June 2014

Inter-observer variability in diagnosing radiological features of aneurysmal subarachnoid hemorrhage; a preliminary single centre study comparing observers from different specialties and levels of training

Usman T Siddiqui
Aga Khan University

Anjum F. Khan
Aga Khan University

Muhammad Shahzad Shamim
Agha Khan University, shahzad.shamim@aku.edu

Rana Shoaib Hamid
Aga Khan University

Muhammad Mehboob Alam
Aga Khan University

Authors
Usman T Siddiqui, Anjum F. Khan, Muhammad Shahzad Shamim, Rana Shoaib Hamid, Muhammad Mehboob Alam, and Emaduddin Siddiqui

Surgical Neurology International

OPEN ACCESS

For entire Editorial Board visit : http://www.surgicalneurologyint.com

Original Article

Inter-observer variability in diagnosing radiological features of aneurysmal subarachnoid hemorrhage; a preliminary single centre study comparing observers from different specialties and levels of training

Usman T. Siddiqui, Anjum F. Khan, Muhammad Shahzad Shamim, Rana Shoaib Hamid[1], Muhammad Mehboob Alam, Muhammad Emaduddin[1]

Section of Neurosurgery, Departments of Surgery and [1]Radiology, Aga Khan University Hospital, Karachi, Pakistan

E-mail: Usman T. Siddiqui - usmantsiddiqui@hotmail.com; Anjum F. Khan - drfaridkhan@hotmail.com; *Muhammad Shahzad Shamim - shahzad.shamim@aku.edu; Rana Shoaib Hamid - d_himself@hotmail.com; Muhammad Mehboob Alam - muhammad.alam@aku.edu; Muhammad Emaduddin - muhammad.emaduddin@aku.edu
*Corresponding Author

Received: 22 October 13 Accepted: 02 April 14 Published: 17 June 14

This article may be cited as:
Siddiqui UT, Khan AF, Shamim MS, Hamid RS, Alam MM, Emaduddin M. Inter-observer variability in diagnosing radiological features of aneurysmal subarachnoid hemorrhage; a preliminary single centre study comparing observers from different specialties and levels of training. Surg Neurol Int 2014;5:96.
Available FREE in open access from: http://www.surgicalneurologyint.com/text.asp?2014/5/1/96/134654

Abstract

Background: A noncontrast computed tomography (CT) scan remains the initial radiological investigation of choice for a patient with suspected aneurysmal subarachnoid hemorrhage (aSAH). This initial scan may be used to derive key information about the underlying aneurysm which may aid in further management. The interpretation, however, is subject to the skill and experience of the interpreting individual. The authors here evaluate the interpretation of such CT scans by different individuals at different levels of training, and in two different specialties (Radiology and Neurosurgery).

Methods: Initial nonontrast CT scan of 35 patients with aSAH was evaluated independently by four different observers. The observers selected for the study included two from Radiology and two from Neurosurgery at different levels of training; a resident currently in mid training and a resident who had recently graduated from training of each specialty. Measured variables included interpreter's suspicion of presence of subarachnoid blood, side of the subarachnoid hemorrhage, location of the aneurysm, the aneurysm's proximity to vessel bifurcation, number of aneurysm(s), contour of aneurysm(s), presence of intraventricular hemorrhage (IVH), intracerebral hemorrhage (ICH), infarction, hydrocephalus and midline shift. To determine the inter-observer variability (IOV), weighted kappa values were calculated.

Results: There was moderate agreement on most of the CT scan findings among all observers. Substantial agreement was found amongst all observers for hydrocephalus, IVH, and ICH. Lowest agreement rates were seen in the location of aneurysm being supra or infra tentorial. There were, however, some noteworthy exceptions. There was substantial to almost perfect agreement between the radiology graduate and radiology resident on most CT findings. The lowest agreement was found between the neurosurgery graduate and the radiology graduate.

