Alberta Stroke Program Early CT Score (ASPECTS) in Real Time: A Prospective Study of its Interobserver Variation among Radiologists in a Training Hospital from a Low-resource Setting

Scott Riley K. Ong, MD,[1] Adovich S. Rivera, MD[2] and Ryan Jason DL. Urgel, MD[1]

[1]Department of Radiology, Philippine General Hospital, University of the Philippines Manila
[2]Institute for Public Health and Medicine, Feinberg School of Medicine, Northwestern University, Chicago, United States of America

ABSTRACT

Background and Objectives. The Alberta Stroke Program Early CT Score (ASPECTS) is a standardized system used to quantify the extent of ischemic involvement in cases of acute middle cerebral artery (MCA) infarct. It aids in clinical decision-making to identify patients who are more likely to benefit from thrombolytic therapy. This study aimed to determine the interobserver variation of ASPECTS among training and expert radiologists in a real-time, low-resource setting.

Patients and Methods. A prospective study was conducted on non-enhanced CT (NECT) images of 79 patients with acute stroke. Patients with hemorrhagic stroke, or ischemic stroke from territories other than the MCA, were excluded. The ASPECTS of each case was assessed by three groups of radiologists—residents, fellows, and an expert. The level of agreement among them was then analyzed.

Results. ASPECT scores were dichotomized into >7 and ≤7. With the expert's reading as gold standard, residents had sensitivity of 0.94 [95% CI: 0.85, 0.99] and specificity of 0.68 [95% CI: 0.46, 0.85], while fellows had sensitivity of 0.96 [95% CI: 0.87, 1.00] and specificity of 0.76 [95% CI: 0.55, 0.91]. There was substantial agreement between residents and expert in overall ASPECTS rating (K = 0.66 [95% CI: 0.48, 0.85]; AC1 = 0.77 [95% CI: 0.62, 0.91]), and substantial to almost perfect agreement between fellows and expert (K = 0.76 [95% CI: 0.59, 0.92]; AC1 = 0.83 [95% CI: 0.71, 0.95]). Differences between the ASPECTS of the expert and trainees were within 2 points in most cases.

Conclusion. ASPECTS is a reliable tool for both training and expert radiologists to quantify the extent of MCA infarcts. Assessment by trainees is comparable with that of the expert reader and is useful for immediate clinical decision-making in low-resource settings.

Key Words: ASPECTS, stroke, middle cerebral artery infarct

INTRODUCTION

Stroke is the second leading cause of mortality globally, and its incidence continues to rise in low- and middle-income countries.[1] It is typically classified as either ischemic or hemorrhagic, with ischemic stroke comprising 70-80% of cases.[2]

For ischemic stroke patients, intravenous thrombolysis within 3 to 4.5 hours from ictus has been shown to effectively increase patient survival and improve functional outcomes.[3-5] Non-enhanced CT (NECT) is the initial imaging modality used in most protocols to triage patients into the appropriate management and treatment plan.[2] The use of

Corresponding author: Scott Riley K. Ong, MD
Department of Radiology
Philippine General Hospital
University of the Philippines Manila
Taft Avenue, Manila 1000, Philippines
Email: scottrileyong@yahoo.com

NECT allows early and accurate detection of intracranial hemorrhage, the presence of which would immediately exclude a patient from thrombolytic therapy. In cases of ischemic stroke, NECT also allows initial evaluation of the extent of infarction. Involvement of more than one-third of the middle cerebral artery (MCA) territory is another exclusion criterion for thrombolytic therapy.[2,3]

Although NECT is highly sensitive in the detection of intracranial hemorrhage, its reliability in detecting the early changes and extent of ischemic stroke has been a subject of debate. Studies have shown that even experienced stroke clinicians and radiologists have difficulty in quantifying whether an MCA infarct involves more or less than one-third of its territory, which in turn determines whether a patient will undergo thrombolysis or not.[3,6,7] For this reason, the Alberta Stroke Program Early CT Score (ASPECTS) was developed to provide a reproducible method of quantifying early ischemic changes seen on NECT in cases of acute MCA infarcts.[3]

