

Early CT changes in patients admitted for thrombectomy

Intrarater and interrater agreement

Behzad Farzin, MD*
 Robert Fahed, MD*
 Francois Guilbert, MD,
 FRCPC
 Alexandre Y. Poppe, MD,
 FRCPC
 Nicole Daneault, MD,
 FRCPC
 André P. Durocher, MD,
 FRCPC
 Sylvain Lanthier, MD,
 FRCPC
 Hayet Boudjani, MD
 Naim N. Khoury, MD
 Daniel Roy, MD, FRCPC
 Alain Weill, MD, FRCPC
 Jean-Christophe Gentric,
 MD
 André L. Batista, MD
 Laurent Létourneau-
 Guillon, MD, FRCPC
 François Bergeron, MD
 Marc-Antoine Henry,
 MD
 Tim E. Darsaut, MD,
 FRCSC
 Jean Raymond, MD,
 FRCPC

Correspondence to
 Dr. Raymond:
jean.raymond@umontreal.ca

Editorial, page 242

Supplemental data
 at Neurology.org

ABSTRACT

Objective: To systematically review the literature and assess agreement on the Alberta Stroke Program Early CT Score (ASPECTS) among clinicians involved in the management of thrombectomy candidates.

Methods: Studies assessing agreement using ASPECTS published from 2000 to 2015 were reviewed. Fifteen raters reviewed and scored the anonymized CT scans of 30 patients recruited in a local thrombectomy trial during 2 independent sessions, in order to study intrarater and interrater agreement. Agreement was measured using intraclass correlation coefficients (ICCs) and Fleiss kappa statistics for ASPECTS and dichotomized ASPECTS at various cutoff values.

Results: The review yielded 30 articles reporting 40 measures of agreement. Populations, methods, analyses, and results were heterogeneous (slight to excellent agreement), precluding a meta-analysis. When analyzed as a categorical variable, intrarater agreement was slight to moderate ($\kappa = 0.042\text{--}0.469$); it reached a substantial level ($\kappa > 0.6$) in 11/15 raters when the score was dichotomized (0–5 vs 6–10). The interrater ICCs varied between 0.672 and 0.811, but agreement was slight to moderate ($\kappa = 0.129\text{--}0.315$). Even in the best of cases, when ASPECTS was dichotomized as 0–5 vs 6–10, interrater agreement did not reach a substantial level ($\kappa = 0.561$), which translates into at least 5 of 15 raters not giving the same dichotomized verdict in 15% of patients.

Conclusions: In patients considered for thrombectomy, there may be insufficient agreement between clinicians for ASPECTS to be reliably used as a criterion for treatment decisions.

Neurology® 2016;87:249–256

GLOSSARY

ASPECTS = Alberta Stroke Program Early CT Score; **EASI** = Endovascular Acute Stroke Intervention Trial; **ICC** = intraclass correlation coefficient; **NIHSS** = NIH Stroke Scale score; **PACS** = Picture Archiving and Communication System.

The Alberta Stroke Program Early CT Score (ASPECTS) was described in 2000.¹ The scale was designed to summarize in one global score (0–10) early signs of ischemia in patients with acute ischemic stroke in a more systematic approach than the one-third middle cerebral artery territory rule previously proposed as an eligibility criterion for clinical studies on IV thrombolysis.¹ Recently, all thrombectomy trials have used ASPECTS for patient selection^{2–4} or for subgroup analyses.⁵ Some national stroke guidelines have incorporated ASPECTS in their recommendations for selecting patients for thrombectomy.^{6,7} It is anticipated that future trials and databases, as well as quality control studies, will require ASPECTS for every patient. Such an important role requires the demonstration that various users come to the same verdict when they examine the same patients.

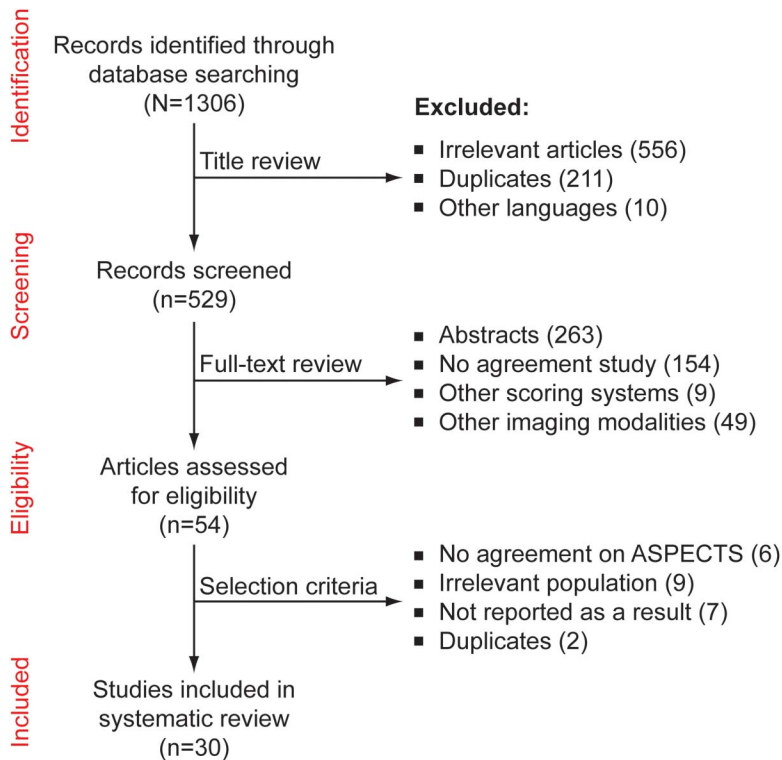
We aimed to systematically review the literature on agreement studies on the ASPECTS system. We then designed an agreement study involving the clinicians managing thrombectomy candidates at one institution in order to study reliability in this clinical context.