Access this article online
Website:
www.surgicalneurologyint.com
DOI:
10.4103/2152-7806.134654
Quick Response Code:

Conclusion: Our study suggests that although agreements were seen in the interpretation of some of the radiological features of aSAH, there is still considerable IOV in the interpretation of most features among physicians belonging to different levels of training and different specialties. Whether these might affect management or outcome is unclear.

Key Words: Aneurysm, computed tomography, inter-observer variability, subarachnoid hemorrhage

INTRODUCTION

Aneurysmal subarachnoid hemorrhage (aSAH) is a common neurosurgical pathology with significant morbidity and mortality. The incidence of aSAH varies in different populations but it is reported to be as high as 22.5 cases per 100,000. These patients mostly present in the emergency room and heavily rely on early diagnosis and initiation of management. The preliminary radiological investigation of choice for the diagnosis of aSAH remains a noncontrast computed tomography (CT) scan (Class1, Level of evidence B), which not only confirms the diagnosis of aSAH, but also provides a number of useful information.[5] The amount of blood on this initial CT scan has also been shown to be associated with later development of delayed cerebral ischemia, angiographic vasospasm, and patient outcome.[1,6,9] Other abnormal findings on the initial CT scan such as extent of hydrocephalus, cerebral edema, intra- or extra-axial hematomas also influence the management, ranging from initial nonoperative management to urgent interventions such as performing a ventriculostomy, or even craniotomy for aneurysm clipping, clot evacuation and/or decompressive craniectomy.[3,12]

It may be added here that in most hospitals of the world, these initial CT scans are interpreted by physicians who are still in their training and their senior colleagues communicate the interpretation of these CT scans on phone. The interpretation of CT scan is subject to inter-observer variability (IOV), and is naturally dependent on the level of training and the specialty of training of the interpreting individual. Very little has been done to explore this variability even though it is likely to affect at least the emergency room management of these critically ill patients. At the time of preparing this draft, there were only four published studies evaluating IOV in aSAH.[7,10,11,13] In this paper, we aim to evaluate, for the first time, the IOV in interpretation of initial CT scans of patients with aSAH, in individuals at two different levels of training and in two different specialties (Radiology and Neurosurgery).

METHODS

Study setting and population

The study was approved by the Institutional Ethics Review Committee (letter attached). We retrospectively reviewed all adult patients admitted to our hospital with aSAH for a 3-year period (between January 2009 and December 2011). The diagnosis of aSAH was based on a positive initial noncontrast head CT scan (done within 48 h of admission for all patients) and subsequent conventional catheter angiography, CT Angiography (CTA) or Digital Subtraction Angiography (DSA). All patients with peri-mesencephalic hemorrhage diagnosed by a typical pattern of hemorrhage on CT made within 3 days after onset of the symptoms and in the absence of an aneurysm on angiography or CTA were excluded. All pediatric patients, patients with subarachnoid hemorrhage (SAH) due to reasons other than aneurysmal rupture, such as head trauma or arteriovenous malformations, were also excluded from the study.

Detail of observers

Each patient's initial noncontrast CT scan was evaluated independently by four different observers and findings were noted on a standardized proforma. The four observers selected for the study included two from Radiology; a resident in mid training (RR: Radiology Resident) and a resident who had just graduated from training (RG: Radiology Graduate). A similar group of two from Neurosurgery was also selected (NR: Neurosurgery Resident, NG: Neurosurgery Graduate). These groups were selected on the basis that in most programs one or more combinations of such groups are usually on call and are among the first to interpret patient's noncontrast CT scan prior to communicating to more senior colleagues. For the purpose of this study, all four observers were blinded to each other's findings and to the subsequent confirmation of findings through angiography.

