potential for improvements in mRS from the strong evidence base that connects time to decision and mRS,'* on page 40 of the OAHSN report.[22] This implies that estimating mRS change indirectly via *'time to decision'*, will be triggered only on the observation of no direct effect on mRS. Of course, it is possible that an improvement in mRS due to e-Stroke might be disguised by confounding factors, which cannot be adequately adjusted for in the regression analysis. However, lack of observation of an effect does not imply that there is an effect or that the effect would be mediated by reduced time to decision. It is also worth noting that the reference cited for *'the strong evidence base that connects time to decision and mRS'* is the GIRFT report,[3] which contains no data on this relationship.

The scenarios seem to be incorrectly expressed as not incremental i.e., given only either with e-Stroke or without it and thus inappropriate for the estimation of INB, which is the difference between with and without e-Stroke. Also, in addition to a conflation between AI and e-Stroke, there seems to be a conflation between e-Stroke and other aspects of the care pathway. There is a general lack of clarity in the specification of scenarios in terms of degree of implementation and what is meant by *'conditions'*. It is also unclear how degree of implementation could be specified in such a way as to affect outcome: presumably this is via independent variables in the regression analyses expressed as *'AI implementation'*, although, as already stated, the precise nature of these variables is unclear.

5. DISCUSSION

5.1 Statement of principal findings

This addendum has been produced to provide an up-date to our 2021 DAR, which assessed the clinical and cost effectiveness of using AI-derived software to support the review of CT brain scans in acute stroke, in the NHS setting.[1] This addendum is intended to be read in conjunction with that report and **not** as a stand-alone document.

The evidence base has not changed substantively since completion of the 2021 DAR and the nature and significance of the evidence gaps, as described at that time,[1] is unchanged. Use of AI-derived software technologies in stroke imaging has grown rapidly; as of April 2023 AI decision support has been implemented at 99 of 107 stroke units in England with all other identified centres actively working on plans to go live before the end of 2023.[43] Despite the rapid and widespread role out of these technologies in clinical settings (99/107 stroke units in England, as of April 2023, with plans to go live in the remaining centres by the end of 2023),[43] there remains a surprising lack of evidence about either the effects of these technologies on the accuracy of human image interpretation and consequent clinical decision making or the overall effects of implementing these technologies on clinical outcomes for patients entering the stroke pathway. There is also a lack of supporting recommendations from current, national, evidence-based guidelines.[44]

The 2022 GIRFT report includes specific recommendations in relation to the use of AI-decision support tools in stroke, under the *heading 'recommendations to ensure rapid access to imaging'*:[3]

'Provide infrastructure, training and technology to share images between hospitals and clinicians to support image interpretation (see also Recommendation 9 from GIRFT's Radiology National Specialty Report - All trusts must meet the RCR standards for the use of IT).'

6c 'Increase regional availability of AI decision-support tools and training.'

6d 'Provide national support for regional roll-out of AI working closely with ISDN footprints.'

However, the GIRFT report does not provide supporting evidence for recommendations or report the use of evidence-based guidelines development method. By contrast, the 2023 up-date of the National Clinical Guideline for Stroke for the UK and Ireland, provides full details of an evidence-based guideline development process and does not include any recommendations about the use of AI-derived software technologies in diagnostic imaging for stroke.[44] The text of the guideline includes one mention of AI decision support. This is as a cautionary note in relation to trial data where RAPID AI was used to select patients, stating that extrapolation of these trial results to other AI products should not be assumed to be appropriate*: 'It should also be noted that the perfusion criteria applied in these*

trials of late-presenting patients (core defined by rCBF below 30% and penumbra by Tmax greater than 6 secs) were mostly based on the use of RAPID™ AI decision-support software from IschemaView (Stanford, USA) and direct extrapolation of these results to other AI systems should not be assumed as appropriate or equivalent to the referenced trials.'[44]

The updated systematic review conducted for this addendum resulted in the identification of five new publications, relating to four studies, which met the inclusion criteria for this assessment.[7-11] One included publication[11] was a full paper relating to a conference abstract[12] which was included in our 2021 DAR,[1] and two[9, 10] are additional publications relating another study[13] included in our 2021 DAR. All of these new publications reported some additional data. In addition to the publications identified by our searches, one unpublished report (stakeholder submission to NICE)[22] and a related conference abstract (provided in the Brainomix submission to NICE)[23] met the inclusion criteria for this assessment and are included in this addendum.

Our 2021 DAR included accuracy data from studies where the AI-derived software technology was evaluated as a stand-alone intervention,[1] rather than as an adjunct or aid to human interpretation (as it would be used in clinical practice, as recommended by the manufacturers and as specified in the inclusion criteria for the assessment. Data from these studies were included because no other accuracy data were available for these technologies. However, it should be noted that these studies do **not** match the intervention specified in the scope[2] and, as discussed in our 2021 DAR,[1] they cannot provide information about how the addition of AI-derived software technologies might affect the performance of human readers at the clinical decision points specified in the three research questions. Following discussions with NICE, it was decided that inclusion of further data of this type would not be informative and no new stand-alone accuracy data have been extracted for this addendum.

All five of the studies included in this addendum were observational comparisons, either between stroke centres (within a network) with and without implementation of an AI-derived software technology,[8] or of time periods before and after implementation of an AI-derived software technology.[7, 10, 11, 22] **It should be noted that the authors of the unpublished report[22, 39] have stated that their report is not intended to be an academic, peer reviewed, scientific research publication.** However, the report describes interim findings from a publicly funded, pre- post-implementation, real-world evaluation of e-Stroke and it was therefore assessed for inclusion, summarised and appraised using the same *a priori* criteria that were applied to all materials identified by our searches or submitted by stakeholders.

The studies included in this addendum were generally poorly reported. No study provided sufficient information to establish both that populations that were comparable (with respect to key baseline characteristics) 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.

With respect to the applicability of the studies included in this addendum to the current decision problem, three studies evaluated the implementation of an AI-derived software technology in the context of providing an automated alert system (i.e., **not** as specified in the scope for this assessment)[7, 8, 10] and one further study did not provide sufficient detail to determine how the AI-derived software technology had been implemented.[11] Only the unpublished report, from a stakeholder submission to NICE,[22] implemented an AI-derived software technology (Brainomix e-Stroke) ostensibly as a decision aid of adjunct to human decision making. However, the authors of this report have stated that: '*Most sites have automatic processing of images, so whilst a scan is processed by e-Stroke that doesn't necessarily mean it has been viewed or used to determine a diagnosis;'*[39] The report did not provide any quantitative data about the proportion of cases in which e-Stroke was viewed or used in decision making and it was not clear whether these data have been collected.[22, 39]

Regarding the use of AI-derived software technologies to guide decisions about IVT (research question 1), all of the new data included in this addendum concerned the implementation of Brainomix e-Stroke (e-ASPECTS and e-CTA) and were provided by two studies.[11, 22] Both studies assessed time from scan to needle and reported no clear difference pre- to post-implementation. One study reported sufficient information to calculate MD -2 (95% CI: -9.05 to 5.05) minutes (n=116 participants).[11] The unpublished report provided a simple comparison of median values (no measures of variance were reported), which indicated that the time from scan to IVT increased following implementation of e-Stroke at 19/26 sites and decreased at 6/26 sites, with the remaining site having no patients who received IVT in the pre-implementation period.[22] One study reported that the proportion of presenting patients who received IVT increased following implementation of e-Stroke, OR 1.69 (95% CI: 1.137 to 2.257).[11] The unpublished report provided some information about the rates of IVT pre- and post-implementation of e-Stroke, however, absolute numbers were not provided and it was unclear whether the reported rates were for all presenting patients or for scanned patients; the reported values indicated that the rate of IVT decreased following implementation of e-Stroke at 17/26 sites and increased at 9/26 sites.[22] Neither study reported any data for clinical outcomes in relation to the implementation of AI-derived software to support decision making about IVT.

Regarding the use of AI-derived software technologies to guide decisions about MT (research questions 2a and 2b), three of the studies included in this addendum provided information about the

effects of implementing Viz LVO[7, 8, 10] and two provided information about the effects of implementing Brainomix e-Stroke.[11, 22] All of the studies of Viz LVO concerned implementation in the context of providing an automated alert system (i.e., **not** as specified in the scope for this assessment). Data from all three of these studies indicated that the implementation of Viz LVO, in this triage-type application, was associated with statistically significant reductions in the time from CTA to arrival at the CSC,[9] time from CTA to groin puncture[8] or time from door to groin puncture.[7] However, only one of these studies reported any clinical outcomes data for patients receiving MT; this study found no statistically significant differences in, pre- to post-implementation of Viz LVO, in the rates of good functional outcome (mRS ≤2) at discharge, good reperfusion (mTICI score 2b-3), symptomatic haemorrhage, or discharge mortality, using either adjusted or unadjusted measures.[9, 10] One study that assessed the implementation of Brainomix e-Stroke also reported information about the effects of implementation on time from scan to MT, which was linked to clinical outcomes data for the same group of patients.[11] However, this was a very small study (n=30) and the way in which e-Stroke had been implemented was unclear; the study found no statistically significant differences, pre-to post-implementation in either the time from scan to MT (MD -29 [95% CI: -78.210 to 20.210]) or the proportion of MT patients achieving a good functional outcome (mRS score ≤2, OR 1.05 [95% CI: 0.216, 5.090]).[11] The proportion of patients receiving MT was 11/399 (2.8%) before implementation and 19/398 (4.8%) after implementation,[11] OR 1.77 (95% CI: 0.83 to 3.77).[11] The unpublished report presented some data indicating that there were small increases in the rates of MT, between 0.57% and 3.46%, following implementation of e-Stroke for patients presenting at 14/16 ASC sites and between 1.42% and 6.51% for patients presenting directly to 5/6 CSC sites.[22] However, the numbers of presenting/imaged patients were not reported, for either of these data sets, and there were substantial differences in the absolute numbers of MT, pre- to post-implementation; for the 16 ASC sites for which data were provided the total numbers or MT were 192 pre-implementation and 512 post-implementation, and for the six CSC sites the reported total numbers of MT were 713 pre-implementation and 451 post-implementation.[22] In order for these absolute numbers to be consistent with the reported increases in rates, there would need to be substantial differences in the denominator (i.e. numbers of patients presenting/imaged and/or time period over which data were collected) between the pre- and post-implementation periods. Examination of the per site thrombectomy trend plots (pages 21-27 of the OAHSN report) shows that the numbers of patients presenting with stroke were similar pre- to post-implementation and that the reported numbers were consistant with the rates of thrombectomy presented. These plots indicated that the pre-implementation periods overlapped with the COVID-19 pandemic period and appeared to indicate that very few (0-2 per month) CTA scans were being carried out during this period.[22] The OAHSN subsequently clarified (in communication with NICE) that the

plots in their report showed only the numbers of CTA scans that were processed by e-Stroke only. It is therefore unclear how many CTA scans were conducted in the pre-implementation periods and/or whether and how rates of CTA scanning and MT may have been affected by the COVID-19 pandemic. The unpublished report included some, very limited, information about time from scan to MT; median times (with no measure of variance) were reported, by CSC, before and after implementation of e-Stroke.[22] A simple comparison of the median values presented indicated that the time from scan to MT increased following implementation of e-Stroke at 4/6 sites and decreased at 2/6 sites.[22] It should be noted that these data were for patients presenting directly to the CSCs only; the report did not include any data to inform the question of whether the implementation of e-Stroke may be associated with any change in time to treatment metrics for patients who were transferred from ASCs. No clinical outcomes (e.g. mRS) were reported for the direct presentation patient group; all mRS data, included in the report, were for the whole treated group (patients who received MT, irrespective of whether they presented directly to a CSC or were transferred from an ASC).[22] The mRS data included in the unpublished report[22] were generally poorly and inconsistently reported. In the case of the 6-month mRS data set, for all patients who received MT, the proportion of MT patients for whom these data were missing was very high (453/666 before implementation and 525/652 after implementation)[39] and was higher for the post-implementation period compared to the pre-implementation period (OR 1.94 [95% CI: 1.509 to 2.504]).[39, 42] Using the total number of MT patients as the denominator, the proportion of patients who achieved a good functional outcome (mRS ≤2) 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 the there was a non-significant increase in the 6-month mortality post-implementation (OR 1.25 [95% CI: 0.609 to 2.551]).[39, 42] The report also included discharge mRS data for the subgroup of patients who received MT >6 hours after the onset of symptoms.[22] As with the 6-month mRS data, there were substantial inconsistencies between the data presented in the report[22] and that which was provided in response to questions from the EAG.[39, 42] As might be expected, the rates of missing data were substantially lower for the discharge mRS data set than for 6-month mRS and were similar (approximately 5%) pre- and post-implementation. Using the total number of patients who received MT >6 hours as the denominator, there were non-significant decreases in both the proportion of patients who achieved a good functional outcome (mRS ≤2) at discharge (OR 0.83 [95% CI: 0.50 to 1.365]) and the mortality rate at discharge (OR 0.81 [95% CI: 0.503 to 1.319]) following implantation of e-stroke.

