Validation of a New Coma Scale: The FOUR Score

Eelco F. M. Wijdicks, MD,[1] William R. Bamlet, MS,[2] Boby V. Maramattom, MD,[1] Edward M. Manno, MD,[1] and Robyn L. McClelland, PhD[2]

The Glasgow Coma Scale (GCS) has been widely adopted. Failure to assess the verbal score in intubated patients and the inability to test brainstem reflexes are shortcomings. We devised a new coma score, the FOUR (Full Outline of UnResponsiveness) score. It consists of four components (eye, motor, brainstem, and respiration), and each component has a maximal score of 4. We prospectively studied the FOUR score in 120 intensive care unit patients and compared it with the GCS score using neuroscience nurses, neurology residents, and neurointensivists. We found that the interrater reliability was excellent with the FOUR score (κ_w = 0.82) and good to excellent for physician rater pairs. The agreement among raters was similar with the GCS (κ_w = 0.82). Patients with the lowest GCS score could be further distinguished using the FOUR score. We conclude that the agreement among raters was good to excellent. The FOUR score provides greater neurological detail than the GCS, recognizes a locked-in syndrome, and is superior to the GCS due to the availability of brainstem reflexes, breathing patterns, and the ability to recognize different stages of herniation. The probability of in-hospital mortality was higher for the lowest total FOUR score when compared with the lowest total GCS score.

Ann Neurol 2005;58:585–593

Assessment of coma is a core clinical skill for physicians. Scales have been constructed to improve communication among health care personnel and also to standardize examination of the unconscious patient. The most commonly used scale is the Glasgow Coma Scale (GCS).[1] Although the originators of the GCS have reported data on practical reliability of the scale,[2] other studies have showed some difficulties in application by untrained nursing staff.[3] Trained personnel tend to apply the GCS better, although interpretation of intermediate scores on the GCS remains difficult for emergency physician.[4]

Other shortcomings of the GCS have been recognized. First, because many comatose patients are intubated, the verbal component cannot be tested. Some clinicians use the lowest possible score; others extrapolate the verbal response based on other neurological findings. Second, abnormal brainstem reflexes, changing breathing patterns, and the need for mechanical ventilation could reflect severity of coma, but the GCS does not include those clinical indicators. Third, the GCS may not detect subtle changes in neurological examination. More recently, a study in traumatic head injury found lack of correlation between outcome and GCS score.[5]

Attempts have been made to modify the GCS; however, most of these scales were more complicated, and were seldom used outside the country of origin.[6,8] Others have suggested simplification of the GCS score after documenting poor interobserver reliability in traumatic brain injury.[9] These concerns and prior attempts to design new scales strongly suggest a new scale is needed that could provide further neurological detail in coma that is easy to use and that could predict outcome. We sought to validate a new coma scale, the FOUR (Full Outline of UnResponsiveness) score, and thus compared it with the GCS.

Patients and Methods

Description of the New Coma Scale

The new coma scale was named the FOUR score (Fig 1). The FOUR score has four testable components, in contrast with the GCS (Table 1). The number of components and the maximal grade in each of the categories is four (E_4, M_4, B_4, R_4). (It is easier to remember than the GCS with its varying number of scores [E_4, M_6, V_5] and is reinforced by the acronym.) The FOUR score detects a locked-in syndrome, as well as the presence of a vegetative state where the eyes can spontaneously open but do not track the examiner's finger. The motor response is obtained preferably at the up-

From the Divisions of [1]Critical Care Neurology and [2]Biostatistics, Mayo Clinic College of Medicine, Rochester, MN.

Received Apr 29, 2005, and in revised form June 11, 2005. Accepted for publication July 17, 2005.

Published online Sep 26, 2005, in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ana.20611

Address correspondence to Dr Wijdicks, Mayo Clinic College of Medicine, Division of Critical Care Neurology, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905. E-mail: wijde@mayo.edu

