Original research

Automated perfusion imaging did not improve reliability of ASPECTS or collateral score in late-window large vessel occlusion stroke

Anass Benomar ,[1] Robert Fahed ,[2] Abdelsimar Tan Omar ,[3,4] Christine Hawkes ,[5] Eef J Hendriks ,[6] Clare A Enriquez ,[5] Muneer Eesa ,[7] Grant Stotts ,[2] Hubert Lee ,[8] Shashank Nagendra ,[9] Alexandre Y Poppe ,[10] Célina Ducroux ,[2] Timothy R Lim ,[1] Karl Narvacan ,[1] Michael A Rizzuto ,[11] Afra Alfalahi ,[1] Hidehisa Nishi ,[12] Pragyan Sarma ,[1] Ze'ev Itsekson Hayosh ,[6] Katrina Ignacio ,[9] William Boisseau ,[13] Eduardo Pimenta Ribeiro Pontes Almeida ,[14] Mohammed A Almekhlafi ,[15] Geneviève Milot ,[16] Aviraj Deshmukh ,[17] Kislay Kishore ,[3] Donatella Tampieri ,[18] Jefffrey Z Wang ,[2] Abhilekh Srivastava ,[19] Daniel Roy ,[20] Federico Carpani ,[21] Nima Kashani ,[22] Claudia Candale-Radu ,[23] Nishita Singh ,[23] Maria Bres-Bullrich ,[24] Robert Sarmiento ,[25] Ryan T Muir ,[5] Carmen Parra-Fariñas ,[26] Stephanie D Reiter ,[5] Yan Deschaintre ,[10] Ravinder-Jeet Singh ,[17] Vivek Bodani ,[6] Aristeidis H Katsanos ,[19] Ronit Agid ,[6] Atif Zafar ,[27] Vitor M Pereira ,[1,28] Julian Spears ,[1,28] Thomas R Marotta ,[1] Pascal Djiadeu ,[29] Sunjay Sharma ,[29] Forough Farrokhyar ,[29] Jose Danilo Bengzon Diestro [1,3,27]

ABSTRACT

Background The Alberta Stroke Program Early CT Score (ASPECTS) and CT angiography (CTA) Collateral Score (CS) are routinely used to estimate ischemic core burden and collateral status in acute ischemic stroke due to large vessel occlusion, but suffer from limited reliability. Whether automated CT perfusion (CTP) maps improve scoring consistency remains unclear.

Objectives To assess the reliability of these scores on non-contrast computed tomography (NCCT) and CTA among numerous raters, with and without access to automated perfusion maps, and their correlations with endovascular thrombectomy (EVT) decisions.

Methods Twenty-nine clinicians independently rated 60 imaging cases from 30 late-window stroke patients with large vessel occlusion, each presented twice: once using NCCT and CTA alone ('non-CTP cases'), and once with the addition of perfusion maps ('CTP cases'). Twenty-eight raters (97%) completed the survey twice. Inter- and intra-reliability was assessed using Gwet's AC1/2 coefficients (κ_G). Correlations with EVT decisions were evaluated using Cramer V.

Results Inter-rater reliability for ASPECTS was substantial and similar between non-CTP (κ_G = 0.65, 95% CI: 0.53 to 0.78) and CTP cases (κ_G = 0.64, 95% CI: 0.51 to 0.77). CS showed an almost perfect reliability in both scenarios (κ_G= 0.82 and 0.83). Intra-rater agreement was at least substantial. Dichotomizing ASPECTS as 0-2/3-10 improved reliability to almost

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Alberta Stroke Program Early CT Score (ASPECTS) and CT angiography (CTA) Collateral Score are commonly used for endovascular thrombectomy (EVT) triage in late-window large vessel occlusion, but their reliability is variable and the incremental value of automated CT perfusion imaging for improving scoring consistency remains uncertain.

WHAT THIS STUDY ADDS

⇒ In a reliability study between 29 clinicians of diverse backgrounds and experiences and assessing 60 late-window large vessel occlusion (LVO) cases, access to automated perfusion maps did not improve the inter- and intra-rater reliability of ASPECTS or CTA Collateral Score (which were both at least substantial when assessed alone), nor their correlation with EVT decisions.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These findings complement the growing body of evidence favoring the use of simpler, reproducible, and more widely available imaging strategies for EVT triage in late-window stroke.

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jnis-2026-025321).

For numbered affiliations see end of article.

Correspondence to
Dr Anass Benomar; anass.benomar@umontreal.ca

FF and JDBD are joint senior authors.

Received 24 March 2026
Accepted 15 April 2026

Check for updates

© Author(s) (or their employer(s)) 2026. No commercial re-use. See rights and permissions. Published by BMJ Group.

To cite: Benomar A, Fahed R, Omar AT, *et al*. *J NeuroIntervent Surg* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/jnis-2026-025321

J NeuroIntervent Surg: first published as 10.1136/jnis-2026-025321 on 29 April 2026. Downloaded from http://jnis.bmj.com/ on May 2, 2026 at SNIS - Society of NeuroInterventional Surgery . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

perfect. ASPECTS strongly correlated with EVT decisions (V=0.80), whereas CS showed moderate correlation (V=0.56), both remaining similar with the addition of CTP (V=0.75 and 0.54, respectively).
Conclusion Automated perfusion imaging did not improve the reliability of ASPECTS or CS. Strong reliability was achieved with NCCT and CTA alone.

