Neuroradiology (2015) 57:1–9
DOI 10.1007/s00234-014-1441-0

DIAGNOSTIC NEURORADIOLOGY

Observer reliability of CT angiography in the assessment of acute ischaemic stroke: data from the Third International Stroke Trial

Grant Mair · Rüdiger von Kummer · Alessandro Adami · Philip M. White · Matthew E. Adams · Bernard Yan · Andrew M. Demchuk · Andrew J. Farrall · Robin J. Sellar · Rajesh Ramaswamy · Daisy Mollison · Elena V. Boyd · Mark A. Rodrigues · Karim Samji · Andrew J. Baird · Geoff Cohen · Eleni Sakka · Jeb Palmer · David Perry · Richard Lindley · Peter A. G. Sandercock · Joanna M. Wardlaw · The IST-3 Collaborative Group

Received: 19 August 2014 / Accepted: 22 September 2014 / Published online: 7 October 2014

Abstract
Introduction CT angiography (CTA) is often used for assessing patients with acute ischaemic stroke. Only limited observer reliability data exist. We tested inter- and intra-observer reliability for the assessment of CTA in acute ischaemic stroke.
Methods We selected 15 cases from the Third International Stroke Trial (IST-3, ISRCTN25765518) with various degrees of arterial obstruction in different intracranial locations on CTA. To assess inter-observer reliability, seven members of the IST-3 expert image reading panel (>5 years experience reading CTA) and seven radiology trainees (<2 years experience) rated all 15 scans independently and blind to clinical data for: presence (versus absence) of any intracranial arterial abnormality (stenosis or occlusion), severity of arterial abnormality using relevant scales (IST-3 angiography score, Thrombolysis in Cerebral Infarction (TICI) score, Clot Burden Score), collateral supply and visibility of a perfusion defect on CTA source images (CTA-SI). Intra-observer reliability was assessed using independently repeated expert panel scan ratings. We assessed observer agreement with Krippendorff's-alpha (K-alpha).

Electronic supplementary material The online version of this article (doi:10.1007/s00234-014-1441-0) contains supplementary material, which is available to authorized users.

G. Mair · A. J. Farrall · R. J. Sellar · D. Mollison · E. Sakka · J. Palmer · J. M. Wardlaw (✉)
Division of Neuroimaging Sciences, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
e-mail: joanna.wardlaw@ed.ac.uk

R. von Kummer
Department of Neuroradiology, Dresden University Stroke Centre, University Hospital, Dresden, Germany

A. Adami
Stroke Center, Department of Neurology, Sacro Cuore-Don Calabria Hospital, Negrar, Italy

P. M. White
Stroke Research Group, Newcastle upon Tyne, UK

M. E. Adams
Department of Neuroradiology, National Hospital for Neurology and Neurosurgery, London, UK

B. Yan
Neurovascular Research Group, Royal Melbourne Hospital, Parkville, Australia

A. M. Demchuk
Department of Clinical Neurosciences, Calgary Stroke Program, Calgary, Canada

R. Ramaswamy · M. A. Rodrigues · K. Samji · A. J. Baird
Department of Radiology, Royal Infirmary of Edinburgh, Edinburgh, UK

E. V. Boyd
Department of Radiology, Northwick Park Hospital, Harrow, UK

G. Cohen · D. Perry · P. A. G. Sandercock
Division of Clinical Neurosciences, University of Edinburgh, Western General Hospital, Edinburgh, UK

R. Lindley
Westmead Hospital Clinical School and The George Institute for Global Health, University of Sydney, Sydney, Australia

Results Among experienced observers, inter-observer agreement was substantial for the identification of any angiographic abnormality (K-alpha=0.70) and with an angiography assessment scale (K-alpha=0.60–0.66). There was less agreement for grades of collateral supply (K-alpha=0.56) or for identification of a perfusion defect on CTA-SI (K-alpha=0.32). Radiology trainees performed as well as expert readers when additional training was undertaken (neuroradiology specialist trainees). Intra-observer agreement among experts provided similar results (K-alpha=0.33–0.72).
Conclusion For most imaging characteristics assessed, CTA has moderate to substantial observer agreement in acute ischaemic stroke. Experienced readers and those with specialist training perform best.

Keywords Cerebral angiography · Stroke · Inter-observer variability · Intra-observer variability

Introduction

Non-contrast CT (NCCT) is the most widely available imaging modality for assessing patients with acute stroke. Many centres now also perform CT angiography (CTA) as part of their stroke imaging protocol [1]. However, only limited observer reliability data exists for the reporting of CTA in acute stroke [2, 3]. A recent consensus statement on angiography grading standards for acute ischaemic stroke recommended that further reliability studies should be performed [4].

The Third International Stroke Trial (IST-3) was a multicentre, randomised controlled trial in 3035 patients that tested whether intravenous recombinant tissue plasminogen activator (rt-PA), given within 6 h of ischaemic stroke, improved functional outcome at 6 months [5]. Standardised brain imaging (predominantly NCCT) was mandatory for all IST-3 patients prior to randomisation in the trial. In some centres, CTA was also routinely obtained.

