



# ASPECTS Interobserver Agreement of 100 Investigators from the TENSION Study

Noel van Horn<sup>1</sup> · Helge Kniep<sup>1</sup> · Gabriel Broocks<sup>1</sup> · Lukas Meyer<sup>1</sup> · Fabian Flottmann<sup>1</sup> · Matthias Bechstein<sup>1</sup> · Julia Götz<sup>1</sup> · Götz Thomalla<sup>2</sup> · Martin Bendszus<sup>3</sup> · Susanne Bonekamp<sup>3</sup> · Johannes Alex Rolf Pfaff<sup>3</sup> · Paulo Roberto Dellani<sup>4</sup> · Jens Fiehler<sup>1</sup> · Uta Hanning<sup>1</sup>

Received: 16 September 2020 / Accepted: 14 December 2020 / Published online: 27 January 2021  
© The Author(s) 2021

## Abstract

**Purpose** Evaluating the extent of cerebral ischemic infarction is essential for treatment decisions and assessment of possible complications in patients with acute ischemic stroke. Patients are often triaged according to image-based early signs of infarction, defined by Alberta Stroke Program Early CT Score (ASPECTS). Our aim was to evaluate interrater reliability in a large group of readers.

**Methods** We retrospectively analyzed 100 investigators who independently evaluated 20 non-contrast computed tomography (NCCT) scans as part of their qualification program for the TENSION study. Test cases were chosen by four neuroradiologists who had previously scored NCCT scans with ASPECTS between 0 and 8 and high interrater agreement. Percent and interrater agreements were calculated for total ASPECTS, as well as for each ASPECTS region.

**Results** Percent agreements for ASPECTS ratings was 28%, with interrater agreement of 0.13 (95% confidence interval, CI 0.09–0.16), at zero tolerance allowance and 66%, with interrater agreement of 0.32 (95% CI: 0.21–0.44), at tolerance allowance set by TENSION inclusion criteria. ASPECTS region with highest level of agreement was the insular cortex (percent agreement=96%, interrater agreement=0.96 (95% CI: 0.94–0.97)) and with lowest level of agreement the M3 region (percent agreement=68%, interrater agreement=0.39 [95% CI: 0.17–0.61]).

**Conclusion** Interrater agreement reliability for total ASPECTS and study enrollment was relatively low but seems sufficient for practical application. Individual region analysis suggests that some are particularly difficult to evaluate, with varying levels of reliability. Potential impairment of the supraganglionic region must be examined carefully, particularly with respect to the decision whether or not to perform mechanical thrombectomy.

**Keywords** Acute stroke therapy · Brain · Endovascular treatment · Interrater reliability · Ischemic stroke · Krippendorff's  $\alpha$

## Introduction

Alberta Stroke Program Early CT Score (ASPECTS) is a diagnostic tool for the assessment of the extent of early ischemic lesions in anterior circulation stroke and is used to select patients for mechanical thrombectomy (MT) [1, 2]. As a semiquantitative grading system, ASPECTS estimates the extent of early ischemic changes on non-contrast computed tomography (CT) for ten regions within the middle cerebral artery territory and has been demonstrated to be predictive of clinical outcome [3].

Clinical guidelines on the management of acute ischemic stroke (AIS) patients with large vessel occlusion recognize a baseline ASPECTS  $\geq 6$  as a key selection criterion for MT [4, 5]; however, different ASPECTS thresholds have been employed to selectively enrol patients in various clinical tri-

The authors Noel van Horn and Helge Kniep contributed equally to the manuscript.

✉ Noel van Horn  
no.vanhorn@uke.de

<sup>1</sup> Department of Diagnostic and Interventional Neuroradiology, University Medical Center Hamburg Eppendorf, Martinistraße 52, 20246 Hamburg, Germany

<sup>2</sup> Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>3</sup> Department of Neuroradiology, University of Heidelberg, Heidelberg, Germany

<sup>4</sup> Forschungszentrum Jülich, Institute of Neurosciences and Medicine, Jülich, Germany

als, thereby complicating stroke assessment and subsequent therapeutic considerations in the clinical routine [6–11]. Of these trials two (MR CLEAN and THRACE) independently demonstrated the benefit of MT based on non-contrast computed tomography (NCCT), the modality most commonly used for initial evaluation [4, 10, 11]. Although ASPECTS has also been shown to be useful in identifying patients eligible for MT who present in the late time window, selection of this subgroup is typically based on advanced imaging criteria [5, 12–15].

The NCCT is an essential diagnostic tool in the management of AIS patients due to its speed and broad availability as well as its accuracy in excluding intracranial hemorrhage; however, the level of interrater agreement in the interpretation of early ischemic lesions on NCCT varies in the literature [16, 17] and is limited, reported by a systematic review analyzing 15 studies on interrater reliability ranging from fair to substantial agreement according to the Fleiss classification [18]. Despite the importance of this diagnostic tool in the selection of patients for MT, no study examining the interrater reliability among a large number of readers has been performed. We hypothesized that the interrater reliability decreases with an increasing number of readers, regardless of expertise or experience.

With the efficacy and safety of thrombectomy in stroke with extended lesion and extended time window (TENSION) trial [19], we sought to evaluate the interrater reliability of a large number of readers in the assessment of total baseline ASPECTS, as well as the individual ASPECTS regions, in patients with severe anterior circulation stroke (ASPECTS 3–5). Our goal was to identify how such potential rater-dependent differences affect the treatment decision-making process.

