



Validation of a New Coma Scale, the FOUR Score, in the Emergency Department

Latha G. Stead · Eelco F. M. Wijdicks · Anjali Bhagra ·
Rahul Kashyap · M. Fernanda Bellolio · David L. Nash ·
Sailaja Enduri · Raquel Schears · Bamlet William

© Humana Press Inc. 2008

Abstract

Objective Full Outline of Unresponsiveness (FOUR) score has previously been validated scale in the Neurosciences Intensive Care Unit. In this study, we sought to validate the use of FOUR score in the emergency department (ED) using non-neurology staff. We also compared its performance to the Glasgow Coma Scale (GCS) and correlated it to functional outcome at hospital discharge and overall survival.

Methods We prospectively rated 69 patients with initial neurologic symptoms presenting to the ED. Three types of examiners performed the FOUR score: ED physician, ED resident, and ED nurse. Patients were followed through hospital discharge; functional outcome was measured using modified Rankin Score (mRS).

Results Interrater reliability for FOUR score and GCS was excellent (respectively, $\kappa_w = 0.88$ and 0.86). Both FOUR score and GCS predicted functional outcome, and

overall survival with and without adjustment for age, sex, and alertness group.

Conclusion The FOUR score can be reliably used in the ED by non-neurology staff. Both FOUR score and GCS performed equally well, but the neurologic detail incorporated in the FOUR score makes it more useful in management and triage of patients.

Keywords FOUR score · Neurologic · Glasgow Coma Scale · Consciousness

Introduction

Coma scales have been created to improve communication between providers and have been used to triage patients with impaired consciousness in and out of the Emergency Department (ED). The Glasgow Coma Scale (GCS) was originated in a Neurosurgical Intensive Care Unit, but found its way elsewhere, and became a standard scale used in the field by first responders, emergency physicians, and neuroscience specialists [1]. Over the years, considerable limitations have been identified on this scale: crucial parts of the neurologic examination of a patient with impaired consciousness were not included (e.g., brainstem reflexes and eye movements) and language evaluation—largely an assessment of orientation rather than consciousness—became useless in intubated patients. More concerning, the performance of the GCS in the ED has mixed results. Recently, pre-hospital GCS scores were compared with its assessment in the ED, and poor agreement was found in patients with traumatic head injury and GCS sum scores <13 points [2]. This disagreement between emergency medical service and emergency physicians confirmed a number of earlier studies that found only good agreement

L. G. Stead (✉) · R. Kashyap · M. F. Bellolio ·
D. L. Nash · S. Enduri · R. Schears
Department of Emergency Medicine, Mayo Clinic, Generose
G 410, 200 First Street, Rochester, MN 55905, USA
e-mail: stead.latha@mayo.edu

E. F. M. Wijdicks
Division of Critical Care Neurology, Mayo Clinic,
Rochester, MN, USA

A. Bhagra
Department of Internal Medicine, Mayo Clinic,
Rochester, MN, USA

B. William
Department of Biostatistics, Mayo Clinic, Rochester,
MN, USA

in alert or near alert (GCS 13–15) patients [3, 4]. Other observational studies found concerning disagreements between ED physicians and ED physicians and nurses when rating consciousness using the GCS [5, 6], and marked differences between level I trauma centers in calculations of the GCS were found [7].

There has been perceived need for an improved coma scale [8]. Such a scale would provide the essentials of a neurological examination and should in a better way indicate the depth of coma. The Full Outline of Unresponsiveness (FOUR) score is a recently developed and validated [9, 10] in the Neurointensive Care Unit and can supplement the GCS [11]. The FOUR score consists of four components—eye, motor, brainstem, and respiration—and each component has a maximal score of 4 (Fig. 1). The Four score has been validated using neurologists and neuroscience nurses as raters, but no studies have investigated its validity outside neurology staff. As use of the FOUR score becomes more widespread, we sought to test its validity in the ED setting and compare its performance to the GCS and correlate it to functional outcome at hospital discharge.