ASPECTS is a standardized scoring system that utilizes a 10-point scale, corresponding to ten anatomically defined regions supplied by the MCA. Four of these represent subcortical structures—the caudate head, lentiform nucleus, internal capsule, and insular ribbon—while six correspond to cortical structures—the anterior MCA cortex (M1), MCA cortex lateral to the insular ribbon (M2), posterior MCA cortex (M3), and the MCA territories immediately superior to M1, M2, and M3 (M4, M5, M6). The score is determined from two standard axial CT images—one at the level of the thalamus and basal ganglia, and one immediately superior to these ganglionic structures. One point is deducted for every involved region, such that a score of 10 indicates a normal scan and a score of 0 indicates ischemia of all the defined territories of the MCA.[3,6,7]

In the pilot study by the ASPECTS study group,[3] ASPECTS value was shown to correlate inversely with the severity of stroke (r = -0.56, p < 0.001). Furthermore, it was shown to be a good predictor of functional outcome (p < 0.001) and symptomatic intracerebral hemorrhage (p = 0.012) following thrombolytic therapy. Patients with lower scores who were treated with the thrombolytic agent Alteplase had higher rates of morbidity and mortality. It was proposed that patients with ASPECTS value of 7 or less would benefit less from thrombolysis due to higher probabilities of functional dependence and hemorrhagic complication.

The reliability of ASPECTS in screening acute stroke patients for thrombolysis is of utmost importance for training institutions in low-resource settings, where a neuroradiologist may not be available round-the-clock and where trainees usually provide the initial interpretation of a CT study for immediate use in treatment decision-making. Studies from high-income countries demonstrated acceptable inter-observer reliability of ASPECTS among stroke neurologists and neuroradiologists, and that it had better reliability compared with the traditional 1/3 MCA rule.[3,4,7]

A previous retrospective study done locally reported substantial agreement between trainees and expert neuroradiologist.[8] This study, however, was conducted in a controlled environment, where readers were given ample time to review the scans and were relatively free from stressors. Our study thus aimed to determine the interobserver variation of ASPECTS among training and expert radiologists prospectively in a real-time setting. It is important to determine real-time agreement to ensure the clinical utility of ASPECTS. Anatomic areas that are common sources of disagreement were also identified so that better attention can be directed to these during the review of scans.

PATIENTS AND METHODS

A prospective study was conducted on NECT images of patients clinically diagnosed with acute stroke in the Philippine General Hospital from July to September 2019. The study protocol was adapted from Coutts et al. with minor modifications.[4]

NECT scans were obtained from patients with lateralizing stroke-like symptoms—specifically, hemiparesis and/or aphasia lasting for more than 5 minutes—as assessed by physicians from the emergency or in-patient departments. Patients should have been functionally independent before the ictus, be scanned within 12 hours from ictus, and be older than 18 years. NECT scans showing hemorrhagic stroke, or ischemic stroke from vascular territories other than the MCA, were excluded.

A minimum sample size of 72 patients was calculated to detect a difference of 0.3 or higher between two groups of raters at an alpha of 0.05, power of 80%, and assuming that equal true marginal rating frequencies are the same.[9] All patients referred for stroke imaging were assessed for eligibility according to the study's inclusion and exclusion criteria and were included in the study until the target sample size was reached.

NECT images were acquired using the 16-slice or 64-slice CT scan machines (GE Healthcare Systems, Illinois, Chicago; Philips Healthcare, Amsterdam, Netherlands) of the institution, and were viewed using the software Philips Intellispace Portal. All residents-in-training and fellows, as well as an expert radiologist, from the CT section of the institution's Department of Radiology, participated in the study.

The images were initially viewed and scored for ASPECTS by the designated resident on-duty during the acquisition period. This was followed by review and scoring by the fellow on-deck. Lastly, the images were reviewed and rated by the expert radiologist, whose reading served as the gold standard for this study. All participants were allowed to know the demographics, clinical history, and symptoms of the patients. They were also allowed to manipulate the window level and width of the images in the viewing station. Due to the nature of workflow in the study site, blinding

of the participating radiologists was not entirely possible, but participants were not allowed to consult each other for their ASPECTS ratings.

Each radiologist evaluated the involvement of each of the ten defined MCA regions used in ASPECTS. Individual findings for each case were recorded in a standardized data collection form as either positive or negative for ischemic involvement, along with the final ASPECTS value. Signs for positive involvement included parenchymal hypo-attenuation and focal brain swelling. At the end of the data collection period, data were manually encoded in an electronic spreadsheet file by the study investigators. The findings and ASPECTS value for each case, as assessed by the resident, fellow, and expert reader, were then analyzed.