Study variables

The proforma included the following variables; presence of subarachnoid blood, side of the SAH, vessel location of the aneurysm, aneurysm's origin to a specific vessel bifurcation, number of aneurysm(s), contour of aneurysm(s), presence of intraventricular hemorrhage (IVH), intracerebral hemorrhage (ICH), infarction, hydrocephalus, midline shift, and the most likely vessel involved. Presence of IVH, ICH, infarction, hydrocephalus, and midline shift was recorded as a Yes/No. The Fisher scale was used to evaluate the amount of blood, which divides it into four classes. Fisher (1): No or minimal blood; (2): Diffuse thin layers, no clots >1 mm thick; (3): Localized blood clots in the subarachnoid space, or vertical layers of blood >1 mm thick (interhemisperic fissure, insular cistern, and ambient cistern); and (4): Diffuse or no SAH, with intracerebral or ventricular blood. This is currently the standard accepted grading scale for blood in aSAH. Hydrocephalus was assessed by subjective assessment of ventricular size and calculation of the ventricle to cortex ratio (VCR).

Statistical analysis

The observations of all four observers were compared with each other making a set of six comparisons for each variable. IOV expresses the degree of agreement among raters or the reproducibility of a measurement. To determine the IOV for the aforementioned variables, weighted kappa values were calculated. IOV was interpreted as poor (<0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80), and almost perfect (0.81-1.00) agreement based on the respective kappa scores. Analysis was performed using Stata v11.2 (StataCorp 1985-2009).

RESULTS

Patient demographics

Thirty-five patients' initial CT-scans were included in the study. Of these, 14 (40%) were male and 21 (60%) were female. Mean age at presentation was 48.4 + 13.3 years. Fifteen (43%) patients presented with a neurological deficit, while 20 (57%) patients were neurologically intact. Mean postictal time was 5.4 + 2.1 days.

Subarachnoid hemorrhage

All observers agreed that none of the scans were normal and no epidural or subdural hematoma was present (kappa value 1.0). Inter-observer agreement values are summarized in Table 1.

There was moderate agreement (Kappa 0.525; 95% CI 0.055-0.995) between the NG and NR. The agreement was considerably higher in the rest of the groups reaching almost perfect agreement (Kappa 0.842; 95% CI 0.540-1.000) between NG and RR. There was only poor to moderate agreement while considering the side

Table 1: Kappa values between 6 observer pairs for interpretation of CT scan findings after SAH