Published by Wiley-Liss, Inc., through Wiley Subscription Services

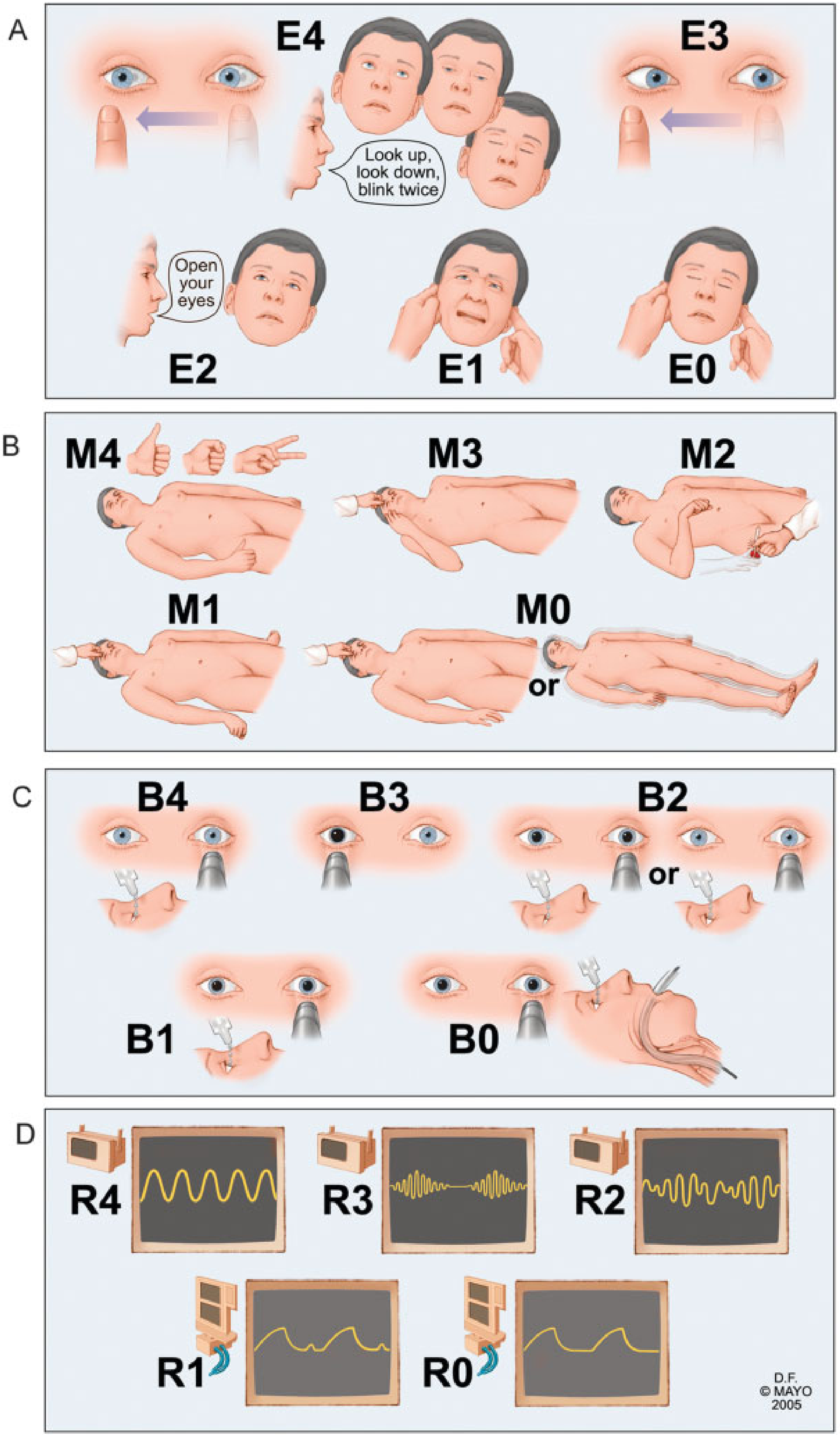
A
E4
E3
Look up, look down, blink twice
Open your eyes
E2
E1
E0
B
M4
M3
M2
M1
M0
or
C
B4
B3
B2
or
B1
B0
D
R4
R3
R2
R1
R0
D.F.
© MAYO
2005

Figure 1

Table 1. Comparison of the FOUR Score with the Glasgow Coma Scale

FOUR Score	Glasgow Coma Scale
Eye response	Eye response
4 = eyelids open or opened, tracking, or blinking to command	4 = eyes open spontaneously
3 = eyelids open but not tracking	3 = eye opening to verbal command
2 = eyelids closed but open to loud voice	2 = eye opening to pain
1 = eyelids closed but open to pain	1 = no eye opening
0 = eyelids remain closed with pain	
Motor response	Motor response
4 = thumbs-up, fist, or peace sign	6 = obeys commands
3 = localizing to pain	5 = localizing pain
2 = flexion response to pain	4 = withdrawal from pain
1 = extension response to pain	3 = flexion response to pain
0 = no response to pain or generalized myoclonus status	2 = extension response to pain
	1 = no motor response
Brainstem reflexes	Verbal response
4 = pupil and corneal reflexes present	5 = oriented
3 = one pupil wide and fixed	4 = confused
2 = pupil or corneal reflexes absent	3 = inappropriate words
1 = pupil and corneal reflexes absent	2 = incomprehensible sounds
0 = absent pupil, corneal, and cough reflex	1 = no verbal response
Respiration	
4 = not intubated, regular breathing pattern	
3 = not intubated, Cheyne–Stokes breathing pattern	
2 = not intubated, irregular breathing	
1 = breathes above ventilator rate	
0 = breathes at ventilator rate or apnea	

FOUR = Full Outline of UnResponsiveness.

per extremities. The motor category includes the presence of myoclonus status epilepticus (persistent, multisegmental, arrhythmic, jerklike movements), a poor prognostic sign in comatose survivors after cardiac resuscitation.[10] The motor component combines decorticate and withdrawal responses. (We think this difference is often difficult to appreciate.) The hand position tests (thumbs-up, fist, and peace sign) have been validated previously and are reliable to assess alertness.[11] Three brainstem reflexes testing mesencephalon, pons, and medulla oblongata function are used in different combinations. The clinical sign of acute third nerve dysfunction (unilateral dilated pupil) is included. The cough reflex mostly becomes absent when both cornea and pupillary reflexes are absent. Breathing patterns are graded. Cheyne–