INTRODUCTION

Endovascular thrombectomy (EVT) has transformed the management of acute ischemic stroke due to large vessel occlusion (LVO), demonstrating clear benefits over medical management. While initial trials focused on treatment within 6 hours of stroke onset, subsequent studies extended eligibility to patients in the late window (6–24 hours), using advanced imaging to guide selection. In this context, early trials such as DAWN and DEFUSE-3 employed automated perfusion imaging to identify candidates for EVT, typically excluding those with large ischemic cores.[1 2] As a result, automated CT perfusion (CTP) solutions were widely incorporated into guidelines and routine stroke workflows to quantify core and penumbra volumes, particularly in the context of late-window presentation.[3 4] However, more recent randomized trials have challenged the superiority of perfusion imaging for EVT triage,[5] with increasing evidence suggesting comparable EVT outcomes when selection is based on simpler and more accessible imaging modalities, such as noncontrast head CT (NCCT) and CT angiography (CTA).[6]

This evolving paradigm has reignited interest in quick visual scoring systems such as the Alberta Stroke Program Early CT Score (ASPECTS) and the CTA Tan Collateral Score (CS), which can potentially serve as pragmatic surrogates for estimating ischemic core volume and cerebral circulation.[7–9] Despite their widespread use, both ASPECTS and CS have been criticized for insufficient inter-rater reliability in some studies, even among experienced clinicians and after dichotomization.[10 11] With automated perfusion maps increasingly embedded in clinical workflows, it remains unclear whether such advanced imaging tools can actually improve the reliability of ASPECTS and CS by providing clearer delineation of ischemic core and hypoperfused tissue.[12]

The Clinical Uncertainty in Large Vessel Occlusion Ischemic Stroke (CULVO) study recently assessed the impact of automated perfusion imaging on EVT decision making for late-window LVO cases, and found that it did not meaningfully improve reliability among 29 expert clinicians from varied specialties and practices.[13] In the present study, derived from CULVO, we sought to assess whether access to automated perfusion maps enhances the inter- and intra-rater reliability of ASPECTS and CS, and how these scores correlate with EVT recommendations. We hypothesized that the inclusion of perfusion imaging would not significantly enhance the reproducibility of these scores, and that they would demonstrate good correlation with EVT decisions.

MATERIALS AND METHODS

Study design

This inter- and intra-rater reliability study was conducted in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[14] The CULVO study protocol has been previously reported.[15] Approval and waiver for patients' consent were obtained (SMH REB# 22–186).

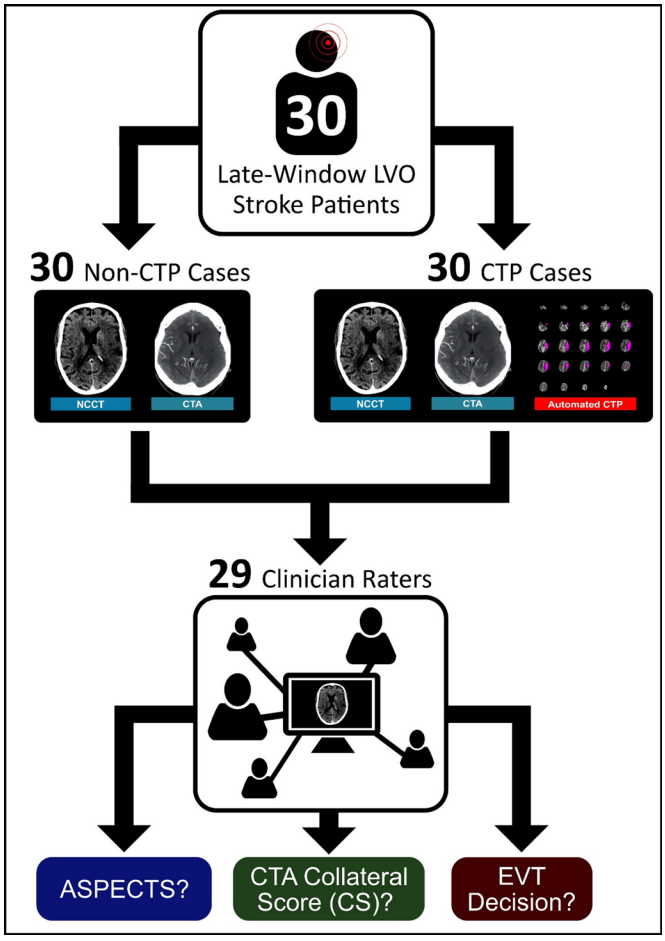


Figure 1 Study protocol. EVT indicates endovascular thrombectomy; LVO, Large vessel occlusion; CTA, CT angiography; CTP, CT perfusion; CS, collateral score.

Case selection

Among 3144 consecutive adult patients undergoing a CT of the head for symptoms of acute stroke between 2018 and 2021 in a single institution (St-Michael's Hospital – Unity Health Toronto, Canada), those presenting with (1) an anterior circulation LVO (M1, carotid terminus, or tandem occlusion); (2) in the late-window (from 6 to 24 hours between stroke onset and time of imaging); (3) a National Institutes of Health Stroke Scale (NIHSS) score of 6 or more; and (4) who underwent NCCT, CTA, and CTP of the head with automated perfusion results (RAPID, iSchemaView, Menlo Park, California, USA) of decent diagnostic quality, were considered eligible for inclusion. To avoid spectrum bias and minimize paradoxes of κ statistics, a representative sample of 30 unique cases was selected spanning a broad distribution of ischemic core volumes in 1:1:1 ratio (small, 0–49 cc; medium, 50–99 cc; and large, >100 cc).[16–18] The rationale for this methodological choice is further explained in the published CULVO protocol.[15] For each unique included patient, two cases were derived, respectively with NCCT and CTA images only ('non-CTP cases'), or with NCCT and CTA supplemented with automated CT perfusion maps ('CTP cases'). The final study portfolio thus included 30 CTP and 30 non-CTP cases (figure 1).

