Interrater Reliability and Predictive Validity of the FOUR Score Coma Scale in a Pediatric Population

CE

Jennifer Cohen

ABSTRACT

The Glasgow Coma Scale (GCS) was developed in 1974 to objectively describe neurological status and predict outcome in neuroscience patients. Through the years, the GCS has become the gold standard for coma assessment. Despite its widespread use, the GCS has many limitations that are well documented in the literature. The Full Outline of Unresponsiveness (FOUR) score is a new coma scale that was recently developed and validated in adults as a proposed replacement for the GCS. The purpose of this study was to compare the interrater reliability and predictive validity of the FOUR score and the GCS in pediatric patients. The interrater reliability for the GCS was good (k_w = .738), and that for the FOUR score was excellent (k_w = .951). Outcome prediction analysis showed that the FOUR score and the GCS are both able to predict in-hospital morbidity and poor outcome at the end of hospitalization. The results from this pediatric study were consistent with the adult studies which suggest that the FOUR score is a reliable and valid tool for use in a wide variety of neuroscience patients.

The Glasgow Coma Scale (GCS) has been the gold standard for assessing a patient's level of consciousness (LOC) and acute changes in neurological status since it was developed in 1974. It is the universally accepted measure of assessment and documentation of neurological findings and is used as a predictor of functional patient outcome. Despite its widespread use, the GCS has some significant limitations, including variations in interrater reliability and predictive validity (Rowley & Fielding, 1991). Over the years, many attempts have been made to improve the GCS to create a more comprehensive yet easy-to-use tool. These modified or alternative coma scales have rarely been published or accepted into practice outside of their originating countries or institutions. Researchers at the Mayo Clinic recently developed the Full Outline of Unresponsiveness (FOUR) score (Wijdicks et al., 2005). Several studies have validated the FOUR score's use in adults and suggest that the FOUR score is a good alternative to the GCS. There have been no studies to validate its use in pediatrics. The purpose of this study was to compare the interrater reliability and predictive validity of the FOUR score and the GCS in pediatrics patients.

Questions or comments about this article may be directed to Jennifer Cohen, MSN CNS RN CCRN CNRN, at jcohen@choc.org. She is the clinical nurse specialist of the Pediatric Intensive Care Unit, CHOC Children's Hospital, Orange, CA.

Background

The adult human brain accounts for approximately 2% of the total body weight, yet it consumes more than 20% of the oxygen used at rest. The brain is, therefore, one of the most metabolically active organs in the body. However, it does not have an effective way to store oxygen and glucose and is dependent upon a constant supply to meet its needs. Even a brief interruption in oxygenation and blood flow to the brain can result in acute changes in LOC. Alterations in LOC can also be caused by chemical or metabolic toxins, infection, or increased intracranial pressure. Whatever the cause, a decreased LOC is characteristic of nervous system dysfunction and is associated with increased morbidity and mortality. Despite advances in technology and state-of-the-art monitoring devices, a thorough clinical assessment is still the key to identifying subtle changes in a patient's neurological status and is fundamental to the management of neuroscience patients. To provide quality patient care, the bedside nurse must therefore be able to accurately and consistently assess and communicate these changes.

