

Evaluation of Coma: A Critical Appraisal of Popular Scoring Systems

Joshua Kornbluth · Anish Bhardwaj

Published online: 22 July 2010

© Springer Science+Business Media, LLC 2010

Abstract Numerous scoring scales have been proposed and validated to evaluate coma for rapid pre-hospital assessment and triage, disease severity, and prognosis for morbidity and mortality. These scoring systems have been predicated on core features that serve as a basis for this review and include ease of use, inter-rater reliability, reproducibility, and predictive value. Here we review the benefits and limitations of the most popular coma scoring systems. The methods include search of Medline, databases, and manual review of article bibliographies. Few of the many available coma scales have gained widespread approval and popularity. The best known and widely accepted scale is the Glasgow Coma Scale (GCS). The Reaction Level Scale (RLS85) has utility and proven benefit, but little acceptance outside of Scandinavia. The newer Full Outline of UnResponsiveness (FOUR) score provides an attractive replacement for all patients with fluctuating levels of consciousness and is gradually gaining wide acceptance.

Keywords Coma · GCS · FOUR score · RLS85 · Consciousness

Introduction

The cornerstone of care of the neurological patient is the physical exam. In acute and severe neurological illness, serial examinations are the simplest, least expensive, and often most reliable tool to assess the clinical course. These bedside evaluations have given us the ability to synthesize

data into norms for patients across a wide spectrum of diseases and severity with a good degree of accuracy. One of the disease states that has benefited from this acquisition and application is coma.

Coma is an alteration of consciousness that represents the final pathway of various pathophysiological processes in disease states (trauma, toxic-metabolic, vascular, neoplastic, seizures) ultimately leading to derangement in cerebral function manifesting as decreased arousal and awareness [1]. Coma represents a medical emergency and its management hinges on the understanding of its etiology, and managing complications that may arise. Perhaps most important is being able to identify in a timely fashion those patients with a reversible cause who may benefit from aggressive treatment and have the potential for a favorable outcome.

Integral to evaluating patients are methods of rapidly and reliably assessing their current status, monitoring for and predicting the potential for worsening of their condition, and inferring potential outcomes to discuss with these patients and their families. Rapidly performed objective scoring systems have been developed for numerous disease states and outcome variables serve this purpose. The utility of a simple assessment scale for the evaluation of disorders of consciousness serves to facilitate communication between healthcare providers, allows rapid and therefore frequent accurate bedside clinical assessments, and benefits research by providing standardized assessment scales.

The ideal scoring system for evaluating coma should be easy to administer and score, be applicable to the greatest number of patients, able to accurately assess level of consciousness, identify rapidly deteriorating patients, and predict morbidity and mortality. Inter-rater reliability for coma scales is generally presented as a weighted κ value. The κ value is a measure of agreement between two or more observers accounting for variability based on chance

J. Kornbluth · A. Bhardwaj (✉)
Department of Neurology, Tufts University School of Medicine,
Box 314, 800 Washington Street, Boston, MA 02111, USA
e-mail: abhardwaj1@tuftsmedicalcenter.org

alone [2]. A value of 1 indicates perfect agreement and a value of 0 indicates agreement by chance alone. Among the scales developed for assessing patients with altered consciousness are the Glasgow Coma Scale (GCS), the Reaction level Scale (RLS85), and the Full Outline of UnResponsiveness (FOUR). This manuscript reviews these scales used clinically in detail including their validity and limitations in assessing and monitoring disorders of consciousness. Ancillary physiologic monitoring to augment these commonly used coma scales at the bedside is not in the scope of this review.

GCS

History and Validation

The most widely used and most studied coma score to date is the GCS, first described by Teasdale and Jennett in 1974 and revised in 1976 with the addition of a sixth point in the motor response subscale for “withdrawal from painful stimulus” [3, 4]. The GCS was initially intended to assess level of consciousness after traumatic brain injury (TBI) in a Neurosurgical Intensive Care Unit in order to facilitate communication among staff regarding patient status [3]. Since then it has become the gold standard against which newer scales are compared and used widely by Emergency Department (ED) staff, Medical and Surgical ICU’s as well as by pre-hospital providers. Moving beyond the developers’ original indication, the GCS has been validated as a useful tool for prediction of outcome after intracranial hemorrhage [5], subarachnoid hemorrhage (SAH) [6], poisonings including ethanol [7–9], neurodegenerative diseases [10], drowning [11, 12], cardiac arrest [13–16], recently tuberculous meningitis [17], and prediction of death in palliative care [18]. The GCS is typically praised for its ease of use, and universal approval. The GCS calculates a score from 3 to 15, with 3 being the worst, allowing for 120 different combinations grouped into 12 possible scores. Points are awarded for eye opening, motor response, and verbal response (Table 1).

As of 2005, more than 4,500 publications made reference to the GCS [19]. The ease and appeal of the GCS has lead it to be incorporated into many trauma scoring systems, namely the Revised Trauma Score (RTS) [20], the APACHE II [21], the Simplified Acute Physiology Score (SAPS), and SAPSII [22], the Circulation, Respiration, Abdomen, Motor, Speech scale (CRAMS) [23], the Traumatic Injury Scoring System (TRISS) [24], and A Severity Characterization of Trauma (ASCOT) scale [25].