Rater selection

Stroke neurologists, dual-trained interventional stroke neurologists, interventional neuroradiologists, and endovascular

J NeuroIntervent Surg: first published as 10.1136/jnis-2026-025321 on 29 April 2026. Downloaded from http://jnis.bmj.com/ on May 2, 2026 at SNIS - Society of NeuroInterventional Surgery . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Patient characteristics

Characteristics*	
Age	
Mean	72.2 (SD, 16.0)
<70 years	8 (27%)
≥70 years	22 (73%)
Sex	
Male	14 (47%)
Female	16 (53%)
NIHSS at presentation	
Mean	16.9 (SD, 6.5)
5–15	14 (47%)
≥16	16 (53%)
Occlusion laterality	
Right	12 (40%)
Left	18 (60%)
Occlusion location	
M1-MCA	16 (53%)
Carotid terminus	11 (37%)
Tandem occlusion	3 (10%)
Time between onset and CT	
Mean	13.0 (SD, 4.6)
6–9 hour	8 (27%)
9–12 hour	6 (20%)
12–16 hour	8 (27%)
16–24 hour	8 (27%)

*Data are n (%) or mean (SD).
NIHSS, National Institutes of Health Stroke Scale.

Table 2 Rater characteristics

Characteristics*	
Specialty background	n (%)
Neurologist	18 (62%)
Radiologist	7 (24%)
Neurosurgeon	4 (14%)
Type of practice†	
Non-academic	2 (7%)
Academic	27 (93%)
Non-interventionist (Stroke neurologists)	12 (41%)
Interventionist	17 (59%)
Experience†	
0–5 years	16 (55%)
6–10 years	4 (14%)
11–20 years	6 (21%)
>20 years	3 (10%)
Junior (0–10 years)	20 (69%)
Senior (>10 years)	9 (31%)
Percentage of CT use‡	
≤50%	2 (7%)
>50%	27 (93%)
Percentage of CT perfusion use‡	
≤50%	7 (24%)
>50%	22 (76%)
Access to automated CT perfusion software	
No	8 (28%)
Yes	21 (72%)
Percentage of automated CT perfusion use‡	
≤50%	11 (38%)
>50%	18 (62%)

*Data are n (%).
†Frequencies and percentages within this category are not mutually exclusive.
‡For decision-making purposes in patients presenting in the late-window (6–24 hours) since stroke onset.

neurosurgeons, were invited to participate as raters on this study through the Canadian Stroke Consortium and Canadian Interventional Neuro Group. Completion of stroke or neurointerventional fellowship training and active practice in EVT-capable centers were required criteria to participate.

Reliability study

An electronic survey was created and sent to raters using the REDCap online platform (https://www.project-redcap.org/) hosted at St-Michael's Hospital – Unity Health Toronto.[19] Raters provided information regarding their specialty background (neurology, radiology, neurosurgery), years of experience, type of practice (academic vs non-academic, interventionalist vs non-interventionalist), including specific questions about the percentage of late-window anterior circulation LVO patients in their practice being assessed with CT and CTP. Each case was accompanied by relevant demographic and clinical information (age, sex, baseline NIHSS, time from symptom onset to imaging, occlusion laterality and location) and neuroimaging dataset (NCCT in optimized window levels, single-phase CTA with multiplanar reformats, and perfusion maps when applicable). For each case, raters were instructed to independently provide: (1) ASPECTS on baseline NCCT; (2) four-point Tan CS on single-phase CTA (ranging from 0, absent, to 3, 100% supply in the occluded vascular territory)[8]; and (3) whether, based on all available clinical and radiological data, they would recommend EVT. Each rater reviewed all 30 duplicated cases under both imaging sets (non-CTP and CTP) in a random order (figure 1). After a washout period of at least 3 weeks, raters again adjudicated cases in a permuted order, for intra-rater reliability purposes.

Statistical analysis

All analyses were performed in R statistical and computing software version 4.5.1 (R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org/). Continuous descriptive variables were tested for normality using the Shapiro–Wilk test and were presented as mean with SD or median with IQR. Categorical variables were presented as frequencies with percentages. All following analyses were conducted for CTP and non-CTP cases separately. The inter- and intra-rater reliability for ASPECTS and CS was analyzed using Gwet's AC1 or AC2 (κ_G) reliability coefficients for binary or ordinal data respectively, with 95% CI.[16] Reliability was also analyzed among prespecified rater and patient subgroups, and after score dichotomization. We estimated, assuming an anticipated κ_G value of 0.6 (substantial) and an anticipated prevalence of EVT

J NeuroIntervent Surg: first published as 10.1136/jnis-2026-025321 on 29 April 2026. Downloaded from http://jnis.bmj.com/ on May 2, 2026 at SNIS - Society of NeuroInterventional Surgery . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 3 Inter- and intra-rater reliability results for ASPECTS and CS

Radiological scores	Inter-rater reliability*		Intra-rater reliability†	
	CTP	Non-CTP	CTP	Non-CTP
ASPECTS				
Raw	0.64 (0.51, 0.77)	0.65 (0.53, 0.78)	0.76 (SD, 0.10)	0.74 (SD, 0.14)
Dichotomized (0–5/6–10)	0.54 (0.39, 0.69)	0.51 (0.36, 0.66)	0.68 (SD, 0.12)	0.70 (SD, 0.10)
Dichotomized (0–2/3–10)	0.82 (0.72, 0.91)	0.80 (0.69, 0.91)	0.85 (SD, 0.11)	0.87 (SD, 0.11)
Dichotomized (0–3/4–10)	0.66 (0.51, 0.82)	0.67 (0.51, 0.84)	0.76 (SD, 0.13)	0.78 (SD, 0.10)
Trichotomized (0–2/3–5/6–10)	0.69 (0.60, 0.78)	0.67 (0.57, 0.77)	0.77 (SD, 0.07)	0.80 (SD, 0.08)
CS				
Raw	0.83 (0.79, 0.86)	0.82 (0.78, 0.86)	0.88 (SD, 0.07)	0.87 (SD, 0.08)
Dichotomized (0–1/2–3)	0.67 (0.53, 0.81)	0.68 (0.54, 0.82)	0.73 (SD, 0.10)	0.76 (SD, 0.14)