## Methods

This retrospective study was approved by the leading ethics committee of the local ethics committee. All patient data, study protocols and procedures were conducted in accordance with the Declaration of Helsinki and informed consent was obtained. All data used for training and validation in this study are available from the corresponding author upon reasonable request.

### TENSION—Study Design

The TENSION trial is a European investigator-initiated, prospective, open label, blinded endpoint (PROBE) [19], randomized, controlled, two-arm trial to examine the safety and effectiveness of MT compared to best medical care alone in the treatment of AIS patients with extensive early ischemic lesions (defined by an ASPECTS of 3–5) in an ex-

tended time window (up to 12 h from onset or last seen well) [20]. Only patients presenting with AIS due to focal occlusion of the M1 segment of the MCA, and/or the intracranial segment of the distal internal carotid artery, determined by either magnetic resonance angiography or computed tomography angiography were included. Subjects who met the inclusion criteria were randomized in a 1:1 ratio to either (1) best medical care alone or (2) best medical care with MT. The primary endpoint was functional outcome after 90 days post-stroke measured by modified Rankin scale (mRS90).

### TENSION—ASPECTS Image Reading Academy

The TENSION ASPECTS Image Reading Academy, a web-based training tool, was developed to qualify investigators for study-specific image assessment and ASPECTS reading (Eppdata, Hamburg, Germany). To that end, 56 cases from our stroke database were identified with a preliminary ASPECTS of 0–8, and evaluated by 4 highly experienced neuroradiologists (>20 years of experience each, GOLD rater) in accordance with the ASPECTS grading system. The image assessment tool allowed the evaluator to window, zoom, and pan individually at a slice thickness of 4 mm. For 30 of the cases at least 3 of the 4 GOLD raters were in agreement for each individual ASPECTS region. Of these, 10 were implemented in the online training tool as training module cases, while the remaining 20 comprised the rating module. Only those who successfully completed the training modules (at least 80% correct answers) were granted participation in rating module and subsequently included for further analysis.

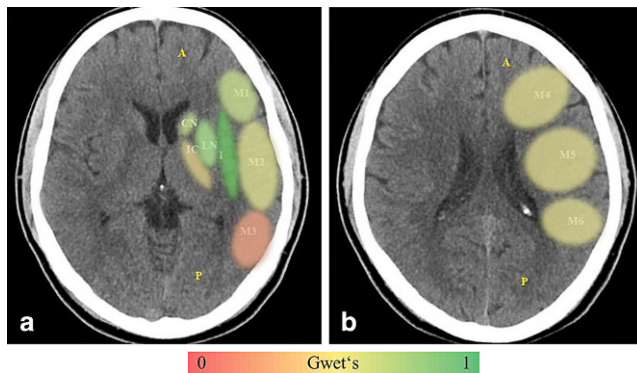
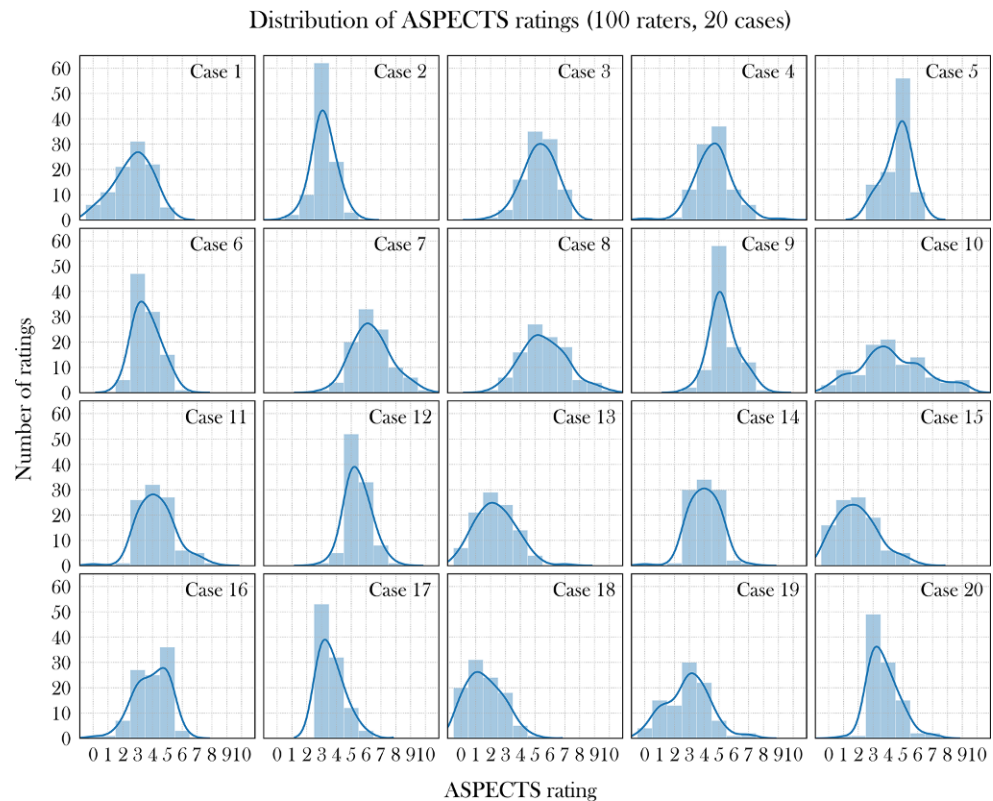
### ASPECTS Scoring

The first 100 TENSION investigators who successfully concluded the above defined ASPECTS training and testing were analyzed. All of them were experienced radiologists whose everyday clinical practice is to select AIS patients for MT. Early ischemic change was defined as the presence of hypodensity and/or loss of grey-white differentiation. Raters were not blinded within the test module to clinical details, affected site, time since onset, age and NIHSS. Raters provided a score for each ASPECTS region in a binary manner with 1, scoring for an affected region, and 0 for an unaffected one. Total ASPECTS score of 0–10 was simultaneously calculated based on these per region ratings and displayed to the raters.

