Further Validation of the FOUR Score Coma Scale by Intensive Care Nurses

CHRIS A. WOLF, RN; EELCO F. M. WIJDICKS, MD; WILLIAM R. BAMLET, MS; AND ROBYN L. MCCLELLAND, PHD

OBJECTIVE: **The FOUR (Full Outline of UnResponsiveness) score is a new coma scale that consists of 4 components (eye, motor, brainstem, and respiration). The scale was recently validated, but variability among nursing staff has been documented.**

PATIENTS AND METHODS: **We prospectively studied the FOUR score in 80 patients with acute neurologic disease in an intensive care unit (ICU) and compared it with the Glasgow Coma Scale (GCS) using 20 experienced and inexperienced neuroscience ICU nurses and nonneuroscience ICU nurses. Each nurse was trained with the use of video examples and instruction cards. Each patient was rated by 2 nurses, with the order randomly assigned.**

RESULTS: **The rater agreement was good to excellent with the FOUR score (weighted κ: eye, 0.84; respiration, 0.92; brainstem, 0.89; and motor, 0.73) and similar to that for the GCS (weighted κ: eye, 0.85; verbal, 0.89; and motor, 0.74). Greater average experience in years was associated with less disagreement, but the difference was not statistically significant.**

CONCLUSION: **The FOUR score provides more neurologic information than the GCS. The FOUR score can be used by any ICU nurses, even those with minimal experience.**

***Mayo Clin Proc.* 2007;82(4):435-438**

CI = confidence interval; FOUR = Full Outline of UnResponsiveness; GCS = Glasgow Coma Scale; ICU = intensive care unit; OR = odds ratio

The FOUR (Full Outline of UnResponsiveness) score was recently developed and validated and is an alternative to the Glasgow Coma Scale (GCS). The FOUR score consists of 4 components—eye, motor, brainstem, and respiration—and each component has a maximal score of 4. A low FOUR score is associated with in-hospital mortality and disability in patients with acute brain injury.[1] Because only 3 intensive care unit (ICU) nurses rated patients in our prior study, we studied both neuroscience ICU nurses' and nonneuroscience ICU nurses' variability with a larger rater pool. We also tested the effect of experience on rater variability using the FOUR score and GCS.

PATIENTS AND METHODS

All raters were ICU nurses at Saint Marys Hospital in Rochester, Minn. Ten were experienced neuroscience ICU nurses, 5 were inexperienced neuroscience ICU nurses, and 5 were ICU nurses with no neuroscience training. In this study, raters were considered experienced if they had a minimum of 2 years of neuroscience experience. Inexperienced neuroscience nurses were defined as nurses within 6 months of graduation or nurses assigned in medical or surgical ICUs with virtually no exposure to monitoring acutely ill neurologic patients. Among the experienced nurses, the median year in the neurological ICU was 4.5 years (range, 2-27 years). With the use of video examples and instruction cards, each nurse was trained for 20 to 30 minutes and shown a patient example. Each nurse was allowed to practice on 1 to 2 patients while being supervised by one of the authors (E.F.M.W.). Patients aged 18 years and older who were admitted to the neuroscience ICU and patients seen in consultation and admitted to medical and surgical ICUs within 24 hours of admission were included in the study. Patients taking sedative agents that could not be temporarily discontinued were excluded.

Twenty patients in each conscious category (alert, drowsy, stupor, coma) were studied (total, 80 patients). Each patient was scored by 2 nurses, with the order randomly assigned. We blocked the randomization by consciousness group as in our prior study.[1] Each rater was provided with a handout with a FOUR score (Figure 1) and GCS description, and nurses assessed the patient within 1 hour of each other. Raters were blinded to other ratings and were not aware of the diagnosis of the patient. Outcome was assessed at 30 days after admission using the Rankin scale to measure outcome.[2]

For both the FOUR score and the GCS overall average, weighted κ scores were calculated to determine the degree of agreement between pairs. κ Statistics (weighted κ) of 0.4 or less are considered poor. Values between 0.4 and 0.6 are considered fair to moderate, values between 0.6 and 0.8 suggest good observer agreement, and values greater than 0.8 suggest excellent agreement.[3] Cronbach α was calculated for each score to assess the internal consistency. Bland and Altman[4] reported that, although a Cronbach α value of 0.7 to 0.8 would be considered satisfactory when the scale is being used as a research tool to compare groups, a minimum of 0.9 would be needed and a value of 0.95 desirable for a clinical application. To investigate the association between level of disagreement and the rater

From the Department of Nursing (C.A.W.), Division of Critical Care Neurology (E.F.M.W.), and Division of Biostatistics (W.R.B.), College of Medicine, Mayo Clinic, Rochester, Minn; and Department of Biostatistics, University of Washington, Seattle (R.L.M.).