*These authors contributed equally to this work.

From the Departments of Radiology (B.F., R.F., F.G., N.N.K., D.R., A.W., J.-C.G., A.L.B., L.L.-G., F.B., M.-A.H., J.R.) and Neurosciences (A.Y.P., N.D., A.P.D., S.L., H.B.), Centre Hospitalier de l'Université de Montréal (CHUM), Notre-Dame Hospital, Montreal; and the Department of Surgery, Division of Neurosurgery (T.E.D.), University of Alberta Hospital, Mackenzie Health Sciences Centre, Edmonton, Canada.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Figure 1 Flow diagram of systematic review



ASPECTS = Alberta Stroke Program Early CT Score.

METHODS Systematic review. A detailed protocol was designed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement prior to conducting the systematic review.⁸ The search strategy is provided in appendix e-1 on the *Neurology*[®] Web site at Neurology.org.

The electronic database search included publications up to October 2015, with no starting date specification. We included any study that reported in the Results estimates of reliability or agreement using ASPECTS on noncontrast CT scan. Articles reporting the same data in duplicate publications and studies on populations not relevant to our research question (such as patients with established infarction examined after 12 hours) were excluded. All articles were reviewed and reports of reliability were assessed by one author (B.F.). Ambiguous items were discussed and resolved through consensus with 2 other authors (J.R. and R.F.).

Agreement study. The study was prepared in accordance with the Guidelines for Reporting Reliability and Agreement Studies.⁹

Patients. All included CT scans were selected from patients who were recruited in a pragmatic randomized trial of thrombectomy, the Endovascular Acute Stroke Intervention Trial (EASI) (ClinicalTrials.gov Identifier: NCT02157532). EASI did not require minimum ASPECTS to include patients in the trial. The image-based exclusion criterion was “established infarction or hemorrhagic transformation of the target symptomatic territory.”

Two authors (R.F. and B.F.) reviewed all 77 EASI patients, and graded the anterior circulation occlusion (n = 68) in a consensus session in order to select 30 scans that captured a distribution of scores. This included approximately 10% of examinations with a poor (<5), 40% with a moderate (5–7), and 50% with

a good score (8–10). CT scans were either performed at the thrombectomy center (the trial site, CHUM Notre-Dame Hospital, Montreal; n = 20) or at an outside referring hospital (n = 10). CT studies that were difficult to interpret because of previous stroke, atrophy, or movement artefacts were not excluded.

Cases and proportions were chosen to resemble a typical thrombectomy case series, to provide a significant number of patients with midrange scores, and to minimize paradoxes of kappa statistics.^{10,11}

Sample size was determined after consultation with the study statistician, taking into account pragmatic considerations (how many cases raters would be willing to examine in a single session) and an estimate of the number required to keep the lower boundary of the confidence interval within a predetermined limit (0.4) assuming good agreement ($\kappa > 0.6$) between raters.¹²

Readings. The CT studies were anonymized and reuploaded to the server of the hospital Picture Archiving and Communication System (PACS). Readers had no access to other imaging studies or to clinical information other than sex, age, symptoms (i.e., left or right motor deficit, aphasia), NIH Stroke Scale score (NIHSS), time of symptom onset, and time of head CT examination. An electronic case report form was created (Microsoft Access; Microsoft, Redmond, CA). The scoring system was displayed on a computer screen adjacent to the PACS, showing 2 brain slices extracted from the ASPECTS Web site (<http://www.aspectsinstroke.com/training-for-aspects/test/>). The readers could then click on any of the 10 ASPECTS-defined brain regions where they thought early ischemic changes (defined as sulcal effacement, a loss of gray–white differentiation, or a hypodensity) were present. The ASPECTS for each scan was then generated and saved automatically.

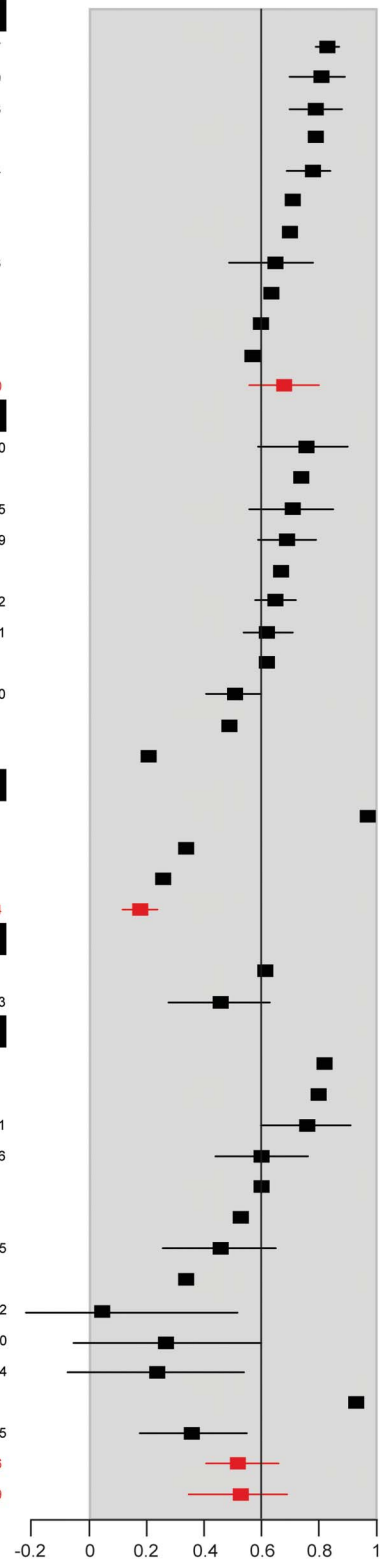
Each reading session was independently performed in a dedicated neuroradiology reading room provided with 2 screens (BARCO E-3620 3 MP medical flat grayscale display). The rater had no time restraints; full access to the entire slice set was given and modification of the window and level of the image were allowed as needed. One author (R.F.) supervised the reading sessions. Each rater performed 2 independent reading sessions, evaluating all 30 cases at one sitting, under the same conditions, at least 3 weeks apart. The order of the studies was the same for the 2 reading sessions.