### Statistical Analysis

Concordance of ASPECTS was evaluated using percent and interrater agreement (Krippendorff's  $\alpha$  [21]) at two differ-

**Fig. 1** Frequency distribution of Alberta Stroke Program Early CT Score (ASPECTS) ratings for each of the 20 cases



**Fig. 2** Illustrates ASPECTS per region agreement on the level of the basal ganglia (a) and immediately above (b) calculated using Gwet's AC1 on NCCT with low (red), moderate (yellow) and high (green) agreement. *I* insula, *IC* internal capsule, *C* caput, *L* lentiform, *M1–M6* media cerebral artery territory M1–M6, ASPECTS Alberta Stroke Program Early CT Score, NCCT non-contrast computed tomography

ent levels of tolerance allowance: a) zero tolerance (because deviations of one scoring point could lead to different therapeutic decisions) and b) tolerance ranges according to the TENSION inclusion criteria (ASPECTS 0–2 vs. 3–5 vs. 6–10). Concordance for binary per region ratings were calculated for percent agreement and interrater agreement using Gwet's AC1 [22]. These metrics have been found to be unbiased in cases of high prevalence [21] (e.g., infarct demarcation in the insular region is present in almost all AIS

patient scans). Values of interrater agreement are demonstrated with 95% confidence intervals (bootstrap using 1000 replicates). Furthermore, percentage shares of concordant ratings were calculated per region and per case (all raters with concordant ratings: 100%; 50/50 scoring: 50%). All data analyses were conducted using R 3.6.2 (R Core Team, 2019, Vienna, Austria) with irrCAC 1.0 and irr 0.84.1 packages [23].

## Results

Total ASPECTS rating frequency distributions per case were visualized using histogram plots (Fig. 1). The percent agreement for total ASPECTS ratings at zero tolerance allowance was 28%, with an interrater agreement of 0.13 (Krippendorff's  $\alpha$ ; 95% CI: 0.09–0.16;  $p$ -value < 0.01), suggesting low agreement between raters. When calculated using the tolerance allowance according to TENSION inclusion criteria ranges (ASPECTS 0–2 vs. 3–5 vs. 6–10), percent agreement was 66%, with an interrater agreement of 0.32 (Krippendorff's  $\alpha$ ; 95% CI: 0.21–0.44;  $p$ -value < 0.01).

The percent and interrater agreement of the individual ASPECTS regions was highest in the insular cortex, with a percent agreement of 96% and an interrater agreement of 0.96 (Gwet's AC1; 95% CI: 0.94–0.97;  $p$ -value < 0.01) (Fig. 2). The lowest per region agreement was found in the M3 region, with a percent agreement of 68% and an inter-

**Table 1** ASPECTS interrater agreement of 100 radiologists

	Percent agreement (%)	Interrater reliability	95% CI	<i>p</i> -value
<i>Total ASPECTS</i>				
Score number	28	0.13	0.09–0.16	<0.01
Inclusion criteria	66	0.32	0.21–0.44	<0.01
<i>Rating per region</i>				
ASPECTS C	83	0.67	0.48–0.86	<0.01
ASPECTS I	96	0.96	0.94–0.97	<0.01
ASPECTS IC	73	0.58	0.37–0.79	<0.01
ASPECTS L	86	0.73	0.59–0.86	<0.01
ASPECTS M1	84	0.76	0.59–0.93	<0.01
ASPECTS M2	80	0.73	0.55–0.90	<0.01
ASPECTS M3	68	0.39	0.17–0.61	<0.01
ASPECTS M4	80	0.66	0.49–0.83	<0.01
ASPECTS M5	78	0.62	0.42–0.82	<0.01
ASPECTS M6	77	0.55	0.35–0.74	<0.01

Table 1 shows percent and interrater reliability with *p*-values. Total ASPECTS was evaluated for score numbers (zero tolerance) and agreement according to TENSION inclusion criteria cut-offs (ASPECTS 0–2; 3–5; 6–10). Interrater reliability for total ASPECTS was calculated using Krippendorff's  $\alpha$ ; binary rating per region based on Gwet's AC1

*I* insula, *IC* internal capsule, *C* caput, *L* lentiform, *M1–M6* media cerebral artery territory M1–M6

rater agreement of 0.39 (Gwet's AC1; 95% CI: 0.17–0.61; *p*-value<0.01) (Table 1). The total and individual region shares of concordant ratings are demonstrated in Fig. 3. In all cases, the highest level of concordant ratings was observed for the insular cortex, with a mean share of 98%, while the lowest was seen in the M3 region (76%).

Baseline demographic and clinical information for the participating cases showed a median age of 65.5 years (95% CI: 52.8–76.3 years) and a median NIHSS of 16 (95% CI: 14–22), with a mean large vessel occlusion distribution located 85% in middle cerebral artery, 25% in internal carotid artery, and 5% in the anterior cerebral artery.