Teasdale and Jennett reported significant consistency between raters of the GCS [3]. Validation of inter-rater reliability was initially described in terms of a

“disagreement rate” [3, 26] that has been defined as “low” when they fall between 0 and 0.299 and high when between 0.3 and 0.5 [27]. According to this definition, Teasdale et al. found that 7 neurosurgeons displayed low disagreement when assessing 12 ICU patients with the GCS (disagreement rates for eye opening = 0.143, verbal response = 0.0054, and motor response = 0.109) [26]. The simplicity of the GCS and its rapidity of administration have made it popular among emergency medical system (EMS) providers for triage and to guide therapies, and has become a component of many algorithms for out-of-hospital triage to trauma centers [23, 28–33]. Consistent with prior studies [34], Menegazzi et al. [30] reported inter-rater reliability of paramedics and ED physicians with $\kappa = 0.48$ for subjectively severe alterations of consciousness, with least discrepancy among evaluators in the mild group ($\kappa = 0.85$). Although these statistics were deemed to show “significant inter-rater reliability” by the authors, their results have less agreement than other studies and when compared to other scoring systems [35, 36]. Furthermore a prospective study reported in 2003 of pre-hospital and ED staff assessment of patients with TBI showed GCS scores to be on average 2 points lower by ED scoring; this discrepancy was not statistically significant and correlation in scoring was shown to be strong and independent of time between scoring [28]. Further validation has been shown among nurses in the ED and ICU’s [27, 34, 35]. As anticipated, Rowley and Fielding observed that more experienced providers consistently made more accurate measurements of the GCS [37]. Interestingly they pointed out that error rates were the highest in the patients with “intermediate” levels of consciousness where accurate monitoring for change of clinical status is critical.

GCS as been utilized as a grading system in other specific brain injury paradigms, for example to assess outcomes following SAH [38]. A simplified grading system based on GCS that compresses the 15-point GCS into five grades that are comparable with other grading systems for SAH, namely the Hunt Hess Scale (HHS) and World Federation of Neurological Surgeons scale (WFNSS) [39]. In a study comprising of 291 consecutive patients, GCS was the best predictor of discharge Glasgow outcome score and had the best inter-rater reliability ($\kappa = 0.46$) compared to HHS ($\kappa = 0.41$) and WFNSS ($\kappa = 0.27$) [39]. Other investigators have suggested a statistically validated scale for patients with poor grade SAH that combines the HHS and the GCS to enhance outcome prediction [40].

Limitations

Based on initial validation studies, the GCS is assumed to be accurate and reproducible; however, many newer studies have found only moderate degrees of inter-rater

Table 1 Scoring on various coma scales

Glasgow Coma Scale	Reaction Level Scale	FOUR Score
Eye opening	1 = Alert; no delay in response	Eye response
4 = Spontaneous	2 = Drowsy or confused; responsive to strong stimulation	4 = Eyelids open or opened, tracking, or blinking to command
3 = To speech	3 = Very drowsy or confused; responsive to strong stimulation	3 = Eyelids open but not tracking
2 = To pain	4 = Unconscious; localizes but does not ward off pain	2 = Eyelids closed but open to loud voice
1 = None	5 = Unconscious; withdrawing movements on pain stimulation	1 = Eyelids closed but open to pain
Best verbal response	6 = Unconscious; stereotype flexion movements on pain stimulation	0 = Eyelids remain closed with pain
5 = Oriented	7 = Unconscious; stereotype extension movements on pain stimulation	Motor response
4 = Confused conversation	8 = Unconscious; no response to pain stimulation	4 = Thumbs up, fist, or peace sign
3 = Inappropriate words		3 = Localizing to pain
2 = Incomprehensible sounds		2 = Flexion response to pain
1 = None		1 = Extension response to pain
Best motor response		0 = No response to pain or generalized myoclonus status
5 = Obeys commands		Brainstem reflexes
5 = Localizes to pain		4 = Pupil and corneal reflexes present
4 = Withdrawal (normal flexion)		3 = One pupil wide and fixed
3 = Abnormal flexion (decorticate)		2 = Pupil or corneal reflexes absent
2 = Extension (decerebrate)		1 = Pupil and corneal reflexes absent
1 = None		0 = Absent pupil, corneal, and cough reflex
		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

agreement at best [35, 37, 41]. Gill et al. [35] reported a study incorporating 116 patients evaluated by two emergency physicians. The weighted κ for total GCS was 0.4 (component scores ranged from 0.48 for verbal response, 0.63 for motor response and 0.72 for eye opening). A later study ($n = 120$ patients) reported a κ for total GCS = 0.32 (component scores were 0.44 for verbal response, 0.54 for motor response and 0.71 for eye opening) [42]. A study by Holdgate et al. [41] demonstrated significant variability, described as >2 points in sum GCS scores, between senior ED physicians and nurses in a tertiary hospital. Unlike the study by Gill et al., [42] there was more reliability in the total GCS and verbal component (weighted $\kappa > 0.75$), compared to both the motor and eye components (weighted κ 0.4–0.75 for both). In this study there was also no difference in GCS calculations across the range of possible scores.

A major concern for providers about the GCS is its inability to accurately assess intubated patients and difficulty in assessing aphasic patients due to the requirement

of a verbal component [43, 44]. This is especially important for the use of the GCS in predictive scores such as the APACHE score. In fact, the GCS has been proven to be the most powerful predictive component of the APACHE score [45, 46]. The methods of dealing with the inability to accurately score the verbal subscore vary significantly and include omitting the subscore, substituting median values, and substituting “1” for the verbal subscore [43]. The presence of aphasia has been specifically addressed by Prasad and Menon [47]. They assessed three methods of dealing with the verbal scoring of an aphasic patient: (1) eliminating the verbal component, (2) pseudoscore with ‘1’, and (3) median value substitution of the other components. Their data agrees with others’ data [48–54], that the motor and eye components alone can accurately substitute for a complete GCS where the eye and motor subscale had 87% accuracy compared to 88% for the model with eye, motor, and verbal scale.

Previous work has been reported using linear regression analysis to predict GCS verbal scores from motor and eye

subscores [51, 52]. These studies have proved the extrapolation of the verbal subscore from the eye and motor components to be applicable to alterations of consciousness due to traumatic and non-traumatic etiologies [53]. Rutledge's original study [52] with mathematical derivations was further validated by Meredith et al. [53]. In these studies the actual subjects were not intubated, a necessary weakness in the testing of their theorems in order to compare extrapolated scores to actual measured scores.