Raters. All raters were clinicians involved in the CT assessment of patients referred for thrombectomy. To promote participation, raters were offered to collaborate in the design and reporting of the present work. Fifteen raters from a single institution included 6 vascular neurologists (4 seniors), 5 interventional neuroradiologists (3 seniors), and 4 diagnostic neuroradiologists (2 seniors). Readers were considered senior if they had more than 10 years’ experience managing acute stroke. All readers routinely use ASPECTS to score CT scans of patients with ischemic stroke. Yet each participant was required to undergo the online ASPECTS training (<http://www.aspectsinstroke.com/training-for-aspects/test/cases/>) within 15 days of the first reading session. Scores of reference from a 16 rater were added to the final dataset for each of the 30 cases, representing the actual ASPECTS given by the neuroradiologist on-call (8 neuroradiologists, all of whom routinely use ASPECTS) at the time of initial patient management.

Statistical analysis. ASPECTS were first analyzed as a continuous variable. The mean ASPECTS were compared using analysis of variance for repeated measures followed by post hoc tests adjusted with the Bonferroni correction. Intrarater and interrater reliability was first estimated using the intraclass correlation coefficient (ICC). The ASPECTS were then analyzed as a categorical

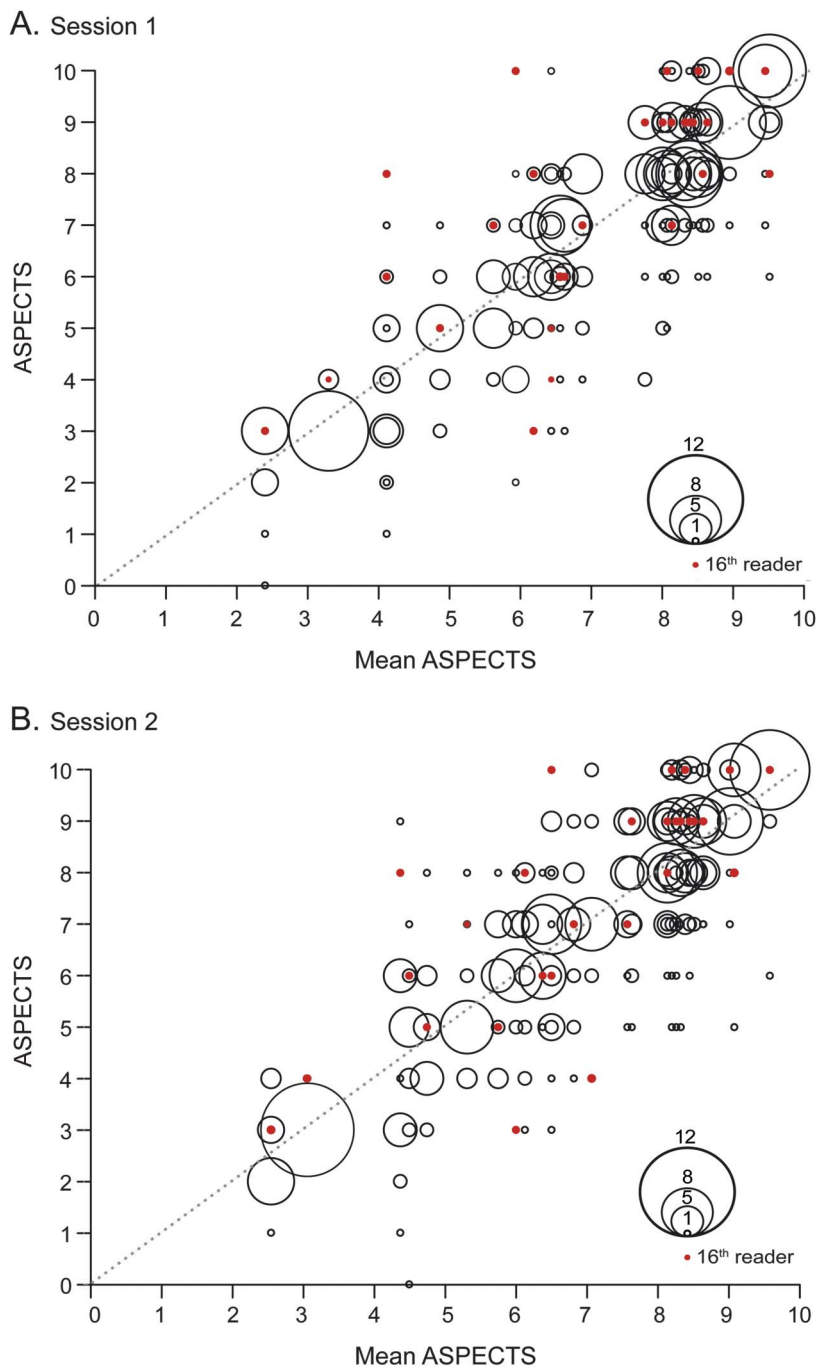
Figure 2 Results of systematic review

| Ref. no. * | Number of patients | NIHSS Median (IQR) Mean ± SD | Number of raters | ASPECTS Median (IQR) | ASPECTS Mean ± SD | Dichotomized/Trichotomized | Value | 95% CI |
|------------------|--------------------|------------------------------|------------------|----------------------|-------------------|----------------------------|-------------|------------------|
| ICC | | | | | | | | |
| 1 | 181 | 6 (0 - 28) | 4 | - | - | - | 0.83 | 0.79-0.87 |
| 2 | 44 | 16 (12 - 19) | 5 | 9 (8 - 10) | - | - | 0.81 | 0.70-0.89 |
| 3 | 37 | 16 (10 - 20) | 5 | - | 8.9 ± 1.9 | - | 0.79 | 0.70-0.88 |
| 4 | 40 | - | 5 | - | - | - | 0.79 | - |
| 5 | 64 | 14 (9 - 19) | 4 | 7.5 (6 - 8) | - | - | 0.78 | 0.69-0.84 |
| 6 | 39 | 16 (9 - 22) | 2 | - | - | - | 0.71 | - |
| 7 | 44 | 17 (13.5 - 20) | 3 | 9 (8 - 10) | - | - | 0.70 | - |
| 8 | 36 | 15 | 3 | - | - | - | 0.65 | 0.49-0.78 |
| 9 | 55 | 15 | 2 | 7 | - | - | 0.63 | - |
