

ORIGINAL
RESEARCH

A.C. Gupta
P.W. Schaefer
Z.A. Chaudhry
T.M. Leslie-Mazwi
R.V. Chandra
R.G. González
J.A. Hirsch
A.J. Yoo



Interobserver Reliability of Baseline Noncontrast CT Alberta Stroke Program Early CT Score for Intra-Arterial Stroke Treatment Selection

BACKGROUND AND PURPOSE: Early ischemic changes on pretreatment NCCT quantified using ASPECTS have been demonstrated to predict outcomes after IAT. We sought to determine the interobserver reliability of ASPECTS for patients with AIS with PAO and to determine whether pretreatment ASPECTS dichotomized at 7 would demonstrate at least substantial κ agreement.

MATERIALS AND METHODS: From our prospective IAT data base, we identified consecutive patients with anterior circulation PAO who underwent IAT over a 6-year period. Only those with an evaluable pretreatment NCCT were included. ASPECTS was graded independently by 2 experienced readers. Interrater agreement was assessed for total ASPECTS, dichotomized ASPECTS (≤ 7 versus > 7), and each ASPECTS region. Statistical analysis included determination of Cohen κ coefficients and concordance correlation coefficients. PABAK coefficients were also calculated.

RESULTS: One hundred fifty-five patients met our study criteria. Median pretreatment ASPECTS was 8 (interquartile range 7–9). Interrater agreement for total ASPECTS was substantial (concordance correlation coefficient = 0.77). The mean ASPECTS difference between readers was 0.2 (95% confidence interval, –2.8 to 2.4). For dichotomized ASPECTS, there was a 76.8% (119/155) observed rate of agreement, with a moderate κ = 0.53 (PABAK = 0.54). By region, agreement was worst in the internal capsule and the cortical areas, ranging from fair to moderate. After adjusting for prevalence and bias, agreement improved to substantial or near perfect in most regions.

CONCLUSIONS: Interobserver reliability is substantial for total ASPECTS but is only moderate for ASPECTS dichotomized at 7. This may limit the utility of dichotomized ASPECTS for IAT selection.

ABBREVIATIONS: AIS = acute ischemic stroke; ASPECTS = Alberta Stroke Program Early CT Score; CI = confidence interval; IAT = intra-arterial stroke therapy; IQR = interquartile range; M1 = MCA stem; M2 = MCA second-order branch; PABAK = prevalence-adjusted, bias-adjusted kappa; PAO = proximal artery occlusion

NCCT is the most widely used imaging technique in the evaluation of acute stroke because of its availability, speed, and accuracy for ruling out intracranial hemorrhage.^{1,2} In patients with AIS, early reperfusion therapy improves outcomes in patients with small infarcts^{3,4} but is harmful in those with extensive injury (eg, greater than 33% of the MCA territory).⁵ Unfortunately, during the first several hours of acute ischemic stroke, signs of tissue injury on NCCT are subtle, and interobserver agreement is limited.⁶ ASPECTS is a semiquantitative grading system developed to improve the assessment of early ischemic changes in the MCA territory on NCCT.⁷

In recent studies, baseline (pretreatment) dichotomized ASPECTS has been demonstrated to predict good outcomes after IAT.^{8–10} Yet before this approach is used clinically to select patients for IAT, it should demonstrate good interobserver agreement. The literature is conflicting in this regard, with κ values ranging from fair to substantial.^{7,11} However, these studies included a broad range of stroke severities and may not necessarily apply to the population eligible for IAT, namely, patients with proximal cerebral artery occlusions.

We sought to evaluate the reliability of baseline ASPECTS,

specifically in patients with anterior circulation PAO undergoing IAT, and how this would affect treatment decision-making for IAT. Specifically, we aimed to determine whether pretreatment ASPECTS dichotomized at 7 would demonstrate at least substantial κ agreement. We also examined the interobserver reliability of ASPECTS by topographic region.