The GCS has been criticized from a purely mathematical point of view by Bhatti and Kapur [55]. They note that there is a mathematical skew and calculation bias towards the motor score as a result of assigning four possible scores to eye responses, five to verbal and six to motor responses. Critics of the GCS have acknowledged this and faulted it for being overly complex and have proposed a modified GCS, with focus on the motor component which would improve ease of use and potentially predictive power when there is inability to obtain a verbal subscore [8, 54]. Kelly et al. tested a simple AVPU responsiveness scale which scores mental status as "alert," "responsive to verbal stimulation," "responsive to painful stimulation," and "unresponsive" in comparison to the GCS. Their data suggested that ward and ICU nursing staff found the AVPU responsiveness scale easier to use than the GCS [8]. This prospective study found that the most frequent difficulty formulating an accurate GCS score occurred for patients with alcohol intoxication in whom there was difficulty in compliance with the assessment or had slurred speech. Responses to this difficulty were to either omit the verbal subscore or to score it inappropriately low. As stated above, Gill et al. [35] showed only modest inter-rater agreement between ED staff assessing GCS whereas disagreement occurred most frequently on assessment of the verbal component rather than the eye or motor response.

In response to the limitations of the verbal component of the GCS a 3-point Simplified Motor Score (defined as obeys commands = 2; localizes pain = 1; withdrawal to pain or worse = 0) was recently derived from the motor component of the GCS and was found to have similar performance for predicting outcomes after TBI when compared with the GCS [56]. Healy et al. [48] have also shown that a motor-only score is a more powerful predictor of mortality and advocated for inclusion of mathematically transformed motor score in lieu of a complete GCS in predictive models. However, they noted the limitation of this motor-only scale for use after pharmacologic paralysis. Further utility of the motor score alone has been proven in terms of its accuracy and reproducibility in pre-hospital triage [49, 50], and its predictive value in TBI [51].

Teoh et al. [57] reported GCS scores with possible permutations (i.e., a single numerical sum score can possibly be made up of more than one subscore profile). They found that specifically the scores of 7, 9, 11, and 14 had the most variability in terms of predictive value and that this can have significant implications for disease severity calculation for the APACHE and SAPS. Furthermore they suggested that for these scores a GCS profile where subscores are specifically addressed should be reported. The inconsistent way in which the GCS is used may make it less suitable for multicenter trials and calls into question its validity between centers for its incorporation into predictive algorithms.

Lastly, the GCS has major limitations for its utility in children particularly those less than 3 years of age and prior to acquisition of language. A pediatric GCS that retains the three major components has been developed for pediatric population (total minimum score of 3 and maximum of 15) [58]. However, there is paucity of studies investigating inter-rater agreement and variability utilizing the pediatric GCS. One study ($n = 73$ with 104 observations) reports inter-observer reliability to be moderate to good for all components, with the grimace score better than the verbal score [59].

RLS85

History and Validation

The RLS85 was a successor to the original RLS82 validated in a pilot study in 1982 [60], and subsequently revised in 1985 [61]. The RLS85 was formulated to specifically overcome the shortcomings of the GCS in scoring intubated patients and patients with swollen eyelids precluding the ability to open one's eyes [61]. Following its validation, it is now used almost exclusively in Sweden. The RLS85 is a hierarchically organized scale with eight possible outcomes or "reaction levels". RLS1-8 is assigned after attempting to arouse a patient to a stable level of consciousness via increasingly the stimuli: light touch/talking, then shaking/shouting, then noxious stimuli [61]. The responses are graded stepwise in terms of depth of coma, with '1' being complete consciousness to '8' for unconscious and unresponsive. The authors indicate that RLS1 is equivalent to "alert", RLS2 is "drowsy or confused", RLS3 is "very drowsy or confused" and RLS4-8 signifies "unconscious" (Tables 1, 3).

The basis for developing a scale predicated on overall responsiveness as opposed to the GCS's multi-scale system was that the latter is inherently prone to modification due to untestable features and multiple possible algorithms used to arrive at the same overall score [62]. As shown above,

the RLS85 is based on the same objective assessments of the GCS but separate modalities are combined into a unified stepwise scale. The RLS85 has been shown to be useful in evaluating mild to severe TBI, cerebrovascular disease and brain neoplasms, drug overdose, cardiovascular disease, and gastrointestinal disease [62–65]. The initial validation study for the RLS85 was tested in a neurosurgical setting as part of a multisite trial throughout Sweden. The inter-rater agreement for the RLS85 regardless of etiology of alteration of consciousness was calculated with a $\kappa = 0.69$ (ranged 0.6–0.82 across the four sites). Sub-score analysis revealed the most discrepancy between nursing assistants, registered nurses, and physicians with regard to recognizing stereotyped flexion ($\kappa = 0.55$) and purposeful withdrawal ($\kappa = 0.51$) [60].

Tesseris et al. [64] compared the GCS, the RLS85 as well as the Edinburgh-2 Coma Scale, modified (E2CS(M)). In their study comparing the evaluation of 46 patients via the RLS85 and GCS, they determined that the RLS85 was reliable and reproducible ($\kappa = 0.633$), and superior to the GCS which only demonstrated a $\kappa = 0.35$ for same patients. They also noted that RLS85 scores were equally reliable regardless of etiology of alteration of consciousness. Based on the fact that the RLS85 was showing superior inter-rater reliability Walther et al. [65] examined the relationship between the GCS and the RLS85 with regards to outcome prediction using APACHE II scores with a far larger cohort ($n = 534$) hypothesizing that the RLS85 scores would improve outcome prediction. They found good agreement between the GCS-based and the RLS-based APACHE II scores as well as improved discrimination of the APACHE II model when cerebral responsiveness was assessed with the RLS instead of the GCS.

Limitations

The principle limitation of the RLS85 is that it is used almost exclusively in Sweden. The Scandinavian Societies of Intensive Care, Anesthesiology, and Neurosurgery have recommended replacement of the GCS with the RLS85 in that country's hospitals [66]. Additionally, research generated from Scandinavian hospitals has been utilizing the RLS85 in lieu of the GCS since shortly after its inception [67–69]. To our knowledge there is only one recent published report utilizing the RLS85 as an objective assessment instead of the GCS outside of Sweden [70].