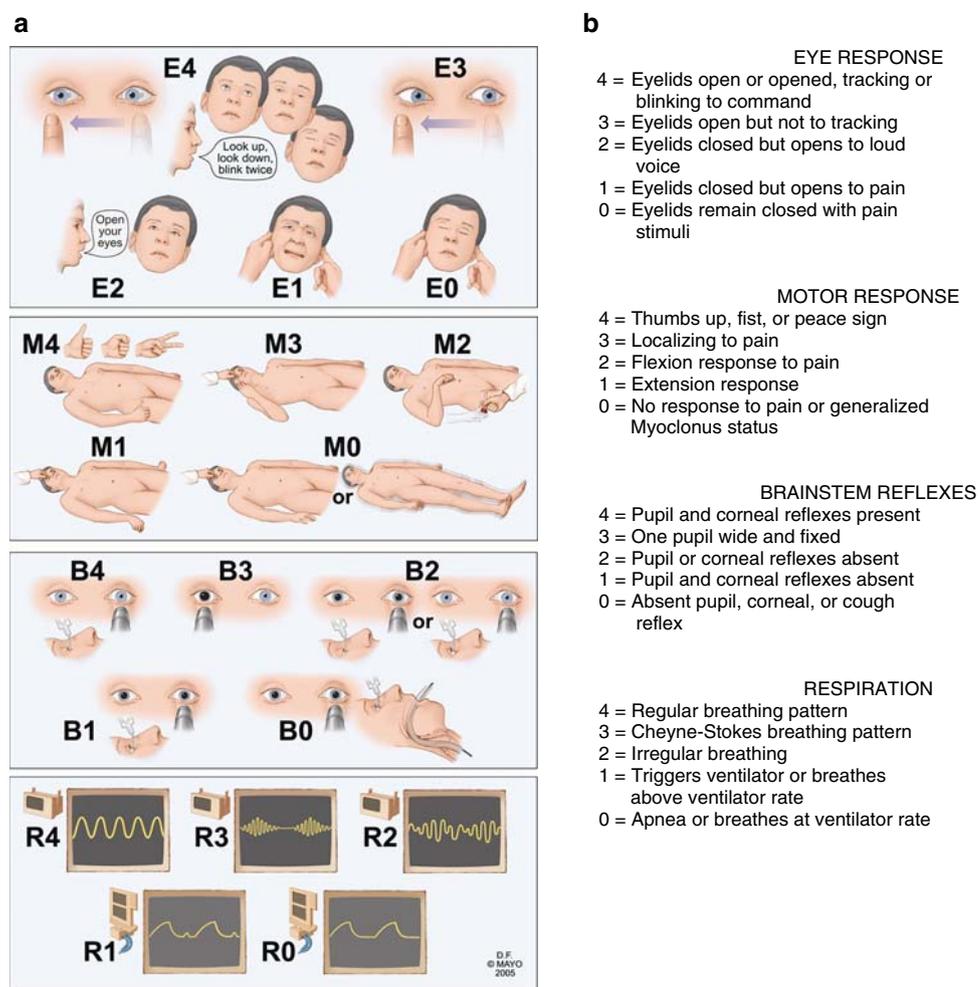
Methods

This is an observational study in adult patients presenting with acute neurologic disease to an ED with an annual census of 79,000.

The study was designed to enroll 120 patients sampled from all four alertness group categories—30 alert patients, 30 comatose, and 60 drowsy/stuporous patients. The raters were selected from three different training type groups (ED physicians, ED residents, and ED nurses). Raters were provided a one-page handout with written instructions describing both the FOUR score and the GCS and were asked to grade a few patients using both the GCS and the FOUR score scale. Written instructions and a scoring sheet were used by each rater during examination of all patients.

The 120 subjects would be comprised of 20 patients scored for each combination of training type (nurse/nurse, nurse/resident, nurse/physician, resident/resident, resident/physician, and physician/physician). Each patient was rated on both scales by two different raters, who performed their examination within 10 min of each other without knowledge of the other's scores. The order of the evaluations was

Fig. 1 Four score visual aid



randomized to reduce bias. (For instance, of the 20 subjects rated by both a resident and a physician, 10 were rated by the resident first followed by physician, and 10 rated by the physician first.) This design allowed us to evaluate both intra-training and inter-training type reliability. We did not evaluate *intra*-rater reliability since it would not be possible for a single rater to score a patient at two time points sufficiently close in time to ensure the patient has not changed status without remembering their previous scores.