The unpublished report also included some information which may be considered relevant to the specified outcome *'ease of use acceptability to clinicians'*.[22] This information was derived from a small

number (maximum n=34) of responses to a digital survey of clinicians at participating sites, which was conducted between July and October 2022 and is summarised in Section 3.2.4 of this addendum.

In summary, the limited evidence available (summarised in the 2021 DAR[1] and this addendum) does not support either the hypothesis that the implementation of AI-derived software technologies (when used alongside clinical judgement, as a decision aid and **not** as an automated alert triage tool, where the AI-derived software technology alone 'triages'/selects patients for automated alert, or other stand-alone application) is associated reductions in time from scan to treatment (IVT or MT). It is unclear whether the implementation of these technologies may be associated with increased rates of intervention (IVT or MT; there is some evidence, from one published study, to support an increase in the rate of IVT and statistically non-significant increase in the rate of MT),[11] however, data from the large UK implementation evaluation indicate that implementation of e-Stroke was not associated with an increase in the rate of IVT and, whilst reported information for some sites suggested an increase in the rates of MT, the overall absolute numbers of MT were similar pre- and post-implementation.[22] Importantly, the available evidence does not support the hypothesis that implementation is associated with improved outcomes for treated patients (IVT or MT), which it has been suggested follow from workflow improvements (such as reduced time to treatment) or from other unspecified changes that may arise following implementation.[22] As was the case in the 2021 DAR,[1] there are no data comparing the pre- and post-implementation outcomes for patients entering the pathway who did not receive IVT of MT. The key question of whether or not the addition of AI-derived software technologies can improve the performance of human readers by increasing the accuracy with which patients are selected for reperfusion interventions remains unaddressed; we have not identified any studies or other submitted evidence which can provide an estimate of the accuracy of any AI-derived software technology when used as a decision aid, in combination with human readers, for detection of any of the target conditions specified for this assessment (Table 1).

The up-date searches, conducted for this addendum, did not identify any new costs or cost effectiveness studies which met the inclusion criteria for this assessment. The unpublished report, provided as a stakeholder submission to NICE,[22] included a description of planned cost-effectiveness analyses (CEAs) a summary and critique of which is provided in Section 4 of this addendum.

5.2 Strengths and limitations of assessment

We have conducted full update literature searches in an attempt to maximise retrieval of potentially relevant studies. As was the case for our 2021 DAR,[1] our searches included sources of unpublished material and all additional materials submitted by any stakeholder were assessed for inclusion against the same *a priori* criteria used for studies retrieved by literature searches.

Clear inclusion criteria were specified in the protocol for this assessment and remained unchanged for this addendum; the review has been registered on PROSPERO (CRD42021269609) and the protocol is available from https://www.nice.org.uk/guidance/gid-dg10044/documents/final-protocol. The eligibility of studies for inclusion is therefore transparent. In addition, we have provided specific reasons for exclusion for all studies which were considered potentially relevant at initial citation screening and were subsequently excluded on assessment of the full publication (Appendix 3). The review process followed recommended methods to minimise the potential for error and/or bias;[4] studies were independently screened for inclusion by two reviewers and data extraction and quality assessment were done by one reviewer and checked by a second. Any disagreements were resolved by consensus.

The main limitations for this assessment remain the paucity and nature of the available evidence.

The very limited additional evidence identified for this addendum was all derived from observational studies of the implementation of AI-derived software technologies. The information provided by studies of this type is limited in that it concerns only treated (i.e., test positive) patients; no information is provided about test negative patients, hence there is no information about the extent to which AI-derived software technologies, as implemented, may miss patients with the target condition(s), or the extent to which changes in false positive rates subsequent to implementation, may adversely affect patients (e.g. by delaying receipt of appropriate interventions) or costs (e.g. more specialist imaging review following referral for MT). In addition, no 'real-world' implementation study, included in this addendum or the 2021 DAR,[1] compared clinical outcomes along with time to intervention, in populations that were comparable (with respect to key baseline characteristics) before and after the implementation of the AI-derived software technology, and where the AI-derived software technology was the only change to the care pathway. Differences in the study population (before and after implementation) and/or additional changes in the care pathway (other than implementation of the AI-derived software technology) mean that the extent to which any observed changes in time to intervention or clinical outcome are attributable to the implementation of the AI-derived software technology is highly uncertain. Studies which report only the effects of implementation of AI-derived software technologies on time to intervention are deficient in that they do not provide the information about clinical outcomes needed to inform decision making; a reduction in time to intervention may not always be advantageous, e.g., if the time saving is associated with a detrimental effect on clinical outcomes.

This assessment did not identify any new evidence that would be sufficient evidence to support further modelling of the cost effectiveness of AI-derived software technologies to support the review of CT brain scans in acute stroke.

5.3 Uncertainties

There is a large volume of literature about AI-derived software technologies, which have been designed to support image interpretation in stroke patients. However, surprisingly few studies have attempted to address the questions that are key to determining whether these technologies are clinically effective or cost-effective interventions. Broadly, there are three potential mechanisms by which using AI-derived software to support the review of CT brain scans in acute stroke could provide a clinically effective and/or cost-effective intervention:

If the addition the AI-derived software technology improves the performance of clinicians who routinely interpret brain images, such that more patients are correctly classified in relation to their suitability for a given intervention.

This option relies on the introduction of the AI-derived software technology affecting the judgment of clinicians to change which patients are selected for a given intervention. In order to show an effect of this type, studies are needed which assess the combined accuracy of AI-derived software technologies and clinicians who routinely interpret brain scans, when these technologies are implemented as they would be used in clinical practice, (defined *a priori* in the scope for this assessment[2] and in the inclusion criteria in the published protocol[40]); we have not identified any studies of this type. Additional data are needed to support a comparison of this combined accuracy with the accuracy of those clinicians when working without the assistance of AI-derived software technologies.

The evidence about the accuracy of AI-derived software technologies as stand-alone interventions, included in the 2021 DAR,[1] is of limited use (even when it includes comparisons with the performance of clinicians alone,[45] because it provides no information about the effects on accuracy of the interaction between human readers and AI-derived software technologies. It has been stated that: *'It is important that clinicians understand when e-Stroke should and should not be relied upon and how to modify their decision-making process to take this into account.'*[22] However, it is unclear what factors would be expected to inform these judgements, and there is currently no information about how these judgements are made in practice or how they affect the performance of clinicians.

It has been suggested that *'You would only need to enable one more MT to easily pay for >1 year licence from a health economic perspective.'*[43] An assessment of cost effectiveness based on an increase in the numbers of patients undergoing MT would be problematic, in that requires both an assumption that additional thrombectomies undertaken are appropriate and result in clinical benefit,

and that there are no associated detrimental effects for patients who do not undergo MT. In the absence of appropriate accuracy data, it would also be impossible to assess the number (and associated costs) of any additional false positives that may be associated with an increased detection rate (rate of MT).

Our update searches for this addendum identified a retrospective study,[46] conducted by one of the trusts (Royal Cornwall Hospitals NHS Trust) whose data contributed to the unpublished report,[22] which compared the recorded e-CTA findings with the radiologist report in 300 consecutive scans. This study is not listed in Table 2 because it does not provide true accuracy data for e-CTA, since the reference standard (adjudication by a sub-specialist vascular radiologist) was only applied where there was a discrepancy between e-CTA and the radiologist report.[46] Assuming that all instances where e-CTA and the radiologist report agreed on the identification of an intracranial LVO were correct, and applying the reference standard adjudication in cases of discrepancy, the results of this study indicated that if decision making had always followed e-CTA findings one additional LVO would have been correctly identified (18/21 compared to 17/21 for the decision making based on the radiologist report).[46] However, this would have been associated with a substantial increase in the number of false positives (34 for e-CTA compared to 2 for the radiologist report),[46] which would be expected to have downstream cost consequences for the numbers of patients referred for specialist review prior to MT and may also have clinical consequences for patients who are incorrectly referred. The study did not meet the inclusion criteria for this assessment because it was not possible to derive an accuracy estimate for e-CTA combined with radiologist from this study; the reference standard was only applied in cases of discrepancy and, in addition, it was not clear to what extent the e-CTA report had been viewed/utilised when the radiologist reports were prepared.[46] The authors of this study noted its limitations, with respect to the provision of accuracy estimates, stating that their reported accuracy estimate for e-CTA alone was potentially biased in favour of e-CTA, since some radiologists may have agreed incorrectly with an e-CTA false positive or false negative.[46] They also highlighted the importance of conducting accuracy studies in real-world settings, were prevalence tends to be lower; the prevalence of LVO in their study was 7%,[46] compared to 53% in the study comparing human readers to e-Stroke which was included in our 2021 DAR,[45] whilst noting that the 7% figure represented only the subset of patients presenting with acute stroke who underwent CTA within the thrombectomy treatment window (unspecified).[46] There is evidence that estimates of sensitivity and specificity can vary substantially with disease prevalence; because the mechanisms of these effects are complex and may be difficult to identify it has been suggested that *'clinicians should use prevalence as a guide when selecting studies that most closely match their situation'*.[47] The authors of this study further noted that over half of the occlusions reported by e-CTA were false positives and that

arbitration by reference standard favoured e-CTA over the original radiology report in only six instances, concluding that: *'Together the results presented suggest 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, but not be unduly influenced by the software should they disagree with it.'*[46] The findings of this study further emphasise the need to understand the interaction between AI-derived software technologies and reviewing radiologists and how the introduction of these technologies may affect the accuracy of clinical judgement in real-world settings. Retrospective re-analyses of stored images from the e-Stroke implementation programme, similar to the above study[46] but with the reference standard (neuroradiologist review) applied to the whole data-set could provide valuable information on this question. However, it is unclear whether any such analyses are planned; the OAHSN report does not describe any plans to conduct analyses of accuracy.[22]

If the addition the AI-derived software technology reduces the time from scan to treatment for time critical interventions such as MT

An assessment of cost effectiveness, based on improvements in time to treatment alone would require a number of assumptions. This is because the implementation of an AI-derived software technology has the potential to change, not only the outcomes of patients who undergo thrombectomy, but also which patients are selected for thrombectomy. Hence, evidence of a beneficial effect of implementation, for patients undergoing thrombectomy, is insufficient to show clinical effectiveness. An assessment of cost effectiveness, based on improvements in time to treatment alone would require the assumption that either the AI-derived software technology has no effect at all on clinical decision making or that any changes to which patients are selected for thrombectomy are such that, when combined with an observed reduction in time to treatment, the net population effect is positive; there is currently no evidence to support either of these assumptions. With respect to the potential positive clinical effects for treated patients based on reductions in time to treatment, evidence is required that the introduction of the AI-derived software technology is associated with a consistent reduction in the time from scan to treatment which is of an order of magnitude sufficient to effect clinical outcomes. As discussed in our 2021 DAR,[1] there is evidence, from an individual-patient-data (IPD) meta-analysis[48] and a multi-centre RCT (the MR CLEAN study),[49] to indicate a negative correlation between time to intervention and functional outcome in patients with LVO who undergo thrombectomy. The results of the IPD meta-analysis indicated that earlier treatment with thrombectomy in addition to pharmacological thrombolysis was associated with lower degrees of disability, as indicated by 90 day mRS, than pharmacological thrombolysis alone and that this benefit remained statistically significant up to 7 hours and 18 minutes from onset of symptoms to

arterial puncture; each hour of reperfusion delay was associated with a reduction in the proportion of patients achieving function independence (mRS 0 to 2), absolute risk difference (ARD) -5.2% (95% CI: -8.3% to -2.1%).[48] Similarly, the MR CLEAN study reported that thrombectomy remained an effective intervention, with respect to the proportion of patients achieving functional independence, up to 6 hours and 18 minutes from onset of symptoms to arterial puncture and that the ARD for achieving a good functional outcome was reduced by 6% for every hour of delay to reperfusion.[49] The evidence about reductions in time to intervention following implementation of AI-derived software technologies is inconsistent and it remains unclear whether any potential reductions in time to intervention that might be achieved as a result of implementing of AI-derived software technologies would be of sufficient magnitude and consistency to translate into improved clinical outcomes in 'real-world' settings. 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.