One of the first standardized neurological assessment tools, the Ommaya "vital sign" card (1966), described a patient's LOC, motor activity, pupillary status, corneal reflexes, blood pressure, pulse, respirations, respiratory effort, and rectal temperature. Teasdale and Jennett (1974) expanded on the work of Ommaya and others and created the Coma Index, later known as the GCS. The GCS was developed to standardize the description of coma depth and severity and to improve communication between healthcare providers with different levels of experience and expertise (Teasdale & Jennett, 1974). The GCS independently measures three aspects of consciousness: eye opening, motor response, and verbal response. In the original 1974 publication, the GCS was introduced as an unnumbered system. In 1976, a modified GCS assigned a numeric value to the responses in each of the three categories (Teasdale & Jennett, 1976). Assigning a total sum to the GCS eye + motor + verbal was never the intention of the originators, yet GCS summation has become standard practice. The GCS sum scores of 3, 8, and 15 now have universal meaning to healthcare providers and have become instrumental in determining a patient's prognosis. The GCS has been incorporated into several intensive care and trauma scoring systems including the Trauma and Injury Severity Score, the Acute Physiology and Chronic Health Evaluation, and the Pediatric Risk of Mortality score, to assess risk of in-hospital mortality and predict future disability (Rutledge, Lentz, Fakhry, & Hunt, 1996). The GCS is also used internationally in both prehospital and hospital environments to predict the likelihood of morbidity, mortality, and long-term outcomes in acutely ill neuroscience patients. The accuracy of the GCS is, therefore, crucial.

To be an effective tool, a coma scale must be practical for use in a wide variety of settings and by healthcare providers with diverse experience (Teasdale & Jennett, 1974). Rowley and Fielding (1991) determined that, by using the GCS, experienced healthcare providers were able to assess a patient with high levels of accuracy and reliability, whereas inexperienced raters demonstrated significant variability, particularly in the intermediate levels of consciousness. Gill, Reiley, and Green (2004) reported only moderate degrees of interrater reliability for the GCS and its three components in the emergency department setting. This is consistent with the results of a national survey that revealed variation among trauma centers' calculations of the GCS (Buechler et al. 1998). Riechers et al. (2005) indicated that many physicians were unable to accurately identify the tool's subcategories and the specific scoring of each. Physicians with advanced certification and training performed significantly better than those with less experience. This wide variation in GCS scoring within organizations among providers with varying levels of expertise and between healthcare organizations is a concern because it questions the accuracy of the databases and outcome research that are based upon the GCS (Ingram, 1994).

Both the reliability and predictive validity of the GCS have been questioned in the literature because individual components of the tool may be impossible to score in many patients.

In addition to reliability issues, the predictive validity of the GCS has also been questioned in the literature. Teasdale and Jennett (1974) indicated the likelihood of shortcomings in the GCS, noting that, for various reasons, the individual components may be impossible to score. First, the ability to accurately assess the GCS motor score is often impacted by the administration of sedatives or neuromuscular blocking agents and the presence of confounders such as spinal cord injury. The GCS is skewed toward motor assessment, with a maximum of 6 points. This is weighted more heavily than the eye (4 points) and verbal (5 points) categories. This will affect the ability to assign an accurate GCS to patients who are receiving medications or who have injuries that interfere with motor assessment.

The eye opening subscore is often difficult to obtain due to the administration of sedatives and paralytics as well as facial swelling associated with traumatic injury or surgery. In 2003, the predictive power of the GCS was compared with that of its individual components using a large trauma data set (n = 204,181); this study suggested that the eye opening category added negligible value to the tool and should be removed from the GCS (Healey et al., 2003).

Another widely documented limitation to the GCS is the verbal category and its usefulness in assessing critically ill, often intubated, neuroscience patients. The presence of an endotracheal tube eliminates the ability to assess a patient's verbal response. The need for intubation and mechanical ventilation, however, may suggest brainstem involvement and can be an important factor in the assessment of coma severity. It is often these patients who are at highest risk for in-patient mortality and decreased functional outcome. Methods used to overcome this limitation have included: assigning all intubated patients the lowest verbal score (1 point), pseudoscoring that predicts the patient's ability to verbalize, or simply designating a non-numeric score of "T" to all intubated patients

(Rutledge et al., 1996). However, these methods of assessing the GCS verbal response in the presence of an endotracheal tube have not been validated and are not universally used and will as a result affect the data reported by different organizations that are used to predict patient outcome.

