Observer Error and Prediction of Outcome - Grading of Head Injury based on Computerised Tomography

J. H. HAVILL, J. W. SLEIGH, G. M. DAVIS, B. J. CHATTERTON, K. W. GILBERT, N. V. MARSH, D. A. KERSEL
Intensive Care Unit, Waikato Hospital, Hamilton, NEW ZEALAND

ABSTRACT

Objective: *To measure inter-observer error of a recently reported computerised tomography scoring system and to assess the ability of the scoring system to predict outcome in head injury patients.*

Methods: *Two radiologists independently graded all CT scans performed during the admission of all head injured patients. They were blinded to the clinical condition of the patient. Patients were followed up at 12 months and given a Glasgow outcome score. Outcomes were matched to the 2 independent assessments done on the first CT scan for each patient.*

Results: *A total of 123 head injury patients were studied. For the diffuse injury categories, there were 410 gradings made. Of these, 32% differed by at least one category. Where at least one of the radiologists identified non-evacuated mass lesions there were 148 gradings. Of these, one radiologist reported an un-evacuated mass lesion in 47%, which was not reported by the other.*

The first CT scan was evaluated on 119 patients. Using the Chi-Squared test, the diffuse injury IV category was the only one to show a strong relationship with outcome as measured by the Glasgow outcome score.

Conclusions: *The prediction of outcome for head injury patients based on CT scans has significant shortcomings. In our study, there was significant variation in grading by experienced radiologists. The separate categories were also poor predictors of outcome at 12 months except for diffuse injury IV. The classification of mass lesions needs modification to be useful.*(**Critical Care and Resuscitation 2001; 3: 15-18**)

Key words: Head injury, observer error, prediction of outcome, computerised tomography score

In 1991, a new classification of head injury based on computerised tomography was introduced following the examination of the initial CT scans and the Glasgow Outcome Scores at discharge from hospital of over 700 severe head injury patients (Table 1).[1] The researchers utilised the status of the mesencephalic cisterns, the degree of midline shift in millimetres, and the presence or absence of one or more surgical masses on the initial CT scan, and correlated the observation to early outcome. They concluded that this method of classification allowed a useful prediction of outcome.

Although considerable attention was applied to avoiding inter-observer error, no information was given to support the accuracy of the gradings from different radiologists. Since this time, various groups have begun

Correspondence to: Associate Professor. J. H. Havill, Intensive Care Unit, Waikato Hospital, Hamilton, New Zealand (e-mail: havillj@hwl.co.nz)

to use this classification to help characterise their cohorts of traumatic brain injury.

Table 1. Diagnostic categories of types of abnormalities visualised on CT scanning

Category	*Definition*
Diffuse Injury I	No visible pathology seen on CT scan
Diffuse Injury II	Cisterns are present with midline shift 0 - 5 mm and/or: lesion densities present no high or mixed density lesion > 25 cc may include bone fragments and foreign bodies
Diffuse Injury III (swelling)	Cisterns compressed or absent with midline shift 0-5 mm, no high or mixed lesion > 25 cc
Diffuse Injury IV (shift)	Midline shift > 5 mm, no high or mixed density lesion > 25 cc
Evacuated mass lesion	Any lesion surgically evacuated
Non-evacuated mass lesion	High or mixed density lesion > 25 cc, not surgically evacuated

As part of a multifaceted observational study of a cohort of traumatic brain injury patients at Waikato Hospital,[2,3,4,5,6] we prospectively used this grading classification, tested the inter-observer error and correlated the findings to outcome at 12 months.

MATERIALS AND METHODS

Patients admitted to the Waikato Hospital intensive care unit (ICU) between 1993 and 1996 with severe brain trauma were studied prospectively. The study criteria included patients with brain trauma who were expected to require sedation and controlled ventilation for more than 24 hr as part of the management of the head injury. The clinical management of the patients has been previously described,[2,8] and correlated substantially to the guidelines in the 1996 consensus report on management of head injury.[8] Consent was gained from the next of kin and ethical approval was given by the Waikato Ethical Committee.

As part of this study, two independent specialist radiologists assessed each CT scan and graded them using the scale shown in table 1. They were unaware of the patient details and outcomes. Altogether there were 3 radiologists involved, as one left the project before completion. Each radiologist was given the original publication describing the grading system (Table 1),[1] but limited cross-checking was done between the individuals as we wanted to check the robustness of the classification as a clinical tool. All available CT scans performed during the patient's stay in ICU were graded.

Patients were followed up at 6 and 12 months and detailed observations of their psychosocial recovery were performed. The Glasgow Outcome Score formed part of a battery of tests and was used to characterise the outcome at 12 months. These tests were carried out by one of the authors (DK) and a Masters Student under the supervision of this author. The outcomes were compared with the CT gradings of the first CT scan.