Materials and Methods

From our prospective observational IAT data base, we identified 169 consecutive patients with AIS with anterior circulation PAO (terminal ICA and/or MCA M1 or M2 segment occlusion) who underwent intraarterial treatment between June 2004 and May 2010. General inclusion criteria for IAT included the following: 1) imaging assessment completed < 8 hours post ictus; 2) NCCT without intracranial hemorrhage; 3) NCCT hypodense ischemic lesion or MRI DWI hyperintense ischemic lesion $< 1/3$ MCA territory; 4) proximal artery occlusion on CT angiography; 5) NIHSS score > 8 or with significant aphasia; 6) if IV rtPA was administered, no treatment response. Additional inclusion criteria for this study included an evaluable pretreatment noncontrast head CT scan. Of the 169 patients, 14 were excluded due to the presence of chronic infarcts ipsilateral to the acute stroke or lack of pretreatment NCCT imaging at our hospital. Clinical and imaging data were retrospectively analyzed. This study was approved by our local institutional review board and was Health Insurance Portability and Accountability Act compliant.

Imaging Protocol and Analysis

NCCT scans of the head were performed in helical mode (1.25-mm thickness, kV 120, mA 250) and reconstructed in the axial plane with 5-mm section thickness. This imaging was evaluated independently

Received September 7, 2011; accepted after revision October 2.

From Division of Diagnostic (A.C.G., P.W.S., Z.A.C., R.G.G., A.J.Y.) and Interventional (A.C.G., Z.A.C., T.M.L.-M., R.V.C., J.A.H., A.J.Y.) Neuroradiology, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts.

Paper previously presented at: Annual Meeting of the Radiological Society of North America, November 27–December 2, 2011; Chicago, Illinois.

Please address correspondence to Albert J. Yoo, MD, 55 Fruit St, GRB241, Boston, MA 02114; e-mail: ajyoo@partners.org

<http://dx.doi.org/10.3174/ajnr.A2942>

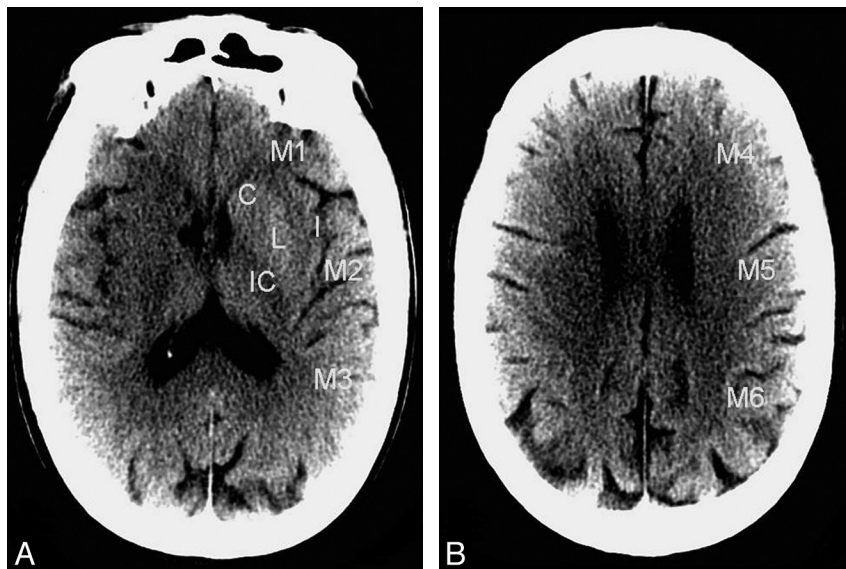


Fig 1. ASPECTS regions on noncontrast CT. A, Ganglionic level; B, supraganglionic level. For each CT scan, all of the images were evaluated for ASPECTS. C indicates caudate nucleus; I, insular ribbon; IC, internal capsule; L, lentiform nucleus; M1–6, cortical/subcortical territories of the MCA.

by 2 experienced neuroradiologists using the ASPECTS system. Contrary to the original scoring system, which utilized only 2 brain sections,⁷ readers graded early ischemic change in each of the 10 ASPECTS regions (Fig 1) according to current methodology, which utilizes all of the scan images.¹² Early ischemic change was defined as tissue hypodensity or loss of gray–white matter differentiation, as these changes have been associated with edema and irreversible injury. Isolated cortical swelling, which was part of the original ASPECTS criteria, was not included, as recent work has demonstrated that it is associated with increased cerebral blood volume and may represent penumbral (threatened but salvageable) tissue.¹³ Discrepancies between readers were resolved by consensus adjudication. Window and level settings were adjusted at the discretion of the readers to increase contrast between normal and ischemic brain. ASPECTS was calculated by subtracting the number of affected regions from a total possible score of 10. Imaging review was performed blinded to all clinical information except stroke side.