CS indicates single-phase CTA Tan Collateral Score (ranging from 0, absent collaterals, to 3, 100% collateral supply in the occluded vascular territory); CTP, neuroimaging cases supplemented with automated perfusion maps; non-CTP, neuroimaging cases with NCCT and CTA imaging only.
*Data are κ_G (95% CI) between 29 raters.
†Data are mean κ_G (SD) from 28 raters who performed the assessment twice.
ASPECTS, Alberta Stroke Program Early CT Score; CS, Collateral Score; CTP, CT perfusion.

recommendation of 0.5, that a sample size of 30 cases was necessary for the lower limit of a 95% two-sided CI to remain above 0.45 for at least six raters.[20] Interpretation of κ_G estimates was according to Landis and Koch.[21] Non-overlapping 95% CI was used as a criterion to claim a statistical difference between κ_G values.[22] Correlations between ASPECTS, CS, and EVT decisions were explored using Cramer V for each rater, with 95% CI estimated using 500 bootstrap iterations. Cramer V estimates were interpreted as poor below 0.5, moderate between 0.5 and 0.7, and strong above 0.7.

RESULTS

Patients

The patient flow chart has been separately reported.[13] Characteristics of the 30 patients included in the study are summarized in table 1. Most patients were over 70 years-old (73%), female (53%), presented with an M1 occlusion (53%), with a mean NIHSS at presentation of 16.9 (SD, 6.5) and a mean time between stroke onset and imaging of 13 hours (SD, 4.6).

Raters

Twenty-nine raters agreed to participate and completed the survey. Of these, 28 (97%) completed the survey twice. Most raters were within their first 10 years of practice (69%), and were from a neurology background (18/29, 62%) including 12 (41%) practicing as stroke neurologists. All raters except two were working in academic settings. All raters had access to CT perfusion if needed, but nearly a quarter of them did not routinely rely on it for triaging late-window stroke patients (24%) or did not have access to automated perfusion software (28%). Rater characteristics are detailed in table 2.

ASPECTS on baseline NCCT

When ASPECTS was dichotomized as low (0–5) or high (6-10), perfect agreement among all raters was achieved in 6 (20%) non-CTP and 7 (23%) CTP cases. When ASPECTS was dichotomized as very low (0–2) or higher (3–10), perfect agreement among all raters was achieved in 14 (47%) non-CTP and 12 (40%) CTP cases. The number of cases (out of 30) for which very low ASPECTS was adjudicated varied between 0 and 10 depending on the rater for non-CTP cases and between 0 and 11 with the addition of automated perfusion maps. For cases adjudicated as higher ASPECTS, the number ranged between 20 and 30 for non-CTP cases and between 19 and 30 for CTP cases.

Main reliability results for ASPECTS and CS are shown in table 3. Overall reliability of ASPECTS between 29 raters was substantial for both non-CTP (κ_G = 0.65; 95% CI, 0.53 to 0.78) and CTP cases (κ_G = 0.64; 95% CI, 0.51 to 0.77), yet with overlapping CI. Dichotomizing ASPECTS as 0–2 vs. 3–10 improved reliability to almost perfect for both non-CTP (κ_G = 0.80, 95% CI, 0.69 to 0.91) and CTP cases (κ_G = 0.82, 95% CI, 0.72 to 0.91). Other dichotomization cut-offs did not improve reliability of the score. No significant differences were identified between CTP and non-CTP cases for all rater and patient subgroups (online supplemental materials).

Intra-rater reliability for non-CTP cases varied between 0.40 and 0.92, with a substantial mean κ_G of 0.74 (SD, 0.14). For cases supplemented with automated perfusion maps, intra-rater reliability varied from 0.53 to 0.95, also with a substantial mean κ_G (0.76; SD, 0.10). Individual and subgroup intra-rater estimates are available in the online supplemental materials tables 1 and 2.

Single-phase CTA collateral score

When CS was dichotomized as poor (0–1) or good (2-3), perfect agreement between 29 raters was achieved in 12 (40%) non-CTP and 9 (30%) CTP cases. The number of cases (out of 30) for which poor CS was adjudicated varied between 11 (37%) and 19 (63%) depending on the rater, and was slightly lower with CTP (from 8 to 19, respectively). The number of cases with good CS varied between 11 (37%) and 19 (63%) for non-CTP cases, and from 11 (37%) to 22 (73%) for CTP cases.

Inter-rater reliability for CS was almost perfect and similar for non-CTP (κ_G = 0.82; 95% CI, 0.78 to 0.86) and CTP cases (κ_G = 0.83; 95% CI, 0.79 to 0.86). This remained true for all patient and rater subgroups (online supplemental materials figure 3 and 4). Dichotomization did not improve reliability. Intra-rater reliability varied between 0.56 and 0.96 for non-CTP cases, and 0.57 and 0.94 for CTP cases, with almost perfect mean reliabilities of 0.87 (SD, 0.08) and 0.88 (SD, 0.07), respectively (table 3, online supplemental materials tables 1 and 2).