Observers Questions	Kappa (CI)					
	NG-RG	NG-NR	NG-RR	RG-NR	RG-RR	NR-RR
SAH	0.525 (0.055-0.995)	0.525 (0.055-0.995)	0.842 (0.540-1.000)	0.718 (0.348-1.000)	0.718 (0.348-1.000)	0.718 (0.348-1.000)
Side of SAH	0.321 (0.248-0.363)	0.478 (0.395-0.577	0.249 (0.191-0.271)	0.528 (0.508-0.566)	0.560 (0.462-0.601)	0.472 (0.324-0.607)
Location of SAH	0.323 (0.138-0.460)	0.137 (−0.403-0.677)	0.271 (−0.295-0.837)	0.531 (0.106-0.957)	0.891 (0.670-1.000)	0.609 (0.208-1.000)
Location supra or infra tentorial	0.183 (−0.031-0.341)	0.362 (0.158-0.522)	0.103 (−0.412-0.617)	0.278 (0.164-0.601)	0.309 (−0.090-0.709)	0.227 (−0.244-0.698)
Location anterior or posterior circulation	0.300 (0.197-0.384)	0.105 (−0.442-0.653)	0.179 (0.078-0.239)	0.298 (−0.129-0.725)	0.751 (0.574-0.927)	0.519 (0.452-0.647)
Aneurysm arising from vessel bifurcation	0.216 (−0.202-0.634)	0.664 (0.314-1.000)	0.672 (0.338-1.000)	0.444 (0.036-0.853)	0.143 (−0.283-0.569)	0.615 (0.234-0.997)
Parenchymal Hemorrhage	0.689 (0.439-0.940)	0.668 (0.414-0.922)	0.668 (0.414-0.922)	0.724 (0.480-0.969)	0.724 (0.480-0.969)	1.000 (1.000-1.000)
Intraventricular Hemorrhage	0.828 (0.643-1.000)	0.943 (0.833-1.000	0.886 (0.732-1.000)	0.886 (0.734-1.000)	0.943 (0.832-1.000)	0.943 (0.833-1.000)
Aneurysm contour	0.000 (0-1.000)	0.000 (0-1.000)	0.000 (0-1.000)	−0.008 (−0.305-0.289)	0.123 (−0.211-0.458)	0.077 (−0.264-0.418)
Number of aneurysms	*	*	*	*	0.000 (0-1.000)	0.000 (0-1.000)
Hydrocephalus	0.721 (0.468-0.974)	0.815 (0.618-1.000)	0.867 (0.689-1.000)	0.684 (0.440-0.927)	0.861 (0.675-1.000)	0.815 (0.618-1.000)
Vessel involved	0.399 (0.254-0.503)	0.296 (0.261-0.358)	0.513 (0.448-0.644)	0.292 (0.024-0.417)	0.518 (0.375-0.702)	0.408 (0.283-0.631)
Fisher grade	0.408 (0.263-0.438)	0.665 (0.394-0.833)	0.613 (0.463-0.738)	0.385 (0.273-0.540)	0.438 (0.429-0.501)	0.702 (0.637-0.743)
Infarction	−0.045 (−0.134-0.044)	−0.040 (−0.121-0.042)	−0.048 (−0.142-0.046)	0.785 (0.380-1.000)	0.525 (0.055-0.995	0.639 (0.186-1.000)

NG: Neurosurgery graduate, RG: Radiology graduate, NR: Neurosurgery resident, RR: Radiology resident, SAH: Subarachnoid hemorrhage. *Insufficient observations to calculate and location of the SAH in most of the groups except between RG and RR, which was almost always substantial to near perfect.

Vessel of origin of aneurysm, proximity to bifurcation, number of aneurysms and contour of aneurysm

There was fair to moderate agreement when identifying the vessel from which the aneurysm was arising with the highest agreement being between RG and RR (Kappa 0.518; 95% CI 0.375-0.702). Though when commenting on the likelihood of proximity to a specific vessel bifurcation, the lowest agreement was seen among the RG and RR (Kappa 0.143; 95% CI −0.283 to 0.569) while the highest agreement was reached between RR and NG (Kappa 0.672; 95% CI 0.338-1.000). Due to the difficulty in identification of the number and contour of aneurysms on a CT, there was insufficient data and a meaningful analysis was not possible.

ICH and IVH

There was substantial agreement between most groups when considering ICH (Kappa 0.689; 95% CI 0.439-0.940) and it was perfect when comparing NR and RR (Kappa 1.0). An almost perfect agreement was seen in all groups when diagnosing IVH (Kappa 0.943; 95% CI 0.833-1.000).

Fisher grade

There was considerably variability when comparing the Fisher grading between the groups. There was substantial agreement between the NR and RR (Kappa 0.702; 95% CI 0.637-0.743) but it fell considerably, to fair, when comparing between the RG and NR (Kappa 0.385; 95% CI 0.273-0.540).

Hydrocephalus

Most of the groups had an almost perfect agreement when diagnosing hydrocephalus except when comparing the NG and RG (Kappa 0.721; 95% CI 0.468-0.974) and RG and NR (Kappa 0.684; 95% CI 0.440-0.927), which was only moderate.

Infarction

There was a remarkable difference in the level of agreement among the various groups. All the comparisons with the NG involved, showed poor agreement the lowest between NG and NR (Kappa -0.040; 95% CI -0.121 to -0.042). The agreement was substantial in the remaining three categories, highest being achieved between the RG and NR (Kappa 0.785; 95% CI 0.380-1.000).