Statistical Analysis

Intrater agreement was measured for total ASPECTS, dichotomized ASPECTS (≤ 7 versus > 7), and each individual topographic region. ASPECTS was dichotomized at 7, given recent evidence supporting this threshold for identifying good clinical response to intra-arterial treatment or reperfusion.^{8–10} For total ASPECTS, agreement was measured by using the concordance correlation coefficient. In addition, Bland and Altman¹⁴ analysis was performed to assess the absolute degree of interrater differences across the entire ASPECTS scale. For dichotomized ASPECTS and for each ASPECTS region, Cohen κ coefficients were calculated. In addition, because κ scores are influenced by prevalence and bias, PABAK scores were calculated and reported with their associated prevalence and bias indices.^{15,16} Interpretation of the κ values followed the proposed standards of Landis and Koch: 0–.20 (slight); 0.21–0.40 (fair); 0.41–0.60 (moderate); 0.61–0.80 (substantial); and 0.81–1.00 (almost perfect).¹⁷ Continuous data with normal distribution were reported as mean \pm standard deviation, ordinal or non-normal data were reported as median (IQR), and categorical data were reported as proportions. Normality was tested by using the Kolmogorov-Smirnov test. Statistical significance was taken at $P < .05$. Statistical analysis was performed by using MedCalc, version 11.2.1 (MedCalc Software, Mariakerke, Belgium).

Histogram of consensus ASPECTS

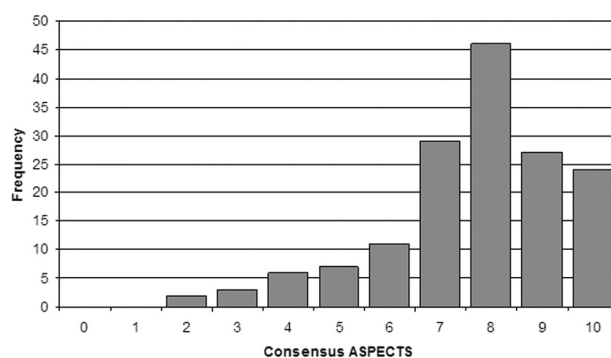


Fig 2. Histogram illustrates the frequency distribution of consensus ASPECTS within the patient population.

Results

A total of 155 patients met study criteria. The cohort included 80 female patients, with an average age of 67.7 ± 17.2 years. Seventy strokes involved the right hemisphere. The median baseline NIHSS score was 17 (IQR 14–20). There were 140 isolated intracranial occlusions: 33 occlusions of the terminal ICA with or without extension into the M1 segment, 92 M1 MCA occlusions, and 15 M2 MCA occlusions. There were 15 tandem occlusions of the cervical carotid and intracranial ICA and/or MCA. The mean duration from stroke onset to NCCT imaging was 199 ± 130 minutes.

The distribution of consensus ASPECTS within the cohort was skewed toward higher ASPECTS (smaller infarcts), with 97 patients demonstrating ASPECTS > 7 (Fig 2). The median baseline ASPECTS was 8 (IQR 7–9). There was substantial interrater agreement for total ASPECTS grading, with a concordance correlation coefficient of 0.77 (95% CI, 0.70–0.83). In the Bland-Altman analysis (Fig 3), there was a small difference between the mean ASPECTS (7.4 versus 7.6; $P < .05$). The limits of agreement (95% CI) for the ASPECTS score differences ranged from -2.8 to $+2.4$. ASPECTS was the same between readers in 34.2% (53/155), within 1 point in 76.8% (119/155), and within 2 points in 91.6% (142/155). There were no differences greater than 3 points.

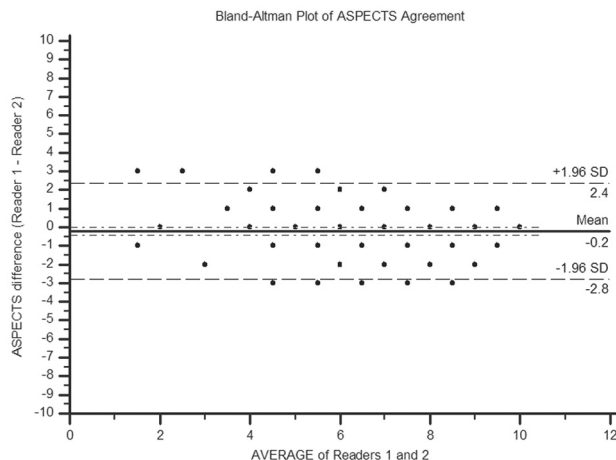


Fig 3. Bland-Altman plot demonstrates the distribution of ASPECTS differences between the 2 readers in this study.