Individual reprints of this article are not available. Address correspondence to Eelco F. M. Wijdicks, Division of Critical Care Neurology, College of Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (e-mail: wijde@mayo.edu).

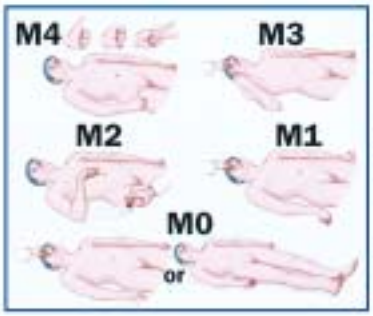
Eye response

- 4 Eyelids open or opened, tracking, or blinking to command
- 3 Eyelids open but not tracking
- 2 Eyelids closed but open to loud voice
- 1 Eyelids closed but open to pain
- 0 Eyelids remain closed with pain

Eye response (E)
Grade the best possible response after at least 3 trials in an attempt to elicit the best level of alertness. A score of **E4** indicates at least 3 voluntary excursions. If eyes are closed, the examiner should open them and examine tracking of a finger or object. Tracking with the opening of 1 eyelid will suffice in cases of eyelid edema or facial trauma. If tracking is absent horizontally, examine vertical tracking. Alternatively, 2 blinks on command should be documented. This will recognize a locked-in syndrome (patient is fully aware). A score of **E3** indicates the absence of voluntary tracking with open eyes. A score of **E2** indicates eyelids opening to loud voice. A score of **E1** indicates eyelids open to pain stimulus. A score of **E0** indicates no eyelids opening to pain.

Motor response

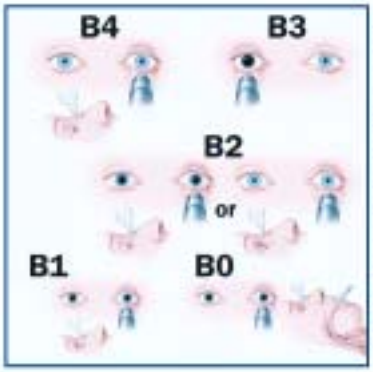
- 4 Thumbs-up, fist, or peace sign to command
- 3 Localizing to pain
- 2 Flexion response to pain
- 1 Extensor posturing
- 0 No response to pain or generalized myoclonus status epilepticus



Motor response (M)
Grade the best possible response of the arms. A score of **M4** indicates that the patient demonstrated at least 1 of 3 hand positions (thumbs-up, fist, or peace sign) with either hand. A score of **M3** indicates that the patient touched the examiner's hand after a painful stimulus compressing the temporomandibular joint or supraorbital nerve (localization). A score of **M2** indicates any flexion movement of the upper limbs. A score of **M1** indicates extensor posturing. A score of **M0** indicates no motor response or myoclonus status epilepticus.

Brainstem reflexes

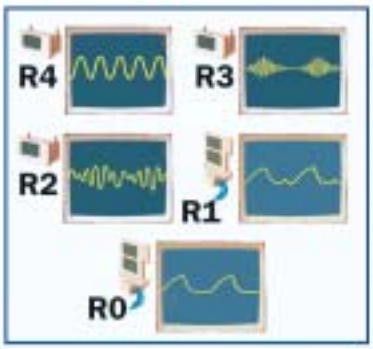
- 4 Pupil and corneal reflexes present
- 3 One pupil wide and fixed
- 2 Pupil or corneal reflexes absent
- 1 Pupil and corneal reflexes absent
- 0 Absent pupil, corneal, and cough reflex



Brainstem reflexes (B)
Grade the best possible response. Examine pupillary and corneal reflexes. Preferably, corneal reflexes are tested by instilling 2-3 drops of sterile saline on the cornea from a distance of 4-6 inches (this minimizes corneal trauma from repeated examinations). Cotton swabs can also be used. The cough reflex to tracheal suctioning is tested only when both of these reflexes are absent. A score of **B4** indicates pupil and cornea reflexes are present. A score of **B3** indicates one pupil wide and fixed. A score of **B2** indicates either pupil or cornea reflexes are absent, **B1** indicates both pupil and cornea reflexes are absent, and a score of **B0** indicates pupil, cornea, and cough reflex (using tracheal suctioning) are absent.