Statistical Analysis

Descriptive statistics of the patients were calculated in terms of demographic and clinical variables. Total ASPECT scores were computed and dichotomized (>7 vs ≤7). Readers were classified as resident, fellow, or expert radiologist. The pairwise agreement of the different reader groups was calculated using simple agreement and Kappa coefficients. An alternative agreement measure, the AC1 index, was also used in case when paradoxical agreement (e.g., high simple agreement but low Cohen's Kappa) was detected. Per item level agreement was done to determine any differences in agreement at the item level. The correlation and agreement of the total ASPECT scores of different readers were assessed through intraclass correlation.

Agreement was classified as no to slight agreement (0.00 to 0.20), fair agreement (0.21 to 0.40), moderate agreement (0.41 to 0.60), substantial agreement (0.61 to 0.80), or almost perfect agreement (0.81 to 1.00).[10]

Ethical Considerations

The protocol for this study was reviewed and approved by the University of the Philippines Manila Research Ethics Board (UPMREB). The study was conducted in compliance with the Helsinki Declaration, the Data Privacy Act of 2012, and the 2017 National Ethical Guidelines for Privacy and Health-Related Research.

Due to the nature of the study, the identities of the patients were known by the interpreting radiologists. However, their identities were concealed through the use of identification numbers when study-related findings were recorded in the data collection forms. Data were used only in the analysis of the study and were accessible only to the study investigators.

Informed consent was obtained from the participating radiologists before the start of the study. Their participation was voluntary, and they received no incentives for participating. Patients were classified as secondary subjects in this study. Since this study involved image interpretation and there was no direct interaction between the study investigators/participants and the patients, the need for informed consent from patients was waived. Patients received no more than minimal risk both from this study and from the routine imaging procedure which they would receive as part of their management.

RESULTS

Patient Demographics

A total of 79 patients with lateralizing stroke-like symptoms were included in this study. Their ages ranged from 23 to 94 years, with a mean of 58 (± 14) years. Thirty-six (45.6%) were males, and forty-three (54.4%) were females. Presenting symptoms were unilateral weakness in 73 (92.4%) patients, facial asymmetry in 57 (72.2%), and slurring of speech in 26 (32.9%). CT scan was done using the 64-slice machines of the institution in 77 (97.5%) patients, while the 16-slice machine was used in 2 (2.5%) patients.

Data and Outcomes

Eighteen residents (median: 4 cases per resident), five fellows (median: 15 cases per fellow), and one expert radiologist participated in the study. All residents had at least 3 months of training in CT scan. The ASPECT scores were dichotomized into >7 and ≤7, representing patients who were predicted to benefit from intravenous thrombolysis and those who were less likely to benefit from the treatment, respectively.

Table 1 shows a comparison of the ASPECTS ratings across the three groups of readers. Evaluation by the expert yielded the lowest mean ASPECT score (8.15), while evaluation by the residents resulted in the highest mean ASPECT score (8.25). Ratings performed by residents and fellows were compared with those done by the expert reader as the gold standard. Computed sensitivity and specificity reflecting the diagnostic performance of residents were 0.94 [95% CI: 0.85, 0.99] and 0.68 [95% CI: 0.46, 0.85], respectively. Meanwhile, sensitivity and specificity of fellows were 0.96 [95% CI: 0.87, 1.00] and 0.76 [95% CI: 0.55, 0.91], respectively.

Table 1. Comparison of ASPECTS rating among expert, fellow, and resident radiologists

	Mean ASPECTS (± SD)	Number of patients with ASPECTS >7
Expert	8.15 (± 2.46)	54 (68.4%)
Fellows	8.19 (± 2.32)	58 (73.4%)
Residents	8.25 (± 2.66)	59 (74.7%)

Interobserver Agreement

Tables 2, 3, and 4 show the agreement coefficients between residents and expert, between fellows and expert, and among the three groups, respectively. Percent agreement was consistently high (>0.80) across all MCA regions. Meanwhile, agreement between residents and expert, and between fellows and expert, ranged from moderate to almost perfect in terms of Cohen's Kappa and AC1 index.

In terms of overall ASPECTS rating, there was substantial agreement between residents and expert, and substantial to almost perfect agreement between fellows and expert. Three-way correlation among the three groups showed overall substantial agreement with Fleiss' Kappa and almost perfect agreement with the AC1 index. Intraclass correlation was good, with a computed value of 0.88 [95% CI: 0.83, 0.92]. Bland-Altman plots (Figures 1 and 2) showed that most ASPECT scores from the residents and fellows are within ±2 of those from the expert.