Fig 1. Instructions for the assessment of the individual categories of the FOUR (Full Outline of UnResponsiveness) score (see Table 1). (A) For eye response (E), grade the best possible response after at least three trials in an attempt to elicit the best level of alertness. A score of E4 indicates at least three voluntary excursions. If eyelids are closed, the examiner should open them and examine tracking of a finger or object. Tracking with the opening of one eyelid will suffice in cases of eyelid edema or facial trauma. If tracking is absent horizontally, examine vertical tracking. Alternatively, two 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 a loud voice. A score of E1 indicates eyelids open to pain stimulus. A score of E0 indicates no eyelid opening to pain. (B) For motor response (M), grade the best possible response of the arms. A score of M4 indicates that the patient demonstrated at least one of three hand positions (thumbs-up, fist, or peace sign) with either hand. A score of M3 (localization) indicates that the patient touched the examiner's hand after a painful stimulus compressing the temporomandibular joint or supraorbital nerve. A score of M2 indicates any flexion movement of the upper limbs. A score of M1 indicates extensor response to pain. A score of M0 indicates no motor response to pain, or myoclonus status epilepticus. (C) For brainstem reflexes (B), grade the best possible response. Examine pupillary and corneal reflexes. Preferably, corneal reflexes are tested by instilling two to three drops sterile saline on the cornea from a distance of 4 to 6 inches (this minimizes corneal trauma from repeated examinations). Sterile 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 corneal 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. A score of B1 indicates both pupil and cornea reflexes are absent. A score of B0 indicates pupil, cornea, and cough reflex (using tracheal suctioning) are absent. (D) For respiration (R), determine spontaneous breathing pattern in a nonintubated patient and grade simply as regular (R_4), or irregular (R_2), Cheyne–Stokes (R_3) breathing. In mechanically ventilated patients, assess the pressure waveform of spontaneous respiratory pattern or the patient triggering of the ventilator (R_1). The ventilator monitor displaying respiratory patterns can be 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 (R_0). Figure reproduced with permission by Mayo Foundation.

Stokes respiration and irregular breathing can represent bihemispheric or lower brainstem dysfunction of respiratory control. In intubated patients, overbreathing the mechanical ventilator represents functioning respiratory centers. With all categories graded 0, the examiner is alerted to consider brain death evaluation. The FOUR score can be graded in a few minutes.

Interobserver Variability Study

Three different types of examiners tested the FOUR score and the GCS: three neuroscience nurses, three neurology residents (third or fourth year) or fellows, and three neurointensivists. Each of the neuroscience nurses and neurointensivists had more than 10 years' clinical experience in a neurological/neurosurgical intensive care unit (ICU) and were reinstructed in GCS using drawings of the eye, motor, and verbal responses. Raters watched a 20-minute instruction on the FOUR score that used videos with patient examples. Subsequently, 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 120 patients.

We recruited patients admitted to the neurointensive care unit. Consultations in other medical or surgical ICUs for "unresponsiveness" were included. Patients receiving sedative agents or neuromuscular function blockers were excluded. Patients were tested on the day of admission or the day "unresponsiveness or mental status changes" triggered a neurology consult by medical or surgical intensivists. To assess the properties of the FOUR score over the full spectrum of patients, we selected patients from four categories. These patients were alert, drowsy, stuporous, or comatose, using definitions of the above states as Ropper[12] described previously. The study was designed to evaluate an equal number of patients ($n = 30$) in each of these 4 categories. Patients were chosen consecutively subject to availability of a study investigator until the subset was complete.

Each patient was rated on both scales by two different raters. The raters performed their examination within 1 hour of each other without knowledge of the other's scores. The experiment was designed so that 20 patients were rated for each combination of training type (nurse/nurse, nurse/resident, nurse/neurointensivist, resident/resident, resident/neurointensivist, and neurointensivist/neurointensivist), resulting in 120 patients studied. The order of the evaluations was randomized to reduce bias. (For instance, of the 20 patients rated by both a resident and an neurointensivist, 10 were rated by the resident first followed by the neurointensivist, and 10 were rated by the neurointensivist first.) This design allowed us to evaluate reliability between types of examiners. We did not evaluate intrarater reliability because it would not be possible for a single rater to score a patient at two time points (sufficiently close in time to ensure neurological status of the patient has not changed) without remembering previous scores.

With this sample size, if 50% of patients had a score better than or equal to 3 on any particular subscale, a κ statistic of 0.8 to 0.85 would have a standard error of 0.12 to 0.13 based on a sample of 20 patients. The overall average κ would have a standard error of 0.02, and a 95% confidence interval (CI) would have estimated κ to within $\pm 4\%$.

We recorded in-hospital mortality and clinical diagnosis of brain death. Morbidity was assessed at 3 months using the modified Rankin Scale.[13] In brief, 0 = no symptoms; 1 = no significant disability despite symptoms; 2 = slight disability, unable to carry out all previous activities, able to take after own affairs; 3 = moderate disability, requiring some help, but able to walk without assistance; 4 = moderately severe disability, unable to walk without assistance, and unable to attend to own bodily needs without assistance; 5 = severe disability, bedridden, incontinent, and requiring constant nursing care; 6 = dead.

This study was approved by the Neurology Research Committee and Mayo Foundation Institutional Research Board. Informed consent was waived in accordance with 45 CFR 46.116 (d), but written Health Insurance Portability and Accountability Act authorization to obtain patient data was obtained.

Statistical Analysis

For both the FOUR score and the GCS, pairwise weighted κ values (for each training type pair) and overall weighted κ values were calculated. For intubated patients, the lowest GCS verbal score was used in the analysis. This approach is conservative for our reliability comparisons in that the agreement among raters will be inflated by these automatically perfect agreements. Cronbach's α was calculated for each score to assess internal consistency, and Spearman's correlation coefficients between the FOUR score and the GCS were calculated to assess construct validity. Finally, sensitivity and specificity of the two measures were compared for prediction of in-hospital mortality and morbidity using a logistic regression model controlling for age, sex, diagnosis, and either total FOUR or total GCS score. The area under the receiver operating characteristic curve was calculated for each model.