## Discussion

Interrater agreement differs remarkably for both total ASPECTS and the individual ASPECTS regions [24]. In our dataset, a wide frequency distribution for total ASPECTS evaluation was found. When readings of the individual regions were compared, the insular cortex showed the highest agreement, whereas the M3 region showed the lowest. This is an important finding, as ASPECTS is a well-established and frequently used diagnostic tool for the selection of patients for MT [4, 5]. The reasons for disagreement among readers, as well as the limitations of this selection tool, should therefore be clarified.

Varying ASPECTS thresholds are used to identify suitable patients for MT [6–11]. The 2019 updated “Guidelines for the Early Management of Patients with Acute Ischemic

Stroke” recommend an ASPECTS threshold of 6 [4]; patients with a lower score are unlikely to experience good outcome [5]. In contrast, a recently conducted meta-analysis pooling 17 studies reported that these patients may benefit from successful reperfusion, with a higher probability of functional independence after 90 days and no greater risk of symptomatic intracranial hemorrhage [25]. Equal consideration should be given to both image evaluation and the patient's clinical condition in the decision for best possible treatment; however, especially in patients with low ASPECTS (i.e., <6), there seems to be some controversy over which imaging and clinical factors most affect clinical outcome. Indeed, most centers make therapeutic strategy decisions on a case by case basis.

Since first introduced in 2000 by Barber et al., varying ASPECTS thresholds have been implemented [1]. In the first subsequent study conducted by the Calgary group, a high level of disagreement among six physicians (three non-neuroradiologists and three neuroradiologists) was reported [26]. A further study by Gupta et al. [16] reported moderate interrater agreement with a  $\kappa=0.53$  (Cohen's  $\kappa$ ) between 2 experienced neuroradiologists for 155 patients, while Finlayson et al. [27] demonstrated very good agreement among 4 readers (2 neuroradiologists and 2 neurologists) for 181 patients using Cronbach's  $\alpha$  accounting for internal consistency, with an  $\alpha=0.83$ . Yet another recently conducted study by Nicholson et al. [24] found that overall agreement was relatively good, with a  $\kappa=0.66$  (Cohen's  $\kappa$ ) at an ASPECTS threshold of 5.

To date, most studies examining interrater agreement have been heterogeneous and have never examined more



## Agreement per Region %

## Total Score Treatment Agreement %

Case number	ASPECTS C	ASPECTS I	ASPECTS IC	ASPECTS L	ASPECTS M1	ASPECTS M2	ASPECTS M3	ASPECTS M4	ASPECTS M5	ASPECTS M6	ASPECTS mean	Standard dev	Enroll y/n	MT outside study	No MT
2	93%	97%	67%	98%	100%	99%	96%	87%	74%	99%	3.14	0.72	86%	96%	90%
29	100%	100%	100%	99%	99%	75%	95%	95%	70%	90%	5.45	0.80	57%	57%	100%
44	100%	98%	100%	100%	85%	100%	68%	95%	99%	89%	3.66	0.80	96%	96%	100%
9	99%	98%	97%	97%	100%	100%	69%	88%	100%	77%	3.61	0.85	96%	100%	96%
7	100%	99%	100%	96%	100%	92%	65%	99%	93%	86%	4.64	0.86	88%	88%	100%
15	99%	98%	100%	99%	97%	92%	93%	94%	70%	92%	5.32	0.90	63%	68%	95%
35	99%	97%	95%	86%	97%	100%	57%	94%	96%	62%	4.01	0.99	95%	97%	98%
53	99%	98%	99%	98%	97%	100%	73%	81%	95%	78%	3.74	1.00	91%	95%	96%
5	100%	98%	72%	97%	59%	60%	95%	90%	91%	99%	5.28	1.07	55%	55%	100%
40	83%	96%	90%	89%	95%	80%	76%	99%	100%	75%	3.94	1.14	78%	96%	82%
47	62%	99%	63%	95%	99%	99%	74%	84%	97%	95%	1.58	1.19	76%	97%	73%
20	80%	99%	89%	79%	90%	62%	95%	100%	55%	98%	4.27	1.24	86%	87%	99%
6	99%	99%	53%	96%	63%	55%	98%	52%	52%	99%	4.64	1.24	80%	82%	98%
1	100%	97%	57%	100%	99%	98%	52%	62%	69%	53%	2.71	1.27	61%	95%	66%
12	54%	95%	74%	87%	91%	70%	85%	92%	95%	95%	6.35	1.29	75%	73%	98%
34	84%	99%	62%	93%	99%	98%	70%	87%	96%	70%	2.34	1.34	57%	97%	54%
36	81%	95%	74%	62%	85%	98%	72%	98%	99%	91%	1.92	1.39	71%	95%	66%
50	96%	99%	56%	96%	75%	93%	53%	75%	81%	66%	2.92	1.46	63%	95%	68%
13	68%	94%	74%	86%	88%	67%	56%	93%	88%	91%	5.62	1.54	52%	51%	99%
16	52%	99%	74%	65%	51%	57%	75%	61%	66%	70%	4.29	2.24	51%	66%	85%
Mean	87%	98%	80%	91%	89%	85%	76%	86%	84%	84%			74%	84%	88%

**Fig. 3** Demonstrates rating concordance among all raters regarding ASPECTS rating of all 20 cases for each ASPECTS region, mean concordance for all cases and concordance for enrolment decision for mechanical thrombectomy. Mean ASPECTS is illustrated according to ASPECTS with standard deviation (standard dev) sorted by lowest to highest standard dev (*blue*, high concordance; *red*, low concordance). *I* insula, *IC* internal capsule, *C* caput, *L* lentiform, *M1–M6* media cerebral artery territory M1–M6, *ASPECTS* Alberta Stroke Program Early CT Score

than six readers, making comparability to our study difficult [16, 27–29]. In order to analyze 100 experienced readers for total ASPECTS, Krippendorff's  $\alpha$  was used as a measure for interrater agreement [30] and showed only slight agreement. The frequency distribution of the total ASPECTS ratings also varied among the individual cases. Widest distribution was observed for case number 10, one of the oldest (78 years) and most severely affected (NIHSS of 21) cases with two occluded vessels (middle cerebral artery and internal carotid artery), which in this constellation may have had a negative impact on image evaluation.