Using CTA to assess intracerebral arterial patency is limited by the lack of a grading scale developed specifically for cross-sectional imaging [6]. To date, most trials incorporating CTA have used one of the catheter angiography scales, e.g. Thrombolysis in Cerebral Infarction (TICI) [7]. However, there are two reasons why application of catheter angiography scales to CTA (or MR angiography) without modification, is problematic: 1) catheter angiography scales assess distal tissue perfusion, but perfusion is not appreciable on CTA (unless it is time resolved) [6] and 2) catheter angiography scales conflate features of cerebral arterial patency, flow and perfusion into one scale thereby potentially increasing sources of observer disagreement. Unlike catheter angiography, standard CTA provides only a snapshot in time and can only be used to assess arterial patency rather than flow. A new IST-3 angiography score was developed in an attempt to overcome the limitations of applying catheter angiography scores to CTA. The IST-3 angiography score aims to assess only those characteristics of angiography that are identifiable on CTA, especially arterial luminal patency at the main point of occlusion [6].

Our primary aim was to investigate inter- and intra-observer reliability of expert readers assessing CTA in acute ischaemic stroke. We also sought to establish how less-experienced readers perform and to evaluate a new CTA grading scale, the IST-3 angiography score.

Materials and methods

The Third International Stroke Trial

IST-3 was an international, multicentre, prospective, randomised, open, blinded endpoint (PROBE) trial of intravenous rt-PA in acute ischaemic stroke. Enrolment, data collection and image analysis have been fully described elsewhere [6, 8]. Briefly, patients with acute stroke of any severity were eligible for inclusion in the trial if intravenous rt-PA (alteplase) could be started within 6 h of known stroke onset, and CT or MR imaging had reliably excluded both intracranial haemorrhage and any structural stroke mimic. Patients were aged 18 years or above with no upper age limit. Informed consent was obtained from all patients. IST-3 is registered, ISRCTN25765518.

Scan acquisition and management

Prior to joining IST-3, all centres had to submit a test scan to ensure adequate image acquisition parameters. Minimum acquisition standards were specified in the trial protocol and centre participation criteria. All scans were checked against quality assurance standards centrally. Due to the large number of participating sites, CT scans were inevitably obtained from many different scanners (and generations of scanner) across the trial.

In centres where CTA was routinely performed in the assessment of acute ischaemic stroke, these images were also submitted to the IST-3 central trials office. Subgroup analysis of IST-3 angiography was pre-specified [6].

Once received by the IST-3 central trials office, all images were anonymised and uploaded to a local server. Image analysis was undertaken using the *Systematic Image Review System 2* (SIRS2). The use of this system for remote multi-reader scan assessment has been fully described [9, 10]. Briefly, SIRS2 provides an environment for viewing images via a web browser (available at www.neuroimage.co.uk). Scan ratings are entered simultaneously and are automatically submitted securely to the trial database. Users are assigned specific image datasets upon which standard

image manipulation functions (e.g. zoom, pan and scroll) can be applied. SIRS2 allows multiple images to be viewed in parallel. Scan ratings were entered on a structured pro forma: www.sbirc.ed.ac.uk/research/imageanalysis.html [6].

Image assessment panel

All imaging in IST-3 were assessed centrally by a panel of expert readers comprising neuroradiologists, neurologists and stroke physicians with extensive experience in the assessment of acute stroke imaging. All readers underwent scan rating training prior to joining the panel by completing the ACCESS study [9, 10]. All readers were completely blinded to all clinical information including stroke symptoms, treatment allocation, time after stroke and any other image data acquired at different time points.

Image analysis in IST-3

Non-contrast CT assessment NCCT was evaluated for the extent, depth and location of acute ischaemia using an IST-3 scale and the Alberta Stroke Program Early CT Score (ASPECTS) [11, 12], ischaemic tissue swelling, the presence and location of any hyperattenuated artery and background pre-stroke brain changes (brain atrophy, leukoaraiosis, prior infarct or haemorrhage) [13, 14], using validated qualitative scales (details Table 1). The IST-3 scale grades infarct location and extent in any arterial territory, with up to eight categories in the middle cerebral artery (MCA) territory. ASPECTS is a 10-point scale developed to assess infarct extent within the MCA territory where points are deducted for each of the ten MCA territory regions involved; the anterior cerebral artery (ACA) and PCA territories can be included by adding 1 point for each.

CTA assessment CTA was firstly categorised as 'normal' or 'abnormal' (when any arterial stenosis and/or occlusion was identified in any intracranial location). A modified version of the TICI and IST-3 angiography (modified Mori) scales was then applied [6, 7, 15]. Both are scalar and range from occlusion (0) through lesser grades of obstruction to normal patency (3 for TICI, 4 for IST-3) as detailed in Table 1.

The *Clot Burden Score* assesses the extent of contrast deficits (as a surrogate for clot) in the internal carotid artery (ICA), MCA and ACA [16]. From an initial score of 10, points are deducted for each vessel segment involved; a score of 0 implies all segments of all named vessels are occluded.

The quality of *collateral vessel supply* in patients with ICA or MCA occlusion was categorised as good, moderate or poor [17].