The authors of the unpublished report, included in this addendum,[22] made the following statement in relation to the collection of data on time from scan to intervention: '*We are measuring the effectiveness of e-Stroke to speed up clinical decision time, our specific metrics and measures have been designed with this in mind. DIDO time at acute stroke centres, is considered a better (more refined) measure than scan to MT, as the latter may be confounded by the variation seen in patient transfer time to the thrombectomy centre caused by pressures on ambulance availability.*'[39] It should be noted that the inclusion of factors effecting transfer time in the scan to MT metric is not an example of 'confounding' but rather is a measure of the true, real-world effects of the intervention; 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 report did include some time from scan to thrombectomy data, for patients presenting directly to CSCs only (no transfer times included), and these data indicated that, even without transfer time considerations, the introduction of the AI-derived software technology did not appear to be associated with consistent reductions in time to treatment; the median time from scan to MT increased, following implementation, at 4/6 sites and decreased at 2/6.[22] As an aside, the median DIDO times provided by the report, for 16/20 ASCs, also did not indicate a consistent reduction in time following implementation (median DIDO times increased, following implementation, at 9/16 sites, were unchanged at one site and decreased at 6/16 sites).[22]

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 (cost saving only)

It has been argued that potential workflow improvements are an important feature of AI-derived software technologies and that, in relation to our 2021 DAR[1]: *'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. 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.'*[43] Whilst it is possible that the image sharing components of these technologies may be associated with improvements in workflow, the 2021 Diagnostic Assessment did not aim to assess the clinical and cost effectiveness of image sharing technologies; if a cost effectiveness evaluation were to be based on workflow-related cost savings alone, it could reasonably be argued that a wider scope (e.g. including potentially less costly image sharing technologies with no AI-derived component) would be appropriate. It is also important to remember that, in the case of AI-derived software technologies any benefits from rapid image sharing are not independent of the effects of the decision aid component of the technology. The type of cost saving described above does not consider any additional costs that may arise as a consequence of increased inappropriate referrals for MT due to false positives associated with the technology.[46] These costs are an important consideration and are dependent upon how AI-derived technologies effect clinical decision making in practice (the combined accuracy of the technology and the clinician compared to that of the clinician alone), for which there remains a lack of data.

Time to treatment decision was not an outcome specified in the scope for the 2021 DAR and this addendum.[2] Therefore, we cannot draw any conclusions about the availability of evidence to support time savings associated with image sharing. The unpublished report, included in this addendum, described some results from a survey of clinicians (clinical discipline unspecified) at participating sites: *'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. Of this, 100% (6/6) CSC staff agreed and 73% (19/26) agreed.'*[22] 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'*.[22] The report did not include any measured values for time to decision to treat and the limited information provided by the survey of clinicians appeared inconsistent.

In summary, accuracy is not the only consideration when assessing the clinical and cost effectiveness of using AI-derived software to support the review of CT brain scans in acute stroke; as with all assessments of clinical tests, it is important to be able to show a link between test characteristics and patient-perceived outcomes. However, it is important to note that other metrics (e.g., time to treatment, effectiveness of treatment) are likely to be affected by which patients are selected for treatment (i.e., accuracy). The application of AI-derived software technologies in the context of decision aids does not lessen the importance of accuracy information, but rather increases the complexity of the information required. The large volume of literature evaluating the stand-alone accuracy and reproducibility of AI-derived software technologies is of limited usefulness, other than to confirm why these technologies are not recommended as a replacement for clinical judgement. It remains crucial to understand how AI-derived software technologies interact with clinical judgement, in real-world settings, and how this may affect the selection of patients for treatment; there remains a lack of evidence to inform this question.

6. CONCLUSIONS

6.1 Implications for service provision

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

There remains a lack of information about the accuracy of AI-derived software technologies when used as decision aids, in combination with clinical judgement. This means that it is not possible to assess to what extent clinical decision making is changed by the availability of AI-derived software technology outputs or what may be the downstream consequences (e.g., for treatments and outcomes) of any such changes.

In addition to consideration of accuracy, both our 2021 DAR[1] and this addendum assessed and included published and unpublished observational comparisons that assessed the effects of implementing AI-derived software technologies in 'real-world' clinical settings. The information provided by studies of this type was limited in that it concerned only treated patients; no information was provided about test negative patients and hence there was no information about the extent to which AI-derived software technologies, as implemented, may miss patients with the target condition(s) i.e., false negatives or the extent to which false positive test results may adversely affect outcomes for some patients. In addition, the limited information available from these studies did not indicate that the implementation of AI-derived software technologies was associated with either consistent reductions in time to treatment or improved clinical outcomes.

This assessment did not identify any new evidence that would be sufficient evidence to support further modelling of the cost effectiveness of AI-derived software technologies to support the review of CT brain scans in acute stroke.

6.2 Suggested research priorities

Given the deficiencies in the evidence base, summarised in Section 6.1, studies are needed (for all AI-derived software technologies) that evaluate these technologies as they would be implemented in clinical practice.

Diagnostic cohort studies should evaluate the performance of AI-derived software technologies, when used as an adjunct/aid to human readers. Ideally such studies should compare the performance of the AI-derived software technology in combination with a human reader to that of the human reader alone, where interpretation by an experienced expert or panel of experts provides the reference

standard. Studies should be conducted in the population and setting in which the AI-derived software technology would be applied in practice such that there is a representative, real-world prevalence of the target condition (e.g., LVO). Studies of this type would allow assessment of whether and to what extent the addition of AI-derived software technologies changes the performance of human readers, in the relevant clinical context.

Observational studies, evaluating the effects of implementing AI-derived software technologies in UK clinical settings, may also be of interest. Again, the precise way in which the technologies are implemented is critical to the utility of such studies for UK decision making. Based on the scope defined for this assessment, AI-derived software technologies would need to be implemented as a real time adjunct/aid to human readers and not as e.g., an automated early alert system. Observational comparative studies provide a lower level of evidence with respect to the effects of an intervention than RCTs. Where observational study designs are used to provide estimates of effect, it is therefore important to control, as far as possible, for potential confounding factors (factors other than the AI-derived software technology that may affect the outcome or outcomes being assessed), for example, by matching participants in the intervention and comparator groups on key risk factors. It is also important that the care pathway remains unchanged, other than with respect to the implementation of the AI-derived software technology. Studies of the effects of implementation of AI-derived software technologies should measure clinical outcomes alongside intermediate outcomes such as time to intervention and should report outcomes for test negative as well as test positive patients (e.g., for the interpretation of CTA to select patients for thrombectomy, outcomes should be reported for both patients who received thrombectomy and those who did not).

Cluster-RCTs, where stroke centres are randomised to implement AI-derived software technologies or to continue with current practice, would offer a more methodologically robust approach to evaluating the effects of implementation.

The advanced stage of implementation of these technologies in the NHS in England (as of April 2023 AI decision support has been implemented at 99 of 107 stroke units in England with all other identified centres actively working on plans to go live before the end of 2023),[43] limits the potential for commissioning some types of high quality primary research (e.g. cluster-RCTs as part of phased implementation). It remains possible, however, to ensure that the potential to collect relevant outcome measures is maximised. Hard outcomes, such as accuracy or time to treatment data, should be informed by measurement rather than by survey/clinical opinion. Intermediate outcomes (e.g., time to treatment) should be chosen with consideration to their potential to effect clinical endpoints and/or costs in a real-world context. Finally, retrospective re-analyses of stored radiology reports,

collected during the implementation of these technologies in England, may have potential to inform estimates of the accuracy of AI-derived software technologies in combination with clinical judgement. It is unclear whether and to what extent any ongoing work by the OAHSN, or any other evaluations arising from the implementation of these technologies in England, will include these types of data collection and analysis; protocol and full methods have not been provided for the OAHSN work. The OAHSN report included a brief description of planned CEAs, a summary and critique of which is provided in Section 4 of this addendum.

7. REFERENCES

[1] Westwood M, Ramaekers B, Grimm S, Armstrong N, Wijnen B, Ahmadu C, et al. Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke: a systematic review and cost effectiveness analysis. A Diagnostic Assessment Report. York: Kleijnen Systematic Reviews Ltd, 2021 [accessed 20.6.23] Available from: https://www.nice.org.uk/guidance/gid-dg10044/documents/diagnostics-assessment-report

[2] National Institute for Health and Care Excellence. Diagnostics Assessment Programme: Software with artificial intelligence derived algorithms for analysing CT brain scans in people with a suspected acute stroke: final scope [Internet]. London: NICE, 2021 [accessed 12.8.21] Available from: https://www.nice.org.uk/guidance/gid-dg10044/documents/final-scope

[3] Hargroves D, Lowe D. Stroke: GIRFT programme national specialty report [PDF provided by NICE]: Getting It Right First Time (GIRFT), 2022 Available from: https://gettingitrightfirsttime.co.uk/girft-reports/

[4] Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]. York: University of York, 2009 [accessed 23.3.11] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[5] National Institute for Health and Care Excellence. Diagnostics assessment programme manual [Internet]. London: NICE, 2011 [accessed 25.5.21]. 130p. Available from: https://www.nice.org.uk/media/default/about/what-we-do/nice-guidance/nice-diagnostics-guidance/diagnostics-assessment-programme-manual.pdf

[6] Cochrane Diagnostic Test Accuracy Working Group. Handbook for DTA Reviews [Internet]: The Cochrane Collaboration, 2009 [accessed 23.3.11] Available from: http://srdta.cochrane.org/handbook-dta-reviews

[7] Figurelle ME, Meyer DM, Perrinez ES, Paulson D, Pannell JS, Santiago-Dieppa DR, et al. Viz.ai implementation of stroke augmented intelligence and communications platform to improve indicators and outcomes for a comprehensive stroke center and network. AJNR Am J Neuroradiol 2023; 44(1):47-53

[8] Matsoukas S, Stein LK, Fifi JT. Artificial intelligence-assisted software significantly decreases all workflow metrics for large vessel occlusion transfer patients, within a large spoke and hub system. Cerebrovasc Dis Extra 2023; 13:41–6

[9] Hassan AE, Ringheanu VM, Rabah R, Preston L, Qureshi AI. Early experience utilizing artificial intelligence shows significant reduction in transfer times and length of stay in a hub and spoke model. Presented at International Stroke Conference 2020; 19-21 Feb 2020; Virtual. Stroke 2020; 51(Suppl 1):TMP62

[10] Hassan AE, Ringheanu VM, Tekle WG. The implementation of artificial intelligence significantly reduces door-in-door-out times in a primary care center prior to transfer. Interv Neuroradiol 2022; Epub 2022 Aug 25:15910199221122848

[11] Gunda B, Neuhaus A, Sipos I, Stang R, Bojti PP, Takacs T, et al. Improved stroke care in a primary stroke centre using AI-decision support. Cerebrovasc Dis Extra 2022; 12(1):28-32

[12] Gunda B, Sipos I, Stang R, Bojti P, Kis B, Harston G. Improved stroke care in a primary stroke

centre using aidecision support. Presented at 12th World Stroke Congress; 12-15 May 2020; Vienna (Austria). Int J Stroke 2020; 15(1 Suppl):107

[13] Hassan AE, Ringheanu VM, Rabah RR, Preston L, Tekle WG, Qureshi AI. Early experience utilizing artificial intelligence shows significant reduction in transfer times and length of stay in a hub and spoke model. Interv Neuroradiol 2020; 26(5):615-22

[14] Adhya J, Li C, Eisenmenger L, Cerejo R, Tayal A, Goldberg M, et al. Positive predictive value and stroke workflow outcomes using automated vessel density (RAPID-CTA) in stroke patients: one year experience. Neuroradiol J 2021; 34(5):476-81

[15] Amukotuwa SA, Straka M, Dehkharghani S, Bammer R. Fast automatic detection of large vessel occlusions on CT angiography. Stroke 2019; 50(12):3431-8

[16] Amukotuwa SA, Straka M, Smith H, Chandra RV, Dehkharghani S, Fischbein NJ, et al. Automated detection of intracranial large vessel occlusions on computed tomography angiography a single center experience. Stroke 2019; 50(10):2790-8

[17] Dehkharghani S, Lansberg M, Venkatsubramanian C, Cereda C, Lima F, Coelho H, et al. High-performance automated anterior circulation CT angiographic clot detection in acute stroke: a multireader comparison. Radiology 2021; 298:665-70

[18] Mair G, White P, Bath PM, Muir KW, Al-Shahi Salman R, Martin C, et al. External validation of e-ASPECTS software for interpreting brain CT in Stroke. Ann Neurol 2022; 92(6):943-57