| 10 | 105 | 6 (3 - 12) | 2 | - | - | - | 0.60 | - |
| 11 | 74 | 17 (12 - 20) | 2 | - | 7.6 | - | 0.57 | - |
| Our study | 30 | 17.5 (14.25 - 23) | 15 | 8 (6 - 9) | - | - | 0.68 | 0.56-0.80 |
| Kw | | | | | | | | |
| 12 | 143 | 4 (2 - 10) | 2 | 10 (9 - 10) | - | - | 0.75 | 0.59 - 0.90 |
| 13 | 53 | 13 (2 - 32) | 2 | 9 | - | - | 0.74 | - |
| 14 | 51 | 17.5 ± 6 | 2 | 8 (6.5 - 9) | - | - | 0.71 | 0.56 - 0.85 |
| 15 | 214 | 5 | 2 | - | - | - | 0.69 | 0.59 - 0.79 |
| 7 | 44 | 17 (13.5 - 20) | 3 | 9 (8 - 10) | - | - | 0.67 | - |
| 16 | 91 | 15 (12 - 20) | 2 | 8 (6 - 9) | - | - | 0.65 | 0.58 - 0.72 |
| 17 | 351 | 6 (13) | 2 | 10 | - | - | 0.62 | 0.54 - 0.71 |
| 18 | 102 | 9.5 (5 - 11) | 2 | 9 (7 - 10) | - | - | 0.62 | - |
| 19 | 432 | 12 (11) | 2 | 10 (1) | - | - | 0.51 | 0.41 - 0.60 |
| 20 | 103 | 15 | 2 | 10 | - | - | 0.49 | - |
| 10 | 105 | 6 (3 - 12) | 2 | - | - | - | 0.21 | - |
| K | | | | | | | | |
| 21 | 30 | 10.3 ± 5.31 | 2 | - | 6.9 | - | 0.97 | - |
| 22 | 164 | 14 | 3 | 10 | - | - | 0.34 | - |
| 23 | 62 | 14 | 2 | 10 | - | - | 0.26 | - |
| Our study | 30 | 17.5 (14.25 - 23) | 15 | 8 (6 - 9) | - | - | 0.17 | 0.12-0.24 |
| Kw | | | | | | | | |
| 3 | 37 | 16 (10 - 20) | 5 | - | 8.9 ± 1.9 | 0-3 vs 4-7 vs 8-10 | 0.61 | - |
| 11 | 74 | 17 (12 - 20) | 2 | - | 7.6 | 0-4 vs 5-7 vs 8-10 | 0.46 | 0.28 - 0.63 |
| K | | | | | | | | |
| 24 | 68 | - | 2 | 8 | - | 0-7 vs 8-10 | 0.82 | - |
| 25 | 106 | 19 (15 - 24) | 2 | 7 (7 - 8) | - | 0-7 vs 8-10 | 0.80 | - |
| 16 | 91 | 15 (12 - 20) | 2 | 8 (6 - 9) | - | 0-5 vs 6-10 | 0.76 | 0.60 - 0.91 |
| 16 | 91 | 15 (12 - 20) | 2 | 8 (6 - 9) | - | 0-7 vs 8-10 | 0.60 | 0.44 - 0.76 |
| 26 | 230 | - | 2 | - | - | 0-6 vs 7-10 | 0.60 | - |
| 27 | 155 | 17 (14 - 20) | 2 | 8 (7 - 9) | - | 0-7 vs 8-10 | 0.53 | - |
| 11 | 74 | 17 (12 - 20) | 2 | - | 7.6 | 0-7 vs 8-10 | 0.46 | 0.26 - 0.65 |
| 28 | 80 | - | 5 | - | - | 0-7 vs 8-10 | 0.34 | - |
| 29 | 34 | 5.5 (3 - 11.75) | 2 | - | - | 0-5 vs 6-10 | 0.16 | -0.21 - 0.52 |
| 29 | 34 | 5.5 (3 - 11.75) | 2 | - | - | 0-6 vs 7-10 | 0.27 | -0.05 - 0.60 |
| 29 | 34 | 5.5 (3 - 11.75) | 2 | - | - | 0-7 vs 8-10 | 0.24 | -0.07 - 0.54 |
| 30 | 62 | 15 (16) | 2 | - | - | 0-7 vs 8-10 | 0.93 | - |
| 11 | 74 | 17 (12 - 20) | 2 | - | 7.6 | 0-4 vs 5-7 vs 8-10 | 0.36 | 0.18 - 0.55 |
| Our study | 30 | 17.5 (14.25 - 23) | 15 | 8 (6 - 9) | - | 0-7 vs 8-10 | 0.52 | 0.41-0.66 |
| Our study | 30 | 17.5 (14.25 - 23) | 15 | 8 (6 - 9) | - | 0-6 vs 7-10 | 0.53 | 0.35-0.69 |