The verbal component of GCS also poses a challenge in the assessment of infants and preverbal children. To account for the developmental variations in the verbal, as well as the motor and cognitive ability of infants and children, a modified pediatric GCS was developed. There have been many attempts through the years to develop and validate additional pediatric coma scales including the Starship Infant Neurological Assessment Tool (Birse, 2006), the CHOP Infant Coma Scale (Durham et al., 2000), and the Adelaide Pediatric Coma Scale (Reilly, Simpson, Sprod, & Thomas, 2004). Despite these attempts, there is currently no agreed-upon "gold standard" for pediatrics (Tatman et al., 1997).

Additional key clinical indicators that are essential to a neurological assessment which are not assessed when using the GCS include pupillary asymmetry, abnormalities in ocular movement, and changes in breathing patterns. Subtle changes in these brainstem reflexes and cranial nerve functions may indicate brainstem injury and neurological impairment (Youman, 1996). Teasdale and Jennett (1974) recognized that the test of brainstem function can be useful in the diagnosis of stupor and coma and yet chose not to incorporate brainstem assessment into the GCS.

To address the many limitations to the GCS, researchers at the Mayo Clinic designed the FOUR score coma scale as a proposed alternative. The FOUR score assigns a value of 0 to 4 to each of four functional categories: eye response, motor response, brainstem reflexes, and respiration. In each of these categories, a score of 0 indicates nonfunctioning status, and a score of 4 represents normal functioning (Wijdicks et al., 2005). In contrast to the GCS, the FOUR score eye response category assesses eye tracking in addition to eye opening. The motor assessment includes response to pain, ability to follow simple commands, and the presence of generalized myoclonus status epilepticus. The FOUR score eliminated verbal response from the scale due to the documented limitations of verbal score in GCS. Instead, a brainstem reflex category was created to assess the function of the mesencephalon, pons, and medulla. Lower brainstem function is evaluated using the respiration category to identify irregular breathing patterns, including Cheyne Stokes respirations.

In the original study published by the Mayo Clinic, the FOUR score was prospectively tested on 120 adult intensive care unit (ICU) patients and compared with the GCS by interrater pairs of neuroscience nurses, residents, and neurointensivists (Wijdicks et al., 2005). Results indicated that the overall interrater reliability of the FOUR score was excellent and that predictive validity was good and similar to the GCS. Intensive care staff indicated that the tool was easy to use and adopt. Developers of the tool suggest that the FOUR score may provide greater neurological detail than the GCS due to its ability to evaluate brain stem reflexes and to recognize changes in breathing patterns and stages of herniation (Wijdicks et al., 2005). Subsequent studies have validated the FOUR score's use by ICU nurses with varying levels of experience (Wolf et al., 2007) and in the emergency department by nonneurology staff (Stead et al., 2009). Researchers at the Mayo Clinic believe that the FOUR score gives a comprehensive and accurate snapshot of a patient's neurological status and are hopeful that it will eventually replace the GCS. The FOUR score has the potential to become the most effective means of evaluating neuroscience patients and assessing their risk for severe injury and mortality.

Methods

The proposed study was submitted to the institutional review board at the 232-bed freestanding children's hospital in Southern California where the study was conducted. Following institutional review board approval, a convenience sample of 60 neuroscience patients, ages 2 to 18 years, was recruited from the pediatric intensive care unit. To assess the use of the GCS and the FOUR score on a variety of patients, the participants were assigned by the principal investigator to one of four categories upon admission: alert (n = 44), drowsy (n = 10), stuporous (n = 3), or comatose (n = 3), as previously defined by Ropper (1986). Patients from each of these categories were included in this study. Patients receiving sedatives and/or neuromuscular blocking agents were excluded. Over the course of 1 year, 60 patients ranging in age from 2 to 18 years with a mean age of 9.4 years (SD = 5.12 years) were enrolled in the study. Patient diagnoses included brain tumors (n = 25), hydrocephalus (n = 7), traumatic brain injury (n = 5), spinal surgery (n = 5), seizure disorder (n = 4), arteriovenous malformations (AVMs) (n = 4), cranial remodeling (n = 3), moyamoya disease (n = 2), encephalitis (n = 2), subdural hematoma (n = 2), and near drowning (n = 1).