The learning curve in using the RLS85 may be slower (less steep) than for other scales. The authors suggest a total of 2–3 h of training time, including watching an instructional video and practice on at least 10 patients [61]. This is a significant difference than the instruction on using the FOUR score where raters watched a 20-min video with an accompanying handout [36].

Innsbruck Coma Scale

History and Validation

In use since 1981 and first published in 1991 the Innsbruck Coma Scale (ICS) was developed for the specific assessment of trauma victims and is almost used exclusively at the University Hospital, Innsbruck, Austria [71]. The total score is analogous to that of the GCS in that there are 8 separate categories, 7 of which are rated from 0 to 3, and one rated 0–2, with 0 being the worst score in each category. This gives a total of 147,356 separate score sums grouped into 23 possible scores. It is similar to the GCS, but excludes verbal response thus overcoming the limitation in intubated, aphasic, and aphonic patients. The ICS also measures pupillary size and reaction, movement and position of the eyes and oral automatisms (Table 2).

Aside from its internal use at the University Hospital, the ICS was validated as a predictive scale and as Benzer et al. [71] points out it fulfills two important criteria for a predictive coma scale: simple rapid assessment, and high accuracy in prediction of non-survival. Diringier and Edwards [72] examined the ICS and its sub-scores using Cronbach's α to test reliability, finding that the oral automatisms subscore was disproportionately unreliable compared to the other subscores. The reliability coefficient for the total ICS = 0.78, contrasted to the GCS in the same study where the reliability coefficient = 0.77. The ICS was modified to be calculated without oral automatisms and evaluated for its predictive power; ICS-Modified was shown to have a better predictive power than the standard ICS.

Limitations

The ICS has not gained widespread popularity and very little is published about it. There is no published study assessing the inter-rater reliability as with most coma scales other than the validation of the modified ICS as above. One criticism of the ICS is that the score rates dilated fixed pupils of greater severity (lower score) than midposition nonreactive pupils [73]. Thus, a patient with brain death having midposition pupils would achieve a better score than one who is not brain dead with dilated pupils. Further work needs to be done validating the ICS before it can gain widespread acceptance.

FOUR

History and Validation

A significant drawback of many coma scales is the inability to accurately and reliably assess brainstem function

Table 2 Innsbruck Coma Scale

<i>Eye opening</i>	
3	Spontaneous
2	To acoustic stimuli
1	To painful stimuli
0	None
<i>Reaction to acoustic stimuli</i>	
3	Turning towards stimuli
2	Better-than-extension movements
1	Extension movements
0	None
<i>Reaction to pain</i>	
3	Defensive movements
2	Better-than-extension movements
1	Extension movements
0	None
<i>Body posture</i>	
3	Normal
2	Better-than-extension movements
1	Extension movements
0	None
<i>Pupil size</i>	
3	Normal
2	Narrow
1	Dilated
0	Completely dilated
<i>Pupil response to light</i>	
3	Sufficient
2	Reduced
1	Minimum
0	None
<i>Position and movements of the eyeballs</i>	
3	Fixing of the eyes
2	Sway of the eyeballs
1	Divergent
0	Divergent fixed
<i>Oral automatisms</i>	
2	Spontaneous
1	To external stimuli
0	None

specifically. Moulton and Pennycook [74] have previously shown that the GCS inadequately assesses the cough reflex regardless of level consciousness score. Many coma scales that include indicators of brainstem function have been proposed to supplant the GCS including the Bouzarth Coma Scale for TBI which incorporates brainstem reflexes [75], the Maryland Coma Scale which includes pupils, caloric reflexes, and grimace [76], the Comprehensive Level of Consciousness Scale which includes pupillary reflexes, eye position, opening, and movement [77], the

Clinical Neurologic Assessment Tool which included chewing and yawning [78], and the Glasgow-Liege scale which combined the GCS with five brainstem reflexes: pupillary, fronto-orbicular, oculocardiac, horizontal, and vertical occulocephalics [79]. These scales generally have been more complex than the GCS and none have gained widespread use.

Recognizing the shortcoming of the GCS, Widjicks et al. [36] published a new scoring system in 2005, the FOUR score (Table 1). Widjicks et al. [80] had first proposed a scoring system for measuring impaired consciousness that overcame some of the shortcomings of the GCS; critiquing that it lacks the ability to identify subtle changes in alteration of consciousness. This system added a continuous performance test where a patient is asked to raise his hand every time he hears a certain letter in a standardized sentence to monitor alertness, and a “hand position test” (“thumbs up → fist → victory sign”) to measure praxis. The FOUR score assesses four variables: eye response, motor response, brainstem reflexes, and respiration pattern (Table 1). The acronym additionally reflects the number of categories and the maximum number of potential points in each category, exemplifying its simplicity and attempt at universal appeal. Each category is awarded 0–4 points with 0 being the worst. There are 625 possible scoring combinations grouped into 17 possible scores from 0 to 16. According to authors the FOUR score is superior to the GCS in that it can account for the intubated patient without substitute scores and identify a locked-in state, and detect the presence of a vegetative state [36]. This is particularly poignant given recent evidence that locked-in syndromes are under-recognized early on [81] and increased patient awareness of the syndrome given recent media attention to the story of Jean-Dominique Bauby [82] who suffered a stroke resulting in a locked-in syndrome.

