Reliability assessment of computerized tomography scanning measurements in intracerebral hematoma

PARAMESWARAN S. BHATTATHIRI, M.CH., BARBARA GREGSON, PH.D.,
KUSKOOR S. MANJUNATH PRASAD, M.S., M. CH., PATRICK MITCHELL, F.R.C.S.,
CALVIN SOH, F.R.C.S.(I), DIPAYAN MITRA, M.D., M.R.C.P., ANIL GHOLKAR, M.B.B.S.,
AND ALEXANDER D. MENDELOW, F.R.C.S., PH.D.

Departments of Neurosurgery and Neuroradiology, Newcastle General Hospital, Newcastle upon Tyne, United Kingdom

Object. As one of the aspects of the International Surgical Trial in Intracerebral Haemorrhage (STICH), prerandomization computerized tomography (CT) scans were collected. In the present study the authors determined the inter- and intraobserver variability of various parameters pertinent to CT scans obtained in patients with intracerebral hematomas (ICHs).

Methods. A protocol was devised to analyze CT scans in a uniform and systematic manner. Each observer evaluated the same set of scans twice, with a minimum 2-month interval between assessments. In addition to noting the side and the sites of involvement, the observers measured the scale present on the scan itself and the length, breadth, height, and depth of the spontaneous ICH as well as the midline shift. The intraclass correlation was very high (κ 0.8–1) for the measurements of volume, depth, and midline shift. Good interobserver agreement (κ 0.8–1) was demonstrated with regard to involvement of basal ganglia or thalamus, presence of intraventricular extension, and the side of the hematoma. Agreement was substantial (κ 0.61–0.8) with regard to identifying primary involvement of particular lobes. Agreement was moderate (κ 0.41–0.6) on the presence or absence of hydrocephalus. When comparing the first and the second sets of readings, the intraobserver agreement was good (80–100%).

Conclusions. The study quantifies the degree of inter- and intraobserver agreement regarding evaluation of CT scans in patients with ICH when conducted in accordance with a set protocol.

KEY WORDS • intracerebral hematoma • computerized tomography • observer variation • reliability • Surgical Trial in Intracerebral Haemorrhage

The definitive management of and the long-term outcome after ICH remain controversial and unclear. Investigators associated with STICH continue to address these issues.

Because much research into ICH hinges on imaging parameters of the ICH, it is important to know the accuracy of these results. In patients with ICH, there is some clinical evidence that it is important to separate superficial (lobar) and deep (basal ganglia) locations, because there are attendant consequences for treatment and prognosis. Hemorrhages occurring at these sites have different causes and different risk factors.³ The approximate size of the hematoma influences management. The depth from the surface, which then affects surgical accessibility, further weights this selection. In clinical practice these assessments are usually performed informally. We have devised simple and reproducible criteria that are evaluated in this study.

Abbreviations used in this paper: CT = computerized tomography; ICH = intracerebral hemorrhage; STICH = Surgical Trial in Intracerebral Haemorrhage.

In a search of the literature, we found only one study in which observer variability was analyzed in relation to determining the location of the ICH.¹⁹ These authors expressed substantial concern about the likelihood of a higher degree of disagreement among the less experienced radiologists, neurologists, and neurosurgeons.

To our knowledge this study is the first of its kind and was devised with the objective of determining the inter- and intraobserver variability among four experienced neurosurgeons and two experienced neuroradiologists in the evaluation of CT scans for intracerebral hematoma with respect to the location, size, and presence/absence of intraventricular hemorrhage and/or hydrocephalus.

CLINICAL MATERIAL AND METHODS

Computerized tomography scans, obtained as a part of the STICH were used as the radiological basis of this study. The database included only the patients randomized into the trial according to the preestablished protocol.¹ An opportunity sample of 43 CT scans was selected from a total of 1033 cases included in the trial. The observers

were blinded to clinical presentation and outcome data at the time of the review, reports, the results of the other observers, and their own earlier determinations. A protocol was devised by which to analyze the CT scans in a uniform and systematic manner. The guidelines shown in Table 1 have been used in the analysis of the prerandomization CT scans obtained in patients recruited for STICH.