[19] Rodrigues GM, Barreira C, Froehler M, Baxter B, Devlin T, Lim J, et al. Multicenter ALADIN: automated large artery occlusion detection in stroke imaging using artificial intelligence. Presented at International Stroke Conference 2019; 6-8 Feb 2019; Honolulu, Hawaii. Stroke 2019; 50(Suppl 1):AWP71

[20] Seker F, Pfaff JAR, Mokli Y, Berberich A, Namias R, Gerry S, et al. Diagnostic accuracy of automated occlusion detection in CT angiography using e-CTA. Int J Stroke 2022; 17(1):77-82

[21] Yahav-Dovrat A, Saban M, Merhav G, Lankri I, Abergel E, Eran A, et al. Evaluation of artificial intelligence-powered identification of large-vessel occlusions in a comprehensive stroke center. AJNR Am J Neuroradiol 2020; 42:247-54

[22] Oxford Academic Health Science Network (AHSN), Brainomix. AI in health and care award: evaluation annual report March 2023 [PDF provided by NICE]: NHS (UK), 2023

[23] Nagaratnam K, Maharjan D, Neuhaus A, Flossmann C, Harston G. Artificial intelligence clinical decision aid tool shortens time in mechanical thrombectomy pathway. Int J Stroke 2021; 16(2S):WSC21-597

[24] Chan N, Sibtain N, Booth T, de Souza P, Bibby S, Mah YH, et al. Machine-learning algorithm in acute stroke: real-world experience. Clin Radiol 2023; 78(2):e45-51

[25] Chang W. Automated detection of large vessel occlusion using relative vessel density: one year experience in a multihospital integrated delivery network. Presented at Radiological Society of North America Annual Meeting 2021; 28 Nov - 2 Dec 2021; Chicago, IL, 28-Nov-2021, 2021 Available from: https://www.rsna.org/-/media/Files/RSNA/Annual-meeting/Future-and-past-

meetings/RSNA2021_Meeting_Program.ashx#page=548

[26] Chang W, Eisenmenger L, Cerejo R, Li C, Goldberg MF. Artificial intelligence for automated detection of large vessel occlusion using relative vessel density (RAPID LVO): one year experience in a multihospital integrated delivery network. Stroke 2022; 53(Suppl 1):Abstract WMP68

[27] Eldaya RW, Kansagra AP, Zei M, Mason E, Holder D, Heitsch L, et al. Performance of automated RAPID intracranial hemorrhage detection in real-world practice: a single-institution experience. J Comput Assist Tomogr 2022; 46(5):770-4

[28] Karamchandani RR, Helms AM, Satyanarayana S, Yang H, Clemente JD, Defilipp G, et al. Automated detection of intracranial large vessel occlusions using Viz.ai software: experience in a large, integrated stroke network. Brain Behav 2023; 13(1):e2808

[29] Leer M, McParland S, Wiggam I. How useful is artificial intelligence to support interpretation of CTA in hyperacute stroke? Presented at 17th UK Stroke Forum; 29 Nov - 1 Dec 2022; Liverpool, UK. Int J Stroke 2023; 18(1 Suppl):68-9

[30] Mair G, White P, Bath PM, Muir K, Martin C, Dye D, et al. Accuracy of artificial intelligence software for CT angiography in stroke. Ann Clin Transl Neurol 2023; 19:19

[31] Mallon DH, Taylor EJR, Vittay OI, Sheeka A, Doig D, Lobotesis K. Comparison of automated ASPECTS, large vessel occlusion detection and CTP analysis provided by Brainomix and RapidAI in patients with suspected ischaemic stroke. J Stroke Cerebrovasc Dis 2022; 31(10):106702

[32] Mannix N, Penhos E, Nimjee S, Shujaat MT, Heaton S, Vitellas C, et al. Real world experience with Viz.Ai automated large vessel occlusion in CT angiogram. Presented at American Academy of Neurology Annual Meeting 2023; 22-27 Apr 2023; Boston, MA, 22-Apr-2023, 2023 Available from: https://index.mirasmart.com/aan2023/PDFfiles/AAN2023-003353.html

[33] Matsoukas S, Chennareddy S, Kalagara R, Scaggiante J, Smith CJ, Bazil MJ, et al. Pilot deployment of Viz-intracranial hemorrhage for intracranial hemorrhage detection: real-world performance in a stroke code cohort. Stroke 2022; 53(9):E418-9

[34] Rodrigues G, Barreira CM, Bouslama M, Haussen DC, Al-Bayati A, Pisani L, et al. Automated large artery occlusion detection in stroke: a single-center validation study of an artificial intelligence algorithm. Cerebrovasc Dis 2022; 51(2):259-64

[35] Schlossman J, Ro D, Salehi S, Chow D, Yu W, Chang PD, et al. Head-to-head comparison of commercial artificial intelligence solutions for detection of large vessel occlusion at a comprehensive stroke center. Front Neurol 2022; 13:1026609

[36] Schmitt N, Mokli Y, Weyland CS, Gerry S, Herweh C, Ringleb PA, et al. Automated detection and segmentation of intracranial hemorrhage suspect hyperdensities in non-contrast-enhanced CT scans of acute stroke patients. Eur Radiol 2022; 32(4):2246-54

[37] Vella N, Hackett N, Baron S, Nagaratnam K. Comparison between artificial intelligence based detection of large vessel occlusion and neuroradiologist reporting. Presented at 17th UK Stroke Forum; 29 Nov - 1 Dec 2022; Liverpool, UK. Int J Stroke 2023; 18(1 Suppl):33-34

[38] Vitellas CA, Mannix NC, Nimjee SM, Shujaat MT, Heaton S, Lee VH. Real world experience with

Viz.AI automated large vessel occlusion detection. Presented at International Stroke Conference 2022; 9-11 Feb 2022; New Orleans, LA. Stroke 2022; 53(Suppl 1):A130

[39] DAP56 AI for stroke. OAHSN final response to EAG questions [as provided by NICE]. 2023: 13.

[40] Westwood M, Ryczek E, Ramaekers B, Grimm S. Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke: Diagnostics Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence – Protocol [Internet]. York: Kleijnen Systematic Reviews Ltd, 2021 [accessed 19.7.23] Available from: https://www.nice.org.uk/guidance/gid-dg10044/documents/final-protocol

[41] Oxford Academic Health Science Network (AHSN). Oxford AHSN evaluation of Brainomix AI technology: mid-term clinician survey [as provided by NICE]. 2023: 13.

[42] National Institute for Health and Care Excellence. DAP56 Additional analysis NICE Assessment AI in stroke [Excel spreadsheet as provided by NICE]. 2023.

[43] National Institute for Clinical Excellence. Diagnostics Assessment Programme: AI software to help clinical decision making in stroke. NHS organisation submission: David Hargroves (NHS England GIRFT Programme) [Document provided by NICE]. London: NICE, 2023 [accessed 26.6.23]

[44] National clinical guideline for stroke for the United Kingdom and Ireland (2023 ed) [Internet]. London: Intercollegiate Stroke Working Party, 2023 [accessed 19.7.23] Available from: www.strokeguideline.org

[45] Seker F, Pfaff JAR, Mokli Y, Berberich A, Namias R, Gerry S, et al. Diagnostic accuracy of automated occlusion detection in CT angiography using e-CTA. Int J Stroke 2021 Feb 11 [accessed 30.7.21]. Available from: https://doi.org/10.1177%2F17474930211992592 [Epub ahead of print];

[46] Andralojc LE, Kim DH, Edwards AJ. Diagnostic accuracy of a decision-support software for the detection of intracranial large-vessel occlusion in CT angiography. Clin Radiol 2023; 78(4):e313-8

[47] Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ 2013; 185(11):E537-44

[48] Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CB, Dippel DW, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. JAMA 2016; 316(12):1279-88

[49] Fransen PS, Berkhemer OA, Lingsma HF, Beumer D, van den Berg LA, Yoo AJ, et al. Time to reperfusion and treatment effect for acute ischemic stroke: a randomized clinical trial. JAMA Neurol 2016; 73(2):190-6

[50] Adhya J, Li C, Goldberg M, Eisenmenger L, Chang W. Positive predictive value of automated vessel density (RAPID-CTA) for detection of large vessel occlusion: one year experience [Abstract 219]. Presented at American Society of Neuroradiology Annual Meeting 2021; 22-26 May 2021; Virtual, 2021 Available from: https://www.asnr.org/wp-content/uploads/2021/07/ASNR21-Proceedings.pdf#page=614American

[51] Degan D, Turinese E, Iannucci G, Lanaro D, De Boni A, Brotto M, et al. Rapid software effectiveness in selecting patients with acute ischemic stroke eligible for endovascular mechanical

thrombectomy. Presented at 52nd Annual Conference of the Italian Society of Neurology; 3-6 Dec 2022; Milano, Italy. Neurol Sci 2022; 43(Suppl 1):S359

[52] Dyer T, Chawda S, Alkilani R, Morgan TN, Hughes M, Rasalingham S. Validation of an artificial intelligence solution for acute triage and rule-out normal of non-contrast CT head scans. Neuroradiology 2022; 64(4):735-43

[53] Elijovich L, Dornbos Iii D, Nickele C, Alexandrov A, Inoa-Acosta V, Arthur AS, et al. Automated emergent large vessel occlusion detection by artificial intelligence improves stroke workflow in a hub and spoke stroke system of care. J Neurointerv Surg 2022; 14(7):704-8

[54] Fraser J, Campbell M, Wise L, Krein B, Silvernail R, Dornbos D, et al. Artificial intelligence in stroke networking: early experience in building networks of care. Presented at 14th Congress of the European Society of Minimally Invasive Neurological Therapy (ESMINT 2022); 7-9 Sept 2022; Nice, France. J Neurointerv Surg 2022; 14(Suppl 2):A22

[55] Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. Neuroradiology 2020; 62:335–40

[56] Giurgiutiu D-v, Filly A, Honce J, kwok k, Mitchell B, Copeland K. Automated analysis for detecting large vessel occlusion on non contrast CT in acute stroke. Presented at the American Stroke Association 2022 International Stroke Conference and State-of-the-Science Stroke Nursing Symposium 2022; 9-11 Feb 2022; New Orleans, US. Stroke 2022; 53(Suppl 1):WP97

[57] Hillal A, Sultani G, Ramgren B, Norrving B, Wasselius J, Ullberg T. Accuracy of automated intracerebral hemorrhage volume measurement on non-contrast computed tomography: a Swedish Stroke Register cohort study. Neuroradiology 2023; 65(3):479-88

[58] James MH, Bernardo CB, Romane G, Christopher PB, John KC, Buthaina H, et al. Enhanced physician performance when using an artificial intelligence model to detect ischemic stroke on computed tomography. medRxiv (BMJ) 2023. Available from: https://www.medrxiv.org/content/medrxiv/early/2023/01/18/2023.01.16.23284632.full.pdf

[59] Khoruzhaya AN, Smorchkova A, Kremneva EI, Kudryavtsev N, Vladzymyrskyy A, Morozov S. Consistency of stroke and haemorrhage diagnosis on native brain CT scans in emergency medicine by clinicians using AI compared to radiologists. Presented at European Congress of Radiology (ECR 2022); 13-17 July 2022; Vienna, Austria., 2022 Available from: https://connect.myesr.org/app/uploads/brochure_38974.pdf?v=1647895402#page=233European

[60] Khunte M, Wu X, Ikuta I, Payabvash S, Malhotra A. Automated detection of intracranial proximal vessel occlusion on computed tomography. Presented at American Society of Neuroradiology Annual Meeting 2020; 30 May - 4 June 2020; Virtual, 2020 Available from: https://www.morressier.com/o/event/5e5d0b2c35f8e2df5e55a268/article/5e8335ba7cb08a046ef7c714

[61] Kundisch A, Honning A, Mutze S, Kreissl L, Spohn F, Lemcke J, et al. Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies. PLoS One 2021; 16(11):e0260560

[62] Love CJ, Bonner J. Should they stay or should they go? the impact of AI-based care coordination on stroke patient transfers across a hospital network. Presented at International Stroke

Conference 2023; 8-10 Feb 2023; Dallas, TX. Stroke 2023; 54(Suppl 1):AWP41

[63] Luijten SPR, Wolff L, Duvekot MHC, van Doormaal P-J, Moudrous W, Kerkhoff H, et al. Diagnostic performance of an algorithm for automated large vessel occlusion detection on CT angiography. J Neurointerv Surg 2022; 14(8):794-8