The administration of the FOUR has a few specific advantages over utilizing the GCS. The FOUR adds to the eye opening of the GCS by testing eye tracking, thus incorporating midbrain and pontine functions. Adding to the motor score of the GCS is an extension of Wijdick’s [80] earlier work incorporating hand gestures into the evaluation. This alternative to the verbal score allows for testing of afferent language processing and remains testable regardless of endotracheal intubation, aphasia, aphonia, or trauma to the vocal apparatus. The bulk of the motor score is similar to the GCS except that no difference is delineated between flexor posturing and normal flexion to pain. Additionally, no motor response and myoclonic status epilepticus are scored equally, reflecting the associated poor outcome after anoxic brain injury [83]. Specific testing of brainstem reflexes via pupillary, corneal, and cough reflexes further allows the practitioner to localize

lesions and track progression of cerebral injury specifically by addressing unilateral fixed mydriasis, a sign alerting to uncal herniation. The authors advocate for utilizing saline drops as opposed to gauze or swabs in testing corneal reflexes in an effort to minimize corneal trauma. The final category of the FOUR evaluates patterns of respiration. This assesses respirations as spontaneous regular or irregular, Cheyne-Stokes, intubated but independently breathing above the ventilator, or absent. If all four categories are graded at zero the authors advocate to consider brain death testing.

The FOUR was initially validated as a prospective study in ICU patients ($n = 120$) assessing inter-rater reliability among neuroscience nurses, neurology residents, and neurointensivists, and compared to the same for the GCS; this is the largest validation study of a new coma score to date [36]. This initial validation proved that inter-rater reliability between the FOUR score and the GCS is equivalent (weighted κ of 0.82; 95% CI). The agreement was highest among neurology residents and lowest among neuroscience nurses [36]. The FOUR score has recently been validated through an observational study comparing experienced neuroscience nurses, inexperienced neuroscience nurses, and non neuroscience-trained nurses [84] as well as separate studies evaluating ED staff [85] and Medical ICU staff [86]. Comparing critical care nursing staff, the overall weighted κ score was 0.85 for the FOUR score and 0.83 for the GCS.

In comparing non-neurology trained ED physicians, ED residents and ED nurses, each of these groups was assigned to evaluate 69 patients presenting with neurologic symptoms and data was compared to the same cohort measuring the more standard GCS. Inter-rater reliability for considered excellent (weighted κ scores of 0.88 for the FOUR score and 0.86 for the GCS) [82]. In a study comprising of medical ICU staff (nurses, residents, fellows, and intensivists) there was excellent inter-rater agreement of FOUR score (weighted κ values of 0.97–0.99 for the FOUR score and 0.96–0.99 for the GCS) [86]. These two studies validate the inter-rater reliability of the FOUR score and substantiate its use by healthcare professionals not specifically trained to recognize neurologic signs. Two recent studies have validated the use of the FOUR score outside of the Mayo Clinic. Weiss and colleagues at the hôpital de la Pitié-Salpêtrière, Paris, France translated the FOUR score into French and assessed its utility and validity in a neurologic critical care unit. A total of 176 FOUR scores were calculated by two neurologists, four experienced nurses and five inexperienced nurses. This was consistent with prior validation studies (weighted κ was 0.86 for the FOUR score and 0.85 for the GCS) [87]. The French team highlighted that the FOUR score was useful, easy to learn and easy to perform. A subsequent study by Akavipat further

validated and endorsed the use of the FOUR score specifically for neurosurgical patients [88]. 100 patients were evaluated to assess inter-rater reliability of each the FOUR score and the GCS, as well as to compare scoring between the two. Patients were assessed by expert clinicians, novice clinicians, experienced nurses, and inexperienced nurses. The exact definition of ‘expert’, ‘novice’, and ‘clinician’ was not reported. Weighted κ scores among the types of rater varied from 0.93 to 0.99 for the FOUR score and 0.9–0.97 for the GCS. The poorest agreement was in the brainstem subscale. The author points out potential pitfalls of brainstem scoring that may be variable among examiners including the loudness voice, intensity of applied noxious stimuli, potential pupil size estimation, and fluctuations between ratings. In addition to validation, the practicality of adopting a new coma scale was assessed via a questionnaire.

Limitations

As the FOUR score is a relatively new scoring model there are relatively few criticisms beyond those highlighted by the original study group. One potential flaw to date is that up until recently the FOUR score had only been validated at the Mayo Clinic (Table 3). As Bellomo et al. [89] recently pointed out, caution is warranted for single-center trials. The studies above have attempted to overcome this limitation but further experience and validation outside the initial study institution is required before the FOUR score can be universally endorsed and utilized as a standard for research protocols. In order to compete with the GCS in its widespread use, further work needs to be done regarding the predictive value of the FOUR score as well.

Conclusions

Critically ill neurologic and neurosurgical patients require frequent and accurate assessment of their neurologic status. Of the many coma scales that have been proposed for this purpose, few of them have gained widespread approval and popularity. The primary purpose of the coma scale remains to facilitate communication of reliable and rapid patient assessment for decision making, prognostication, patient “hand-off” and to be a measured variable in research protocols. The best known and widely accepted scale is clearly the GCS. The RLS85 has utility and proven benefit, but little acceptance outside of Scandinavia. The newer FOUR score is slowly gaining acceptance outside of the Mayo Clinic hospitals and makes up for limitations of prior scales and provides an attractive replacement for all patients with fluctuating levels of consciousness for assessment of the neuraxis (Tables 1, 3).