J NeuroIntervent Surg: first published as 10.1136/jnis-2026-025321 on 29 April 2026. Downloaded from http://jnis.bmj.com/ on May 2, 2026 at SNIS - Society of NeuroInterventional Surgery . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

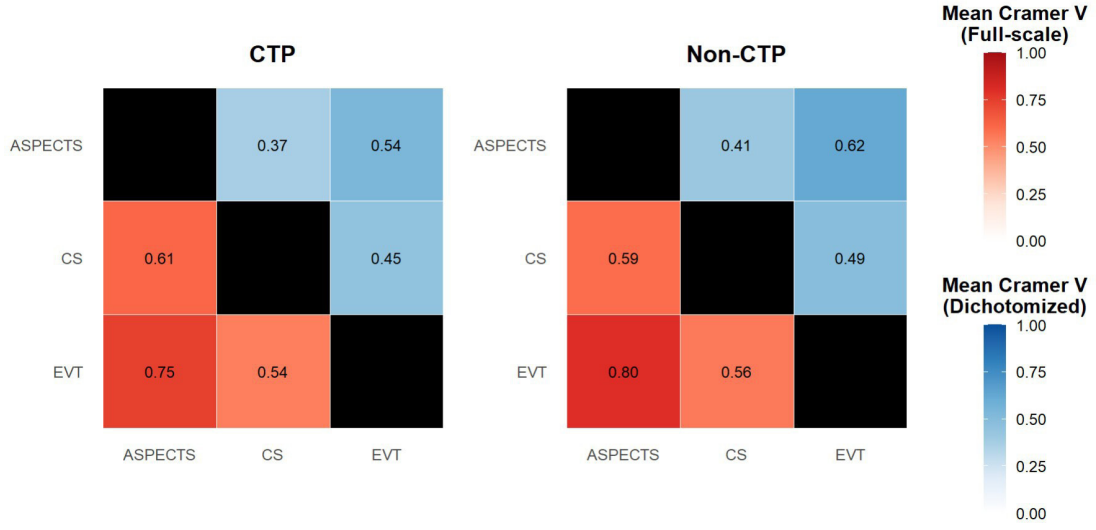


Figure 2 Correlations between raw and dichotomized ASPECTS, CS, and EVT decisions. Mean Cramer V was used to estimate correlations between raw (red-colored) or dichotomized (blue-colored) radiological scores and EVT recommendations, for cases with NCCT and CTA only ('non-CTP') and those supplemented with automated perfusion maps ('CTP'). Dichotomization of ASPECTS was 0–5 vs 6–10, and of CS as 0–1 vs 2–3. CS: single-phase CTA Tan collateral scoring system (ranging from 0, absent collaterals, to 3, 100% collateral supply in the occluded vascular territory), EVT, Endovascular thrombectomy.

Correlation between radiological scores and EVT recommendations

Descriptive statistics and reliability analyses regarding EVT decisions were separately explored in a dedicated paper.[13] Individual and mean correlation estimates between ASPECTS, CS, and EVT recommendations are shown in online supplemental materials tables 3 and 4. Correlations between ASPECTS and EVT decisions varied between 0.62 and 0.95 for CTP cases, and between 0.56 and 1 for non-CTP cases, depending on the rater, with a strong mean Cramer V of 0.75 (SD, 0.09) and 0.80 (SD, 0.11), respectively (figure 2). Correlations between CS and EVT decisions varied between 0.23 and 0.80 for CTP cases, and 0.23 and 0.82 for non-CTP cases, with a moderate mean Cramer V of 0.54 (SD, 0.14) and 0.56 (SD, 0.16), respectively. Correlation between ASPECTS and CS was better than that of CS with EVT decisions, and lower than that of ASPECTS with EVT decisions, with a mean Cramer V of 0.61 (SD, 0.09) and 0.59 (SD, 0.06) for CTP and non-CTP cases, respectively. The use of dichotomized scores did not improve correlation with EVT recommendations (figure 2).

DISCUSSION

In our study we assessed the reliability for ASPECTS, and CS across 29 raters, comparing interpretations with and without automated perfusion maps. We also assessed the correlation of these radiological scores with EVT decisions, with and without CTP. A few observations emerge.

First, adding CTP maps did not significantly enhance the reliability of either imaging score. This was true for all rater and patient subgroups. To our knowledge, no prior studies specifically assessed the comparative reliabilities of these scores on NCCT and CTA in combination –or not– with CTP. A few studies evaluated the reliabilities of ASPECTS and CS derived from multiphase CTP source images, and demonstrated higher reliability for CTP-based ASPECTS and similar reliability for CTP-based CS when compared with routine NCCT- and CTA-based scores.[23 24] Our results cannot be validly compared with these studies for methodological reasons as CTP-based ASPECTS or CS were assessed separately from (and not in combination with) NCCT/CTA results, and prior studies did not use automated CTP maps. In practical terms, our findings suggest that, whether or not automated perfusion maps are available, clinicians are likely to assign the same ASPECTS and CS when reviewing the same patient.

Second, both ASPECTS and CS proved to be reproducible among various clinicians of diverse backgrounds and experience levels, even without dichotomization, showing substantial (κ_G = 0.65) and almost perfect (κ_G = 0.82) reliability, respectively. Dichotomizing ASPECTS as 0–2 versus 3–10 yielded an almost perfect agreement, a finding of interest in the context of recent large ischemic core thrombectomy trials that use this threshold. Because reliability is a prerequisite for clinical usefulness, a score that cannot be reproduced across (and within) readers cannot be meaningfully applied in clinical trials or practice. A systematic review from 2016 of 30 studies assessing ASPECTS reliability found κ values ranging from 0.17 (slight) to 0.97 (almost perfect), with great heterogeneity in terms of methods, populations, number of readers, choice of coefficient, and presence of 95% CI.[10] A more recent meta-analysis of 20 studies focusing on regional and total ASPECTS found substantial pooled reliability (κ=0.67), which is in line with our results, however, with the omission of a few recent papers such as that of van Horn *et al*, which found slight agreement (0.13) between 100 investigators for 20 cases.[25] The reported reliability of CS in the literature was generally higher than that of ASPECTS and varied according to CTA method (single-phase, multiphase, dynamic/time-resolved) and scoring system used (eg, Tan, modified/dichotomized Tan, Miteff, Rosenthal, Menon). Studies specifically assessing Tan single-phase CS yielded moderate to substantial agreements, placing our observed reliability estimates at the higher end of the spectrum.[11 24 26 27] Higher reliability, within the same range as ours, was usually found when CS was adjudicated on multiphase or dynamic CTA.[28 29]