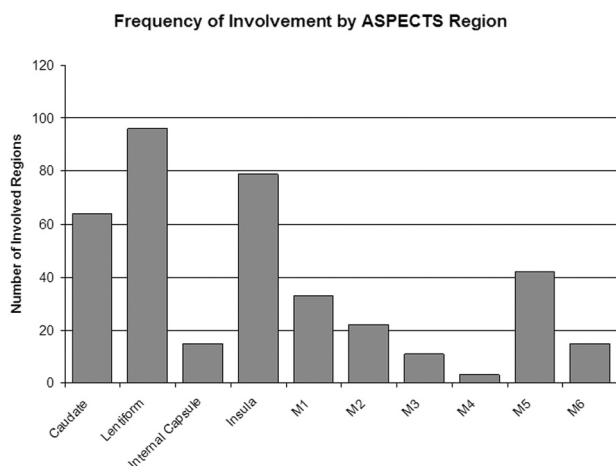


Fig 4. Histogram reveals the prevalence of ischemic involvement by ASPECTS region in our patient cohort.

For dichotomized baseline ASPECTS (≤ 7 versus > 7), there was a 76.8% (119/155) observed rate of agreement. Accounting for chance, there was moderate agreement, with $\kappa = 0.53$ (PABAK = 0.54; bias index = 0.04; prevalence index = 0.11).

Fig 4 illustrates the prevalence of ischemic involvement for each ASPECTS region using the consensus reads. The basal ganglia and insula were the most frequently affected regions. The table provides the κ and PABAK values for the individual ASPECTS regions. Interrater agreement was variable, with the worst performance in the internal capsule and the cortical regions. The internal capsule and M6 region demonstrated only fair agreement. When accounting for bias and prevalence, agreement improved to at least substantial in all regions except the lentiform nucleus (moderate).

Discussion

Among highly experienced readers, there was substantial interobserver agreement for grading baseline ASPECTS on acute noncontrast head CT in the setting of anterior circulation proximal artery occlusion. Most differences in ASPECTS were within 1–2 points in this study. However, only moderate interrater agreement was achieved for assessing dichotomized

Interobserver agreement for individual ASPECTS regions				
Region	Cohen κ (95%CI)	PABAK	Prevalence Index	Bias Index
Caudate	0.713 (0.602, 0.824)	0.716	0.110	0.026
Lentiform	0.530 (0.395, 0.664)	0.535	0.200	0.168
Internal capsule	0.280 (0.028, 0.533)	0.665	0.742	0.129
Insula	0.731 (0.624, 0.838)	0.729	0.026	0.084
M1	0.592 (0.434, 0.750)	0.716	0.561	0.103
M2	0.416 (0.184, 0.648)	0.729	0.735	0.071
M3	0.436 (0.143, 0.730)	0.832	0.839	0.032
M4	0.433 (–0.056, 0.922)	0.935	0.942	0.032
M5	0.508 (0.347, 0.669)	0.626	0.490	0.032
M6	0.383 (0.115, 0.651)	0.768	0.794	0.077

ASPECTS > 7 versus ≤ 7 . In approximately 25% of our cohort, treatment decisions for IAT based on dichotomized ASPECTS would have been affected by the interrater reliability.

Our findings with respect to total ASPECTS scoring support a previous study that also demonstrated substantial agreement (weighted $\kappa = 0.69$) across the entire ASPECTS scale.¹⁸ In that study, the authors found a mean interrater difference of zero, with a standard deviation of 1.1 points. Similarly, we found a mean difference of 0.2, with 76.8% of scores being within 1 point.