Forest plot summarizes the agreement measures retrieved in the systematic review. ASPECTS = Alberta Stroke Program Early CT Score; CI = confidence interval; ICC = intraclass correlation coefficient; IQR = interquartile range; NIHSS = NIH Stroke Scale score.

Figure 3 Graphic representation of the distribution of the Alberta Stroke Program Early CT Score (ASPECTS)



For each patient, represented on the x axis by the mean value of the ASPECTS given by all raters, the distribution of ASPECTS values given by raters is represented on the y-axis by bubbles. The bubble area is proportional to the number of raters who gave the same score (see the bubble scale). Red dots represent the ASPECTS given to each patient by the radiologist on-call. Perfect agreement would have been represented by large bubbles aligned along the diagonal formed by red dots.

variable, and agreement, with or without dichotomization using various cutoff points, was assessed using Fleiss kappa statistics with 95% bias-corrected confidence intervals obtained by 1,000 bootstrap resampling. All analyses were done with Stata version IC 12.0 and a significance level of 5%. Slight, fair, moderate, and substantial agreement categories were qualified according to Landis and Koch.¹³

RESULTS Systematic review. A total of 1,306 titles were identified; 529 were retained for full text review; 54 articles mentioned the reliability of ASPECTS. Twenty-four studies were excluded after applying selection criteria (see Methods). The flow diagram is provided in figure 1. A list of articles is provided online (appendix e-1, pages 3–4). Overall, 40 measures of agreement were extracted from a total of 30 eligible studies.

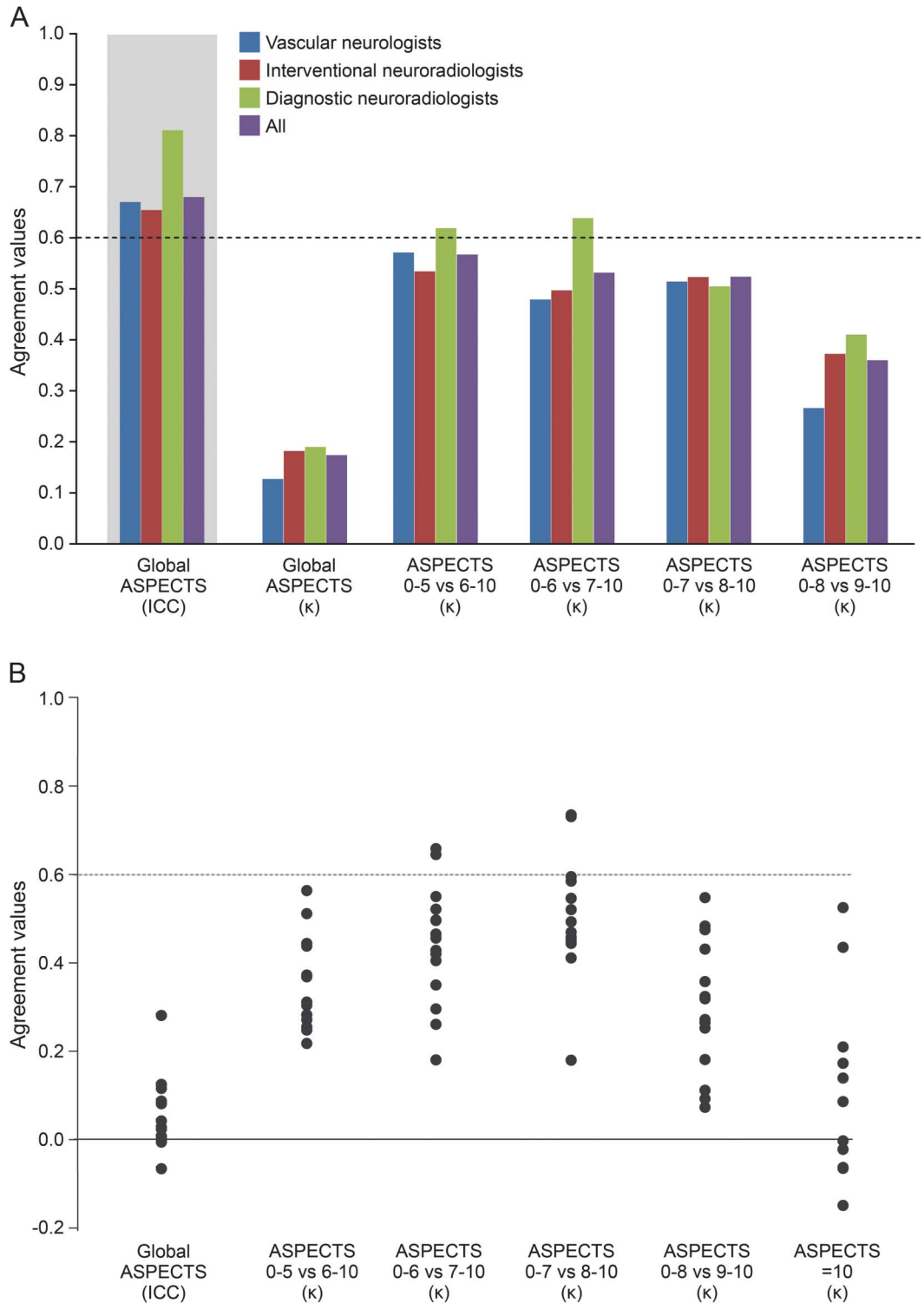
Reporting was often incomplete; methods, populations, and results varied widely (figure 2). Marked heterogeneity precluded a meta-analysis. The mean number of patients was 97.7 ± 84.8 (range 30–432). The median number of raters was 2 (interquartile range 2–3). Intrarater agreement was infrequently studied ($n = 1$). When the ASPECTS was analyzed as a continuous variable, ICCs varied between 0.57 and 0.83. When studied as a categorical variable, kappas varied between 0.26 and 0.97; weighted kappas varied between 0.21 and 0.75. When studied as a dichotomized score, kappas varied between 0.16 and 0.93 (summarized in figure 2).