Table 3 A comparison of coma scales

Scale	Benefits	Limitations
GCS	Widespread use and understanding Rapid assessment possible Proven use in pre-hospital, ED, and inpatient settings Validation and incorporation into predictive scores (i.e., APACHE, TRISS)	Differential scoring unfairly weights Motor subscale Unable to accurately assess intubated/aphasic/aphonic patients, or patients with swollen or injured eyes Not all subscores are equally validated No direct brainstem evaluation Unable to accurately identify locked-in syndrome Known to be unreliable in inexperienced scorers Varied inter-rater reliability
RLS85	Rapid, stepwise assessment without separate subscores Can accurately assess intubated/aphasic/aphonic patients Able to assess patients with swollen/injured eyes	Limited experience or use outside of Scandinavia Numeric values not necessarily separated by steps of equal value Unable to accurately identify locked-in syndrome extensive training necessary to master accurate scoring
ICS	Simple and rapid assessment Specific predictor of mortality	Almost no use outside of University Hospital, Innsbruck Oral automatisms are disproportionately unreliable as a subscore for predictive value
FOUR Score	Rapid assessment possible Direct observation of brainstem function Motor and Eye responses incorporate language testing Gestures allow for testing responsiveness in intubated/aphasic/aphonic patients Easy to learn and apply Able to identify patients with “locked-in syndrome” Consistent inter-rater reliability	Limited use outside of Mayo Clinic system, but gaining wider acceptance

Future Directions

A rapid bedside evaluation is the basis for all of the popular coma scales. Supportive data in the form of neuromonitoring including biomarkers, physiologic, electrophysiologic, and radiographic information are readily available and as the assays of these become faster and more reliable they will become attractive compliments to established coma scales.

References

1. Plum and Posner's diagnosis of stupor and coma. 4th ed. USA: Oxford University Press; 2007.
2. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;37:360–3.
3. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet.* 1974;2:81–4.
4. Teasdale GM, Jennett B. Assessment and prognosis of coma after head injury. *Acta Neurochir.* 1976;34:45–55.
5. Tuhim S, Dambrosia JM, Price TR, Price TR, Mohr JP, Wolf PA, Hier DB, Kase CS. Intracerebral hemorrhage: external validation and extension of a model for prediction of 30-day survival. *Ann Neurol.* 1991;29:658–63.
6. Hirai S, Ono J, Yamaura A. Evaluation of consciousness by the Glasgow Coma Scale in aneurysmal subarachnoid hemorrhage. Appraisal using logistic regression analysis. *Cerebrovasc Dis.* 1997;99:S42–3.
7. Chan B, Gaudry P, Grattan-Smith TM, McNeil R. The use of Glasgow Coma Scale in poisoning. *J Emerg Med.* 1993; 11:579–82.
8. Kelly CA, Upex A, Bateman DN. Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the Glasgow Coma Scale. *Ann Emerg Med.* 2004;44:108–13.
9. Sam KG, Kondabolu K, Pati D, Kamath A, Kumar PG, Rao PG. Poisoning severity score, APACHE II and GCS: effective clinical indices for estimating severity and predicting outcome of acute organophosphorus and carbamate poisoning. *J Forensic Leg Med.* 2009;16:239–47.
10. Benesch CG, McDaniel KD, Cox C, Hamill RW. End-stage Alzheimer's disease. Glasgow Coma Scale and the neurologic examination. *Arch Neurol.* 1993;50:1309–15.
11. Bratton SL, Jardine DS, Murray JP. Serial neurologic examinations after near drowning and outcome. *Arch Pediatr Adolesc Med.* 1994;148:167–70.
12. Ballesteros MA, Gutiérrez-Cuadra M, Muñoz P, Miñambres E. Prognostic factors and outcome after drowning in an adult population. *Acta Anaesthesiol Scand.* 2009;53:935–40.
13. Mullie A, Verstringe P, Buylaert W, Houbrechts H, Michem N, Delooz H, Verbruggen H, Van den Broeck L, Corne L, Lauwaert D, et al. Predictive value of Glasgow coma score for awakening after out-of-hospital cardiac arrest. Cerebral Resuscitation Study Group of the Belgian Society for Intensive Care. *Lancet.* 1988;1(8578):137–40.
14. Schefold JC, Storm C, Krüger A, Ploner CJ, Hasper D. The Glasgow Coma Score is a predictor of good outcome in cardiac arrest patients treated with therapeutic hypothermia. *Resuscitation.* 2009;80:658–61.