Each observer read the same set of scans twice, with a minimum 2-month interval. All the observers used the same electronic caliper to measure the hematoma. The measurements were then converted into the actual values with reference to the scale on the scan, when available, by using mathematical conversion on a Microsoft Excel worksheet. We analyzed the interobserver variability among six observers (four neurosurgeons [two consultants and two specialist registrars] and two neuroradiologists [one consultant and one specialist registrar]) in the assessment of the localization and measurement of ICH demonstrated on CT scans. To quantify the possible intraobserver variation, the scans were reevaluated by each the observer for a second time after a minimum 2-month period.

Statistical Analysis

The data were collected and maintained on the Microsoft Excel worksheets; they were then exported into the SPSS software for Windows (SPSS, Inc., Chicago, IL) for the final analysis.

Interobserver and intraobserver agreement and the corresponding 95% confidence intervals were calculated using the unweighted κ statistic or intraclass correlation, as appropriate.^{5,9} The κ statistic is the observed amount of

agreement adjusted for the amount of agreement expected by chance alone: κ is defined as $(Po - Pe)/(1 - Pe)$, where Po is the observed proportional agreement (number of all actual pair-wise agreements divided by all possible pair-wise agreements) and Pe is the proportion of agreement expected by chance. A κ value of 0 indicates agreement no better than chance. Values of 0 to 0.2 generally are considered to indicate poor agreement; 0.21 to 0.4, fair agreement; 0.41 to 0.6, moderate agreement; 0.61 to 0.8, substantial agreement; and 0.81 to 1, almost perfect agreement.¹⁴ Intraclass correlation was calculated for continuous variables.

RESULTS

A set of 43 CT scans was chosen from the STICH database of 1033 cases. The observers were not provided with any clinical information about the patients and were blinded to other reports on the scan and their first reading. The set of scans was rated independently by each observer, and the measurement was undertaken according to the preset criteria. When a scale was not present on the CT scan, either anteroposterior or biparietal diameters were measured, together with the similar anthropological measurement as approximated by the individual observer. The measured data were then entered onto Excel worksheets, and the actual values of each measure were calculated.

Each observer measured the length, breadth, height, depth, and midline shift of the hematoma on every scan. The actual values for length, breadth, depth, midline shift, and volume were then calculated using the formula integrated in the Excel worksheet. The volume was calculated according to the method advocated by Broderick, et al.⁴

TABLE 1
Guideline for the analysis of CT scans obtained in patients with ICH

A)	all dimensions are finally expressed in centimeters
B)	when available, the scale on the CT scan is used for converting the measurement on the film to the actual measurement
C)	length
i)	the slice in which the hematoma visually appears largest is chosen for measurement
ii)	the longest dimension is measured
iii)	streak or striplike extensions, if connected to the hematoma, are considered part of the hematoma; small specks, if distinctly separate from the main hematoma, are excluded
D)	breadth
i)	measured from the same slice as the length
ii)	the largest dimension of the hematoma at a rt angle to the plane of length measurement is taken
E)	depth
i)	the slice is used in which the edge of the hematoma is closest to the inner table of the skull & not necessarily the slice used for length & breadth
ii)	the shortest distance between the inner table of the skull & the nearest point on the hematoma is documented as depth
F)	midline shift
	the farthest point on the anterior interventricular septum/falx measured as perpendicular from a line joining the most anterior & posterior visible points on the falx is recorded as midline shift
G)	height
i)	the table position of the lowest slice in which the hematoma is 1st seen is taken as the 1st point
ii)	the table position of the first slice, after the highest slice in which the hematoma is seen, is taken as the 2nd point
iii)	the table position of the highest slice depicting the hematoma is sometimes used as the 2nd point if the clot was only visible as faint specks
iv)	the distance between the 2 points is documented as height of the hematoma
H)	volume
	the formula (length $\times$ breadth $\times$ height)/2 is used to calculate volume in milliliters
I)	hydrocephalus
i)	subjective impression of abnormal ventricular dilation is recorded as hydrocephalus
ii)	factors contributing to this impression include blood in the ventricles, grossly dilated temporal horns, ballooned or rounded frontal/occipital horns, periventricular lucency, & large ventricles in the absence of cortical atrophy
J)	location
	the sites documented separately have been frontal, temporal, parietal, & occipital lobes; the basal ganglia or its components & the internal capsule

TABLE 2
Interobserver and intraobserver intraclass correlation for the measurement of volume depth and midline shift*

Variable	Intraclass Correlations Between 1st and 2nd Ratings						Interrater ICC
rater no.	1	2	3	4	5	6	
volume	0.875	0.945	0.930	0.940	0.838	0.828	0.885
depth	0.948	0.987	0.982	0.980	0.979	0.946	0.955
midline shift	0.827	0.910	0.853	0.948	0.833	0.652	0.833

* ICC = intraclass correlation.