The study was stopped early after 1 year of recruitment with only 69 patients having been evaluated because of slow recruitment. For the purposes of this manuscript, the cohort was a convenience sample of 69 adult patients presenting to the ED with neurologic signs and symptoms during a 12-month period. Patients with altered consciousness were categorized as alert, or non-alert. Patients were prospectively followed through hospital discharge, and their functional outcome measured via the Modified Rankin score (mRS). The mRS is a 6-point score that measures functional outcome [12].

This study was approved by the Emergency Medicine Research Committee and Mayo Foundation Institutional Research Board. All patients were prospectively enrolled and provided informed consent.

Statistical Analysis

For both the FOUR score and GCS, overall average weighted kappa scores were calculated to determine the degree of agreement between raters. κ_w (weighted kappa) of 0.4 or less is considered poor, values between 0.4 and 0.6 are considered fair to moderate, κ_w between 0.6 and 0.8 suggest good observer agreement, and κ_w values >0.8 suggest excellent agreement [13, 14]. In order to assess the level of agreement between raters, we calculated the absolute difference in total FOUR (and GCS) score between the two raters for each of the 69 patients. We used the Kruskal–Wallis test to compare total FOUR (and GCS) score across the three levels of rater group. We investigated prediction of in-hospital mortality and morbidity using a logistic regression model controlling for age, sex, alertness group, and either total FOUR score or total GCS score. The relationship between total score for each scale and overall survival was investigated using Cox proportional hazards

regression models, and analyses for both scales considered unadjusted models as well as models which adjusted for age, sex, and alertness group.

Results

The study cohort comprised 69 patients, with the following neurologic complaints: Suspicion CNS infection (1%); stroke (22%); seizure (14%); subarachnoid hemorrhage (3%); altered consciousness or encephalopathy (51%), and traumatic head injury (9%).

There were 32 alert and 44 non-alert patients. Thirty-one percent of the evaluations were done by nurses, 36% by residents, and 33% by physicians. There was no difference in the total score assigned by nurses, residents, or physicians for either the FOUR score ($P = 0.777$) or the GCS ($P = 0.125$), with a median FOUR score 16, 16, and 16, and a median GCS of 14.5, 14, and 14.5 for nurses, residents, and physicians, respectively. The within subject differences between the first and second rater total scores were small for both scores. The median absolute value of the difference for both the FOUR and GCS scores was equal to 0 (mean \pm SD, respectively, 0.36 ± 0.76 , 0.46 ± 0.87).

Table 1 presents the rater agreement characteristics for both scales. The overall weighted kappa for the FOUR score was 0.882 and for the GCS 0.862. Intra-class correlation coefficients were 0.975 and 0.964, respectively.

Table 2 presents the relationships between total score for each of the two scales and the outcomes of in-hospital death and mRS. Considering the FOUR scale total score, we see that for every 1-point increase in total score there is a 0.67 (95% CI = 0.53–0.84) times lower risk of experiencing in-hospital mortality under the unadjusted model. A 1-point increase in total score was related to a better functional outcome (OR = 0.43, 95% CI = 0.26–0.71) when we consider all poor outcomes defined as Rankin scores of 3, 4, 5, and 6. Considering the GCS scale total score, we see that for every 1-point increase in total score there is a 0.68 (95% CI = 0.56–0.83) times lower risk of experiencing in-hospital mortality under the unadjusted model. A 1-point increase in total score also has a protective effect (OR = 0.49, 95% CI = 0.32–0.75) when we

Table 1 Rater agreement—weighted kappa and intra-class correlation coefficients (ICC)

	FOUR score					GCS			
	Eye	Respiration	Brainstem	Motor	Total	Eye	Verbal	Motor	Total
Weighed kappa	0.859	0.756	0.943	0.825	0.882	0.871	0.869	0.807	0.862
ICC	0.945	0.843	0.969	0.926	0.975	0.934	0.921	0.911	0.964

Adjusted for age, sex, and alertness group

Table 2 FOUR score and GCS relation to functional outcome (mRS)

Outcome (N = 69)	N	FOUR score			GCS		
		OR	95% CI	P-value	OR	95% CI	P-value
In-hospital death	9	0.67	0.53–0.84	<0.001	0.68	0.56–0.83	<0.001
mRS 3–6							
Unadjusted	25	0.43	0.26–0.71	<0.001	0.49	0.32–0.75	0.001
Adjusted ^a	25	0.59	0.36–0.95	0.029	0.61	0.41–0.92	0.018