[64] Marigold R. Benefits of a regional CT artificial intelligence work steam in identifying patients for transfer to a comprehensive stroke centre (CSC) for mechanical thrombectomy. Presented at 17th UK Stroke Forum; 29 Nov - 1 Dec 2022; Liverpool UK. Int J Stroke 2023; 18(1 Suppl):26

[65] Mathieson P, Nagaratnam K, Podlasek A, Slade P, Ford G, Harston G. Influence of e-CTA software on CT angiography-based decision making. Presented at 8th European Stroke Organisation Conference (ESOC 2022): 4-6 May 2022; Lyon, France. Eur Stroke J 2022; 7(1 Suppl):497-8

[66] Matsoukas S, Morey J, Lock G, Chada D, Shigematsu T, Marayati NF, et al. AI software detection of large vessel occlusion stroke on CT angiography: a real-world prospective diagnostic test accuracy study. J Neurointerv Surg 2023; 15:52–6

[67] Nagamine M, Zolyan A, Elstrott S, McLouth J, Vu V, Yu W, et al. Evaluation of the implementation of an AI tool for large vessel occlusion: impact on radiologists' workflow and patient outcomes. Stroke 2022; 53(Suppl 1)

[68] Nicholas CF, Aubrey R, Alan SB, John D, Alexandra RP. Artificial intelligence application expediates surgical team evaluation as part of hemorrhagic stroke workflow metrics. medRxiv (BMJ) 2023 [accessed 5.6.23]:2023.02.24.23286438. Available from: https://www.medrxiv.org/content/medrxiv/early/2023/03/01/2023.02.24.23286438.full.pdf

[69] Olive-Gadea M, Garcia-Tornel A, Requena M, De Dios Lascuevas M, Boned S, Rubiera M, et al. Artificial intelligence assessment of non-contrast CT increases intracranial occlusion detection as compared with ct-angiograhy readings. Stroke 2022; 53(Suppl 1):Abstract WMP63

[70] Salokhiddinov M, Nuruddinov R, Khaqberdiev J, Nazarov A. Clinical application of the realtime artificial intelligence in the detection of hemorrhagic stroke. Presented at 14th World Stroke Congress; 26-29 Oct 2022; Singapore Int J Stroke 2022; 17(3 Suppl):268

[71] Scavasine VC, Ferreti LA, da Costa RT, Leitao CA, Teixeira BC, Zetola VdHF, et al. Automated evaluation of ASPECTS from brain computerized tomography of patients with acute ischemic stroke. J Neuroimaging 2023; 33(1):134-7

[72] Stib MT, Vasquez J, Dong MP, Kim YH, Subzwari SS, Triedman HJ, et al. Detecting large vessel occlusion at multiphase CT angiography by using a deep convolutional neural network. Radiology 2020; 297:640-49

[73] Voter AF, Meram E, Garrett JW, Yu JPJ. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of intracranial hemorrhage. J Am Coll Radiol 2021; 8(8)

[74] Weyland CS, Papanagiotou P, Schmitt N, Joly O, Bellot P, Mokli Y, et al. Hyperdense artery sign in patients with acute ischemic stroke-automated detection with artificial intelligence-driven software. Front Neurol 2022; 13:807145

APPENDIX 1: LITERATURE SEARCH STRATEGIES

For full details of all previous searches please see the 2021 DAR.[1]

Updated clinical effectiveness searches

Database	Dates covered	Hits
Embase	1974-2023/05/22	3,122
MEDLINE + Pre-MEDLINE	1946-2023/05/22	1,738
Northern Light	2010-2023/wk18	157
MedRxiv	up to 2023/05/23	507
Total		**5,524**

Embase (Ovid): 1974-2023/05/22
Searched: 23.5.23

1 exp brain ischemia/ (217857)
2 exp brain hemorrhage/ (176182)
3 basal ganglion hemorrhage/ (778)
4 cerebrovascular accident/ (283653)
5 brain infarction/ (63257)
6 blood vessel occlusion/ (14414)
7 (Stroke$ or apople$ or cerebral vasc$ or cerebrovasc$ or cerebro vasc$ or poststroke$ or encephalorrhag$ or hematencephalon$ or large vessel occlusion$).ti,ab,ot. (583779)
8 ((brain or blood flow) adj2 disturb$).ti,ab,ot. (3220)
9 ((sinus or sagittal) adj3 thromb$).ti,ab,ot. (8499)
10 (isch?emi$ adj3 (seizure$ or attack$ or thrombo$ or embolic or encephalopath$ or neural)).ti,ab,ot. (46947)
11 ((Bleed$ or h?emorrhag$) adj2 corpus callosum).ti,ab,ot. (29)
12 ((brain or cerebr$ or cerebell$ or cortical or Intraparenchymal or intracortical or vertebrobasil$ or hemispher$ or intracran$ or intra-cran$ or intracerebral or intratentorial or intra-tentorial or intraventricular or intra-ventricular or periventricular or peri-ventricular or supratentorial or supra-tentorial or anterior circulat$ or posterior circulat$ or basal gangli$ or global or focal or parenchymal or subarachnoid or sub-arachnoid or putaminal or putamen or posterior fossa or intra-axial or intraaxial or lacunar) adj3 (arrest$ or attack$ or isch?emi$ or infarct$ or insufficien$ or emboli$ or occlus$ or hypox$ or vasospasm or obstruction or vasculopath$ or failure$ or thromb$ or h?emorrhag$ or microh?emorrhag$ or accident$ or h?ematoma$ or bleed$ or microbleed$ or insult$)).ti,ab,ot. (331365)
13 (CVA or CVAS or MCA$ or ICH or ICHs or CVST or CVSTs or CVDST or CVT or CVDSTs or CVTs or LVO or LVOs).ti,ab. (96273)
14 or/1-13 (997566)
15 ((diagnos$ or predict$ or specificity or sensitiv$) adj4 (criteria or criterion or guideline$ or pattern$ or trend$ or utili$ or management or prevalence or initiat$ or distribution$ or coverage or variety or selection or spread or alternative$ or frequen$)).ti,ab,ot. (578943)
16 diagnosis/ or early diagnosis/ (1518976)

17 exp brain scintiscanning/ (10472)
18 Neurologic examination/ (80379)
19 Computer assisted tomography/ (887259)
20 Brain radiography/ (9148)
21 ((Brain or cerebral or neurologic$ or CT or head) adj2 (scan$ or scintigraph$ or examination$ or angiograph$ or image analys$ or perfusion$ or radiograph$)).ti,ab,ot,hw. (453449)
22 (CAT scan$ or CTA or CTP or CTAs or CTPs or neuroimag$ or neuro-imag$ or (comput$ adj2 tomograph$)).ti,ab,ot,hw. (1533067)
23 (Gamma encephalograph$ or Gammaencephalograph$ or Radio encephalograph$ or Radioencephalograph$).ti,ab,ot,hw. (48)
24 or/15-23 (3498425)
25 exp artificial intelligence/ (81257)
26 automated pattern recognition/ (17278)
27 decision support system/ (27431)
28 computer assisted diagnosis/ (42105)
29 Convolutional neural network/ (24664)
30 (Artificial intelligence or AI or machine intelligence or computer-aided triage$ or support vector machine$ or relevance vector machine$).ti,ab,ot. (112020)
31 ((automat$ or computer) adj2 (analys$ or diagnos$ or detect$)).ti,ab,ot. (64446)
32 ((deep or machine) adj learning).ti,ab,ot. (135200)
33 (decision support$ adj (software or tool$)).ti,ab,ot. (6136)
34 (CNN or CNNs or convNet or (convolut$ adj2 neural network$) or convolutional ANNs or convolutional ANN or convolutional NNs or convolutional NN).ti,ab. (31822)
35 automat$ hierarch$ evaluat$.ti,ab. (1)
36 (Aidoc or e-CTA or e-ASPECTS or e-stroke or brainomix or brainscan or "brainscan.ai" or icobrain or icometrix or qER or Qure or Zebra$ or e-CTP or briefcase or rapid CTA or rapid LVO or rapid core or rapid ASPECTS or rapid ICH or rapidai or blackford or "viz.ai" or viz or "ct perfusion 4d" or cercare or cina$ or Avicenna or accipio$ or maxQ AI or biomind or "biomind.ai" or ischemaview or rapid CTP or "qure.ai").ti,ab. (156010)
37 or/25-36 (543205)
38 14 and 24 and 37 (3290)
39 (letter or editorial or note).pt. (3020548)
40 38 not 39 (3183)
41 animal/ (1608366)
42 animal experiment/ (3059803)
43 (rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).ti,ab,ot,hw. (7616899)
44 or/41-43 (7616899)
45 exp human/ (25485707)
46 human experiment/ (649899)
47 or/45-46 (25488107)
48 44 not (44 and 47) (5716882)
49 40 not 48 (3122)

MEDLINE and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions (Ovid): 1946-2023/05/22
Searched: 23.05.23

1 exp Brain Ischemia/ (124019)
2 exp Intracranial Hemorrhages/ (79855)
3 Stroke/ (130988)
4 (Stroke$ or apople$ or cerebral vasc$ or cerebrovasc$ or cerebro vasc$ or poststroke$ or encephalorrhag$ or hematencephalon$ or large vessel occlusion$).ti,ab,ot. (367334)
5 ((brain or blood flow) adj2 disturb$).ti,ab,ot. (2291)
6 ((sinus or sagittal) adj3 thromb$).ti,ab,ot. (5772)
7 (isch?emi$ adj3 (seizure$ or attack$ or thrombo$ or embolic or encephalopath$ or neural)).ti,ab,ot. (29663)
8 ((Bleed$ or h?emorrhag$) adj2 corpus callosum).ti,ab,ot. (24)
9 ((brain or cerebr$ or cerebell$ or cortical or Intraparenchymal or intracortical or vertebrobasil$ or hemispher$ or intracran$ or intra-cran$ or intracerebral or intratentorial or intra-tentorial or intraventricular or intra-ventricular or periventricular or peri-ventricular or supratentorial or supra-tentorial or anterior circulat$ or posterior circulat$ or basal gangli$ or global or focal or parenchymal or subarachnoid or sub-arachnoid or putaminal or putamen or posterior fossa or intra-axial or intraaxial or lacunar) adj3 (arrest$ or attack$ or isch?emi$ or infarct$ or insufficien$ or emboli$ or occlus$ or hypox$ or vasospasm or obstruction or vasculopath$ or failure$ or thromb$ or h?emorrhag$ or microh?emorrhag$ or accident$ or h?ematoma$ or bleed$ or microbleed$ or insult$)).ti,ab,ot. (229471)
10 (CVA or CVAS or MCA$ or ICH or ICHs or CVST or CVSTs or CVDST or CVT or CVDSTs or CVTs or LVO or LVOs).ti,ab. (55387)
11 or/1-10 (612343)
12 ((diagnos$ or predict$ or specificity or sensitiv$) adj4 (criteria or criterion or guideline$ or pattern$ or trend$ or utili$ or management or prevalence or initiat$ or distribution$ or coverage or variety or selection or spread or alternative$ or frequen$)).ti,ab,ot. (394521)
13 Diagnosis/ (17527)
14 Early Diagnosis/ (30195)
15 Brain/dg [Diagnostic Imaging] (61501)
16 Stroke/dg [Diagnostic Imaging] (9607)
17 Radiography/ (326915)
18 exp Radionuclide Imaging/ (235437)
19 Neurologic Examination/ (28032)
20 Tomography, X-Ray Computed/ (417032)
21 ((Brain or cerebral or neurologic$ or CT or head) adj2 (scan$ or scintigraph$ or examination$ or angiograph$ or image analys$ or perfusion$ or radiograph$)).ti,ab,ot,hw. (247199)
22 (CAT scan$ or CTA or CTP or CTAs or CTPs or neuroimag$ or neuro-imag$ or (comput$ adj2 tomograph$)).ti,ab,ot,hw. (530476)
23 (Gamma encephalograph$ or Gammaencephalograph$ or Radio encephalograph$ or Radioencephalograph$).ti,ab,ot,hw. (155)
24 or/12-23 (1756130)
25 exp Artificial Intelligence/ (172310)
26 Pattern Recognition, Automated/ (26475)