CTA source images (CTA-SI) were assessed for deficits in contrast enhancement of brain tissue as a surrogate of impaired cerebral blood flow (CBF) and low cerebral blood volume (CBV), indicative of infarction. Extent of any perfusion deficit on CTA-SI was categorised using ASPECTS [18, 19].

Observer reliability analysis

Selection of cases We identified 15 cases from the IST-3 angiography subgroup that had both NCCT and concurrent CTA performed pre-randomisation. Time-resolved CTA was not included. These 15 cases were chosen to represent a range of angiographic findings (e.g. presence/absence of arterial obstruction in various locations, clot burden) including normal appearances based on the consensus opinion of three senior neuroradiologists (details Table 2). In three of these cases, angiography was deemed to be normal. In the remaining 12 cases, arterial obstruction of varying severity (TICI 1-2b) was identified in an ICA (n=4), in an MCA (n=7) or in the basilar artery (n=1). Clot Burden scores ranged from 1 to 10. There were no significant differences between the full IST-3 CTA subgroup (n=269) and the 15 cases selected for reliability analysis for the following variables (full subgroup data are presented): age (median 81 years), sex (56 % female), National Institutes of Health Stroke Scale (median 10) and time from stroke onset to scan (median 170 min).

Selection of readers We identified 14 readers comprised of seven (of the original ten) expert IST-3 angiography panel members (each with greater than 5 years of experience in assessing CTA in acute stroke) and seven non-expert readers (radiology trainees with less than 2 years of experience in assessing CTA).

Scan rating All 15 cases for reliability analysis were independently rated by the 14 readers. These scan ratings (of both NCCT and CTA) were performed purely to assess reader reliability and were undertaken in addition to, and separate from, scan ratings performed during the main IST-3 trial and the IST-3 angiography subgroup analysis.

Inter-observer reliability comparisons Three distinct inter-observer analyses were performed.

1. The expert panel inter-observer reliability analysis compared seven observers, i.e. maximum 315 pairs of readings for each imaging characteristic assessed (21 reader pairs×15 cases).
2. The non-expert panel (n=7) was assessed in a separate but identical analysis as for the expert panel (maximum 315 pairs of readings).
3. To ascertain whether additional neuroradiology training might improve the inter-observer reliability of non-expert CTA readers, the results of neuroradiology specialist

Table 1 Scoring options for imaging characteristics available to users on SIRS

Imaging characteristic	Options presented to image reader
Acute ischaemia	None
	Left or right sided only, both left and right sided
	Midline
ASPECTS [13]	Ten MCA territory divisions available:
	Cortical divisions (M1–M6)
	Insula, caudate head, lentiform nucleus, internal capsule
IST-3 ischaemia score [12]	0=None
(condensed code)	1=Small cortical, border zone or lacunar infarct, etc.
	2=More than half ACA territory or >2 cm within basal ganglia, etc.
	3=More than half cerebellar hemisphere or peripheral MCA territory, etc.
	4=Entire MCA territory, etc.
Depth of ischaemia	0=None
	1=Loss of grey-white matter differentiation
	2=Hypodense to normal brain
Brain swelling	0=None
	1=Effacement of local sulci
	2=1+minor effacement of lateral ventricle
	3=1+complete effacement of lateral ventricle
	4=1+effacement of lateral and third ventricles
	5=4+midline shift
	6=5+effacement of basal cisterns
TICI (modified) [7]	0=No flow/patency
	1=Minimal flow/patency
	2a=Partial flow <50 % of expected territory
	2b=Partial flow >50 % of expected territory
	3=Complete flow/patency
IST-3 angiography score [6]	0=No patency
(modified Mori [16])	1=Contrast penetrates obstruction but minimal enters distal artery
	2a=Luminal patency of <50 % with partial filling (<half) of major branches
	2b=Luminal patency of <50 % with partial filling (>half) of major branches
	3=Patency of >50 % of lumen and filling of most branches
	4=Complete patency—normal
Clot Burden Score [17]	Six arterial locations are assessed for the presence of clot:
	Infraclinoid (1 point) and supraclinoid ICA (2 points)
	Proximal (2 points) and distal M1 MCA (2 points)
	M2 branches of MCA (1 point each)
	ACA (1 point)
Collateral supply (MCA) [18]	Good=entire MCA distal to occlusion fills with contrast
	Moderate=some of the MCA branches in the sylvian fissure fill with contrast
	Poor=only distal superficial MCA branches fill with contrast
Brain atrophy [14]	0=None, 1=modest, 2=severe
	Assessed for both central and cortical regions
Leukoaraiosis [15]	0=None, 1=periventricular only, 2=from ventricle to cortical surface
	Assessed for both anterior and posterior regions

SIRS Systematic Image Review System, *ASPECTS* Alberta Stroke Program Early CT Score, *MCA* middle cerebral artery, *ACA* anterior cerebral artery, *TICI* Thrombolysis in Cerebral Infarction

The coding forms are available at www.sbirc.ed.ac.uk/research/imageanalysis.html

trainees (n=3) from within the non-expert panel were separately examined as a subgroup, i.e. maximum 45 pairs (3×15) of readings for each imaging characteristic.