DISCUSSION

Trainees were slightly more likely to give ASPECT scores of >7. In terms of diagnostic performance, both residents and fellows demonstrated high sensitivities but moderate specificities. These suggest that patients rated as ASPECTS ≤7 by trainees may be confidently excluded as good candidates for thrombolytic treatment. However, there is less confidence in their assessment for patients rated as ASPECTS >7, especially when the possibility of missed diagnosis or underdiagnosis is considered.

Table 2. Agreement statistics between residents and expert reader for ASPECT scoring in the various middle cerebral artery regions

Region	Percent Agreement	Cohen's Kappa	AC1 index
Caudate head	0.96 [0.92, 1.00]	0.80 [0.58, 1.00]	0.95 [0.90, 1.00]
Lentiform nucleus	0.86 [0.78, 0.94]	0.60 [0.38, 0.81]	0.79 [0.66, 0.92]
Internal capsule	0.87 [0.80, 0.95]	0.59 [0.36, 0.82]	0.82 [0.70, 0.94]
Insular ribbon	0.85 [0.77, 0.93]	0.63 [0.44, 0.82]	0.74 [0.59, 0.89]
M1	0.90 [0.83, 0.97]	0.58 [0.32, 0.84]	0.87 [0.77, 0.96]
M2	0.92 [0.86, 0.98]	0.74 [0.54, 0.94]	0.89 [0.80, 0.98]
M3	0.92 [0.86, 0.98]	0.66 [0.40, 0.92]	0.90 [0.82, 0.99]
M4	0.85 [0.77, 0.93]	0.45 [0.19, 0.71]	0.79 [0.66, 0.92]
M5	0.87 [0.80, 0.95]	0.64 [0.43, 0.85]	0.80 [0.68, 0.93]
M6	0.90 [0.83, 0.97]	0.61 [0.36, 0.86]	0.86 [0.76, 0.96]
ASPECTS >7 versus ≤7	0.86 [0.78, 0.94]	0.66 [0.48, 0.85]	0.77 [0.62, 0.91]

Table 3. Agreement statistics between fellows and expert reader for ASPECT scoring in the various middle cerebral artery regions

Region	Percent Agreement	Cohen's Kappa	AC1 index
Caudate head	0.96 [0.92, 1.00]	0.82 [0.62, 1.00]	0.95 [0.90, 1.00]
Lentiform nucleus	0.85 [0.77, 0.93]	0.58 [0.37, 0.80]	0.76 [0.62, 0.90]
Internal capsule	0.90 [0.83, 0.97]	0.65 [0.43, 0.88]	0.86 [0.75, 0.96]
Insular ribbon	0.87 [0.80, 0.95]	0.71 [0.54, 0.88]	0.78 [0.64, 0.92]
M1	0.87 [0.80, 0.95]	0.58 [0.35, 0.80]	0.82 [0.70, 0.94]
M2	0.91 [0.85, 0.98]	0.69 [0.48, 0.90]	0.88 [0.78, 0.97]
M3	0.94 [0.88, 0.99]	0.67 [0.40, 0.94]	0.92 [0.85, 0.99]
M4	0.86 [0.78, 0.94]	0.54 [0.29, 0.78]	0.80 [0.68, 0.93]
M5	0.82 [0.74, 0.91]	0.48 [0.24, 0.71]	0.73 [0.59, 0.88]
M6	0.91 [0.85, 0.98]	0.58 [0.30, 0.86]	0.89 [0.80, 0.98]
ASPECTS >7 versus ≤7	0.90 [0.83, 0.97]	0.76 [0.59, 0.92]	0.83 [0.71, 0.95]

Table 4. Agreement statistics among residents, fellows, and expert reader for ASPECT scoring in the various middle cerebral artery regions