^a Adjusted for age, sex, and alertness group

mRS, modified Rankin Score

Table 3 FOUR score and GCS relation with overall survival

Outcome (N = 69)	N	FOUR score			GCS		
		HR	95% CI	P-value	HR	95% CI	P-value
Unadjusted	25	0.84	0.79–0.89	<0.001	0.80	0.73–0.87	<0.001
Adjusted ^a	25	0.93	0.84–1.01	0.101	0.88	0.78–1.00	0.050

^a Adjusted for age, sex, and alertness group

consider all poor outcomes defined as Rankin scores of 3, 4, 5, and 6. Both relationships remain after adjusting for age, gender, and alertness group.

Table 3 presents the relationships between total score for each of the two scales and overall survival. Considering the FOUR scale total score, we see that for every 1-point increase in total score there is a 0.84 (95% CI = 0.79–0.89) times lower risk of mortality. After adjusting for age, sex, alertness group, HR = 0.93 (95% CI = 0.84–1.01) which is no longer significant at the 0.05 level. Considering the GCS scale total score, we see that for every 1-point increase in total score there is a 0.80 (95% CI = 0.73–0.87) times lower risk of mortality. After adjusting for age, sex, alertness group, HR = 0.88 (95% CI = 0.78–1.00) which is no longer significant at the 0.05 level.

Table 4 presents a summary of within subject differences for the FOUR score and the GCS by evaluator type (including ED nurse, ED attending physician, and ED resident) and alertness group (alert versus drowsy/stuporous/comatose). The within subject differences between raters appear to be comparable across evaluator type combinations for both the FOUR score (mean difference ranging from 0.10 to 0.67) and for the GCS (mean difference ranging from 0.0 to 1.0). Mean within subject differences for non-alert patients were higher for both FOUR score and GCS (0.64 and 0.72, respectively) then were observed for alert patients (0.03 and 0.15, respectively).

Table 4 FOUR score and GCS within subject differences by evaluator type and alertness group

	N	FOUR score		GCS	
		Mean	SD	Mean	SD
<i>Evaluator type</i>					
Nurse/nurse	10	0.10	0.31	0.00	0.00
Nurse/resident	10	0.40	0.69	0.50	0.70
Nurse/physician	13	0.46	0.96	0.69	0.94
Resident/resident	12	0.67	0.88	1.00	1.27
Resident/physician	15	0.33	0.89	0.40	0.82
Physician/physician	9	0.11	0.33	0.00	0.00
<i>Alertness group</i>					
Alert	32	0.03	0.17	0.15	0.57
Non-alert	37	0.64	0.94	0.72	0.99

Discussion

This is the first study of the FOUR score outside the Neurosciences Intensive Care Unit using non-neurology staff as raters. The advantages of the FOUR score have been outlined previously [9, 10]. This new coma scale includes important clinical neurological findings in patients with impaired consciousness and this study shows that can be assessed by emergency physicians, residents, and nurses in the ED with excellent agreement. Our raters with no specific neurological training were able to identify key neurologic signs in patients with impaired consciousness. Furthermore, this study confirmed prior studies that the FOUR score is a robust predictor of in-hospital mortality, functional outcome at hospital discharge, and overall survival in patients seen for neurologic complaints [9, 10].

The GCS has remained the “gold standard” for assessment of impaired consciousness in all patient populations. Studies in the ED have not only involved validation of the scale, but also attempts at modifications (e.g., simplified motor scale) eliminating the eye and verbal response [15]. We would argue that further simplification of the GCS diminishes neurologic assessment despite better interrater reliability. Our study also shows that the GCS has an excellent interrater agreement and at least in our ED is much higher than prior reports. This can be explained by inclusion of a high proportion of alert patients because most validation studies perform worse in patients with marked decline in consciousness.