A κ statistic of 0.4 or less is considered poor, values between 0.4 and 0.6 are considered fair to moderate, those between 0.6 and 0.8 suggest good interobserver agreement, and values greater than 0.8 suggest excellent agreement.[14]

Survey of Raters

Face validity of the FOUR score was assessed by asking the raters to comment on its use. The following statements were graded using a five-point Likert scale (1 = strongly disagree; 5 = strongly agree): (1) The FOUR score is clinically relevant and easy to use; (2) The FOUR score is obtained in a matter of minutes; (3) The FOUR score is a good alternative to GCS; (4) The FOUR score is a better score than GCS when looking for depth of coma or patient deterioration; and (5) The FOUR score is a coma assessment scale I would use if it becomes generally accepted.

Results

Patient Characteristics

From July to September 2004, 120 patients were enrolled. The average age of patients was 58.9 years (median, 60 years; range, 45–70 years); 52% were men.

Table 2. *Rater Agreement with the FOUR Score and the Glasgow Coma Scale as Indicated by Weighted κ Values*

Rater Pair	No. of Patients	FOUR Score					GCS			
		Eye	Motor	Brainstem	Respiration	Total	Eye	Motor	Verbal	Total
N/N	20	0.48	0.66	0.43	0.50	0.70	0.50	0.75	0.84	0.72
N/R	20	0.84	0.87	0.79	0.54	0.75	0.85	0.66	0.87	0.77
N/NI	20	0.77	0.74	0.83	0.83	0.81	0.62	0.77	0.87	0.80
R/R	21	0.79	0.92	1.00	1.00	0.93	0.93	0.86	1.00	0.92
R/NI	19	0.89	0.83	0.84	0.81	0.85	0.89	0.79	0.91	0.89
NI/NI	20	0.87	0.69	0.89	0.81	0.81	0.79	0.69	0.80	0.75
Overall (95% CI)	120	0.78 (0.70–0.87)	0.80 (0.72–0.88)	0.81 (0.70–0.91)	0.78 (0.68–0.88)	0.82 (0.77–0.88)	0.77 (0.69–0.85)	0.77 (0.68–0.85)	0.88 (0.81–0.96)	0.82 (0.76–0.87)

FOUR = Full Outline of Unresponsiveness; GCS = Glasgow Coma Scale; N = nurse; R = resident; NI = neurointensivist; CI = confidence interval.

The diagnoses of the patients selected for the study were ischemic or hemorrhagic stroke (29 patients; 24%), traumatic head injury (25 patients; 21%), craniotomy for brain tumor (13 patients; 11%), aneurysmal subarachnoid hemorrhage (12 patients; 10%), postanoxic-ischemic encephalopathy (10 patients; 8%), spinal surgery (including trauma; 8 patients; 7%), seizures and status epilepticus (7 patients; 6%), other encephalopathies (4 patients; 3%), central nervous system infection (4 patients; 3%), acute neuromuscular disease (3 patients; 2%), and miscellaneous acute neurological conditions (5 patients; 4%). In two patients, eye opening and brainstem reflexes could not be tested reliably because of posttraumatic eye swelling; thus, the best score was entered. Fifty-seven patients (48%) were intubated and mechanically ventilated.

Interrater Reliability of the FOUR Score

The overall reliability was excellent for both the FOUR score (κ_w = 0.82; 95% CI, 0.77–0.88) and the GCS (κ_w = 0.82; 95% CI, 0.76–0.87). The rater agreement was good to excellent for physician rater pairs. The highest degree of agreement was among the neurology residents, and agreement was lowest among the neuroscience nurses for both scales (Table 2). The frequency of scores for each scale is shown in Figure 2. Five patients became brain dead, and one patient had a locked-in syndrome. Myoclonus status epilepticus (in all patients after cardiac resuscitation) was noted in five comatose patients, who also had no motor response to pain. For the FOUR score, 82 of 240 (34%) observations had a maximal score of 16. The brainstem component had the highest maximal scores (188/240 observations; 78%). The distribution of the scores for the eye and motor components of the FOUR score was comparable with the distribution with the GCS. A GCS total score of 3 was recorded on 34 occasions. Of these, only nine were scored at the lowest FOUR score of 0. In the remaining 25 instances, the brainstem reflexes and respiration components provided additional information that allowed the raters to distinguish among the patients' total scores (FOUR score of 1–8).

Cronbach's α showed a high degree of internal consistency for FOUR score (α = 0.86 for the first rater; α = 0.87 for the second rater) and the GCS (α = 0.88 for the first rater; α = 0.84 for the second rater). Spearman's correlation coefficients between GCS and FOUR scores were high (ρ = 0.92 for both first and second ratings).

The rater agreement by level of conscious group was comparable between the two scales. The total weighted κ scores of the FOUR score for the alert, drowsy, stuporous, and comatose groups were κ_w = 0.57 (95% CI, 0.24–0.90), κ_w = 0.74 (95% CI, 0.53–0.94), κ_w = 0.75 (95% CI, 0.61–0.88), and κ_w = 0.70 (95% CI, 0.57–0.83), respectively. The total weighted κ values of the GCS for the alert, drowsy, stuporous, and comatose groups were κ_w = 0.59 (95% CI, 0.33–0.86), κ_w = 0.69 (95% CI, 0.54–0.85), κ_w = 0.72 (95% CI, 0.57–0.86), and κ_w = 0.69 (95% CI, 0.54–0.84), respectively.

The rater agreement by diagnosis of traumatic head injury was comparable between the two scales. The total weighted κ scores of the FOUR score for the traumatic and nontraumatic head injury groups were κ_w = 0.73 (95% CI, 0.57–0.88) and κ_w = 0.84 (95% CI, 0.79–0.90), respectively. The total weighted κ values of the GCS for traumatic and nontraumatic head injury groups were κ_w = 0.71 (95% CI, 0.55–0.86) and κ_w = 0.84 (95% CI, 0.79–0.90), respectively.