Intra-observer reliability comparisons We used expert panel readings performed during the primary IST-3 angiography rating (i.e. prior to this observer reliability study) for intra-observer

Table 2 Consensus opinion of three senior neuroradiologists for CTA results of the 15 selected cases

Scan	CTA abnormal (yes/no)	Location of CTA abnormality	TICI score (0–3)	IST-3 angiography score (0–4)	Clot Burden Score (10–0)	Collaterals (good, moderate, poor)	CTA-SI deficit (yes/no)
1	Yes	ICA	1	1	7	Poor	Yes
2	No	–	3	4	10	Good	No
3	Yes	M1	2a	2a	7	Moderate	Yes
4	Yes	M1	2b	2b	8	Good	Yes
5	Yes	M2	2a	2a	9	Moderate	Yes
6	Yes	ICA	2b	3	8	Good	Yes
7	Yes	ICA	1	1	1	Good	Yes
8	Yes	M1	2b	3	8	Good	No
9	No	–	3	4	10	Good	No
10	Yes	M1	2a	2a	7	Moderate	No
11	Yes	ICA	2b	2b	7	Good	Yes
12	Yes	Basilar	2b	3	9	Good	Yes
13	Yes	M2	1	1	8	Moderate	Yes
14	Yes	M1	2b	2b	8	Good	Yes
15	No	–	3	4	10	Good	Yes

TICI Thrombolysis in Cerebral Infarction, *CTA-SI* CT angiography source image, *ICA* internal carotid artery, *M1* horizontal segment of the middle cerebral artery (MCA), *M2* sylvian branch of the MCA

reliability analysis. Each of the seven experts had read at least one of the same 15 cases in the primary angiography assessment, and that reading could be compared with their subsequent readings performed specifically for this observer reliability analysis (Fig. 1). The readers had no knowledge of their previous scan assessment or even that a previous assessment of the same case was undertaken. All scan reads for intra-observer analysis were separated in time by at least 4 weeks, but in many cases, up to 1 year passed between scan reads.

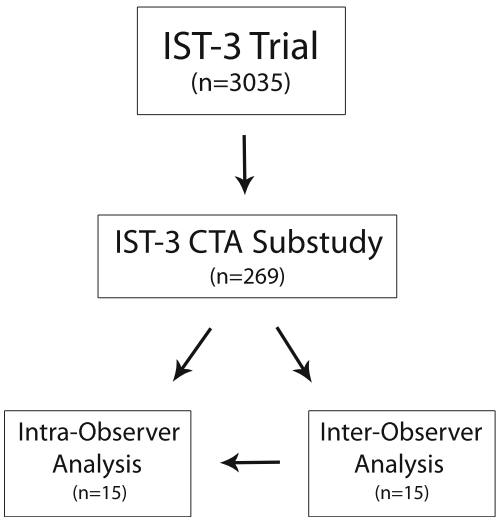

Fig. 1 Flowchart demonstrating the source of the 15 cases for reliability analysis. All 15 cases for inter-observer analysis were chosen from within the IST-3 angiography subgroup. The seven expert readers for inter-observer analysis were all members of the IST-3 image reading panel for the CTA substudy. Intra-observer analysis compares expert panel scan reads (from the same individuals) performed during the substudy with those performed for inter-observer analysis

Statistical analysis

All reliability analyses were performed using Krippendorff's alpha (K-alpha) with 1000 bootstrap samples for each. K-alpha results range from −1.0 to +1.0 where +1.0 equates to perfect agreement, 0.0 means no agreement and −1.0 implies perfect disagreement [20]. We have adopted the Landis and Koch approach for interpreting these results such as K-alpha 0.00–0.20=slight agreement, 0.21–0.40=fair agreement, 0.41–0.60=moderate agreement, 0.61–0.80=substantial agreement and 0.81–1.00=almost perfect agreement [21]. Differences in K-alpha between expert and non-expert groups (including between neuroradiology specialist trainees and others) and between imaging characteristics assessed on NCCT and CTA were not tested for significance.

All analyses were performed using IBM SPSS Statistics software, version 20.0 (IBM Corporation, Armonk, NY, USA). SPSS does not provide native support for K-alpha; an appropriate macro was applied [22].

Results

CTA inter-observer agreement

Inter-observer reliability analyses for the assessment of CTA among expert and non-expert readers (n=7 for both groups)

are displayed in Table 3. The IST-3 angiography expert panel had moderate to substantial agreement between readers for all CTA measures (K-alpha 0.56–0.70) except identification of a perfusion deficit on CTA-SI (K-alpha=0.32). Non-expert readers had only fair to moderate agreement (K-alpha 0.25–0.61) for all CTA variables. Among non-experts, neuroradiology specialist trainees' (n=3) agreement compared more favourably with the expert group (K-alpha 0.36–0.78).

Inter-observer agreement among experts and among non-experts with additional training was greatest for assessing whether CTA was 'normal' or 'abnormal' (i.e. any intracranial arterial stenosis or occlusion). For assessing the extent of angiographic abnormality, IST-3 scoring performed better than TICI in all groups, although the 95 % confidence intervals overlap, so this difference is unlikely to be significant. Expert panel IST-3 angiography scores for all 15 cases are presented in Table 4. Note that while there are some discrepancies, most of the disagreements do not extend by more than 2 points.