In line with our findings, Nicholson et al. [24] reported the highest level of agreement for the insular cortex, with a  $\kappa=0.56$  (Cohen's  $\kappa$ ), the lowest being observed in the M3 ( $\kappa=0.34$ ) and internal capsule ( $\kappa=0.44$ ) regions. Similarly, the four observers in the study conducted by Finlayson et al. [27] had the highest level of concordance for the lentiform nucleus, with a Cronbach's  $\alpha$  of 0.82, and the lowest for the internal capsule ( $\alpha=0.41$ ).

Both anatomical and methodological factors could explain why the agreement among different ASPECTS territories is so varied. In general, the basal ganglia are well-defined and rather easy to distinguish due to the grey-white contrast provided by the attenuation of the surrounding structures on NCCT. The high level of agreement for the insular cortex might be due to the contrast provided laterally by the liquor-isodense Sylvian fissure and bordering brain parenchyma of the operculum and medially by the relatively hypoattenuated external capsule. Furthermore, vulnerability to ischemic changes due to hypoperfusion varies between different locations of the brain, with the insular cortex, precentral gyrus and basal ganglia being the most sensitive [31]. On the contrary, the internal capsule is by nature distinguishable due its basic hypoattenuation compared to surrounding structures in NCCT; however, further hypoattenuation in the case of infarction could be overlooked. Moreover, if the surrounding caudate and lentiform nuclei also undergo loss of attenuation, the internal capsule

may become more difficult to define, particularly if they are affected at different rates, despite a shared blood supply.

In the subganglionic and supraganglionic nuclei territories (M1–M6), differentiation of the cortex is a main criterion for early infarct detection. In order to provide a realistic diagnostic scenario, the training tool used in the study enables all raters to individually adjust window width and center for every single NCCT evaluated. Due to their location bordering the skull, beam hardening artifacts can make differentiation of these locations more difficult. Another reason the M3 region may have such a low level of agreement could be due to its relatively infrequent occurrence [32]. These territories can be of high functional eloquence. Excluding patients from MT based on involvement of these regions should therefore be carefully considered, particularly in the light of the limited interrater reliability of these locations.

We believe that this study makes an important contribution to the field of AIS therapy because it shows that despite its widespread application and acceptance, the ASPECTS grading tool is not without flaws. This is underlined by the high degree of interrater variability in both total ASPECTS scoring and in the evaluation of the individual ASPECTS regions among a large number of experienced neuroradiologists. At the zero tolerance level, the total ASPECTS scoring interrater agreement reliability was rather low. At a tolerance level according to the TENSION inclusion criteria (0–2, 3–5, 6–10), percent agreement (66%) seems sufficient for clinical routine and for study enrolment but did not meet the criteria from a strictly statistical point of view: however, responses for patient enrolment are not evenly balanced with extreme prevalence in most cases. This imbalance leads to a very low expected disagreement (which is influenced by the prevalence) and results in low reliability statistics (as the observed disagreement is higher). In both settings of a clinical trial and the everyday routine, a one-point scoring deviation may determine whether a patient suffering from AIS undergoes MT or best medical treatment alone. Moreover, with the knowledge gained from an increasing number of studies, more and more limitations to ASPECTS are being unmasked. For example, the score lacks clearly defined territory borders, unequally weights different territories, and is limited to the anterior circulation [33]. Perhaps different levels of sub-scoring for each individual region could further improve ASPECTS ratings.

When ASPECTS is applied to advanced imaging, interrater reliability improves. For example, ASPECTS evaluated on computed tomography perfusion [27] and diffusion weighted magnetic resonance imaging [34], which are based on physiological and hemodynamic changes, is more sensitive compared to NCCT-ASPECTS for the assessment of imaging abnormalities. Furthermore, automated software and deep learning algorithms are currently being tested on

computed tomography perfusion-based imaging protocols to maintain or even improve interrater reliability, including that among non-neuroradiologists [35]. Preliminary results show that automated computed tomography perfusion of AIS patients in the emergency setting tends to overestimate the infarct core, as confirmed by follow-up CT after 24h [36].

One major limitation of our study is the patient population. The observed rating results are probably poorer than in real life because of the unusually low ASPECTS population average compared to typical stroke patients. This likely leads to an inherently higher probability of disagreement. This was in part corrected for by the study inclusion selection criteria, which were based on a relatively high agreement among the four GOLD raters. Moreover, the ASPECTS range of 3–5 and the need for decision making to enrol the patient into the TENSION study could have influenced the observers. Lastly, readings were typically done on regular computer monitors, with the software running on different web browsers and operating systems, likely under suboptimal environmental/lighting conditions.