Inter-group comparison

When looking at each of the six individual groups there were some notable findings. The highest agreement rates were noted between the RG and RR through almost all the analyzed variables. The second highest agreement rates were seen between RR and NR and they were relatively similar when compared with the RG and RR group. Surprisingly though the lowest results were noted in most of the variables when the agreement rate was analyzed between the RG and NG and a relatively similar result was seen when comparing the RR with the NG.

DISCUSSION

The study assesses the IOV in the interpretation of the initial noncontrast CT scan obtained after an aSAH. This is the first study that compares physicians from two different specialties, Radiology and Neurosurgery, and at two different levels of training, that is, mid-training resident and fresh residency program graduate. The justification for selection of these groups is mentioned in the "Methods" section.

There was considerable variation in the inter-observer agreement for different variables. The highest agreement was noted in all the groups when evaluating presence of subarachnoid blood, ICH, IVH, and hydrocephalus. The lowest agreement was seen in the prediction of the vessel of origin of ruptured aneurysm and the whether the aneurysm was at or near a vessel bifurcation.

There was also considerable variation when we compared the Fisher grade to evaluate the amount of blood in the subarachnoid space. This is consistent with most other studies, which have evaluated the IOV in Fisher grading. There is also published literature on the IOV with the quantitative assessment (Hijdra Scoring) of the amount of blood. Van Norden *et al.* noted higher inter-observer agreement rates on Hijdra scores (kappa 0.67-0.75) than Fisher grades (kappa 0.37-0.55).[4,13]

While some studies have found a high variability in inter-observer agreement when analyzing hydrocephalus, we found a substantial to almost perfect agreement. The high variability in those studies was attributed to the subjective interpretation of the presence or absence of hydrocephalus.[2,7] There are numerous quantitative methods to evaluate hydrocephalus but VCR is now being increasingly used as a substitute for subjective analysis. Van Zagten *et al*. found high inter-observer agreement for VCR (ICC 0.82; 95% CI 0.75-0.94) when examining brain atrophy.[14] This is further supported by low agreement when a similar study evaluating brain atrophy used subjective estimation based on categorical scales (kappa 0.11-0.59).[8]

We do not think we have presented enough evidence to draw any meaningful conclusions regarding the IOV of diagnosing aSAH, or any of its associated radiological features. Our study is limited in both its methodology and data. These are mere observations of four individuals and cannot be generalized by any stretch of imagination. More observers from each specialty in each group, from different training programs, evaluated at the same time of the day and on a larger number of CT scans would have been a better way to test our hypothesis. Moreover, it may also show whether the level of training and experience influences the IOV for interpreting aSAH for either Radiology or Neurosurgery residents. It may also help in exploring the minimal level of training required to safely diagnose aSAH and its associated radiological features. We have in all honesty, just explored the tip of the iceberg. However, we do think our modest effort raises interesting food for thought. First, that such a study is possible. The analysis that we have attempted to use, in an effort to compare six different groups, to the best of our knowledge, has not been used previously in either Radiology or Neurosurgery literature. It is obvious that the topic of this study is extremely important and a common topic for corridor discussion among both Radiologists and Neurosurgeons; and often a subject of disagreement in multi-disciplinary meetings, but surprisingly, very few investigators have tried to explore it scientifically.

Once it is established that such an IOV does exist across levels of training and specialties, it might be useful to go back to the drawing board and explore whether the discrepancy is in the method of teaching, or whether the differences faze out gradually with increasing experience and interaction. It is also unclear whether this IOV does in fact affect patient outcome to a significant extent, as given the high morbidity and mortality of aSAH, senior colleagues leave little for the residents to decide on their own. It seems though that at least the emergency room management especially in terms of efficiency, might be at risk. However, to answer these questions in more detail, this topic requires a more extensive exploration.