RESULTS

We studied 123 patients with severe brain trauma (median post-resuscitation GCS score of 6, range 3-11). The time between the accident and the first CT scan were 0-1 hr (2%), > 1 - 2 hr (21%), > 2 - 3 hr (31%), > 3 - 4 hr (14%) and > 4 hr (32%). Scans on 4 patients were lost. There were a total of 205 scans (410 gradings) where no mass lesions were seen on the CT scan by either radiologist, and therefore the scan was classified as 'diffuse injury' (DI) I, II, III, or IV (Table 2). The classification of 65 scans (130 gradings) differed by 1 category and 4 scans (8 gradings) differed by 2 categories (i.e. 32% differed by one category and 2% differed by 2 categories).

Table 2. Inter-observer error between two radiologists while grading CT scans. Diffuse Injury I, II, III, and IV categories*

	Number of Assessments		*Percentage*
Same grading	272	(136 pairs)	66%
Difference by 1 category	130	(65 pairs)	32%
Difference by 2 categories	8	(4 pairs)	2%
Total	410	(205 pairs)	100%

* where neither radiologist scored a mass lesion

There were 74 scans (148 gradings) where a non-evacuated mass was recorded by at least one of the radiologists, shown in Table 3. The observers differed in grading 35 scans (70 gradings) i.e. 47%. Where the differing radiologist did not score the presence of a non-evacuated mass, the alternative scores included: DI II (9), DI III (11) and DI IV (11) (the remainder were scored as evacuated masses).

Table 4 uses a similar format to that described by Marshall *et al*,[1] except the outcomes were recorded at 12 months rather than at discharge. The total combined evaluations from both radiologists on the first scans are recorded - thus each patient is represented twice. The scans on 4 patients were lost and therefore 8 assessments could not be performed. This left 238 gradings on the first CT scans of 119 patients for analysis. Ten patients were lost to follow up at 12 months. Hence, 218 comparisons between the scan gradings and the outcomes were made.

Table 3. Inter-observer error between two radiologists while grading CT Scans. Non-evacuated mass categories*

	Number of Assessments		*Percentage*
Same grading	78	(39 pairs)	53%
Different grading**	70	(35 pairs)	47%
Total	148	(74 pairs)	100%

* where at least one radiologist scored a non-evacuated lesion.
** where only one radiologist scored a non-evacuated mass, the alternative scores included: Diffuse Injury II (9), Diffuse Injury III (11), Diffuse Injury IV (11).

Using the chi-squared test for homogeneity, there was an overall significant relationship between outcome and category ($p = 0.0002$). However, one quarter of the total chi-squared value was derived from the cell with the excess deaths in the DI IV group. The DI IV column was the only column that had a significant chi-squared value. Thus although a statistically significant relationship between the CT scores and the outcome was recorded, the utility of this in clinical prediction was weak, as 75% of the patients were in the DI categories II and III.

The above results refer to overall outcomes by Glasgow outcome score (GOS) category. Mortality increased over the DI categories: DI I (4.3%); DI II (8.4%); DI III (28%); DI IV (56%). The non-evacuated lesion mortality was 23%. However the good/moderate GOS were; DI I 69%, DI II 55%, DI III 41%, DI IV 37% and non-evacuated mass 50%.

DISCUSSION

The radiologists involved in this study are all experienced specialists who regularly evaluate CT scans. They are regarded as competent by their peers and work in a large department where they constantly review each other's work. Our results show that competent radiologists will vary in their grading of DI, using the classification as shown in Table 1, by at least 1 category in almost one third of cases. They also vary in their assessment of mass lesions. In our study, one radiologist recorded a non-evacuated mass lesion in 47% of assessments where the other radiologist did not. In almost all of these gradings the alternative radiologist chose a DI category.

Furthermore, we did not find that the first CT scan gradings were particularly helpful in predicting outcome. There was a tendency for those with normal CT scans to approach statistical significance. Although only one this group died, and none were vegetative, the Glasgow outcome scores were distributed evenly over the moderate and severe outcome groups. A similar distribution was also noted by Marshall *et al.*[1] The groups DI II and DI III did not show a statistically significant relationship with a wide spread of outcomes from good to dead. The DI IV group was the only one to show a statistical relationship, however the numbers were small. Yet, clinicians who deal regularly with head injured patients would agree that the absence of mesencephalic cisterns is likely to result in a bad outcome, and thus the statistical result is supported in clinical practice.