27 Neural Networks, Computer/ (47283)
28 (Artificial intelligence or AI or machine intelligence or computer-aided triage$ or support vector machine$ or relevance vector machine$).ti,ab,ot. (86341)
29 ((automat$ or computer) adj2 (analys$ or diagnos$ or detect$)).ti,ab,ot. (49177)
30 ((deep or machine) adj learning).ti,ab,ot. (109328)
31 (decision support$ adj (software or tool$)).ti,ab,ot. (4295)
32 (CNN or CNNs or convNet or (convolut$ adj2 neural network$) or convolutional ANNs or convolutional ANN or convolutional NNs or convolutional NN).ti,ab. (25791)
33 automat$ hierarch$ evaluat$.ti,ab. (1)
34 (Aidoc or e-CTA or e-ASPECTS or e-stroke or brainomix or brainscan or "brainscan.ai" or icobrain or icometrix or qER or Qure or Zebra$ or e-CTP or briefcase or rapid CTA or rapid LVO or rapid core or rapid ASPECTS or rapid ICH or rapidai or blackford or "viz.ai" or viz or "ct perfusion 4d" or cercare or cina$ or Avicenna or accipio$ or maxQ AI or biomind or "biomind.ai" or ischemaview or rapid CTP or "qure.ai").ti,ab. (124587)
35 or/25-34 (459853)
36 11 and 24 and 35 (1796)
37 (letter or editorial or note).pt. (1867211)
38 exp animals/ not (exp animals/ and humans/) (5123632)
39 36 not (37 or 38) (1738)

Northern Light Life Sciences Conference Abstracts (Ovid): 2010–2023/wk18
Searched: 23.5.23

1 exp Brain Ischemia/ (6403)
2 exp Intracranial Hemorrhages/ (14794)
3 Stroke/ (44370)
4 (Stroke$ or apople$ or cerebral vasc$ or cerebrovasc$ or cerebro vasc$ or poststroke$ or encephalorrhag$ or hematencephalon$ or large vessel occlusion$).ti,ab. (57837)
5 (CVA or CVAS or MCA$ or ICH or ICHs or CVST or CVSTs or CVDST or CVT or CVDSTs or CVTs or LVO or LVOs).ti,ab. (9606)
6 ((brain or cerebr$ or cerebell$ or cortical or Intraparenchymal or intracortical or vertebrobasil$ or hemispher$ or intracran$ or intra-cran$ or intracerebral or intratentorial or intra-tentorial or intraventricular or intra-ventricular or periventricular or peri-ventricular or supratentorial or supra-tentorial or anterior circulant$ or posterior circulat$ or basal ganglion$ or global or focal or parenchymal or subarachnoid or sub-arachnoid or putaminal or putamen or posterior fossa or intra-axial or intraaxial or lacunar) adj3 (arrest$ or attack$ or isch?emi$ or infarct$ or insufficien$ or emboli$ or occlus$ or hypox$ or vasospasm or obstruction or vasculopath$ or failure$ or thromb$ or h?emorrhag$ or microh?emorrhag$ or accident$ or h?ematoma$ or bleed$ or microbleed$ or insult$)).ti,ab. (20268)
7 ((brain or blood flow) adj2 disturb$).ti,ab. (117)
8 ((sinus or sagittal) adj3 thromb$).ti,ab. (642)
9 (isch?emi$ adj3 (seizure$ or attack$ or thrombo$ or embolic or encephalopath$ or neural)).ti,ab. (2427)
10 ((Bleed$ or h?emorrhag$) adj2 corpus callosum).ti,ab. (1)
11 or/1-10 (99564)
12 Diagnosis/ (0)

13 Early Diagnosis/ (25140)
14 Radiography/ (0)
15 exp Radionuclide Imaging/ (0)
16 Neurologic Examination/ (0)
17 Tomography, X-Ray Computed/ (0)
18 ((Brain or cerebral or neurologic$ or CT or head) adj2 (scan$ or scintigraph$ or examination$ or angiograph$ or image analys$ or perfusion$ or radiograph$)).ti,ab. (21101)
19 (CAT scan$ or CTA or CTAs or CTP or CTPs or neuroimag$ or neuro-imag$ or (comput$ adj2 tomograph$)).ti,ab. (27247)
20 (Gamma encephalograph$ or Gammaencephalograph$ or Radio encephalograph$ or Radioencephalograph$).ti,ab. (0)
21 ((diagnos$ or predict$ or specificity or sensitiv$) adj4 (criteria or criterion or guideline$ or pattern$ or trend$ or utili$ or management or prevalence or initiat$ or distribution$ or coverage or variety or selection or spread or alternative$ or frequen$)).ti,ab. (36241)
22 or/12-21 (103275)
23 exp Artificial Intelligence/ (0)
24 Pattern Recognition, Automated/ (0)
25 Neural Networks, Computer/ (0)
26 (CNN or CNNs or convNet or (convolut$ adj2 neural network$) or convolutional ANNs or convolutional ANN or convolutional NNs or convolutional NN).ti,ab. (1999)
27 (Artificial intelligence or AI or machine intelligence or computer-aided triage$ or support vector machine$ or relevance vector machine$).ti,ab. (10825)
28 ((automat$ or computer) adj2 (analys$ or diagnos$ or detect$)).ti,ab. (5094)
29 ((deep or machine) adj learning).ti,ab. (14970)
30 automat$ hierarch$ evaluat$.ti,ab. (0)
31 (decision support$ adj (software or tool$)).ti,ab. (971)
32 (Aidoc or e-CTA or e-ASPECTS or e-stroke or brainomix or brainscan or "brainscan.ai" or icobrain or icometrix or qER or Qure or Zebra$ or e-CTP or briefcase or rapid CTA or rapid LVO or rapid core or rapid ASPECTS or rapid ICH or rapidai or blackford or "viz.ai" or viz or "ct perfusion 4d" or cercare or cina$ or Avicenna or accipio$ or maxQ AI or biomind or "biomind.ai" or ischemaview or rapid CTP or "qure.ai").ti,ab. (8541)
33 or/23-32 (39048)
34 11 and 22 and 33 (157)

medRxiv (Internet): up to 2023/05/23
https://www.medrxiv.org/
Searched: 23.5.23

Advanced search

Limit: posted between "01 Oct, 2021 and 23 May, 2023"

Full text or abstract or title (match whole all)	**Update (23.05.23)**
stroke* Aidoc	0
Stroke* e-CTA	0
Stroke* e-ASPECTS	1
e-stroke	14

Stroke* brainomix	2
Stroke* brainscan	2
Stroke* brainscan.ai	0
stroke icobrain	1
Stroke* icometrix	4/5
Stroke* qER	0
Stroke* Qure	1
Stroke* Zebra*	1
Stroke* e-CTP	0
Stroke* briefcase	0
Stroke* rapid CTA	21
Stroke* rapid LVO	10/18
Stroke* rapid core	249/270
Stroke* rapid ASPECTS	160/288
Stroke* rapid ICH	11/36
Stroke* rapidai	1/3
Stroke* blackford	1
Stroke* viz.ai	0/6
Stroke* viz	11/19
Stroke* ct perfusion 4d	13/17
Stroke* cercare	0
Stroke* cina*	1/2
Stroke* Avicenna	0
Stroke* accipio*	0
Stroke* maxQ AI	0
Stroke* biomind	0
Stroke* biomind.ai	0
Stroke* ischemaview	0/2
Stroke* rapid CTP	3/9
Stroke* qure.ai	0
Total	**719**
Total without dupes	**507**

APPENDIX 2: DATA EXTRACTION TABLES

Table A2.1: Baseline study details

Study Details	Selection criteria	Participant details	AI intervention
Figurelle 2023[7] **Publication type:** Full paper **Country:** USA **Funding:** None declared **Recruitment:** June 2020 to January 2021 and January 2021 to June 2021 (retrospective) **Number of participants:** 82	**Inclusion criteria**: Sequential acute stroke cases requiring embolectomy **Exclusion criteria:** NR **Research Question:** (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	**None reported**	

Gunda 2022[11] **Publication type:** Full paper **Country:** Hungary **Funding:** e-Stroke Suite was provided by Angels Initiative (endorsed by ESO) **Competing interests:** The lead author is the Chief Medical Officer at Brainomix **Recruitment:** May to December 2017 and May to December 2018 (retrospective) **Number of participants:** 797	**Inclusion criteria**: Consecutive patients admitted to the Department of Neurology, Semmelweis University (Budapest, Hungary) with acute ischemic stroke. **Exclusion criteria:** NR **Research Question:** (Q1) Is AI-derived software assisted review of non-enhanced CT brain scans for guiding thrombolysis treatment decisions for people with suspected acute stroke a clinically effective intervention? (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	**Patients who received IVT** ***Intervention:*** **Mean (SD) age, years:** 65.1±13.5 **Male (%):** 40 (55.6) **Median (unspecified) admission NIHSS:** 6 (3 to 10.25) ***Comparator:*** **Mean (SD) age, years:** 67.6±13.3 **Male (%):** 21 (45.7) **Median (unspecified) admission NIHSS:** 8 (5 to 13) **Patients who received MT** ***Intervention:*** **Mean (SD) age, years:** 62.3±15.3 **Male (%):** 10 (52.6) **Median (unspecified) admission NIHSS:** 13 (10 to 15.5) ***Comparator:*** **Mean (SD) age, years:** 55.8±18.1 **Male (%):** 6 (54.5) **Median (unspecified) admission NIHSS:** 15 (13.5 to 18.5)	Brainomix eASPECTS and eCTA

Study Details	Selection criteria	Participant details	AI intervention
Hassan 2022[10] Hassan 2020[9] **Publication type:** Full paper **Country:** USA **Funding:** None declared **Recruitment:** February 2017 to November 2018 (retrospective) **Number of participants:** 63	**Inclusion criteria:** Consecutive patients who presented a single PSC, with signs of LVO on CTA, who were transferred to a CSC. **Exclusion criteria:** NR **Research Question:** (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	***Intervention:*** **Mean (SD) age, years:** 68.7 (14.9) **Male (%):** 13 (37.1) **Ethnicity:** White 3 (14.3); Hispanic 30 (85.7); African American 0 (0); Asian 0 (0) **AF (%):** 8 (22.9) **Diabetes (%):** 13 (37.1) **Smoking (%):** 4 (11.14) **Hypertension (%):** 28 (80.0) **History of stroke/TIA:** 7 (20.0) **CAD:** 9 (25.7) **Baseline NIHSS, mean (SD):** 16.0 (6.0) ***Comparator:*** **Mean (SD) age, years:** 71.6 (12.3) **Male (%):** 15 (53.6) **Ethnicity:** White 5 (17.9); Hispanic 23 (82.1); African American 0 (0); Asian 0 (0) **AF (%):** 10 (35.7) **Diabetes (%):** 12 (42.9) **Smoking (%):** 2 (7.1) **Hypertension (%):** 25 (89.3) **History of stroke/TIA:** 6 (21.4) **CAD:** 7 (25.0) **Baseline NIHSS, mean (SD):** 18.3 (7.4) **There were no significant differences, in baseline characteristics, between groups**	Viz LVO

Study Details	Selection criteria	Participant details	AI intervention
Matsoukas 2023[8] **Publication type:** Full paper **Country:** USA **Funding:** One author was supported by an AHA grant and one author had received research funding from Viz.ai. **Recruitment:** January 2020 to December 2021 (retrospective) **Number of participants:** 63	**Inclusion criteria:** Patients with suspected/confirmed LVO who were transferred from PSCs within and outside the healthcare system during the study period **Exclusion criteria:** Patients placed on 'LVO watch' due to mild symptoms, patients with missing time metrics. **Research Question:** (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	***Intervention:*** **Median age, years:** 70 **Male (%):** 17 (44.7) **Ethnicity:** White 15 (39.4); Hispanic 5 (13.2); African American 11 (28.9); Asian 0 (0); Other/unknown 7 (18.4) **Baseline NIHSS, mean (SD):** 13.3 (8.7) ***Comparator:*** **Median age, years:** 71.5 **Male (%):** 19 (47.5) **Ethnicity:** White 5 (12.5); Hispanic 8 (20.0); African American 16 (40); Asian 1 (2.5); Other/unknown 10 (25.0) **Baseline NIHSS, mean (SD):** 15.5 (8.5) **There were no significant differences, in baseline characteristics, between groups**	Viz LVO

Study Details	Selection criteria	Participant details	AI intervention
OAHSN 2023[22, 39] **Publication type:** Unpublished report **Country:** UK **Funding:** The report is an interim report, provided as part of a 3-year evaluation, funded by an AI in Health and Care Award, managed by the AAC in partnership with the NIHR. **Recruitment:** Unclear; the date of implemenation varied across participating sites, ranging from 25/02/2020 to 25/10/2021. The time periods over which data were collected also varied between sites (from 13 to 33 months for the pre-implementation period and from 12 to 31 months for the post-implementation period). **Number of participants:** Unclear; the total number of patients presenting with stroke was reported as 67,810.	**Inclusion criteria:** Not reported **Exclusion criteria:** Not reported **Research Question:** (Q1) Is AI-derived software assisted review of non-enhanced CT brain scans for guiding thrombolysis treatment decisions for people with suspected acute stroke a clinically effective intervention? (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention? (Q2b) Is AI-derived software assisted review of CT perfusion brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke after a CTA brain scan a clinically effective intervention? **It is not clear which data included in the report may be applicable to individual research questions.**	**None reported**	e-Stroke