The assessment of CTA collateral supply and the identification of a perfusion deficit on CTA-SI scored lowest for inter-observer agreement in all groups.

CTA intra-observer agreement

Intra-observer agreement for the expert panel assessment of CTA (K-alpha 0.33–0.72) was generally similar to their inter-observer results (above). However, the wide confidence intervals for intra-observer analysis suggest these results may be underpowered (Table 3).

Observer agreement for NCCT versus CTA

Online appendix 1 provides results of the inter- and intra-observer reliability analyses for NCCT findings of the angiography expert panel. These results show fair to moderate agreement for most imaging characteristics. Identification and classification of ischaemia (using either ASPECTS or the IST-3 ischaemia score) showed the best agreement (K-alpha 0.56–0.66 for both inter- and intra-observer analyses). Identification of a hyperattenuated artery showed only fair inter-observer agreement but almost perfect intra-observer agreement (K-alpha 0.37 and 0.83, respectively).

Figure 2 compares IST-3 expert panel inter-observer agreement for NCCT and CTA findings. Agreement was generally greater for the assessment of CTA features (K-alpha 0.32–0.70) than for NCCT features (K-alpha 0.13–0.66) although the ranges were similar. In addition, four of the top five agreement scores were for imaging characteristics assessed on CTA.

Discussion

In this study, where 14 observers with differing levels of experience assessed a purposive sample of 15 examinations, we show that CTA features have slightly higher levels of agreement than non-contrast CT features. Imaging characteristics that are likely to have the greatest clinical impact (e.g. the presence and severity of arterial occlusion) are reported with the highest inter-observer agreement, both by experienced (K-alpha >0.60) and inexperienced observers. There was less agreement over arterial collateral supply and use of CTA-SI to identify perfusion deficits, even among experienced observers (K-alpha 0.30–0.60). Despite being comparatively inexperienced, the participating radiology trainees that had undertaken additional neuroradiology training (neuroradiology fellows) performed as well as experts in the assessment of CTA. This implies that, with adequate training, CTA can be reliably assessed even by readers with less experience.

Table 3 Observer reliability analyses for CT angiography (CTA) among expert and non-expert readers

	IST-3 expert panel readers		Non-expert readers	
	n=7		n=7	Neuroradiology trainees, n=3
Total stroke cases	15 Cases			
Inter or intra-observer	Inter (315 pairs)	Intra (15 pairs)	Inter (315 pairs)	Inter (45 pairs)
CTA abnormal (yes/no)	0.70 (0.52–0.85)	0.54 (0.08–1.00)	0.49 (0.32–0.66)	0.78 (0.56–0.94)
TICI score (0–3)	0.60 (0.53–0.66)	0.60 (0.12–0.94)	0.39 (0.28–0.50)	0.63 (0.42–0.81)
IST-3 angiography score (0–4)	0.66 (0.60–0.70)	0.63 (0.16–0.91)	0.43 (0.32–0.52)	0.72 (0.52–0.87)
Clot Burden Score (10–0)	0.63 (0.56–0.70)	0.33 (−0.28–0.87)	0.61 (0.52–0.69)	0.76 (0.60–0.90)
Collaterals (good, moderate, poor)	0.56 (0.44–0.67)	0.72 (0.54–0.91)	0.30 (0.14–0.42)	0.36 (0.06–0.64)
CTA-SI deficit (yes/no)	0.32 (0.13–0.52)	0.57 (−0.29–1.00)	0.25 (0.07–0.45)	0.60 (0.35–0.80)

Both inter and intra-observer results are provided. Results represent K-alpha (95 % confidence interval). Maximum number of possible result pairs in each calculation is provided

TICI Thrombolysis in Cerebral Infarction, *CTA-SI* CT angiography source images

Table 4 Comparison of IST-3 angiography scores from 15 scans as assessed by seven members of the IST-3 expert panel

Scan	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6	Reader 7
1	1	0	0	1	0	0	1
2	4	4	4	4	4	4	4
3	2b	0	1	2a	0	0	0
4	2b	0	0	2b	0	0	0
5	2a	0	0	2b	0	2a	2b
6	3	2b	1	3	4	3	4
7	1	1	0	2b	0	0	1
8	3	4	4	3	4	3	4
9	4	4	4	4	4	4	4
10	2a	1	4	2b	0	3	1
11	2b	0	1	2b	0	0	0
12	1	4	3	3	1	0	4
13	1	0	0	1	0	0	0
14	2b	1	2b	2b	0	0	2a
15	3	4	4	4	4	4	4

K-alpha for this inter-observer analysis (315 pairs of readings) is 0.60 (95 % CI=0.53–0.66). 0= No patency, 1=contrast penetrates obstruction but minimal enters the distal artery, 2a=luminal patency of <50 % with partial filling (<half) of major branches, 2b=luminal patency of <50 % with partial filling (>half) of major branches, 3=patency of >50 % of lumen and filling of most branches and 4=complete patency

The IST-3 angiography score is an adaptation of earlier scores (TICI, Mori). It is designed to overcome the limitations of using a catheter angiography score for the assessment of CTA by primarily assessing residual arterial calibre at the point of stenosis and contrast penetration into the major distal vessels only and makes no attempt to assess distal tissue perfusion [6]. The present work represents the first external testing of observer reliability for the IST-3 angiography score, and it compares favourably with TICI.