Respiration

- 4 Not intubated, regular breathing pattern
- 3 Not intubated, Cheyne-Stokes breathing pattern
- 2 Not intubated, irregular breathing pattern
- 1 Breathes above ventilator rate
- 0 Breathes at ventilator rate or apnea



Respiration (R)
Determine spontaneous breathing pattern in a nonintubated patient and grade simply as regular **R4**, irregular **R2**, or Cheyne-Stokes **R3** breathing. In mechanically ventilated patients, assess the pressure waveform of spontaneous respiratory pattern or the patient triggering of the ventilator **R1**. The ventilator monitor displaying respiratory patterns is used to identify the patient-generated breaths on the ventilator. No adjustments are made to the ventilator while the patient is graded, but grading is done preferably with $Paco_2$ within normal limits. A standard apnea (oxygen-diffusion) test may be needed when patient breathes at ventilator rate **R0**.

FIGURE 1. Full Outline of UnResponsiveness (FOUR) score.

TABLE 1. **Nurse Rater Agreement With the FOUR Score and Glasgow Coma Scale as Indicated by Weighted κ Values***

Alertness group	No. of patients	FOUR score					Glasgow Coma Scale score			
		Eye	Respiration	Brainstem	Motor	Total	Eye	Verbal	Motor	Total
Overall	80	0.84	0.92	0.89	0.73	0.85	0.85	0.89	0.74	0.83
ICC		0.90	0.95	0.95	0.84	0.96	0.89	0.94	0.86	0.94
Evaluator pairs										
2 Experienced	20	0.94	0.87	1.0	0.91	0.92	0.82	0.97	0.81	0.86
1 Experienced and 1 inexperienced	41	0.83	0.91	0.85	0.64	0.83	0.86	0.87	0.70	0.81
2 Inexperienced	19	0.72	0.95	0.89	0.72	0.82	0.87	0.82	0.77	0.83

*FOUR = Full Outline of UnResponsiveness; ICC = interrater correlation coefficients.

experience level, we calculated the absolute difference in the total FOUR score between the 2 raters for each of the 80 patients. We modeled this as a function of rater experience (expressed as categories) alone and then adjusted for age, sex, diagnosis (trauma vs nontrauma), and alertness group. This approach was also applied for the total GCS score.

RESULTS

Our study consisted of 37 women and 43 men. The average age was 64 years (range, 27-96 years). The diagnoses or procedures of the selected patients were hemorrhagic stroke (n=14), subarachnoid hemorrhage (n=9), craniotomy for brain tumor (n=8), encephalopathy (n=8), ischemic stroke (n=7), subdural hematoma (n=5), seizures/status epilepticus (n=5), vascular surgery (n=4), postanoxic encephalopathy (n=3), spine surgery (n=3), trauma (n=2), central nervous system infection (n=2), neuromuscular disease (n=1), and miscellaneous acute neurologic disorders (n=9).

The rater agreement is shown in Table 1. The overall weighted κ score was 0.85 for the FOUR score and 0.83 for the GCS. Cronbach α was high for the FOUR score (0.95) and for the GCS score (0.94). Using the FOUR score, the overall weighted κ score for 2 experienced nursing pairs was higher (0.92) than that with the GCS (0.86). The weighted κ values remained good to excellent but declined in pairs that included an inexperienced nurse (Table 1).

Both the FOUR score and the GCS were associated with in-hospital death and poor outcome. Twenty-three patients (29%) died, and 61 patients (76%) had a poor outcome (modified Rankin Scale, 3-6). Considering the total FOUR score, for every 1-point increase in total score, there is an estimated 38% reduction in the odds of in-hospital mortality (odds ratio [OR], 0.62; 95% confidence interval [CI], 0.51-0.75). This relationship remained (OR, 0.73; 95% CI, 0.56-0.95) after adjusting for age, sex, alertness group, and diagnosis (traumatic vs nontraumatic). A lower odds of poor outcome (OR, 0.58; 95% CI, 0.41-0.82) was also observed; however, the effect did not remain significant after adjustment (OR, 0.81; 95% CI, 0.53-1.23).With the GCS total score, for every 1-point increase in total score, there is an estimated 55% reduction in the odds of in-hospital mortality (OR, 0.45; 95% CI, 0.31-0.66). This relationship remained (OR, 0.56; 95% CI, 0.34-0.93) after adjustment. A lower odds of poor outcome (OR, 0.67; 95% CI, 0.54-0.83) was also observed with the GCS; similarly, the effect did not remain significant after adjustment (OR, 0.96; 95% CI, 0.67-1.38).

A GCS score of 3 was assigned to 14 (18%) of the 80 patients. All 14 patients had a poor outcome (13 with in-hospital death). Patients with a GCS score of 3 had the following FOUR sum scores: 0, n=5; 1, n=3; 3, n=4; and 5, n=2. The minimum FOUR score was assigned to 5 patients, all of whom died.