15. Edgren E, Hedstrand U, Kelsey S, Sutton-Tyrrell K, Safar P. Assessment of neurological prognosis in comatose survivors of cardiac arrest. BRCT I Study Group. *Lancet*. 1994;343(8905):1055–9.
16. Geocadin RG, Buitrago MM, Torbey MT, Chandra-Strobos N, Williams MA, Kaplan PW. Neurologic prognosis and withdrawal of life support after resuscitation from cardiac arrest. *Neurology*. 2006;67:105–8.
17. Chou CH, Lin GM, Ku CH, Chang FY. Comparison of the APACHE II, GCS and MRC scores in predicting outcomes in patients with tuberculous meningitis. *Int J Tuberc Lung Dis*. 2010;14:86–92.
18. Coleman NL, Brieva JL, Crowfoot E. Prediction of death after withdrawal of life-sustaining treatments. *Crit Care Resusc*. 2008;10:278–84.
19. Laureys S, Piret S, Ledoux D. Quantifying consciousness. *Lancet*. 2005;4:789–90.
20. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flannagan ME. A revision of the trauma score. *J Trauma*. 1989;29:623–9.
21. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–29.
22. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993;270:2957–63.
23. Gormican SP. CRAMS scale: field triage of trauma victims. *Ann Emerg Med*. 1982;11:132–5.
24. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. *J Trauma*. 1987;27:370–8.
25. Champion HR, Copes WS, Sacco WJ, Frey CF, Holcroft JW, Hoyt DB, Weigelt JA. Improved predictions from a severity characterization of trauma (ASCOT) over Trauma and Injury Severity Score (TRISS): results of an independent evaluation. *J Trauma*. 1996;40:42–8. discussion 48–9.
26. Teasdale G, Knill-Jones R, van der Sande J. Observer variability in assessing impaired consciousness and coma. *J Neurol Neurosurg Psychiatry*. 1978;41:603–10.
27. Juarez VJ, Lyons M. Interrater reliability of the Glasgow Coma Scale. *J Neurosci Nurs*. 1995;27:283–6.
28. Bazarjian JJ, Eirich MA, Salhanick SD. The relationship between pre-hospital and emergency department Glasgow coma scale scores. *Brain Inj*. 2003;17:553–60.
29. Norwood SH, McAuley CE, Berne JD, Vallina VL, Creath RG, McLarty J. A prehospital glasgow coma scale score $<$ or $=14$ accurately predicts the need for full trauma team activation and patient hospitalization after motor vehicle collisions. *J Trauma*. 2002;53:503–7.
30. 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.
31. Davis DP, Vadeboncoeur TF, Ochs M, Poste JC, Vilke GM, Hoyt DB. The association between field Glasgow Coma Scale score and outcome in patients undergoing paramedic rapid sequence intubation. *J Emerg Med*. 2005;29:391–7.
32. 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.
33. Campbell JE. Basic trauma life support for paramedics and other advanced providers. 4th ed. Upper Saddle River, NJ: Prentice Hall; 1999.
34. Fielding K, Rowley G. Reliability of assessments by skilled observers using the Glasgow Coma Scale. *Aust J Adv Nurs*. 1990;7:13–7.
35. Gill MR, Reiley DG, Green SM. Interrater reliability of Glasgow Coma Scale scores in the emergency department. *Ann Emerg Med*. 2004;43:215–23.
36. Wijdicks EFM, Bamlet WR, Maramotom BV, Manno EM, McClelland RL. Validation of a new coma scale: the FOUR score. *Ann Neurol*. 2005;58:585–93.
37. Rowley G, Fielding K. Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet*. 1991;337(8740):535–8.
38. Rosen DS, Macdonald. Subarachnoid hemorrhage grading scales: a systematic review. *Neurocrit Care*. 2005;2:110–8.
39. Oshiro EM, Walter KA, Piantadosi S, Witham TF, Tamargo RJ. A new subarachnoid hemorrhage grading system based on the Glasgow Coma Scale: a comparison with the Hunt and Hess and World Federation of Neurological Surgeons Scales in a clinical series. *Neurosurgery*. 1997;41:140–7.
40. Starke RM, Komotar RJ, Kim GH, Kellner CP, Otten ML, Han DK, Schmidt MJ, Sciacca RR, Mayer SA, Connolly SE. Evaluation of a revised Glasgow Coma Scale in predicting long-term outcome of poor grade aneurysmal subarachnoid hemorrhage patients. *J Clin Neurosci*. 2009;16:894–9.
41. 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.
42. Gill MR, Martens K, Lynch EL, Salih A, Green SM. Interrater reliability of 3 simplified neurologic scales applied to adults presenting to the emergency department with altered levels of consciousness. *Ann Emerg Med*. 2007;47:403–7.
43. 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.
44. Marion DW, Carlier PM. Problems with initial Glasgow Coma Scale assessment caused by prehospital treatment of patients with head injuries: results of a national survey. *J Trauma*. 1994;36:89–95.
45. Niskanen M, Kari A, Nikki P, Iisalo E, Kaukinen L, Rauhala V, Saarela E, Halinen M. Acute physiology and chronic health evaluation (APACHE II) and Glasgow coma scores as predictors of outcome from intensive care after cardiac arrest. *Crit Care Med*. 1991;19:1465–73.
46. Bastos PG, Sun X, Wagner DP, Wu AW, Knaus WA. Glasgow Coma Scale score in the evaluation of outcome in the intensive care unit: findings from the Acute Physiology and Chronic Health Evaluation III Study. *Crit Care Med*. 1993;21:1459–65.
47. Prasad K, Menon GR. Comparison of the three strategies of verbal scoring of the Glasgow Coma Scale in patients with stroke. *Cerebrovasc Dis*. 1998;8:79–85.
48. Healey K, Osler TM, Rogers FB, Healey MA, Glance LG, Kilgo PD, Shackford SR, Meredith JW. Improving the Glasgow Coma Scale score: motor score alone is a better predictor. *J Trauma*. 2003;54:671–8.
49. Ross SE, Leipold C, Terregino C, O'Malley KF. Efficacy of the motor component of the Glasgow Coma Scale in trauma triage. *J Trauma*. 1998;45:42–4.
50. Meredith W, Rutledge R, Hansen AR, Oller DW, Thomason M, Cunningham P, Baker CC. Field triage of trauma patients based upon the ability to follow commands: a study in 29, 573 injured patients. *J Trauma*. 1995;38:129–35.
51. Al-Salamah MA, McDowell I, Stiell IG, Stiell IG, Wells GA, Perry J, Al-Sultan M, Nesbitt L, OPALS Study Group. Initial emergency department trauma scores from the OPALS study: the case for the motor score in blunt trauma. *Acad Emerg Med*. 2004;11:834–42.