Thirty-five pediatric critical care nurse raters participated in this study. The nurses ranged in age from 23 to 60 years, with a range of less than 1 year to 40 years of clinical experience in pediatric critical care nursing. Of the 35, 12 were educated at the associate degree level, 20 at the baccalaureate level, and 3 were master's prepared. Ten of the nurses held national certification in pediatric critical care nursing. Each nurse signed an agreement to participate, which included a statement that they agreed not to discuss their assessments with the other nurses in the study.

Nurse raters all participated in education, provided by the principal investigator, related to the use of the GCS and the FOUR score assessment tool. Raters were given a copy of the pediatric GCS tool and the FOUR score instructional card (see Figure 1) for reference during the assessment of patients. Each patient was assessed by two nurse raters using both the GCS and the FOUR score within 30 min of admission or at the time that a neurosurgical or neurology consult was ordered. The nurses completed the assessment at the same point in time and recorded their scores on separate rater scorecards. The timing was critical to minimize the chance that there was a change in the patients' neurological status between assessments. To decrease bias, the order in which each nurse rater performed the assessment (FOUR or GCS) was randomized, and the raters were blinded to each other's scores.

In-hospital mortality and clinical diagnosis of brain death were documented by the principal investigator, and patient morbidity was assessed upon discharge from the hospital using the Modified Rankin Scale. The Modified Rankin Scale is one of the most commonly used tools to assess functional neurological outcome in adults (Wilson et al., 2005). Despite lack of formal validation in children, this scale has been used in several pediatric studies to allow for direct comparison with adult studies. Outcome was assessed, using the Modified Rankin, according to the following: 0 = no symptoms; 1 = no significant disability despite symptoms, able to carry out all usual duties and activities; 2 = slight disability, unable to carry out all previous activities, but able to look after own affairs without assistance; 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 and attention; and 6 = dead (Lindley et al., 1994).

Results

The interrater reliability of the GCS and the FOUR score was evaluated using the weighted kappa (k_w) coefficient (Sims & Wright, 2005). A k_w statistic of .4 or less is considered poor, values between .4 and .6 are considered fair to moderate, those between .6 and .8 suggest good interobserver agreement, and values greater than .8 suggest excellent agreement (Landis & Koch, 1977). The nurse rater agreement is shown in Table 1. The overall reliability was good for the total GCS score (k_w = .74, 95% confidence intervals [CI] = .59–.87) and excellent for the total FOUR score (k_w = .95, 95% CI = .91–.99). Logistics regression analyses were performed to determine the ability of the two scales to predict in-hospital morbidity and poor outcome (Rankin = 3–6) at the end of hospitalization. Of the 60 patients in the study, 4 (6.7%) died and 23 (38.3%), which includes the 4 who died, had a poor outcome. The average ratings of the two raters on each scale were used in the regression analyses. Table 2 shows the odds ratios, CIs, and percent of cases correctly classified for the total FOUR and GCS scores. The odds ratios for the FOUR are somewhat lower than those for the GCS. In previous studies, lower odds ratios have been related to positive predictive value for a higher chance of a positive outcome with increased total score values (Stead et al., 2009; Wijdicks et al., 2005). The proportion of cases correctly classified for both in-hospital mortality and poor outcome was similar for both the FOUR and the GCS.

