Inter-observer Variation of the Alberta Stoke Program Early CT Score (ASPECTS) among Radiologists in the Philippine General Hospital

Ryan Jason DL. Urgel[1] and Alvin C. Camacho[1]

[1]Section of CT-MRI, Department of Radiology, College of Medicine and Philippine General Hospital, University of the Philippines Manila

ABSTRACT

Objective. To determine the inter-observer variation of ASPECTS among radiologists in the Philippine General Hospital (PGH), particularly between trainee radiologists and an expert reader.

Methods. Thirty (30) cranial CT scan studies of clinically-diagnosed, non-hemorrhagic stroke patients were analyzed by 9 trainee radiologists (3 fellows, and 3 senior and 3 junior residents) and one expert reader. Data analysis involved determining the levels of agreement within and across groups, and against the expert reader.

Results and Conclusion. There was moderate agreement (kappa = 0.60) between the junior residents and the expert reader, and substantial agreement between the senior residents and the expert reader (kappa = 0.70), as well as between the fellows and the expert reader (kappa = 0.63). Over-all, there was substantial agreement between the trainee radiologists and the expert reader (kappa = 0.63). It can be concluded that the interpretation of trainee radiologists in PGH, particularly that of a senior resident or a fellow, is comparable with that of an expert reader, and can, thus, be useful in cases where an interpretation of a CT scan procedure in a clinically-diagnosed stroke patient is needed.

Key Words: ASPECTS, middle cerebral artery infarct, radiology

Corresponding author: Ryan Jason de Lara Urgel, MD
Department of Radiology
College of Medicine
Philippine General Hospital
University of the Philippines Manila
Taft Avenue, Manila 1000, Philippines
Telephone: +632 5548400 local 3100
Email: urgelmd@gmail.com

INTRODUCTION

Stroke is a neurological condition that has been classified as an emergency, as its early detection may lead to life-saving measures. It can either be ischemia or infarction, or hemorrhage of a part of the brain that can be salvaged, if identified early and given the right intervention. These two causes lead to different management decisions. In a study by Navarro *et al.*, published in 2014, it was noted that stroke is the second leading cause of death in the Philippines, with a prevalence of 0.9%. Of these, 70% are from ischemic strokes, whereas the remaining 30% are from hemorrhagic stroke.[1]

Both ischemia and infarction refer to the blockage of a vessel, particularly an artery, which lead to decreased perfusion in the region it supplies. Of interest in this study is the infarction of the middle cerebral artery (MCA) territory. As the MCA is the largest of the intracerebral vessels and supplies almost the entire outer aspect of the brain, its occlusion is the most common type of anterior circulation infarcts[2] and can lead to more widespread and prominent sensorimotor deficits.

The goal of treatment of ischemic strokes is to quickly restore the perfusion in the affected regions of the brain,

especially the region adjacent to the infarcted portion, or the ischemic penumbra. In particular, intravenous (IV) thrombolytic therapy is instituted in cases detected early on in the course of the disease.[3] A "golden hour" has been established in literature, which describes the greatest benefit obtained from re-canalization therapy, or IV thrombolytic therapy, among patients who presented to the hospital, underwent the proper imaging studies, and were given the right therapy, all within an hour.[4]

It is for this reason that the Alberta Stroke Program Early CT Score (ASPECTS) was developed. This scoring system was made to standardize the detection and reporting of the extent of an infarction, and to help the clinicians predict the possible response to thrombolysis.[5] This has supplanted the practice of clinicians in using the one-third rule for acute MCA stroke, wherein the volume of the infarcted region is subjectively estimated and must be less than 1/3 of the MCA territory before treating patients with IV thrombolysis.[6]

Standardized axial sections of the brain are used in doing the ASPECTS. One image is taken at the level of the thalamus and basal ganglia; another is taken adjacent to the most superior margin of the basal ganglia.[6] Assessment is done by determining the extent of the infarction, be it focal swelling or parenchymal hypo-attenuation, subtracting from ten (10) the number of regions involved. Specifically, the regions include the following: caudate head, lentiform nucleus, internal capsule, insular ribbon, anterior MCA cortex, MCA cortex lateral to the insular ribbon, posterior MCA cortex, and anterior, lateral and posterior MCA territories.[6] A final score of 7 or less has been associated with the development of malignant MCA infarction[7] – that is, the condition wherein there is a rapid deterioration of the patient due to the direct and indirect effects of an MCA infarct.[8]