It is important to realise that the statistics determined all outcomes as measured by the GOS and the mortality did increase steadily as the DI categories increased. But when looking at other outcomes, it was clear that, even in

Table 4. Outcome at 12 months in relation to first CT scan gradings.*

GOS	Diffuse injury I	Diffuse injury II	Diffuse injury III	Diffuse injury IV	Evacuated mass	Non-evacuated mass
Good	10	47	16	4	1	16
Moderate	6	12	7	2	0	1
Severe	3	27	15	1	0	6
Vegetative	0	0	0	0	0	0
Dead	1 (4.3%)	9 (8.4%)	16 (28%)	9 (56%)	1	8 (23%)
Lost to follow-up	3	12	2	0	0	3
Total	23	107	56	16	2	34

* each patient had two independent gradings of the first CT scan. Four CT scans were missing, leaving a total of 238 gradings.
GOS = Glasgow outcome score

the worst categories (DI III and DI IV), there was an approximately 40% good/moderate outcome. Thus, when counseling relatives of a trauma victim one has a large chance of being incorrect if the CT is used as a prognostic tool.

The DI categories II and III are likely to have within them the patients who have experienced significant hypoxia as well as diffuse axonal injury. Thus pre-hospital aspects such as, time to resuscitation, time to get to hospital, presence of hypoxia at the accident scene, and mechanism of injury could well be important in causing variation in outcome. We have described the pre-hospital aspects of our series in a previous publication.[2] Our patients came from a predominantly rural area where over 60% were injured more than 30 kilometres from the first admission hospital, and 43% took more than 90 min to arrive in hospital. Hypoxia was deemed to be present in 50% of the patients at the accident scene. Most of the accidents occurred from high speed vehicle accidents.

Another variation factor in assessing the first CT scan was the time between the accident and the first CT scan. A very early scan can often miss the extent of eventual swelling, and it is well recognised that bleeding often becomes more apparent on subsequent scans. Our times varied between 0-1 hr and > 4 hr, although more than 75% were performed after 2 hr delay.

In our series, the presence of a non-evacuated mass was not predictive of a bad outcome, as more patients had a good outcome than bad outcome as assessed by the GOS. This did not correspond to the results in the original paper by Marshall *et al.*[1] They had only 100 non-evacuated masses compared with 276 evacuated masses where their evacuated mass classification referred to a surgically evacuated mass lesion. However, we found that some mass lesions (e.g. extradural, subdural), if removed expeditiously, do not lead to bad outcomes. On the other hand, large intra-cerebral masses of blood do have a bad outcome and often lead to death, even if evacuation takes place. Because 34 of our first CT scan gradings were non-evacuated masses, and there were only 2 gradings where the mass had already been evacuated, we were unable to compare our results with those in the original article by Marshall *et al.*[1] However, we do believe that it is not appropriate to add all masses into the categories as described, because extradurals and subdurals probably have a much better prognosis than intracerebral bleeding where amounts are over 25 mL.

In conclusion, our results suggest that the classification of head injury based on computerised tomography as described by Marshall *et al*[1] has significant shortcomings. There is significant variation in grading by experienced radiologists. The categories are poor predictors of outcome except for DI IV and the classification of all mass lesions into one group is simplistic and probably inappropriate.

Acknowledgments

This work was supported by the Neurological Foundation of New Zealand (Grant No. 35/92/135, to J Havill et al), and the Waikato Medical Research Foundation. We wish to also acknowledge the help of Mary Lapine, Research Assistant.

Received: 17 January 2001
Accepted: 12 February 2001

REFERENCES

1. Marshall LF, Marshall SB, Klauber MR, et al. A new classification of head injury based on computerized tomography. J Neurosurg 1991;75:S14-S20.
2. Havill JH, Sleigh JW, Kersel DA, Marsh NV. Prehospital treatment of head injuries. Can we do better? Emerg Med 1998;10:123-128.
3. Havill JH, Sleigh JW, Kersel DA, Marsh NV. Profile and costs of head injuries admitted to the Waikato Hospital. NZ Med J 1998;111:161-163.
4. Marsh NV, Kersel DA, Havill JH, Sleigh JW. Caregiver burden at 6 months following severe brain injury. Brain Injury 1998;12:123-128.
5. Marsh NV, Kersel DA, Havill JH, Sleigh JW. Caregiver burden at one year following severe traumatic brain injury. Brain Injury 1998;12:1045-1059.
6. Sleigh JW, Havill JH, Kersel DA, Marsh NV, Frith RW, Ulyatt D. Somatosensory evoked potentials in severe traumatic brain injury: a blinded study. J Neurosurgery 1999;91:577-580.
7. Havill JH, Sleigh JW. Management and outcomes of patients with brain trauma in a tertiary referral trauma hospital without neurosurgeons on site. Anaesth Intensive Care 1998;26:642-647.
8. Anonymous. Guidelines for the management of severe head injury. J Neurotrauma 1996;13:643-645.