Receiver operator characteristic curve analysis compares the true rate of the outcome (sensitivity) with the false rate (100 − specificity) and was used in this study to compare the ability of the FOUR score and the GCS to predict in-hospital mortality and poor outcome (Rankin = 3–6) at the end of hospitalization. The closer the area under the curve is to 1.00, the greater the likelihood that the tool is able to identify the outcome state (Schonjans, 2008). The average ratings of the two raters were used for the receiver operator characteristic curve analyses. For in-hospital mortality, the area under the curve for the FOUR was .81 (95% CI = .69–.90) and that for the GCS was .77 (95% CI = .64–.87). The differences in areas under the curve were not statistically significant. Sensitivity and specificity were maximized for both the FOUR and the GCS at a score of 13 (FOUR sensitivity = .75, specificity = .86; GCS sensitivity = .75, specificity = .79). For poor outcome at the end of hospitalization, the area under the curve for the FOUR was .78 (95% CI = .65–.88) and that for the GCS was .76 (95% CI = .64–.86). Differences

FIGURE 1 The FOUR Score Instructional Card

MAYO CLINIC

FOUR Score

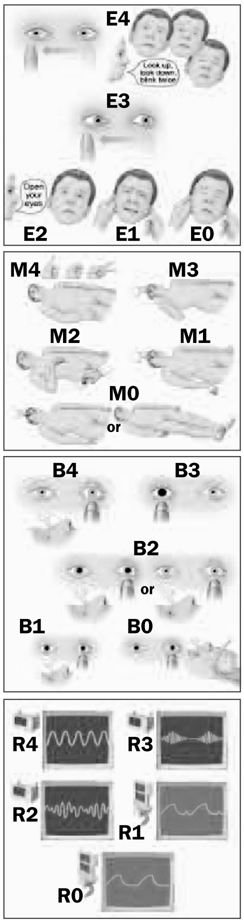
Eye Response
- 4 Eyelids open or opened, tracking or blinking to command
- 3 Eyelids open but not tracking
- 2 Eyelids closed but opens to loud voice
- 1 Eyelids closed but opens to pain
- 0 Eyelids remain closed with pain

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

Brainstem Reflexes
- 4 Pupil and corneal reflexes present
- 3 One pupil wide and fixed
- 2 Pupil or corneal reflexes absent
- 1 Pupil andcorneal reflexes absent
- 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 pattern
- 1 Breathes above ventilator rate
- 0 Breathes at ventilator rate or apnea

Wijdicks EFM, Bamlet WR, Maramattom BV, Manno EM, McClelland RL. Validation of a new Coma Scale: the FOUR score. Annals of Neurology, 2005, 58:585-593



Instructions for the Assessment of the Individual Categories of the FOUR Score

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 (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 (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 (R)
Determine spontaneous breathing patternin a nonintubated patient, and grade simply as regular **R4**, Cheyne-Stokes **R3**, or irregular **R2** 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 PaCO2 within normal limits. A standard apnea (oxygen-diffusion) test may be needed when patient breathes at ventilator rate **R0**.

Note. FOUR = Full Outline of Unresponsiveness. Copyright by the Mayo Foundation for Medical Education and Research. Reproduced with permission. All rights reserved. Published originally in Wijdicks et al. (2005).

were not significant. Sensitivity and specificity were maximized for the FOUR at a score of 15 (sensitivity = .61, specificity = .95) and that for the GCS was at a score of 14 (sensitivity = .57, specificity = .92).

At the completion of the data collection process, each nurse participant completed a 5-point Likert survey to assess nurse satisfaction with the FOUR score. All 35 nurse raters agreed or strongly agreed (Likert score of 4 or 5) with the following statements that addressed their perception of the clinical usefulness of the FOUR score: (a) The FOUR score is clinically relevant and easy to use, (b) the FOUR score is a good alternative to the

TABLE 1. Weighted Kappa Values, Standard Error, and 95% Confidence Intervals for Interrater Agreement on the FOUR Score and Glasgow Coma Scale (*N* = 60 Patients)

	FOUR Score					Glasgow Coma Scale			
	Eye	Motor	Brainstem	Respiration	Total	Eye	Motor	Verbal	Total
Weighted κ	.98	.86	1.00	1.00	.95	.62	.71	.60	.74
SE	.03	.08	.00	.00	.02	.13	.09	.10	.08
95% CI	.93–1.00	.70–1.00	1.00–1.00	1.00–1.00	.91–.99	.36–.88	.53–.89	. 40–.80	.59–.87

Note. FOUR = Full Outline of Unresponsiveness.