## Conclusion

Interrater agreement reliability for total ASPECTS and study enrollment was relatively low but seems sufficient for practical application. Per region analysis suggests that some ASPECTS regions are particularly difficult to evaluate, with varying levels of interrater reliability. Based on our observations, a decision to exclude patients from MT based on involvement of the supraganglionic regions should be considered with particular care. We consider these observations to be practically relevant as they are based on the findings of a large number of raters, all with considerable job experience and who are all highly motivated to achieve good rating results and, more importantly, the best patient outcome.

**Funding** European Union Horizon 2020 Research and Innovation Programme (Grant Agreement No.: 754640)

**Author Contribution** NVH: study design. Acquisition of data. Image processing. Data analysis. Statistical analysis. Drafting the manuscript and revising it critically.

HK: Study design. Acquisition of data. Image processing. Data analysis. Statistical analysis. Drafting the manuscript and revising it critically.

GB: Data analysis. Statistical analysis. Acquisition of Data. Drafting the manuscript and revising it critically.

LM: Acquisition of data. Drafting the manuscript and revising it critically.

FF: Acquisition of data. Drafting the manuscript and revising it critically.

MB: Data analysis. Statistical analysis. Drafting the manuscript and revising it critically.

JG: Study design. Acquisition of data. Image analysis. Data analysis. Drafting the manuscript and revising it critically.

GT: Study design. Data analysis. Drafting the manuscript and revising it critically.

MB: Acquisition of data. Data analysis. Drafting the manuscript and revising it critically.

SB: Acquisition of data. Data analysis. Drafting the manuscript and revising it critically.

JARP: Data analysis. Image analysis. Drafting the manuscript and revising it critically.

PRD: Data analysis. Image analysis. Drafting the manuscript and revising it critically.

JF: Study design. Acquisition of data. Image analysis. Data analysis. Drafting the manuscript and revising it critically.

UH: Study design. Acquisition of data. Image analysis. Data analysis. Drafting the manuscript and revising it critically.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Compliance with ethical guidelines

**Conflict of interest** G. Thomalla: consulting fees from Acandis, grant support and lecture fees from Bayer, lecture fees from Boehringer Ingelheim, Bristol-Myers Squibb/Pfizer, and Daiichi Sankyo, and consulting fees and lecture fees from Stryker. Grants from Bundesministerium für Wirtschaft und Energie (BMWi), Deutsche Forschungsgemeinschaft (DFG), European Union (EU), German Innovation Fund, Corona Foundation. J.A.R. Pfaff: personal fees from Stryker, outside the submitted work. J. Fiehler: consultant for Acandis, Boehringer Ingelheim, Codman, Microvention, Sequent, Stryker. Speaker for Bayer Healthcare, Bracco, Covidien/ev3, Penumbra, Philips, Siemens. Grants from Bundesministeriums für Wirtschaft und Energie (BMWi), Bundesministerium für Bildung und Forschung (BMBF), Deutsche Forschungsgemeinschaft (DFG), European Union (EU), Covidien, Stryker (THRILL study), Microvention (ERASER study), Philips. N. van Horn, H. Kniep, G. Broocks, L. Meyer, F. Flottmann, M. Bechstein, J. Götz, M. Bendszus, S. Bonekamp, P.R. Dellani, and U. Hanning declare that they have no competing interests.

**Ethical standards** For this article no studies with human participants or animals were performed by any of the authors. All studies mentioned were in accordance with the ethical standards indicated in each case. This retrospective study was approved by the leading ethics committee of the medical faculty of the University of Heidelberg (S-248/2016).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Barber PA, Demchuk AM, Zhang J, Buchan AM. 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*. 2000;355:1670–4.
- Yoo AJ, Zaidat OO, Chaudhry ZA, Berkhemer OA, González RG, Goyal M, Demchuk AM, Menon BK, Muallem E, Ueda D, Buell H, Sit SP, Bose A; Penumbra Pivotal and Penumbra Imaging Collaborative Study (PICS) Investigators. Impact of pretreatment noncontrast CT Alberta Stroke Program Early CT Score on clinical outcome after intra-arterial stroke therapy. *Stroke*. 2014;45:746–51.
- Goyal M, Menon BK, Coutts SB, Hill MD, Demchuk AM; Penumbra Pivotal Stroke Trial Investigators, Calgary Stroke Program, and the Seaman MR Research Center. Effect of baseline CT scan appearance and time to recanalization on clinical outcomes in endovascular thrombectomy of acute ischemic strokes. *Stroke*. 2011;42:93–7.
- Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, Biller J, Brown M, Demerschalk BM, Hoh B, Jauch EC, Kidwell CS, Leslie-Mazwi TM, Ovbiagele B, Scott PA, Sheth KN, Southerland AM, Summers DV, Tirschwell DL. Guidelines for the Early Management of Patients With Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*. 2019;50:e344–418.
- Turc G, Bhogal P, Fischer U, Khatri P, Lobotesis K, Mazighi M, Schellinger PD, Toni D, de Vries J, White P, Fiehler J. European Stroke Organisation (ESO)—European Society for Minimally Invasive Neurological Therapy (ESMINT) Guidelines on Mechanical Thrombectomy in Acute Ischemic Stroke. *J Neurointerv Surg*. 2019; <https://doi.org/10.1136/neurintsurg-2018-014569>.
- Jovin TG, Chamorro A, Cobo E, de Miquel MA, Molina CA, Rovira A, San Román L, Serena J, Abilleira S, Ribó M, Millán M, Urra X, Cardona P, López-Cancio E, Tomasello A, Castaño C, Blasco J, Aja L, Dorado L, Quesada H, Rubiera M, Hernandez-Pérez M, Goyal M, Demchuk AM, von Kummer R, Gallofré M, Dávalos A; REVASCAT Trial Investigators. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med*. 2015;372:2296–306.
- Saver JL, Goyal M, Bonafe A, Diener HC, Levy EI, Pereira VM, Albers GW, Cognard C, Cohen DJ, Hacke W, Jansen O, Jovin TG, Mattle HP, Nogueira RG, Siddiqui AH, Yavagal DR, Baxter BW, Devlin TG, Lopes DK, Reddy VK, du Mesnil de Rochemont R, Singer OC, Jahan R; SWIFT PRIME Investigators. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med*. 2015;372:2285–95.
- Campbell BC, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, Yan B, Dowling RJ, Parsons MW, Oxley TJ, Wu TY, Brooks M, Simpson MA, Miteff F, Levi CR, Krause M, Harrington TJ, Faulder KC, Steinfort BS, Priglinger M, Ang T, Scroop R, Barber PA, McGuinness B, Wijeratne T, Phan TG, Chong W, Chandra RV, Bladin CF, Badve M, Rice H, de Villiers L, Ma H, Desmond PM, Donnan GA, Davis SM; EXTEND-IA Investigators. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*. 2015;372:1009–18.
- Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, Roy D, Jovin TG, Willinsky RA, Sapkota BL, Dowlathshahi D, Frei DF, Kamal NR, Montanera WJ, Poppe AY, Ryckborst KJ, Silver FL, Shuaib A, Tampieri D, Williams D, Bang OY, Baxter BW, Burns PA, Choe H, Heo JH, Holmstedt CA, Jankowitz B, Kelly M, Linares G, Mandzia JL, Shankar J, Sohn SI, Swartz RH, Barber PA, Coutts SB, Smith EE, Morrish WF, Weill A, Subramaniam S, Mitha AP, Wong JH, Lowerison MW, Sajobi TT, Hill MD; ESCAPE Trial Investigators. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. 2015;372:1019–30.