Variation of ASPECTS between real-time and expert ratings was assessed in a study done in 2014. It showed an inter-observer agreement between the real-time analysis of the CT images of acute ischemic stroke by the stroke fellow or neurologist with the expert reading done by either another stroke fellow or neurologist, or by a neuroradiologist. Analysis of the inter-observer variability with a trainee has not been done.[9]

As the evaluation of acute ischemic stroke using ASPECTS can and will affect patient management, it is important to determine the reliability of such scoring system when applied by a trainee radiologist, be it a Computed Tomography (CT) scan Fellow, or a senior or junior Radiology resident. The objective of this study, therefore, is to determine the inter-observer variation of ASPECTS among the trainee radiologists and an expert reader of the Department of Radiology, Philippine General Hospital.

MATERIAL AND METHODS

Prior to the start of the data gathering, each study participant was given a short introduction as to what ASPECTS is, including the process of analysis. He/she then reviewed several cases, evaluating the ASPECTS for each case. Cases included normal brains (but were initially thought of as stroke due to symptomatology) and acute ischemic strokes of the different MCA territories. These cases were taken from those scanned in the 16- and 64-slice CT scan machines of the Department of Radiology, Philippine General Hospital. No additional information was given to the study participants, save for the laterality of the neurologic deficit. During the image or case analysis, the study participants were able to adjust the window width and level of the images of each case, in the dedicated workstation provided for the study participants. Each participant analyzed the images on separate occasions, away from the other participants. The interpretation was done under the supervision of the primary author, in a relatively controlled environment without time limit. The expert reader, who was not aware of the answers of the study participants, then analyzed the cases, with the same conditions. The expert reader in this study is a diagnostic neuroradiologist by training, a Philippine board-certified CT-MRI radiologist, and a diplomate of the American Board of Radiology. His reading was considered as the gold standard for this study. Data analysis was then performed to determine the inter-observer variation between the expert reader and the trainee radiologists, and between the different levels of trainee radiologists.

The trainee radiologists were categorized into three groups: the CT scan Fellow, the senior radiology resident, and the junior radiology resident. By definition, a CT scan fellow is a diagnostic radiology residency graduate, who is already a board-certified radiologist in the Philippines, and is presently undergoing training in the Section of CT-MRI of the Department of Radiology, PGH. A senior radiology resident is in his/her 3rd or 4th year of diagnostic radiology residency training at PGH, with at least three months of exposure in interpreting CT scan images. A junior radiology resident is in his/her 1st or 2nd year of residency training at PGH, with at least a month of exposure in reading CT scan images. A PGH diagnostic radiology resident with no exposure in interpreting CT scan images was not included in this study.

There are 8 CT scan fellows, 11 senior and 6 junior radiology residents in the Department of Radiology of PGH. A maximum of three participants was included per group, and was randomly selected using a list of residents and fellows provided by the administrative office of the Department of Radiology. The number of study participants per group was based on reliability studies using the statistic Kappa, where increasing the number more than 3 has a minor effect on the power of hypothesis tests or the width of confidence intervals.[10] The selected participants were allowed to discontinue their participation at any point in the study.

Thirty cases were reviewed and analyzed individually by the study participants. Each participant, after signing

informed consent, was given a standardized data collection form, where he or she placed the findings noted and the final ASPECT score for each case. Again, the cases were viewed on a dedicated workstation in the Reading room of the Department of Radiology, with appropriate room luminance, minimizing any variations that these may bring about.

Ethical Consideration

The proposal of this study was submitted to the UP Manila Research Ethics Board (UPMREB) for ethics review and approval. Approval for implementation of the study was given last April 24, 2017. Each study participant signed informed consent prior to the start of data collection. A code number was assigned to each study participant, such that data analysis of the inter-observer variation will involve only the trainee classification. Included cases for review were anonymized. Participants were unaware of the patient identity of the cases reviewed. The patient's chart, or CT scan request, was reviewed only to determine the symptom side. No additional risk was incurred by the study participants in this study. No direct material or non-material benefits (which include incentives or compensations) were given to the study participants. No external source of funding was made. All expenses were shouldered by the principal investigator. There was no conflict of interest.