The continuous variables associated with the volume, depth, and midline shift were compared using intraclass correlation. Table 2 provides a summary of intraobserver intraclass correlation and interobserver intraclass correlation. Among the observers, both ratings showed an almost perfect agreement (intraclass correlation 0.81–1) in all but one case, in which it was 0.652 for Observer 6 in the assessment of midline shift. When the data were analyzed for intraclass correlation among the observers, there was a very high level of agreement.

The variation of the measurements among the observers was slightly higher when volume was compared with measured length as seen on the graphs in Fig. 1. This could be attributed to the involvement of four different measurements (length, breadth, height, and scale) for the calculation of volume. The graph in Fig. 2 upper shows the interobserver variability for measuring the depth and Fig. 2 lower depicts variability in midline shift.

The location of the hematoma was recorded as side, site of primary involvement and sites of secondary involve-

ment. The sites of involvement were to be labeled either as frontal, temporal, parietal, occipital, basal ganglia, internal capsule, or thalamus. Table 3 provides a summary of data regarding interobserver variability with respect to the side, location, intraventricular extension, and hydrocephalus. The interobserver variability for denoting the side had a κ value of 0.87. The κ values remained high (0.85 and 0.82, respectively) for involvement of basal ganglia/thalamus and intraventricular extension. There was relative interobserver disagreement (variability κ 0.44) in the assessment of hydrocephalus.

To quantify intraobserver variation, the scans were re-evaluated by each observer after a 2-month period. Table 4 shows the percentage of agreement between the two assessments conducted by the same observer (Observers 1–6) and its corresponding κ values. Analysis of the results showed a substantial-to-almost-perfect agreement (κ 0.61–1) for side, intraventricular extension, involvement of basal ganglia/thalamic sites, and primary involvement of the lobes for all observers; however, there was signifi-

Fig. 1. Graphs. Upper: Interobserver variability for hematoma volume. Triangles indicate data points with missing values; diamonds represent the mean value for the particular CT scan. Lower: The dispersion of measured length showing very little variability among observers.

Fig. 2. Graphs. Upper: Interobserver variability for measuring depth. Triangles indicate data points with missing values and diamonds represent the mean value for the particular CT scan. Lower: Interobserver variability for measuring midline shift. Triangles show data points with missing values and diamonds the mean value for the particular CT scan.

TABLE 3

*Interobserver agreement for location of the hematoma, intraventricular extension, and hydrocephalus**

Issue	No. of Observers Disagreeing With the Majority				% of Total Agreement	κ Value
	0	1	2	3		
side	38	0	5	0	88	0.87
intraventricular extension	33	8	2	0	77	0.82
hydrocephalus	22	10	10	1	51	0.44
primary site as bg/thal	36	4	2	1	84	0.85
primary site as lobar	31	10	1	1	72	0.78

* bg/thal = basal ganglia/thalamus.

cant variation in the rating of hydrocephalus for one observer (Observer 3) with a low κ value of 0.545. The values for the remaining observers showed substantial agreement (0.61–0.8). The Youden plot (Fig. 3) shows the intraobserver variation and distribution of the calculated volume measurements among all the six observers.

DISCUSSION

Our goal was to quantify the inter- and intraobserver variability with regard to assessment of CT scans. We quantified the variability among the observers on categorical parameters such as location and presence of hydrocephalus and continuous measurements were undertaken for volume, depth, and midline shift. Although the measurements conducted were simple, repeatable, and included only the basic parameters of the ICH, the results indicated the possible margins of variability within the same observer and among various raters. The study may be limited by the fact that the observer, although not provided with his previous reading, may still not be completely blinded to his own previous measurements. Additionally the set of scans is a cross-section obtained from all around the world and may not be truly pertinent to any individual center with a single scanner.

Intracerebral hemorrhage remains a common and devastating clinical problem. The value of CT scan measurements in prognostication with ICH is well known.¹⁸ The scan is also used to delineate the underlying pathological (vascular) causes of ICH.¹³

Kendall and Radue¹² have described the typical CT scan features of a recent intracerebral hematoma as a homogeneous region of increased attenuation surrounded by a well-demarcated relatively narrow low attenuation ring.