Third, ASPECTS and CS were shown to be influential in clinical decision making, as they respectively strongly and moderately correlated with EVT decisions (figure 2). Results related to EVT recommendations were explored and discussed in great detail in the main CULVO study.[15] Taken together, our results

J NeuroIntervent Surg: first published as 10.1136/jnis-2026-025321 on 29 April 2026. Downloaded from http://jnis.bmj.com/ on May 2, 2026 at SNIS - Society of NeuroInterventional Surgery . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

convey the reassuring message that NCCT-based ASPECTS and CTA-based CS, two visual scores that any stroke clinicians can readily assess with minimal training, can be trusted as reliable tools when triaging EVT patients. These findings complement the growing body of evidence supporting simpler imaging approaches for late-window stroke evaluation.[6] This is particularly relevant in real-world practice, where access to automated perfusion imaging remains variable, even in high-income countries, and is often limited or absent in resource-constrained settings.

Our study had several limitations. First, only representative images from axial NCCT and single-phase CTA were provided to raters. These might have decreased the variability that could have been seen in a more realistic context of evaluation, but was a necessary compromise to recruit many diversified raters. Second, while the initial case selection involved consecutive late-window stroke patients with anterior LVO, the number of selected cases had to be reduced to make the prospect of rating twice the portfolio acceptable to numerous raters. The final case portfolio was carefully selected to include a broad spectrum of LVO cases based on their ischemic core volumes, a necessary step to avoid the known paradoxes of kappa statistics.[16] Fourth, while CTP and non-CTP cases were presented in random order, there is a potential risk that raters could recall and link CTP and non-CTP cases based on shared clinical information (carry-over effect). Fourth, our study only focused on single-phase CTA collateral score as proposed by Tan *et al*.[8] The collateral status might be less physiologically accurate and underestimated compared with multiphase CTA or time-resolved CTP, and reliability results might differ with other CTA methods or scoring systems. The choice of single-phase CTA was driven by its wider availability and consistent routine use across various facilities. Fifth, since all participating raters were expert stroke clinicians, results cannot be validly extrapolated to physicians with different training scopes. Whether CTP would be more useful in other rater populations remains to be explored. Finally, the reliability of automated ASPECTS or CS, of regional ASPECTS, as well as the performance of these scores in improving patient selection or predicting patient clinical outcomes or final core volumes – aided or not by perfusion maps – were beyond the scope of this study and could be assessed in future studies.

CONCLUSION

Automated perfusion imaging did not improve the reliability of ASPECTS or CS, nor did it enhance their correlations with EVT decisions. Sufficient reliability was already achieved with NCCT and CTA alone, particularly when ASPECTS was dichotomized using clinically relevant thresholds.

Author affiliations
[1]Department of Medical Imaging, Division of Diagnostic and Therapeutic Neuroradiology, St Michael's Hospital, Toronto, Ontario, Canada
[2]Department of Medicine, Division of Neurology, University of Ottawa, Ottawa, Ontario, Canada
[3]Department of Surgery, Division of Neurosurgery, McMaster University Faculty of Health Sciences, Hamilton, Ontario, Canada
[4]Division of Neurosurgery, University of the Philippines-Philippine General Hospital Department of Surgery, Manila, Philippines
[5]Department of Medicine, Division of Neurology, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada
[6]Joint Department of Medical Imaging (JDMI), Division of Interventional Neuroradiology, Toronto Western Hospital, Toronto, Ontario, Canada
[7]Department of Radiology, University of Calgary, Calgary, Alberta, Canada
[8]Department of Surgery, Division of Neurosurgery, Royal Inland Hospital, Kamloops, British Columbia, Canada
[9]Department of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada
[10]Department of Neurosciences, Université de Montréal, Montreal, Quebec, Canada
[11]Department of Surgery, Division of Neurosurgery, The University of British Columbia Faculty of Medicine, Vancouver, British Columbia, Canada
[12]Department of Neurosurgery, Juntendo University Hospital, Bunkyo-ku, Japan
[13]Department of Interventional Neuroradiology, Rothschild Foundation Hospital, Paris, France
[14]Department of Medical Imaging, McMaster University, Toronto, Ontario, Canada
[15]Departments of Clinical Neurosciences, Radiology, and Community Health Sciences, University of Calgary, Calgary, Alberta, Canada
[16]Department of Surgery, Division of Neurosurgery, Universite Laval Faculte de Medecine, Quebec, Quebec, Canada
[17]Division of Clinical Sciences, Northern Ontario School of Medicine University, Sudbury, Ontario, Canada
[18]Department of Radiology, Queen's University, Kingston, Ontario, Canada
[19]Department of Medicine, Division of Neurology, McMaster University, Hamilton, Ontario, Canada
[20]Department of Radiology, Radiation Oncology and Nuclear Medicine, Université de Montréal, Montreal, Quebec, Canada
[21]Department of Medicine, Division of Neurology, Toronto Western Hospital, Toronto, Ontario, Canada
[22]Department of Medical Imaging, Royal University Hospital, Saskatoon, Saskatchewan, Canada
[23]Department of Internal Medicine, Section of Neurology, University of Manitoba, Winnipeg, Manitoba, Canada
[24]Department of Clinical Neurological Sciences, Western University, London, Ontario, Canada
[25]Department of Medicine, Queen's University, Kingston, Ontario, Canada
[26]Department of Diagnostic and Interventional Radiology, Divisions of Neuroradiology & Neurointervention, The Hospital for Sick Children, Toronto, Ontario, Canada
[27]Department of Medicine, Division of Neurology, St Michael's Hospital, Toronto, Ontario, Canada
[28]Department of Surgery, Division of Neurosurgery, St Michael's Hospital, Toronto, Ontario, Canada
[29]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