Data Analysis

All numeric measurements were summarized using the mean and standard deviation, while frequencies and percentages were used to describe the categorical data. Analysis of variance was done to analyze the ASPECTS given by the expert reader, fellows, and residents. This was followed by post-hoc tests using the Bonferroni method, if significant.

The ASPECT scores were also dichotomized into >7 and ≤ 7 and the level of agreement between groups of raters on this binary outcome was computed using the kappa statistic. Interpretation will follow Landis and Koch[11]:

< 0.2	slight agreement
0.21 – 0.4	fair agreement
0.41 – 0.6	moderate agreement
0.61 – 0.8	substantial agreement
0.81 – 1.0	almost perfect agreement

Over-all reliability across all raters, regardless of the group was computed using the inter-class correlation coefficient.

Limitations of the Study

The interpretation of the images was done in a relatively controlled environment. The participants were asked to analyze the images at their most convenient time, when no other stressors, such as fulfilling of their other tasks, present. In actual practice, the interpretation of the images may be affected when these other stressors are present.

RESULTS AND DISCUSSION

There were 10 raters involved in the study: 1 expert reader, 3 fellows, 3 senior and 3 junior residents. Each rater was tasked to analyze 30 cases, determining possible involvement in the 10 regions previously mentioned, and to assign an over-all ASPECTS. The ASPECTS were dichotomized into >7 and ≤7, and the levels of agreement were computed.

Table 1 shows that there is a significant difference in the mean ASPECT score assigned by at least two of the groups. Post-Hoc analyses using the Bonferonni method indicate that the significant differences found are due to the average scores coming from the junior residents. Generally, the junior residents assigned significantly higher ASPECT scores compared to the senior residents, fellows and expert reader.

In a small majority (56%) of the cases analyzed, the ASPECT scores given by all 10 participants fell into the same category. There were 5 cases (17%) with low levels of agreement, where the group was divided almost evenly (6 to 4).

Following the Landis and Koch (1977) interpretation of Kappa Statistics, there is a moderate agreement in the

Table 1. Comparison of mean aspects across rater groups

Rater Level	Mean ± SD	p-value
Senior Residents	5.4 ± 2.77	<0.0001
Junior Residents	6.2 ± 2.62	
Fellows	5.3 ± 2.83	
Expert	5.5 ± 3.13	
Post-Hoc Analyses		
Senior Residents vs. Junior Residents		0.0012
Senior Residents vs.Fellows		0.7868
Senior Residents vs. Expert		0.9907
Junior Residents vs. Fellows		0.0004
Junior Residents vs. Expert		0.0406
Fellows vs. Expert		0.4300

Significant : p-value < 0.05

Table 2. Kappa statistics for aspects and region involved, by level of expertise

Region	Rater Type		
	Junior	Senior	Fellow
ASPECTS	0.5370	0.6591	0.6214
Caudate Head	0.5132	0.5767	0.7575
Lentiform Nucleus	0.5500	0.5833	0.5520
Internal Capsule	0.2546	0.5475	0.2000
Insular Ribbon	0.5982	0.6296	0.7456
M1	0.5370	0.7777	0.5556
M2	0.3728	0.2440	0.5325
M3	0.3651	0.5652	0.6496
M4	0.2877	0.1984	0.4219
M5	0.4219	0.7456	0.2241
M6	0.4907	0.5950	0.2889

M1 – anterior MCA cortex; M2 – MCA cortex lateral to the insular ribbon; M3 –posterior MCA cortex; M4 – anterior MCA territory; M5 – lateral MCA territory; M6 – posterior MCA territory