To the best of our knowledge, there are only a few previous studies of CTA reliability in stroke; all had fewer than seven observers, and none tested all the CTA signs assessed in our study. Knauth et al. reported an inter-reader kappa=0.78 for two readers identifying the correct location of occlusion on CTA in acute ischaemic stroke [23]. Suh et al. compared TICI versus a modified TICI score and found both scales were moderately repeatable (intra-class coefficients (ICC), 0.67 and 0.73, respectively) across five readers [24]. We did not replicate the inter- and intra- observer reliability demonstrated by Puetz and colleagues in their original report of the Clot Burden Score (six readers, ICC=0.87 and 0.96, respectively) despite similar reader numbers [16]. Similarly, in the original report defining their classification of collateral status, Miteff and colleagues demonstrated an inter-observer reliability of

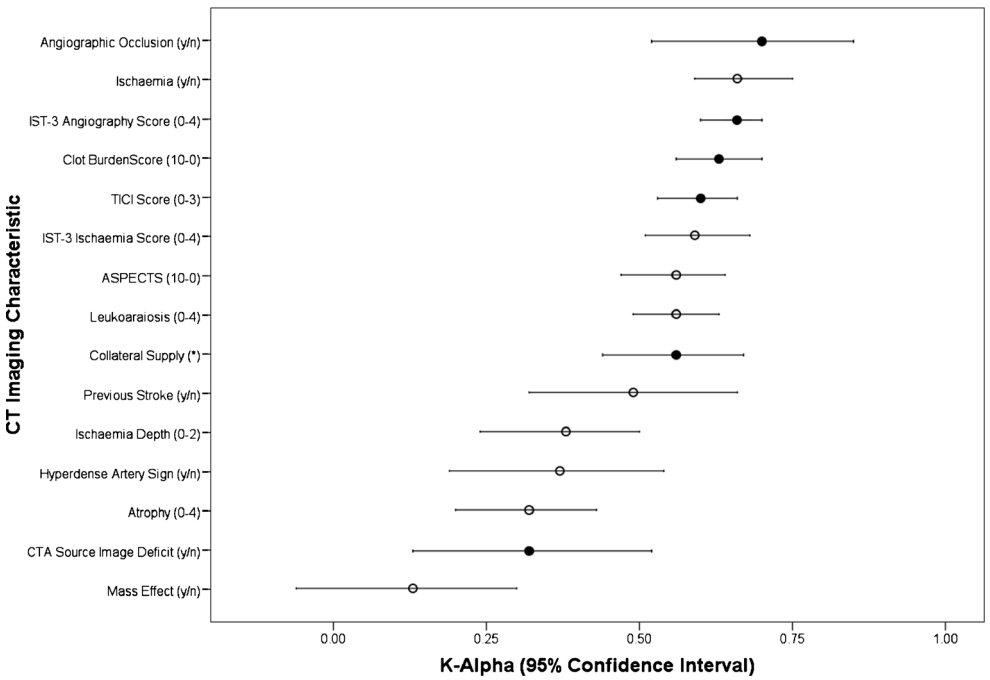

Fig. 2 IST-3 angiography expert panel inter-observer reliability results for imaging characteristics of both non-contrast CT and CTA. K-Alpha of 1.0=perfect agreement, 0.0=no agreement, −1.0=perfect disagreement. *Closed circles* represent imaging characteristics identified at CTA. *Open circles* represent imaging characteristics identified on non-contrast CT. Ischaemia defined as loss of grey-white matter differentiation or parenchymal hypodensity. *TICI* Thrombolysis in Cerebral Infarction, *ASPECTS* Alberta Stroke Program Early CT Score. *Collateral supply is ranked as good, moderate or poor.

kappa=0.93 for two observers [17]. We were unable to replicate those findings, but our results are more consistent with other methods of assessing leptomeningeal flow as demonstrated on a systematic review (0.49–0.87) [3]. Neither did we replicate the results from three recent articles, each with four readers, that demonstrated improved detection of infarct using CTA-SI over NCCT alone; Hopyan et al. improved reader agreement from kappa 0.28–0.44 to 0.34–0.57 [25], Finlayson et al. showed an increase in ICC from 0.83 to 0.88 [26] while van Seeters and colleagues improved their ICC range from 0.54–0.62 to 0.57–0.76 [27].

These previous studies represent a mixture of kappa statistics and ICC and are not directly comparable with our K-alpha results; any comparisons should be treated with caution. Nevertheless, kappa, ICC and K-alpha work on the same numerical scale and are therefore broadly similar. We opted to use K-alpha for several reasons. Kappa is only suitable for assessing two observers rating nominal data and even then may not be the most suitable test [28, 29]; we had up to seven observers per analysis and a mixture of nominal and ordinal data. K-alpha has been shown to provide a more robust measure of observer variance than kappa or ICC and provides several advantages to the user; it allows comparisons between any number of observers, it can handle both categorical and ordinal data, it is less prone to the influences of observer bias and result prevalence and it can still be computed in the presence of missing data [20, 30].