All nine raters agreed or strongly agreed (Likert grade 4 or 5) with the five statements that addressed the clinical usefulness of the FOUR score.

Outcome Prediction of the FOUR Score

Twenty-five (21%) patients died, and 72 patients (60%) had a poor outcome (modified Rankin Scale, 3–6).Table 3 presents the relations between total score and patient outcome for each of the two scales. Considering the FOUR scale total score, for every 1-point increase in total score, there is an estimated 20% reduction in the odds of in-hospital mortality (odds ratio [OR] = 0.80; 95% CI, 0.72–0.88). A 1-point increase in total score is also associated with lower odds of poor

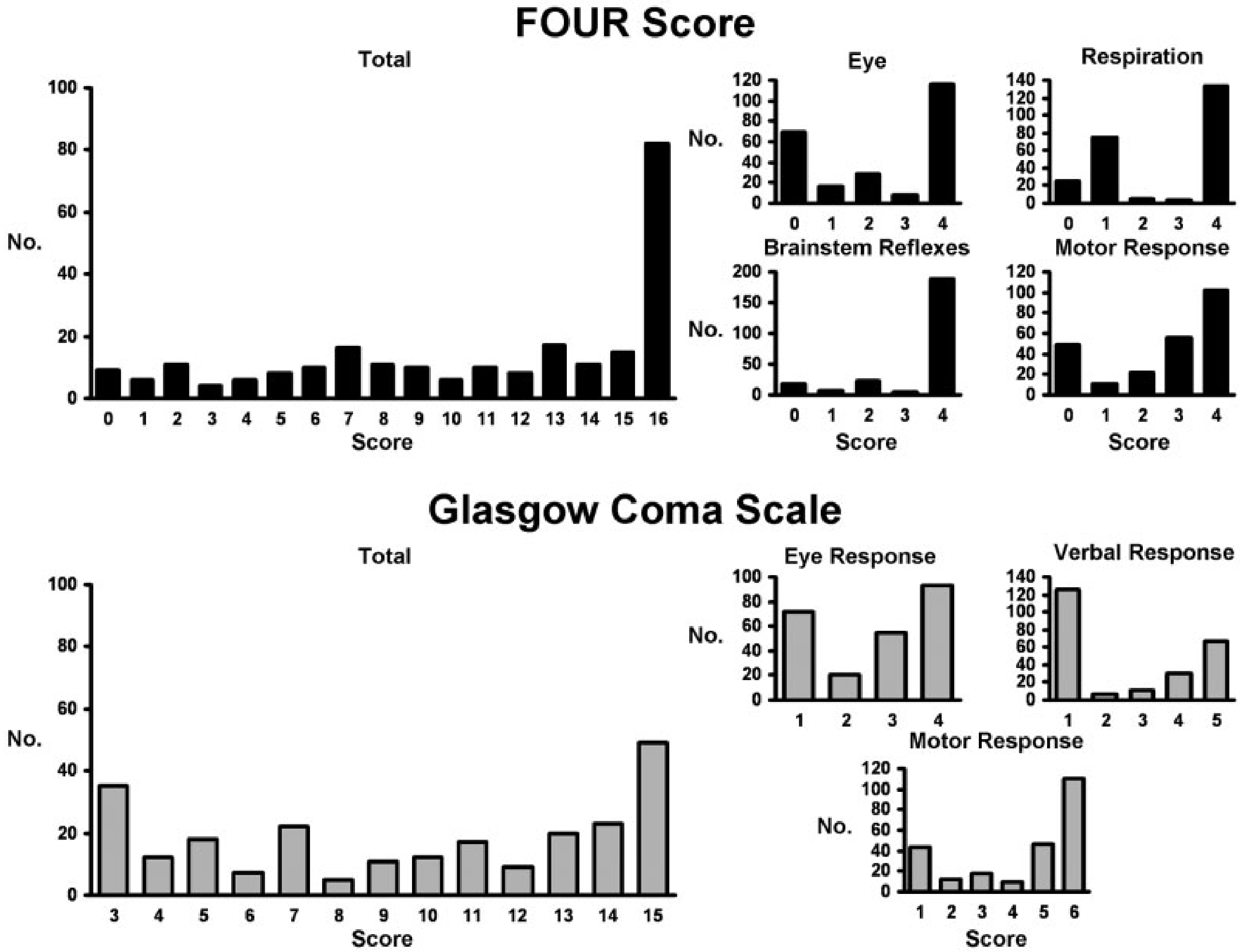

Fig 2. Frequency of rated categories for 240 observations in FOUR (Full Outline of UnResponsiveness) score and Glasgow Coma Scale (GCS).

outcome defined as a modified Rankin scale of 3 or more (OR = 0.84; 95% CI, 0.77–0.92). Both relations remain after adjusting for age, sex, alertness group, and diagnosis (traumatic vs nontraumatic).

With the GCS scale total score, for every 1-point increase in total score, there is an estimated 26% reduced odds of experiencing in-hospital mortality under the unadjusted model (OR = 0.74; 95% CI, 0.65–0.85). This relation remains after adjusting for age, sex, alertness group, and diagnosis (traumatic vs nontraumatic). A 1-point increase in total score is also associated with lower odds of poor outcome (OR = 0.83; 95% CI, 0.76–0.92). This effect is attenuated slightly after considering the adjusted model (OR = 0.89; 95% CI, 0.76–1.03).