The association of various factors with the average difference in scores between 2 raters grading the same patient is shown in Table 2. Regarding experience level, rater pairs with at least 1 inexperienced rater had approximately half a point greater discrepancy between the 2 raters on average (0.61 with 1 inexperienced rater and 0.58 with 2 inexperienced raters) compared with 2 experienced raters; however, this difference was not statistically significant. We also modeled experience as a continuous variable (using the average experience of the 2 raters), and conclusions were similar (data not shown). That is, although greater average experience was associated with less disagreement, this difference was not statistically significant. Age, sex, and trauma were investigated and were not significantly associated with average disagreement. Consciousness category was the only variable that was significantly associated with greater discrepancies. There was significantly less disagreement between the 2 raters for patients in the alert category vs patients in the other categories (drowsy, 0.94; stupor, 1.24; coma, 1.21). Results were similar for the GCS, in that rater experience, age, sex, and trauma were not significantly associated with average disagreement, and there was significantly less disagreement for

TABLE 2. **Linear Regression Model for the Absolute Interrater Difference in Coma Scores***

	GCS		FOUR	
Rater pair	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
2 Experienced	Reference	…	Reference	…
1 Experienced and 1 inexperienced	0.30 (–0.34 to 0.94)	.36	0.61 (–0.03 to 1.25)	.06
2 Inexperienced	0.09 (–0.65 to 0.83)	.80	0.58 (–0.17 to 1.32)	.13
Age (1 y)	–0.01 (–0.03 to 0.01)	.22	–0.003 (–0.180 to 0.170)	.97
Men vs women	–0.13 (–0.67 to 0.40)	.63	0.22 (–0.32 to 0.76)	.42
Trauma vs nontrauma	–0.66 (–1.62 to 0.29)	.17	–0.03 (–0.99 to 0.93)	.95
Group				
Alert	Reference	…	Reference	…
Drowsy	1.18 (0.46 to 1.90)	.001	0.94 (0.22 to 1.67)	.01
Stupor	1.05 (0.30 to 1.80)	.006	1.24 (0.48 to 2.00)	.001
Coma	1.38 (0.58 to 2.18)	<.001	1.21 (0.41 to 2.01)	.003

*CI = confidence interval; FOUR = Full Outline of UnResponsiveness; GCS = Glasgow Coma Scale.

patients in the alert group (drowsy, 1.18; stupor, 1.05; coma, 1.38).

DISCUSSION

A less than optimal agreement was found between the nursing pairs in our prior validation study of the newly devised coma scale, the FOUR score.[1] Although the agreement between nursing staff and residents (weighted κ, 0.75) and nursing staff and neurointensivists (weighted κ, 0.81) rating pairs was high, we found the most variability within the nursing staff, with the highest variability in the eye component of both the FOUR score (weighted κ, 0.48) and the GCS score (weighted κ, 0.50). We found this disagreement among nurses concerning, but the current new large validation study in ICU nursing staff shows that both experienced and inexperienced nurses are able to use both scales with a high degree of agreement. Inexperience in monitoring neurologic patients did not significantly reduce the interobserver agreement, and the scores remained in the good to excellent weighted κ category. Therefore, we conclude that the FOUR score can be used reliably by nurses with limited experience in the neuroscience ICU using a similar patient mix. In addition, the study confirms its association with patient outcome.

When we compared our prior study results with the current results, we found that the overall agreement in the eye component and the brainstem component was markedly better for nurses in the current study. It is likely that familiarity with the FOUR score has resulted in improvement by nurses using this new scale. An earlier study using the GCS found that experience improves rater agreement; however, our study documents that for both the FOUR score and the GCS score, there is little diminishing effect of inexperience on observer agreement.[5] This could be a consequence of teaching the GCS in nursing curriculum and its frequent use in ICUs and other areas of the hospital. The good to excellent agreement between the experienced and inexperienced nurses with the FOUR score, irrespective of years of experience, is notable. This can be explained by incorporating simple daily neurologic tests to assess levels of unconsciousness into the new scale.

CONCLUSION

The FOUR score has major advantages. The 4 components provide important details of the neurologic examination such as brainstem reflexes and eye movements. It recognizes uncal herniation, a locked-in syndrome, and the beginning of a vegetative state. This detail is not provided by the GCS. In fact, 1 of the 3 components of the GCS (verbal score) cannot be used in patients who have undergone intubation. We believe our results suggest that the FOUR score could be used outside the neuroscience ICU by any nurses because experience in the neuroscience ICU is not a strong determinant of reliability.

REFERENCES

1. Wijdicks EFM, Bamlet WR, Maramattom BV, Manno EM, McClelland RL. Validation of a new coma scale: the FOUR score. *Ann Neurol*. 2005;58:585-593.
2. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*. 1988;19:604-607.
3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
4. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ*. 1997; 314:572.
5. Rowley G, Fielding K. Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet*. 1991;337:535-538.