Agreement study. Sixteen of the 30 stroke patients were men (53.3%; mean age = 70.5 ± 12.7 years). The median (IQR) NIHSS at admission was 17.5 (14.25–23). The mean delay between symptoms onset and CT was 190.3 ± 96.2 minutes and the median (IQR) ASPECTS (according to the official reports) was 8 (6–9). The mean ASPECTS given by the readers did not differ between the 2 sessions (6.98 ± 2.16 vs 7.05 ± 2.16), between specialties, or between junior and senior raters. The proportions (minimum–maximum according to various raters) of CTs with an ASPECTS = 10/ ≥ 9 / ≥ 8 / ≥ 7 / ≥ 6 / ≤ 5 were 10% (0%–23%)/27% (0%–47%)/49% (21%–70%)/65% (48%–77%)/77% (62%–87%)/23% (13%–38%), respectively.

The distribution of the ASPECTS given by readers for each patient is graphically displayed in figure 3. The potential clinical effect of the discrepancies in ASPECTS readers attributed to each patient by the readers is summarized in table e-1.

The ICC and kappa values of the interrater agreement for all raters, and by specialty, for the ASPECTS, and the kappa values for various dichotomies, are displayed in figure 4A (details provided in table e-2). There was no significant difference between junior and senior raters (ICC = 0.621, 95% CI [0.471–0.767] and ICC = 0.703, 95% CI [0.576–0.823], respectively). The interrater agreement was slight to moderate for all readers and each subspecialty, with no significant difference between specialties, or between junior and senior raters (see table e-2). None of the kappa values reached a substantial level (0.6) for all raters. Even in the best of

Figure 4 Results of interrater agreement study

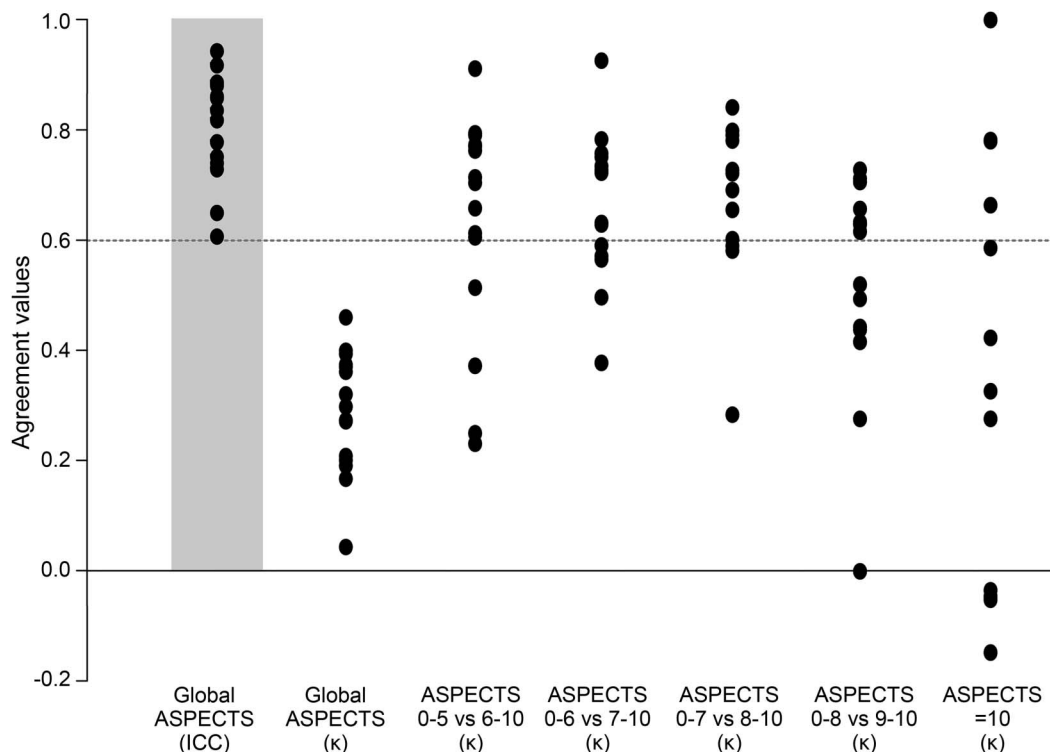


(A) Graphic display of interrater agreement. (B) Graphic display of agreement with the 16th reader (radiologist on-call), with dichotomized Alberta Stroke Program Early CT Score (ASPECTS) at various cut points. ICC = intraclass correlation coefficient.

cases, when ASPECTS was dichotomized as 0–5 vs 6–10, interrater agreement did not reach a substantial level ($\kappa = 0.561$), which translates into at least 5 of 15 raters not giving the same dichotomized verdict in 15% of patients.