This area of surrounding hypodensity is currently considered as the zone of a possible penumbra, which may represent a recoverable part of the brain.¹⁷

There is a strong correlation with the location and site of the hematoma with the clinical deficits, comorbidity and the overall outcome. In a multivariate analysis,² sensory deficit was significantly associated with ICH in the thalamus; lacunar syndrome and hypertension with ICH in the internal capsule–basal ganglia; seizures, nonsudden stroke onset, and hypertension with lobar ICH; ataxia and sensory deficit with cerebellar ICH; cranial nerve palsy with brainstem ICH; and limb weakness, diabetes, and altered consciousness with multiple topographic involvement. Although the overall in-hospital mortality rate was 31%, it varied among sites: 65% for multiple topographic involvement, 44% for intraventricular ICH, and 40% for brainstem ICH to 16% for ICH in the internal capsule–basal ganglia.

Although there have been attempts to elucidate CT scanning and clinical criteria for conservative treatment of supratentorial traumatic intracerebral hematomas,²⁰ no definitive criteria exist for spontaneous intracerebral hematomas.

In addition to the clinical presentation, the treatment protocols for patients with these lesions are highly dependent on the ICH's size and location as well as the surgeon's experience.^{16,21} Although size or volume may influence surgery-related decisions,¹⁵ the actual surgical accessibility of the clot is deemed more important. Lobar hematomas and those located nearer to the cortical surface are much more likely to be surgically treated than those located deeper in the thalamus and putamen.⁸

Overall, it seems that neurosurgeons remain optimistic about the benefits of surgery, but most are more strongly influenced by the lesion's surgical accessibility than by other patient- or hemorrhage-related features.⁸

Extension of the bleeding into the ventricular system has been found to be associated with a 45% mortality rate compared with 9% in ICHs confined to brain parenchyma ($p < 0.01$).¹⁸ This has also been proven in a multivariate analysis.¹¹

In a previous study investigators concluded that hydrocephalus is an independent predictor of death after ICH.⁷ Thus, numerous clinical and radiological variables have been shown to be associated with survival and functional recovery following ICH. Clinical and radiological findings following ICH may assist rehabilitation specialists in the development of treatment goals, anticipating long-term patient care needs, and educating and training care-

TABLE 4

*Intraobserver variation for location of hematoma, intraventricular extension, and hydrocephalus**

Variable	Percentage of Agreement Between 1st and 2nd Ratings						κ Values					
	1	2	3	4	5	6	1	2	3	4	5	6
rater no.	1	2	3	4	5	6						
side	100	93	100	100	100	100	1.000	0.849	1.000	1.000	1.000	1.000
IVH	86	93	86	95	91	98	0.720	0.858	0.711	0.906	0.813	0.952
hydrocephalus	91	93	84	88	93	95	0.797	0.728	0.545	0.758	0.758	0.907
bg/thal sites	98	95	91	100	91	86	0.950	0.898	0.779	1.000	0.779	0.699
primary lobar site	95	88	91	100	86	95	0.900	0.755	0.801	1.000	0.720	0.905

* IVH = intraventricular hematoma.

Fig. 3. Graph demonstrating a comparison of intraobserver variability for calculated volume of all the six observers.

givers.⁶ One remaining controversy is the extent to which these factors influence patient outcome in cases of ICH involving the decision to undertake surgical or conservative treatment, and many authors have indicated the need for a large prospective randomized controlled trial to address these issues.¹⁰ This concern is being addressed by members of the STICH.

To devise definite prognosticating parameters for hematoma in STICH, there has to be a simple and methodical means of examining the CT scans. It is also important to determine the variability that could occur among observers who are experienced enough to interpret the scans and make a management decision.

CONCLUSIONS

In this study we quantify inter- and intraobserver variability in the evaluation of CT scans obtained in patients with ICH, when conducted systematically by using precise and preset criteria.

The absence of any such systematically conducted analysis of CT scans for ICH in the contemporary literature makes this study unique and its results a benchmark against which to measure future comparative studies in terms of observer reliability.

Acknowledgments

Dr. Bhattathiri received funding from a grant provided by the Medical Research Council. The Northern Brainwave Appeal provided Dr. Manjunath Prasad with a fellowship and equipment to permit measurement of ICH.