Other strengths of our work include more readers than in previous studies; calculation of both inter- and intra-reader reliability; use of a robust, standardised image analysis platform, previously shown to provide consistent multiuser reporting [9, 10]; complete blinding of readers to all clinical information and to any other scan assessments and use of representative cases from a multicentre trial which increases the generalisability and real world relevance of our results.

Our work also has some limitations. Firstly, in contrast to previous work [10], we did not formally produce a single reference standard for the 'correct' interpretation of the 15 scans to compare with other readers. Use of a reference standard would have allowed us to assess reader accuracy in addition to reader reliability. The results in Table 2 represent the consensus opinion of three senior neuroradiologists but are nevertheless still open to interpretation error. By confirming high observer agreement among a group of seven experienced readers, including several senior neuroradiologists, we believe that our results are as informative as reader comparisons set against any reference standard created from the same data. We do however acknowledge the possibility that the expert panel was reliable in making false diagnoses but feel this is highly unlikely. Secondly, several of the characteristics we tested in our intra-observer analyses are probably underpowered.

Conclusions

Experienced observers report CTA in acute ischaemic stroke with substantial levels of agreement for most imaging characteristics. Non-expert readers perform well if given specialist training. The IST-3 angiography score is reported as reliably as TICI and has some face validity and practical advantages for the assessment of CTA.

Acknowledgments We thank all IST-3 collaborators, including the National Co-ordinators and Participating Centres, Steering Committee, Data Monitoring Committee and Event Adjudication Committee. The IST-3 collaborative group wishes chiefly to acknowledge all patients who participated in the study, and the many individuals not specifically mentioned in the paper that have provided support.

IST-3 was funded from many sources; these are listed in online appendix 2. The views expressed in this work are those of the authors and not of the funding sources.

Ethical standards and patient consent We declare that all human studies have been approved by the Multicentre Research Ethics Committees, Scotland (reference MREC/99/0/78) and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. We declare that all patients gave informed consent prior to inclusion in this study.

Conflict of interest RvK consults for Lundbeck, Penumbra, Covidien, Brainsgate and Boehringer Ingelheim. BY consults for Boehringer Ingelheim, Bayer and Pfizer. RIL consults for Boehringer Ingelheim and Covidien. PAGS consults for Boehringer Ingelheim.

References

1. McDonald JS, Fan J, Kallmes DF, Cloft HJ (2013) Pretreatment advanced imaging in patients with stroke treated with IV Thrombolysis: evaluation of a multihospital data base. AJNR Am J Neuroradiol. doi:10.3174/ajnr.A3797:
2. Tomsick T (2007) TIMI, TIBI, TICI: I came, I saw, I got confused. AJNR Am J Neuroradiol 28:382–384
3. McVerry F, Liebeskind DS, Muir KW (2012) Systematic review of methods for assessing leptomeningeal collateral flow. AJNR Am J Neuroradiol 33:576–582
4. Zaidat OO, Yoo AJ, Khatri P, Tomsick TA, von Kummer R, Saver JL, Marks MP, Prabhakaran S, Kallmes DF, Fitzsimmons BFM, Mocco JD, Wardlaw JM, Barnwell SL, Jovin TG, Linfante I, Siddiqui AH, Alexander MJ, Hirsch JA, Wintermark M, Albers G, Woo HH, Heck DV, Lev M, Aviv R, Hacke W, Warach S, Broderick J, Derdeyn CP, Furlan A, Nogueira RG, Yavagal DR, Goyal M, Demchuck AM, Bendszus M, Liebeskind DS, Cerebral Angiographic Revascularization Grading (CARG) Collaborators, STIR Revascularization working group, STIR Thrombolysis in Cerebral Infarction (TICI) Task Force (2013) Recommendations on angiographic revascularization grading standards for acute ischemic stroke. A consensus statement. Stroke. doi:10.1161/STROKEAHA.113.001972:
5. The IST-3 Collaborative Group (2012) The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. Lancet 379:2352–2363