Previous studies have shown that baseline ASPECTS > 7 identifies patients who are more likely to benefit from endovascular treatment.^{8–10} However, based on our findings, the utility of this approach may be adversely affected by its limited interrater reliability. The literature is conflicting with respect to this question. The original ASPECTS article by Barber et al⁷ reported substantial to near-perfect agreement for ASPECTS dichotomized at 7, with κ values ranging from 0.71 to 0.89. However, another study by Mak et al¹¹ demonstrated a κ score of 0.34 (fair), with a strikingly low 42% rate of observed agreement. When adjusting for prevalence and bias, agreement improved to moderate (PABAK = 0.44). Our results confirm only a moderate level of reliability ($\kappa = 0.53$, PABAK = 0.54) for dichotomized ASPECTS in the setting of proximal occlusions.

It is unclear exactly why the interrater agreement for dichotomized ASPECTS was so different between the study by Barber et al⁷ and ours. However, this difference is probably related to variable selection bias and consequently differing patient cohorts. Timing of imaging is 1 factor that can influence lesion detection. In their study, patients underwent baseline NCCT within 3 hours of stroke onset and were treated with intravenous alteplase, while in the present study, at least half of the patients were imaged beyond 3 hours (median 188.5 minutes, IQR 83–278 minutes) and were treated with IAT. However, the increased time to imaging in our study would be expected to yield more conspicuous lesions and hence a better agreement rather than a worse one. Another possible explanation for the varying agreement is a difference in ASPECTS distribution. Although the median baseline ASPECTS was 8 in both studies, a broader distribution (ie, larger variance) in 1 study would result in a smaller proportion of scores around the threshold of 7 and therefore fewer potential discrepancies. Unfortunately, Barber et al⁷ did not report the proportions of individual scores in their study. In our cohort, almost half

(48.4% [75/155]) of the consensus scores were 7 or 8 and thus prone to interrater disagreement.

This last issue underscores the problem with using a binary classification scheme. Clearly, dichotomization will penalize scores that are discrepant by only 1 point if they are around the threshold value. Therefore, despite our excellent agreement for total ASPECTS, we achieved only moderate agreement for dichotomized ASPECTS because of the clustering of scores around 7. This limitation may particularly affect favorable IAT candidates (ie, those who present with minimal tissue injury). If a small infarct is present in the setting of a proximal occlusion, it typically involves the caudate nucleus, lentiform, and/or insula, which are end-artery territories. As such, the baseline ASPECTS is often 7 or 8 in these patients. Also in such patients, the moderate level of agreement that we observed for the lentiform nucleus may further contribute to the relatively low rate of agreement for dichotomized ASPECTS.

Our findings with respect to individual ASPECTS regions are novel, as no study, to our knowledge, has reported interobserver agreement by region. Notably, there was only fair to moderate agreement in the internal capsule and the cortical regions. However, this was mostly related to the low prevalence of ischemic involvement in our study. When accounting for prevalence and bias, these areas demonstrated substantial or near-perfect agreement, suggesting that there was strong consensus between the readers as to what was considered abnormal. It is somewhat surprising that the agreement for total ASPECTS was high despite the low level of agreement in these multiple regions. This suggests that overall agreement is determined largely by agreement within the more prevalent ASPECTS regions (eg, basal ganglia, insula).

This study is limited primarily by our highly selected patient population, all of whom underwent IAT. This explains why our ASPECTS distribution was skewed toward smaller infarcts, which are thought to be more likely to benefit from treatment. However, it is in this specific patient subset where interobserver agreement is most important. Moreover, for the analysis of dichotomized ASPECTS, and of the individual ASPECTS regions, we were able to mitigate this issue by adjusting for prevalence. Another limitation is that we did not assess intraobserver reliability. However, we believe that interobserver agreement is more relevant with respect to the clinical utility of ASPECTS. In order for NCCT ASPECTS to be widely adopted for selecting patients for IAT, it must have good reliability between readers at different medical centers. A further limitation is that our findings cannot be extended to general clinical practice, where stroke-imaging studies are often interpreted by physicians who are not expert at ASPECTS grading. Previous studies have shown that the accuracy of ASPECTS evaluation is dependent on the level of rater experience.^{7,19,20} However, we wanted to assess the reliability of ASPECTS under the most ideal conditions (ie, using readers with significant experience with stroke imaging and ASPECTS grading). In support of this last point, a recent study by Wardlaw et al²¹ found that the reliability of grading acute ischemia on NCCT (versus an expert neuroradiologist) was improved if the reader was a neuroradiologist and if the scan was acquired later in the 6-hour window. In our study, both readers were neuroradiologists and at least half of the scans were acquired after 3 hours. Future studies are needed to verify our findings

in an unselected population of patients with AIS with proximal occlusions.