10. Bracard S, Ducrocq X, Mas JL, Soudant M, Oppenheim C, Moulin T, Guillemin F; THRACE investigators. Mechanical thrombectomy after intravenous alteplase versus alteplase alone after stroke (THRACE): a randomised controlled trial. *Lancet Neurol.* 2016;15:1138–47.
11. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, Schonewille WJ, Vos JA, Nederkoorn PJ, Wermer MJ, van Walderveen MA, Staals J, Hofmeijer J, van Oostayen JA, Lycklama à Nijeholt GJ, Boiten J, Brouwer PA, Emmer BJ, de Bruijn SF, van Dijk LC, Kappelle LJ, Lo RH, van Dijk EJ, de Vries J, de Kort PL, van Rooij WJ, van den Berg JS, van Hasselt BA, Aerden LA, Dallinga RJ, Visser MC, Bot JC, Vroomen PC, Eshghi O, Schreuder TH, Heijboer RJ, Keizer K, Tielbeek AV, den Hertog HM, Gerrits DG, van den Berg-Vos RM, Karas GB, Steyerberg EW, Flach HZ, Marquering HA, Sprengers ME, Jenniskens SF, Beenen LF, van den Berg R, Koudstaal PJ, van Zwam WH, Roos YB, van der Lugt A, van Oostenbrugge RJ, Majoie CB, Dippel DW; MR CLEAN Investigators. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* 2015;372:11–20. Erratum in: *N Engl J Med.* 2015;372:394.
12. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, Yavagal DR, Ribo M, Cognard C, Hanel RA, Sila CA, Hassan AE, Millan M, Levy EI, Mitchell P, Chen M, English JD, Shah QA, Silver FL, Pereira VM, Mehta BP, Baxter BW, Abraham MG, Cardona P, Veznedaroglu E, Hellinger FR, Feng L, Kirmani JF, Lopes DK, Jankowitz BT, Frankel MR, Costalat V, Vora NA, Yoo AJ, Malik AM, Furlan AJ, Rubiera M, Agha-Abraham A, Olivot JM, Tekle WG, Shields R, Graves T, Lewis RJ, Smith WS, Liebeskind DS, Saver JL, Jovin TG; DAWN Trial Investigators. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *N Engl J Med.* 2018;378:11–21.
13. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, McTaggart RA, Torbey MT, Kim-Tenser M, Leslie-Mazwi T, Sarraj A, Kasner SE, Ansari SA, Yeatts SD, Hamilton S, Mlynash M, Heit JJ, Zaharchuk G, Kim S, Carrozzella J, Palesch YY, Demchuk AM, Bammner R, Lavori PW, Broderick JP, Lansberg MG; DEFUSE 3 Investigators. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *N Engl J Med.* 2018;378:708–18.
14. Meyers PM, Schumacher HC, Higashida RT, Barnwell SL, Creager MA, Gupta R, McDougall CG, Pandey DK, Sacks D, Wechsler LR; American Heart Association. Indications for the performance of intracranial endovascular neurointerventional procedures: a scientific statement from the American Heart Association Council on Cardiovascular Radiology and Intervention, Stroke Council, Council on Cardiovascular Surgery and Anesthesia, Interdisciplinary Council on Peripheral Vascular Disease, and Interdisciplinary Council on Quality of Care and Outcomes Research. *Circulation.* 2009;119:2235–49.
15. Nagel S, Herweh C, Pfaff JAR, Schieber S, Schöenberger S, Möhlenbruch MA, Bendszus M, Ringleb PA. Simplified selection criteria for patients with longer or unknown time to treatment predict good outcome after mechanical thrombectomy. *J Neurointerv Surg.* 2019;11:559–62.
16. Gupta AC, Schaefer PW, Chaudhry ZA, Leslie-Mazwi TM, Chandra RV, González RG, Hirsch JA, Yoo AJ. Interobserver reliability of baseline noncontrast CT Alberta Stroke Program Early CT Score for intra-arterial stroke treatment selection. *AJNR Am J Neuroradiol.* 2012;33:1046–9.
17. Coutts SB, Demchuk AM, Barber PA, Hu WY, Simon JE, Buchan AM, Hill MD; VISION Study Group. Interobserver variation of ASPECTS in real time. *Stroke.* 2004;35:e103–5.
18. Fleiss J, Levin B, Paik MC. Statistical methods for rates and proportions. Hoboken: John Wiley & Sons, Inc; 2003.