Region	Percent Agreement	Fleiss' Kappa	AC1 index
Caudate head	0.95 [0.91, 0.99]	0.76 [0.58, 0.95]	0.94 [0.88, 0.99]
Lentiform nucleus	0.92 [0.88, 0.97]	0.76 [0.62, 0.91]	0.89 [0.81, 0.96]
Internal capsule	0.95 [0.91, 0.99]	0.83 [0.70, 0.96]	0.93 [0.87, 0.99]
Insular ribbon	0.92 [0.87, 0.97]	0.79 [0.67, 0.92]	0.86 [0.77, 0.95]
M1	0.88 [0.82, 0.94]	0.62 [0.43, 0.80]	0.83 [0.74, 0.92]
M2	0.96 [0.92, 0.99]	0.83 [0.69, 0.97]	0.94 [0.89, 1.00]
M3	0.97 [0.95, 1.00]	0.89 [0.76, 1.00]	0.97 [0.93, 1.00]
M4	0.94 [0.90, 0.98]	0.77 [0.60, 0.94]	0.92 [0.86, 0.98]
M5	0.93 [0.89, 0.98]	0.80 [0.67, 0.93]	0.90 [0.82, 0.97]
M6	0.92 [0.88, 0.97]	0.69 [0.52, 0.86]	0.90 [0.83, 0.97]
ASPECTS >7 versus ≤7	0.91 [0.86, 0.96]	0.76 [0.62, 0.89]	0.85 [0.76, 0.94]
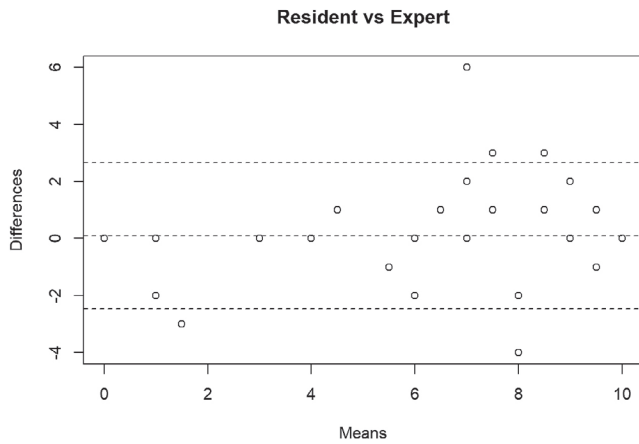
Figure 1. Bland-Altman plot of the differences in ASPECT scores between residents and expert reader.
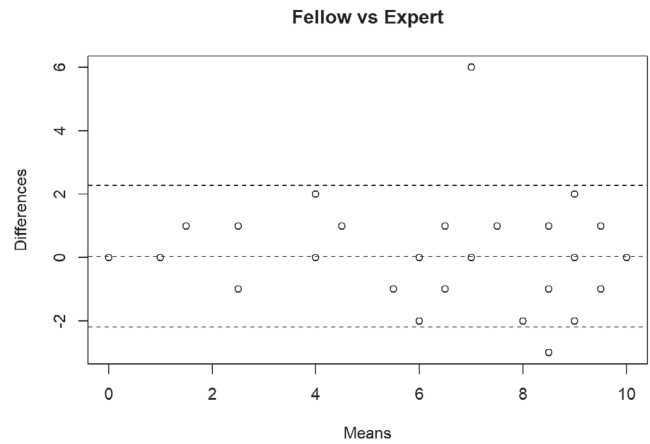
Figure 2. Bland-Altman plot of the differences in ASPECT scores between fellows and expert reader.

Interobserver analyses showed substantial agreement between residents and expert radiologist, and substantial to almost perfect agreement between fellows and expert. In comparison, in a study by Kobkitsuksakul et al.,[11] the levels of agreement were reported to be only slight to moderate between the senior radiology resident and the expert consensus, and moderate to substantial between the neuroradiology fellow and the experts. Our results were more similar to those from the retrospective study by Urgel and Camacho,[8] which reported overall substantial agreement between trainee radiologists (residents and fellows) and the expert.

Agreement coefficients between trainees and the expert were consistently lowest for the lentiform nucleus, internal capsule, M1, and M4. In comparison, the level of agreement between trainees and expert was lowest for the internal capsule and M4 in the study by Urgel and Camacho.[8] It was also lowest for the internal capsule, followed by M4-M6, in the study by Finlayson et al. comparing the scoring done by neuroradiologists and neurologists.[12] Further attention can thus be directed to these anatomic areas during the examination of scans to reduce disagreement.

Differences in the diagnostic performance of trainees may be partly attributed to their experience and level of expertise. As one encounters more cases in training and practice, improvement in the accuracy of his/her interpretation is expected. Since our study was performed in a real-time setting, external factors, such as work-related stressors, reader's fatigue, and urgency to finish one's interpretation and report, may have also affected the performance of the residents and fellows.

The cycle of rotations among residents may also have affected their accuracy and precision. Due to the nature of our study, some residents had just been assigned to CT following a rotation from another section, whereas others had been rotating in CT for more than a month. On the other hand, the fellows who participated in this study were in dedicated CT training.