Agreement with the 16th rater (the actual clinical verdict) was slight for 14/15 readers. It reached a substantial level for 2/15 raters when dichotomization (0–7 vs 8–10) was used (figure 4B). The recommendation to undergo thrombectomy, if based on

Figure 5 Results of intrarater agreement study



Graphic representation of intrarater agreement for each of the 15 readers, expressed as intraclass correlation coefficients (ICCs) or kappa values for the global Alberta Stroke Program Early CT Score (ASPECTS), and kappas obtained for dichotomized scores using various cut points.

ASPECTS ≥ 6 , would have differed from the actual verdict given for at least 5 of 15 raters, and on at least 25% of patients.

Intrarater agreement, expressed as ICCs or kappa values for ASPECTS, and kappas for various dichotomies, are summarized in figure 5. ICCs varied between 0.599 and 0.943 (mean 0.806 ± 0.1); when analyzed as a categorical variable, agreement (according to kappa) was slight to moderate for all raters. Raters scored the same ASPECTS in both sessions in 40% of cases; the same regions were identified in 26% of cases. Intrarater agreement reached a substantial level for some readers when the score was dichotomized. In practice, the same clinician would give a different dichotomized verdict when rating the same patient twice in 13%–15% of cases.

The distribution of each zone of the ASPECTS marked as positive for early ischemic changes and agreement between raters (first session) are summarized in figure e-1. Agreement was fair to moderate at best.

DISCUSSION Systematic review of the literature revealed marked heterogeneity of populations, methods, results, and interpretations, precluding a conclusion on the reliability of ASPECTS. Testing our own clinical practice involving patients recruited in

a thrombectomy trial did not show sufficient agreement between clinicians for ASPECTS to be reliably used for treatment decisions.

Heterogeneity of the literature can be explained by diverse populations, as well as testing methods of various stringencies. The number of raters was often too small to rigorously assess clinical agreement. In addition, methods varied widely.

The scoring system transforms multiple categorical decisions (presence/absence of early ischemic changes in various locations) into a numerical value. Thus, the same score may be given by 2 raters attributing changes to different regions. ASPECTS is sometimes analyzed as a continuous variable, the value being used to weigh the proportion of the brain that is already affected by ischemia (even if ASPECTS only assesses approximately 50% of the middle cerebral artery territory).¹⁴ The score can also be used as a patient comparative tool between studies or as a surrogate imaging marker in prognostic studies.¹⁵ The higher the ASPECTS, the better the prognosis.¹ Considered in this fashion, the reliability of ASPECTS has been measured with ICCs, a method that will take into account any degree of proximity between 2 scores attributed to the same patient. ICCs do not correct for chance agreement. They make ASPECTS look reliable, as exemplified in our study, where

interrater and intrarater ICCs reached substantial levels in many cases. ASPECTS can also be analyzed as a categorical variable. Expecting the same score may be too exacting, and kappa statistics will typically make ASPECTS look unreliable, as shown in the present study, in which interrater agreement was slight. ASPECTS has also been analyzed as an ordinal variable, where agreement was measured using weighted kappas, with varying weighting methods and diverse results and interpretations.^{16,17}

When used as a selection criterion for trials, or for clinical decision-making, ASPECTS is dichotomized as a categorical variable, because decisions are made for or against mechanical thrombectomy, and therefore agreement should be measured using kappa statistics. Our study showed that, if some raters can be quite consistent when reading the same CT scan twice, interrater agreement was at best moderate, and only when the score was dichotomized near the bottom of the scale (0–5 vs 6–10). Even then, a clinical decision, such as performing thrombectomy or not, if only based on ASPECTS, would differ from one clinician to the other in a substantial proportion of patients, as exemplified by the modest agreement, for any dichotomy, between raters and the actual score officially attributed by the radiologist on-call at the time of patient presentation (the 16th rater).

The detection of subtle signs of early ischemia can be challenging, even for experienced physicians.^{18,19} The modest repeatability of ASPECTS raises the question of the propriety of using the scoring system for selecting patients, as was done in recent trials,^{2–4} and as has been proposed in some recent clinical practice guidelines.^{6,7} Although it was acknowledged since its inception that “validation of this score in a randomized controlled study is needed,”¹ this has never been done.

Our study differs from most publications: many concerned a population of patients eligible for IV thrombolysis; the current study specifically concerned candidates for thrombectomy. These patients often present with more severe symptoms, higher NIHSS scores, and more frequent early ischemic signs on brain CT. We did not exclude suboptimal scans, as was often done in other publications. Raters were more numerous; they were not selected for expertise. They are the actual clinicians involved in managing patients. Many studies did not include an assessment of intrarater agreement. This is often a humbling but important learning experience. The population was artificially constructed and a different set of cases could have led to different results. The present study was limited to anterior circulation occlusion, as posterior circulation occlusions are not scored using the ASPECTS. Readings were performed in an optimal environment. The electronic form displayed a visual

reminder of the ASPECTS zones and points were added automatically. The mean reading session time was 60 minutes and raters assessed 30 cases in a row. Readers knew they were being examined, which may have led to some Hawthorne effect.²⁰ These conditions differ from the context in which decisions regarding emergency thrombectomy are made.

Generalizability of our study may be affected by the fact that all raters were from a single institution. We invite clinicians who consider using the ASPECTS scoring system for clinical decisions to test reliability in their own center.