Table 3. Kappa statistics for aspects and region involved

	All Levels	F & S	F & J	S & J	E & F	E & S	E & J
ASPECTS	0.6289	0.6345	0.6063	0.6000	0.6296	0.6990	0.6005
Caudate Head	0.6498	0.6924	0.6235	0.5882	0.7748	0.6500	0.5315
Lentiform Nucleus	0.5518	0.5676	0.5172	0.5311	0.5996	0.6206	0.5530
Internal Capsule	0.3464	0.3278	0.3051	0.3400	0.2674	0.5794	0.3173
Insular Ribbon	0.6523	0.7304	0.6139	0.5723	0.7654	0.6834	0.6408
M1	0.6256	0.6667	0.5735	0.5968	0.6444	0.7221	0.6021
M2	0.4352	0.4031	0.4534	0.3554	0.5615	0.4019	0.4203
M3	0.5671	0.6150	0.5237	0.4886	0.7120	0.6017	0.4566
M4	0.3789	0.3910	0.3554	0.2685	0.4765	0.3626	0.4000
M5	0.4875	0.4991	0.3441	0.5345	0.3702	0.7946	0.4480
M6	0.4919	0.4784	0.4119	0.5719	0.4000	0.5636	0.4430

F – Fellows; S – Senior residents; J – Junior residents; E – Expert reader; M1 – anterior MCA cortex; M2 – MCA cortex lateral to the insular ribbon; M3 –posterior MCA cortex; M4 – anterior MCA territory; M5 – lateral MCA territory; M6 – posterior MCA territory

junior resident group on overall ASPECTS, while there is a substantial agreement in the senior residents and fellows groups. Generally, the junior residents group manifested lower levels of agreement in their findings for the individual regions compared to the fellows and senior residents, with only fair agreement on the internal capsule, M2, M3 and M4. Moderate agreement was found across the junior residents in all other regions. Fellows generally showed moderate to substantial agreement in their findings for the specific regions, except in the internal capsule, M5 and M6, wherein the agreement was fair. Senior residents showed lower levels of agreement on M2 and M4, and moderate to substantial agreement on all other regions.

Overall, the differences in the interpretation within groups and across groups can be attributed to the following factors:

Total number of cases encountered in practice or during training: In general, the greater the number of cases encountered and analyzed, the greater the probability of determining the actual involved regions, the more homogeneous the results will be within and across groups. As such, junior residents will likely show a more varied interpretation than senior residents or fellows. Likewise, within the junior residents group, it was noted that one of them only had a one-month rotation in the CT scan, which can limit the participant's ability to accurately interpret the CT scan images.

Time the study participant performed the analysis: This is with regard to the moment when the study participant did the interpretation. Most did the analysis after their usual work, which could have contributed to the variation in the results. Since these are trainee radiologists, eye strain and fatigue may have factored in the interpretation.

Recent and current rotations: If a trainee radiologist constantly reads and interprets CT scan studies, he or she will become more adept doing so. The junior residents included in this study have had a month or two of CT rotation and were currently rotating in other modalities; as such, they were only able to analyze CT scan images occasionally. Senior residents were either currently rotating in or had a recent rotation in CT. A recent rotation of the senior residents in MRI may also influence their interpretation, as the same modality is also used in the diagnosis of stroke. The CT fellows included in this study have daily exposure to CT scan cases.

A similar pattern can be found when comparing the results of the three groups to the expert reader. In the overall ASPECTS category, there was moderate agreement (kappa = 0.60) between the junior residents and the expert reader, and substantial agreement between the senior residents and the expert reader (kappa = 0.70), as well as between the fellows and the expert reader (kappa = 0.63). For the fellows, the levels of agreement between them and the expert reader were particularly higher for caudate nucleus, insular ribbon and M3. For senior residents, stronger levels were found in the M1 and M5 regions. For junior residents, the results indicated a general pattern of fair to moderate agreement with the expert reader.

There is a better agreement between the interpretation of the senior residents and fellows with the expert reader. This can be attributed to the expertise of the study participants, which can, again, be related to the total number of cases encountered and analyzed in practice and during training. This agreement is similar to that found in the study by Coutts, *et al.*, where a substantial agreement is noted between a stroke fellow or neurologist and an expert reader.

Between groups, it was found that more often, the level of agreement in scores was higher between fellows and senior residents compared to fellows and junior residents or between seniors and juniors for that matter. Over-all, levels of agreement were noticeably lower for the internal capsule and M4, and there was moderate to substantial agreement across all 10 raters for the remaining regions.