Conclusions

Among patients with anterior circulation proximal artery occlusions who are eligible for intra-arterial therapy, interrater reliability for ASPECTS grading is substantial across the entire ASPECTS scale. However, when using dichotomized ASPECTS (≤ 7 versus > 7) for treatment selection, agreement is only moderate, which limits the utility of this approach. In our cohort, approximately 25% of treatment decisions would be affected by interrater reliability.

Disclosures: Joshua Hirsch—UNRELATED: Consultancy: Phillips, CareFusion, Comments: Participated in focus group of NI specialists for NI suites of the future (Phillips); participated on NextGen team for product development (CareFusion); Royalties: CareFusion, Comments: as above; Stock/Stock Options: IntraTech, Nevro, NFocus, Comments: Development stage companies that are working on NI products (IntraTech and N-Focus) and nerve imaging (Nevro). Albert Yoo—UNRELATED: Grants/Grants Pending: Penumbra* * Money paid to institution

References

1. Effect of intravenous recombinant tissue plasminogen activator on ischemic stroke lesion size measured by computed tomography. NINDS; The National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study Group. *Stroke* 2000;31:2912–19
2. Wardlaw JM. Should all patients with suspected stroke have brain MRI instead of CT? *Nat Clin Pract Cardiovasc Med* 2007;4:362–63
3. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *N Engl J Med* 1995;333:1581–87
4. Furlan A, Higashida R, Wechsler L, et al. Intra-arterial prourokinase for acute ischemic stroke. The PROACT II study: a randomized controlled trial. *Prolyse in Acute Cerebral Thromboembolism*. *JAMA* 1999;282:2003–11
5. von Kummer R, Allen KL, Holle R, et al. Acute stroke: usefulness of early CT findings before thrombolytic therapy. *Radiology* 1997;205:327–33
6. Wardlaw JM, Mielke O. Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment—systematic review. *Radiology* 2005;235:444–53
7. Barber PA, Demchuk AM, Zhang J, et al. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet* 2000;355:1670–74
8. Goyal M, Menon BK, Coutts SB, et al. Effect of baseline CT scan appearance and time to recanalization on clinical outcomes in endovascular thrombectomy of acute ischemic strokes. *Stroke* 2011;42:93–97
9. Hill MD, Rowley HA, Adler F, et al. Selection of acute ischemic stroke patients for intra-arterial thrombolysis with pro-urokinase by using ASPECTS. *Stroke* 2003;34:1925–31
10. Hill MD, Demchuk AM, Tomsick TA, et al. Using the baseline CT scan to select acute stroke patients for IV-IA therapy. *AJNR Am J Neuroradiol* 2006;27:1612–16
11. Mak HK, Yau KK, Khong PL, et al. Hypodensity of $>1/3$ middle cerebral artery territory versus Alberta Stroke Programme Early CT Score (ASPECTS): comparison of two methods of quantitative evaluation of early CT changes in hyperacute ischemic stroke in the community setting. *Stroke* 2003;34:1194–96
12. Puetz V, Dzialowski I, Hill MD, et al. The Alberta Stroke Program Early CT Score in clinical practice: what have we learned? *Int J Stroke* 2009;4:354–64
13. Butcher KS, Lee SB, Parsons MW, et al. Differential prognosis of isolated cortical swelling and hypoattenuation on CT in acute stroke. *Stroke* 2007;38:941–47
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10
15. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423–29
16. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257–68
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74
18. Coutts SB, Demchuk AM, Barber PA, et al. Interobserver variation of ASPECTS in real time. *Stroke* 2004;35:e103–05
19. Pexman JH, Barber PA, Hill MD, et al. 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
20. Coutts SB, Hill MD, Demchuk AM, et al. ASPECTS reading requires training and experience. *Stroke* 2003;34:e179; author reply e179
21. Wardlaw JM, von Kummer R, Farrall AJ, et al. A large web-based observer reliability study of early ischaemic signs on computed tomography. The Acute Cerebral CT Evaluation of Stroke Study (ACCESS). *PLoS One* 2010;5:e15757