GCS, and (c) the FOUR score is an assessment tool that I would use if it becomes generally accepted.

Discussion

Data from this pediatric study are consistent with that obtained in the adult studies. Interrater reliability among nurse raters using the FOUR score was better than their interrater reliability on the GCS. Surprisingly, there was perfect agreement ($k_w = 1.0$, with a 95% CI = 1.0 and *SE* = 0) in scoring patients in the FOUR score's brainstem and respiration categories. This is particularly interesting considering that the nurse raters were provided with minimal education on the use of the FOUR score and had no previous experience assessing patients using these categories. A total GCS of <8 was documented by both raters on five patients. Two additional patients were assigned a GCS of <8 by one rater, but the second rater assessed the GCS to be >8. An interesting finding was that, although the nurse raters disagreed on the score using the GCS in both of these cases, the same nurse raters were in agreement using the FOUR score for those patients.

The nurse raters who participated in this study consistently agreed that the FOUR score is easy to remember, with the lowest score in every category assigned a 0 and the highest score in each category assigned a 4 versus the GCS which has a scoring system that is more difficult to remember. The GCS requires some experience with the tool to know that the lowest score in each category is 1 and that the high score varies between the categories. Of note, one patient in this study was incorrectly assigned a GCS of 0 by two separate experienced nurses without knowledge of each other's score. The same nurses correctly assigned that patient a FOUR score of 0.

A limitation of this study was the small number of patients within the stuporous and comatose categories. Future studies should focus on assessing the validity of the FOUR score in critically ill pediatric patients. Additional studies need to be conducted on the use of the FOUR score to add to the existing body of knowledge and provide further evidence to support its use in assessing neuroscience patients of all ages.

Summary

This is the first study to evaluate this promising new neurological assessment tool for its application to the pediatric patient population. Study results demonstrated that the FOUR score and the GCS

TABLE 2. Odds Ratios, Confidence Intervals, and Percent of Cases Correctly Classified for Total FOUR and Total GCS Scores for In-Hospital Death and Poor Outcome

	In-hospital Death, 4 Events (*N* = 60)		Poor Outcome End of Hospitalization, Modified Rankin Scale 3–6, 23 events (*N* = 60)	
	OR (95% CI)	Cases Correctly Classified (%)	OR (95% CI)	Cases Correctly Classified (%)
FOUR score total[a]	.68 (.50–.93)	95.0	.31 (.13–.72)	78.3
GCS score total[a]	.77 (.62–.95)	95.0	.58 (.40–.85)	76.7

Note. GCS = Glasgow Coma Scale; FOUR = Full Outline of Unresponsiveness.
[a]Averaged score of the two raters.

were comparable in predicting outcome in this population. The high level of agreement between nurse raters using the FOUR score suggests that the tool is consistent and reliable and that nurses with differing levels of experience and expertise are more likely to correctly assess the patient and assign the same score using the FOUR score than the GCS. On the basis of the findings of this study and in consideration of the results from the previous studies, the FOUR score appears to be an easier tool to use and provides a more comprehensive neurological assessment. An assessment tool that is easy to use, is able to predict outcome, and has proven interrater reliability may significantly affect the assessment of critically ill children with neurological conditions. The widespread adoption of such a tool may enhance the ability to accurately predict survivability, impacting the treatment and management of these patients and their families.

References

Birse, J. (2006). Pilot testing of the starship infant neurological assessment tool. *Journal of Neuroscience Nursing*, *38*(4), 206–212.

Buechler, C., Blostein, P., Koestner, A., Hurt, K., Schaars, M., & McKernan, J. (1998). Variation among trauma centers' calculation of Glasgow Coma Scale score: Results of a national survey. *Journal of Trauma, Injury, Infection, and Critical Care*, *45*(3), 429–432.