CONCLUSION AND RECOMMENDATIONS

A radiologist is an important component of the Brain Attack Team – a group formed to handle patients who present with and were diagnosed with stroke in the emergency room immediately after or within a few hours of the onset of

neurologic symptoms. In an institution where a Brain Attack Team is in place, or where the immediate interpretation of a CT scan study is needed, the agreement between the reading of an expert reader and the trainee radiologist is of utmost importance. This is because the initial reading of a trainee radiologist will have an impact on the management of a patient, should the expert or official reader be unavailable to do the immediate interpretation. The analysis of the trainee radiologists becomes important as it helps identify which patients will benefit from thrombolysis, or which patients will have a poor prognosis. A substantial agreement between the trainee radiologists and the expert reader was determined in this study. With such results, the initial interpretation of a trainee radiologist, particularly that of a senior resident or a fellow, can be deemed useful and helpful in the management of a stroke patient.

Although this study established the different levels of agreement within and across the different groups, this was done in a relatively controlled environment. In actual practice, interpretation of the CT images may be potentially affected when it is done early in the morning (at the end of a duty shift), and when a lot of stressors are present (such as the presence of an attending neurologist and other members of the Brain Attack Team, while doing the interpretation). As such, a study where real-time analysis of the CT scan images of stroke patients but will involve trainee radiologists can be performed, similar to the study by Coutts, *et al*.

Statement of Authorship

All authors have approved the final version submitted.

Author Disclosure

All the authors declared no conflicts of interest.

Funding Source

This paper was funded by the principal investigator.

REFERENCES

1. Navarro JC, Baroque AC 2nd, Lokin JK, et al. The real stroke burden in the Philippines [abstract]. Int J Stroke [Online]. 2014 [cited 2017 Apr 1];9(5):640-1. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24844610

2. Jichici D, Baird AE. Anterior Circulation Stroke [Online]. WebMD LLC; 2015 [updated 2015 Dec 8; cited 2017 Apr 1]. Available from: http://emedicine.medscape.com/article/1159900-overview#a1

3. Caplan LR. Patient education: ischemic stroke treatment (beyond the basics) [Online]. UpToDate, Inc.; 2017 [updated 2018 Jan 17; cited 2017 Apr 1]. Available from: http://www.uptodate.com/contents/ischemic-stroke-treatment-beyond-the-basics#H1

4. Saver JL, Smith EE, Fonarow GC, et al. The "Golden Hour" and acute brain ischemia. Stroke [Online]. 2010 [cited 2017 Apr 1];41:1431-39. Available from: http://stroke.ahajournals.org/content/41/7/1431

5. Understanding alberta stroke program early CT score (ASPECTS) [Online]. University of Calgary - Faculty of Medicine [cited 2017 Apr 1]. Available from: http://www.aspectsinstroke.com/

6. Warwick 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 strokes. AJNR [Online]. 2001 [cited 2017 Apr 1];22:1534-42. Available from: http://www.ajnr.org/content/22/8/1534.long

7. MacCallum C, Churilov L, Mitchell P, et al. Low alberta stroke program early CT score (ASPECTS) associated with malignant middle cerebral artery infarction (Abstract). Cerebrovasc Dis [Online]. 2014 [cited 2017 Apr 1];38(1):39-45. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25228461

8. Treadwell SD, Thanvi B. Malignant middle cerebral artery (MCA) infarction: pathophysiology, diagnosis and management (Abstract). Postgrad Med J [Online]. 2010 Apr [cited 2017 Apr 1];86(1014):235-42. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20354047

9. Coutts SB, Demchuk AM, Barber PA, et al. Interobserver variation of ASPECTS in real time. Stoke [Online]. 2004 Apr 22 [cited 2017 Apr 1];35:e103-e105. Available from: http://stroke.ahajournals.org/cgi/pmidlookup?view=long&pmid=15073381

10. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation and sample size requirements. Physical Therapy [Online]. 2005 Mar 1 [cited 2017 Apr 1];85:257-268. Available from: https://academic.oup.com/ptj/article/85/3/257/2805022

11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics [Online].1977 Apr [cited 2017 Apr 1];33:159-174. Available from: https://www.researchgate.net/publication/22310512_The_Measurement_Of_Observer_Agreement_For_Categorical_Data.