AAC = Accelerated Access Collaborative; AF = atrial fibrillation; AI = artificial intelligence; AIS = acute ischaemic stroke; ASPECTS = Alberta stroke program; CAD = coronary artery disease; CSC = comprehensive stroke centre; CT = computed tomography; CTA = computed tomography angiography; IVT = intravenous thrombolysis; LVO = large vessel occlusion; MT = mechanical thrombectomy; NIHR = National Institute for Health Research; NIHSS = National Institute of Health Stroke Scale; NR = not reported; PSC = primary stroke centre; Q = question; SD = standard deviation; TIA = transient ischemic attack; USA = United States of America

Table A2.2: Details of AI-derived software technology and comparator

Study details	Imaging details	AI-derived software technology	Reference standard/Comparator
Figurelle 2023[7] **Research Question:** (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	**No details were reported**	**AI-derived software technology:** Viz.ai, version not reported (Viz.ai Inc., San Francisco, CA) **Analysis:** *'When a stroke code is called (from either internal or external sites in our network), the stroke provider evaluates the patient and accesses the Viz.ai images on a hand-held device. The neurointerventional team is contacted using the secure app, and care determinations such as appropriateness for rtPA and thrombectomy are made, including decisions to transfer patients for intervention.'*	**Comparator image interpretation:** Unclear (standard stroke care before implementation of AI decision support software)
Gunda 2022[11] **Research Question:** (Q1) Is AI-derived software assisted review of non-enhanced CT brain scans for guiding thrombolysis treatment decisions for people with suspected acute stroke a clinically effective intervention? (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	NCCT scans were obtained using a 16-slice scanner (Philips, Amsterdam, The Netherlands); slice thickness 2 mm for NCCT and 1 mm for CTA. DICOM images from NCCT and CTA were pre-processed and corrected for positional transformations.	**AI-derived software technology:** e-ASPECTS and e-CTA, version not reported (Brainomix, Oxford, UK) **Analysis:** e-ASPECTS automatically segmented regions of the MCA territory and characterised the tissue as ischaemic or normal appearing; outputs included ASPECTS and acute ischemic volume. e-CTA outputs included LVO location, ratio of collateral flow compared to the contralateral side, and a collateral score (with 0 as no	**Comparator image interpretation:** Unclear (standard stroke care before implementation of AI decision support software)

Study details	Imaging details	AI-derived software technology	Reference standard/Comparator
		flow and 3 as complete collateral blood supply). *'No changes other than the introduction of the e-Stroke Suite were made to service delivery over the duration of the project.'*	
Hassan 2022[10] Hassan 2020[9] **Research Question:** (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	**No details were reported**	**AI-derived software technology:** Viz LVO, version not reported (Viz.ai Inc., San Francisco CA) **Analysis:** *'Once signs of LVO were detected by the software, the triage alerts of the CTA scan were simultaneously sent to the mobile devices of the ED physician, radiologist, neurologist, and interventionalist. The patient was then transferred and treated based on the unified communication of all the involved parties.'*	***Comparator image interpretation:*** *'Patients were initially treated by the emergency department (ED) physician and referred for a CTA that was conducted by the technologist After this scan was read by the radiologist, the information regarding vessel occlusion was sent to the ED physician and the neurologist who recommended care. This information was then referred to the interventionalist, the patient was transferred to the CSC, and the thrombectomy was performed.'*
Matsoukas 2023[8] **Research Question:** (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention?	**No details were reported**	**AI-derived software technology:** Viz LVO, version not reported (Viz.ai Inc., San Francisco CA) **Analysis:** *'Viz LVO is trained to identify LVOs in the supraclinoid internal carotid artery (ophthalmic, choroidal, and communicating segments) and the M1 (horizontal part) of the MCA. However, it does not assess the extracranial circulation, the posterior circulation, or the infraclinoid internal carotid*	**Comparator image interpretation:** Unclear (standard stroke care without implementation of AI decision support software)

Study details	Imaging details	AI-derived software technology	Reference standard/Comparator
		artery. In instances where a partial or complete occlusion is suspected, or when a vessel's caliber is less than the reference threshold, an LVO is suspected, and an alert is automatically sent to the stroke team. For every CTA scan that is processed by Viz, a positive or negative LVO notification is provided, rather than the exact location of the occlusion.'	
OAHSN 2023[22, 39] **Research Question:** (Q1) Is AI-derived software assisted review of non-enhanced CT brain scans for guiding thrombolysis treatment decisions for people with suspected acute stroke a clinically effective intervention? (Q2a) Is AI-derived software assisted review of CTA brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke a clinically effective intervention? (Q2b) Is AI-derived software assisted review of CT perfusion brain scans for guiding mechanical thrombectomy treatment decisions for people with an ischaemic stroke after a CTA brain scan a clinically effective intervention?	**No details were reported**	**AI-derived software technology:** e-Stroke (e-ASPECTS and e-CTA at all participating sites and e-CTP and the six participating CSCs), version not reported (Brainomix, Oxford, UK) **Analysis:** Unclear '*Most sites have automatic processing of images, so whilst a scan is processed by e-Stroke that doesn't necessarily mean it has been viewed or used to determine a diagnosis.'* '*Clinician will vary depending on site and we would need to seek further information on this. '*	**Comparator image interpretation:** Unclear (standard stroke care without implementation of AI decision support software)

Study details	Imaging details	AI-derived software technology	Reference standard/Comparator
It is not clear which data included in the report may be applicable to individual research questions.			

AI = artificial intelligence; ASPECTS = Alberta Stroke Program Early CT Score; CSC = comprehensive stroke centre; CT = computed tomography; CTA = computed tomography angiography; CT angiography; CTP = computed tomography perfusion; LVO = large vessel occlusion; NCCT = non-contrast computed tomography; Q = question; rtPA: tissue plasminogen activator; UK = United Kingdom

APPENDIX 3: DETAILS OF EXCLUDED STUDIES WITH RATIONALE

To be included in the review studies had to fulfil the following criteria:

Population:	Adults (≥18 years old) attending a secondary care stroke centre with: (Q1) suspected acute stroke and who were last known to be well within 24 hours; (Q2a) AIS, who were last known to be well within 6 hours; (Q2b) suspected acute stroke, who were last known to be well more than 6 hours previously, but within 24 hours, and in whom ischaemic stroke has been confirmed on plain CT
Index Test:	AI-derived software: Aidoc ICH, Aidoc LVO, Aidoc mobile (Aidoc); Accipio (MaxQ AI); e-ASPECTS, e-CTP, e-CTA (Brainomix); icobrain ct (Icometrix); Biomind (Biomind.ai); Brainscan; Cercare stroke (Cercare Medical); CINA ICH, CINA LVO, CINA ASPECTS (Avicenna); CT Perfusion 4D (GE Healthcare); qER (Qure.ai); Rapid ASPECTS, Rapid ICH, Rapid CTA, Rapid LVO, Rapid CTP), RapidAI (iSchemaView); Viv ICH, Viz LVO, Viz CTP (Viz.ai); Zebra-Med (Zebra Medical Vision) (Q1) AI-derived software assisted review of plain CT by a healthcare professional other than a neuroradiologist (Q2a) AI-derived software assisted CTA by a healthcare professional other than a neuroradiologist (Q2b) AI-derived software assisted CTA and CTP review by a healthcare professional other than a neuroradiologist
Reference Standard:	Unassisted, (Q1) plain CT, (Q2a) CTA, (Q2b) CTP, review by a neuroradiologist, or by a consensus panel
Comparator:	(Q1) Unassisted plain CT review by a neuroradiologist or other healthcare professional (Q2a) Unassisted CTA review by a neuroradiologist or other healthcare professional (Q2b) AI-derived software assisted CTA and AI-derived software assisted CT perfusion brain scan review by a neuroradiologist or other healthcare professional **OR** Unassisted CTA and AI-derived software assisted CT perfusion brain scan review by a neuroradiologist or other healthcare professional

Outcome: Test accuracy (the numbers of true positive, false negative, false positive and true negative test results), for the target condition: (Q1) ICH or ischaemic stroke; (Q2a) LVO/occlusion of the proximal anterior circulation; (Q2b) LVO/occlusion of the proximal anterior circulation for CTA and presence of salvageable tissue for CTP

Clinical/patient-perceived outcomes: mortality, function (e.g., modified Rankin score), health-related quality of life, procedure-related adverse events (e.g., bleed subsequent to thrombolysis), length of hospital stay

The table below summarises studies which were screened for inclusion based on full text publication but did not fulfil one or more of the above criteria along with details of the reason for exclusion.

Study Details	Brief description of study and reason for exclusion
Adhya 2021[50]	**Duplicate publication:** Conference abstract relating to a study[14] included in the 2021 DAR; no additional data
Andralojc 2023[46]	**Index test and reference standard**: Journal article reporting the stand-alone accuracy of an AI-derived software technology for the detection of LVO, which has not been listed in Table 2 because: index test (not clear that the AI-derived software technology was viewed/used in all included radiology reports) and reference standard (reference standard was only applied where the radiology report and the AI-derived software technology were discordant)
Degan 2022[51]	**Outcomes:** Conference abstract assessing the stand-alone accuracy of RAPID CTP for detection of LVO, which has not been listed in Table 2 because: outcomes (insufficient information to extract 2x2 data)
Dyer 2022[52]	**Population and index test:** Journal article the stand-alone accuracy of an AI-derived software technology for the detection of ICH, which has not been listed in Table 2 because: population (all emergency head CT, not stroke specific) and index test (AI-derived software technology not specified)
Elijovich 2022[53]	**Study design and comparator:** Journal article reporting a comparison of workflow metrics (including time to treatment) between patients with an LVO that was flagged by an automated alert system (Viz LVO) and patients with an LVO who were missed by the automated alert system combined with historical controls; no pre-AI/without AI comparator
Fraser 2022[54]	**Study design, compactor and outcomes:** Conference abstract, reporting a retrospective review CT/CTA scans analysed by Viz.ai; co pre-AI or without AI comparator and no specified outcomes reported
Ginat 2020[55]	**Population:** Journal article the stand-alone accuracy of an AI-derived software technology (Aidoc) for the detection of ICH, which has not been listed in Table 2 because: population (all emergency head CT, not stroke specific)
Giurgiutiu 2022[56]	**Outcomes:** Conference abstract assessing the stand-alone accuracy of RAPID detection of hyperdense vessel sign on NCCT for early prediction of LVO on CTA with insufficient information to extract 2x2 data

Study Details	Brief description of study and reason for exclusion
Hillal 2023[57]	**Participants and reference standard:** Journal article reporting agreement between qER and human readers with respect to ICH volume. Participants (all had non-traumatic ICH) and reference standard (target condition was haemorrhage volume)
James 2023[58]	**Intervention and reference standard:** Journal article reporting the effects of an un-specified novel AI model on the performance of clinicians for detecting ischemic core and determining infarct volume, as defined by MRI. Intervention not specified and reference standard was MRI
Khoruzhaya 2022[59]	**Population, intervention, reference standard and outcomes:** Conference abstract comparing accuracy of a group of radiologists to accuracy of a group of clinicians using and unspecified AI algorithm, for the detection of stroke and ICH on 'native brain CT'. Population included normal patients, intervention not specified, reference standard unclear and insufficient information to extract 2x2 data
Khunte 2020[60]	**Outcomes:** Conference abstract reporting on stand-alone performance of AIdoc for the detection of intracranial proximal vessel occlusion, which has not been listed in Table 2 because: outcomes (insufficient information to extract 2x2 data)
Kundisch 2021[61]	**Population**: Journal article reporting stan-alone accuracy of AIdoc compared to stand-alone accuracy of radiology reports for detecting ICH on NCCT. Population (all head CT, not stroke specific)
Love 2023[62]	**Outcomes:** Conference abstract reporting transfer rates pre- and post-implementation of Viz.ai. Outcomes (no data reported for specified outcomes)
Luijten 2022[63]	**Intervention:** Journal article reporting stand-alone accuracy of an unspecified automated algorithm for the detection of LVO. Intervention (not one of the specified AI interventions)
Marigold 2023[64]	**Outcomes:** Conference abstract, describing the implementation of Rapid AI, which does not report any numerical outcomes data. Outcomes (no data reported for specified outcomes)
Mathieson 2022[65]	**Population and outcomes:** Conference abstract, reporting a vignette study which assessed the performance of clinicians (radiologists and stroke physicians) to detect LVOs, based on vignettes and NCCT reports and CTAs with or without Brainomix e-CTA, from an enriched sample of scans that included 50% LVOs and stroke and non-stroke diagnoses; it was unclear how scans were selected for use in this study and the study used a *'non-stroke'* control group, rather than confirmed stroke, no LVO as would be indicated by the inclusion criteria for this assessment. The methods section of the abstract reports that 17 readers participated, and that each reader reviewed 20 cases, further stating that CTAs were randomly assigned to with or without eCTA and that readers could repeat scoring after two weeks with the inverse allocation of decision support. The results section stated that 220 case reviews were completed (unclear whether this number referred to 220 unique cases or to multiple readings of a smaller set of cases by different readers) and reported single point estimates of sensitivity and specificity for the detection of LVO with and without eCTA; it was unclear how these