Receiver operating characteristic curves were estimated to compare prediction of in-hospital mortality between the two scales. The area under the curve was equivalent for the two scales at 0.81. The sum of sensitivity and specificity was maximized at a FOUR total score of 9 (sensitivity = 0.75; specificity = 0.76) and a GCS total score of 7 (sensitivity = 0.80; specificity = 0.80). Prediction was less accurate for poor outcome (Rankin, 3–6), with area under the curve at 0.72 for both scales.

To visualize the relation between outcome and total score, we examined scatterplots with superimposed local regression smoothers. We used a model-based smoothing with generalized additive models (Fig 3). Although these relations are comparable, overall there are some subtle differences that do exist. For example, the probability of in-hospital mortality is higher for the lowest total FOUR scores when compared with GCS. This is due to our observation that the patients with GCS scores of 3 may be further separated using the FOUR scale. For both scales there appears to be a range of values above which the risk for in-hospital mortality is close to 0 (GCS > 8; FOUR > 12). Risk for a poor outcome (Rankin, 3–6) in general declines more gradually and is comparable between the scales.

Table 3. Prediction of Outcome (in-hospital death and Rankin scores of 3–6)

	In-Hospital Death (N = 120; 25 events)			Rankin Score of 3 to 6 (N = 120; 72 events)		
	OR[a] (95% CI)	OR[b] (95% CI)	OR[c] (95% CI)	OR[a] (95% CI)	OR[b] (95% CI)	OR[c] (95% CI)
FOUR score total	0.80 (0.72–0.88)	0.80 (0.68–0.93)	0.79 (0.68–0.93)	0.84 (0.77–0.92)	0.86 (0.75–0.97)	0.86 (0.75–0.98)
FOUR categories[d]						
Eye	1.03 (0.68–1.58)	1.37 (0.79–2.35)	1.42 (0.82–2.48)	0.78 (0.55–1.10)	0.89 (0.58–1.36)	0.93 (0.60–1.43)
Motor	0.74 (0.47–1.17)	0.74 (0.44–1.24)	0.71 (0.43–1.20)	0.88 (0.58–1.34)	0.85 (0.54–1.35)	0.83 (0.52–1.32)
Brainstem	0.64 (0.41–1.02)	0.55 (0.33–0.91)	0.54 (0.33–0.91)	0.80 (0.48–1.36)	0.71 (0.41–1.22)	0.71 (0.41–1.22)
Respiration	0.76 (0.49–1.18)	0.73 (0.46–1.15)	0.71 (0.45–1.13)	0.90 (0.64–1.27)	0.91 (0.63–1.32)	0.91 (0.63–1.32)
GCS score total	0.74 (0.65–0.85)	0.73 (0.58–0.90)	0.72 (0.57–0.90)	0.83 (0.76–0.92)	0.88 (0.76–1.03)	0.89 (0.76–1.03)
GCS categories[d]						
Eye	1.13 (0.63–2.03)	1.26 (0.64–2.49)	1.34 (0.66–2.73)	0.67 (0.40–1.12)	0.71 (0.39–1.29)	0.74 (0.40–1.37)
Motor	0.64 (0.45–0.91)	0.61 (0.41–0.90)	0.57 (0.38–0.87)	0.97 (0.69–1.37)	0.97 (0.68–1.40)	0.93 (0.64–1.36)
Vocal	0.65 (0.37–1.14)	0.65 (0.35–1.20)	0.66 (0.36–1.24)	0.82 (0.60–1.13)	0.89 (0.64–1.25)	0.92 (0.65–1.30)

[a]Unadjusted logistic regression model.
[b]Logistic regression model adjusted for age, sex, and consciousness group.
[c]Logistic regression model adjusted for age, sex, consciousness group, and diagnosis (trauma vs nontrauma).
[d]Estimates are based on a multivariate model, which includes all individual categories.

OR = odds ratio; CI = confidence interval; FOUR = full outline of unresponsiveness; GCS = Glasgow Coma Scale.

Discussion

The FOUR score is simple to use, includes the minimal necessities of neurological testing in impaired consciousness, and specifically recognizes certain unconscious states. Our interobserver study is the largest validation study of a new coma scale with 120 pairwise ratings involving 3 neuroscience nurses, 3 neurology residents/fellows, and 3 neurointensivists. The patient population studied is indicative of 3 months of admissions to ICUs and includes a diversity of acute neurological conditions, not exclusively trauma.

Prior validation studies on GCS have included 47 neurosurgical patients[6] and 64 neurointensive care patients,[15] but 2 other studies involved less patients.[1,3] The design of these studies, which involved the application of a single stimulus with multiple observers rating at the same time, is not reflective of how the scale is used in practice and eliminates an important source of variability. In addition, we attempted to ensure that the two ratings of each patient occurred as closely in time as possible to minimize the possibility that the patient's condition had changed. Prior studies have involved newly graduated nurses or student nurses,[3] research psychologists,[15] paramedics,[4] and occupational therapy graduate students supervised by a medical director of the neurointensive care unit.[16] Our raters were chosen because in practice they would examine these patients, communicate their findings to each other, and make decisions.

The interrater reliability of the FOUR score and the GCS were of equivalent magnitude. This is remarkable because the raters had only minimal experience with the FOUR score. In our study, the observer agreement was highest among residents, followed by neurointensivists, then nurses. There was perfect agreement among the residents in rating respiration and brainstem reflexes, which is an important finding when communication with the attending consultant is sought.