Social media Anass Benomar, X @AnassBenomarMD; Abdelsimar Tan Omar, X @atomar_md; Christine Hawkes, X @CMHawkes; Muneer Eesa, X @Muneereesa; Alexandre Y Poppe, X @PoppeAY; Karl Narvacan, X @karlnarvacan; Mohammed A Almekhlafi, X @AlmekhlafiMa; Aviraj Deshmukh, X @IamAvirajD; Kislay Kishore, X @kislaykishore9; Nima Kashani, X @nimakashani; Nishita Singh, X @nishita_singh3; Carmen Parra-Fariñas, X @CarmenPFarinas; Vitor M Pereira, X @VitorMendesPer1; Jose Danilo Bengzon Diestro, X @DanniDiestro

Contributors All the authors meet the following criteria: 1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; 2. Drafting the work or revising it critically for important intellectual content; 3. Final approval of the version to be published; 4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JDBD is the overall guarantor of the study. He accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding JDBD is a recipient of the Fergus Mills Scholarship of McMaster University for the Clinical Uncertainty in Large Vessel Occlusion (CULVO) study. This sponsor was not directly or indirectly involved in any of the study steps.

Competing interests JDBD: Honoraria from Medtronic. Travel grant and consulting fees from Microvention; MAA: Member of the scientific advisory board of Palmera Medical; TRM: Principal of eVasc Neurovascular. The rest of the authors have no competing interests to disclose.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Unity Health-St. Michael's Hospital (SMH REB# 22-186). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; internally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines,

J NeuroIntervent Surg: first published as 10.1136/jnis-2026-025321 on 29 April 2026. Downloaded from http://jnis.bmj.com/ on May 2, 2026 at SNIS - Society of NeuroInterventional Surgery . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Anass Benomar https://orcid.org/0000-0002-3735-1686
Robert Fahed https://orcid.org/0000-0002-1887-5097
Abdelsimar Tan Omar https://orcid.org/0000-0002-2965-8506
Christine Hawkes https://orcid.org/0000-0001-5524-9709
Eef J Hendriks https://orcid.org/0000-0003-2100-7992
Clare A Enriquez https://orcid.org/0000-0001-5177-0521
Muneer Eesa https://orcid.org/0000-0001-5297-9618
Grant Stotts https://orcid.org/0000-0001-8226-4883
Hubert Lee https://orcid.org/0000-0003-0058-1020
Shashank Nagendra https://orcid.org/0000-0002-1538-7549
Alexandre Y Poppe https://orcid.org/0000-0002-1419-2635
Célina Ducroux https://orcid.org/0000-0001-8938-9494
Timothy R Lim https://orcid.org/0000-0002-4900-4892
Karl Narvacan https://orcid.org/0000-0002-6237-4775
Michael A Rizzuto https://orcid.org/0000-0001-7743-4454
Afra Alfalahi https://orcid.org/0000-0002-0013-0951
Hidehisa Nishi https://orcid.org/0000-0003-1763-2517
Pragyan Sarma https://orcid.org/0000-0001-9827-7937
Ze'ev Itsekson Hayosh https://orcid.org/0000-0001-8320-4129
Katrina Ignacio https://orcid.org/0000-0001-7263-8474
William Boisseau https://orcid.org/0000-0001-8120-5128
Eduardo Pimenta Ribeiro Pontes Almeida https://orcid.org/0000-0001-8196-866X
Mohammed A Almekhlafi https://orcid.org/0000-0001-9550-8197
Geneviève Milot https://orcid.org/0000-0002-2847-8482
Aviraj Deshmukh https://orcid.org/0000-0002-5603-6307
Kislay Kishore https://orcid.org/0000-0001-5754-9270
Donatella Tampieri https://orcid.org/0000-0002-0329-3262
Jefffrey Z Wang https://orcid.org/0000-0003-4119-6524
Abhilekh Srivastava https://orcid.org/0009-0005-2617-4222
Daniel Roy https://orcid.org/0000-0002-6066-4453
Federico Carpani https://orcid.org/0000-0003-4696-5994
Nima Kashani https://orcid.org/0000-0002-9752-784X
Claudia Candale-Radu https://orcid.org/0000-0001-7294-7701
Nishita Singh https://orcid.org/0000-0003-3647-0904
Maria Bres-Bullrich https://orcid.org/0000-0003-4911-9937
Robert Sarmiento https://orcid.org/0000-0001-7708-1804
Ryan T Muir https://orcid.org/0000-0002-9349-1069
Carmen Parra-Fariñas https://orcid.org/0000-0002-6444-2188
Stephanie D Reiter https://orcid.org/0000-0001-5647-8466
Yan Deschaintre https://orcid.org/0000-0002-1037-2700
Ravinder-Jeet Singh https://orcid.org/0000-0003-1201-8067
Vivek Bodani https://orcid.org/0009-0005-3727-1060
Aristeidis H Katsanos https://orcid.org/0000-0002-6359-0023
Ronit Agid https://orcid.org/0000-0003-4686-1685
Atif Zafar https://orcid.org/0000-0002-2738-5225
Vitor M Pereira https://orcid.org/0000-0002-6804-3985
Julian Spears https://orcid.org/0000-0002-4111-6929
Thomas R Marotta https://orcid.org/0000-0002-2480-2054
Pascal Djiadeu https://orcid.org/0000-0001-9708-6530
Sunjay Sharma https://orcid.org/0000-0001-7826-9075
Forough Farrokhyar https://orcid.org/0000-0001-9928-9016
Jose Danilo Bengzon Diestro https://orcid.org/0000-0001-8450-2021