Study Details	Brief description of study and reason for exclusion
	estimates had been derived from the multiple readers and cases included and whether or not they included results from repeat scoring. Population (includes non-stroke diagnoses) and outcomes (insufficient information to define the intervention and to extract 2x2 data; the total number of unique cases and numbers and type of individual diagnoses included were not reported). Study authors were contacted (31/05/2023), with a view to obtaining further information/clarification of methods; no response was received. Following the submission of the EAG addendum, additional information about this study was provided by Brainomix in response to questions prepared by the NICE team. The Brainomix response indicated that: • Cases were selected to ensure a balanced mix with and without LVO (10 of each) and that LVO cases were selected from a historical registry with occlusion locations to match the distribution in a published IPD meta-analysis of the effectiveness of mechanical thrombectomy. • Twenty patient cases with either stroke or non-stroke diagnoses were selected where the patient was 18 years of age or over, CTA image quality was adequate, demographic information and follow-up imaging were available and there was no evidence of ICH, and that (within these selected cases) ten were chosen with LVOs and ten without LVO. These two descriptions appear to be inconsistent and it is difficult to see how a sample selected as described in the second point could have contained sufficient patients with LVO. The response also provided further detail on the randomisation process. *'Readers viewed the 20 images in two sessions, spaced by at least two weeks. At the first session, half of the cases were selected at random to be presented with decision support from AI software (e-CTA), and the remaining half were presented with no decision support. At the second session, the allocation of decision support was reversed, so that all cases were reviewed with and without decision support from e-CTA.'* It was not clear whether this randomisation procedure was performed separately for each reader and only 9/17 readers completed the second read (allocation of decision support reversed). The accuracy estimates presented were calculated by treating each reading, including repeat readings, as an independent data point; thus data from 20 patients were used to derive accuracy estimates [redacted]; no per reader estimates were presented. Cases where the reader was uncertain and needed a second opinion were excluded from the analyses and there were fewer of these cases in the with AI data set. The effect of excluding 'uncertain' cases from the analyses depends upon the distribution of 'uncertain' results between positive and negative cases and how

Study Details	Brief description of study and reason for exclusion
	'uncertain' results would have been classified in a real-world clinical scenario (with a second opinion).
Matsoukas 2023[66]	**Reference standard:** Journal article reporting stand-alone accuracy of Viz LVO for detection of LVO, which has not been listed in Table 2 because: reference standard not as specified in scope (single radiologist review, not review by panel or expert review by neuroradiologist)
Nagamine 2022[67]	**Intervention:** Conference abstract reporting stand-alone accuracy of an unspecified AI tool for detection of LVO. Intervention not specified
Nicholas 2023[68]	**Outcomes:** Journal article reporting workflow metric for ICH patients pre- and post-implementation of Viz.ai. Outcomes (no data reported for specified outcomes)
Olive-Gadea 2022[69]	**Intervention, reference standard and outcomes:** Conference abstract reporting stand-alone accuracy of an unspecified AI-based software to detect LVO. Intervention (not one of the specified AI interventions), reference standard unclear (appeared to be CTP) and outcomes (insufficient information to extract 2x2 data)
Salokhiddinov 2022[70]	**Intervention:** Conference abstract reporting stand-alone accuracy of an unspecified AI-based algorithm for the detection of ICH on NCCT in patients with suspected acute stroke. Intervention not specified
Scavasine 2023[71]	**Outcomes:** Journal article reporting the accuracy of two individual neuroradiologists and two individual emergency department physicians, with and without e-ASPECTS for dichotomisation of ASPECTS scoring at a threshold of 10. Outcomes (no clinical decision threshold reported for any of the questions defined by the inclusion criteria for this assessment)
Stib 2020[72]	**Intervention:** Journal article reporting the development and validation of a 'convolutional neural network' to detect LVO. Intervention (not one of the specified AI interventions)
Voter 2021[73]	**Population:** Journal article reporting stand-alone accuracy of Aidoc for the detection of ICH. Population (all head CT, not stroke specific)
Weyland 2022[74]	**Outcomes:** Stand-alone accuracy of Brainomix and stand-alone accuracy of human readers for detection of hyperdense artery sign in patients with and without confirmed LVO. Outcomes (no data reported for specified outcomes)

APPENDIX 4: EAG QUESTIONS TO OAHSN

Reporting transparency:

- Please provide details of the stakeholder who requested that the publication of guidance be delayed and who submitted the OAHSN report
- Please provide authorship details for the OAHSN report
- The OAHSN report has 'branding' for both the OAHSN and Brainomix; please provide details of the role of Brainomix in this work (no details are included in the report)
- Please provide details of the commissioner and funding for the OAHSN report (no details are included in the report)

General information:

- Where data have been derived from clinician interviews or surveys (as indicated on page 5 of the report), please provide full details of survey methods including copies of any survey instruments/questionnaires/interview protocols used
- Please provide the dates of e-Stoke implementation for each site
- Please provide details of which components of e-Stroke (which of the e-Stroke tools) were implemented (at each site if there were differences between sites) – in particular, provide data on the extent of use of AI in interpretation of plain CT, CTA or CTP (number and proportion of eligible patients by site)
- Please provide details of how e-Stroke was incorporated in the care pathway (at each site if there were differences between sites) – please include details of type/seniority of clinician and the number and order of clinicians if more than one involved
- Please provide any information that you have (patient characteristics) to inform the question of whether patient populations were comparable before and after implementation of e-Stroke
- Noting that the implementation period appears to have overlapped the COVID-19 pandemic, please provide any information that you have about any changes in the care pathway that occurred (other than the implementation of e-Stroke) during the study period
- Page 4 refers to 'phase three of the evaluation' - please provide details of what constitutes phase 3 as opposed to phases 1 or 2 or any subsequent phase
- Page 5 under the heading Quantitative analysis states: 'We continue to receive clinical audit data via the Stroke Sentinel National Audit Programme (SSNAP)…'– please indicate if this is the only

source of quantitative data or, if not what the other sources are. Please also precisely specify the design of the quantitative analysis plan as well as all outcomes and methods of analysis.

Content (time to treatment and clinical outcomes):

- In relation to the table on page 8 of the report ('Rates and time to treatment – all ASCs, pre and post implementation of e-Stroke'), please provide:
    - Details of which components of e-Stroke were being used (at each site, if different) to inform judgements about IVT
    - The numbers of imaged patients (patients scanned) and the numbers who received IVT (pre- and post-implementation)
    - The IQR or range associated with the reported median times from scan to IVT (pre- and post-implementation) as well as the means and standard deviations
    - The dates (pre- and post-implementation), for each site, over which IVT data were collected
    - Any information that you have about the acute clinical outcomes (e.g., rates of haemorrhage) for patients who received IVT
    - Any information that you have about clinical outcomes (e.g., mRS) at specified follow-up times, for all imaged patients and for patients who received IVT

- In relation to the plot on page 16 of the report ('Median ambulance transfer times by Hospital and Is After Go Live'), please provide:
    - The IQR or range associated with the reported median transfer times (pre- and post-implementation) as well as the means and standard deviations
    - The dates (pre- and post-implementation), for each site, over which transfer time data were collected

- In relation to the table on page 19 of the report ('Rates of MT and DIDO times all ASCs – pre and post implementation of e-Stroke')
    - Noting that DIDO was not amongst the outcomes specified in the scope and protocol for this assessment; please provide an equivalent table reporting time from scan to MT for all ACSs
    - When providing these data, please include:

- The numbers of scanned patients and the numbers who received MT (pre- and post-implementation)
- The IQR or range associated with the median times from scan to MT (pre- and post-implementation) as well as the means and standard deviations
- The dates (pre- and post-implementation), for each site, over which MT data were collected
- Details of which components of e-Stroke were being used (at each site, if different) to inform judgements about transfer for MT

- In relation to the table on page 25 ('CSC Scan to MT times'), please provide:
  - Details of which components of e-Stroke were being used (at each site, if different) to inform judgements about MT
  - The numbers of scanned patients (pre- and post- implementation)
  - The IQR or range associated with the reported median times from scan to MT (pre- and post-implementation) as well as the means and standard deviations
  - The dates (pre- and post-implementation), for each site, over which MT data were collected

- In relation to the plot on page 31 of the report ('MT 6 months post discharge'), please provide:
  - Clarification on whether these data are for all patients who received MT (transferred from ASCs and presenting directly to CSCs)
  - Details of which CSCs contributed to these data and, if possible, site-level data
  - The numbers of patients who received MT and the numbers for whom 6-month mRS data were available (completeness of data)
  - The absolute numbers of patients in each mRS category, at 6-months post-discharge (pre- and post-implementation)
  - If possible, please provide mRS outcomes data stratified by patients who presented directly to CSCs and patients who were transferred from ASCs
  - The plot on page 35 shows mRS by whether received MT or not. Please provide mRS outcomes both pre- and post- implementation of e-Stroke for all imaged patients, including those patients who did not receive MT and those who received IVT

- In relation to the plot on page 34 of the report ('mRS score at discharge for patients following MT >6 hours Pre-Post AI'), please provide:

  - Clarification on whether these data are for all patients who received MT (transferred from ASCs and presenting directly to CSCs)
  - Details of which CSCs contributed to these data and, if possible, site-level data
  - The numbers of patients who received MT >6 hours after the onset of symptoms and the numbers for whom discharge mRS data were available (completeness of data)
  - The absolute numbers of patients in each mRS category, at discharge (pre- and post-implementation)
  - If possible, please provide mRS outcomes data stratified by patients who presented directly to CSCs and patients who were transferred from ASCs
  - Please also provide 6-month mRS data for these patients
  - Please provide the equivalent data sets for patients who received MT ≤6 hours after symptom onset

- In relation to the plot on page 34 ('Patient outcomes based on treatment mRS score on discharge October 2021 – September 2022'), the potential effects of combining IVT with MT are of interest, however, the data provided appear to be for a partial post-implementation period only; please provide comparative pre- and post- implementation data. For both data sets (pre- and post-implementation), please provide:
  - Clarification on whether these data are for all patients who received MT or MT+IVT (transferred from ASCs and presenting directly to CSCs)
  - Details of which CSCs contributed to these data and, if possible, site-level data
  - The numbers of patients who received MT and MT+IVT, and the numbers for whom discharge mRS data were available (completeness of data)
  - The absolute numbers of patients, who received MT and MT+IVT, in each mRS category at discharge (pre- and post-implementation)
  - If possible, please provide these mRS outcomes data stratified by patients who presented directly to CSCs and patients who were transferred from ASCs
  - If possible, please also provide these mRS data for the 6-month post-discharge time point

Content (clinician opinion/experience):

The outcomes specified in the scope and protocol for this assessment included 'ease of use/acceptability to clinicians' (measures not specified). The following plots could provide some information relevant to this outcome:

- The plot on page 28 of the report ('How would you define the general attitude of clinicians towards e-Stroke in your site').
- The plot on page 29 of the report ('How has e-Stroke affected confidence in your decision making')
- The plot on page 36 of the report ('Do you have concerns about the accuracy of e-Stroke')
- The plot on page 36 of the report ('Which functionality do you have concerns about').
- In relation to each of these plots, please provide:
    - Details of who was asked this question and how (e.g., in the context of a larger survey or interview, in person or digital)
    - The date/time since implementation at which the question was asked
    - The absolute numbers of people surveyed and respondents in each category (other, stroke nurse/ACP, radiologist/radiographer, interventional radiologist, consultant physician, stroke consultant)
    - Information about the representativeness of the survey - the numbers of respondents from each site, if possible, by category (other, stroke nurse/ACP, radiologist/radiographer, interventional radiologist, consultant physician, stroke consultant)