Bland-Altman plots (Figures 1 and 2) showed that the difference between the ASPECT scores of the expert and trainees was only 1 to 2 points in most cases. Although such difference appeared marginal, this mattered when ASPECT scores were at or near the border score of 7, and such scenarios accounted for most of the false positive and false negative assessment by the residents and fellows. This issue has similarly been raised in the study by Kobkitsuksakul et al.,[11] wherein a middle range of ASPECTS value was proposed to account for cases with borderline ASPECT scores. Rather than outright including or excluding these cases for thrombolysis based on a single cut-off ASPECTS value in a binary classification scheme, it may be more appropriate to categorize these as equivocal and further scrutinize their eligibility for thrombolysis.

A trichotomous scheme was tested by Coutts et al.,[4] but results showed no significant impact on the mean differences between the rater groups. This study, however, was done to compare the performance of treating physicians (stroke fellow or stroke neurologists), rather than radiology trainees, against expert readers. Additionally, the middle category used spanned a wide range of ASPECTS values from 3 to 7. Further studies to investigate the utility of a trichotomous scheme in interpreting ASPECT scores may be considered.

One limitation of our study was the participation of only one expert radiologist, whose reading served as the gold standard. However, this was not expected to be a significant problem as our study aimed to simulate real-time workflow, where each CT study in our institution is signed out by a single certified radiologist, rather than a consensus team, and it was not our study objective to determine gold standard reliability. Another limitation was the participation of residents in various levels of training, which may have resulted in more variation within their group. This was not avoided due to the nature of our institution's workflow and training program.

CONCLUSION

This study showed substantial agreement between residents and expert radiologist, and substantial to almost perfect agreement between fellows and expert, in NECT ASPECTS evaluation of acute MCA infarcts. Assessment by trainees is thus comparable with that of the expert reader and is reliable in guiding clinicians for immediate decision-making in low-resource settings. Caution, however, should be exercised in cases with ASPECTS >7 due to the possibility of underdiagnosis by trainees, and in cases with borderline scores in a binary classification scheme due to the possibility of marginal but management-defining error.

Statement of Authorship

All authors participated in the conceptualization and design of the study, data interpretation and analysis, and approved the final version submitted. Data collection was performed by Dr. Ong and Dr. Urgel.

Author Disclosure

All authors declared no conflicts of interest.

Funding Source

This paper was self-funded by the authors.

REFERENCES

1. Johnson W, Onuma O, Owolabi M, Sachdev S. Stroke: a global response is needed. Bull World Health Organ 2016; 94(9): 634-634A.
2. Osborn AG. Arterial Anatomy and Stroke. In: Osborn's Brain: Imaging, Anatomy and Pathology, 1st ed. Salt Lake City, UT: Amirys, 2012.
3. 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. Lancet 2000; 355(9221): 2170-4.
4. Coutts SB, Demchuk AM, Barber PA, Hu WY, Simon JE, Buchan AM, Hill MD. Interobserver variation of ASPECTS in real time. Stroke 2004; 35(5): e103-5.
5. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al.; on behalf of the American Heart Association Stroke Council. 2018 Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2018; 49: e46-e99.
6. Schröder J, Thomalla G. A Critical Review of Alberta Stroke Program Early CT Score for Evaluation of Acute Stroke Imaging. Front Neurol 2017; 7: 245.
7. Pexman JHW, Barber PA, Hill MD, Sevick RJ, Demchuk AM, Hudon ME, et al. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for Assessing CT Scans in Patients with Acute Stroke. Am J Neuroradiol 2001; 22(8): 1534-42.
8. Urgel RJDL, Camacho AC. Inter-observer Variation of the Alberta Stroke Program Early CT Score (ASPECTS) among Radiologists in the Philippine General Hospital. Acta Med Phil 2019; 53(1): 12-16.
9. Bujang MA, Baharum N. Guidelines of the minimum sample size requirements for Cohen's Kappa. Epidemiol Biostat Public Health 2017; 14(2): e12267.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1): 159-74.
11. Kobkitsuksakul C, Tritanon O, Suraratdecha V. Interobserver agreement between senior radiology resident, neuroradiology fellow, and experienced neuroradiologist in the rating of Alberta Stroke Program Early Computed Tomography Score (ASPECTS). Diagn Interv Radiol 2018; 24: 104-7.
12. Finlayson O, John V, Yeung R, Dowlatshahi D, Howard P, Zhang L, et al. Interobserver Agreement of ASPECT Score Distribution for Noncontrast CT, CT Angiography, and CT Perfusion in Acute Stroke. Stroke 2013; 44: 234-6.