19. Bendszus M, Bonekamp S, Berge E, Boutitie F, Brouwer P, Gizewski E, Krajina A, Pierot L, Randall G, Simonsen CZ, Zelenák K, Fiehler J, Thomalla G. A randomized controlled trial to test efficacy and safety of thrombectomy in stroke with extended lesion and extended time window. *Int J Stroke.* 2019;14:87–93.
20. Bendszus M, Fiehler J, Thomalla G. New Interventional Stroke Trials. *Clin Neuroradiol.* 2019;29:1.
21. Feng GC. Mistakes and how to avoid mistakes in using intercoder reliability indices. *Methodology.* 2015;11:13–22.
22. Gwet K. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. 2010.
23. R Core Team. A language and environment for statistical computing. 2018. <https://www.R-project.org/>. Accessed 13 Nov 2019.
24. Nicholson P, Hilditch CA, Neuhaus A, Seyedsaadat SM, Benson JC, Mark I, Tsang COA, Schaafsma J, Kallmes DF, Krings T, Brinjikji W. Per-region interobserver agreement of Alberta Stroke Program Early CT Scores (ASPECTS). *J Neurointerv Surg.* 2020;12:1069–71.
25. Cagnazzo F, Derraz I, Dargazanli C, Lefevre PH, Gascou G, Riquelme C, Bonafe A, Costalat V. Mechanical thrombectomy in patients with acute ischemic stroke and ASPECTS  $\leq 6$ : a meta-analysis. *J Neurointerv Surg.* 2020;12:350–5.
26. Pexman JH, Barber PA, Hill MD, Sevick RJ, Demchuk AM, Hudon ME, Hu WY, Buchan AM. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. *AJNR Am J Neuroradiol.* 2001;22:1534–42.
27. Finlayson O, John V, Yeung R, Dowlatabadi D, Howard P, Zhang L, Swartz R, Aviv RI. Interobserver agreement of ASPECT score distribution for noncontrast CT, CT angiography, and CT perfusion in acute stroke. *Stroke.* 2013;44:234–6.
28. Ghandehari K, Rezvani MR, Shakeri MT, Mohammadifard M, Ehsanbakhsh A, Mohammadifard M, Mirgholami A, Boostani R, Ghandehari K, Izadi-Mood Z. Inter-rater reliability of modified Alberta Stroke program early computerized tomography score in patients with brain infarction. *J Res Med Sci.* 2011;16:1326–31.
29. Naylor J, Churilov L, Rane N, Chen Z, Campbell BCV, Yan B. Reliability and Utility of the Alberta Stroke Program Early Computed Tomography Score in Hyperacute Stroke. *J Stroke Cerebrovasc Dis.* 2017;26:2547–52.
30. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data—which coefficients and confidence intervals are appropriate? *BMC Med Res Methodol.* 2016;16:93.
31. Payabvash S, Souza LC, Wang Y, Schaefer PW, Furie KL, Halpern EF, Gonzalez RG, Lev MH. Regional ischemic vulnerability of the brain to hypoperfusion: the need for location specific computed tomography perfusion thresholds in acute stroke patients. *Stroke.* 2011;42:1255–60.
32. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;37:360–3.
33. Bristow MS, Simon JE, Brown RA, Eliasziw M, Hill MD, Coutts SB, Frayne R, Demchuk AM, Mitchell JR. MR perfusion and diffusion in acute ischemic stroke: human gray and white matter have different thresholds for infarction. *J Cereb Blood Flow Metab.* 2005;25:1280–7.
34. Mitomi M, Kimura K, Aoki J, Iguchi Y. Comparison of CT and DWI findings in ischemic stroke patients within 3 hours of onset. *J Stroke Cerebrovasc Dis.* 2014;23:37–42.
35. Sundaram VK, Goldstein J, Wheelwright D, Aggarwal A, Pawha PS, Doshi A, Fifi JT, Leacy R, Mocco J, Puig J, Nael K. Automated ASPECTS in Acute Ischemic Stroke: A Comparative Analysis with CT Perfusion. *AJNR Am J Neuroradiol.* 2019;40:2033–8.
36. Tsang ACO, Lenck S, Hilditch C, Nicholson P, Brinjikji W, Krings T, Pereira VM, Silver FL, Schaafsma JD. Automated CT Perfusion Imaging Versus Non-contrast CT for Ischemic Core Assessment in Large Vessel Occlusion. *Clin Neuroradiol.* 2020;30:109–14.