In patients considered for thrombectomy, there may be insufficient agreement between clinicians for ASPECTS to be reliably used as a criterion for treatment decisions.

AUTHOR CONTRIBUTIONS

B. Farzin: literature search, data analysis, writing. R. Fahed: figures, study design, data collection, data analysis, writing. F. Guilbert: study design, data collection, data analysis, data interpretation. A. Poppe: data collection, data interpretation, writing. N. Daneault: data collection, data interpretation, writing. A. Durocher: data collection, data interpretation, writing. S. Lanthier: data collection, data interpretation, writing. H. Boudjani: data collection, data interpretation, writing. N. Khoury: study design, data collection, data interpretation, writing. D. Roy: data collection, data interpretation, writing. A. Weill: data collection, data interpretation, writing. J.C. Gentic: data collection, data interpretation, writing. A. Batista: data collection, data interpretation, writing. L. Letourneau-Guillon: data collection, data interpretation, writing. F. Bergeron: data collection, data interpretation, writing. M.A. Henry: data collection, data interpretation, writing. T.E. Darsaut: data analysis, data interpretation, writing. J. Raymond: literature search, study design, data analysis, data interpretation, writing.

ACKNOWLEDGMENT

Miguel Chagnon (chagnon@DMS.UMontreal.CA), statistical consultation services (SCS), Department of Mathematics and Statistics, University of Montreal, Canada, conducted all the statistical analyses.

STUDY FUNDING

Robert Fahed is the recipient of a research scholarship delivered by the Fondation pour la Recherche Médicale, Paris, France (grant DEA20140630151). The Fondation pour la Recherche Médicale had no role in this study. J. Raymond has full access to all the data in the study and had final responsibility for the decision to submit for publication.

DISCLOSURE

The authors report no disclosures relevant to the manuscript. Go to Neurology.org for full disclosures.

Received November 19, 2015. Accepted in final form February 24, 2016.

REFERENCES

1. 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;355:1670–1674.
2. Goyal M, Demchuk AM, Menon BK, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med* 2015;372:1019–1030.

3. Jovin TG, Chamorro A, Cobo E, et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med* 2015;372:2296–2306.
4. Saver JL, Goyal M, Bonafe A, et al. Stent-retriever thrombectomy after intravenous t-Pa vs. t-Pa alone in stroke. *N Engl J Med* 2015;372:2285–2295.
5. Berkhemer OA, Fransen PS, Beumer D, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med* 2015;372:11–20.
6. Casaubon LK, Boulanger JM, Blacquiere D, et al; Heart and Stroke Foundation of Canada Canadian Stroke Best Practices Advisory Committee. Canadian stroke best practice recommendations: hyperacute stroke care guidelines, update 2015. *Int J Stroke* 2015;10:924–940.
7. Powers WJ, Derdeyn CP, Biller J, et al; American Heart Association Stroke Council. 2015 American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2015;46:3020–3035.
8. Moher D, Liberati A, Tetzlaff J, Altman DG; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006–1012.
9. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.
10. Feinstein AR, Cicchetti DV. High agreement but low kappa: I: the problems of two paradoxes. *J Clin Epidemiol* 1990;43:543–549.
11. Cicchetti DV, Feinstein AR. High agreement but low kappa: II: resolving the paradoxes. *J Clin Epidemiol* 1990;43:551–558.
12. 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:Article 31.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
14. Phan TG, Donnan GA, Koga M, et al. The ASPECTS template is weighted in favor of the striatocapsular region. *Neuroimage* 2006;31:477–481.
15. Baek JH, Kim K, Lee YB, et al. Predicting stroke outcome using clinical- versus imaging-based scoring system. *J Stroke Cerebrovasc Dis* 2015;24:642–648.
16. Alexander LD, Pettersen JA, Hopyan JJ, Sahlas DJ, Black SE. Long-term prediction of functional outcome after stroke using the Alberta Stroke Program Early Computed Tomography Score in the subacute stage. *J Stroke Cerebrovasc Dis* 2012;21:737–744.
17. McTaggart RA, Jovin TG, Lansberg MG, et al. Alberta Stroke Program Early Computed Tomographic Scoring performance in a series of patients undergoing computed tomography and MRI: reader agreement, modality agreement, and outcome prediction. *Stroke* 2015;46:407–412.
18. Wardlaw JM, Dorman PJ, Lewis SC, Sandercock PA. Can stroke physicians and neuroradiologists identify signs of early cerebral infarction on CT? *J Neurol Neurosurg Psychiatry* 1999;67:651–653.
19. Dippel DW, Du Ry van Beest Holle M, van Kooten F, Koudstaal PJ. The validity and reliability of signs of early infarction on CT in acute ischaemic stroke. *Neuroradiology* 2000;42:629–633.
20. Wickstrom G, Bendix T. The “Hawthorne effect”: what did the original Hawthorne studies actually show? *Scand J Work Environ Health* 2000;26:363–367.

MACRA Changes How Neurologists Are Paid

What Is MACRA, and Why Is It Necessary?

The **Medicare Access and CHIP Reauthorization Act of 2015—MACRA**—replaced the Medicare Sustainable Growth Rate formula for calculating physician payments. The rapidly rising costs of health care in the US are unsustainable. Changes in the health care system are essential and must happen now.

How Will MACRA Affect Physician Payment?

Our health care system is moving from “fee-for-service” payments to a wider array of “value-based” payment models that put the patient at the center of care by calling for the improvement of the quality, safety, and overall experience of patient care while demonstrating cost-effectiveness by providing care that is less expensive and delivers similar or improved clinical outcomes.

Learn more at AAN.com/view/MACRA.