Durham, S. R., Clancy, R. R., Leuthardt, E., Sun, P., Kamerling, S., Dominguez, T., et al. (2000). CHOP Infant Coma Scale ("Infant Face Scale"): A novel coma scale for children less than two years of age. *Journal of Neurotrauma*, *17*(9), 729–737.

Gill, M., Reiley, D., & Green, S. (2004). Interrater reliability of Glasgow Coma Scale scores in the emergency department. *Annals of Emergency Medicine*, *43*(2), 215–223.

Healey, C., Osler, T., Rogers, F., Healey, M., Glance, L., Kilgo, P., et al. (2003). Improving the Glasgow Coma Scale score: Motor score alone is a better predictor. *Journal of Trauma, Injury, Infection, and Critical Care*, *54*(4), 671–680.

Ingram, N. (1994). Knowledge and level of consciousness: Application to nursing practice. *Journal of Advanced Nursing*, *20*(5), 881–884.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Lindley, R., Waddell, F., Livingstone, M., Sandercock, P., Dennis, M., & Slattery, J., et al. (1994). Can simple questions assess outcomes after stroke? *Cerebrovascular Disease*, *4*, 314–324.

Reilly, P. L., Simpson, D. A., Sprod, R., & Thomas, L. (2004). Assessing the conscious level in infants and young children: A paediatric version of the Glasgow Coma Scale. *Child's Nervous System*, *4*(1), 30–33.

Riechers, R., Ramage, A., Brown, W., Kalehua, A., Rhee, P., Ecklund, J., et al. (2005). Physician knowledge of the Glasgow Coma Scale. *Journal of Neurotrauma*, *22*(11), 1327–1334.

Ropper, A. H. (1986). Lateral displacement of the brain and level of consciousness in patients with an acute hemispheral mass. *New England Journal of Medicine*, *314*, 953–958.

Rowley, G., & Fielding, K. (1991). Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet*, *337*, 535–538.

Rutledge, R., Lentz, C., Fakhry, S., & Hunt, J. (1996). 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. *Journal of Trauma, Injury, Infection, and Critical Care*, *41*(3), 514–522.

Schonjans, F. (2008). *Receiver operating characteristic curve analysis* [Software]. Mariakerke, Belgium: MedCalc.

Sims, J., & Wright, C. (2005). The Kappa statistics in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257–268.

Stead, L., Wijdicks, E., Bhagra, A., Kashyap, R., Bellolio, M., Nash, D., et al. (2009). Validation of a new coma scale, the FOUR score, in the emergency department. *Neurocritical Care*, *10*, 50–54.

Tatman, A., Warren, A., Williams, A., Powell, J., & Whitehouse, W. (1997). Development of a modified paediatric coma scale in intensive care clinical practice. *Archives of Disease in Childhood*, *77*, 519–521.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet*, *2*, 81–84.

Teasdale, G., & Jennett, B. (1976). Assessment and prognosis of coma after head injury. *Acta Neurochirgica*, *34*, 45–55.

Wijdicks, E., Bamlet, W., Maramattom, B., Manno, E., & McClelland, R. (2005). Validation of a new coma scale: The FOUR score. *Annals of Neurology*, *58*, 585–593.

Wilson J., Hareendran, A., Grant, M., Hendry, A., Potter, J., Bone, I., et al. (2005). Reliability of the modified Rankin scale across multiple raters. *Stroke*, *36*, 777.

Wolf, C., Wijdicks, E., Bamlet, W., & McClelland R. (2007). Further validation of the FOUR score coma scale by intensive care nurses. *Mayo Clinic Proceedings*, *82*(4), 435–438.

Youman, J. (1996). *Neurological surgery: A comprehensive reference guide to the diagnosis and management of neurosurgical problems* (4th ed.). Philadelphia: WB Saunders.