The FOUR score was developed to fill in a need for an easy to use rapid assessment of all essential neurologic signs in patients with impaired consciousness. It ignores disorientation or confusion used in the verbal scale, but provides a good assessment of eye movements, brainstem reflexes, and respiratory drive in ventilated patients. The FOUR score has the potential to recognize a locked-in

syndrome, uncal herniation, brain death, and less severe neurologic injury. A more comprehensive assessment of a patient with an impaired consciousness could assist in initial decision making, assess the need for additional consultation (neurosurgeon) and more effectively triage patient to the most appropriate Intensive Care Unit, neuro-radiology suite, or operating theater.

Limitations

One of the limitations was that the target enrollment cohort was not reached, and approximately half of the studied patient population included alert patients. This increases the chance of interobserver agreement because no neurologic abnormality will have to be identified. A study of a larger group of stuporous or comatose patients would be desirable. However prospective scale validation studies are very difficult to perform in the ED environment with a diverse population of patients and varying work schedules of potential raters. Such a study is easier to perform in a neurological Intensive Care Unit with patients with acute neurologic disease.

This was a single center study, so the generalizability to other EDs has not been yet proved.

References

1. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet*. 1974;2:81–4. doi:[10.1016/S0140-6736\(74\)91639-0](https://doi.org/10.1016/S0140-6736(74)91639-0).
2. Kerby JD, MacLennan PA, Burton JN, McGwin G Jr, Rue LWIII. Agreement between prehospital and emergency department Glasgow coma scores. *J Trauma*. 2007;63:1026–31.
3. Menegazzi JJ, Davis EA, Sucov AN, Paris PM. Reliability of the Glasgow Coma Scale when used by emergency physicians and paramedics. *J Trauma*. 1993;34:46–8. doi:[10.1097/00005373-199301000-00008](https://doi.org/10.1097/00005373-199301000-00008).
4. Báez AA, Giráldez EM, De Peña JM. Precision and reliability of the Glasgow Coma Scale score among a cohort of Latin American prehospital emergency care providers. *Prehosp Disaster Med*. 2007;22:230–2.
5. Gill MR, Reiley DG, Green SM. Interrater reliability of Glasgow Coma Scale scores in the emergency department. *Ann Emerg Med*. 2004;43:215–23. doi:[10.1016/S0196-0644\(03\)00814-X](https://doi.org/10.1016/S0196-0644(03)00814-X).
6. Holdgate A, Ching N, Angonese L. Variability in agreement between physicians and nurses when measuring the Glasgow Coma Scale in the emergency department limits its clinical usefulness. *Emerg Med Australas*. 2006;18:379–84. doi:[10.1111/j.1742-6723.2006.00867.x](https://doi.org/10.1111/j.1742-6723.2006.00867.x).
7. Buechler CM, Blostein PA, Koestner A, Hurt K, Schaars M, McKernan J. Variation among trauma centers' calculation of Glasgow Coma Scale score: results of a national survey. *J Trauma*. 1998;45:429–32. doi:[10.1097/00005373-199809000-00001](https://doi.org/10.1097/00005373-199809000-00001).
8. Wijdicks EF. Clinical scales for comatose patients: the Glasgow Coma Scale in historical context and the new FOUR score. *Rev Neurol Dis*. 2006;3:109–17.
9. Wijdicks EF, Bamlet WR, Maramattom BV, et al. Validation of a new coma scale: The FOUR score. *Ann Neurol*. 2005;58:585–93. doi:[10.1002/ana.20611](https://doi.org/10.1002/ana.20611).
10. Wolf CA, Wijdicks EF, Bamlet WR, et al. Further validation of the FOUR score coma scale by intensive care nurses. *Mayo Clin Proc*. 2007;82:435–8.
11. Sternbach GL. The Glasgow Coma Scale. *J Emerg Med*. 2000;19:67–71. doi:[10.1016/S0736-4679\(00\)00182-7](https://doi.org/10.1016/S0736-4679(00)00182-7).
12. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scott Med J*. 1957;2(5):200–15.
13. van Swieten JC, Koudstaal PJ, Visser MC, et al. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*. 1988;19:604–7.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biostatistics*. 1977;33:159–74.
15. Haukoos JS, Gill MR, Rabon RE, Gravitz CS, Green SM. Validation of the simplified motor score for the prediction of brain injury outcomes after trauma. *Ann Emerg Med*. 2007;50:18–24. doi:[10.1016/j.annemergmed.2006.10.004](https://doi.org/10.1016/j.annemergmed.2006.10.004).