The FOUR score, unlike the GCS, does not include a verbal response, and thus is more valuable in ICU practices that typically have a large number of intubated patients. In this prospective study using the FOUR score, all patients, except the two patients with periorbital swelling from trauma, could be assessed reliably. In contrast, the GCS, which uses a verbal score as one of the three components, was less useful in 48% of patients because they were intubated. Most likely, the verbal agreement ($\kappa = 0.88$) in our study is artificially high because it merely requires that the patient have an endotracheal tube inserted, which obviates the need to further examine the verbal response. This would be expected because the verbal component has been recognized as the least reliable component of the GCS.[16] The reliability of testing brainstem reflexes has rarely been studied in a large population of patients but was excellent among our raters. In one prior study, pupillary responses and oculocephalic responses were tested in 28 patients, and fair interobserver agreement was found for only the oculocephalic responses ($\kappa = 0.49$).[17] Examination of some brainstem reflexes has been incorporated in the modified GCS (Glasgow–Liege Coma Scale). These reflexes included rapid neck movements to obtain oculovestibular reflexes and eyeball pressure to obtain oculocardiac reflexes. The interobserver agreement among 6 neurosurgeons testing 30 patients had a κ_w of 0.69.[18,19] We did not incorporate these reflexes that could further jeopardize patients who had additional spinal trauma and hemodynamic instability. Abnormal respiratory breathing and ventilatory drive may have localizing value in comatose patients, but we acknowledge important variables such as acute pulmonary disease and ventilator settings. Our study shows patterns of breathing can be easily mastered by physicians and interpreted satisfactorily by neuroscience nurses.

Most discrepant scores were found among the neu-

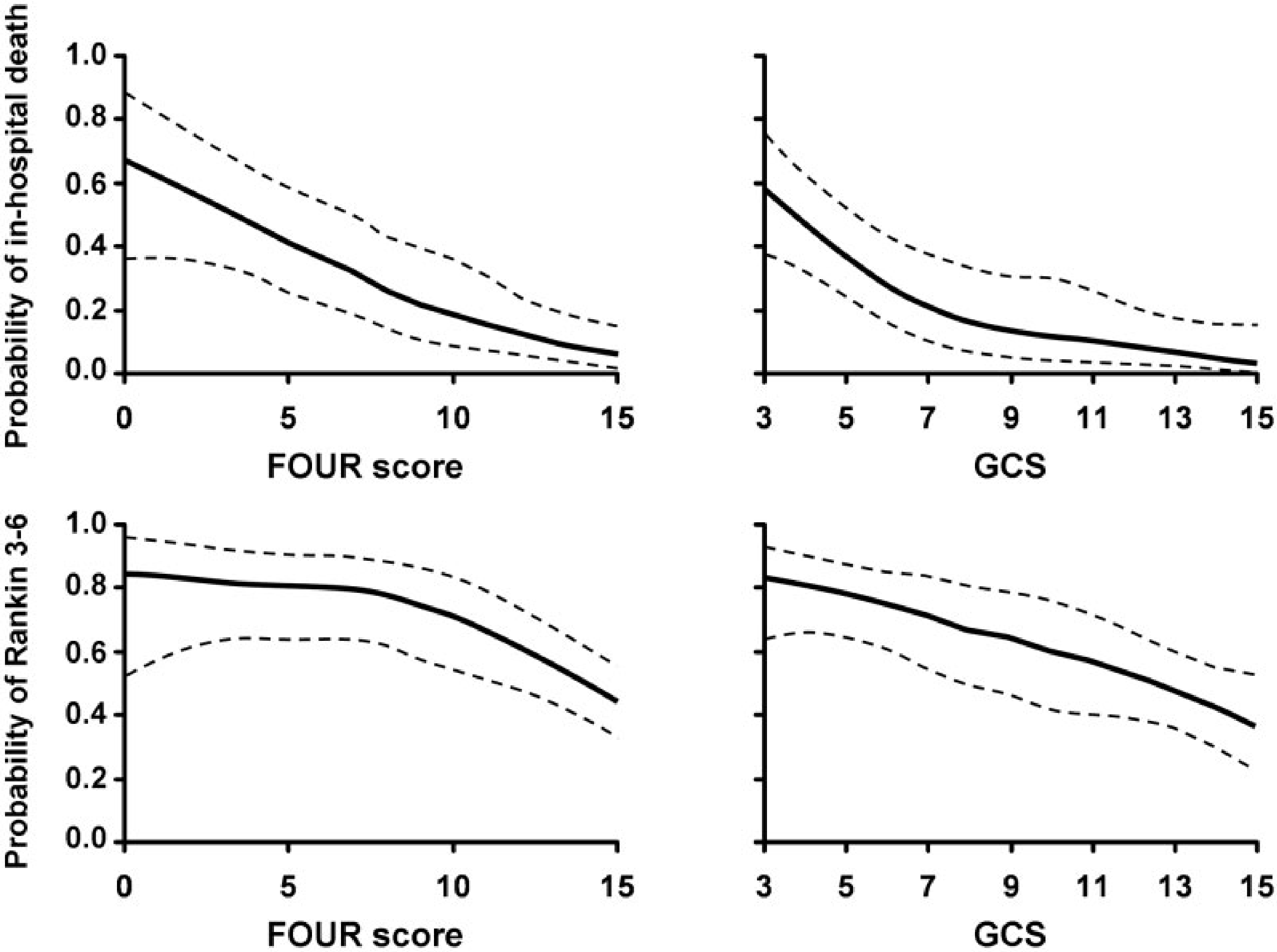

Fig 3. Relation of outcome (in-hospital death or Rankin 3–6) to total score of both FOUR (Full Outline of UnResponsiveness) score and Glasgow Coma Scale (GCS; with pointwise 95% confidence intervals).

roscience nurses in some categories of both scales. In the neuroscience nurses pairs, the lowest agreement was with grading eye responses in both the FOUR score and GCS and with the interpretation of brainstem reflexes. Ratings of the eye responses are influenced by factors such as intensity of pain and loudness of voice, fluctuating alertness in between ratings, or time spent to obtain the response. The differences in observer agreement may nevertheless indicate that the nursing staff with more training should use the scales. However, the interrater reliability among nurses was still fair to good, and in none of the subcategories was a poor interrater reliability found.