REFERENCES

1 Albers GW, Marks MP, Kemp S, *et al*. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *N Engl J Med* 2018;378:708–18.
2 Nogueira RG, Jadhav AP, Haussen DC, *et al*. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *N Engl J Med* 2018;378:11–21.
3 Powers WJ, Rabinstein AA, Ackerson T, *et al*. 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.
4 Vagal A, Wintermark M, Nael K, *et al*. Automated CT perfusion imaging for acute ischemic stroke: Pearls and pitfalls for real-world use. *Neurology (ECronicon)* 2019;93:888–98.
5 Chen H, Lee JS, Michel P, *et al*. Endovascular Stroke Thrombectomy for Patients With Large Ischemic Core: A Review. *JAMA Neurol* 2024;81:1085–93.
6 Nguyen TN, Abdalkader M, Nagel S, *et al*. Noncontrast Computed Tomography vs Computed Tomography Perfusion or Magnetic Resonance Imaging Selection in Late Presentation of Stroke With Large-Vessel Occlusion. *JAMA Neurol* 2022;79:22–31.
7 Hang Y, Wang CD, Ni H, *et al*. Alberta Stroke Program Early CT Score and collateral status predict target mismatch in large vessel occlusion with delayed time windows. *J Neurointerv Surg* 2023;15:876–80.
8 Tan IYL, Demchuk AM, Hopyan J, *et al*. CT angiography clot burden score and collateral score: correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct. *AJNR Am J Neuroradiol* 2009;30:525–31.
9 Dehkharghani S, Bammer R, Straka M, *et al*. Performance of CT ASPECTS and Collateral Score in Risk Stratification: Can Target Perfusion Profiles Be Predicted without Perfusion Imaging? *AJNR Am J Neuroradiol* 2016;37:1399–404.
10 Farzin B, Fahed R, Guilbert F, *et al*. Early CT changes in patients admitted for thrombectomy: Intrarater and interrater agreement. *Neurology (ECronicon)* 2016;87:249–56.
11 Dolotova DD, Blagosklonova ER, Muslimov RS, *et al*. Inter-Rater Reliability of Collateral Status Assessment Based on CT Angiography: A Retrospective Study of Middle Cerebral Artery Ischaemic Stroke. *J Clin Med* 2023;12:5470.
12 Diestro JDB. Rethinking the role of CT perfusion in the management of emergent large vessel ischemic stroke. *J Neurointerv Surg* 2023;15:833–4.
13 Diestro JDB, Fahed R, Omar AT, *et al*. Clinical uncertainty in large vessel occlusion ischemic stroke: does automated perfusion imaging make a difference? An intra-rater and inter-rater agreement study. *J Neurointerv Surg* 2025;17:192–9.
14 Kottner J, Audigé L, Brorson S, *et al*. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.
15 Diestro JDB, Fahed R, Benomar A, *et al*. Clinical Uncertainty in Large Vessel Occlusion ischemic stroke (CULVO): Does automated perfusion scanning make a difference? Protocol of an intrarater and interrater agreement study. *PLoS One* 2024;19:e0297520.
16 Benomar A, Zarour E, Létourneau-Guillon L, *et al*. Measuring Interrater Reliability. *Radiology* 2023;309:e230492.
17 Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551–8.
18 Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543–9.
19 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
20 Donner A, Rotondi MA. Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. *Int J Biostat* 2010;6:31.
21 Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;33:159.
22 Bacchus E, Kate MP, Benomar A, *et al*. Inter-rater reliability of the simplified Modified Rankin Scale as an outcome measure for treated cerebral aneurysm patients. *Neurochirurgie* 2022;68:488–92.
23 Finlayson O, John V, Yeung R, *et al*. Interobserver agreement of ASPECT score distribution for noncontrast CT, CT angiography, and CT perfusion in acute stroke. *Stroke* 2013;44:234–6.
24 Kauw F, Dankbaar JW, Martin BW, *et al*. Collateral Status in Ischemic Stroke: A Comparison of Computed Tomography Angiography, Computed Tomography Perfusion, and Digital Subtraction Angiography. *J Comput Assist Tomogr* 2020;44:984–92.
25 van Horn N, Kniep H, Broocks G, *et al*. ASPECTS Interobserver Agreement of 100 Investigators from the TENSION Study. *Clin Neuroradiol* 2021;31:1093–100.
26 Wolff L, Su J, Van Loon D, *et al*. Inter-rater reliability for assessing intracranial collaterals in patients with acute ischemic stroke: comparing 29 raters and an artificial intelligence-based software. *Neuroradiology* 2022;64:2277–84.
27 Yang W, Soomro J, Jansen IGH, *et al*. Collateral Capacity Assessment : Robustness and Interobserver Agreement of Two Grading Scales and Agreement with Quantitative Scoring. *Clin Neuroradiol* 2023;33:353–9.
28 van den Wijngaard IR, Holswilder G, Wermer MJH, *et al*. Assessment of Collateral Status by Dynamic CT Angiography in Acute MCA Stroke: Timing of Acquisition and Relationship with Final Infarct Volume. *AJNR Am J Neuroradiol* 2016;37:1231–6.
29 Reid M, Famuyide AO, Forkert ND, *et al*. Accuracy and Reliability of Multiphase CTA Perfusion for Identifying Ischemic Core. *Clin Neuroradiol* 2019;29:543–52.

J NeuroIntervent Surg: first published as 10.1136/jnis-2026-025321 on 29 April 2026. Downloaded from http://jnis.bmj.com/ on May 2, 2026 at SNIS - Society of NeuroInterventional Surgery . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.