6. Wardlaw JM, von Kummer R, Carpenter T, Parsons M, Lindley R, Cohen G, Murray V, Kobayashi A, Peeters A, Chappell F, Sandercock P (2013) Protocol for the perfusion and angiography imaging sub-study of the Third International Stroke Trial (IST-3) of alteplase treatment within six hours of acute ischemic stroke. Int J Stroke. doi:10.1111/j.1747-4949.2012.00946.x
7. Higashida RT, Furlan AJ (2003) Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. Stroke 34:1923–1924
8. Sandercock P, Lindley R, Wardlaw J, Dennis M, Lewis S, Venables G, Kobayashi A, Czlonkowska A, Berge E, Slot KB, Murray V, Peeters A, Hankey G, Matz K, Brainin M, Ricci S, Celani MG, Righetti E, Cantisani T, Gubitz G, Phillips S, Arauz A, Prasad K, Correia M, Lyrer P (2008) Third international stroke trial (IST-3) of thrombolysis for acute ischaemic stroke. Trials 9:37
9. Wardlaw JM, Farrall AJ, Perry D, von Kummer R, Mielke O, Moulin T, Ciccone A, Hill M, for the Acute Cerebral CT Evaluation of Stroke Study (ACCESS) Study Group (2007) Factors influencing the detection of early computed tomography signs of cerebral ischemia. An internet-based, international multiobserver study. Stroke 38:1250–1256
10. Wardlaw JM, von Kummer R, Farrall AJ, Chappell FM, Hill M, Perry D (2010) A large web-based observer reliability study of early ischaemic signs on computed tomography. The Acute Cerebral CT Evaluation of Stroke Study (ACCESS). PLoS One 5:e15757
11. Wardlaw JM, Sellar RJ (1994) A simple practical classification of cerebral infarcts on CT and its interobserver reliability. AJNR Am J Neuroradiol 15:1933–1939
12. Barber PA, Demchuk AM, Zhang J, Buchan AM (2000) Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. Lancet 355:1670–1674
13. Farrell C, Chappell F, Armitage PA, Keston P, MacLullich A, Shenkin S, Wardlaw JM (2008) Development and initial testing of normal reference MR images for the brain at ages 65–70 and 75–80 years. Eur Radiol 19:177–183
14. van Swieten JC, Hijdra A, Koudstaal PJ, van Gijn J (1990) Grading white matter lesions on CT and MRI: a simple scale. J Neurol Neurosurg Psychiatry 53:1080–1083
15. Mori E, Yoneda Y, Tabuchi M, Yoshida T, Ohkawa S, Ohsumi Y, Kitano K, Tsutsumi A, Yamadori A (1992) Intravenous recombinant tissue plasminogen activator in acute carotid artery territory stroke. Neurology 42:976–982
16. Puetz V, Dzialowski I, Hill MD, Subramaniam S, Sylaja PN, Krol A, O'Reilly C, Hudon ME, Hu WY, Coutts SB, Barber PA, Watson T, Roy J, Demchuk AM (2008) Intracranial thrombus extent predicts clinical outcome, final infarct size and hemorrhagic transformation in ischemic stroke: the clot burden score. Int J Stroke 3:230–236
17. Miteff F, Levi CR, Bateman GA, Spratt N, McElduff P, Parsons MW (2009) The independent predictive utility of computed tomography angiographic collateral status in acute ischaemic stroke. Brain 132:2231–2238
18. Lev MH, Segal AZ, Farkas J, Hossain ST, Putman C, Hunter GJ, Budzik R, Harris GJ, Buonanno FS, Ezzeddine MA, Chang Y, Koroshetz WJ, Gonzalez RG, Schwamm LH (2001) Utility of perfusion-weighted CT imaging in acute middle cerebral artery stroke treated with intra-arterial thrombolysis: prediction of final infarct volume and clinical outcome. Stroke 32:2021–2028
19. Aviv RI, Shelef I, Malam S, Chakraborty S, Sahlas DJ, Tomlinson G, Symons S, Fox AJ (2007) Early stroke detection and extent: impact of experience and the role of computed tomography angiography source images. Clin Radiol 62:447–452
20. Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. Commun Methods Meas 1:77–89
21. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
22. Hayes AF (2012) K-Alpha SPSS macro, Version 3.1. Available from: http://afhayes.com/spss-sas-and-mplus-macros-and-code.html. Accessed: August 3 2014
23. Knauth M, von Kummer R, Jansen O, Hahnel S, Dorfler A, Sartor K (1997) Potential of CT angiography in acute ischemic stroke. AJNR Am J Neuroradiol 18:1001–1010
24. Suh SH, Cloft HJ, Fugate JE, Rabinstein AA, Liebeskind DS, Kallmes DF (2013) Clarifying differences among thrombolysis in cerebral infarction scale variants: is the artery half open or half closed? Stroke 44:1166–1168
25. Hopyan J, Ciarallo A, Dowlatshahi D, Howard P, John V, Yeung R, Zhang L, Kim J, Macfarlane G, Lee TY, Aviv RI (2010) Certainty of stroke diagnosis: incremental benefit with CT perfusion over noncontrast CT and CT angiography. Radiology 255:142–153
26. Finlayson O, John V, Yeung R, Dowlatshahi D, Howard P, Zhang L, Swartz R, Aviv RI (2013) Interobserver agreement of ASPECT Score distribution for noncontrast CT, CT angiography, and CT perfusion in acute stroke. Stroke 44:234–236
27. van Seeters T, Biessels GJ, Niesten JM, van der Schaaf I, Dankbaar JW, Horsch AD, Mali WP, Kappelle LJ, van der Graaf Y, Velthuis BK (2013) Reliability of visual assessment of non-contrast CT, CT angiography source images and CT perfusion in patients with suspected ischemic stroke. PLoS One 8:e75615
28. Byrt T, Bishop J, Carlin JB (1993) Bias, prevalence and kappa. J Clin Epidemiol 46:423–429
29. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL (2013) Clinicians are right not to like Cohen's kappa. BMJ 346:f2125
30. Hallgren KA (2012) Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 8:23–34