There are significant advantages over the GCS score. The FOUR score remains testable in neurologically critically ill patients who are intubated. Intubation is a common procedure in the field, emergency department, and ICU that invalidates one of the three components of the GCS. The FOUR score tests essential brainstem reflexes and provides information about stages of brainstem injury that is unavailable with the GCS. The FOUR score recognizes a locked-in syndrome and a possible vegetative state. The FOUR score includes signs suggesting uncal herniation. Attention to respiratory patterns in the FOUR score not only may indicate a need for respiratory support in stuporous or comatose patients, but also provides information about the presence of a respiratory drive. The FOUR score further characterizes the severity of the comatose state in patients with the lowest GCS score. Finally, the probability of in-hospital mortality was higher for the lowest total FOUR scores when compared with the GCS.

Teasdale and Jennett[1] noted more than 30 years ago that "some may have reservations about a system which appears to undervalue the niceties of a full neurological examination." Our study shows that greater neurological refinement in the FOUR score is valid when confronted with a patient who has an impaired consciousness. Using our new scale, the examiner has an aid to describe these essential clinical features.

Our study permits implementation of the FOUR score in the neurological-neurosurgical ICU and may provide the opportunity for longitudinal studies. This study was done exclusively in neuroscience professionals. It would be of interest to test the FOUR score in emergency physicians, trauma surgeons, medical or surgical intensivists, and allied nursing staff. The FOUR score may be pertinent to future clinical trials when a more complete appreciation of the severity of coma is needed.

We thank the raters (D. Gusa, A. Miers, D. Pfrimmer, and Drs J. Fulgham, D. Sencakova, T. Steeves, and S. Diesing) for their willingness to participate in this study. We also thank D. Factor of Mayo Clinic Illustration and Design for his contribution and R. Nelson for secretarial assistance.

References

1. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974;2:81–84.
2. Teasdale G, Knill-Jones R, van der Sande J. Observer variability in assessing impaired consciousness and coma. J Neurol Neurosurg Psychiatry 1978;41:603–610.
3. Rowley G, Fielding K. Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. Lancet 1991;337:535–538.
4. Menegazzi JJ, Davis EA, Sucov AN, et al. Reliability of the Glasgow Coma Scale when used by emergency physicians and paramedics. J Trauma 1993;34:46–48.
5. Balestreri M, Czosnyka M, Chatfield DA, et al. Predictive value of Glasgow Coma Scale after brain trauma: change in trend over the past ten years. J Neurol Neurosurg Psychiatry 2004;75:161–162.
6. Starmark JE, Stalhammar D, Holmgren E, et al. A comparison of the Glasgow Coma Scale and the Reaction Level Scale (RLS85). J Neurosurg 1988;69:699–706.
7. Stanczak DE, White JG 3rd, Gouview WD, et al. Assessment of level of consciousness following severe neurological insult. A comparison of the psychometric qualities of the Glasgow Coma Scale and the Comprehensive Level of Consciousness Scale. J Neurosurg 1984;60:955–960.
8. Benzer A, Mitterschiffthaler G, Marosi M, et al. Prediction of non-survival after trauma: Innsbruck Coma Scale. Lancet 1991;338:977–978.
9. Gill M, Windemuth R, Steele R, et al. A comparison of the Glasgow Coma Scale score to simplified alternative scores for the prediction of traumatic brain injury outcomes. Ann Emerg Med 2005;45:37–42.
10. Wijdicks EFM, Parisi JE, Sharbrough FW. Prognostic value of myoclonus status in comatose survivors of cardiac arrest. Ann Neurol 1994;35:239–243.
11. Wijdicks EFM, Kokmen E, O'Brien PC. Measurement of impaired consciousness in the neurological intensive care unit: a new test. J Neurol Neurosurg Psychiatry 1998;64:117–119.
12. Ropper AH. Lateral displacement of the brain and level of consciousness in patients with an acute hemispheral mass. N Engl J Med 1986;314:953–958.
13. van Swieten JC, Koudstaal PJ, Visser MC, et al. Interobserver agreement for the assessment of handicap in stroke patients. Stroke 1988;19:604–607.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
15. Mayer SA, Dennis LJ, Peery S, et al. Quantification of lethargy in the neuro-ICU: the 60-Second Test. Neurology 2003;61:543–545.
16. Diringer MN, Edwards DF. Does modification of the Innsbruck and the Glasgow Coma Scales improve their ability to predict functional outcome? Arch Neurol 1997;54:606–611.
17. van den Berge JH, Schouten HJ, Boomstra S, et al. Interobserver agreement in assessment of ocular signs in coma. J Neurol Neurosurg Psychiatry 1979;42:1163–1168.
18. Born JD, Albert A, Hans P, et al. Relative prognostic value of best motor response and brain stem reflexes in patients with severe head injury. Neurosurgery 1985;16:595–601.
19. Born JD, Hans P, Albert A, et al. Interobserver agreement in assessment of motor response and brain stem reflexes. Neurosurgery 1987;20:513–517.