52. Rutledge R, Lentz CW, Fakhry S, Hunt J. Appropriate use of the Glasgow Coma Scale in intubated patients: a linear regression prediction of the Glasgow Verbal Score from the Glasgow Eye and Motor Scores. *J Trauma*. 1996;41:514.
53. Meredith W, Rutledge R, Fakhry SM, Emery S, Kromhout-Schiro S. The conundrum of the Glasgow Coma Scale in intubated patients: a linear regression prediction of the Glasgow verbal score from the Glasgow eye and motor scores. *J Trauma*. 1998;44:839–44.
54. Gill M, Windemuth R, Steele R, Green SM. 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.
55. Bhatti GB, Kapoor N. The Glasgow Coma Scale: a mathematical critique. *Acta Neurochir*. 1993;120:132–5.
56. 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.
57. Teoh LS, Gowardman JR, Larsen PD, Green R, Galletly DC. Glasgow Coma Scale: variation in mortality among permutations of specific total scores. *Intensive Care Med*. 2000;26(2):157–61.
58. Morray JP, Tyler DC, Jones TK, Stuntz JT, Lemire RJ. Coma scale for use in brain-injured children. *Crit Care Med*. 1984;12:1018–20.
59. Tatman A, Warren A, Williams A, Powell JE, Whitehouse W. Development of a modified paediatric coma scale in intensive care clinical practice. *Arch Dis Child*. 1997;77:519–21.
60. Starmark JE, Holmgren E, Stålharumar D. Estimation of reaction levels in coma patients. Edinburgh: ICRAN; 1982.
61. Starmark JE, Stålhammar D, Holmgren E. The Reaction Level Scale (RLS85). Manual and guidelines. *Acta Neurochir*. 1988;91:12–20.
62. Stålhammar D, Starmark JE, Holmgren E, Eriksson N, Nordström CH, Fedders O. Assessment of responsiveness in acute cerebral disorders. A multicentre study on the reaction level scale (RLS 85). *Acta Neurochir*. 1988;90:73–80.
63. Johnstone AJ, Lohlnun JC, Miller JD, McIntosh CA, Gregori A, Brown R, Jones PA, Anderson SI, Tocher JL. A comparison of the Glasgow Coma Scale and the Swedish Reaction Level Scale. *Brain Inj*. 1993;7:501–6.
64. Tesseris J, Pantazidis N, Routsis C, Fragoulakis D. A comparative study of the Reaction Level Scale (RLS85) with Glasgow Coma Scale (GCS) and Edinburgh-2 Coma Scale (modified) (E2CS(M)). *Acta Neurochir*. 1991;110:65–76.
65. Walther SM, Jonasson U, Gill H. Comparison of the Glasgow Coma Scale and the Reaction Level Scale for assessment of cerebral responsiveness in the critically ill. *Intensive Care Med*. 2003;29:933–8.
66. Starmark JE, Stålhammar D. A new reaction level scale is recommended in Sweden. *Lakartidningen*. 1990;87:1466–9.
67. Frowein RA, Terhaag D, auf der Haar K, Richard KE, Firsching R. Rehabilitation after severe head injury. *Acta Neurochir Suppl*. 1992;55:72–4.
68. Rosén H, Rosengren L, Herlitz J, Blomstrand C. Increased serum levels of the S-100 protein are associated with hypoxic brain damage after cardiac arrest. *Stroke*. 1998;29:473–7.
69. Starmark JE, Heath A. Severity grading in self-poisoning. *Hum Toxicol*. 1988;7:551–5.
70. Costa B, Candido J, Girbal A, Afonso L, Ribeiro J, Gonçalves V, Luis V, Antunes J. Coma of vascular aetiology: evaluation and prognosis. *Neurol Res*. 1992;14:100–2.
71. Benzer A, Mitterschiffthaler G, Marosi M, Luef G, Pühringer F, De La Renotiere K, Lehner H, Schmutzhard E. Prediction of non-survival after trauma: Innsbruck Coma Scale. *Lancet*. 1991;338(8773):977–8.
72. Diringner MN, Edwards DF. Does modification of the Innsbruck and the Glasgow Coma Scales improve their ability to predict functional outcome? *Arch Neurol*. 1997;54(5):606–11.
73. De'Clari F. Innsbruck coma scale. *Lancet*. 1991;338(8781):1537.
74. Moulton C, Pennycook AG. Relation between Glasgow coma score and cough reflex. *Lancet*. 1994;343(8908):1261–2.
75. Bouzarth WF. Neurosurgical watch sheet for craniocerebral trauma. *J Trauma*. 1968;8:29–31.
76. Salzman M, Schepp RS, Ducker TB. Calculated recovery rates in severe head trauma. *Neurosurgery*. 1981;8:301–8.
77. Stanczak DE, White JG, Gouvieux WD, Moehle KA, Daniel M, Novack T, Long CJ. 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;60:955–60.
78. Crosby L, Parsons LC. Clinical neurologic assessment tool: development and testing of an instrument to index neurologic status. *Heart Lung*. 1989;18:121–9.
79. Born JD. The Glasgow-Liège Scale. Prognostic value and evolution of motor response and brain stem reflexes after severe head injury. *Acta Neurochir*. 1988;91:1–11.
80. Wijdicks EF, Kokmen E, O'Brien PC. Measurement of impaired consciousness in the neurological intensive care unit: a new test. *J Neurol Neurosurg Psychiatry*. 1998;64:117–9.
81. Laureys S, Pellas F, Van Eeckhout P, Ghorbel S, Schnakers C, Perrin F, Berré J, Faymonville ME, Pantke KH, Damas F, Lamy M, Moonen G, Goldman S. The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *Prog Brain Res*. 2005;150:495–511.
82. Bauby J-D. *Le Scaphandre et le Papillon* (French original of *The Diving Bell and the Butterfly*). New York: French & European Pubns; 2007.
83. Wijdicks EF, Hijdra A, Young GB, Bassetti CL, Wiebe S. Quality Standards Subcommittee of the American Academy of Neurology. Practice parameter: prediction of outcome in comatose survivors after cardiopulmonary resuscitation (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*. 2006;67:203–10.
84. Wolf CA, Wijdicks EF, Bamlet WR, McClelland RL. Further validation of the FOUR score coma scale by intensive care nurses. *Mayo Clin Proc*. 2007;82:435–8.
85. Stead LG, Wijdicks EF, Bhagra A, Kashyap R, Bellolio MF, Nash DL, Enduri S, Scheers R, William B. Validation of a new coma scale, the FOUR score, in the emergency department. *Neurocrit Care*. 2009;10:50–4.
86. Iyer VN, Mandrekar JN, Danielson RD, Zubkov AY, Elmer JL, Wijdicks EF. Validity of the FOUR score coma scale in the medical intensive care unit. *Mayo Clin Proc*. 2009;84:694–701.
87. Weiss N, Mutlu G, Essardy F, Nacabal C, Sauves C, Bally C, Houbert M, Lecorre C, Germack V, Demeret S, Pierrot-Deseilligny C, Bolgert F. The French version of the FOUR score: a new coma score. *Rev Neurol*. 2009;165:796–802.
88. Akavipat P. Endorsement of the FOUR score for consciousness assessment in neurosurgical patients. *Neurol Med Chir*. 2009;49:565–71.
89. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single center trials. *Crit Care Med*. 2009;37:314–9.