CONCLUSION

Our study suggests that although agreements were seen in the interpretation of some of the radiological features of aSAH, there is still considerable IOV in the interpretation of most features among physicians belonging to different levels of training and different specialties. Whether these might affect management or outcome is unclear.

REFERENCES

1. Adams HP Jr, Kassell NF, Torner JC, Haley EC Jr. Predicting cerebral ischemia after aneurysmal subarachnoid hemorrhage: Influences of clinical condition, CT results, and antifibrinolytic therapy. A report of the Cooperative Aneurysm Study Neurology 1987 Oct; 37:1586-91.
2. Bhattathiri PS, Gregson B, Prasad KS, Mitchell P, Soh C, Mitra D, *et al*. Reliability assessment of computerized tomography scanning measurements in intracerebral hematoma. Neurosurg Focus 2003 Oct 15;15:E6.
3. Bullock MR, Chesnut R, Ghajar J, Gordon D, Hartl R, Newell DW, *et al*. Surgical management of acute subdural hematomas. Neurosurgery 2006 Mar; 58(3 Suppl):S16-24; discussion Si-iv.
4. Claassen J, Bernardini GL, Kreiter K, Bates J, Du YE, Copeland D, *et al*. Effect of cisternal and ventricular blood on risk of delayed cerebral ischemia after subarachnoid hemorrhage: The Fisher scale revisited. Stroke. 2001 Sep; 32:2012-20.
5. Connolly ES Jr., Rabinstein AA, Carhuapoma JR, Derdeyn CP, Dion J, Higashida RT, *et al*. Guidelines for the management of aneurysmal subarachnoid hemorrhage: A guideline for healthcare professionals from the American Heart Association/american Stroke Association. Stroke. 2012 Jun; 43:1711-37.
6. Hijdra A, van Gijn J, Nagelkerke NJ, Vermeulen M, van Crevel H. Prediction of delayed cerebral ischemia, rebleeding, and outcome after aneurysmal subarachnoid hemorrhage. Stroke 1988 Oct; 19:1250-6.
7. Ibrahim GM, Weidauer S, Macdonald RL. Interobserver variability in the interpretation of computed tomography following aneurysmal subarachnoid hemorrhage. J Neurosurg 2011 Dec; 115:1191-6.
8. Leonardi M, Ferro S, Agati R, Fiorani L, Righini A, Cristina E, *et al*. Interobserver variability in CT assessment of brain atrophy. Neuroradiology 1994;36:17-9.
9. Qureshi AI, Sung GY, Razumovsky AY, Lane K, Straw RN, Ulatowski JA. Early identification of patients at risk for symptomatic vasospasm after aneurysmal subarachnoid hemorrhage. Crit Care Med 2000 Apr; 28:984-90.
10. Svensson E, Starmark JE, Ekholm S, von Essen C, Johansson A. Analysis of interobserver disagreement in the assessment of subarachnoid blood and acute hydrocephalus on CT scans. Neurol Res 1996;18:487-94.
11. van der Jagt M, Hasan D, Bijvoet HW, Pieterman H, Koudstaal PJ, Avezaat CJ. Interobserver variability of cisternal blood on CT after aneurysmal subarachnoid hemorrhage Neurology 2000;54:2156-8.
12. van Gijn J, Hijdra A, Wijdicks EF, Vermeulen M, van Crevel H. Acute hydrocephalus after aneurysmal subarachnoid hemorrhage. J Neurosurg 1985;63:355-62.
13. van Norden AG, van Dijk GW, van Huizen MD, Algra A, Rinkel GJ. Interobserver agreement and predictive value for outcome of two rating scales for the amount of extravasated blood after aneurysmal subarachnoid haemorrhage. J Neurol 2006 Sep; 253:1217-20.
14. van Zagten M, Kessels F, Boiten J, Lodder J. Interobserver agreement in the assessment of cerebral atrophy on CT using bicaudate and sylvian-fissure ratios. Neuroradiology 1999; 41:261-4.