References

1. Anonymous: **Protocol 99PRT/7 International STICH (Surgical Trial in Intracerebral Haemorrhage)**. [http://www.

thelancet.com/info/info.isa?n1=authorinfo&n2=protocol+review&uid=14545] (Accessed September 24, 2003)

2. Arboix A, Comes E, Garcia-Eroles L, et al: Site of bleeding and early outcome in primary intracerebral hemorrhage. **Acta Neurol Scand** 105:282–288, 2002
3. Beldman TFJ, Rinkel GJE, Algra A: Prediction of long-term outcome after primary intracerebral haemorrhage: the importance of the site of the lesion. **Cerebrovasc Dis** 7:85–88, 1997
4. Broderick JP, Brott TG, Duldner JE, et al: Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality. **Stroke** 24:987–993, 1993
5. Cohen J: A coefficient of agreement for nominal scales. **Educ Psychol Meas** 20:37–46, 1960
6. Diamond P, Gale S, Stewart K: Primary intracerebral haemorrhage—clinical and radiologic predictors of survival and functional outcome. **Disabil Rehabil** 25:689–698, 2003
7. Diringer MN, Edwards DF, Zazulia AR: Hydrocephalus: a previously unrecognized predictor of poor outcome from supratentorial intracerebral hemorrhage. **Stroke** 29:1352–1357, 1998
8. Fernandes HM, Mendelow AD: Spontaneous intracerebral haemorrhage: a surgical dilemma. **Br J Neurosurg** 13:389–394, 1999
9. Fleiss JL, Cohen J, Everitt BS: Large sample standard errors of kappa and weighted kappa. **Psychol Bull** 72:323–327, 1969
10. Hardemark HG, Wesslen N, Persson L: Influence of clinical factors, CT findings and early management on outcome in supratentorial intracerebral hemorrhage. **Cerebrovasc Dis** 9:10–21, 1999
11. Karnik R, Valentin A, Ammerer HP, et al: Outcome in patients with intracerebral hemorrhage: predictors of survival. **Wien Klin Wochenschr** 112:169–173, 2000
12. Kendall BE, Radue EW: Computed tomography in spontaneous intracerebral haematomas. **Br J Radiol** 51:563–573, 1978
13. Laissy JP, Normand G, Monroc M, et al: Spontaneous intracerebral hematomas from vascular causes. Predictive value of CT compared with angiography. **Neuroradiology** 33:291–295, 1991
14. Landis JR, Koch GG: The measurement of observer agreement for categorical data. **Biometrics** 33:159–174, 1977
15. Mosdal C, Jensen G, Sommer W, et al: Spontaneous intracerebral haematomas. Clinical and computer tomographic findings and long-term outcome after surgical treatment. **Acta Neurochir** 83:92–98, 1986
16. Piotrowski WP, Rochowanski E: Operative results in hypertensive intracerebral hematomas in patients over 60. **Gerontology** 42:339–347, 1996
17. Siddique MS, Fernandes HM, Wooldridge TD, et al: Reversible ischemia around intracerebral hemorrhage: a single-photon emission computerized tomography study. **J Neurosurg** 96:736–741, 2002
18. Steiner I, Gomori JM, Melamed E: The prognostic value of the CT scan in conservatively treated patients with intracerebral hematoma. **Stroke** 15:279–282, 1984
19. Wermer MJ, Rinkel GJ, Van Rooij WJ, et al: Interobserver agreement in the assessment of lobar versus deep location of intracerebral haematomas on CT. **J Neuroradiol** 29:271–274, 2002
20. Wong CW: CT and clinical criteria for conservative treatment of supratentorial traumatic intracerebral haematomas. **Acta Neurochir** 135:131–135, 1995
21. Zderkiewicz E: [Spontaneous intracerebral hematomas.] **Neurol Neurochir Pol** 31:999–1011, 1997 (Polish)

Manuscript received August 19, 2003.

Accepted in final form September 16, 2003.

Address reprint requests to: A. David Mendelow, F.R.C.S., Ph.D., Department of Neurosurgery, Newcastle General Hospital, Westgate Road, Newcastle upon Tyne NE4 6BE, United Kingdom. email: Sanathana.Parameswaran@nuth.northy.nhs.uk.